# Springer Undergraduate Mathematics Series

## Advisory Board



For other titles published in this series, go to
www.springer.com/series/3423

Andrew Pressley

# Elementary Differential Geometry

Second Edition

Springer

Andrew Pressley
Department of Mathematics
King's College London
Strand, London WC2R 2LS
United Kingdom
andrew.pressley@.kcl.ac.uk

Springer Undergraduate Mathematics Series ISSN 1615-2085
ISBN 978-1-84882-890-2 e-ISBN 978-1-84882-891-9
DOI 10.1007/978-1-84882-891-9
Springer London Dordrecht Heidelberg New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009942256

Mathematics Subject Classification (2000): 53-01, 53A04, 53A05, 53A35



*Cover design*: Deblik

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

The Differential Geometry in the title of this book is the study of the geometry of curves and surfaces in three-dimensional space using calculus techniques. This topic contains some of the most beautiful results in Mathematics, and yet most of them can be understood without extensive background knowledge. Thus, for virtually all of this book, the only pre-requisites are a good working knowledge of Calculus (including partial differentiation), Vectors and Linear Algebra (including matrices and determinants).

Many of the results about curves and surfaces that we shall discuss are prototypes of more general results that apply in higher-dimensional situations. For example, the Gauss–Bonnet theorem, treated in Chapter 11, is the prototype of a large number of results that relate 'local' and 'global' properties of geometric objects. The study of such relationships formed one of the major themes of 20th century Mathematics.

We want to emphasise, however, that the *methods* used in this book are *not* necessarily those which generalise to higher-dimensional situations. (For readers in the know, there is, for example, no mention of 'connections' in the remainder of this book.) Rather, we have tried at all times to use the simplest approach that will yield the desired results. Not only does this keep the pre-requisites to an absolute minimum, it also enables us to avoid some of the conceptual difficulties often encountered in the study of Differential Geometry in higher dimensions. We hope that this approach will make this beautiful subject accessible to a wider audience.

It is a cliché, but true nevertheless, that Mathematics can be learned only by doing it, and not just by reading about it. Accordingly, the book contains over 200 exercises. Readers should attempt as many of these as their stamina permits. Full solutions to all the exercises are given at the end of the book, but these should be consulted only after the reader has obtained his or her own solution, or in case of desperation. We have tried to minimise the number of instances of the latter by including hints to many of the less routine exercises.

## Preface to the Second Edition

Few books get smaller when their second edition appears, and this is not one of those few. The largest addition is a new chapter devoted to hyperbolic (or non-Euclidean) geometry. Quite reasonably, most elementary treatments of this subject mimic Euclid's axiomatic treatment of ordinary plane geometry. A much quicker route to the main results is available, however, once the basics of the differential geometry of surfaces have been established, and it seemed a pity not to take advantage of it.

The other two most significant changes were suggested by commentators on the first edition. One was to treat the tangent plane more geometrically - this then allows one to define things like the first and second fundamental forms and the Weingarten map as geometric objects (rather than just as matrices). The second was to make use of parallel transport. I only partly agreed with this suggestion as I wanted to preserve the elementary nature of the book, but in this edition I have given a definition of parallel transport and related it to geodesics and Gaussian curvature. (However, for the experts reading this, I have stopped just short of introducing connections.)

There are many other smaller changes that are too numerous to list, but perhaps I should mention new sections on map-colouring (as an application of Gauss-Bonnet), and a self-contained treatment of spherical geometry. Apart from its intrinsic interest, spherical geometry provides the simplest 'non-Euclidean' geometry and it is in many respects analogous to its hyperbolic cousin. I have also corrected a number of errors in the first edition that were spotted either by me or by correspondents (mostly the latter).

For teachers thinking about using this book, I would suggest that there are now three routes through it that can be travelled in a single semester, terminating with *one* of chapters 11, 12 or 13, and taking in along the way the necessary basic material from chapters 1–10. For example, the new section on spherical geometry might be covered only if the final destination is hyperbolic geometry.

As in the first edition, solutions to all the exercises are provided at the end of the book. This feature was almost universally approved of by student commentators, and almost as universally disapproved of by teachers! Being one myself, I do understand the teachers' point of view, and to address it

I have devised a large number of new exercises that will be accessible online to all users of the book, together with a solutions manual for teachers, at www.springer.com.

I would like to thank all those who sent comments on the first edition, from beginning students through to experts - you know who you are! Even if I did not act on all your suggestions, I took them all seriously, and I hope that readers of this second edition will agree with me that the changes that resulted make the book more useful and more enjoyable (and not just longer).

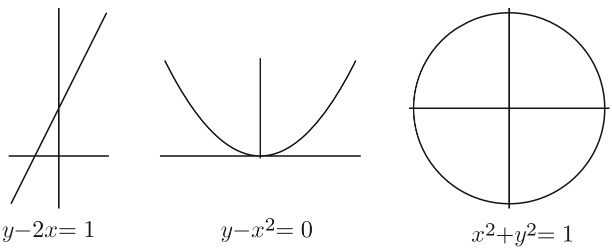
# Contents

# 1
# Curves in the plane and in space

In this chapter, we discuss two mathematical formulations of the intuitive notion of a curve. The precise relation between them turns out to be quite subtle, so we begin by giving some examples of curves of each type and practical ways of passing between them.

## 1.1 What is a curve?

If asked to give an example of a curve, you might give a straight line, say $y - 2x = 1$ (even though this is not 'curved'!), or a circle, say $x^2 + y^2 = 1$, or perhaps a parabola, say $y - x^2 = 0$.





Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_1,

All of these curves are described by means of their Cartesian equation

$$f(x, y) = c,$$

where $f$ is a function of $x$ and $y$ and $c$ is a constant. From this point of view, a curve is a set of points, namely

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}. \tag{1.1}$$

These examples are all curves in the plane $\mathbb{R}^2$, but we can also consider curves in $\mathbb{R}^3$ – for example, the $x$-axis in $\mathbb{R}^3$ is the straight line given by

$$y = 0, \quad z = 0,$$

and more generally a curve in $\mathbb{R}^3$ might be defined by a pair of equations

$$f_1(x, y, z) = c_1, \qquad f_2(x, y, z) = c_2.$$

Curves of this kind are called *level curves*, the idea being that the curve in Eq. 1.1, for example, is the set of points $(x, y)$ in the plane at which the quantity $f(x, y)$ reaches the 'level' $c$.

But there is another way to think about curves which turns out to be more useful in many situations. For this, a curve is viewed as the path traced out by a moving point. Thus, if $\boldsymbol{\gamma}(t)$ is the position of the point at time $t$, the curve is described by a function $\boldsymbol{\gamma}$ of a scalar parameter $t$ with vector values (in $\mathbb{R}^2$ for a plane curve, in $\mathbb{R}^3$ for a curve in space). We use this idea to give our first formal definition of a curve in $\mathbb{R}^n$ (we shall be interested only in the cases $n = 2$ or 3, but it is convenient to treat both cases simultaneously).

## Definition 1.1.1

A *parametrized curve* in $\mathbb{R}^n$ is a map $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^n$, for some $\alpha, \beta$ with $-\infty \leq \alpha < \beta \leq \infty$.

The symbol $(\alpha, \beta)$ denotes the open interval

$$(\alpha, \beta) = \{t \in \mathbb{R} \mid \alpha < t < \beta\}.$$

A parametrized curve, whose image is contained in a level curve $\mathcal{C}$, is called a *parametrization* of (part of) $\mathcal{C}$. The following examples illustrate how to pass from level curves to parametrized curves and back again in practice.

## Example 1.1.2

Let us find a parametrization $\boldsymbol{\gamma}(t)$ of the parabola $y = x^2$. If $\boldsymbol{\gamma}(t) = (\gamma_1(t), \gamma_2(t))$, the components $\gamma_1$ and $\gamma_2$ of $\boldsymbol{\gamma}$ must satisfy

$$\gamma_2(t) = \gamma_1(t)^2 \tag{1.2}$$

for all values of $t$ in the interval $(\alpha, \beta)$ where $\boldsymbol{\gamma}$ is defined (yet to be decided), and ideally every point on the parabola should be equal to $(\gamma_1(t), \gamma_2(t))$ for some value of $t \in (\alpha, \beta)$. Of course, there is an obvious solution to Eq. 1.2: take $\gamma_1(t) = t, \gamma_2(t) = t^2$. To get every point on the parabola we must allow $t$ to take every real number value (since the $x$-coordinate of $\boldsymbol{\gamma}(t)$ is just $t$, and the $x$-coordinate of a point on the parabola can be any real number), so we must take $(\alpha, \beta)$ to be $(-\infty, \infty)$. Thus, the desired parametrization is

$$\boldsymbol{\gamma} : (-\infty, \infty) \to \mathbb{R}^2, \quad \boldsymbol{\gamma}(t) = (t, t^2).$$

But this is not the only parametrization of the parabola. Another choice is $\boldsymbol{\gamma}(t) = (t^3, t^6)$ (with $(\alpha, \beta) = (-\infty, \infty)$). Yet another is $(2t, 4t^2)$, and of course there are (infinitely many) others. So the parametrization of a given level curve is not unique.

## Example 1.1.3

Now we try the circle $x^2 + y^2 = 1$. It is tempting to take $x = t$ as in the previous example, so that $y = \sqrt{1 - t^2}$ (we could have taken $y = -\sqrt{1 - t^2}$). So we get the parametrization

$$\boldsymbol{\gamma}(t) = (t, \sqrt{1 - t^2}).$$

But this is only a parametrization of the upper half of the circle because $\sqrt{1 - t^2}$ is always $\geq 0$. Similarly, if we had taken $y = -\sqrt{1 - t^2}$, we would only have covered the lower half of the circle.

If we want a parametrization of the whole circle, we must try again. We need functions $\gamma_1(t)$ and $\gamma_2(t)$ such that

$$\gamma_1(t)^2 + \gamma_2(t)^2 = 1 \tag{1.3}$$

for all $t \in (\alpha, \beta)$, and such that *every* point on the circle is equal to $(\gamma_1(t), \gamma_2(t))$ for some $t \in (\alpha, \beta)$. There is an obvious solution to Eq. 1.3: $\gamma_1(t) = \cos t$ and $\gamma_2(t) = \sin t$ (since $\cos^2 t + \sin^2 t = 1$ for all values of $t$). We can take $(\alpha, \beta) = (-\infty, \infty)$, although this is overkill: any open interval $(\alpha, \beta)$ whose length is greater than $2\pi$ will suffice.

The next example shows how to pass from parametrized curves to level curves.

## Example 1.1.4

Take the parametrized curve (called an *astroid*)

$$\boldsymbol{\gamma}(t) = (\cos^3 t, \sin^3 t), \quad t \in \mathbb{R}.$$

Since $\cos^2 t + \sin^2 t = 1$ for all $t$, the coordinates $x = \cos^3 t$, $y = \sin^3 t$ of the point $\boldsymbol{\gamma}(t)$ satisfy

$$x^{2/3} + y^{2/3} = 1.$$

This level curve coincides with the image of the map $\boldsymbol{\gamma}$. See Exercise 1.1.5 for a picture of the astroid.

In this book, we shall be studying parametrized curves (and later, surfaces) using methods of calculus. Such curves and surfaces will be described almost exclusively in terms of *smooth* functions: a function $f : (\alpha, \beta) \to \mathbb{R}$ is said to be smooth if the derivative $\frac{d^n f}{dt^n}$ exists for all $n \geq 1$ and all $t \in (\alpha, \beta)$. If $f(t)$ and $g(t)$ are smooth functions, it follows from standard results of calculus that the sum $f(t)+g(t)$, product $f(t)g(t)$, quotient $f(t)/g(t)$, and composite $f(g(t))$ are smooth functions, where they are defined.

To differentiate a *vector-valued* function such as $\boldsymbol{\gamma}(t)$ (as in Definition 1.1.1), we differentiate componentwise: if

$$\boldsymbol{\gamma}(t) = (\gamma_1(t), \gamma_2(t), \ldots, \gamma_n(t)),$$

then

$$\frac{d\boldsymbol{\gamma}}{dt} = \left(\frac{d\gamma_1}{dt}, \frac{d\gamma_2}{dt}, \ldots, \frac{d\gamma_n}{dt}\right), \quad \frac{d^2\boldsymbol{\gamma}}{dt^2} = \left(\frac{d^2\gamma_1}{dt^2}, \frac{d^2\gamma_2}{dt^2}, \ldots, \frac{d^2\gamma_n}{dt^2}\right), \quad \text{etc.}$$

To save space, we often denote $d\boldsymbol{\gamma}/dt$ by $\dot{\boldsymbol{\gamma}}(t)$, $d^2\boldsymbol{\gamma}/dt^2$ by $\ddot{\boldsymbol{\gamma}}(t)$, etc. We say that $\boldsymbol{\gamma}$ is *smooth* if the derivatives $d^n\boldsymbol{\gamma}/dt^n$ exist for all $n \geq 1$ and all $t \in (\alpha, \beta)$; this is equivalent to requiring that each of the components $\gamma_1, \gamma_2, \ldots, \gamma_n$ of $\boldsymbol{\gamma}$ is smooth.

*From now on, all parametrized curves studied in this book will be assumed to be smooth.*

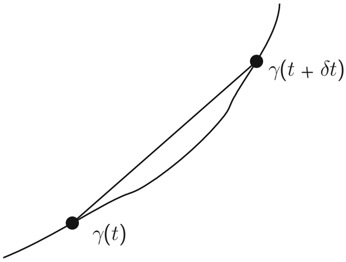
## Definition 1.1.5

If $\boldsymbol{\gamma}$ is a parametrized curve, its first derivative $\dot{\boldsymbol{\gamma}}(t)$ is called the *tangent vector* of $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(t)$.

To see the reason for this terminology, note that the vector

$$\frac{\boldsymbol{\gamma}(t+\delta t)-\boldsymbol{\gamma}(t)}{\delta t}$$

is parallel to the chord joining the points $\boldsymbol{\gamma}(t)$ and $\boldsymbol{\gamma}(t+\delta t)$ of the image $\mathcal{C}$ of $\boldsymbol{\gamma}$:




As $\delta t$ tends to zero the length of the chord also tends to zero, but we expect that the *direction* of the chord becomes parallel to that of the tangent to $\mathcal{C}$ at $\boldsymbol{\gamma}(t)$. But the direction of the chord is the same as that of the vector

$$\frac{\boldsymbol{\gamma}(t+\delta t)-\boldsymbol{\gamma}(t)}{\delta t},$$

which tends to $d\boldsymbol{\gamma}/dt$ as $\delta t$ tends to zero. Of course, this only determines a well-defined direction tangent to the curve if $d\boldsymbol{\gamma}/dt$ is non-zero. If that condition holds, we define the *tangent line* to $\mathcal{C}$ at a point $\mathbf{p}$ of $\mathcal{C}$ to be the straight line passing through $\mathbf{p}$ and parallel to the vector $d\boldsymbol{\gamma}/dt$.

The following result is intuitively clear:

## Proposition 1.1.6

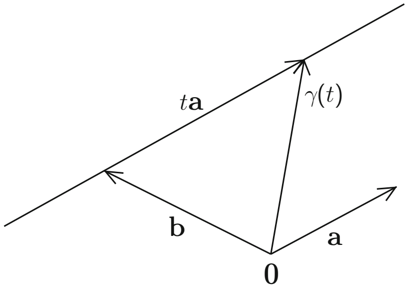
If the tangent vector of a parametrized curve is constant, the image of the curve is (part of) a straight line.

## Proof

If $\dot{\boldsymbol{\gamma}}(t) = \mathbf{a}$ for all $t$, where $\mathbf{a}$ is a constant vector, we have, integrating componentwise,

$$\boldsymbol{\gamma}(t) = \int \frac{d\boldsymbol{\gamma}}{dt}dt = \int \mathbf{a}\,dt = t\,\mathbf{a} + \mathbf{b},$$

where $\mathbf{b}$ is another constant vector. If $\mathbf{a} \neq 0$, this is the parametric equation of the straight line parallel to $\mathbf{a}$ and passing through the point $\mathbf{b}$:




If $\mathbf{a} = 0$, the image of $\boldsymbol{\gamma}$ is a single point (namely, $\mathbf{b}$). □

Before proceeding further with our study of curves, we should point out a potential source of confusion in the discussion of parametrized curves. This is regarding the question what is a 'point' of such a curve? The difficulty can be seen in the following example.

## Example 1.1.7

The *limaçon* is the parametrized curve

$$\boldsymbol{\gamma}(t) = ((1 + 2\cos t)\cos t, (1 + 2\cos t)\sin t), \qquad t \in \mathbb{R}$$

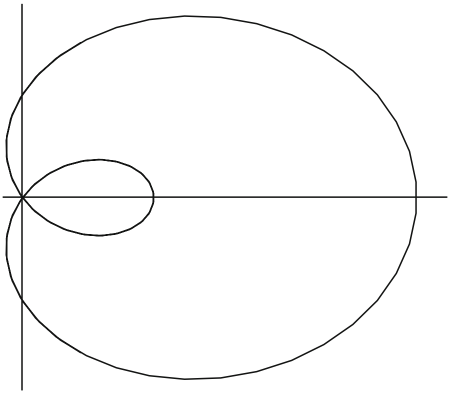
(see the diagram below). Note that $\boldsymbol{\gamma}$ has a self-intersection at the origin in the sense that $\boldsymbol{\gamma}(t) = \mathbf{0}$ for $t = 2\pi/3$ and for $t = 4\pi/3$. The tangent vector is

$$\dot{\boldsymbol{\gamma}}(t) = (-\sin t - 2\sin 2t, \cos t + 2\cos 2t).$$

In particular,

$$\dot{\boldsymbol{\gamma}}(2\pi/3) = (\sqrt{3}/2, -3/2), \quad \dot{\boldsymbol{\gamma}}(4\pi/3) = (-\sqrt{3}/2, -3/2).$$

So what is the tangent vector of this curve at the origin? Although $\dot{\boldsymbol{\gamma}}(t)$ is well-defined for all values of $t$, it takes different values at $t = 2\pi/3$ and $t = 4\pi/3$, both of which correspond to the point $\mathbf{0}$ on the curve.



This example shows that we must be careful while talking about a 'point' of a parametrized curve $\boldsymbol{\gamma}$: strictly speaking, this should be the same thing as a value of the curve parameter $t$, and not the corresponding geometric point $\boldsymbol{\gamma}(t) \in \mathbb{R}^n$. Thus, Definition 1.1.5 should more properly read "If $\boldsymbol{\gamma}$ is a parametrized curve, its first derivative $\dot{\boldsymbol{\gamma}}(t)$ is called the *tangent vector* of $\boldsymbol{\gamma}$ at the parameter value $t$." However, it seems to us that to insist on this distinction takes away from the geometric viewpoint, and we shall sometimes repeat the 'error' committed in the statement of Definition 1.1.5. This should not lead to confusion if the preceding remarks are kept in mind.

## *EXERCISES*

1.1.1 Is $\boldsymbol{\gamma}(t) = (t^2, t^4)$ a parametrization of the parabola $y = x^2$?

1.1.2 Find parametrizations of the following level curves:

(i) $y^2 - x^2 = 1$;

(ii) $\frac{x^2}{4} + \frac{y^2}{9} = 1$.

1.1.3 Find the Cartesian equations of the following parametrized curves:

(i) $\boldsymbol{\gamma}(t) = (\cos^2 t, \sin^2 t)$;

(ii) $\boldsymbol{\gamma}(t) = (e^t, t^2)$.

1.1.4 Calculate the tangent vectors of the curves in Exercise 1.1.3.

1.1.5 Sketch the astroid in Example 1.1.4. Calculate its tangent vector at each point. At which points is the tangent vector zero?

1.1.6 Consider the ellipse

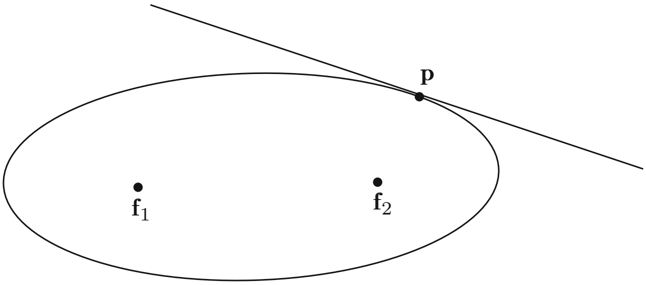
$$\frac{x^2}{p^2} + \frac{y^2}{q^2} = 1,$$

where $p > q > 0$ (see below). The *eccentricity* of the ellipse is $\epsilon = \sqrt{1 - \frac{q^2}{p^2}}$ and the points $(\pm\epsilon p, 0)$ on the $x$-axis are called the *foci* of the ellipse, which we denote by $\mathbf{f}_1$ and $\mathbf{f}_2$. Verify that $\boldsymbol{\gamma}(t) = (p\cos t, q\sin t)$ is a parametrization of the ellipse. Prove that

(i) The sum of the distances from $\mathbf{f}_1$ and $\mathbf{f}_2$ to any point $\mathbf{p}$ on the ellipse does not depend on $\mathbf{p}$.

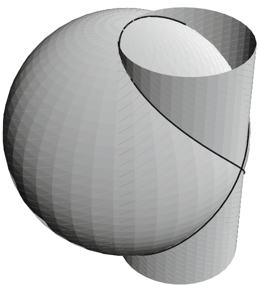
(ii) The product of the distances from $\mathbf{f}_1$ and $\mathbf{f}_2$ to the tangent line at any point $\mathbf{p}$ of the ellipse does not depend on $\mathbf{p}$.

(iii) If $\mathbf{p}$ is any point on the ellipse, the line joining $\mathbf{f}_1$ and $\mathbf{p}$ and that joining $\mathbf{f}_2$ and $\mathbf{p}$ make equal angles with the tangent line to the ellipse at $\mathbf{p}$.




1.1.7 A *cycloid* is the plane curve traced out by a point on the circumference of a circle as it rolls without slipping along a straight line. Show that, if the straight line is the $x$-axis and the circle has radius $a > 0$, the cycloid can be parametrized as

$$\boldsymbol{\gamma}(t) = a(t - \sin t, 1 - \cos t).$$



1.1.8 Show that $\boldsymbol{\gamma}(t) = (\cos^2 t - \frac{1}{2}, \sin t\cos t, \sin t)$ is a parametrization of the curve of intersection of the circular cylinder of radius $\frac{1}{2}$ and axis the $z$-axis with the sphere of radius 1 and centre $(-\frac{1}{2}, 0, 0)$. This is called *Viviani's Curve* – see above.

1.1.9 The *normal line* to a curve at a point $\mathbf{p}$ is the straight line passing through $\mathbf{p}$ perpendicular to the tangent line at $\mathbf{p}$. Find the tangent and normal lines to the curve $\boldsymbol{\gamma}(t) = (2\cos t - \cos 2t, 2\sin t - \sin 2t)$ at the point corresponding to $t = \pi/4$.

## 1.2 Arc-length

We recall that, if $\mathbf{v} = (v_1, \ldots, v_n)$ is a vector in $\mathbb{R}^n$, its *length* is

$$\| \mathbf{v} \| = \sqrt{v_1^2 + \cdots + v_n^2}.$$

If $\mathbf{u}$ is another vector in $\mathbb{R}^n$, $\| \mathbf{u} - \mathbf{v} \|$ is the length of the straight line segment joining the points $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$.

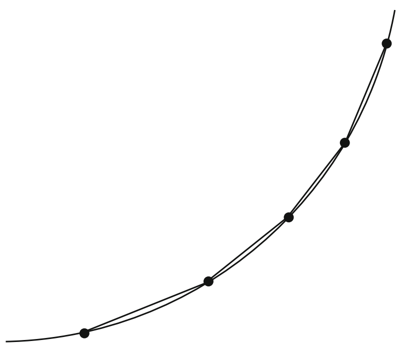
To find a formula for the length of a parametrized curve $\boldsymbol{\gamma}$, note that, if $\delta t$ is very small, the part of the image $\mathcal{C}$ of $\boldsymbol{\gamma}$ between $\boldsymbol{\gamma}(t)$ and $\boldsymbol{\gamma}(t + \delta t)$ is nearly a straight line, so its length is approximately

$$\| \boldsymbol{\gamma}(t + \delta t) - \boldsymbol{\gamma}(t) \|.$$

Again, since $\delta t$ is small, $(\boldsymbol{\gamma}(t + \delta t) - \boldsymbol{\gamma}(t))/\delta t$ is nearly equal to $\dot{\boldsymbol{\gamma}}(t)$, so the length is approximately

$$\| \dot{\boldsymbol{\gamma}}(t) \| \, \delta t. \tag{1.4}$$

If we want to calculate the length of a (not necessarily small) part of $\mathcal{C}$, we can divide it into segments, each of which corresponds to a small increment $\delta t$ in $t$, calculate the length of each segment using (1.4), and add up the results. Letting $\delta t$ tend to zero should then give the exact length.



This motivates the following definition:

## Definition 1.2.1

The *arc-length* of a curve $\boldsymbol{\gamma}$ starting at the point $\boldsymbol{\gamma}(t_0)$ is the function $s(t)$ given by

$$s(t) = \int_{t_0}^{t} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du.$$

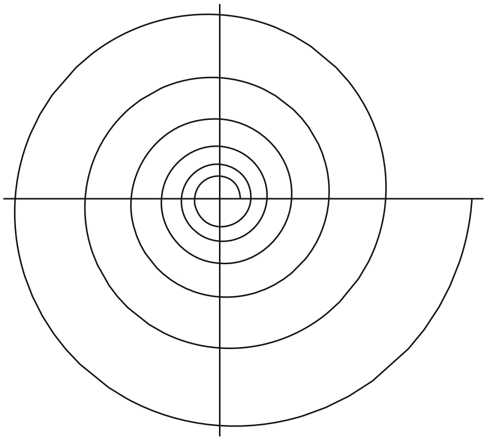
Thus, $s(t_0) = 0$ and $s(t)$ is positive or negative according to whether $t$ is larger or smaller than $t_0$. If we choose a different starting point $\boldsymbol{\gamma}(\tilde{t}_0)$, the resulting arc-length $\tilde{s}$ differs from $s$ by the constant $\int_{t_0}^{\tilde{t}_0} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du$ because

$$\int_{t_0}^{t} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du = \int_{\tilde{t}_0}^{t} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du + \int_{t_0}^{\tilde{t}_0} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du.$$

## Example 1.2.2

For a *logarithmic spiral*

$$\boldsymbol{\gamma}(t) = (e^{kt}\cos t, e^{kt}\sin t),$$



where $k$ is a non-zero constant, we have

$$\dot{\boldsymbol{\gamma}} = (e^{kt}(k\cos t - \sin t), e^{kt}(k\sin t + \cos t)),$$
$$\therefore \quad \parallel \dot{\boldsymbol{\gamma}} \parallel^2 = e^{2kt}(k\cos t - \sin t)^2 + e^{2kt}(k\sin t + \cos t)^2 = (k^2+1)e^{2kt}.$$

Hence, the arc-length of $\boldsymbol{\gamma}$ starting at $\boldsymbol{\gamma}(0) = (1, 0)$ (for example) is

$$s = \int_0^t \sqrt{k^2+1}e^{ku}\, du = \frac{\sqrt{k^2+1}}{k}(e^{kt} - 1).$$

The arc-length is a differentiable function. Indeed, if $s$ is the arc-length of a curve $\boldsymbol{\gamma}$ starting at $\boldsymbol{\gamma}(t_0)$, we have

$$\frac{ds}{dt} = \frac{d}{dt}\int_{t_0}^{t} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du = \parallel \dot{\boldsymbol{\gamma}}(t) \parallel. \tag{1.5}$$

Thinking of $\boldsymbol{\gamma}(t)$ as the position of a moving point at time $t$, $ds/dt$ is the speed of the point (rate of change of distance along the curve). This suggests the following definition.

## Definition 1.2.3

If $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^n$ is a parametrized curve, its *speed* at the point $\boldsymbol{\gamma}(t)$ is $\parallel \dot{\boldsymbol{\gamma}}(t) \parallel$, and $\boldsymbol{\gamma}$ is said to be a *unit-speed* curve if $\dot{\boldsymbol{\gamma}}(t)$ is a unit vector for all $t \in (\alpha, \beta)$.

We shall see many examples of formulas and results relating to curves that take on a much simpler form when the curve is unit-speed. The reason for this simplification is given in the next proposition. Although this admittedly looks uninteresting at first sight, it will be extremely useful for what follows.

We recall that the *dot product* (or scalar product) of vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ in $\mathbb{R}^n$ is

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i.$$

If $\mathbf{a}$ and $\mathbf{b}$ are smooth functions of a parameter $t$, we shall make use of the 'product formula'

$$\frac{d}{dt}(\mathbf{a} \cdot \mathbf{b}) = \frac{d\mathbf{a}}{dt} \cdot \mathbf{b} + \mathbf{a} \cdot \frac{d\mathbf{b}}{dt}.$$

This follows easily from the definition of the dot product and the usual product formula for scalar functions,

$$\frac{d}{dt}(a_i b_i) = \frac{da_i}{dt} b_i + a_i \frac{db_i}{dt}.$$

## Proposition 1.2.4

Let $\mathbf{n}(t)$ be a unit vector that is a smooth function of a parameter $t$. Then, the dot product

$$\dot{\mathbf{n}}(t) \cdot \mathbf{n}(t) = 0$$

for all $t$, i.e., $\dot{\mathbf{n}}(t)$ is zero or perpendicular to $\mathbf{n}(t)$ for all $t$.

In particular, if $\boldsymbol{\gamma}$ is a unit-speed curve, then $\ddot{\boldsymbol{\gamma}}$ is zero or perpendicular to $\dot{\boldsymbol{\gamma}}$.

## Proof

Using the product formula to differentiate both sides of the equation $\mathbf{n} \cdot \mathbf{n} = 1$ with respect to $t$ gives

$$\dot{\mathbf{n}} \cdot \mathbf{n} + \mathbf{n} \cdot \dot{\mathbf{n}} = 0,$$

so $2\dot{\mathbf{n}} \cdot \mathbf{n} = 0$. The last part follows by taking $\mathbf{n} = \dot{\boldsymbol{\gamma}}$. □

## *EXERCISES*

1.2.1 Calculate the arc-length of the *catenary* $\boldsymbol{\gamma}(t) = (t, \cosh t)$ starting at the point $(0, 1)$. This curve has the shape of a heavy chain suspended at its ends – see Exercise 2.2.4.

1.2.2 Show that the following curves are unit-speed:

(i) $\boldsymbol{\gamma}(t) = \left(\frac{1}{3}(1+t)^{3/2}, \frac{1}{3}(1-t)^{3/2}, \frac{t}{\sqrt{2}}\right)$.

(ii) $\boldsymbol{\gamma}(t) = \left(\frac{4}{5}\cos t, 1 - \sin t, -\frac{3}{5}\cos t\right)$.

1.2.3 A plane curve is given by

$$\boldsymbol{\gamma}(\theta) = (r\cos\theta, r\sin\theta),$$

where $r$ is a smooth function of $\theta$ (so that $(r, \theta)$ are the polar coordinates of $\boldsymbol{\gamma}(\theta)$). Under what conditions is $\boldsymbol{\gamma}$ regular? Find all functions $r(\theta)$ for which $\boldsymbol{\gamma}$ is unit-speed. Show that, if $\boldsymbol{\gamma}$ is unit-speed, the image of $\boldsymbol{\gamma}$ is a circle; what is its radius?

1.2.4 This exercise shows that *a straight line is the shortest curve joining two given points*. Let $\mathbf{p}$ and $\mathbf{q}$ be the two points, and let $\boldsymbol{\gamma}$ be a curve passing through both, say $\boldsymbol{\gamma}(a) = \mathbf{p}$, $\boldsymbol{\gamma}(b) = \mathbf{q}$, where $a < b$. Show that, if $\mathbf{u}$ is any unit vector,

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{u} \leq \parallel \dot{\boldsymbol{\gamma}} \parallel$$

and deduce that

$$(\mathbf{q} - \mathbf{p}) \cdot \mathbf{u} \leq \int_a^b \parallel \dot{\boldsymbol{\gamma}} \parallel \, dt.$$

By taking $\mathbf{u} = (\mathbf{q} - \mathbf{p}) / \parallel \mathbf{q} - \mathbf{p} \parallel$, show that the length of the part of $\boldsymbol{\gamma}$ between $\mathbf{p}$ and $\mathbf{q}$ is at least the straight line distance $\parallel \mathbf{q} - \mathbf{p} \parallel$.

## 1.3 Reparametrization

We saw in Examples 1.1.2 and 1.1.3 that a given level curve can have many parametrizations, and it is important to understand the relation between them.

### Definition 1.3.1

A parametrized curve $\tilde{\boldsymbol{\gamma}} : (\tilde{\alpha}, \tilde{\beta}) \to \mathbb{R}^n$ is a *reparametrization* of a parametrized curve $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^n$ if there is a smooth bijective map $\phi : (\tilde{\alpha}, \tilde{\beta}) \to (\alpha, \beta)$ (the *reparametrization map*) such that the inverse map $\phi^{-1} : (\alpha, \beta) \to (\tilde{\alpha}, \tilde{\beta})$ is also smooth and

$$\tilde{\boldsymbol{\gamma}}(\tilde{t}) = \boldsymbol{\gamma}(\phi(\tilde{t})) \quad \text{for all} \quad \tilde{t} \in (\tilde{\alpha}, \tilde{\beta}). \tag{1.6}$$

Note that, since $\phi$ has a smooth inverse, $\boldsymbol{\gamma}$ is a reparametrization of $\tilde{\boldsymbol{\gamma}}$:

$$\tilde{\boldsymbol{\gamma}}(\phi^{-1}(t)) = \boldsymbol{\gamma}(\phi(\phi^{-1}(t))) = \boldsymbol{\gamma}(t) \text{ for all } t \in (\alpha, \beta).$$

Two curves that are reparametrizations of each other have the same image, so they should have the same geometric properties.

### Example 1.3.2

In Example 1.1.3, we found that the circle $x^2 + y^2 = 1$ has a parametrization $\boldsymbol{\gamma}(t) = (\cos t, \sin t)$. Another parametrization is

$$\tilde{\boldsymbol{\gamma}}(t) = (\sin t, \cos t)$$

(since $\sin^2 t + \cos^2 t = 1$). To see that $\tilde{\boldsymbol{\gamma}}$ is a reparametrization of $\boldsymbol{\gamma}$, we have to find a reparametrization map $\phi$ such that

$$(\cos \phi(t), \sin \phi(t)) = (\sin t, \cos t).$$

One solution is $\phi(t) = \pi/2 - t$.

As we remarked in Section 1.2, the analysis of a curve is simplified when it is known to be unit-speed. It is therefore important to know exactly which curves have unit-speed reparametrizations.

### Definition 1.3.3

A point $\boldsymbol{\gamma}(t)$ of a parametrized curve $\boldsymbol{\gamma}$ is called a *regular point* if $\dot{\boldsymbol{\gamma}}(t) \neq \mathbf{0}$; otherwise $\boldsymbol{\gamma}(t)$ is a *singular point* of $\boldsymbol{\gamma}$. A curve is *regular* if all of its points are regular.

Before we show the relation between regularity and unit-speed reparametrization, we note two simple properties of regular curves. Although these results are not particularly appealing, they are very important for what is to follow.

### Proposition 1.3.4

Any reparametrization of a regular curve is regular.

### Proof

Suppose that $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ are related as in Definition 1.3.1, let $t = \phi(\tilde{t})$ and $\psi = \phi^{-1}$ so that $\tilde{t} = \psi(t)$. Differentiating both sides of the equation $\phi(\psi(t)) = t$ with respect to $t$ and using the chain rule gives

$$\frac{d\phi}{d\tilde{t}}\frac{d\psi}{dt} = 1.$$

This shows that $d\phi/d\tilde{t}$ is never zero. Since $\tilde{\boldsymbol{\gamma}}(\tilde{t}) = \boldsymbol{\gamma}(\phi(\tilde{t}))$, another application of the chain rule gives

$$\frac{d\tilde{\boldsymbol{\gamma}}}{d\tilde{t}} = \frac{d\boldsymbol{\gamma}}{dt}\frac{d\phi}{d\tilde{t}},$$

which shows that $d\tilde{\boldsymbol{\gamma}}/d\tilde{t}$ is never zero, if $d\boldsymbol{\gamma}/dt$ is never zero. $\square$

### Proposition 1.3.5

If $\boldsymbol{\gamma}(t)$ is a regular curve, its arc-length $s$ (see Definition 1.2.1), starting at any point of $\boldsymbol{\gamma}$, is a smooth function of $t$.

### Proof

We have already seen that (whether or not $\boldsymbol{\gamma}$ is regular) $s$ is a differentiable function of $t$ and

$$\frac{ds}{dt} = \parallel \dot{\boldsymbol{\gamma}}(t) \parallel.$$

To simplify the notation, assume from now onwards that $\boldsymbol{\gamma}$ is a plane curve, say

$$\boldsymbol{\gamma}(t) = (u(t), v(t)),$$

where $u$ and $v$ are smooth functions of $t$, so that

$$\frac{ds}{dt} = \sqrt{\dot{u}^2 + \dot{v}^2}.$$

The crucial point is that the function $f(x) = \sqrt{x}$ is a *smooth* function on the open interval $(0, \infty)$. Indeed, it is easy to prove by induction on $n \geq 1$ that

$$\frac{d^n f}{dx^n} = (-1)^{n-1} \frac{1.3.5.\ldots.(2n-1)}{2^n} x^{-(2n+1)/2}.$$

Since $u$ and $v$ are smooth functions of $t$, so are $\dot{u}$ and $\dot{v}$ and hence is $\dot{u}^2 + \dot{v}^2$. Since $\boldsymbol{\gamma}$ is regular, $\dot{u}^2 + \dot{v}^2 > 0$ for all values of $t$, so the composite function

$$\frac{ds}{dt} = f(\dot{u}^2 + \dot{v}^2)$$

is a smooth function of $t$, and hence $s$ itself is smooth. □

The main result we want is the following.

## Proposition 1.3.6

A parametrized curve has a unit-speed reparametrization if and only if it is regular.

## Proof

Suppose first that a parametrized curve $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^n$ has a unit-speed reparametrization $\tilde{\boldsymbol{\gamma}}$, with reparametrization map $\phi$. Letting $t = \phi(\tilde{t})$, we have $\tilde{\boldsymbol{\gamma}}(\tilde{t}) = \boldsymbol{\gamma}(t)$ and so

$$\begin{aligned} \frac{d\tilde{\boldsymbol{\gamma}}}{d\tilde{t}} &= \frac{d\boldsymbol{\gamma}}{dt}\frac{dt}{d\tilde{t}}, \\ \therefore \quad \left\| \frac{d\tilde{\boldsymbol{\gamma}}}{d\tilde{t}} \right\| &= \left\| \frac{d\boldsymbol{\gamma}}{dt} \right\| \left| \frac{dt}{d\tilde{t}} \right|. \end{aligned}$$

Since $\tilde{\boldsymbol{\gamma}}$ is unit-speed, $\| d\tilde{\boldsymbol{\gamma}}/d\tilde{t} \| = 1$, so $d\boldsymbol{\gamma}/dt$ cannot be zero.

Conversely, suppose that the tangent vector $d\boldsymbol{\gamma}/dt$ is never zero. By Eq. 1.5, $ds/dt > 0$ for all $t$, where $s$ is the arc-length of $\boldsymbol{\gamma}$ starting at any point of the curve, and by Proposition 1.3.5 $s$ is a smooth function of $t$. It follows from the inverse function theorem that $s : (\alpha, \beta) \to \mathbb{R}$ is injective, that its image is an open interval $(\tilde{\alpha}, \tilde{\beta})$, and that the inverse map $s^{-1} : (\tilde{\alpha}, \tilde{\beta}) \to (\alpha, \beta)$ is smooth. (Readers unfamiliar with the inverse function theorem should accept these statements for now; the theorem will be discussed informally in Section 1.5 and formally in Section 5.6.) We take $\phi = s^{-1}$ and let $\tilde{\boldsymbol{\gamma}}$ be the corresponding reparametrization of $\boldsymbol{\gamma}$, so that $\tilde{\boldsymbol{\gamma}}(s) = \boldsymbol{\gamma}(t)$ (see Eq. 1.6). Then,

$$\frac{d\tilde{\boldsymbol{\gamma}}}{ds}\frac{ds}{dt} = \frac{d\boldsymbol{\gamma}}{dt},$$

$$\therefore \quad \left\| \frac{d\tilde{\gamma}}{ds} \right\| \frac{ds}{dt} = \left\| \frac{d\gamma}{dt} \right\| = \frac{ds}{dt} \quad \text{(by Eq. 1.5)},$$
$$\therefore \quad \left\| \frac{d\tilde{\gamma}}{ds} \right\| = 1.$$

□

The proof of Proposition 1.3.6 shows that the arc-length is essentially the only unit-speed parameter on a regular curve:

## Corollary 1.3.7

Let $\boldsymbol{\gamma}$ be a regular curve and let $\tilde{\boldsymbol{\gamma}}$ be a unit-speed reparametrization of $\boldsymbol{\gamma}$:

$$\tilde{\boldsymbol{\gamma}}(u(t)) = \boldsymbol{\gamma}(t) \quad \text{for all } t,$$

where $u$ is a smooth function of $t$. Then, if $s$ is the arc-length of $\boldsymbol{\gamma}$ (starting at any point), we have

$$u = \pm s + c, \tag{1.7}$$

where $c$ is a constant. Conversely, if $u$ is given by Eq. 1.7 for some value of $c$ and with either sign, then $\tilde{\boldsymbol{\gamma}}$ is a unit-speed reparametrization of $\boldsymbol{\gamma}$.

## Proof

The calculation in the first part of the proof of Proposition 1.3.6 shows that $u$ gives a unit-speed reparametrization of $\boldsymbol{\gamma}$ if and only if

$$\frac{du}{dt} = \pm \left\| \frac{d\boldsymbol{\gamma}}{dt} \right\| = \pm \frac{ds}{dt} \quad \text{(by Eq. 1.5)},$$

which is equivalent to $u = \pm s + c$ for some constant $c$. □

Although every regular curve has a unit-speed reparametrization, this may be very complicated, or even impossible, to write down 'explicitly', as the following examples show.

## Example 1.3.8

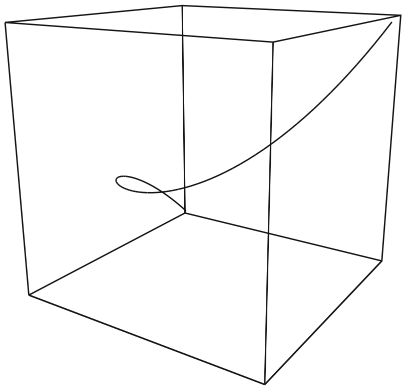
For the logarithmic spiral $\boldsymbol{\gamma}(t) = (e^{kt}\cos t, e^{kt}\sin t)$, we found in Example 1.2.2 that $\parallel \dot{\boldsymbol{\gamma}} \parallel^2 = (k^2+1)e^{2kt}$. This is never zero, so $\boldsymbol{\gamma}$ is regular. The arc-length of $\boldsymbol{\gamma}$ starting at $(1,0)$ was found to be $s = \sqrt{k^2+1}(e^{kt}-1)/k$. Hence, $t = \frac{1}{k}\ln\left(\frac{ks}{\sqrt{k^2+1}} + 1\right)$, so a unit-speed reparametrization of $\boldsymbol{\gamma}$ is given by the rather unwieldy formula

$$\tilde{\boldsymbol{\gamma}}(s) = \left(\left(\frac{ks}{\sqrt{k^2+1}}+1\right)\cos\left(\frac{1}{k}\ln\left(\frac{ks}{\sqrt{k^2+1}}+1\right)\right),\right.$$
$$\left.\left(\frac{ks}{\sqrt{k^2+1}}+1\right)\sin\left(\frac{1}{k}\ln\left(\frac{ks}{\sqrt{k^2+1}}+1\right)\right)\right).$$

## Example 1.3.9

The *twisted cubic* is the space curve given by

$$\boldsymbol{\gamma}(t) = (t, t^2, t^3), \quad t \in \mathbb{R}.$$



We have $\dot{\boldsymbol{\gamma}}(t) = (1, 2t, 3t^2)$ and so

$$\| \dot{\boldsymbol{\gamma}}(t) \| = \sqrt{1+4t^2+9t^4}.$$

This is never zero, so $\boldsymbol{\gamma}$ is regular. The arc-length starting at $\boldsymbol{\gamma}(0) = \mathbf{0}$ is

$$s = \int_0^t \sqrt{1+4u^2+9u^4}\, du.$$

This integral cannot be evaluated in terms of familiar functions like logarithms and exponentials, and trigonometric functions. (It is an example of an *elliptic integral.*)

Our final example shows that a given level curve can have both regular and non-regular parametrizations.

## Example 1.3.10

For the parametrization $\boldsymbol{\gamma}(t) = (t, t^2)$ of the parabola $y = x^2$, $\dot{\boldsymbol{\gamma}}(t) = (1, 2t)$ is obviously never zero, so $\boldsymbol{\gamma}$ is regular. But $\tilde{\boldsymbol{\gamma}}(t) = (t^3, t^6)$ is also a parametrization of the same parabola. This time, $\dot{\tilde{\boldsymbol{\gamma}}} = (3t^2, 6t^5)$, and this is zero when $t = 0$, so $\tilde{\boldsymbol{\gamma}}$ is *not* regular.

## EXERCISES

1.3.1 Which of the following curves are regular?

(i) $\boldsymbol{\gamma}(t) = (\cos^2 t, \sin^2 t)$ for $t \in \mathbb{R}$.

(ii) The same curve as in (i), but with $0 < t < \pi/2$.

(iii) $\boldsymbol{\gamma}(t) = (t, \cosh t)$ for $t \in \mathbb{R}$.

Find unit-speed reparametrizations of the regular curve(s).

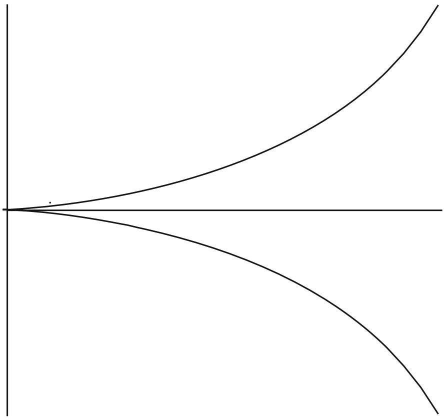
1.3.2 The *cissoid of Diocles* (see below) is the curve whose equation in terms of polar coordinates $(r, \theta)$ is

$$r = \sin\theta \tan\theta, \quad -\pi/2 < \theta < \pi/2.$$

Write down a parametrization of the cissoid using $\theta$ as a parameter and show that

$$\boldsymbol{\gamma}(t) = \left(t^2, \frac{t^3}{\sqrt{1-t^2}}\right), \quad -1 < t < 1$$

is a reparametrization of it.



1.3.3 The simplest type of singular point of a curve $\boldsymbol{\gamma}$ is an *ordinary cusp*: a point $\mathbf{p}$ of $\boldsymbol{\gamma}$, corresponding to a parameter value $t_0$, say, is an ordinary cusp if $\dot{\boldsymbol{\gamma}}(t_0) = \mathbf{0}$ and the vectors $\ddot{\boldsymbol{\gamma}}(t_0)$ and $\dddot{\boldsymbol{\gamma}}(t_0)$ are linearly independent (in particular, these vectors must both be non-zero). Show that:

(i) The curve $\boldsymbol{\gamma}(t) = (t^m, t^n)$, where $m$ and $n$ are positive integers, has an ordinary cusp at the origin if and only if $(m, n) = (2, 3)$ or $(3, 2)$.

(ii) The cissoid in Exercise 1.3.2 has an ordinary cusp at the origin.

(iii) If $\boldsymbol{\gamma}$ has an ordinary cusp at a point $\mathbf{p}$, so does any reparametrization of $\boldsymbol{\gamma}$.

1.3.4 Show that:

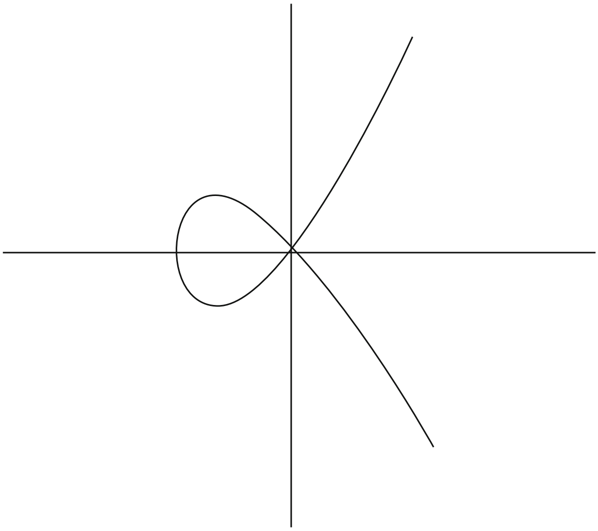
(i) If $\tilde{\boldsymbol{\gamma}}$ is a reparametrization of a curve $\boldsymbol{\gamma}$, then $\boldsymbol{\gamma}$ is a reparametrization of $\tilde{\boldsymbol{\gamma}}$.

(ii) If $\tilde{\boldsymbol{\gamma}}$ is a reparametrization of $\boldsymbol{\gamma}$, and $\hat{\boldsymbol{\gamma}}$ is a reparametrization of $\tilde{\boldsymbol{\gamma}}$, then $\hat{\boldsymbol{\gamma}}$ is a reparametrization of $\boldsymbol{\gamma}$.

## 1.4 Closed curves

It is obvious that some curves 'close up', like a circle or an ellipse, while some do not, like a straight line or a parabola. If a point moves, say at constant speed, around a curve that closes up, it will return to its starting point after some time interval, and will then trace out the same curve all over again. On the other hand, if a point moves at constant speed along a straight line or a parabola, it will never return to its starting point. But there are some intermediate cases like

$$\boldsymbol{\gamma}(t) = (t^2 - 1, t^3 - t);$$



a point moving at constant speed along this curve may return to its starting point if the starting point is the origin, but will not do so otherwise. So a careful definition of what it means for a curve to 'close up' is needed.

## Definition 1.4.1

Let $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$ be a smooth curve and let $T \in \mathbb{R}$. We say that $\boldsymbol{\gamma}$ is *$T$-periodic* if

$$\boldsymbol{\gamma}(t+T) = \boldsymbol{\gamma}(t) \qquad \text{for all } t \in \mathbb{R}.$$

If $\boldsymbol{\gamma}$ is not constant and is $T$-periodic for some $T \neq 0$, then $\boldsymbol{\gamma}$ is said to be *closed*.

Thus, if $\boldsymbol{\gamma}$ is $T$-periodic, a point moving around $\boldsymbol{\gamma}$ returns to its starting point after time $T$, whatever the starting point is. Of course, every curve is 0-periodic.

### Remark

If $\boldsymbol{\gamma}$ is $T$-periodic, it is clear that $\boldsymbol{\gamma}$ is determined by its restriction to any interval of length $|T|$. Conversely, closed curves are often given to us as curves defined on a closed interval, say $\boldsymbol{\gamma} : [a, b] \to \mathbb{R}^n$. If $\boldsymbol{\gamma}$ and all its derivatives take the same value at $a$ and $b$,[1] there is a unique way to extend $\boldsymbol{\gamma}$ to a $(b-a)$-periodic (smooth) curve $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$. Thus, the discussion below can be applied to curves defined on closed intervals.

It is clear that if a curve $\boldsymbol{\gamma}$ is $T$-periodic then it is $(-T)$-periodic because

$$\boldsymbol{\gamma}(t-T) = \boldsymbol{\gamma}((t-T)+T) = \boldsymbol{\gamma}(t).$$

It follows that if $\boldsymbol{\gamma}$ is $T$-periodic for some $T \neq 0$, then it is $T$-periodic for some $T > 0$.

## Definition 1.4.2

The *period* of a closed curve $\boldsymbol{\gamma}$ is the smallest positive number $T$ such that $\boldsymbol{\gamma}$ is $T$-periodic.

It is actually not quite obvious that this number $T$ exists (remember that not every set of positive real numbers has a smallest element). A proof that it does exist can be found in the exercises.

## Example 1.4.3

The ellipse $\boldsymbol{\gamma}(t) = (p\cos t, q\sin t)$ (Exercise 1.1.6) is a closed curve with period $2\pi$ because both of its components are (by well-known properties of trigonometric functions).

[1] The derivatives at the endpoints $a$ and $b$ must be defined in the one-sided sense.

If $\boldsymbol{\gamma}$ is a regular closed curve, a unit-speed reparametrization of $\boldsymbol{\gamma}$ is always closed. To see this, note that since every point in the image of a closed curve $\boldsymbol{\gamma}$ of period $T$ is traced out as the parameter $t$ of $\boldsymbol{\gamma}$ varies through any interval of length $T$, for example, $0 \leq t \leq T$, it is reasonable to define the *length of* $\boldsymbol{\gamma}$ to be

$$\ell(\boldsymbol{\gamma}) = \int_0^T \parallel \dot{\boldsymbol{\gamma}}(t) \parallel dt.$$

By the proof of Proposition 1.3.6, using the arc-length

$$s = \int_0^t \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du$$

of $\boldsymbol{\gamma}$ as the parameter gives a unit-speed reparametrization $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ (so that $\tilde{\boldsymbol{\gamma}}(s) = \boldsymbol{\gamma}(t)$). Note that

$$s(t+T) = \int_0^{t+T} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du = \int_0^T \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du + \int_T^{t+T} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel du = \ell(\boldsymbol{\gamma}) + s(t),$$

since, putting $v = u - T$ and using $\boldsymbol{\gamma}(u-T) = \boldsymbol{\gamma}(u)$ (and hence by differentiation $\dot{\boldsymbol{\gamma}}(u-T) = \dot{\boldsymbol{\gamma}}(u)$), we get

$$\int_T^{t+T} \parallel \dot{\boldsymbol{\gamma}}(u) \parallel \, du = \int_0^t \parallel \dot{\boldsymbol{\gamma}}(v) \parallel \, dv = s(t).$$

Hence,

$$\tilde{\boldsymbol{\gamma}}(s(t)) = \tilde{\boldsymbol{\gamma}}(s(t')) \iff \boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(t') \iff t' - t = kT \iff s(t') - s(t) = k\ell(\boldsymbol{\gamma}),$$

where $k$ is an integer. This shows that $\tilde{\boldsymbol{\gamma}}$ is a closed curve with period $\ell(\boldsymbol{\gamma})$. Note that, since $\tilde{\boldsymbol{\gamma}}$ is unit-speed, this is also the length of $\tilde{\boldsymbol{\gamma}}$. In short, *we can always assume that a closed curve is unit-speed and that its period is equal to its length.*

Returning to the curve illustrated at the beginning of this section, it is clearly not closed; nevertheless, if a point starts at the origin and moves at constant speed around the loop in the region $x < 0$ it will return to its starting point. This suggests the following definition.

## Definition 1.4.4

A curve $\boldsymbol{\gamma}$ is said to have a *self-intersection* at a point $\mathbf{p}$ of the curve if there exist parameter values $a \neq b$ such that

(i) $\boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(b) = \mathbf{p}$, and

(ii) if $\boldsymbol{\gamma}$ is closed with period $T$, then $a - b$ is not an integer multiple of $T$.

### Example 1.4.5

The limaçon in Example 1.1.7 is a closed curve with period $2\pi$. It is clear from the picture that it has exactly one self-intersection, at the origin. (This can also be verified analytically – cf. Exercise 1.4.1 and its solution.)

## *EXERCISES*

1.4.1 Show that the *Cayley sextic*

$$\boldsymbol{\gamma}(t) = (\cos^3 t \cos 3t, \cos^3 t \sin 3t), \qquad t \in \mathbb{R},$$

is a closed curve which has exactly one self-intersection. What is its period? (The name of this curve derives from the fact that its Cartesian equation involves a polynomial of degree 6.)

1.4.2 Give an example to show that a reparametrization of a closed curve need not be closed.

1.4.3 Show that if a curve $\boldsymbol{\gamma}$ is $T_1$-periodic and $T_2$-periodic, then it is $(k_1T_1 + k_2T_2)$-periodic for any integers $k_1, k_2$.

1.4.4 Let $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$ be a curve and suppose that $T_0$ is the smallest positive number such that $\boldsymbol{\gamma}$ is $T_0$-periodic. Prove that $\boldsymbol{\gamma}$ is $T$-periodic if and only if $T = kT_0$ for some integer $k$.

1.4.5 Suppose that a *non-constant* function $\gamma : \mathbb{R} \to \mathbb{R}$ is $T$-periodic for some $T \neq 0$. This exercise shows that there is a *smallest* positive $T_0$ such that $\gamma$ is $T_0$-periodic. The proof uses a little real analysis. Suppose for a contradiction that there is no such $T_0$.

(i) Show that there is a sequence $T_1, T_2, T_3, \ldots$ such that $T_1 > T_2 > T_3 > \cdots > 0$ and that $\gamma$ is $T_r$-periodic for all $r \geq 1$.

(ii) Show that the sequence $\{T_r\}$ in (i) can be chosen so that $T_r \to 0$ as $r \to \infty$.

(iii) Show that the existence of a sequence $\{T_r\}$ as in (i) such that $T_r \to 0$ as $r \to \infty$ implies that $\gamma$ is constant.

1.4.6 Let $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$ be a non-constant curve that is $T$-periodic for some $T > 0$. Show that $\boldsymbol{\gamma}$ is closed.

## 1.5 Level curves versus parametrized curves

We shall now try to clarify the relation between the two types of curves we have considered in previous sections.

Level curves in the generality we have defined them are not always the kind of objects we would want to call curves. For example, the level 'curve' $x^2 + y^2 = 0$ is a single point. The correct conditions to impose on a function $f(x, y)$ in order that $f(x, y) = c$, where $c$ is a constant, will be an acceptable level curve in the plane are contained in the following theorem, which shows that such level curves can be parametrized. Note that we might as well assume that $c = 0$ (since we can replace $f$ by $f - c$).

### Theorem 1.5.1

Let $f(x, y)$ be a smooth function of two variables (which means that all the partial derivatives of $f$, of all orders, exist and are continuous functions). Assume that, at every point of the level curve

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = 0\},$$

$\partial f/\partial x$ and $\partial f/\partial y$ are not both zero. If $\mathbf{p}$ is a point of $\mathcal{C}$, with coordinates $(x_0, y_0)$, say, there is a regular parametrized curve $\boldsymbol{\gamma}(t)$, defined on an open interval containing 0, such that $\boldsymbol{\gamma}$ passes through $\mathbf{p}$ when $t = 0$ and $\boldsymbol{\gamma}(t)$ is contained in $\mathcal{C}$ for all $t$.

The proof of this theorem makes use of the inverse function theorem (one version of which has already been used in the proof of Proposition 1.3.6). For the moment, we shall only try to convince the reader of the truth of this theorem. The proof will be given later (Exercise 5.6.2) after the inverse function theorem has been formally introduced and used in our discussion of surfaces.

To understand the significance of the conditions on $f$ in Theorem 1.5.1, suppose that $(x_0 + \Delta x, y_0 + \Delta y)$ is a point of $\mathcal{C}$ near $\mathbf{p}$, so that

$$f(x_0 + \Delta x, y_0 + \Delta y) = 0.$$

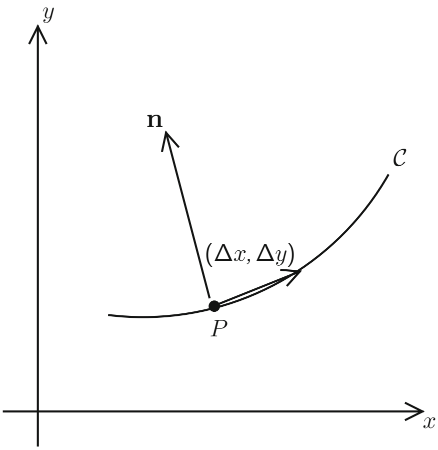
By the two-variable form of Taylor's theorem,

$$f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) + \Delta x \frac{\partial f}{\partial x} + \Delta y \frac{\partial f}{\partial y},$$

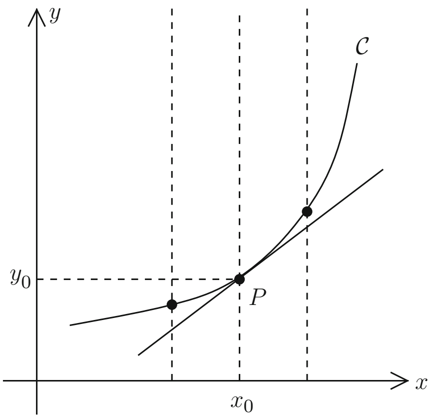
neglecting products of the small quantities $\Delta x$ and $\Delta y$ (the partial derivatives are evaluated at $(x_0, y_0)$). Hence,

$$\Delta x \frac{\partial f}{\partial x} + \Delta y \frac{\partial f}{\partial y} = 0. \tag{1.8}$$

Since $\Delta x$ and $\Delta y$ are small, the vector $(\Delta x, \Delta y)$ is nearly tangent to $\mathcal{C}$ at $\mathbf{p}$, so Eq. 1.8 says that *the vector* $\mathbf{n} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$ *is perpendicular to* $\mathcal{C}$ *at* $\mathbf{p}$.



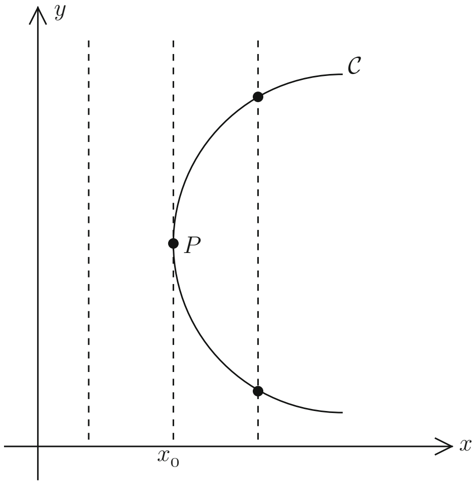

The hypothesis in Theorem 1.5.1 tells us that the vector $\mathbf{n}$ is non-zero at every point of $\mathcal{C}$. Suppose, for example, that $\frac{\partial f}{\partial y} \neq 0$ at $\mathbf{p}$. Then, $\mathbf{n}$ is not parallel to the $x$-axis at $\mathbf{p}$, so the tangent to $\mathcal{C}$ at $\mathbf{p}$ is not parallel to the $y$-axis.




This implies that vertical lines $x =$ constant near $x = x_0$ all intersect $\mathcal{C}$ in a unique point $(x, y)$ near $\mathbf{p}$. In other words, *the equation*

$$f(x, y) = 0 \tag{1.9}$$

*has a unique solution $y$ near $y_0$ for every $x$ near $x_0$.* Note that this may fail to be the case if the tangent to $\mathcal{C}$ at $\mathbf{p}$ is parallel to the $y$-axis (i.e., if $\partial f/\partial y = 0$):

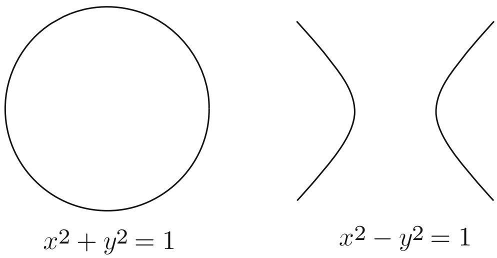



In this example, lines $x =$ constant just to the left of $x = x_0$ do not meet $\mathcal{C}$ near $\mathbf{p}$, while those just to the right of $x = x_0$ meet $\mathcal{C}$ in more than one point near $\mathbf{p}$.

The italicized statement about $f$ in the last paragraph means that there is a function $g(x)$, defined for $x$ near $x_0$, such that $y = g(x)$ is the unique solution of Eq. 1.9 near $y_0$. We can now define a parametrization $\boldsymbol{\gamma}$ of the part of $\mathcal{C}$ near $\mathbf{p}$ by

$$\boldsymbol{\gamma}(t) = (t, g(t)).$$

If we accept that $g$ is smooth (which follows from the inverse function theorem), then $\boldsymbol{\gamma}$ is certainly regular since $\dot{\boldsymbol{\gamma}} = (1, \dot{g})$ is obviously never zero. This 'proves' Theorem 1.5.1.




It is actually possible to prove slightly more than we have stated in Theorem 1.5.1. Suppose that $f(x, y)$ satisfies the conditions in the theorem, and assume in addition that the level curve $\mathcal{C}$ given by $f(x, y) = 0$ is *connected.*

For readers unfamiliar with point set topology, this means roughly that $\mathcal{C}$ is in 'one piece'. For example, the circle $x^2 + y^2 = 1$ is connected, but the hyperbola $x^2 - y^2 = 1$ is not (see above). With these assumptions on $f$, there is a regular parametrized curve $\boldsymbol{\gamma}$ whose image is *the whole* of $\mathcal{C}$. Moreover, if $\mathcal{C}$ is not closed $\boldsymbol{\gamma}$ can be taken to be injective; if $\mathcal{C}$ is closed, then $\boldsymbol{\gamma}$ maps some closed interval $[\alpha, \beta]$ *onto* $\mathcal{C}$, $\boldsymbol{\gamma}(\alpha) = \boldsymbol{\gamma}(\beta)$ and $\boldsymbol{\gamma}$ is injective on the open interval $(\alpha, \beta)$.

A similar argument can be used to pass from parametrized curves to level curves:

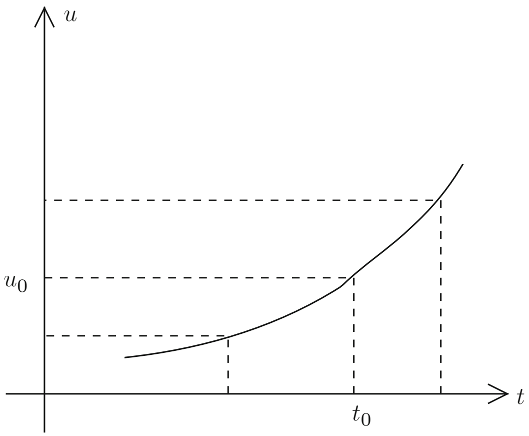
## Theorem 1.5.2

Let $\boldsymbol{\gamma}$ be a regular parametrized plane curve, and let $\boldsymbol{\gamma}(t_0) = (x_0, y_0)$ be a point in the image of $\boldsymbol{\gamma}$. Then, there is a smooth real-valued function $f(x, y)$, defined for $x$ and $y$ in open intervals containing $x_0$ and $y_0$, respectively, and satisfying the conditions in Theorem 1.5.1, such that $\boldsymbol{\gamma}(t)$ is contained in the level curve $f(x, y) = 0$ for all values of $t$ in some open interval containing $t_0$.

The proof of Theorem 1.5.2 is similar to that of Theorem 1.5.1. Let

$$\boldsymbol{\gamma}(t) = (u(t), v(t)),$$

where $u$ and $v$ are smooth functions. Since $\boldsymbol{\gamma}$ is regular, at least one of $\dot{u}(t_0)$ and $\dot{v}(t_0)$ is non-zero, say $\dot{u}(t_0)$. This means that the graph of $u$ as a function of $t$ is not parallel to the $t$-axis at $t_0$:




As in the proof of Theorem 1.5.1, this implies that any line parallel to the $t$-axis close to $u = x_0$ intersects the graph of $u$ at a unique point $u(t)$ with $t$ close to $t_0$. This gives a function $h(x)$, defined for $x$ in an open interval containing

$x_0$, such that $t = h(x)$ is the unique solution of $u(t) = x$ if $x$ is near $x_0$ and $t$ is near $t_0$. The inverse function theorem tells us that $h$ is smooth. The function

$$f(x, y) = y - v(h(x))$$

has the properties we want.

It is not in general possible to find a *single* function $f(x, y)$ satisfying the conditions in Theorem 1.5.1 such that the image of $\boldsymbol{\gamma}$ is contained in the level curve $f(x, y) = 0$, for $\boldsymbol{\gamma}$ may have self-intersections like the limaçon in Example 1.1.7. It follows from the inverse function theorem that no single function $f$ satisfying the conditions in Theorem 1.5.1 can be found that describes a curve near such a self-intersection.

## EXERCISES

1.5.1 Show that the curve $\mathcal{C}$ with Cartesian equation

$$y^2 = x(1 - x^2)$$

is not connected. For what range of values of $t$ is

$$\boldsymbol{\gamma}(t) = (t, \sqrt{t - t^3})$$

a parametrization of $\mathcal{C}$? What is the image of this parametrization?

1.5.2 State an analogue of Theorem 1.5.1 for level curves in $\mathbb{R}^3$ given by $f(x, y, z) = g(x, y, z) = 0$.

1.5.3 State and prove an analogue of Theorem 1.5.2 for curves in $\mathbb{R}^3$ (or even $\mathbb{R}^n$). (This is easy.)

*In the remainder of this book, we shall speak simply of 'curves', unless there is serious danger of confusion as to which type (level or parametrized) is intended.*

# 2
# How much does a curve curve?

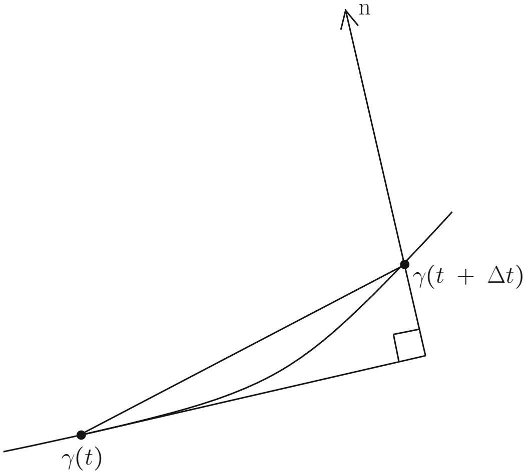
In this chapter, we associate two scalar functions, its curvature and torsion, to any curve in $\mathbb{R}^3$. The curvature measures the extent to which a curve is not contained in a straight line (so that straight lines have zero curvature), and the torsion measures the extent to which a curve is not contained in a plane (so that plane curves have zero torsion). It turns out that the curvature and torsion together determine the shape of a curve.

## 2.1 Curvature

We are going to try to find a measure of how 'curved' a curve is, and to simplify matters we shall work with plane curves initially. Since a straight line should certainly have zero curvature, a measure of the curvature of a plane curve at a point $\mathbf{p}$ of the curve should be its deviation from the tangent line at $\mathbf{p}$.

Suppose then that $\boldsymbol{\gamma}$ is a unit-speed curve in $\mathbb{R}^2$. As the parameter $t$ of $\boldsymbol{\gamma}$ changes to $t + \Delta t$, the curve moves away from its tangent line at $\boldsymbol{\gamma}(t)$ by a distance $(\boldsymbol{\gamma}(t + \Delta t) - \boldsymbol{\gamma}(t)) \cdot \mathbf{n}$, where $\mathbf{n}$ is a unit vector perpendicular to the tangent vector $\dot{\boldsymbol{\gamma}}(t)$ of $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(t)$.

Andr *Elementary Differential Geometry: Second Edition*,
Spr 10.1007/978-1-84882-891-9_2,




By Taylor's theorem,

$$\boldsymbol{\gamma}(t+\Delta t) = \boldsymbol{\gamma}(t) + \dot{\boldsymbol{\gamma}}(t)\Delta t + \frac{1}{2}\ddot{\boldsymbol{\gamma}}(t)(\Delta t)^2 + \text{ remainder}, \tag{2.1}$$

where (remainder)/$(\Delta t)^2$ tends to zero as $\Delta t$ tends to zero. Since $\dot{\boldsymbol{\gamma}} \cdot \mathbf{n} = 0$, the deviation of $\boldsymbol{\gamma}$ from its tangent line at $\boldsymbol{\gamma}(t)$ is

$$\frac{1}{2}\ddot{\boldsymbol{\gamma}}(t) \cdot \mathbf{n}(\Delta t)^2 + \text{ remainder}.$$

Since $\boldsymbol{\gamma}$ is unit-speed, $\ddot{\boldsymbol{\gamma}}$ is perpendicular to $\dot{\boldsymbol{\gamma}}$ and therefore parallel to $\mathbf{n}$. Hence, neglecting the remainder terms, the magnitude of the deviation of $\boldsymbol{\gamma}$ from its tangent line is

$$\frac{1}{2} \parallel \ddot{\boldsymbol{\gamma}}(t) \parallel (\Delta t)^2.$$

This suggests the following definition:

### Definition 2.1.1

If $\boldsymbol{\gamma}$ is a unit-speed curve with parameter $t$, its *curvature* $\kappa(t)$ at the point $\boldsymbol{\gamma}(t)$ is defined to be $\parallel \ddot{\boldsymbol{\gamma}}(t) \parallel$.

Note that we make this definition for unit-speed curves in $\mathbb{R}^n$ for all $n \geq 2$. Note also that this definition is consistent with Proposition 1.1.6, which tells us that if $\ddot{\boldsymbol{\gamma}} = \mathbf{0}$ everywhere then $\boldsymbol{\gamma}$ is part of a straight line, and so should certainly have zero curvature.

Let us see if Definition 2.1.1 is consistent with what we expect for the curvature of circles. Consider the circle in $\mathbb{R}^2$ centred at $(x_0, y_0)$ and of radius $R$. This has a unit-speed parametrization

$$\boldsymbol{\gamma}(t) = \left(x_0 + R\cos\frac{t}{R}, y_0 + R\sin\frac{t}{R}\right).$$

We have $\dot{\boldsymbol{\gamma}}(t) = \left(-\sin\frac{t}{R}, \cos\frac{t}{R}\right)$, and so

$$\| \dot{\boldsymbol{\gamma}}(t) \| = \sqrt{\left(-\sin\frac{t}{R}\right)^2 + \left(\cos\frac{t}{R}\right)^2} = 1,$$

confirming that $\boldsymbol{\gamma}$ is unit-speed, and hence $\ddot{\boldsymbol{\gamma}}(t) = \left(-\frac{1}{R}\cos\frac{t}{R}, -\frac{1}{R}\sin\frac{t}{R}\right)$, so the curvature

$$\| \ddot{\boldsymbol{\gamma}}(t) \| = \sqrt{\left(-\frac{1}{R}\cos\frac{t}{R}\right)^2 + \left(-\frac{1}{R}\sin\frac{t}{R}\right)^2} = \frac{1}{R}$$

is the reciprocal of the radius of the circle. This is in accordance with our expectation that small circles should have large curvature and large circles small curvature.

So far we have only considered unit-speed curves. If $\boldsymbol{\gamma}$ is any *regular* curve, then by Proposition 1.3.6, $\boldsymbol{\gamma}$ has a unit-speed parametrization $\tilde{\boldsymbol{\gamma}}$, say, and we can define the curvature of $\boldsymbol{\gamma}$ to be that of $\tilde{\boldsymbol{\gamma}}$. For this to make sense, we need to know that if $\hat{\boldsymbol{\gamma}}$ is another unit-speed parametrization of $\boldsymbol{\gamma}$, the curvatures of $\tilde{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}$ are the same. To see this, note that $\hat{\boldsymbol{\gamma}}$ will be a reparametrization of $\tilde{\boldsymbol{\gamma}}$ (Exercise 1.3.4), so by Corollary 1.3.7,

$$\tilde{\boldsymbol{\gamma}}(t) = \hat{\boldsymbol{\gamma}}(u),$$

where $u = \pm t + c$ and $c$ is a constant. Then, by the chain rule, $\frac{d\tilde{\boldsymbol{\gamma}}}{dt} = \pm\frac{d\hat{\boldsymbol{\gamma}}}{du}$, so

$$\frac{d^2\tilde{\boldsymbol{\gamma}}}{dt}^2 = \pm\frac{d}{du}\left(\pm\frac{d\hat{\boldsymbol{\gamma}}}{du}\right) = \frac{d^2\hat{\boldsymbol{\gamma}}}{du^2},$$

which shows that $\tilde{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}$ do indeed have the same curvature.

Although every regular curve $\boldsymbol{\gamma}$ has a unit-speed reparametrization, it may be complicated or impossible to write it down *explicitly* (see Examples 1.3.8 and 1.3.9), and so it is desirable to have a formula for the curvature of $\boldsymbol{\gamma}$ in terms of $\boldsymbol{\gamma}$ itself rather than a reparametrization of it.

## Proposition 2.1.2

Let $\boldsymbol{\gamma}(t)$ be a regular curve in $\mathbb{R}^3$. Then, its curvature is

$$\kappa = \frac{\| \ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}} \|}{\| \dot{\boldsymbol{\gamma}} \|^3}, \tag{2.2}$$

where the $\times$ indicates the vector (or cross) product and the dot denotes $d/dt$.

Of course, since a curve in $\mathbb{R}^2$ can be viewed as a curve in the $xy$-plane (say) in $\mathbb{R}^3$, Eq. 2.2 can also be used to calculate the curvature of plane curves.

## Proof

Let $s$ be a unit-speed parameter for $\boldsymbol{\gamma}$. Then, by the chain rule,

$$\dot{\boldsymbol{\gamma}} = \frac{d\boldsymbol{\gamma}}{dt} = \frac{d\boldsymbol{\gamma}}{ds}\frac{ds}{dt},$$

so

$$\kappa = \left\| \frac{d^2\boldsymbol{\gamma}}{ds^2} \right\| = \left\| \frac{d}{ds}\left(\frac{d\boldsymbol{\gamma}/dt}{ds/dt}\right) \right\| = \left\| \frac{\frac{d}{dt}\left(\frac{d\boldsymbol{\gamma}/dt}{ds/dt}\right)}{ds/dt} \right\| = \left\| \frac{\frac{ds}{dt}\frac{d^2\boldsymbol{\gamma}}{dt^2} - \frac{d^2s}{dt^2}\frac{d\boldsymbol{\gamma}}{dt}}{(ds/dt)^3} \right\|. \quad (2.3)$$

Now,

$$\left(\frac{ds}{dt}\right)^2 = \| \dot{\boldsymbol{\gamma}} \|^2 = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}},$$

and differentiating with respect to $t$ gives

$$\frac{ds}{dt}\frac{d^{2s}}{dt^2} = \dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}}.$$

Using this and Eq. 2.3, we get

$$\kappa = \left\| \frac{\left(\frac{ds}{dt}\right)^2 \ddot{\boldsymbol{\gamma}} - \frac{d^2s}{dt^2}\frac{ds}{dt}\dot{\boldsymbol{\gamma}}}{(ds/dt)^4} \right\| = \frac{\| (\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}})\ddot{\boldsymbol{\gamma}} - (\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}})\dot{\boldsymbol{\gamma}} \|}{\| \dot{\boldsymbol{\gamma}} \|^4}.$$

Using the vector triple product identity

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$$

(where $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$), we get

$$\dot{\boldsymbol{\gamma}} \times (\ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}}) = (\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}})\ddot{\boldsymbol{\gamma}} - (\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}})\dot{\boldsymbol{\gamma}}.$$

Further, $\dot{\boldsymbol{\gamma}}$ and $\ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}}$ are perpendicular vectors, so

$$\| \dot{\boldsymbol{\gamma}} \times (\ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}}) \| = \| \dot{\boldsymbol{\gamma}} \| \| \ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}} \|.$$

Hence,

$$\frac{\| (\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}})\ddot{\boldsymbol{\gamma}} - (\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}})\dot{\boldsymbol{\gamma}} \|}{\| \dot{\boldsymbol{\gamma}} \|^4} = \frac{\| \dot{\boldsymbol{\gamma}} \times (\ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}}) \|}{\| \dot{\boldsymbol{\gamma}} \|^4} = \frac{\| \dot{\boldsymbol{\gamma}} \| \| \ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}} \|}{\| \dot{\boldsymbol{\gamma}} \|^4} = \frac{\| \ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}} \|}{\| \dot{\boldsymbol{\gamma}} \|^3}. \quad \square$$

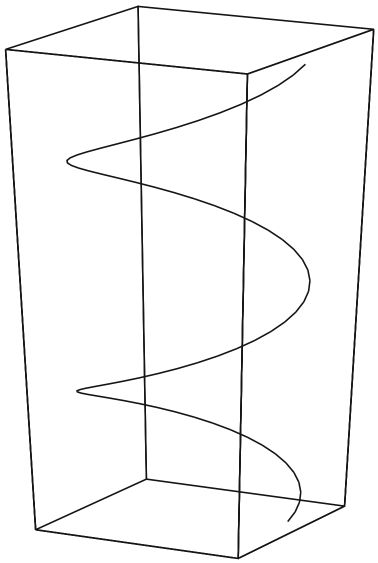
Note that formula (2.2) makes sense provided that $\dot{\boldsymbol{\gamma}} \neq \mathbf{0}$. Thus, the curvature is defined at all regular points of the curve.

### Example 2.1.3

A *circular helix* with axis the $z$-axis is a curve of the form

$$\boldsymbol{\gamma}(\theta) = (a\cos\theta, a\sin\theta, b\theta), \quad \theta \in \mathbb{R},$$

where $a$ and $b$ are constants.



If $(x, y, z)$ is a point on the helix, so that

$$x = a\cos\theta, \ y = a\sin\theta, \ z = b\theta,$$

for some value of $\theta$, then $x^2+y^2 = a^2$, showing that the helix lies on the cylinder with axis the $z$-axis and radius $|a|$; the positive number $|a|$ is called the *radius* of the helix. As $\theta$ increases by $2\pi$, the point $(a\cos\theta, a\sin\theta, b\theta)$ rotates once round the $z$-axis and moves parallel to the $z$-axis by $2\pi b$; the positive number $2\pi|b|$ is called the *pitch* of the helix.

Let us compute the curvature of the helix using the formula in Proposition 2.1.2. Denoting $d/d\theta$ by a dot, we have $\dot{\boldsymbol{\gamma}}(\theta) = (-a\sin\theta, a\cos\theta, b)$ so

$$\| \dot{\boldsymbol{\gamma}}(\theta) \| = \sqrt{a^2 + b^2}.$$

This shows that $\dot{\boldsymbol{\gamma}}(\theta)$ is never zero, so $\boldsymbol{\gamma}$ is regular (unless $a = b = 0$, in which case the image of the helix is a single point). Hence, the formula in

Proposition 2.1.2 applies, and we have $\ddot{\gamma} = (-a\cos\theta, -a\sin\theta, 0)$ so $\ddot{\gamma} \times \dot{\gamma} = (-ab\sin\theta, ab\cos\theta, -a^2)$ and hence

$$\kappa = \frac{\| (-ab\sin\theta, ab\cos\theta, -a^2) \|}{\| (-a\sin\theta, a\cos\theta, b) \|^3} = \frac{(a^2b^2 + a^4)^{1/2}}{(a^2+b^2)^{3/2}} = \frac{|a|}{a^2+b^2}. \tag{2.4}$$

Thus, the curvature of the helix is constant.

Let us examine some limiting cases to see if this result agrees with what we already know. First, suppose that $b = 0$ (but $a \neq 0$). Then, the helix is simply a circle in the $xy$-plane of radius $|a|$, so by the calculation following Definition 2.1.1 its curvature is $1/|a|$. On the other hand, the formula (2.4) gives the curvature as

$$\frac{|a|}{a^2+0^2} = \frac{|a|}{a^2} = \frac{|a|}{|a|^2} = \frac{1}{|a|}.$$

Next, suppose that $a = 0$ (but $b \neq 0$). Then, the image of the helix is just the $z$-axis, a straight line, so the curvature is zero. And formula (2.4) gives zero when $a = 0$ too.

## *EXERCISES*

2.1.1 Compute the curvature of the following curves:

(i) $\gamma(t) = \left(\frac{1}{3}(1+t)^{3/2} \frac{1}{3}(1-t)^{3/2}, \frac{t}{\sqrt{2}}\right)$.

(ii) $\gamma(t) = \left(\frac{4}{5}\cos t, 1 - \sin t, -\frac{3}{5}\cos t\right)$.

(iii) $\gamma(t) = (t, \cosh t)$.

(iv) $\gamma(t) = (\cos^3 t, \sin^3 t)$.

For the astroid in (iv), show that the curvature tends to $\infty$ as we approach one of the points $(\pm 1, 0)$, $(0, \pm 1)$. Compare with the sketch found in Exercise 1.1.5.

2.1.2 Show that, if the curvature $\kappa(t)$ of a regular curve $\gamma(t)$ is $> 0$ everywhere, then $\kappa(t)$ is a smooth function of $t$. Give an example to show that this may not be the case without the assumption that $\kappa > 0$.

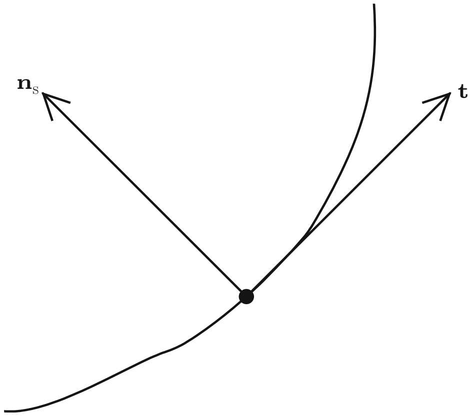
## 2.2 Plane curves

For plane curves, it is possible to refine the definition of curvature slightly and give it an appealing geometric interpretation.

Suppose that $\boldsymbol{\gamma}(s)$ is a unit-speed curve in $\mathbb{R}^2$. Denoting $d/ds$ by a dot, let

$$\mathbf{t} = \dot{\boldsymbol{\gamma}}$$

be the tangent vector of $\boldsymbol{\gamma}$; note that $\mathbf{t}$ is a unit vector. There are two unit vectors perpendicular to $\mathbf{t}$; we make a choice by defining $\mathbf{n}_s$, the *signed unit normal* of $\boldsymbol{\gamma}$, to be the unit vector obtained by rotating $\mathbf{t}$ anticlockwise by $\pi/2$.



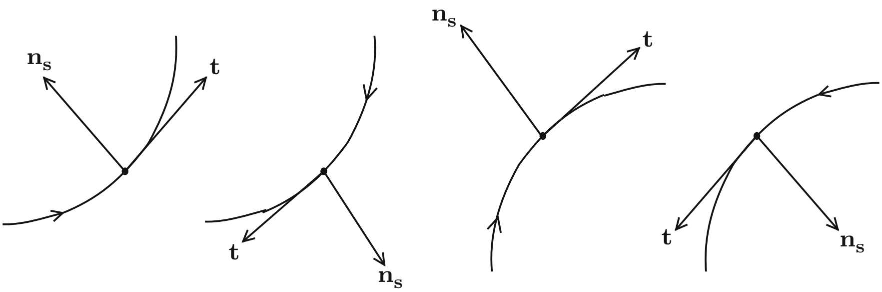

By Proposition 1.2.4, $\dot{\mathbf{t}} = \ddot{\boldsymbol{\gamma}}$ is perpendicular to $\mathbf{t}$, and hence parallel to $\mathbf{n}_s$. Thus, there is a scalar $\kappa_s$ such that

$$\ddot{\boldsymbol{\gamma}} = \kappa_s \mathbf{n}_s;$$

$\kappa_s$ is called the *signed curvature* of $\boldsymbol{\gamma}$ (it can be positive, negative or zero). Note that, since $\| \mathbf{n}_s \| = 1$, we have

$$\kappa = \| \ddot{\boldsymbol{\gamma}} \| = \| \kappa_s \mathbf{n}_s \| = |\kappa_s|, \tag{2.5}$$

so the curvature of $\boldsymbol{\gamma}$ is the absolute value of its signed curvature. The following diagrams show how the sign of the signed curvature is determined (in each case, the arrow on the curve indicates the direction of increasing $s$); $\kappa_s$ is negative for the two middle diagrams and positive for the other two.




$\kappa_s > 0$ $\kappa_s < 0$ $\kappa_s < 0$ $\kappa_s > 0$

I $\boldsymbol{\gamma}(t)$ e
t $\mathbf{t}$, s $\mathbf{n}_s$ and s $\kappa_s$ of $\boldsymbol{\gamma}$ t
t $\tilde{\boldsymbol{\gamma}}(s)$ $s$ is $\boldsymbol{\gamma}$.
Thus

$$\mathbf{t} = \frac{d\boldsymbol{\gamma}/dt}{ds/dt} = \frac{d\boldsymbol{\gamma}/dt}{\parallel d\boldsymbol{\gamma}/dt \parallel},$$

$\mathbf{n}_s$ is $\mathbf{t}$ ant $\pi/2$, and

$$\frac{d\mathbf{t}}{dt} = \frac{d\mathbf{t}}{ds}\frac{ds}{dt} = \kappa_s \frac{ds}{dt}\mathbf{n}_s = \kappa_s \left\| \frac{d\boldsymbol{\gamma}}{dt} \right\| \mathbf{n}_s.$$

The in t
r $\boldsymbol{\gamma}$ is
of $\dot{\boldsymbol{\gamma}}(s)$ $\varphi(s)$

$$\dot{\boldsymbol{\gamma}}(s) \quad \varphi(s), \mathrm{s} \quad \varphi(s) \ . \qquad (\ $$

The $\varphi(s)$ e
any int $\pi$. The
a *smooth* c

## Proposition 2.2.1

Le $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^2$ be $s_0 \in (\alpha, \beta)$ $\varphi_0$ be
t

$$\dot{\boldsymbol{\gamma}}(s_0) \quad \varphi_0, \mathrm{s} \quad \varphi_0).$$

The $\varphi : (\alpha, \beta) \to \mathbb{R}$ s $\varphi(s_0)$ $\varphi_0$
and t 2.6 holds $s \in (\alpha, \beta)$

## Proof

Le

$$\dot{\boldsymbol{\gamma}}(s) \quad f(s), g(s)$$

not

$$f(s)^2 + g(s)^2 = 1 \ \ \mathrm{f} \quad s \qquad (\ $$

s $\boldsymbol{\gamma}$ is

$$\varphi(s) \quad \varphi_0 + \int_{s_0}^{s} (f\dot{g} - g\dot{f})\, dt.$$

Obvious $\varphi(s_0)$ $\varphi_0$. Mor $f$ and $g$ ar
$\dot{\varphi} = f\dot{g} - g\dot{f}$, and he $\varphi$.

Let

$$F = f\cos\varphi + g\sin\varphi, \qquad G = f\sin\varphi - g\cos\varphi.$$

Then,

$$\dot{F} = (\dot{f} + g\dot{\varphi})\cos\varphi + (\dot{g} - f\dot{\varphi})\sin\varphi.$$

But

$$\dot{f} + g\dot{\varphi} = \dot{f}(1 - g^2) + fg\dot{g} = f(f\dot{f} + g\dot{g}) = 0,$$

where the second equality used Eq. 2.7 and the last equality used its consequence

$$f\dot{f} + g\dot{g} = 0.$$

Similarly, $\dot{g} - f\dot{\varphi} = 0$. Hence, $\dot{F} = 0$ and $F$ is constant. A similar argument shows that $G$ is constant. But

$$F(s_0) = f(s_0)\cos\varphi_0 + g(s_0)\sin\varphi_0 = \cos^2\varphi_0 + \sin^2\varphi_0 = 1,$$

and similarly $G(s_0) = 0$. It follows that

$$f\cos\varphi + g\sin\varphi = 1, \qquad f\sin\varphi - g\cos\varphi = 0$$

for all $s$. These equations imply that $f = \cos\varphi$, $g = \sin\varphi$, and hence that the smooth function $\varphi$ satisfies Eq. 2.6.

As to the uniqueness, if $\psi$ is another smooth function such that $\psi(s_0) = \varphi_0$ and $\dot{\boldsymbol{\gamma}}(s) = (\cos\psi(s), \sin\psi(s))$ for $s \in (\alpha, \beta)$, there is an integer $n(s)$ such that

$$\psi(s) - \varphi(s) = 2\pi n(s) \qquad \text{for all } s \in (\alpha, \beta).$$

Because $\varphi$ and $\psi$ are smooth, $n$ is a smooth, hence continuous, function of $s$. This implies that $n$ is a constant: otherwise we would have $n(s_0) \neq n(s_1)$ for some $s_1 \in (\alpha, \beta)$, and then by the intermediate value theorem the continuous function $n(s)$ would have to take all values between $n(s_0)$ and $n(s_1)$ when $s$ is between $s_0$ and $s_1$. But most real numbers between $n(s_0)$ and $n(s_1)$ are not integers! Thus, $n$ is actually independent of $s$, and since $\psi(s_0) = \varphi(s_0) = \varphi_0$, we must have $n = 0$ and hence $\psi(s) = \varphi(s)$ for all $s \in (\alpha, \beta)$. $\square$

## Definition 2.2.2

The smooth function $\varphi$ in Proposition 2.2.1 is called the *turning angle* of $\boldsymbol{\gamma}$ determined by the condition $\varphi(s_0) = \varphi_0$.

We are now in a position to give the geometric interpretation of the signed curvature that we promised earlier.

### Proposition 2.2.3

Let $\boldsymbol{\gamma}(s)$ be a unit-speed plane curve, and let $\varphi(s)$ be a turning angle for $\boldsymbol{\gamma}$. Then,

$$\kappa_s = \frac{d\varphi}{ds}.$$

Thus, *the signed curvature is the rate at which the tangent vector of the curve rotates.* As the diagrams following Eq. 2.5 show, the signed curvature is positive or negative accordingly as $\mathbf{t}$ rotates anticlockwise or clockwise as one moves along the curve in the direction of increasing $s$.

### Proof

By Eq. 2.6 the tangent vector $\mathbf{t} = (\cos\varphi, \sin\varphi)$, so

$$\dot{\mathbf{t}} = \dot{\varphi}(-\sin\varphi, \cos\varphi).$$

Since $\mathbf{n}_s = (-\sin\varphi, \cos\varphi)$, the equation $\dot{\mathbf{t}} = \kappa_s \mathbf{n}_s$ gives the stated result. □

### Example 2.2.4

Let us find the signed curvature of the catenary (Exercise 1.2.1). Using the parametrization $\boldsymbol{\gamma}(t) = (t, \cosh t)$ we get $\dot{\boldsymbol{\gamma}} = (1, \sinh t)$ and hence

$$s = \int_0^t \sqrt{1 + \sinh^2 t}\, dt = \sinh t,$$

so if $\varphi$ is the angle between $\dot{\boldsymbol{\gamma}}$ and the $x$-axis,

$$\begin{aligned} \tan\varphi = \sinh t \quad &= \quad s, \\ \therefore \qquad \sec^2\varphi \frac{d\varphi}{ds} \quad &= \quad 1, \\ \therefore \quad \kappa_s = \frac{d\varphi}{ds} = \frac{1}{\sec^2\varphi} \quad &= \quad \frac{1}{1 + \tan^2\varphi} = \frac{1}{1 + s^2}. \end{aligned}$$

Proposition 2.2.3 has an interesting consequence in terms of the *total signed curvature* of a unit-speed *closed* curve $\boldsymbol{\gamma}$ of length $\ell$, namely

$$\int_0^\ell \kappa_s(s)\, ds. \tag{2.8}$$

### Corollary 2.2.5

The total signed curvature of a closed plane curve is an integer multiple of $2\pi$.

### Proof

Let $\boldsymbol{\gamma}$ be a unit-speed closed plane curve and let $\ell$ be its length. By Proposition 2.2.3, the total signed curvature of $\boldsymbol{\gamma}$ is

$$\int_0^\ell \frac{d\varphi}{ds}\,ds = \varphi(\ell) - \varphi(0),$$

where $\varphi$ is a turning angle for $\boldsymbol{\gamma}$. Now, $\boldsymbol{\gamma}$ is $l$-periodic (see Section 1.4):

$$\boldsymbol{\gamma}(s+\ell) = \boldsymbol{\gamma}(s).$$

Differentiating both sides gives

$$\dot{\boldsymbol{\gamma}}(s+\ell) = \dot{\boldsymbol{\gamma}}(s),$$

and in particular $\dot{\boldsymbol{\gamma}}(\ell) = \dot{\boldsymbol{\gamma}}(0)$. Hence, by Eq. 2.6,

$$(\cos\varphi(\ell), \sin\varphi(\ell)) = (\cos\varphi(0), \sin\varphi(0)),$$

which implies that $\varphi(\ell) - \varphi(0)$ is an integer multiple of $2\pi$. □

The next result shows that a unit-speed plane curve is essentially determined once we know its signed curvature at each point of the curve. The meaning of 'essentially' here is 'up to a direct isometry of $\mathbb{R}^2$', i.e., a map $M : \mathbb{R}^2 \to \mathbb{R}^2$ of the form

$$M = T_{\mathbf{a}} \circ \rho_\theta,$$

where $\rho_\theta$ is an anticlockwise rotation by an angle $\theta$ about the origin,

$$\rho_\theta(x, y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta),$$

and $T_{\mathbf{a}}$ is the translation by the vector $\mathbf{a}$,

$$T_{\mathbf{a}}(\mathbf{v}) = \mathbf{v} + \mathbf{a},$$

for any vectors $(x, y)$ and $\mathbf{v}$ in $\mathbb{R}^2$ (see Appendix 1).

### Theorem 2.2.6

Let $k : (\alpha, \beta) \to \mathbb{R}$ be any smooth function. Then, there is a unit-speed curve $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^2$ whose signed curvature is $k$.

Further, if $\tilde{\boldsymbol{\gamma}} : (\alpha, \beta) \to \mathbb{R}^2$ is any other unit-speed curve whose signed curvature is $k$, there is a direct isometry $M$ of $\mathbb{R}^2$ such that

$$\tilde{\boldsymbol{\gamma}}(s) = M(\boldsymbol{\gamma}(s)) \quad \text{for all } s \in (\alpha, \beta).$$

## Proof

For the first part, fix $s_0 \in (\alpha, \beta)$ and define, for any $s \in (\alpha, \beta)$,

$$\begin{aligned} \varphi(s) &= \int_{s_0}^{s} k(u)du, \quad \text{(cf. Proposition 2.2.3)}, \\ \boldsymbol{\gamma}(s) &= \left( \int_{s_0}^{s} \cos\varphi(t)dt, \int_{s_0}^{s} \sin\varphi(t)dt \right). \end{aligned}$$

Then, the tangent vector of $\boldsymbol{\gamma}$ is

$$\dot{\boldsymbol{\gamma}}(s) = (\cos\varphi(s), \sin\varphi(s)),$$

which is a unit vector making an angle $\varphi(s)$ with the $x$-axis. Thus, $\boldsymbol{\gamma}$ is unit-speed and, by Proposition 2.2.3, its signed curvature is

$$\frac{d\varphi}{ds} = \frac{d}{ds} \int_{s_0}^{s} k(u)du = k(s).$$

For the second part, let $\tilde{\varphi}(s)$ be a smooth turning angle for $\tilde{\boldsymbol{\gamma}}$. Thus,

$$\begin{aligned} \dot{\tilde{\boldsymbol{\gamma}}}(s) &= (\cos\tilde{\varphi}(s), \sin\tilde{\varphi}(s)), \\ \therefore \quad \tilde{\boldsymbol{\gamma}}(s) &= \left( \int_{s_0}^{s} \cos\tilde{\varphi}(t)dt, \int_{s_0}^{s} \sin\tilde{\varphi}(t)dt \right) + \tilde{\boldsymbol{\gamma}}(s_0). \end{aligned} \tag{2.9}$$

By Proposition 2.2.3, $k(s) = d\tilde{\varphi}/ds$ so

$$\tilde{\varphi}(s) = \int_{s_0}^{s} k(u)du + \tilde{\varphi}(s_0).$$

Inserting this into Eq. 2.9, and writing $\mathbf{a}$ for the constant vector $\tilde{\boldsymbol{\gamma}}(s_0)$ and $\theta$ for the constant scalar $\tilde{\varphi}(s_0)$, we get

$$\begin{aligned} \tilde{\boldsymbol{\gamma}}(s) &= T_{\mathbf{a}} \left( \int_{s_0}^{s} \cos(\varphi(t) + \theta)dt, \int_{s_0}^{s} \sin(\varphi(t) + \theta)dt \right) \\ &= T_{\mathbf{a}} \left( \cos\theta \int_{s_0}^{s} \cos\varphi(t)dt - \sin\theta \int_{s_0}^{s} \sin\varphi(t)dt, \right. \\ &\qquad \left. \sin\theta \int_{s_0}^{s} \cos\varphi(t)dt + \cos\theta \int_{s_0}^{s} \sin\varphi(t)dt \right) \\ &= T_{\mathbf{a}}\rho_\theta \left( \int_{s_0}^{s} \cos\varphi(t)dt, \int_{s_0}^{s} \sin\varphi(t)dt \right) \\ &= T_{\mathbf{a}}\rho_\theta(\boldsymbol{\gamma}(s)). \end{aligned}$$

$\square$

## Example 2.2.7

Any regular plane curve $\boldsymbol{\gamma}$ whose curvature is a positive *constant* is part of a circle. To see this, let $\kappa$ be the curvature of $\boldsymbol{\gamma}$, and let $\kappa_s$ be its signed curvature. Then, by Eq. 2.5,

$$\kappa_s = \pm\kappa.$$

A priori, we could have $\kappa_s = \kappa$ at some points of the curve and $\kappa_s = -\kappa$ at others, but in fact this cannot happen since $\kappa_s$ is a continuous function of $s$ (see Exercise 2.2.2), so the intermediate value theorem tells us that, if $\kappa_s$ takes both the value $\kappa$ and the value $-\kappa$, it must take all values between. Thus, either $\kappa_s = \kappa$ at all points of the curve, or $\kappa_s = -\kappa$ at all points of the curve. In particular, $\kappa_s$ is constant.

The idea now is to show that, whatever the value of $\kappa_s$, we can find a parametrized circle whose signed curvature is $\kappa_s$. The theorem then tells us that *every* curve whose signed curvature is $\kappa_s$ can be obtained by applying a direct isometry to this circle. Since rotations and translations obviously take circles to circles, it follows that *every* curve whose signed curvature is constant is (part of) a circle.

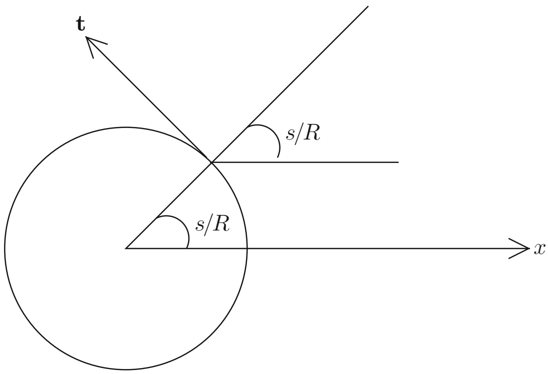
A unit-speed parametrization of the circle with centre the origin and radius $R$ is

$$\boldsymbol{\gamma}(s) = \left(R\cos\frac{s}{R}, R\sin\frac{s}{R}\right).$$

Its tangent vector

$$\mathbf{t} = \dot{\boldsymbol{\gamma}} = \left(-\sin\frac{s}{R}, \cos\frac{s}{R}\right)$$

is the unit vector making an angle $\pi/2 + s/R$ with the positive $x$-axis:




Hence, the signed curvature of $\boldsymbol{\gamma}$ is

$$\frac{d}{ds}\left(\frac{\pi}{2} + \frac{s}{R}\right) = \frac{1}{R}.$$

Thus, if $\kappa_s > 0$, the circle of radius $1/\kappa_s$ has signed curvature $\kappa_s$.

If $\kappa_s < 0$, it is easy to check that the curve

$$\tilde{\gamma}(s) = \left(R\cos\frac{s}{R}, -R\sin\frac{s}{R}\right)$$

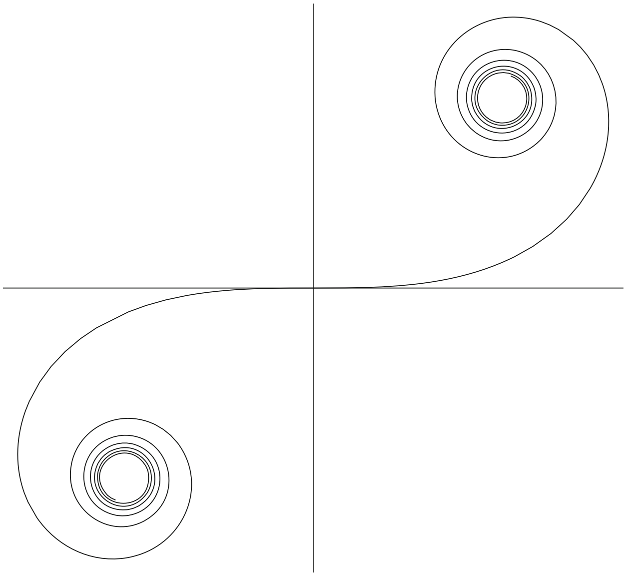
(which is just another parametrization of the circle with centre the origin and radius $R$) has signed curvature $-1/R$. Thus, if $R = -1/\kappa_s$ we again get a circle with signed curvature $\kappa_s$.

These calculations should be compared to the analogous ones for curvature (as opposed to signed curvature) following Definition 2.1.1.

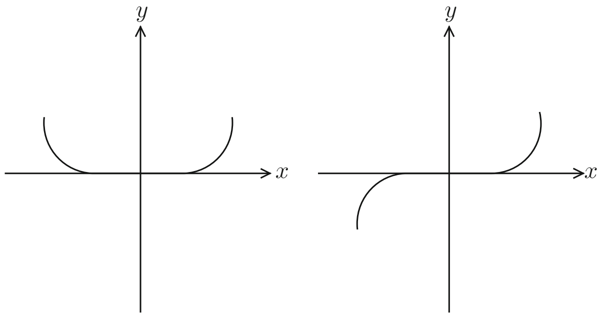
### Example 2.2.8

Theorem 2.2.6 shows that we can find a plane curve with any given smooth function as its signed curvature. But simple curvatures can lead to complicated curves. For example, let the signed curvature be $\kappa_s(s) = s$. Following the proof of Theorem 2.2.6, and taking $s_0 = 0$, we get $\varphi(s) = \int_0^s u\,du = \frac{s^2}{2}$ so

$$\boldsymbol{\gamma}(s) = \left(\int_0^s \cos\left(\frac{t^2}{2}\right)dt, \int_0^s \sin\left(\frac{t^2}{2}\right)dt\right).$$



These integrals cannot be evaluated in terms of 'elementary' functions. (They arise in the theory of diffraction of light, where they are called *Fresnel's integrals*, and the curve $\boldsymbol{\gamma}$ is called *Cornu's Spiral*, although it was first considered by Euler.) The picture of $\boldsymbol{\gamma}$ above is obtained by computing the integrals numerically.




It is natural to ask whether Theorem 2.2.6 remains true if we replace 'signed curvature' by 'curvature'. The first part holds if (and only if) we assume that $k \geq 0$, for then $\boldsymbol{\gamma}$ can be chosen to have signed curvature $k$ and so will have curvature $k$ as well. The second part of Theorem 2.2.6, however, no longer holds. For, we can take a (smooth) curve $\boldsymbol{\gamma}$ that coincides with the $x$-axis for $-1 \leq x \leq 1$ (say), and is otherwise above the $x$-axis. (The reader who wishes to write down such a curve explicitly will find the solution of Exercise 9.4.3 helpful.) We now reflect the part of the curve with $x \leq 0$ in the $x$-axis. The new curve has the same curvature as $\boldsymbol{\gamma}$, but obviously cannot be obtained by applying an isometry to $\boldsymbol{\gamma}$. See Exercise 2.2.3 for a version of Theorem 2.2.6 that *is* valid for curvature instead of signed curvature.

## *EXERCISES*

2.2.1 Show that, if $\boldsymbol{\gamma}$ is a unit-speed plane curve,

$$\dot{\mathbf{n}}_s = -\kappa_s \mathbf{t}.$$

2.2.2 Show that the signed curvature of any regular plane curve $\boldsymbol{\gamma}(t)$ is a smooth function of $t$. (Compare with Exercise 2.1.2.)

2.2.3 Let $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ be two plane curves. Show that, if $\tilde{\boldsymbol{\gamma}}$ is obtained from $\boldsymbol{\gamma}$ by applying an isometry $M$ of $\mathbb{R}^2$, the signed curvatures $\kappa_s$ and $\tilde{\kappa}_s$ of $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ are equal if $M$ is direct but that $\tilde{\kappa}_s = -\kappa_s$ if $M$ is opposite (in particular, $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ have the same curvature). Show, conversely, that if $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ have the same nowhere-vanishing curvature, then $\tilde{\boldsymbol{\gamma}}$ can be obtained from $\boldsymbol{\gamma}$ by applying an isometry of $\mathbb{R}^2$.

2.2.4 Let $k$ be the signed curvature of a plane curve $\mathcal{C}$ expressed in terms of its arc-length. Show that, if $\mathcal{C}_a$ is the image of $\mathcal{C}$ under the dilation

$\mathbf{v} \mapsto a\mathbf{v}$ of the plane (where $a$ is a non-zero constant), the signed curvature of $\mathcal{C}_a$ in terms of *its* arc-length $s$ is $\frac{1}{a}k(\frac{s}{a})$.

A heavy chain suspended at its ends hanging loosely takes the form of a plane curve $\mathcal{C}$. Show that, if $s$ is the arc-length of $\mathcal{C}$ measured from its lowest point, $\varphi$ the angle between the tangent of $\mathcal{C}$ and the horizontal, and $T$ the tension in the chain, then

$$T \cos\varphi = \lambda, \quad T \sin\varphi = \mu s,$$

where $\lambda, \mu$ are non-zero constants (we assume that the chain has constant mass per unit length). Show that the signed curvature of $\mathcal{C}$ is

$$\kappa_s = \frac{1}{a}\left(1 + \frac{s^2}{a^2}\right)^{-1},$$

where $a = \lambda/\mu$, and deduce that $\mathcal{C}$ can be obtained from the catenary in Example 2.2.4 by applying a dilation and an isometry of the plane.

2.2.5 Let $\boldsymbol{\gamma}(t)$ be a regular plane curve and let $\lambda$ be a constant. The *parallel curve* $\boldsymbol{\gamma}^\lambda$ of $\boldsymbol{\gamma}$ is defined by

$$\boldsymbol{\gamma}^\lambda(t) = \boldsymbol{\gamma}(t) + \lambda \mathbf{n}_s(t).$$

Show that, if $\lambda\kappa_s(t) \neq 1$ for all values of $t$, then $\boldsymbol{\gamma}^\lambda$ is a regular curve and that its signed curvature is $\kappa_s/|1 - \lambda\kappa_s|$.

2.2.6 Another approach to the curvature of a unit-speed plane curve $\boldsymbol{\gamma}$ at a point $\boldsymbol{\gamma}(s_0)$ is to look for the 'best approximating circle' at this point. We can then *define* the curvature of $\boldsymbol{\gamma}$ to be the reciprocal of the radius of this circle.

Carry out this programme by showing that the centre of the circle which passes through three nearby points $\boldsymbol{\gamma}(s_0)$ and $\boldsymbol{\gamma}(s_0 \pm \delta s)$ on $\boldsymbol{\gamma}$ approaches the point

$$\boldsymbol{\epsilon}(s_0) = \boldsymbol{\gamma}(s_0) + \frac{1}{\kappa_s(s_0)}\mathbf{n}_s(s_0)$$

as $\delta s$ tends to zero. The circle $\mathcal{C}$ with centre $\boldsymbol{\epsilon}(s_0)$ passing through $\boldsymbol{\gamma}(s_0)$ is called the *osculating circle* to $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(s_0)$, and $\boldsymbol{\epsilon}(s_0)$ is called the *centre of curvature* of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(s_0)$. The radius of $\mathcal{C}$ is $1/|\kappa_s(s_0)| = 1/\kappa(s_0)$, where $\kappa$ is the curvature of $\boldsymbol{\gamma}$ – this is called the *radius of curvature* of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(s_0)$.

2.2.7 With the notation in the preceding exercise, we regard $\boldsymbol{\epsilon}$ as the parametrization of a new curve, called the *evolute* of $\boldsymbol{\gamma}$ (if $\boldsymbol{\gamma}$ is any regular plane curve, its evolute is defined to be that of a unit-speed reparametrization of $\boldsymbol{\gamma}$). Assume that $\dot{\kappa}_s(s) \neq 0$ for all values of $s$ (a dot denoting $d/ds$), say $\dot{\kappa}_s > 0$ for all $s$ (this can be achieved by replacing $s$ by $-s$ if necessary). Show that the arc-length of $\boldsymbol{\epsilon}$ is $-\frac{1}{\kappa_s(s)}$ (up to adding a constant), and calculate the signed curvature of $\boldsymbol{\epsilon}$. Show also that all the normal lines to $\boldsymbol{\gamma}$ are tangent to $\boldsymbol{\epsilon}$ (for this reason, the evolute of $\boldsymbol{\gamma}$ is sometimes described as the 'envelope' of the normal lines to $\boldsymbol{\gamma}$). Show that the evolute of the cycloid

$$\boldsymbol{\gamma}(t) = a(t - \sin t, 1 - \cos t), \qquad 0 < t < 2\pi,$$

where $a > 0$ is a constant, is

$$\boldsymbol{\epsilon}(t) = a(t + \sin t, -1 + \cos t)$$

(see Exercise 1.1.7) and that, after a suitable reparametrization, $\boldsymbol{\epsilon}$ can be obtained from $\boldsymbol{\gamma}$ by a translation of the plane.

2.2.8 A string of length $\ell$ is attached to the point $\boldsymbol{\gamma}(0)$ of a unit-speed plane curve $\boldsymbol{\gamma}(s)$. Show that when the string is wound onto the curve while being kept taught, its endpoint traces out the curve

$$\boldsymbol{\iota}(s) = \boldsymbol{\gamma}(s) + (\ell - s)\dot{\boldsymbol{\gamma}}(s),$$

where $0 < s < \ell$ and a dot denotes $d/ds$. The curve $\boldsymbol{\iota}$ is called the *involute* of $\boldsymbol{\gamma}$ (if $\boldsymbol{\gamma}$ is any regular plane curve, we define its involute to be that of a unit-speed reparametrization of $\boldsymbol{\gamma}$). Suppose that the signed curvature $\kappa_s$ of $\boldsymbol{\gamma}$ is never zero, say $\kappa_s(s) > 0$ for all $s$. Show that the signed curvature of $\boldsymbol{\iota}$ is $1/(\ell - s)$.

2.2.9 Show that the involute of the catenary

$$\boldsymbol{\gamma}(t) = (t, \cosh t)$$

with $l = 0$ (see the preceding exercise) is the *tractrix*

$$x = \cosh^{-1}\left(\frac{1}{y}\right) - \sqrt{1 - y^2}.$$

See Section 8.3 for a simple geometric characterization of this curve.

2.2.10 A unit-speed plane curve $\boldsymbol{\gamma}(s)$ rolls without slipping along a straight line $\ell$ parallel to a unit vector $\mathbf{a}$, and initially touches $\ell$ at a point $\mathbf{p} = \boldsymbol{\gamma}(0)$. Let $\mathbf{q}$ be a point fixed *relative to* $\boldsymbol{\gamma}$. Let $\boldsymbol{\Gamma}(s)$ be the point to which $\mathbf{q}$ has moved when $\boldsymbol{\gamma}$ has rolled a distance $s$

along $\ell$ (note that $\mathbf{\Gamma}$ will not usually be unit-speed). Let $\theta(s)$ be the angle between $\mathbf{a}$ and the tangent vector $\dot{\boldsymbol{\gamma}}$. Show that

$$\mathbf{\Gamma}(s) = \mathbf{p} + s\mathbf{a} + \rho_{-\theta(s)}(\mathbf{q} - \boldsymbol{\gamma}(s)),$$

where $\rho_\varphi$ is the rotation about the origin through an angle $\varphi$. Show further that

$$\dot{\mathbf{\Gamma}}(s) \cdot \rho_{-\theta(s)}(\mathbf{q} - \boldsymbol{\gamma}(s)) = 0.$$

Geometrically, this means that a point on $\mathbf{\Gamma}$ moves as if it is rotating about the instantaneous point of contact of the rolling curve with $\ell$. See Exercise 1.1.7 for a special case.

## 2.3 Space curves

Our main interest in this book is in curves (and surfaces) in $\mathbb{R}^3$, i.e., space curves. While a plane curve is essentially determined by its curvature (see Theorem 2.2.6), this is no longer true for space curves. For example, a circle of radius 1 in the $xy$-plane and a circular helix with $a = b = 1/2$ (see Example 2.1.3) both have curvature 1 everywhere, but it is obviously impossible to change one curve into the other by any isometry of $\mathbb{R}^3$. We shall define another type of curvature for space curves, called the *torsion*, and we shall prove that the curvature and torsion of a curve together determine the curve up to a direct isometry of $\mathbb{R}^3$.

Let $\boldsymbol{\gamma}(s)$ be a unit-speed curve in $\mathbb{R}^3$, and let $\mathbf{t} = \dot{\boldsymbol{\gamma}}$ be its unit tangent vector. *If the curvature* $\kappa(s)$ *is non-zero*, we define the *principal normal* of $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(s)$ to be the vector

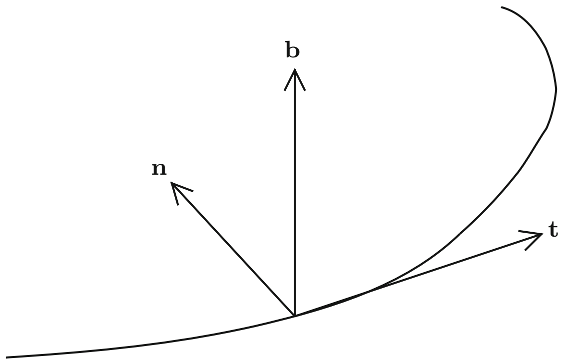
$$\mathbf{n}(s) = \frac{1}{\kappa(s)}\dot{\mathbf{t}}(s). \tag{2.10}$$

Since $\| \dot{\mathbf{t}} \| = \kappa$, $\mathbf{n}$ is a unit vector. Further, by Proposition 1.2.4, $\mathbf{t} \cdot \dot{\mathbf{t}} = 0$, so $\mathbf{t}$ and $\mathbf{n}$ are actually perpendicular unit vectors. It follows that

$$\mathbf{b} = \mathbf{t} \times \mathbf{n} \tag{2.11}$$

is a unit vector perpendicular to both $\mathbf{t}$ and $\mathbf{n}$. The vector $\mathbf{b}(s)$ is called the *binormal* vector of $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(s)$. Thus, $\{\mathbf{t}, \mathbf{n}, \mathbf{b}\}$ is an orthonormal basis of $\mathbb{R}^3$, and is *right-handed*, i.e.,

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}, \quad \mathbf{n} = \mathbf{b} \times \mathbf{t}, \quad \mathbf{t} = \mathbf{n} \times \mathbf{b}.$$




Since $\mathbf{b}(s)$ is a unit vector for all $s$, $\dot{\mathbf{b}}$ is perpendicular to $\mathbf{b}$. Now we use the 'product rule' for differentiating the vector product of vector-valued functions $\mathbf{u}$ and $\mathbf{v}$ of a parameter $s$:

$$\frac{d}{ds}(\mathbf{u} \times \mathbf{v}) = \frac{d\mathbf{u}}{ds} \times \mathbf{v} + \mathbf{u} \times \frac{d\mathbf{v}}{ds}.$$

(This is easily proved by writing out both sides in component form and using the usual product rule for differentiating scalar functions.) Applying this to $\mathbf{b} = \mathbf{t} \times \mathbf{n}$ gives

$$\dot{\mathbf{b}} = \dot{\mathbf{t}} \times \mathbf{n} + \mathbf{t} \times \dot{\mathbf{n}} = \mathbf{t} \times \dot{\mathbf{n}}, \tag{2.12}$$

since by the definition (2.10) of $\mathbf{n}$, $\dot{\mathbf{t}} \times \mathbf{n} = \kappa \mathbf{n} \times \mathbf{n} = \mathbf{0}$. Equation 2.12 shows that $\dot{\mathbf{b}}$ is perpendicular to $\mathbf{t}$. Being perpendicular to both $\mathbf{t}$ and $\mathbf{b}$, $\dot{\mathbf{b}}$ must be parallel to $\mathbf{n}$, so

$$\dot{\mathbf{b}} = -\tau \mathbf{n}, \tag{2.13}$$

for some scalar $\tau$, which is called the *torsion* of $\boldsymbol{\gamma}$ (inserting the minus sign here will reduce the total number of minus signs later). Note that the torsion is only defined if the curvature is non-zero.

Of course, we define the torsion of an arbitrary regular curve $\boldsymbol{\gamma}$ to be that of a unit-speed reparametrization of $\boldsymbol{\gamma}$. As in the case of the curvature, to see that this makes sense, we have to investigate how the torsion is affected by a change in the unit-speed parameter of $\boldsymbol{\gamma}$ of the form

$$u = \pm s + c,$$

where $c$ is a constant. But this change of parameter clearly has the following effect on the vectors introduced above:

$$\mathbf{t} \mapsto \pm \mathbf{t}, \ \dot{\mathbf{t}} \mapsto \dot{\mathbf{t}}, \ \mathbf{n} \mapsto \mathbf{n}, \ \mathbf{b} \mapsto \pm \mathbf{b}, \ \dot{\mathbf{b}} \mapsto \dot{\mathbf{b}};$$

it follows from Eq. 2.13 that $\tau \mapsto \tau$. Thus, *the curvature and torsion are well-defined for any regular curve.*

Just as we did for the curvature in Proposition 2.1.2, it is possible to give a formula for the torsion of a regular space curve $\boldsymbol{\gamma}$ in terms of $\boldsymbol{\gamma}$ itself, rather than in terms of a unit-speed reparametrization:

## Proposition 2.3.1

Let $\boldsymbol{\gamma}(t)$ be a regular curve in $\mathbb{R}^3$ with nowhere-vanishing curvature. Then, denoting $d/dt$ by a dot, its torsion is given by

$$\tau = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\| \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} \|^2}. \tag{2.14}$$

Note that this formula shows that $\tau(t)$ is defined at all points $\boldsymbol{\gamma}(t)$ of the curve at which its curvature $\kappa(t)$ is non-zero, since by Proposition 2.1.2 this is the condition for the denominator on the right-hand side to be non-zero.

## Proof

We could 'derive' Eq. 2.14 by imitating the proof of Proposition 2.1.2. But it is easier and clearer to proceed as follows, even though this method has the disadvantage that one must know the formula (2.14) for $\tau$ in advance.

We first treat the case in which $\boldsymbol{\gamma}$ is unit-speed. Using Eqs. 2.11 and 2.13,

$$\tau = -\mathbf{n} \cdot \dot{\mathbf{b}} = -\mathbf{n} \cdot (\mathbf{t} \times \mathbf{n})^{\cdot} = -\mathbf{n} \cdot (\dot{\mathbf{t}} \times \mathbf{n} + \mathbf{t} \times \dot{\mathbf{n}}) = -\mathbf{n} \cdot (\mathbf{t} \times \dot{\mathbf{n}}).$$

Now, $\mathbf{n} = \frac{1}{\kappa}\dot{\mathbf{t}} = \frac{1}{\kappa}\ddot{\boldsymbol{\gamma}}$, so

$$\tau = -\frac{1}{\kappa}\ddot{\boldsymbol{\gamma}} \cdot \left(\dot{\boldsymbol{\gamma}} \times \frac{d}{dt}\left(\frac{1}{\kappa}\ddot{\boldsymbol{\gamma}}\right)\right) = -\frac{1}{\kappa}\ddot{\boldsymbol{\gamma}} \cdot \left(\dot{\boldsymbol{\gamma}} \times \left(\frac{1}{\kappa}\dddot{\boldsymbol{\gamma}} - \frac{\dot{\kappa}}{\kappa^2}\ddot{\boldsymbol{\gamma}}\right)\right) = \frac{1}{\kappa^2}\dddot{\boldsymbol{\gamma}} \cdot (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}),$$

since $\ddot{\boldsymbol{\gamma}} \cdot (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) = 0$ and $\ddot{\boldsymbol{\gamma}} \cdot (\dot{\boldsymbol{\gamma}} \times \dddot{\boldsymbol{\gamma}}) = -\dddot{\boldsymbol{\gamma}} \cdot (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}})$. This agrees with Eq. 2.14, for, since $\boldsymbol{\gamma}$ is unit-speed, $\dot{\boldsymbol{\gamma}}$ and $\ddot{\boldsymbol{\gamma}}$ are perpendicular, so

$$\| \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} \| = \| \dot{\boldsymbol{\gamma}} \| \| \ddot{\boldsymbol{\gamma}} \| = \| \ddot{\boldsymbol{\gamma}} \| = \kappa.$$

In the general case, let $s$ be arc-length along $\boldsymbol{\gamma}$. Then,

$$\frac{d\boldsymbol{\gamma}}{dt} = \frac{ds}{dt}\frac{d\boldsymbol{\gamma}}{ds}, \quad \frac{d^2\boldsymbol{\gamma}}{dt^2} = \left(\frac{ds}{dt}\right)^2 \frac{d^2\boldsymbol{\gamma}}{ds^2} + \frac{d^2 s}{dt^2}\frac{d\boldsymbol{\gamma}}{ds},$$
$$\frac{d^3\boldsymbol{\gamma}}{dt^3} = \left(\frac{ds}{dt}\right)^3 \frac{d^3\boldsymbol{\gamma}}{ds^3} + 3\frac{ds}{dt}\frac{d^2 s}{dt^2}\frac{d^2\boldsymbol{\gamma}}{ds^2} + \frac{d^3 s}{dt^3}\frac{d\boldsymbol{\gamma}}{ds}.$$

Hence,

$$\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} = \left(\frac{ds}{dt}\right)^3 \left(\frac{d\boldsymbol{\gamma}}{ds} \times \frac{d^2\boldsymbol{\gamma}}{ds^2}\right),$$
$$\dddot{\boldsymbol{\gamma}} \cdot (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) = \left(\frac{ds}{dt}\right)^6 \left(\frac{d^3\boldsymbol{\gamma}}{ds^3} \cdot \left(\frac{d\boldsymbol{\gamma}}{ds} \times \frac{d^2\boldsymbol{\gamma}}{ds^2}\right)\right).$$

So the torsion of $\boldsymbol{\gamma}$ is

$$\tau = \frac{\left(\frac{d^3\boldsymbol{\gamma}}{ds^3} \cdot \left(\frac{d\boldsymbol{\gamma}}{ds} \times \frac{d^2\boldsymbol{\gamma}}{ds^2}\right)\right)}{\left\| \frac{d\boldsymbol{\gamma}}{ds} \times \frac{d^2\boldsymbol{\gamma}}{ds^2} \right\|^2} = \frac{\dddot{\boldsymbol{\gamma}} \cdot (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}})}{\| \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} \|^2}. \qquad \square$$

## Example 2.3.2

We compute the torsion of the circular helix $\boldsymbol{\gamma}(\theta) = (a\cos\theta, a\sin\theta, b\theta)$ studied in Example 2.1.3. We have

$$\begin{aligned}
\dot{\boldsymbol{\gamma}}(\theta) &= (-a\sin\theta, a\cos\theta, b), \quad \ddot{\boldsymbol{\gamma}}(\theta) = (-a\cos\theta, -a\sin\theta, 0), \\
\dddot{\boldsymbol{\gamma}}(\theta) &= (a\sin\theta, -a\cos\theta, 0), \quad \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} = (ab\sin\theta, -ab\cos\theta, a^2), \\
\| \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} \|^2 &= a^2(a^2+b^2), \quad (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} = a^2 b,
\end{aligned}$$

so the torsion

$$\tau = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\| \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} \|^2} = \frac{a^2 b}{a^2(a^2+b^2)} = \frac{b}{a^2+b^2}.$$

Note that the torsion of the circular helix in Example 2.3.2 becomes zero when $b = 0$, in which case the helix is just a circle in the $xy$-plane. This gives us a clue to the geometrical interpretation of torsion, contained in the next proposition.

## Proposition 2.3.3

Let $\boldsymbol{\gamma}$ be a regular curve in $\mathbb{R}^3$ with nowhere vanishing curvature (so that the torsion $\tau$ of $\boldsymbol{\gamma}$ is defined). Then, the image of $\boldsymbol{\gamma}$ is contained in a plane if and only if $\tau$ is zero at every point of the curve.

## Proof

We can assume that $\boldsymbol{\gamma}$ is unit-speed (for this can be achieved by reparametrizing $\boldsymbol{\gamma}$, and reparametrizing changes neither the torsion nor the fact that $\boldsymbol{\gamma}$ is, or is not, contained in a plane). We denote the parameter of $\boldsymbol{\gamma}$ by $s$ and $d/ds$ by a dot as usual.

Suppose first that the image of $\boldsymbol{\gamma}$ is contained in the plane $\mathbf{v} \cdot \mathbf{N} = d$, where $\mathbf{N}$ is a constant vector and $d$ is a constant scalar and $\mathbf{v} \in \mathbb{R}^3$. We can assume that $\mathbf{N}$ is a unit vector. Differentiating $\boldsymbol{\gamma} \cdot \mathbf{N} = d$ with respect to $s$, we get

$$\begin{aligned}
& \mathbf{t} \cdot \mathbf{N} = 0, && (2.15) \\
\therefore \quad & \dot{\mathbf{t}} \cdot \mathbf{N} = 0 \qquad (\text{since } \dot{\mathbf{N}} = \mathbf{0}), &&
\end{aligned}$$

$$\begin{aligned} \therefore \quad \kappa\mathbf{n}\cdot\mathbf{N} &= 0 \qquad (\text{since } \dot{\mathbf{t}} = \kappa\mathbf{n}), \\ \therefore \quad \mathbf{n}\cdot\mathbf{N} &= 0 \qquad (\text{since } \kappa \neq 0). \end{aligned} \tag{2.16}$$

Equations 2.15 and 2.16 show that $\mathbf{t}$ and $\mathbf{n}$ are perpendicular to $\mathbf{N}$. It follows that $\mathbf{b} = \mathbf{t}\times\mathbf{n}$ is parallel to $\mathbf{N}$. Since $\mathbf{N}$ and $\mathbf{b}$ are both unit vectors, and $\mathbf{b}(s)$ is a smooth (hence continuous) function of $s$, we must have $\mathbf{b}(s) = \mathbf{N}$ for all $s$ or $\mathbf{b}(s) = -\mathbf{N}$ for all $s$. In both cases, $\mathbf{b}$ is a constant vector. But then $\dot{\mathbf{b}} = \mathbf{0}$, so $\tau = 0$.

Conversely, suppose that $\tau = 0$ everywhere. By Eq. 2.13, $\dot{\mathbf{b}} = \mathbf{0}$, so $\mathbf{b}$ is a constant vector. The first part of the proof suggests that $\boldsymbol{\gamma}$ should be contained in a plane $\mathbf{v}\cdot\mathbf{b} = \text{constant}$. We therefore consider

$$\frac{d}{ds}(\boldsymbol{\gamma}\cdot\mathbf{b}) = \dot{\boldsymbol{\gamma}}\cdot\mathbf{b} = \mathbf{t}\cdot\mathbf{b} = 0,$$

so $\boldsymbol{\gamma}\cdot\mathbf{b}$ is a constant (scalar), say $d$. This means that $\boldsymbol{\gamma}$ is indeed contained in the plane $\mathbf{v}\cdot\mathbf{b} = d$. □

There is a gap in our calculations which we would like to fill. Namely, we know that, for a unit-speed curve, we have

$$\dot{\mathbf{t}} = \kappa\mathbf{n} \quad \text{and} \quad \dot{\mathbf{b}} = -\tau\mathbf{n}$$

(these were our definitions of $\mathbf{n}$ and $\tau$, respectively), but we have not computed $\dot{\mathbf{n}}$. This is not difficult. Since $\{\mathbf{t},\mathbf{n},\mathbf{b}\}$ is a right-handed orthonormal basis of $\mathbb{R}^3$,

$$\mathbf{t}\times\mathbf{n} = \mathbf{b}, \quad \mathbf{n}\times\mathbf{b} = \mathbf{t}, \quad \mathbf{b}\times\mathbf{t} = \mathbf{n}.$$

Hence,

$$\dot{\mathbf{n}} = \dot{\mathbf{b}}\times\mathbf{t} + \mathbf{b}\times\dot{\mathbf{t}} = -\tau\mathbf{n}\times\mathbf{t} + \kappa\mathbf{b}\times\mathbf{n} = -\kappa\mathbf{t} + \tau\mathbf{b}.$$

Putting all these together, we get the following theorem.

## Theorem 2.3.4

Let $\boldsymbol{\gamma}$ be a unit-speed curve in $\mathbb{R}^3$ with nowhere vanishing curvature. Then,

$$\begin{aligned} \dot{\mathbf{t}} &= \kappa\mathbf{n} \\ \dot{\mathbf{n}} &= -\kappa\mathbf{t} + \tau\mathbf{b} \\ \dot{\mathbf{b}} &= -\tau\mathbf{n}. \end{aligned} \tag{2.17}$$

Equations 2.17 are called the *Frenet–Serret equations*. Notice that the matrix

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

which expresses $\dot{\mathbf{t}}$, $\dot{\mathbf{n}}$ and $\dot{\mathbf{b}}$ in terms of $\mathbf{t}$, $\mathbf{n}$ and $\mathbf{b}$ is *skew-symmetric*, i.e., it is equal to the negative of its transpose. This helps when trying to remember the equations. (The 'reason' for this skew-symmetry can be seen in Exercise 2.3.6.)

Here is a simple application of Frenet–Serret:

## Proposition 2.3.5

Let $\boldsymbol{\gamma}$ be a unit-speed curve in $\mathbb{R}^3$ with constant curvature and zero torsion. Then, $\boldsymbol{\gamma}$ is a parametrization of (part of) a circle.

## Proof

This result is actually an immediate consequence of Example 2.2.7 and Proposition 2.3.3, but the following proof is instructive and gives more information, namely the centre and radius of the circle and the plane in which it lies.

By the proof of Proposition 2.3.3, the binormal $\mathbf{b}$ is a constant vector and $\boldsymbol{\gamma}$ is contained in a plane $\Pi$, say, perpendicular to $\mathbf{b}$. Now

$$\frac{d}{ds}\left(\boldsymbol{\gamma}+\frac{1}{\kappa}\mathbf{n}\right)=\mathbf{t}+\frac{1}{\kappa}\dot{\mathbf{n}}=\mathbf{0},$$

using the fact that the curvature $\kappa$ is constant and the Frenet–Serret equation

$$\dot{\mathbf{n}}=-\kappa\mathbf{t}+\tau\mathbf{b}=-\kappa\mathbf{t}\quad(\text{since } \tau=0)$$

(the reason for considering $\boldsymbol{\gamma}+\frac{1}{\kappa}\mathbf{n}$ can be found in Exercise 2.2.6). Hence, $\boldsymbol{\gamma}+\frac{1}{\kappa}\mathbf{n}$ is a constant vector, say $\mathbf{a}$, and we have

$$\parallel \boldsymbol{\gamma}-\mathbf{a} \parallel = \parallel -\frac{1}{\kappa}\mathbf{n} \parallel = \frac{1}{\kappa}.$$

This shows that $\boldsymbol{\gamma}$ lies on the sphere $\mathcal{S}$, say, with centre $\mathbf{a}$ and radius $1/\kappa$. The intersection of $\Pi$ and $\mathcal{S}$ is a circle, say $\mathcal{C}$, and we have shown that $\boldsymbol{\gamma}$ is a parametrization of part of $\mathcal{C}$. If $r$ is the radius of $\mathcal{C}$, we have $\kappa=1/r$ so $r=1/\kappa$ is also the radius of $\mathcal{S}$. It follows that $\mathcal{C}$ is a *great circle* on $\mathcal{S}$, i.e., that $\Pi$ passes through the centre $\mathbf{a}$ of $\mathcal{S}$. Thus, $\mathbf{a}$ is the centre of $\mathcal{C}$ and the equation of $\Pi$ is $\mathbf{v}\cdot\mathbf{b}=\mathbf{a}\cdot\mathbf{b}$. □

We conclude this chapter with the analogue of Theorem 2.2.6 for space curves.

## Theorem 2.3.6

Let $\boldsymbol{\gamma}(s)$ and $\tilde{\boldsymbol{\gamma}}(s)$ be two unit-speed curves in $\mathbb{R}^3$ with the same curvature $\kappa(s) > 0$ and the same torsion $\tau(s)$ for all $s$. Then, there is a direct isometry $M$ of $\mathbb{R}^3$ such that

$$\tilde{\boldsymbol{\gamma}}(s) = M(\boldsymbol{\gamma}(s)) \quad \text{for all } s.$$

Further, if $k$ and $t$ are smooth functions with $k > 0$ everywhere, there is a unit-speed curve in $\mathbb{R}^3$ whose curvature is $k$ and whose torsion is $t$.

## Proof

Let $\mathbf{t}, \mathbf{n}$ and $\mathbf{b}$ be the tangent vector, principal normal and binormal of $\boldsymbol{\gamma}$, and let $\tilde{\mathbf{t}}, \tilde{\mathbf{n}}$ and $\tilde{\mathbf{b}}$ be those of $\tilde{\boldsymbol{\gamma}}$. Let $s_0$ be a fixed value of the parameter $s$, let $\theta$ be the angle between $\mathbf{t}(s_0)$ and $\tilde{\mathbf{t}}(s_0)$ and let $\rho$ be the rotation through an angle $\theta$ around the axis passing through the origin and perpendicular to both of these vectors. Then, $\rho(\mathbf{t}(s_0)) = \tilde{\mathbf{t}}(s_0)$; let $\hat{\mathbf{n}} = \rho(\mathbf{n}(s_0))$, $\hat{\mathbf{b}} = \rho(\mathbf{b}(s_0))$. If $\varphi$ is the angle between $\hat{\mathbf{n}}$ and $\tilde{\mathbf{n}}(s_0)$, let $\rho'$ be the rotation through an angle $\varphi$ around the axis passing through the origin parallel to $\tilde{\mathbf{t}}(s_0)$. Then, $\rho'$ fixes $\tilde{\mathbf{t}}(s_0)$ and takes $\hat{\mathbf{n}}$ to $\tilde{\mathbf{n}}(s_0)$. Since $\{\mathbf{t}(s_0), \mathbf{n}(s_0), \mathbf{b}(s_0)\}$ and $\{\tilde{\mathbf{t}}(s_0), \tilde{\mathbf{n}}(s_0), \tilde{\mathbf{b}}(s_0)\}$ are both right-handed orthonormal bases of $\mathbb{R}^3$, $\rho' \circ \rho$ takes the vectors $\mathbf{t}(s_0), \mathbf{n}(s_0), \mathbf{b}(s_0)$ to the vectors $\tilde{\mathbf{t}}(s_0), \tilde{\mathbf{n}}(s_0), \tilde{\mathbf{b}}(s_0)$, respectively. Now let $M$ be the direct isometry $M = T_{\tilde{\boldsymbol{\gamma}}(s_0) - \boldsymbol{\gamma}(s_0)} \circ \rho' \circ \rho$. By Exercise 2.3.5, the curve $\boldsymbol{\Gamma} = M(\boldsymbol{\gamma})$ is unit-speed, and if $\mathbf{T}$, $\mathbf{N}$ and $\mathbf{B}$ denote its unit tangent vector, principal normal and binormal, we have

$$\boldsymbol{\Gamma}(s_0) = \tilde{\boldsymbol{\gamma}}(s_0), \ \mathbf{T}(s_0) = \tilde{\mathbf{t}}(s_0), \ \mathbf{N}(s_0) = \tilde{\mathbf{n}}(s_0), \ \mathbf{B}(s_0) = \tilde{\mathbf{b}}(s_0). \tag{2.18}$$

The trick now is to consider the expression

$$A(s) = \tilde{\mathbf{t}} \cdot \mathbf{T} + \tilde{\mathbf{n}} \cdot \mathbf{N} + \tilde{\mathbf{b}} \cdot \mathbf{B}.$$

In view of Eq. 2.18, we have $A(s_0) = 3$. On the other hand, since $\tilde{\mathbf{t}}$ and $\mathbf{T}$ are unit vectors, $\tilde{\mathbf{t}} \cdot \mathbf{T} \leq 1$, with equality holding if and only if $\tilde{\mathbf{t}} = \mathbf{T}$; and similarly for $\tilde{\mathbf{n}} \cdot \mathbf{N}$ and $\tilde{\mathbf{b}} \cdot \mathbf{B}$. It follows that $A(s) \leq 3$, with equality holding if and only if $\tilde{\mathbf{t}} = \mathbf{T}, \tilde{\mathbf{n}} = \mathbf{N}$ and $\tilde{\mathbf{b}} = \mathbf{B}$. Thus, if we can prove that $A$ is constant, it will follow in particular that $\tilde{\mathbf{t}} = \mathbf{T}$, i.e., that $\dot{\tilde{\boldsymbol{\gamma}}} = \dot{\boldsymbol{\Gamma}}$, and hence that $\tilde{\boldsymbol{\gamma}}(s) - \boldsymbol{\Gamma}(s)$ is a constant. But by Eq. 2.18 again, this constant vector must be zero, so $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\Gamma}$.

For the first part of the theorem, we are therefore reduced to proving that $A$ is constant. But, using the Frenet–Serret equations,

$$\begin{aligned}
\dot{A} &= \dot{\tilde{\mathbf{t}}} \cdot \mathbf{T} + \dot{\tilde{\mathbf{n}}} \cdot \mathbf{N} + \dot{\tilde{\mathbf{b}}} \cdot \mathbf{B} + \tilde{\mathbf{t}} \cdot \dot{\mathbf{T}} + \tilde{\mathbf{n}} \cdot \dot{\mathbf{N}} + \tilde{\mathbf{b}} \cdot \dot{\mathbf{B}} \\
&= \kappa \tilde{\mathbf{n}} \cdot \mathbf{T} + (-\kappa \tilde{\mathbf{t}} + \tau \tilde{\mathbf{b}}) \cdot \mathbf{N} + (-\tau \tilde{\mathbf{n}}) \cdot \mathbf{B} + \tilde{\mathbf{t}} \cdot \kappa \mathbf{N} \\
&\quad + \tilde{\mathbf{n}} \cdot (-\kappa \mathbf{T} + \tau \mathbf{B}) + \tilde{\mathbf{b}} \cdot (-\tau \mathbf{N}),
\end{aligned}$$

and this vanishes since the terms cancel in pairs.

For the second part of the theorem, we observe first that it follows from the theory of ordinary differential equations that the equations

$$\dot{\mathbf{T}} = k\mathbf{N}, \tag{2.19}$$

$$\dot{\mathbf{N}} = -k\mathbf{T} + t\mathbf{B}, \tag{2.20}$$

$$\dot{\mathbf{B}} = -t\mathbf{N} \tag{2.21}$$

have a unique solution $\mathbf{T}(s), \mathbf{N}(s), \mathbf{B}(s)$ such that $\mathbf{T}(s_0), \mathbf{N}(s_0), \mathbf{B}(s_0)$ are the standard orthonormal vectors $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$, $\mathbf{k} = (0, 0, 1)$, respectively. Since the matrix

$$\begin{pmatrix} 0 & k & 0 \\ -k & 0 & t \\ 0 & -t & 0 \end{pmatrix}$$

expressing $\dot{\mathbf{T}}, \dot{\mathbf{N}}$ and $\dot{\mathbf{B}}$ in terms of $\mathbf{T}, \mathbf{N}$ and $\mathbf{B}$ is skew-symmetric, it follows that the vectors $\mathbf{T}$, $\mathbf{N}$ and $\mathbf{B}$ are orthonormal for all values of $s$ (see Exercise 2.3.6).

Now define

$$\boldsymbol{\gamma}(s) = \int_{s_0}^{s} \mathbf{T}(u)du.$$

Then, $\dot{\boldsymbol{\gamma}} = \mathbf{T}$, so since $\mathbf{T}$ is a unit vector, $\boldsymbol{\gamma}$ is unit-speed. Next, $\dot{\mathbf{T}} = k\mathbf{N}$ by Eq. 2.19, so since $\mathbf{N}$ is a unit vector, $k$ is the curvature of $\boldsymbol{\gamma}$ and $\mathbf{N}$ is its principal normal. Next, since $\mathbf{B}$ is a unit vector perpendicular to $\mathbf{T}$ and $\mathbf{N}$, $\mathbf{B} = \lambda\mathbf{T} \times \mathbf{N}$ where $\lambda$ is a smooth function of $s$ that is equal to $\pm 1$ for all $s$. Since $\mathbf{k} = \mathbf{i} \times \mathbf{j}$, we have $\lambda(s_0) = 1$, so it follows that $\lambda(s) = 1$ for all $s$. Hence, $\mathbf{B}$ is the binormal of $\boldsymbol{\gamma}$ and by Eq. 2.21, $t$ is its torsion. □

## *EXERCISES*

2.3.1 Compute $\kappa, \tau, \mathbf{t}, \mathbf{n}$ and $\mathbf{b}$ for each of the following curves, and verify that the Frenet–Serret equations are satisfied:

(i) $\boldsymbol{\gamma}(t) = \left(\frac{1}{3}(1+t)^{3/2} \frac{1}{3}(1-t)^{3/2}, \frac{t}{\sqrt{2}}\right)$.

(ii) $\boldsymbol{\gamma}(t) = \left(\frac{4}{5}\cos t, 1 - \sin t, -\frac{3}{5}\cos t\right)$.

Show that the curve in (ii) is a circle, and find its centre, radius and the plane in which it lies.

2.3.2 Describe all curves in $\mathbb{R}^3$ which have *constant* curvature $\kappa > 0$ and *constant* torsion $\tau$.

2.3.3 A regular curve $\boldsymbol{\gamma}$ in $\mathbb{R}^3$ with curvature $> 0$ is called a *generalized helix* if its tangent vector makes a fixed angle $\theta$ with a fixed unit vector $\mathbf{a}$. Show that the torsion $\tau$ and curvature $\kappa$ of $\boldsymbol{\gamma}$ are related by $\tau = \pm\kappa \cot\theta$. Show conversely that, if the torsion and curvature of a regular curve are related by $\tau = \lambda\kappa$ where $\lambda$ is a constant, then the curve is a generalized helix.
In view of this result, Examples 2.1.3 and 2.3.2 show that a circular helix is a generalized helix. Verify this directly.

2.3.4 Let $\boldsymbol{\gamma}(t)$ be a unit-speed curve with $\kappa(t) > 0$ and $\tau(t) \neq 0$ for all $t$. Show that, if $\boldsymbol{\gamma}$ is *spherical*, i.e., if it lies on the surface of a sphere, then

$$\frac{\tau}{\kappa} = \frac{d}{ds}\left(\frac{\dot{\kappa}}{\tau\kappa^2}\right). \tag{2.22}$$

Conversely, show that if Eq. 2.22 holds, then

$$\rho^2 + (\dot{\rho}\sigma)^2 = r^2$$

for some (positive) constant $r$, where $\rho = 1/\kappa$ and $\sigma = 1/\tau$, and deduce that $\boldsymbol{\gamma}$ lies on a sphere of radius $r$. Verify that Eq. 2.22 holds for Viviani's curve (Exercise 1.1.8).

2.3.5 Let $P$ be an $n \times n$ orthogonal matrix and let $\mathbf{a} \in \mathbb{R}^n$, so that $M(\mathbf{v}) = P\mathbf{v} + \mathbf{a}$ is an isometry of $\mathbb{R}^3$ (see Appendix 1). Show that, if $\boldsymbol{\gamma}$ is a unit-speed curve in $\mathbb{R}^n$, the curve $\boldsymbol{\Gamma} = M(\boldsymbol{\gamma})$ is also unit-speed. Show also that, if $\mathbf{t}, \mathbf{n}, \mathbf{b}$ and $\mathbf{T}, \mathbf{N}, \mathbf{B}$ are the tangent vector, principal normal and binormal of $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$, respectively, then $\mathbf{T} = P\mathbf{t}$, $\mathbf{N} = P\mathbf{n}$ and $\mathbf{B} = P\mathbf{b}$.

2.3.6 Let $(a_{ij})$ be a skew-symmetric $3 \times 3$ matrix (i.e., $a_{ij} = -a_{ji}$ for all $i, j$). Let $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ be smooth functions of a parameter $s$ satisfying the differential equations

$$\dot{\mathbf{v}}_i = \sum_{j=1}^{3} a_{ij}\mathbf{v}_j,$$

for $i = 1, 2$ and 3, and suppose that for some parameter value $s_0$ the vectors $\mathbf{v}_1(s_0), \mathbf{v}_2(s_0)$ and $\mathbf{v}_3(s_0)$ are orthonormal. Show that the vectors $\mathbf{v}_1(s), \mathbf{v}_2(s)$ and $\mathbf{v}_3(s)$ are orthonormal for all values of $s$.

*For the remainder of this book,*
*all parametrized curves will be assumed to be regular.*

# 3
# Global properties of curves

All the properties of curves that we have discussed so far are 'local': they depend only on the behaviour of a curve near a given point and not on the 'global' shape of the curve. Proving global results about curves often requires concepts from *topology*, in addition to the calculus techniques we have used in the first two chapters of this book. Since we are not assuming that readers of this book have extensive familiarity with topological ideas, we will not be able to give complete proofs of some of the global results about curves that we discuss in this chapter.

## 3.1 Simple closed curves

In this chapter, we shall consider plane curves of the following type.

### Definition 3.1.1

A *simple closed curve* in $\mathbb{R}^2$ is a closed curve in $\mathbb{R}^2$ that has no self-intersections.

It is a standard, but highly non-trivial, result of the topology of $\mathbb{R}^2$, called the *Jordan Curve Theorem*, that any simple closed curve in the plane has an 'interior' and an 'exterior': more precisely, the *complement* of the image of $\boldsymbol{\gamma}$

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_3,

(i.e., the set of points of $\mathbb{R}^2$ that are *not* in the image of $\boldsymbol{\gamma}$) is the disjoint union of two subsets of $\mathbb{R}^2$, denoted by $\mathrm{int}(\boldsymbol{\gamma})$ and $\mathrm{ext}(\boldsymbol{\gamma})$, with the following properties:

(i) $\mathrm{int}(\boldsymbol{\gamma})$ is *bounded*, i.e., it is contained inside a circle of sufficiently large radius.

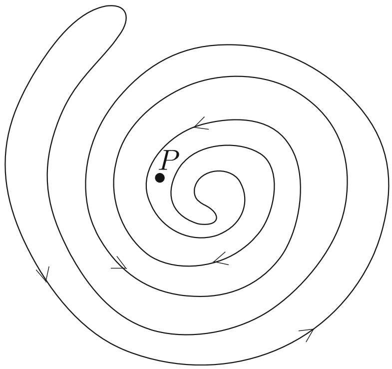
(ii) $\mathrm{ext}(\boldsymbol{\gamma})$ is unbounded.

(iii) Both of the regions $\mathrm{int}(\boldsymbol{\gamma})$ and $\mathrm{ext}(\boldsymbol{\gamma})$ are *connected*, i.e., they have the property that any two points in the same region can be joined by a curve contained entirely in the region (but any curve joining a point of $\mathrm{int}(\boldsymbol{\gamma})$ to a point of $\mathrm{ext}(\boldsymbol{\gamma})$ must cross the curve $\boldsymbol{\gamma}$).

## Example 3.1.2

The ellipse $\boldsymbol{\gamma}(t) = (p\cos t, q\sin t)$, where $p$ and $q$ are non-zero constants, is a simple closed curve with period $2\pi$. The interior and exterior of $\boldsymbol{\gamma}$ are, of course, given by $\left\{(x,y) \in \mathbb{R}^2 \mid \frac{x^2}{p^2} + \frac{y^2}{q^2} < 1\right\}$ and $\left\{(x,y) \in \mathbb{R}^2 \mid \frac{x^2}{p^2} + \frac{y^2}{q^2} > 1\right\}$, respectively.

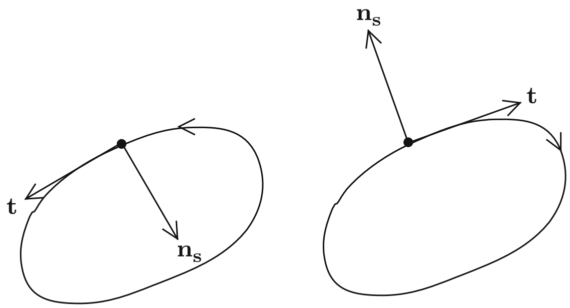
Not all examples of simple closed curves have such an obvious interior and exterior, however. Is the point $\mathbf{p}$ in the interior or the exterior of the simple closed curve shown below?




## Example 3.1.3

The limaçon in Example 1.1.7 is closed but is not a simple closed curve as it has a self-intersection – see Exercise 3.1.1.

The fact that a simple closed curve has an interior and an exterior enables us to distinguish between the two possible orientations of $\boldsymbol{\gamma}$. We shall say that $\boldsymbol{\gamma}$ is *positively-oriented* if the signed unit normal $\mathbf{n}_s$ of $\boldsymbol{\gamma}$ (see Section 2.2) points into $\text{int}(\boldsymbol{\gamma})$ at every point of $\boldsymbol{\gamma}$. This can always be achieved by replacing the parameter $t$ of $\boldsymbol{\gamma}$ by $-t$, if necessary. In the diagrams below, the arrow indicates the direction of increasing parameter. Is the simple closed curve shown above positively-oriented?




Positively-oriented Not positively-oriented

We conclude this section by stating the following important result.

### Theorem 3.1.4 (Hopf's Umlaufsatz)

The total signed curvature of a simple closed curve in $\mathbb{R}^2$ is $\pm 2\pi$.

The proof of Theorem 3.1.4 would take us a little further into the realm of topology than is appropriate for this book. A heuristic proof (of a slightly more general result) is given in Section 13.1.

Note that Corollary 2.2.5 shows that the total signed curvature of any closed curve in $\mathbb{R}^2$ is an integer multiple of $2\pi$. The point of Hopf's theorem is that if the curve is *simple* closed, this integer must be $\pm 1$. The German word 'Umlaufsatz' means 'rotation theorem': from the proof of Corollary 2.2.5 we see that Hopf's theorem says that any turning angle $\varphi$ of a simple closed curve changes by $\pm 2\pi$ on going once round the curve, which means that the tangent vector rotates by $\pm 2\pi$. The reader might like to check that this property holds for the maze-like simple closed curve preceding Example 3.1.3.

## *EXERCISES*

3.1.1 Show that
$$\boldsymbol{\gamma}(t) = ((1 + a\cos t)\cos t, (1 + a\cos t)\sin t),$$
where $a$ is a constant, is a simple closed curve if $|a| < 1$, but that if $|a| > 1$ its complement is the disjoint union of three connected subsets of $\mathbb{R}^2$, two of which are bounded and one is unbounded. What happens if $a = \pm 1$?

## 3.2 The isoperimetric inequality

The *area* contained by a simple closed curve $\boldsymbol{\gamma}$ is

$$\mathcal{A}(\boldsymbol{\gamma}) = \int_{\mathrm{int}(\boldsymbol{\gamma})} dxdy. \tag{3.1}$$

This can be computed by using the following theorem.

**Green's Theorem** *Let $f(x,y)$ and $g(x,y)$ be smooth functions (i.e., functions with continuous partial derivatives of all orders), and let $\boldsymbol{\gamma}$ be a positively-oriented simple closed curve. Then,*

$$\int_{\mathrm{int}(\boldsymbol{\gamma})} \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dxdy = \int_{\boldsymbol{\gamma}} f(x,y)dx + g(x,y)dy.$$

A proof can be found in standard books on multivariable calculus.

### Proposition 3.2.1

If $\boldsymbol{\gamma}(t) = (x(t), y(t))$ is a positively-oriented simple closed curve in $\mathbb{R}^2$ with period $T$, then

$$\mathcal{A}(\boldsymbol{\gamma}) = \frac{1}{2}\int_0^T (x\dot{y} - y\dot{x})dt. \tag{3.2}$$

### Proof

Taking $f = -\frac{1}{2}y$, $g = \frac{1}{2}x$ in Green's theorem, we get

$$\mathcal{A}(\boldsymbol{\gamma}) = \frac{1}{2}\int_{\boldsymbol{\gamma}} xdy - ydx,$$

which gives Eq. 3.2 immediately. □

Note that, although the formula in Eq. 3.2 involves the parameter $t$ of $\boldsymbol{\gamma}$, it is clear from the Definition 3.1.1 that $\mathcal{A}(\boldsymbol{\gamma})$ is unchanged if $\boldsymbol{\gamma}$ is reparametrized.

One of the most famous global results about plane curves is the following theorem.

### Theorem 3.2.2 (Isoperimetric Inequality)

Let $\boldsymbol{\gamma}$ be a simple closed curve, let $\ell(\boldsymbol{\gamma})$ be its length and let $\mathcal{A}(\boldsymbol{\gamma})$ be the area contained by it. Then,

$$\mathcal{A}(\boldsymbol{\gamma}) \leq \frac{1}{4\pi}\ell(\boldsymbol{\gamma})^2,$$

and equality holds if and only if $\boldsymbol{\gamma}$ is a circle.

Of course, it is obvious that equality holds when $\boldsymbol{\gamma}$ is a circle, since in that case $\ell(\boldsymbol{\gamma}) = 2\pi R$ and $\mathcal{A}(\boldsymbol{\gamma}) = \pi R^2$, where $R$ is the radius of the circle.

To prove this theorem, we need the following result from analysis:

### Proposition 3.2.3 (Wirtinger's Inequality)

Let $F : [0, \pi] \to \mathbb{R}$ be a smooth function such that $F(0) = F(\pi) = 0$. Then,

$$\int_0^\pi \left(\frac{dF}{dt}\right)^2 dt \geq \int_0^\pi F(t)^2 \, dt,$$

and equality holds if and only if $F(t) = D \sin t$ for all $t \in [0, \pi]$, where $D$ is a constant.

Assuming this result for the moment, we show how to deduce the isoperimetric inequality from it.

### Proof

We start by making some assumptions about $\boldsymbol{\gamma}$ that will simplify the proof. First, we can, if we wish, assume that $\boldsymbol{\gamma}$ is parametrized by arc-length $s$. However, because of the $\pi$ that appears in Theorem 3.2.2, it turns out to be more convenient to assume that the period of $\boldsymbol{\gamma}$ is $\pi$. If we change the parameter of $\boldsymbol{\gamma}$ from $s$ to

$$t = \frac{\pi s}{\ell(\boldsymbol{\gamma})}, \tag{3.3}$$

the resulting curve is still simple closed, and has period $\pi$ because when $s$ increases by $\ell(\boldsymbol{\gamma})$, $t$ increases by $\pi$. We shall therefore assume that $\boldsymbol{\gamma}$ is parametrized using the parameter $t$ in Eq. 3.3 from now on.

For the second simplification, we note that both $\ell(\boldsymbol{\gamma})$ and $\mathcal{A}(\boldsymbol{\gamma})$ are unchanged if $\boldsymbol{\gamma}$ is subjected to a translation $\boldsymbol{\gamma}(t) \mapsto \boldsymbol{\gamma}(t) + \mathbf{b}$, where $\mathbf{b}$ is any constant vector (see Exercise 3.2.1). Taking $\mathbf{b} = -\boldsymbol{\gamma}(0)$, we might as well assume that $\boldsymbol{\gamma}(0) = \mathbf{0}$ to begin with, i.e., we assume that $\boldsymbol{\gamma}$ begins and ends at the origin.

To prove Theorem 3.2.2, we shall calculate $\ell(\boldsymbol{\gamma})$ and $\mathcal{A}(\boldsymbol{\gamma})$ by using polar coordinates

$$x = r\cos\theta, \quad y = r\sin\theta.$$

Using the chain rule, it is easy to show that

$$\dot{x}^2 + \dot{y}^2 = \dot{r}^2 + r^2\dot{\theta}^2, \qquad x\dot{y} - y\dot{x} = r^2\dot{\theta},$$

with $d/dt$ denoted by a dot. Then, using Eq. 3.3,

$$\dot{r}^2 + r^2\dot{\theta}^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = \left(\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2\right)\left(\frac{ds}{dt}\right)^2 = \frac{\ell(\boldsymbol{\gamma})^2}{\pi^2}, \quad (3.4)$$

since $(dx/ds)^2 + (dy/ds)^2 = 1$. Further, by Eq. 3.2, we have

$$\mathcal{A}(\boldsymbol{\gamma}) = \frac{1}{2}\int_0^\pi (x\dot{y} - y\dot{x})\,dt = \frac{1}{2}\int_0^\pi r^2\dot{\theta}\,dt. \quad (3.5)$$

To prove Theorem 3.2.2, we have to show that

$$\frac{\ell(\boldsymbol{\gamma})^2}{4\pi} - \mathcal{A}(\boldsymbol{\gamma}) \geq 0,$$

with equality holding if and only if $\boldsymbol{\gamma}$ is a circle. By Eq. 3.4,

$$\int_0^\pi (\dot{r}^2 + r^2\dot{\theta}^2)\,dt = \frac{\ell(\boldsymbol{\gamma})^2}{\pi}.$$

Hence, using Eq. 3.5,

$$\frac{\ell(\boldsymbol{\gamma})^2}{4\pi} - \mathcal{A}(\boldsymbol{\gamma}) = \frac{1}{4}\int_0^\pi (\dot{r}^2 + r^2\dot{\theta}^2)dt - \frac{1}{2}\int_0^\pi r^2\dot{\theta}\,dt = \frac{1}{4}\mathcal{I},$$

where

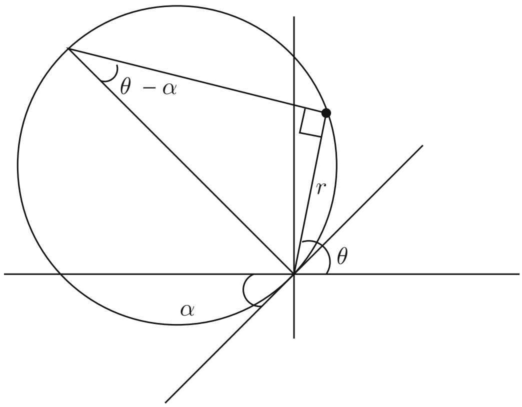
$$\mathcal{I} = \int_0^\pi (\dot{r}^2 + r^2\dot{\theta}^2 - 2r^2\dot{\theta})\,dt. \quad (3.6)$$

Thus, to prove Theorem 3.2.2, we have to show that $\mathcal{I} \geq 0$, and that $\mathcal{I} = 0$ if and only if $\boldsymbol{\gamma}$ is a circle.

By simple algebra,

$$\mathcal{I} = \int_0^\pi r^2(\dot{\theta} - 1)^2\,dt + \int_0^\pi (\dot{r}^2 - r^2)\,dt. \quad (3.7)$$

The first integral on the right-hand side of Eq. 3.7 is obviously $\geq 0$, and the second integral is $\geq 0$ by Wirtinger's inequality (we are taking $F = r$: note that $r(0) = r(\pi) = 0$ since $\boldsymbol{\gamma}(0) = \boldsymbol{\gamma}(\pi) = \mathbf{0}$). Hence, $\mathcal{I} \geq 0$. Further, since both integrals on the right-hand side of Eq. 3.7 are $\geq 0$, their sum $\mathcal{I}$ is zero if and only if both of these integrals are zero. But the first integral is zero only if $\dot{\theta} = 1$ for all $t$, and the second is zero only if $r = D\sin t$ for some constant $D$ (by Wirtinger again). So $\theta = t + \alpha$, where $\alpha$ is a constant, and hence $r = D\sin(\theta - \alpha)$. It is easy to see that this is the polar equation of a circle of diameter $D$, thus completing the proof of Theorem 3.2.2 (see the diagram below). $\square$




We now prove Wirtinger's inequality.

Let $G(t) = F(t)/\sin t$. Then, denoting $d/dt$ by a dot as usual,

$$\begin{aligned}\int_0^\pi \dot{F}^2\,dt &= \int_0^\pi (\dot{G}\sin t + G\cos t)^2\,dt\\ &= \int_0^\pi \dot{G}^2\sin^2 t\,dt + 2\int_0^\pi G\dot{G}\sin t\cos t\,dt + \int_0^\pi G^2\cos^2 t\,dt.\end{aligned}$$

Integrating by parts[1]:

$$\begin{aligned}2\int_0^\pi G\dot{G}\sin t\cos t\,dt &= G^2\sin t\cos t\big|_0^\pi - \int_0^\pi G^2(\cos^2 t - \sin^2 t)\,dt\\ &= \int_0^\pi G^2(\sin^2 t - \cos^2 t)\,dt.\end{aligned}$$

Hence,

$$\begin{aligned}\int_0^\pi \dot{F}^2 dt &= \int_0^\pi \dot{G}^2\sin^2 t\,dt + \int_0^\pi G^2(\sin^2 t - \cos^2 t)\,dt + \int_0^\pi G^2\cos^2 t\,dt\\ &= \int_0^\pi (G^2 + \dot{G}^2)\sin^2 t\,dt = \int_0^\pi F^2 dt + \int_0^\pi \dot{G}^2\sin^2 t\,dt,\end{aligned}$$

and so

$$\int_0^\pi \dot{F}^2\,dt - \int_0^\pi F^2\,dt = \int_0^\pi \dot{G}^2\sin^2 t\,dt.$$

The integral on the right-hand side is obviously $\geq 0$, and it is zero if and only if $\dot{G} = 0$ for all $t$, i.e., if and only if $G(t)$ is equal to a constant, say $D$, for all $t$, which means that $F(t) = D\sin t$. □

---

[1] In performing the integration by parts, we assume that $G$ is continuously differentiable (for we assume that the function $G(t)^2\sin t\cos t$ is equal to the integral of its derivative). Unfortunately, $G(t)$ is not even defined when $t = 0$ or $\pi$, as the ratio $F(t)/\sin t$ is $0/0$ there. So we must show that $G$ can be defined at these points so as to become continuously differentiable everywhere. This can be done by using l'Hospital's rule.

### *EXERCISES*

3.2.1 Show that the length $\ell(\boldsymbol{\gamma})$ and the area $\mathcal{A}(\boldsymbol{\gamma})$ are unchanged by applying an isometry to $\boldsymbol{\gamma}$.

3.2.2 By applying the isoperimetric inequality to the ellipse

$$\frac{x^2}{p^2} + \frac{y^2}{q^2} = 1$$

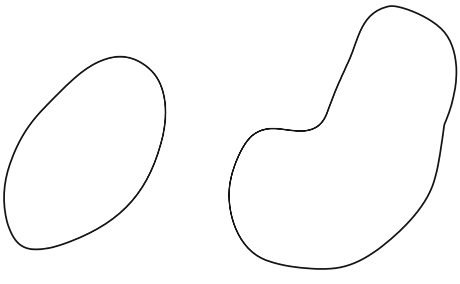
(where $p$ and $q$ are positive constants), prove that

$$\int_0^{2\pi} \sqrt{p^2 \sin^2 t + q^2 \cos^2 t}\, dt \geq 2\pi\sqrt{pq},$$

with equality holding if and only if $p = q$.

## 3.3 The four vertex theorem

We conclude this chapter with a famous result about convex curves in the plane. A simple closed curve $\boldsymbol{\gamma}$ is called *convex* if its interior $\operatorname{int}(\boldsymbol{\gamma})$ is convex, in the sense that the straight line segment joining any two points of $\operatorname{int}(\boldsymbol{\gamma})$ is contained entirely in $\operatorname{int}(\boldsymbol{\gamma})$.




### Definition 3.3.1

A *vertex* of a curve $\boldsymbol{\gamma}(t)$ in $\mathbb{R}^2$ is a point where its signed curvature $\kappa_s$ has a stationary point, i.e., where $d\kappa_s/dt = 0$.

It is easy to see that this definition is independent of the parametrization of $\boldsymbol{\gamma}$.

## Example 3.3.2

The signed curvature of the ellipse $\boldsymbol{\gamma}(t) = (p\cos t, q\sin t)$, where $p$ and $q$ are positive constants, is easily found to be

$$\kappa_s(t) = \frac{pq}{(p^2\sin^2 t + q^2\cos^2 t)^{3/2}}.$$

Then,

$$\frac{d\kappa_s}{dt} = \frac{3pq(q^2 - p^2)\sin t\cos t}{(p^2\sin^2 t + q^2\cos^2 t)^{5/2}}$$

vanishes at exactly four points of the ellipse, namely the points with $t = 0, \pi/2, \pi$ and $3\pi/2$, which are the ends of the two axes of the ellipse.

The following theorem says that this is the smallest number of vertices a convex simple closed curve can have.

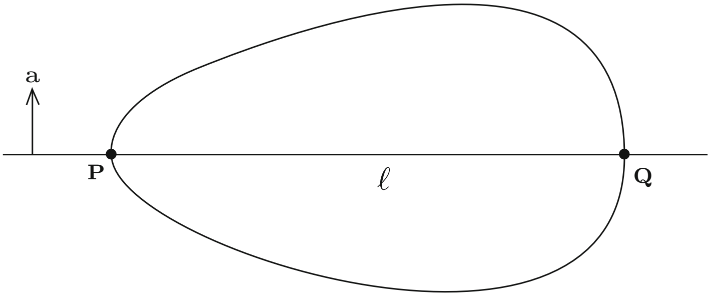
## Theorem 3.3.3 (Four Vertex Theorem)

Every convex simple closed curve in $\mathbb{R}^2$ has at least four vertices.

The conclusion of this theorem actually remains true without the assumption of convexity, but the proof is then more difficult than the one we are about to give.

## Proof

Let $\boldsymbol{\gamma}$ be a parametrization of a convex simple closed curve in $\mathbb{R}^2$, and let $\ell$ be its length. Assume for a contradiction that $\boldsymbol{\gamma}$ has fewer than four vertices. We show first that there is a straight line $L$ that divides $\boldsymbol{\gamma}$ into two segments, in one of which $\dot{\kappa}_s > 0$ and in the other $\dot{\kappa}_s \leq 0$ (or possibly $\dot{\kappa}_s \geq 0$ on one and $\dot{\kappa}_s < 0$ on the other). Indeed, $\kappa_s$ attains all of its values on the closed interval $[0, \ell]$, so $\kappa_s$ must attain its maximum and minimum values at some points $\mathbf{p}$ and $\mathbf{q}$ of $\boldsymbol{\gamma}$. We can assume that $\mathbf{p} \neq \mathbf{q}$, since otherwise $\kappa_s$ would be constant, $\boldsymbol{\gamma}$ would be a circle (by Example 2.2.7), and every point of $\boldsymbol{\gamma}$ would be a vertex. If $\mathbf{p}$ and $\mathbf{q}$ were the only vertices of $\boldsymbol{\gamma}$, we would have $\dot{\kappa}_s > 0$ on one of the segments into which the line through $\mathbf{p}$ and $\mathbf{q}$ divides $\boldsymbol{\gamma}$ and $\dot{\kappa}_s < 0$ on the other. Suppose now that there is just one more vertex, say $\mathbf{r}$. Then, $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ divide $\boldsymbol{\gamma}$ into three segments, on each of which either $\dot{\kappa}_s > 0$ or $\dot{\kappa}_s < 0$. It follows that there are two adjacent segments on which $\dot{\kappa}_s > 0$ or two on which $\dot{\kappa}_s < 0$ (except at the point at which the two segments meet). This proves our assertion. $\square$




Let $\mathbf{a}$ be a unit vector perpendicular to $L$, so that $\boldsymbol{\gamma}\cdot\mathbf{a} > 0$ on one side of $L$ and $\boldsymbol{\gamma}\cdot\mathbf{a} < 0$ on the other. Then, the quantity $\dot{\kappa}_s(\boldsymbol{\gamma}\cdot\mathbf{a})$ is either always $> 0$ or always $< 0$, except at the two points in which $L$ intersects the curve. It follows that

$$\int_0^\ell \dot{\kappa}_s(\boldsymbol{\gamma}\cdot\mathbf{a})\,dt \neq 0, \tag{3.8}$$

as this integral is definitely $> 0$ in the first case and $< 0$ in the second. But, using the equation $\dot{\mathbf{n}}_s = -\kappa_s\mathbf{t}$ (see Exercise 2.2.1), we get

$$\dot{\kappa}_s\boldsymbol{\gamma} = (\kappa_s\boldsymbol{\gamma})^{\boldsymbol{\cdot}} - \kappa_s\dot{\boldsymbol{\gamma}} = (\kappa_s\boldsymbol{\gamma} + \mathbf{n}_s)^{\boldsymbol{\cdot}},$$

so the integrand on the left-hand side of (3.8) is the derivative of $(\kappa_s\boldsymbol{\gamma}+\mathbf{n}_s)\cdot\mathbf{a} = \lambda$, say. Since $\boldsymbol{\gamma}$ is $\ell$-periodic,

$$\boldsymbol{\gamma}(t+\ell) = \boldsymbol{\gamma}(t) \quad \text{for all} \quad t,$$

differentiating with respect to $t$ shows that the tangent vector $\mathbf{t}$ of $\boldsymbol{\gamma}$ is also $\ell$-periodic:

$$\mathbf{t}(t+\ell) = \dot{\boldsymbol{\gamma}}(t+\ell) = \dot{\boldsymbol{\gamma}}(t) = \mathbf{t}(t).$$

Rotating by $\pi/2$ gives

$$\mathbf{n}_s(t+\ell) = \mathbf{n}_s(t),$$

and hence $\kappa_s(t+\ell) = \kappa_s(t)$. It follows that $\lambda(t+\ell) = \lambda(t)$ for all $t$, so the integral in (3.8) is equal to

$$\int_0^\ell \dot{\lambda}(t)\,dt = \lambda(\ell) - \lambda(0) = 0.$$

This contradiction proves that $\boldsymbol{\gamma}$ must have at least four vertices. $\square$

## *EXERCISES*

3.3.1 Show that the ellipse in Example 3.1.2 is convex.

3.3.2 Show that the limacęn in Example 1.1.7 has only two vertices (cf. Example 3.1.3).

3.3.3 Show that a plane curve $\boldsymbol{\gamma}$ has a vertex at $t = t_0$ if and only if the evolute $\boldsymbol{\epsilon}$ of $\boldsymbol{\gamma}$ (Exercise 2.2.7) has a singular point at $t = t_0$.

# 4
# Surfaces in three dimensions

In this chapter, we introduce several diffe nt ways to mathematically formulate the notion of a surface. Although the simplest of these, that of a surface patch, is all that is needed for most of the book, it does not describe adequately most of the objects that we would want to call surfaces. For example, a sphere is not a surface patch, but it can be described by 'gluing' two surface patches together suitably. The idea behind this gluing procedure is simple enough, but making it precise turns out to be a little complicated. We have tried to minimize the trauma by collecting the most demanding proofs in a separate section (Section 5.6).

## 4.1 What is a surface?

A surface is a subset of $\mathbb{R}^3$ that looks like a piece of $\mathbb{R}^2$ in the vicinity of any given point, just as the surface of the Earth, although actually nearly spherical, appears to be a flat plane to an observer on the surface who sees only to the horizon. To make the phrases 'looks like' and 'in the vicinity' precise, we must first introduce some preliminary material. We describe this for $\mathbb{R}^n$ for any $n \geq 1$, although we shall need it only for $n = 1, 2$, or 3.

First, a subset $U$ of $\mathbb{R}^n$ is called *open* if, whenever $\mathbf{a}$ is a point in $U$, there is a positive number $\epsilon$ such that every point $\mathbf{u} \in \mathbb{R}^n$ within a distance $\epsilon$ of $\mathbf{a}$ is also in $U$:

$$\mathbf{a} \in U \text{ and } \parallel \mathbf{u} - \mathbf{a} \parallel < \epsilon \quad \Longrightarrow \quad \mathbf{u} \in U.$$

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_4,

For example, the whole of $\mathbb{R}^n$ is an open set, as is

$$\mathcal{D}_r(\mathbf{a}) = \{\mathbf{u} \in \mathbb{R}^n \mid \parallel \mathbf{u} - \mathbf{a} \parallel < r\},$$

the *open ball* with centre $\mathbf{a}$ and radius $r > 0$. (If $n = 1$, an open ball is called an *open interval*; if $n = 2$ it is called an *open disc*.) However,

$$\overline{\mathcal{D}}_r(\mathbf{a}) = \{\mathbf{u} \in \mathbb{R}^n \mid \parallel \mathbf{u} - \mathbf{a} \parallel \leq r\}$$

is *not* open because however small the positive number $\epsilon$ is, there is a point within a distance $\epsilon$ of the point $(a_1 + r, a_2, \ldots, a_n) \in \overline{\mathcal{D}}_r(\mathbf{a})$ (say) that is not in $\overline{\mathcal{D}}_r(\mathbf{a})$ (for example, the point $(a_1 + r + \frac{\epsilon}{2}, a_2, \ldots, a_n)$).

Next, if $X$ and $Y$ are subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, a map $f : X \to Y$ is said to be continuous at a point $\mathbf{a} \in X$ if points in $X$ near $\mathbf{a}$ are mapped by $f$ to points in $Y$ near $f(\mathbf{a})$. More precisely, $f$ is *continuous at* $\mathbf{a}$ if, given any number $\epsilon > 0$, there is a number $\delta > 0$ such that

$$\mathbf{u} \in X \quad \text{and} \quad \parallel \mathbf{u} - \mathbf{a} \parallel < \delta \Longrightarrow \parallel f(\mathbf{u}) - f(\mathbf{a}) \parallel < \epsilon.$$

Then $f$ is said to be *continuous* if it is continuous at every point of $X$. Composites of continuous maps are continuous.

In view of the definition of an open set, this is equivalent to the following: $f$ is continuous if and only if, for any open set $V$ of $\mathbb{R}^n$, there is an open set $U$ of $\mathbb{R}^m$ such that $U \cap X = \{x \in X \mid f(x) \in V\}$.

If $f : X \to Y$ is continuous and bijective, and if its inverse map $f^{-1} : Y \to X$ is also continuous, then $f$ is called a *homeomorphism* and $X$ and $Y$ are said to be *homeomorphic*.

We are now in a position to make our first attempt at defining the notion of a surface in $\mathbb{R}^3$.

## Definition 4.1.1

A subset $\mathcal{S}$ of $\mathbb{R}^3$ is a *surface* if, for every point $\mathbf{p} \in \mathcal{S}$, there is an open set $U$ in $\mathbb{R}^2$ and an open set $W$ in $\mathbb{R}^3$ containing $\mathbf{p}$ such that $\mathcal{S} \cap W$ is homeomorphic to $U$. A subset of a surface $\mathcal{S}$ of the form $\mathcal{S} \cap W$, where $W$ is an open subset of $\mathbb{R}^3$, is called an *open subset* of $\mathcal{S}$. A homeomorphism $\boldsymbol{\sigma} : U \to \mathcal{S} \cap W$ as in this definition is called a *surface patch* or *parametrization* of the open subset $\mathcal{S} \cap W$ of $\mathcal{S}$. A collection of such surface patches whose images cover the whole of $\mathcal{S}$ is called an *atlas* of $\mathcal{S}$.

## Example 4.1.2

Every plane in $\mathbb{R}^3$ is a surface with an atlas consisting of a single surface patch. In fact, let $\mathbf{a}$ be a point on the plane, and let $\mathbf{p}$ and $\mathbf{q}$ be two unit vectors that are parallel to the plane and perpendicular to each other. If $\mathbf{v}$ is any point of the plane, $\mathbf{v} - \mathbf{a}$ is parallel to the plane, and so

$$\mathbf{v} - \mathbf{a} = u\mathbf{p} + v\mathbf{q}$$

for some scalars $u$ and $v$. Thus, the desired surface patch is

$$\boldsymbol{\sigma}(u, v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q},$$

and its inverse map is

$$\boldsymbol{\sigma}^{-1}(\mathbf{v}) = ((\mathbf{v} - \mathbf{a}) \cdot \mathbf{p}, (\mathbf{v} - \mathbf{a}) \cdot \mathbf{q}).$$

These formulas make it clear that $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^{-1}$ are continuous, and hence that $\boldsymbol{\sigma}$ is a homeomorphism. (We shall not verify this in detail.)

The following example shows why we have to consider surfaces, and not just surface patches.

## Example 4.1.3

A *circular cylinder* is the set of points of $\mathbb{R}^3$ that are at a fixed distance (the *radius* of the cylinder) from a fixed straight line (its *axis*). For example, the circular cylinder of radius 1 and axis the $z$-axis, which we shall call the *unit cylinder*, is

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 = 1\}.$$

The simplest parametrization of $\mathcal{S}$ is

$$\boldsymbol{\sigma}(u, v) = (\cos u, \sin u, v).$$

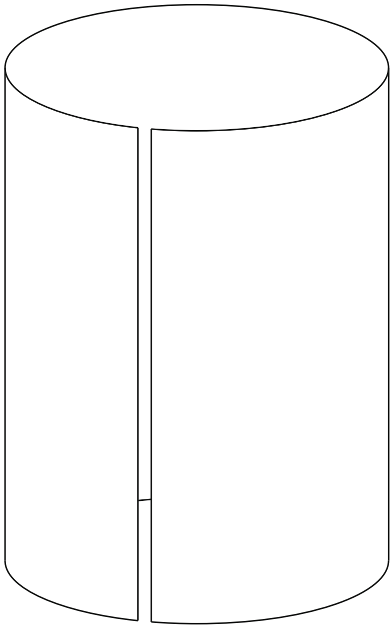
Clearly, $\boldsymbol{\sigma}(u, v) \in \mathcal{S}$ for all $(u, v) \in \mathbb{R}^2$, and every point of $\mathcal{S}$ is of this form. Moreover, $\boldsymbol{\sigma}$ is continuous. However, $\boldsymbol{\sigma}$ is not injective, and so is not a homeomorphism, because $\boldsymbol{\sigma}(u, v) = \boldsymbol{\sigma}(u + 2\pi, v)$ for all $(u, v)$. To get an injective map we can restrict $u$ to lie in an interval of length $\leq 2\pi$, say $0 \leq u < 2\pi$. However, although the restriction $\boldsymbol{\sigma}|_V$ of $\boldsymbol{\sigma}$ to

$$V = \{(u, v) \in \mathbb{R}^2 \,|\, 0 \leq u < 2\pi\}$$

is injective, $V$ is not an open subset of $\mathbb{R}^2$ and so $\boldsymbol{\sigma}|_V$ is not a surface patch. The largest open subset of $\mathbb{R}^2$ contained in $V$ is

$$U = \{(u, v) \in \mathbb{R}^2 \,|\, 0 < u < 2\pi\},$$

and the restriction $\boldsymbol{\sigma}|_U$ of $\boldsymbol{\sigma}$ to $U$ *is* a surface patch. However, $\boldsymbol{\sigma}|_U$ does not cover the whole of $\mathcal{S}$, but only the open subset obtained by removing the line $x = 1, y = 0$ from $\mathcal{S}$.



To get an atlas for $\mathcal{S}$ we therefore need at least one more surface patch. We can take $\boldsymbol{\sigma}|_{\tilde{U}}$, where

$$\tilde{U} = \{(u,v) \in \mathbb{R}^2 \mid -\pi < u < \pi\};$$

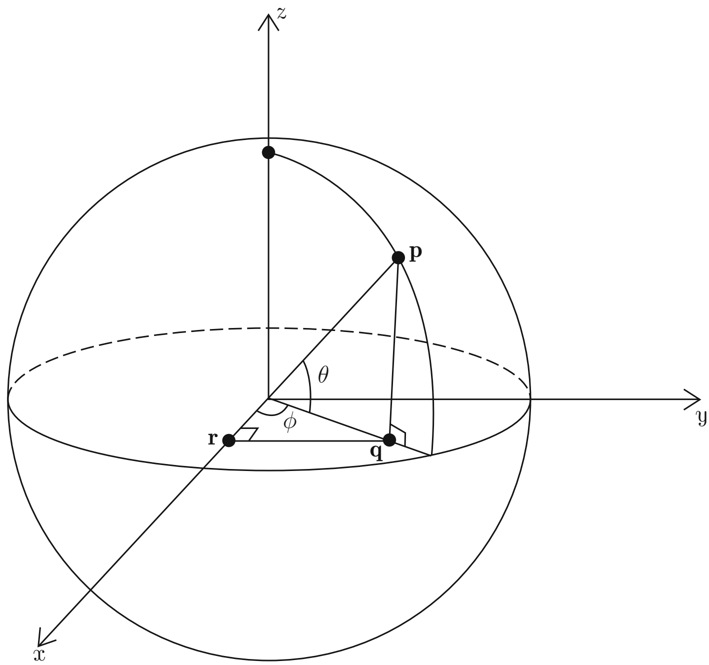
this covers the open subset of $\mathcal{S}$ obtained by removing the line $x = -1, y = 0$. Every point of $\mathcal{S}$ is in the image of at least one of the surface patches $\boldsymbol{\sigma}|_U, \boldsymbol{\sigma}|_{\tilde{U}}$, so $\{\boldsymbol{\sigma}|_U, \boldsymbol{\sigma}|_{\tilde{U}}\}$ is an atlas for $\mathcal{S}$, and $\mathcal{S}$ is a surface.

## Example 4.1.4

A *sphere* is the set of points of $\mathbb{R}^3$ that are a fixed distance (the *radius* of the sphere) from a fixed point (its *centre*). For example, the sphere of radius 1 and centre the origin, called the *unit sphere*, is

$$S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}.$$

The most popular parametrization of $S^2$ is that given by latitude $\theta$ and longitude $\varphi$: if $\mathbf{p}$ is a point of the sphere, the line through $\mathbf{p}$ parallel to the $z$-axis intersects the $xy$-plane at a point $\mathbf{q}$, say; then, $\theta$ is the angle between $\mathbf{q}$ and $\mathbf{p}$ and $\varphi$ is the angle between $\mathbf{q}$ and the positive $x$-axis. The circles on the sphere corresponding to a constant value of $\theta$ are called *parallels*; those corresponding to a constant value of $\varphi$ are called *meridians*.




To obtain an explicit formula for this parametrization, we must express $\mathbf{p}$ in terms of the angles $\theta$ and $\varphi$. From the right-angled triangle with vertices $\mathbf{0}$, $\mathbf{p}$ and $\mathbf{q}$, we see that the $z$-component of $\mathbf{p}$ is $\sin\theta$. The $x$- and $y$-components can be found from the right-angled triangle in the $xy$-plane with vertices $\mathbf{0}$, $\mathbf{q}$ and $\mathbf{r}$, where $\mathbf{r}$ is the foot of the perpendicular from $\mathbf{q}$ to the $x$-axis. The length of the hypotenuse of this triangle is $\| \mathbf{q} \| = \cos\theta$, so the $x$- and $y$-components of $\mathbf{p}$ are

$$\| \mathbf{q} \| \cos\varphi = \cos\theta\cos\varphi \quad \text{and} \quad \| \mathbf{q} \| \sin\varphi = \cos\theta\sin\varphi,$$

respectively. Putting all these together gives

$$\mathbf{p} = (\cos\theta\cos\varphi, \cos\theta\sin\varphi, \sin\theta).$$

Denote the right-hand side of this equation by $\boldsymbol{\sigma}(\theta,\varphi)$; this is the *latitude–longitude parametrization* of $S^2$.

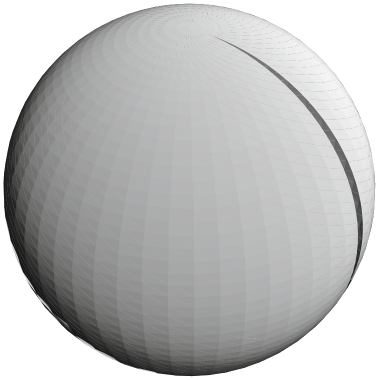
As in the case of the cylinder, $\boldsymbol{\sigma}$ is not injective since (for example) $\boldsymbol{\sigma}(\theta,\varphi) = \boldsymbol{\sigma}(\theta,\varphi+2\pi)$. In fact, a little thought shows that to cover the whole sphere, it is sufficient to take

$$-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}, \quad 0 \leq \varphi \leq 2\pi.$$

However, the set of points $(\theta, \varphi)$ satisfying these inequalities is not an open subset of $\mathbb{R}^2$. The largest open set consistent with the above inequalities is

$$U = \left\{ (\theta, \varphi) \;\middle|\; -\frac{\pi}{2} < \theta < \frac{\pi}{2},\ 0 < \varphi < 2\pi \right\};$$

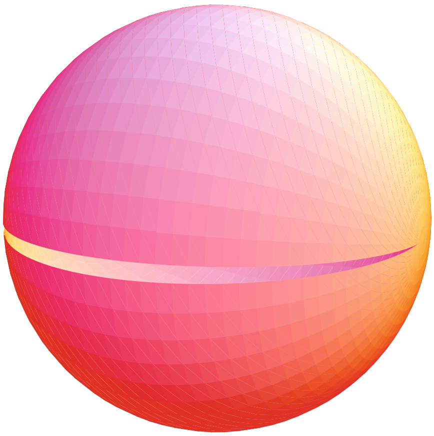
however, the image of $\boldsymbol{\sigma}|_U$ is not the whole of the sphere, but the open subset obtained by removing the semicircle $\mathcal{C}$ consisting of the points of the sphere of the form $(x, 0, z)$ with $x \geq 0$.



To show that the sphere is a surface, we must therefore produce at least one more surface patch covering the part of the sphere omitted by $\boldsymbol{\sigma}$. One possibility is the patch $\tilde{\boldsymbol{\sigma}}$ obtained by first rotating $\boldsymbol{\sigma}$ by $\pi$ about the $z$-axis and then by $\pi/2$ about the $x$-axis. Explicitly, $\tilde{\boldsymbol{\sigma}} : U \to \mathbb{R}^3$ is given by

$$\tilde{\boldsymbol{\sigma}}(\theta, \varphi) = (-\cos\theta\cos\varphi, -\sin\theta, -\cos\theta\sin\varphi)$$

(the open set $U$ is the same as for $\boldsymbol{\sigma}$). The image of $\tilde{\boldsymbol{\sigma}}$ is the open subset of $S^2$ obtained by removing the semicircle $\tilde{\mathcal{C}}$ consisting of the points of the sphere of the form $(x, y, 0)$ with $x \leq 0$.



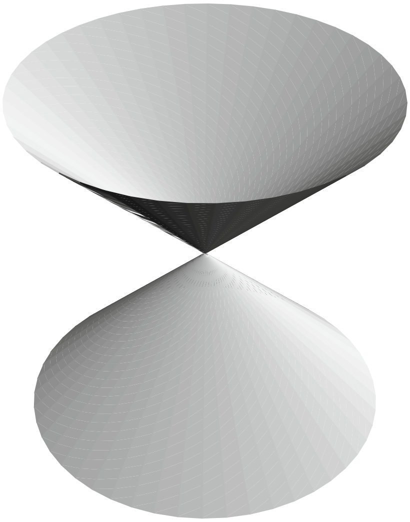
It is clear that $\mathcal{C}$ and $\tilde{\mathcal{C}}$ do not intersect, so the union of the images of $\boldsymbol{\sigma}|_U$ and $\tilde{\boldsymbol{\sigma}}|_U$ is the whole sphere.

Our last example (for the moment) is a subset of $\mathbb{R}^3$ that is nearly, but not quite, a surface.

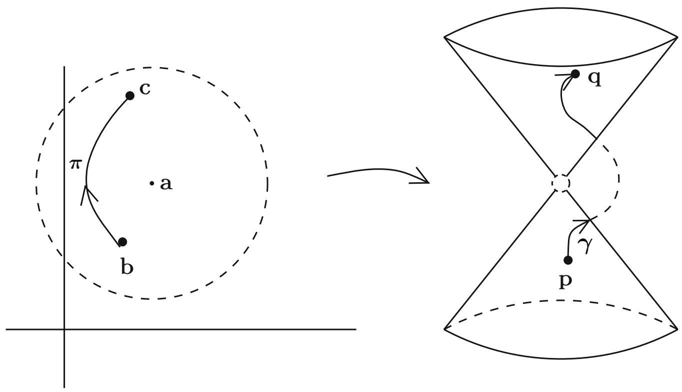
## Example 4.1.5

The *circular cone* with *vertex* a point $\mathbf{v}$, *axis* a straight line $\ell$ passing through $\mathbf{v}$, and *angle* $\alpha$, where $0 < \alpha < \pi/2$, is the set of points $\mathbf{p}$ in $\mathbb{R}^3$ such that the straight line through $\mathbf{v}$ and $\mathbf{p}$ makes an angle $\alpha$ with the line $\ell$. For example, if $\mathbf{v}$ is the origin, $\ell$ is the $z$-axis and $\alpha = \pi/4$, the circular cone is

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

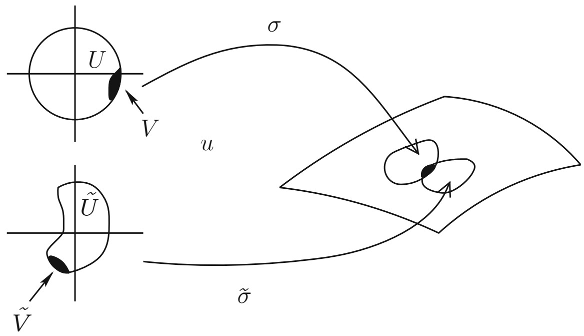


To see that this is *not* a surface, suppose that $\boldsymbol{\sigma} : U \to \mathcal{S} \cap W$ is a surface patch containing the vertex $(0, 0, 0)$ of the cone, and let $\mathbf{a} \in U$ correspond to the vertex. We can assume that $U$ is an open ball with centre $\mathbf{a}$, since any open set $U$ containing $\mathbf{a}$ must contain such an open ball. The open set $W$ must obviously contain a point $\mathbf{p}$ in the lower half $\mathcal{S}_-$ of $\mathcal{S}$ where $z < 0$ and a point $\mathbf{q}$ in the upper half $\mathcal{S}_+$ where $z > 0$; let $\mathbf{b}$ and $\mathbf{c}$ be the corresponding points in $U$. It is clear that there is a curve $\boldsymbol{\pi}$ in $U$ passing through $\mathbf{b}$ and $\mathbf{c}$, but not passing through $\mathbf{a}$. This is mapped by $\boldsymbol{\sigma}$ into the curve $\boldsymbol{\gamma} = \boldsymbol{\sigma} \circ \boldsymbol{\pi}$ lying entirely in $\mathcal{S}$, passing through $\mathbf{p}$ and $\mathbf{q}$, and not passing through the vertex. (It is true that $\boldsymbol{\gamma}$ will, in general, only be continuous, and not smooth, but this does not affect the argument.) This is clearly impossible. (Readers familiar with point set topology will be able to make this heuristic argument rigorous.)




If we remove the vertex, however, we do get a surface $\mathcal{S}_- \cup \mathcal{S}_+$. It has an atlas consisting of the two surface patches $\boldsymbol{\sigma}_\pm : U \to \mathbb{R}^3$, where $U = \mathbb{R}^2 \backslash \{(0,0)\}$, given by the inverse of projection onto the $xy$ plane:

$$\boldsymbol{\sigma}_\pm(u,v) = (u, v, \pm\sqrt{u^2+v^2}).$$




As the example of the sphere shows, a point $\mathbf{a}$ of a surface $\mathcal{S}$ will generally lie in the image of more than one surface patch. In general, suppose then that $\boldsymbol{\sigma} : U \to \mathcal{S} \cap W$ and $\tilde{\boldsymbol{\sigma}} : \tilde{U} \to \mathcal{S} \cap \tilde{W}$ are two patches such that $\mathbf{a} \in \mathcal{S} \cap W \cap \tilde{W}$. Since $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are homeomorphisms, $\boldsymbol{\sigma}^{-1}(\mathcal{S} \cap W \cap \tilde{W})$ and $\tilde{\boldsymbol{\sigma}}^{-1}(\mathcal{S} \cap W \cap \tilde{W})$ are open sets $V \subseteq U$ and $\tilde{V} \subseteq \tilde{U}$, respectively. The composite homeomorphism $\boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}} : \tilde{V} \to V$ is called the *transition map* from $\boldsymbol{\sigma}$ to $\tilde{\boldsymbol{\sigma}}$. If we denote this map by $\Phi$, we have

$$\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v}) = \boldsymbol{\sigma}(\Phi(\tilde{u}, \tilde{v}))$$

for all $(\tilde{u}, \tilde{v}) \in \tilde{V}$.

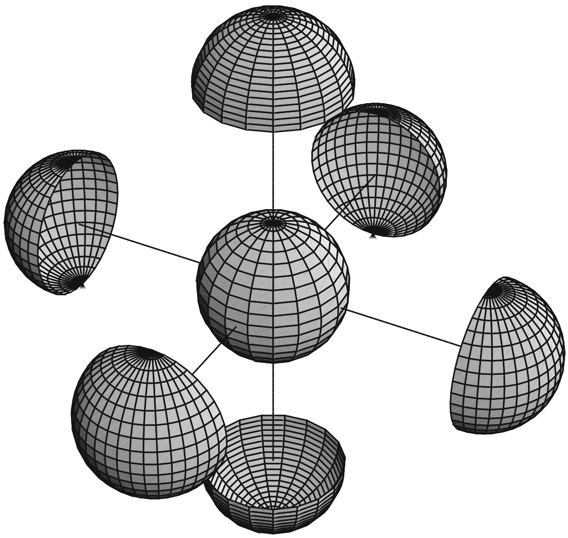
## EXERCISES

4.1.1 Show that any open disc in the $xy$-plane is a surface.

4.1.2 Define surface patches $\boldsymbol{\sigma}_{\pm}^{x} : U \to \mathbb{R}^3$ for $S^2$ by solving the equation $x^2 + y^2 + z^2 = 1$ for $x$ in terms of $y$ and $z$:

$$\boldsymbol{\sigma}_{\pm}^{x}(u, v) = (\pm\sqrt{1 - u^2 - v^2}, u, v),$$

defined on the open set $U = \{(u, v) \in \mathbb{R}^2 \mid u^2 + v^2 < 1\}$. Define $\boldsymbol{\sigma}_{\pm}^{y}$ and $\boldsymbol{\sigma}_{\pm}^{z}$ similarly (with the same $U$) by solving for $y$ and $z$, respectively. Show that these six patches give $S^2$ the structure of a surface.



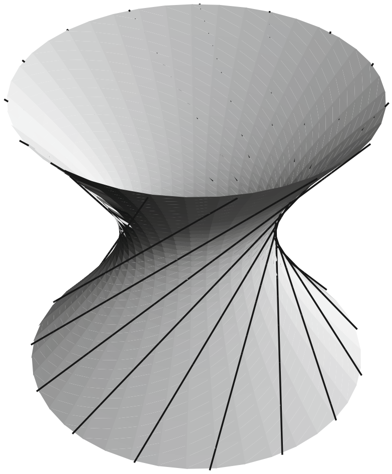
4.1.3 The *hyperboloid of one sheet* is

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 - z^2 = 1\}.$$

Show that, for every $\theta$, the straight line

$$(x - z)\cos\theta = (1 - y)\sin\theta, \quad (x + z)\sin\theta = (1 + y)\cos\theta$$

is contained in $\mathcal{S}$, and that every point of the hyperboloid lies on one of these lines. Deduce that $\mathcal{S}$ can be covered by a single surface patch, and hence is a surface. (Compare the case of the cylinder in Example 4.1.3.)



Find a second family of straight lines on $\mathcal{S}$, and show that no two lines of the same family intersect, while every line of the first family intersects every line of the second family with one exception. One says that the surface $\mathcal{S}$ is *doubly ruled*.

4.1.4 Show that the unit cylinder can be covered by a single surface patch, but that the unit sphere cannot. (The second part requires some point set topology.)

4.1.5 Show that every open subset of a surface is a surface.

## 4.2 Smooth surfaces

In Differential Geometry we use calculus to analyse surfaces (and other geometric objects). We must be able to make sense of the statement that a function on a surface is differentiable, for example. For this, we have to consider surfaces with some extra structure.

First, if $U$ is an open subset of $\mathbb{R}^m$, we say that a map $\mathbf{f} : U \to \mathbb{R}^n$ is *smooth* if each of the $n$ components of $\mathbf{f}$, which are functions $U \to \mathbb{R}$, have continuous partial derivatives of all orders. The partial derivatives of $\mathbf{f}$ are then computed componentwise. For example, if $m = 2$ and $n = 3$, and

$$\mathbf{f}(u, v) = (f_1(u, v), f_2(u, v), f_3(u, v)),$$

then

$$\frac{\partial \mathbf{f}}{\partial u} = \left(\frac{\partial f_1}{\partial u}, \frac{\partial f_2}{\partial u}, \frac{\partial f_3}{\partial u}\right), \quad \frac{\partial \mathbf{f}}{\partial v} = \left(\frac{\partial f_1}{\partial v}, \frac{\partial f_2}{\partial v}, \frac{\partial f_3}{\partial v}\right),$$

and similarly for higher derivatives. We often use the following abbreviations:

$$\frac{\partial \mathbf{f}}{\partial u} = \mathbf{f}_u, \quad \frac{\partial \mathbf{f}}{\partial v} = \mathbf{f}_v,$$
$$\frac{\partial^2 \mathbf{f}}{\partial u^2} = \mathbf{f}_{uu}, \quad \frac{\partial^2 \mathbf{f}}{\partial u \partial v} = \mathbf{f}_{uv}, \quad \frac{\partial^2 \mathbf{f}}{\partial v \partial u} = \mathbf{f}_{vu}, \quad \frac{\partial^2 \mathbf{f}}{\partial v^2} = \mathbf{f}_{vv},$$

and so on. From advanced calculus we know that $\mathbf{f}_{uv} = \mathbf{f}_{vu}$ if $f$ is smooth.

It now makes sense to say that a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is smooth. But we shall require one further condition.

### Definition 4.2.1

A surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is called *regular* if it is smooth and the vectors $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent at all points $(u, v) \in U$. Equivalently, $\boldsymbol{\sigma}$ should be smooth and the vector product $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$ should be non-zero at every point of $U$.

The reason for imposing this condition will appear in Section 4.4.

We can finally define the class of surfaces to be studied in this book.

### Definition 4.2.2

If $\mathcal{S}$ is a surface, an *allowable surface patch* for $\mathcal{S}$ is a regular surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ such that $\boldsymbol{\sigma}$ is a homeomorphism from $U$ to an open subset of $\mathcal{S}$. A *smooth surface* is a surface $\mathcal{S}$ such that, for any point $\mathbf{p} \in \mathcal{S}$, there is an allowable surface patch $\boldsymbol{\sigma}$ as above such that $\mathbf{p} \in \boldsymbol{\sigma}(U)$. A collection $\mathcal{A}$ of allowable surface patches for a surface $\mathcal{S}$ such that every point of $\mathcal{S}$ is in the image of at least one patch in $\mathcal{A}$ is called an *atlas* for the smooth surface $\mathcal{S}$.

### Example 4.2.3

The plane in Example 4.1.2 is a smooth surface. For

$$\boldsymbol{\sigma}(u, v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q}$$

is clearly smooth and $\boldsymbol{\sigma}_u = \mathbf{p}$ and $\boldsymbol{\sigma}_v = \mathbf{q}$ are linearly independent because $\mathbf{p}$ and $\mathbf{q}$ were chosen to be perpendicular unit vectors.

### Example 4.2.4

The unit cylinder (Example 4.1.3) is a smooth surface. Indeed,

$$\boldsymbol{\sigma}(u, v) = (\cos u, \sin u, v)$$

is clearly smooth and

$$\boldsymbol{\sigma}_u = (-\sin u, \cos u, 0), \quad \boldsymbol{\sigma}_v = (0, 0, 1)$$

are obviously linearly independent for all $(u, v)$, so $\boldsymbol{\sigma}|_U$ and $\boldsymbol{\sigma}|_{\tilde{U}}$ are regular surface patches.

## Example 4.2.5

For the unit sphere $S^2$ in Example 4.1.4, it is again clear that $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are smooth. As for regularity,

$$\boldsymbol{\sigma}_\theta = (-\sin\theta\cos\varphi, -\sin\theta\sin\varphi, \cos\theta), \quad \boldsymbol{\sigma}_\varphi = (-\cos\theta\sin\varphi, \cos\theta\cos\varphi, 0),$$
$$\boldsymbol{\sigma}_\theta \times \boldsymbol{\sigma}_\varphi = (-\cos^2\theta\cos\varphi, -\cos^2\theta\sin\varphi, -\sin\theta\cos\theta)$$

and hence $\| \boldsymbol{\sigma}_\theta \times \boldsymbol{\sigma}_\varphi \| = |\cos\theta|$. But if $(\theta, \varphi) \in U$, then $-\pi/2 < \theta < \pi/2$ so $\cos\theta \neq 0$. Similarly, one checks that $\tilde{\boldsymbol{\sigma}}$ is regular.

In Exercise 4.1.2 we gave another family of allowable surface patches covering the unit sphere $S^2$ (it is easy to check that they are regular – see Exercise 4.2.2). An obvious question is: which of these two atlases should we use to study the sphere? The answer is that we can use either, or both. The eight patches in Exercise 4.1.2 and Example 4.1.4 together form a third atlas. In most situations (although not in all – see Definition 4.5.1), one might as well use the *maximal atlas* for a given surface $\mathcal{S}$ consisting of *all* of its allowable surface patches. The maximal atlas is independent of any arbitrary choices.

Although not at first sight very interesting, the next two results are very important for what is to follow.

## Proposition 4.2.6

The transition maps of a smooth surface are smooth.

The proof of this will be given in Section 5.6. The next result is a kind of converse.

## Proposition 4.2.7

Let $U$ and $\tilde{U}$ be open subsets of $\mathbb{R}^2$ and let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular surface patch. Let $\Phi : \tilde{U} \to U$ be a bijective smooth map with smooth inverse map $\Phi^{-1} : U \to \tilde{U}$. Then, $\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi : \tilde{U} \to \mathbb{R}^3$ is a regular surface patch.

## Proof

The patch $\tilde{\boldsymbol{\sigma}}$ is smooth because any composite of smooth maps is smooth. As for regularity, let $(u, v) = \Phi(\tilde{u}, \tilde{v})$. By the chain rule,

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}} = \frac{\partial u}{\partial \tilde{u}}\boldsymbol{\sigma}_u + \frac{\partial v}{\partial \tilde{u}}\boldsymbol{\sigma}_v, \quad \tilde{\boldsymbol{\sigma}}_{\tilde{v}} = \frac{\partial u}{\partial \tilde{v}}\boldsymbol{\sigma}_u + \frac{\partial v}{\partial \tilde{v}}\boldsymbol{\sigma}_v,$$

so

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} = \left(\frac{\partial u}{\partial \tilde{u}}\frac{\partial v}{\partial \tilde{v}} - \frac{\partial u}{\partial \tilde{v}}\frac{\partial v}{\partial \tilde{u}}\right)\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v. \tag{4.1}$$

The scalar on the right-hand side of this equation is the determinant of the *Jacobian matrix*

$$J(\Phi) = \begin{pmatrix} \frac{\partial u}{\partial \tilde{u}} & \frac{\partial v}{\partial \tilde{u}} \\ \frac{\partial u}{\partial \tilde{v}} & \frac{\partial v}{\partial \tilde{v}} \end{pmatrix}$$

of $\Phi$. We recall from calculus that, if $\Psi$ and $\tilde{\Psi}$ are two smooth maps between open sets of $\mathbb{R}^2$,

$$J(\tilde{\Psi} \circ \Psi) = J(\tilde{\Psi})J(\Psi).$$

(In fact, this is equivalent to the chain rule that expresses the first partial derivatives of $\tilde{\Psi} \circ \Psi$ in terms of those of $\tilde{\Psi}$ and $\Psi$.) Taking $\Psi = \Phi$ and $\tilde{\Psi} = \Phi^{-1}$, we see that $J(\Phi^{-1}) = J(\Phi)^{-1}$. In particular, $J(\Phi)$ is invertible, so its determinant is non-zero and Eq. 4.1 shows that $\tilde{\boldsymbol{\sigma}}$ is regular. □

If regular surface patches $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are related as in this proposition, we say that $\tilde{\boldsymbol{\sigma}}$ is a *reparametrization* of $\boldsymbol{\sigma}$, and that $\Phi$ is a *reparametrization map*. Note that $\boldsymbol{\sigma}$ is then a reparametrization of $\tilde{\boldsymbol{\sigma}}$, since $\boldsymbol{\sigma} = \tilde{\boldsymbol{\sigma}} \circ \Phi^{-1}$.

Note also that, if $\boldsymbol{\sigma} : U \to \mathcal{S} \cap W$ and $\tilde{\boldsymbol{\sigma}} : \tilde{U} \to \mathcal{S} \cap \tilde{W}$ are two allowable surface patches of a smooth surface $\mathcal{S}$, and if $V \subseteq U$ and $\tilde{V} \subseteq \tilde{U}$ are the open subsets such that $\boldsymbol{\sigma}(V) = \tilde{\boldsymbol{\sigma}}(\tilde{V}) = \mathcal{S} \cap W \cap \tilde{W}$, then $\Phi = \boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}} : \tilde{V} \to V$ is bijective, smooth and has a smooth inverse by Proposition 4.2.6. Thus, $\tilde{\boldsymbol{\sigma}}$ *is a reparametrization of* $\boldsymbol{\sigma}$ *where they are both defined.*

These observations give rise to a very important principle that we shall use throughout the book. The principle is that *we can define a property of any smooth surface provided we can define it for any regular surface patch in such a way that it is unchanged when the patch is reparametrized.* We shall give an important application of this principle in the next section.

*For the rest of this book, by a surface we shall mean a smooth surface $\mathcal{S}$, and by a surface patch for $\mathcal{S}$ we shall mean an allowable surface patch for $\mathcal{S}$.*

Unless we indicate otherwise, we shall also assume that all surfaces we consider are *connected*, which means that any two points of $\mathcal{S}$ can be joined by a curve lying entirely in $\mathcal{S}$. This is not a serious restriction, for it is not difficult to prove that any surface $\mathcal{S}$ is a disjoint union of connected surfaces, each of which is an open subset of $\mathcal{S}$, so $\mathcal{S}$ can be studied by studying each of its connected parts separately (see Exercise 4.2.9). All the surfaces we have encountered so far are connected except the double cone of Example 4.1.5, which breaks into the union of two disjoint half cones $\mathcal{S}_{\pm}$ when the vertex is removed, as it must be to have a surface.

## EXERCISES

4.2.1 Show that, if $f(x, y)$ is a smooth function, its *graph*

$$\{(x, y, z) \in \mathbb{R}^3 \mid z = f(x, y)\}$$

is a smooth surface with atlas consisting of the single regular surface patch

$$\boldsymbol{\sigma}(u, v) = (u, v, f(u, v)).$$

In fact, every surface is "locally" of this type – see Exercise 5.6.4.

4.2.2 Verify that the six surface patches for $S^2$ in Exercise 4.1.2 are regular. Calculate the transition maps between them and verify that these maps are smooth.

4.2.3 Which of the following are regular surface patches (in each case, $u, v \in \mathbb{R}$):

(i) $\boldsymbol{\sigma}(u, v) = (u, v, uv)$.

(ii) $\boldsymbol{\sigma}(u, v) = (u, v^2, v^3)$.

(iii) $\boldsymbol{\sigma}(u, v) = (u + u^2, v, v^2)$?

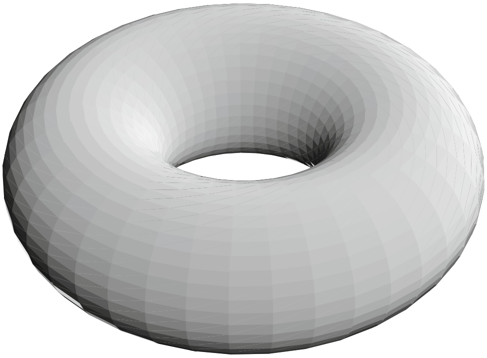
4.2.4 Show that the ellipsoid

$$\frac{x^2}{p^2} + \frac{y^2}{q^2} + \frac{z^2}{r^2} = 1,$$

where $p$, $q$ and $r$ are non-zero constants, is a smooth surface.

4.2.5 A *torus* is obtained by rotating a circle $\mathcal{C}$ in a plane $\Pi$ around a straight line $l$ in $\Pi$ that does not intersect $\mathcal{C}$. Take $\Pi$ to be the $xz$-plane, $l$ to be the $z$-axis, $a > 0$ the distance of the centre of $\mathcal{C}$ from $l$, and $b < a$ the radius of $\mathcal{C}$. Show that the torus is a smooth surface with parametrization

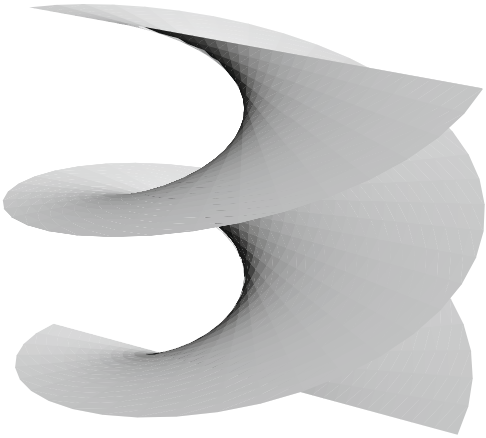
$$\boldsymbol{\sigma}(\theta, \varphi) = ((a + b\cos\theta)\cos\varphi, (a + b\cos\theta)\sin\varphi, b\sin\theta).$$



4.2.6 A *helicoid* is the surface swept out by an aeroplane propeller, when both the aeroplane and its propeller move at constant speed (see the picture below). If the aeroplane is flying along the $z$-axis, show that the helicoid can be parametrized as

$$\boldsymbol{\sigma}(u, v) = (v \cos u, v \sin u, \lambda u),$$

where $\lambda$ is a constant. Show that the cotangent of the angle that the standard unit normal of $\boldsymbol{\sigma}$ at a point $\mathbf{p}$ makes with the $z$-axis is proportional to the distance of $\mathbf{p}$ from the $z$-axis.

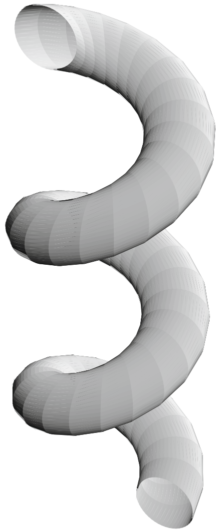


4.2.7 Let $\boldsymbol{\gamma}$ be a unit-speed curve in $\mathbb{R}^3$ with nowhere vanishing curvature. The *tube* of radius $a > 0$ around $\boldsymbol{\gamma}$ is the surface parametrized by

$$\boldsymbol{\sigma}(s, \theta) = \boldsymbol{\gamma}(s) + a(\mathbf{n}(s) \cos \theta + \mathbf{b}(s) \sin \theta),$$

where $\mathbf{n}$ is the principal normal of $\boldsymbol{\gamma}$ and $\mathbf{b}$ is its binormal. Give a geometrical description of this surface. Prove that $\boldsymbol{\sigma}$ is regular if the curvature $\kappa$ of $\boldsymbol{\gamma}$ is less than $a^{-1}$ everywhere.

Note that, even if $\boldsymbol{\sigma}$ is regular, the surface $\boldsymbol{\sigma}$ will have self-intersections if the curve $\boldsymbol{\gamma}$ comes within a distance $2a$ of itself. This illustrates the fact that regularity is a *local* property: if $(s, \theta)$ is restricted to lie in a sufficiently small open subset $U$ of $\mathbb{R}^2$, $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ will be smooth and injective (so there will be no self-intersections) – see Exercise 5.6.3. We shall see other instances of this later (for example, Example 12.2.5).



The tube around a circular helix

4.2.8 Show that translations and invertible linear transformations of $\mathbb{R}^3$ take smooth surfaces to smooth surfaces.

4.2.9 Show that every open subset of a smooth surface is a smooth surface.

## 4.3 Smooth maps

We want to define the notion of a *smooth map* $f : \mathcal{S}_1 \to \mathcal{S}_2$, where $\mathcal{S}_1$ and $\mathcal{S}_2$ are smooth surfaces. It is not obvious how to do this, because so far we only know how to define smooth maps between open subsets of Euclidean spaces.

In view of the principle stated at the end of the preceding section, we can assume that $\mathcal{S}_1$ and $\mathcal{S}_2$ are covered by single surface patches $\boldsymbol{\sigma}_1 : U_1 \to \mathbb{R}^3$ and $\boldsymbol{\sigma}_2 : U_2 \to \mathbb{R}^3$ provided we verify that the definition we give is unaffected by a reparametrization of $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$. Since $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are bijective, any map $f : \mathcal{S}_1 \to \mathcal{S}_2$ gives rise to the map $\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1 : U_1 \to U_2$, and we say that $f$ is *smooth* if this map is smooth (we already know what it means for a map between open subsets of $\mathbb{R}^2$ to be smooth). Now suppose that $\tilde{\boldsymbol{\sigma}}_1 : \tilde{U}_1 \to \mathbb{R}^3$ and $\tilde{\boldsymbol{\sigma}}_2 : \tilde{U}_2 \to \mathbb{R}^3$ are reparametrizations of $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, with reparametrization maps $\Phi_1 : \tilde{U}_1 \to U_1$ and $\Phi_2 : \tilde{U}_2 \to U_2$, respectively. We have to show that the corresponding map $\tilde{\boldsymbol{\sigma}}_2^{-1} \circ f \circ \tilde{\boldsymbol{\sigma}}_1 : \tilde{U}_1 \to \tilde{U}_2$ is smooth if $\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1 : U_1 \to U_2$ is smooth. But this is true, since

$$\begin{aligned}
\tilde{\boldsymbol{\sigma}}_2^{-1} \circ f \circ \tilde{\boldsymbol{\sigma}}_1 &= \tilde{\boldsymbol{\sigma}}_2^{-1} \circ (\boldsymbol{\sigma}_2 \circ \boldsymbol{\sigma}_2^{-1}) \circ f \circ (\boldsymbol{\sigma}_1 \circ \boldsymbol{\sigma}_1^{-1}) \circ \tilde{\boldsymbol{\sigma}}_1 \\
&= (\tilde{\boldsymbol{\sigma}}_2^{-1} \circ \boldsymbol{\sigma}_2) \circ (\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1) \circ (\boldsymbol{\sigma}_1^{-1} \circ \tilde{\boldsymbol{\sigma}}_1) \\
&= \Phi_2^{-1} \circ (\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1) \circ \Phi_1,
\end{aligned}$$

and $\Phi_1$ and $\Phi_2^{-1}$ are smooth maps (between open subsets of $\mathbb{R}^2$). The reader should check that composites of smooth maps between surfaces are smooth.

We shall be especially interested in smooth maps $f : \mathcal{S}_1 \to \mathcal{S}_2$, which are bijective and whose inverse map $f^{-1} : \mathcal{S}_2 \to \mathcal{S}_1$ is smooth. Such maps are called *diffeomorphisms*, and $\mathcal{S}_1$ and $\mathcal{S}_2$ are said to be *diffeomorphic* if there is a diffeomorphism between them. The following observation will be useful.

## Proposition 4.3.1

Let $f : \mathcal{S}_1 \to \mathcal{S}_2$ be a diffeomorphism. If $\boldsymbol{\sigma}_1$ is an allowable surface patch on $\mathcal{S}_1$, then $f \circ \boldsymbol{\sigma}_1$ is an allowable surface patch on $\mathcal{S}_2$.

## Proof

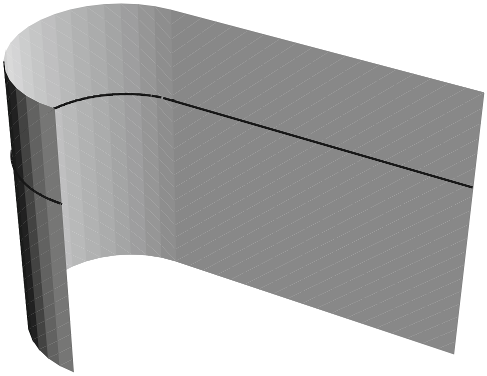
We can assume that $\mathcal{S}_1$ and $\mathcal{S}_2$ are covered by single allowable patches $\boldsymbol{\sigma}_1 : U_1 \to \mathbb{R}^3$ and $\boldsymbol{\sigma}_2 : U_2 \to \mathbb{R}^3$, respectively. Since $f$ is a diffeomorphism, $f(\boldsymbol{\sigma}_1(u, v)) = \boldsymbol{\sigma}_2(F(u, v))$, where $F : U_1 \to U_2$ is bijective, smooth and $F^{-1}$ is smooth. The result now follows from Proposition 4.2.6. □

It will actually be useful to consider smooth maps which satisfy a condition slightly weaker than being a diffeomorphism. A smooth map $f : \mathcal{S}_1 \to \mathcal{S}_2$ between smooth surfaces is called a *local diffeomorphism* if, for any point $\mathbf{p} \in \mathcal{S}_1$, there is an open subset $\mathcal{O}$ of $\mathcal{S}_1$ such that $f(\mathcal{O})$ is an open subset of $\mathcal{S}_2$ and $f|_{\mathcal{O}} : \mathcal{O} \to f(\mathcal{O})$ is a diffeomorphism (note that open subsets of surfaces are surfaces – see Exercise 4.2.9). It is obvious that every diffeomorphism is a local diffeomorphism (take $\mathcal{O} = \mathcal{S}_1$). Moreover, Proposition 4.3.1 holds if $f$ is a local diffeomorphism, provided that the restriction of $f$ to the image of $\boldsymbol{\sigma}_1$ is injective.

### Example 4.3.2

We consider the map from the $yz$-plane onto the unit cylinder $\mathcal{S}$, which wraps each line parallel to the $y$-axis around the 'waist' of the cylinder at height $z$ above the $xy$-plane. This map is given by

$$f(0, y, z) = (\cos y, \sin y, z).$$



Clearly, $f$ is not a diffeomorphism because it is not injective – the plane wraps around $\mathcal{S}$ infinitely many times. To see that $f$ is a local diffeomorphism, we parametrize the $yz$-plane in the obvious way by the single surface patch $\boldsymbol{\pi}(u, v) = (0, u, v)$, and use the atlas $\{\boldsymbol{\sigma}|_U, \boldsymbol{\sigma}|_{\tilde{U}}\}$ of $\mathcal{S}$ in Example 4.1.3. Let $\mathbf{p}$ be any point in the $yz$-plane, say $(0, a, b)$. If $a$ is not an even multiple of $2\pi$, there is an integer $n$ such that $2n\pi < a < 2(n+1)\pi$ and we have

$$f(\boldsymbol{\pi}(u, v)) = \boldsymbol{\sigma}(u - 2n\pi, v) \quad \text{if} \quad 2n\pi < u < 2(n+1)\pi,$$

showing that $f$ is a diffeomorphism from the open subset

$$\mathcal{O} = \{(0, y, z) \,|\, 2n\pi < y < 2(n+1)\pi\}$$

of the plane to the open subset

$$f(\mathcal{O}) = \{(x, y, z) \in \mathcal{S} \,|\, x \neq 1\}$$

of $\mathcal{S}$. If $a$ is not an odd multiple of $\pi$, a similar argument works with $\boldsymbol{\sigma}|_U$ replaced by $\boldsymbol{\sigma}|_{\tilde{U}}$ (we leave the details to the reader). Since $a$ cannot both be an even and an odd multiple of $\pi$, we have shown that for all points $\mathbf{p}$ of the plane, there is an open subset $\mathcal{O}$ of the plane containing $\mathbf{p}$ that $f$ maps diffeomorphically onto an open subset of $\mathcal{S}$.

## *EXERCISES*

4.3.1 If $\mathcal{S}$ is a smooth surface, define the notion of a *smooth function* $\mathcal{S} \to \mathbb{R}$. Show that, if $\mathcal{S}$ is a smooth surface, each component of the inclusion map $\mathcal{S} \to \mathbb{R}^3$ is a smooth function $\mathcal{S} \to \mathbb{R}$.

4.3.2 Let $\mathcal{S}$ be the half-cone $x^2+y^2 = z^2$, $z > 0$ (see Example 4.1.5). Define a map $f$ from the half-plane $\{(0, y, z) \mid y > 0\}$ to $\mathcal{S}$ by $f(0, y, z) = (y \cos z, y \sin z, y)$. Show that $f$ is a local diffeomorphism but not a diffeomorphism.

# 4.4 Tangents and derivatives

A natural way to study a surface $\mathcal{S}$ is via the (smooth) curves $\boldsymbol{\gamma}$ that lie in $\mathcal{S}$. This enables us to define the notion of a tangent vector to a surface.

### Definition 4.4.1

A *tangent vector* to a surface $\mathcal{S}$ at a point $\mathbf{p} \in \mathcal{S}$ is the tangent vector at $\mathbf{p}$ of a curve in $\mathcal{S}$ passing through $\mathbf{p}$. The *tangent space* $T_{\mathbf{p}}\mathcal{S}$ of $\mathcal{S}$ at $\mathbf{p}$ is the set of all tangent vectors to $\mathcal{S}$ at $\mathbf{p}$.

To understand the tangent space $T_{\mathbf{p}}\mathcal{S}$, choose a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ of $\mathcal{S}$ such that $\mathbf{p}$ is in the image of $\boldsymbol{\sigma}$, say $\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$. If a curve $\boldsymbol{\gamma}$ lies in $\mathcal{S}$ and passes through $\mathbf{p}$ when $t = t_0$, say, there are functions $u(t)$ and $v(t)$ such that

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)) \qquad (4.2)$$

for all values of $t$ close to $t_0$, and $u(t_0) = u_0, v(t_0) = v_0$. The functions $u$ and $v$ are necessarily smooth (this will be proved in Section 5.6); conversely, it is obvious that if $t \mapsto (u(t), v(t))$ is smooth, then Eq. 4.2 defines a curve lying in $\mathcal{S}$.

### Proposition 4.4.2

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a patch of a surface $\mathcal{S}$ containing a point $\mathbf{p} \in \mathcal{S}$, and let $(u, v)$ be coordinates in $U$. The tangent space to $\mathcal{S}$ at $\mathbf{p}$ is the vector subspace of $\mathbb{R}^3$ spanned by the vectors $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ (the derivatives are evaluated at the point $(u_0, v_0) \in U$ such that $\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$).

## Proof

Let $\boldsymbol{\gamma}$ be a smooth curve in $\mathcal{S}$, say

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)).$$

Denoting $d/dt$ by a dot, we have, by the chain rule,

$$\dot{\boldsymbol{\gamma}} = \boldsymbol{\sigma}_u \dot{u} + \boldsymbol{\sigma}_v \dot{v}.$$

Thus, $\dot{\boldsymbol{\gamma}}$ is a linear combination of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. (Perhaps it is worth spelling out what this last equation means: the vectors $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are smooth functions of $u$ and $v$; in these formulas one replaces $u$ and $v$ by the functions $u(t)$ and $v(t)$; then the right-hand side of the equation becomes a function of $t$ only, and the equation says that this function is equal to $d\boldsymbol{\gamma}/dt$.)

Conversely, any vector in the vector subspace of $\mathbb{R}^3$ spanned by $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ is of the form $\lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v$ for some scalars $\lambda$ and $\mu$. Define

$$\gamma(t) = \boldsymbol{\sigma}(u_0 + \lambda t, v_0 + \mu t).$$

Then, $\boldsymbol{\gamma}$ is a smooth curve in $\mathcal{S}$ and at $t = 0$, i.e., at the point $\mathbf{p} \in \mathcal{S}$, we have

$$\dot{\boldsymbol{\gamma}} = \lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v.$$

This shows that every vector in the span of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ is the tangent vector at $\mathbf{p}$ of some curve in $\mathcal{S}$. □

Since $\boldsymbol{\sigma}$ is assumed to be regular, $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent so the tangent space is two-dimensional, and will be called the *tangent plane* from now on. Note that Definition 4.4.1 shows that the tangent plane is independent of the choice of patch containing $\mathbf{p}$, even though this is not immediately obvious from Proposition 4.4.2 (see Exercise 4.4.2). Note also that the vectors $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ that form a basis of the tangent plane at some point $\boldsymbol{\sigma}(u_0, v_0)$ of the surface are the tangent vectors of the *parameter curves* on the surface, i.e., the curves $u \mapsto \boldsymbol{\sigma}(u, v_0)$ and $v \mapsto \boldsymbol{\sigma}(u_0, v)$ (we shall sometimes describe these curves as 'the parameter curves $u = u_0$ and $v = v_0$').

As a first application of the tangent plane to a smooth surface, we shall explain what is meant by the *derivative* of a smooth map between surfaces. Suppose then that $f : \mathcal{S} \to \tilde{\mathcal{S}}$ is such a map. The derivative of $f$ at a point $\mathbf{p} \in \mathcal{S}$ should measure how the point $f(\mathbf{p}) \in \tilde{\mathcal{S}}$ changes when $\mathbf{p}$ moves to a nearby point $\mathbf{q}$, say, of $\mathcal{S}$. If the points $\mathbf{p}$ and $\mathbf{q}$ are very close together, the straight line through them should be nearly tangent to $\mathcal{S}$ at $\mathbf{p}$. So we should expect that the derivative of $f$ at $\mathbf{p}$ associates to any tangent vector to $\mathcal{S}$ at $\mathbf{p}$ a tangent vector to $\tilde{\mathcal{S}}$ at $f(\mathbf{p})$, in other words, the derivative of $f$ at $\mathbf{p}$ should be a map $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$.

To give a precise definition of $D_{\mathbf{p}}f$, let $\mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ be a tangent vector to $\mathcal{S}$ at $\mathbf{p}$. By definition, $\mathbf{w}$ is the tangent vector at $\mathbf{p}$ of a curve $\boldsymbol{\gamma}$ in $\mathcal{S}$ passing through $\mathbf{p}$, say $\mathbf{w} = \dot{\boldsymbol{\gamma}}(t_0)$. Then, $\tilde{\boldsymbol{\gamma}} = f \circ \boldsymbol{\gamma}$ is a curve in $\tilde{\mathcal{S}}$ passing through $f(\mathbf{p})$ when $t = t_0$, so $\tilde{\mathbf{w}} = \dot{\tilde{\boldsymbol{\gamma}}}(t_0) \in T_{f(\mathbf{p})}\tilde{\mathcal{S}}$.

## Definition 4.4.3

With the above notation, the *derivative* $D_{\mathbf{p}}f$ of $f$ at the point $\mathbf{p} \in \mathcal{S}$ is the map $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ such that $D_{\mathbf{p}}f(\mathbf{w}) = \tilde{\mathbf{w}}$ for any tangent vector $\mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$.

The first thing we must do now is to show that this definition makes sense, i.e., that $D_{\mathbf{p}}f(\mathbf{w})$ depends only on $f$, $\mathbf{p}$, and $\mathbf{w}$: there are (infinitely) many curves $\boldsymbol{\gamma}$ with the correct tangent vector $\mathbf{w}$ at $\mathbf{p}$ and a priori $D_{\mathbf{p}}f(\mathbf{w})$ could depend on which curve is chosen.

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a surface patch of $\mathcal{S}$ containing $\mathbf{p}$, say $\mathbf{p} = \boldsymbol{\sigma}(u_0, v_0)$, and let $\alpha, \beta$ be the smooth functions on $U$ such that

$$f(\boldsymbol{\sigma}(u,v)) = \tilde{\boldsymbol{\sigma}}(\alpha(u,v), \beta(u,v)).$$

Let $\mathbf{w} = \lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v$ be the tangent vector at $\mathbf{p}$ of a curve $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$, where $u$ and $v$ are smooth functions such that $\dot{u}(t_0) = \lambda$, $\dot{v}(t_0) = \mu$. Since the corresponding curve on $\tilde{\mathcal{S}}$ is $\tilde{\boldsymbol{\gamma}}(t) = \tilde{\boldsymbol{\sigma}}(\tilde{u}(t), \tilde{v}(t))$, where $\tilde{u}(t) = \alpha(u(t), v(t))$ and $\tilde{v}(t) = \beta(u(t), v(t))$, we have

$$D_{\mathbf{p}}f(\mathbf{w}) = \dot{\tilde{u}}\tilde{\boldsymbol{\sigma}}_{\tilde{u}} + \dot{\tilde{v}}\tilde{\boldsymbol{\sigma}}_{\tilde{v}} = (\dot{u}\alpha_u + \dot{v}\alpha_v)\tilde{\boldsymbol{\sigma}}_{\tilde{u}} + (\dot{u}\beta_u + \dot{v}\beta_v)\tilde{\boldsymbol{\sigma}}_{\tilde{v}},$$

the derivatives of $u$ and $v$ being evaluated at $t_0$. Thus,

$$D_{\mathbf{p}}f(\mathbf{w}) = (\lambda\alpha_u + \mu\alpha_v)\tilde{\boldsymbol{\sigma}}_{\tilde{u}} + (\lambda\beta_u + \mu\beta_v)\tilde{\boldsymbol{\sigma}}_{\tilde{v}}. \tag{4.3}$$

The right-hand side depends only on $\mathbf{p}$, $f$, $\lambda$ and $\mu$, i.e., on $\mathbf{p}$, $f$ and $\mathbf{w}$, as we want.

Equation 4.3 also establishes the following proposition.

## Proposition 4.4.4

If $f : \mathcal{S} \to \tilde{\mathcal{S}}$ is a smooth map between surfaces and $\mathbf{p} \in \mathcal{S}$, the derivative $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is a linear map.

In fact, Eq. 4.3 shows that the matrix of the linear map $D_{\mathbf{p}}f$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}f$ and the basis $\{\tilde{\boldsymbol{\sigma}}_{\tilde{u}}, \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\}$ of $T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is the Jacobian matrix

$$\begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix}$$

of the smooth map $(u,v) \mapsto (\alpha(u,v), \beta(u,v))$.

### Proposition 4.4.5

(i) If $\mathcal{S}$ is a surface and $\mathbf{p} \in \mathcal{S}$, the derivative at $\mathbf{p}$ of the identity map $\mathcal{S} \to \mathcal{S}$ is the identity map $T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S}$.

(ii) If $\mathcal{S}_1, \mathcal{S}_2$ and $\mathcal{S}_3$ are surfaces and $f_1 : \mathcal{S}_1 \to \mathcal{S}_2$ and $f_2 : \mathcal{S}_2 \to \mathcal{S}_3$ are smooth maps, then for all $\mathbf{p} \in \mathcal{S}_1$,

$$D_{\mathbf{p}}(f_2 \circ f_1) = D_{f_1(\mathbf{p})} f_2 \circ D_{\mathbf{p}} f_1.$$

(iii) If $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a diffeomorphism, then for all $\mathbf{p} \in \mathcal{S}_1$ the linear map $D_{\mathbf{p}} f : T_{\mathbf{p}}\mathcal{S}_1 \to T_{f(\mathbf{p})}\mathcal{S}_2$ is invertible.

### Proof

Part (i) is obvious. For (ii), let $\mathbf{w} \in T_{\mathbf{p}}\mathcal{S}_1$ be the tangent vector at $\mathbf{p}$ of a curve $\boldsymbol{\gamma}_1$ on $\mathcal{S}_1$. Then, $\boldsymbol{\gamma}_2 = f_1 \circ \boldsymbol{\gamma}_1$ is a curve on $\mathcal{S}_2$ with tangent vector $D_{\mathbf{p}} f_1(\mathbf{w})$ at $f_1(\mathbf{p})$, so $\boldsymbol{\gamma}_3 = f_2 \circ \boldsymbol{\gamma}_2 = (f_2 \circ f_1) \circ \boldsymbol{\gamma}_1$ is a curve on $\mathcal{S}_3$ with tangent vector $D_{f_1(\mathbf{p})} f_2(D_{\mathbf{p}} f_1(\mathbf{w}))$ at $f_2(f_1(\mathbf{p}))$. But the tangent vector of $\boldsymbol{\gamma}_3$ at $\mathbf{p}$ is also $D_{\mathbf{p}}(f_2 \circ f_1)(\mathbf{w})$.

Finally, for (iii) let $g : \mathcal{S}_2 \to \mathcal{S}_1$ be the inverse map of $f$, so that $g \circ f$ and $f \circ g$ are the identity maps $\mathcal{S}_1 \to \mathcal{S}_1$ and $\mathcal{S}_2 \to \mathcal{S}_2$, respectively. Parts (i) and (ii) show that $D_{f(\mathbf{p})} g$ is the inverse of the linear map $D_{\mathbf{p}} f$. $\square$

We can now give a simple criterion for a smooth map to be a local diffeomorphism.

### Proposition 4.4.6

Let $\mathcal{S}$ and $\tilde{\mathcal{S}}$ be surfaces and let $f : \mathcal{S} \to \tilde{\mathcal{S}}$ be a smooth map. Then, $f$ is a local diffeomorphism if and only if, for all $\mathbf{p} \in \mathcal{S}$, the linear map $D_{\mathbf{p}} f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is invertible.

### Proof

Suppose first that $f$ is a local diffeomorphism and let $\mathbf{p} \in \mathcal{S}$. Then, there is an open subset $\mathcal{O}$ of $\mathcal{S}$ containing $\mathbf{p}$ such that $f(\mathcal{O})$ is an open subset of $\tilde{\mathcal{S}}$ and $f|_{\mathcal{O}} : \mathcal{O} \to f(\mathcal{O})$ is a diffeomorphism. By Proposition 4.4.5(iii), $D_{\mathbf{p}} f$ is invertible (note that it is obvious that $T_{\mathbf{p}}\mathcal{S} = T_{\mathbf{p}}\mathcal{O}$).

The proof of the 'if' part requires the inverse function theorem and will be given in Section 5.6. $\square$

## EXERCISES

4.4.1 Find the equation of the tangent plane of each of the following surface patches at the indicated points:

(i) $\boldsymbol{\sigma}(u,v) = (u, v, u^2 - v^2)$, $(1,1,0)$.

(ii) $\boldsymbol{\sigma}(r,\theta) = (r\cosh\theta, r\sinh\theta, r^2)$, $(1,0,1)$.

4.4.2 Show that, if $\boldsymbol{\sigma}(u,v)$ is a surface patch, the set of linear combinations of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ is unchanged when $\boldsymbol{\sigma}$ is reparametrized.

4.4.3 Let $\mathcal{S}$ be a surface, let $\mathbf{p} \in \mathcal{S}$ and let $F : \mathbb{R}^3 \to \mathbb{R}$ be a smooth function. Let $\boldsymbol{\nabla}_{\mathcal{S}} F$ be the perpendicular projection of the gradient $\boldsymbol{\nabla} F = (F_x, F_y, F_z)$ of $F$ onto $T_{\mathbf{p}}\mathcal{S}$. Show that, if $\boldsymbol{\gamma}$ is any curve on $\mathcal{S}$ passing through $\mathbf{p}$ when $t = t_0$, say,

$$(\boldsymbol{\nabla}_{\mathcal{S}} F)\cdot\dot{\boldsymbol{\gamma}}(t_0) = \left.\frac{d}{dt}\right|_{t=t_0} F(\boldsymbol{\gamma}(t)).$$

Deduce that $\boldsymbol{\nabla}_{\mathcal{S}} F = \mathbf{0}$ if the restriction of $F$ to $\mathcal{S}$ has a local maximum or a local minimum at $\mathbf{p}$.

4.4.4 Let $f : \mathcal{S}_1 \to \mathcal{S}_2$ be a local diffeomorphism and let $\boldsymbol{\gamma}$ be a regular curve on $\mathcal{S}_1$. Show that $f \circ \boldsymbol{\gamma}$ is a regular curve on $\mathcal{S}_2$.

# 4.5 Normals and orientability

Since the tangent plane $T_{\mathbf{p}}\mathcal{S}$ of a surface $\mathcal{S}$ at a point $\mathbf{p} \in \mathcal{S}$ passes through the origin of $\mathbb{R}^3$, it is completely determined by giving a unit vector perpendicular to it, called a *unit normal* to $\mathcal{S}$ at $\mathbf{p}$. There are, of course, two such vectors, but Proposition 4.4.2 shows that choosing a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ containing $\mathbf{p}$ leads to a definite choice, namely

$$\mathbf{N}_{\boldsymbol{\sigma}} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel} \tag{4.4}$$

(with the derivatives evaluated at the point of $U$ corresponding to $\mathbf{p}$), for this is clearly a unit vector perpendicular to every linear combination of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. This is called the *standard unit normal* of the surface patch $\boldsymbol{\sigma}$ at $\mathbf{p}$. Unlike the tangent plane, however, $\mathbf{N}_{\boldsymbol{\sigma}}$ is *not* quite independent of the choice of patch $\boldsymbol{\sigma}$ containing $\mathbf{p}$. In fact, if $\tilde{\boldsymbol{\sigma}} : \tilde{U} \to \mathbb{R}^3$ is another surface patch in the atlas of $\mathcal{S}$ containing $\mathbf{p}$, we showed in the proof of Proposition 4.2.7 that

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} = \det(J(\Phi))\, \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v,$$

where $J(\Phi)$ is the Jacobian matrix of the transition map $\Phi$ from $\boldsymbol{\sigma}$ to $\tilde{\boldsymbol{\sigma}}$. So the standard unit normal of $\tilde{\boldsymbol{\sigma}}$ is

$$\mathbf{N}_{\tilde{\boldsymbol{\sigma}}} = \frac{\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}}}{\| \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \|} = \pm \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \|} = \pm \mathbf{N}_{\boldsymbol{\sigma}},$$

where the sign is that of the determinant of $J(\Phi)$. This leads to the following definition.

## Definition 4.5.1

A surface $\mathcal{S}$ is *orientable* if there exists an atlas $\mathcal{A}$ for $\mathcal{S}$ with the property that, if $\Phi$ is the transition map between any two surface patches in $\mathcal{A}$, then $\det(J(\Phi)) > 0$ where $\Phi$ is defined.

The preceding discussion gives the following proposition.

## Proposition 4.5.2

Let $\mathcal{S}$ be an orientable surface equipped with an atlas $\mathcal{A}$ as in Definition 4.5.1. Then, there is a smooth choice of unit normal at any point of $\mathcal{S}$: take the standard unit normal of any surface patch in $\mathcal{A}$.

An *oriented surface* is a surface $\mathcal{S}$ together with a smooth choice of unit normal $\mathbf{N}$ at each point, i.e., a smooth map $\mathbf{N} : \mathcal{S} \to \mathbb{R}^3$ (meaning that each of the three components of $\mathbf{N}$ is a smooth function $\mathcal{S} \to \mathbb{R}$) such that, for all $\mathbf{p} \in \mathcal{S}$, $\mathbf{N}(\mathbf{p})$ is a unit vector perpendicular to $T_{\mathbf{p}}\mathcal{S}$. Any oriented surface is orientable! To see this, start with the maximal atlas of $\mathcal{S}$ and retain a patch $\boldsymbol{\sigma}(u, v)$ if $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$ is a positive multiple of $\mathbf{N}$ at all points in the image of $\boldsymbol{\sigma}$, otherwise discard it. The patches that remain form an atlas $\mathcal{A}$ satisfying the condition in Definition 4.5.1. We leave the details of this to the reader (the argument is similar to that used in the next example). From now onwards, *whenever we are dealing with an oriented surface $\mathcal{S}$, we shall only use surface patches for $\mathcal{S}$ whose standard unit normal is the same as the chosen normal of $\mathcal{S}$.*

Most of the surfaces we shall discuss are orientable. Here is one that is not.

## Example 4.5.3

The *Möbius band* is the surface obtained by rotating a straight line segment $l$ around its midpoint $\mathbf{p}$ at the same time as $\mathbf{p}$ moves around a circle $\mathcal{C}$, in such a way that as $\mathbf{p}$ moves once around $\mathcal{C}$, $l$ makes a half-turn about $\mathbf{p}$. If we take $\mathcal{C}$

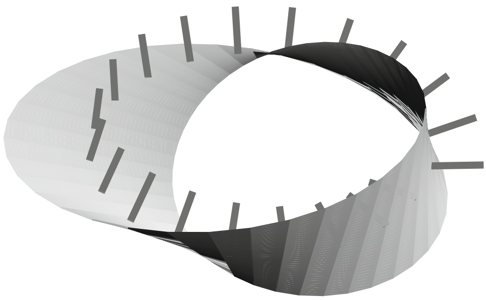
to be the circle $x^2 + y^2 = 1$ in the $xy$-plane, and $l$ to be a segment of length 1 that is initially parallel to the $z$-axis with its midpoint $\mathbf{p}$ at $(1, 0, 0)$, then after $\mathbf{p}$ has rotated by an angle $\theta$ around the $z$-axis, $l$ should have rotated by $\theta/2$ around $\mathbf{p}$ in the plane containing $\mathbf{p}$ and the $z$-axis. The point of $l$ initially at $(1, 0, t)$ is then at the point

$$\boldsymbol{\sigma}(t, \theta) = \left(\left(1 - t\sin\frac{\theta}{2}\right)\cos\theta, \left(1 - t\sin\frac{\theta}{2}\right)\sin\theta, \, t\cos\frac{\theta}{2}\right).$$

We take the domain of definition of $\boldsymbol{\sigma}$ to be

$$U = \{(t, \theta) \in \mathbb{R}^2 \mid -1/2 < t < 1/2, \; 0 < \theta < 2\pi\}.$$

We can define a second patch $\tilde{\boldsymbol{\sigma}}$ by the same formula as $\boldsymbol{\sigma}$ but with domain of definition $\tilde{U} = \{(t, \theta) \in \mathbb{R}^2 \mid -1/2 < t < 1/2, \; -\pi < \theta < \pi\}$. It can be checked that these two patches form an atlas for the Möbius band consisting of regular surface patches, making the Möbius band into a smooth surface $\mathcal{S}$ (see Exercise 4.5.1).



We compute the standard unit normal $\mathbf{N}_{\boldsymbol{\sigma}}$ at points on the median circle (where $t = 0$). At such points, we have

$$\boldsymbol{\sigma}_t = \left(-\sin\frac{\theta}{2}\cos\theta, -\sin\frac{\theta}{2}\sin\theta, \cos\frac{\theta}{2}\right), \quad \boldsymbol{\sigma}_\theta = (-\sin\theta, \cos\theta, 0),$$

so

$$\boldsymbol{\sigma}_t \times \boldsymbol{\sigma}_\theta = \left(-\cos\theta\cos\frac{\theta}{2}, -\sin\theta\cos\frac{\theta}{2}, -\sin\frac{\theta}{2}\right).$$

This is a unit vector, so it is equal to $\mathbf{N}_{\boldsymbol{\sigma}}$.

If the Möbius band was orientable, there would be a well-defined unit normal $\mathbf{N}$ defined at every point of $\mathcal{S}$ and varying smoothly over $\mathcal{S}$. At a point $\boldsymbol{\sigma}(0, \theta)$ on the median circle, we would have

$$\mathbf{N} = \lambda(\theta)\mathbf{N}_{\boldsymbol{\sigma}},$$

where $\lambda : (0, 2\pi) \to \mathbb{R}$ is smooth and $\lambda(\theta) = \pm 1$ for all $\theta$. It follows that either $\lambda(\theta) = +1$ for all $\theta \in (0, 2\pi)$, or $\lambda(\theta) = -1$ for all $\theta \in (0, 2\pi)$. Replacing $\mathbf{N}$ by

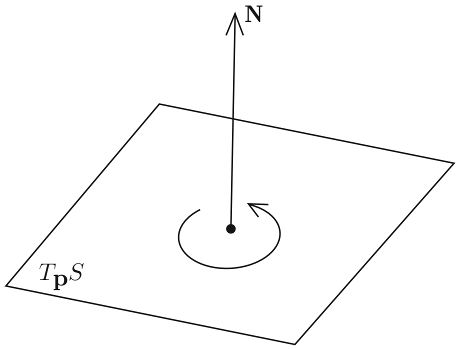
$-\mathbf{N}$ if necessary, we can assume that $\lambda = 1$. At the point $\boldsymbol{\sigma}(0,0) = \boldsymbol{\sigma}(0,2\pi)$, we would have (since $\mathbf{N}$ is smooth)

$$\mathbf{N} = \lim_{\theta \downarrow 0} \mathbf{N}_{\boldsymbol{\sigma}} = (-1,0,0) \text{ and also } \mathbf{N} = \lim_{\theta \uparrow 2\pi} \mathbf{N}_{\boldsymbol{\sigma}} = (1,0,0).$$

This contradiction shows that the Möbius band is not orientable.

If a surface $\mathcal{S}$ is oriented, it is possible to give a sign to the angle between two tangent vectors at a point of $\mathcal{S}$. This will be important on a number of occasions later in this book.

Let $\mathbf{p} \in \mathcal{S}$ and let $\mathbf{N}$ be the chosen unit normal at $\mathbf{p}$. A rotation in the tangent plane $T_{\mathbf{p}}\mathcal{S}$ is said to be in the *positive sense*, or *anticlockwise*, if rotation in this sense of a right-handed screw held perpendicular to $T_{\mathbf{p}}\mathcal{S}$ would cause it to advance in the direction of $\mathbf{N}$. Put another way, the choice of $\mathbf{N}$ enables us to distinguish the two 'sides' of $T_{\mathbf{p}}\mathcal{S}$: the 'positive' side is the half-space into which $\mathbf{N}$ points. Then, if we view $T_{\mathbf{p}}\mathcal{S}$ from a point on the positive side, a positive rotation in $T_{\mathbf{p}}\mathcal{S}$ would be seen as anticlockwise in the usual sense.




If $\mathbf{v}$ and $\mathbf{w}$ are non-zero vectors in $T_{\mathbf{p}}\mathcal{S}$, the *oriented angle* (which we shall sometimes just call the angle) between $\mathbf{v}$ and $\mathbf{w}$ is the angle through which $\mathbf{v}$ must be rotated in the positive sense in order for the resulting vector to be a positive scalar multiple of $\mathbf{w}$. We shall denote this angle by $\widehat{\mathbf{v}\mathbf{w}}$. Note that

$$\widehat{\mathbf{w}\mathbf{v}} = -\widehat{\mathbf{v}\mathbf{w}},$$

and that the sign of $\widehat{\mathbf{v}\mathbf{w}}$ will change if we change the choice of unit normal to $T_{\mathbf{p}}\mathcal{S}$. Note also that $\widehat{\mathbf{v}\mathbf{w}}$ is determined only up to the addition of an integer multiple of $2\pi$.

## Example 4.5.4

It is clear that at a point $\mathbf{p} \in S^2$, the tangent plane is perpendicular to $\mathbf{p}$. Since $\mathbf{p}$ is a unit vector, the two unit normals at $\mathbf{p}$ are $\mathbf{p}$, the 'outward' normal, and $-\mathbf{p}$, the 'inward' normal. There are two smooth choices of unit normal on $S^2$: either always inwards or always outwards.

For example, the reader should check that, if we take the inward normal and if $\mathbf{p}$ is the point $(1, 0, 0)$ on the equator, the oriented angle between the tangent vectors $\mathbf{v} = (0, 1, 0)$ and $\mathbf{w} = (0, 0, 1)$ is $\widehat{\mathbf{vw}} = -\pi/2$ (or $3\pi/2, 7\pi/2$, etc.). If we used the outward pointing normal instead, this oriented angle would change sign.

## *EXERCISES*

4.5.1 Calculate the transition map $\Phi$ between the two surface patches for the Möbius band in Example 4.5.3. Show that it is defined on the union of two disjoint rectangles in $\mathbb{R}^2$, and that the determinant of the Jacobian matrix of $\Phi$ is equal to $+1$ on one of the rectangles and to $-1$ on the other.

4.5.2 Suppose that two smooth surfaces $\mathcal{S}$ and $\tilde{\mathcal{S}}$ are diffeomorphic and that $\mathcal{S}$ is orientable. Prove that $\tilde{\mathcal{S}}$ is orientable.

# 5
# Examples of surfaces

In this chapter we describe some of the simplest classes of surfaces. Others will be introduced later in the book.

## 5.1 Level surfaces

As we have already seen (Examples 4.1.3–5 and Exercise 4.1.3), surfaces are often given to us as *level surfaces*

$$\{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = 0\},$$

where $f$ is a smooth function. In those examples, we constructed atlases by fairly ad hoc methods. The following result gives general conditions under which a level surface is a smooth surface. In fact, it deals with a slightly more general situation in which different regions of a surface may be defined by different functions.

### Theorem 5.1.1

Let $\mathcal{S}$ be a subset of $\mathbb{R}^3$ with the following property: for each point $\mathbf{p} \in \mathcal{S}$, there is an open subset $W$ of $\mathbb{R}^3$ containing $\mathbf{p}$ and a smooth function $f : W \to \mathbb{R}$ such that

(i) $\mathcal{S} \cap W = \{(x, y, z) \in W \mid f(x, y, z) = 0\}$;

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_5,

(ii) The gradient $\nabla f = (f_x, f_y, f_z)$ of $f$ does not vanish at $\mathbf{p}$.

Then, $\mathcal{S}$ is a smooth surface.

We postpone the proof to Section 5.6.

### Example 5.1.2

For the unit sphere $S^2$, we can take $W = \mathbb{R}^3$ and use the single function $f(x,y,z) = x^2+y^2+z^2-1$. Then, $\nabla f = (2x, 2y, 2z)$ so $\| \nabla f \| = 2$ at all points of $S^2$. In particular, $\nabla f$ is non-zero everywhere on $S^2$. Hence, Theorem 5.1.1 tells us that $S^2$ is a smooth surface.

### Example 5.1.3

For the circular cone of Example 4.1.5, $f(x,y,z) = x^2 + y^2 - z^2$. Hence, $\nabla f = (2x, 2y, -2z)$, and this vanishes only at the vertex $(0,0,0)$. Theorem 5.1.1 applies with $W = \{(x,y,z) \in \mathbb{R}^3 \mid z \neq 0\}$, so the circular cone with the vertex removed is a smooth surface, as we have already seen.

A large class of level surfaces is studied in the next section.

## *EXERCISES*

5.1.1 Show that the following are smooth surfaces:

(i) $x^2 + y^2 + z^4 = 1$.

(ii) $(x^2 + y^2 + z^2 + a^2 - b^2)^2 = 4a^2(x^2 + y^2)$, where $a > b > 0$ are constants.

Show that the surface in (ii) is, in fact, the torus of Exercise 4.2.5.

5.1.2 Consider the surface $\mathcal{S}$ defined by $f(x,y,z) = 0$, where $f$ is a smooth function such that $\nabla f$ does not vanish at any point of $\mathcal{S}$. Show that $\nabla f$ is perpendicular to the tangent plane at every point of $\mathcal{S}$, and deduce that $\mathcal{S}$ is orientable. Suppose now that $F : \mathbb{R}^3 \to \mathbb{R}$ is a smooth function. Show that, if the restriction of $F$ to $\mathcal{S}$ has a local maximum or a local minimum at $\mathbf{p}$ then, at $\mathbf{p}$, $\nabla F = \lambda \nabla f$ for some scalar $\lambda$. (This is called *Lagrange's Method of Undetermined Multipliers*.)

5.1.3 Show that the smallest value of $x^2 + y^2 + z^2$ subject to the condition $xyz = 1$ is 3, and that the points $(x,y,z)$ that give this minimum value lie at the vertices of a regular tetrahedron in $\mathbb{R}^3$.

## 5.2 Quadric surfaces

The simplest level surfaces, namely planes, have Cartesian equations of the form $ax + by + cz = d$, where $a, b, c, d$ are constants. From this point of view, the next simplest surfaces should be those whose Cartesian equations are given by quadratic expressions in $x, y$ and $z$.

In this section, we identify any vector $\mathbf{v} = (x, y, z) \in \mathbb{R}^3$ with the column matrix $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, which we also denote by $\mathbf{v}$. Note that, if $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, $\mathbf{v}^t\mathbf{w}$ is a $1 \times 1$ matrix, i.e., a number, namely the dot product $\mathbf{v} \cdot \mathbf{w}$.

### Definition 5.2.1

A *quadric* is the subset of $\mathbb{R}^3$ defined by an equation of the form

$$\mathbf{v}^t A\mathbf{v} + \mathbf{b}^t\mathbf{v} + c = 0,$$

where $\mathbf{v} = (x, y, z)$, $A$ is a constant symmetric $3\times3$ matrix, $\mathbf{b} \in \mathbb{R}^3$ is a constant vector, and $c \in \mathbb{R}$ is a constant scalar.

To see this more explicitly, let

$$A = \begin{pmatrix} a_1 & a_4 & a_6 \\ a_4 & a_2 & a_5 \\ a_6 & a_5 & a_3 \end{pmatrix}, \quad \mathbf{b} = (b_1, b_2, b_3).$$

Then, the equation of the quadric is

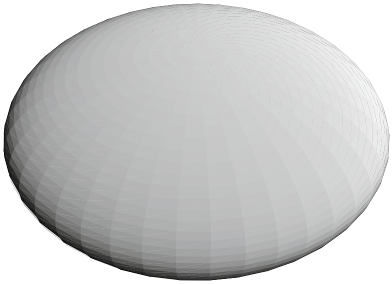
$$a_1x^2 + a_2y^2 + a_3z^2 + 2a_4xy + 2a_5yz + 2a_6xz + b_1x + b_2y + b_3z + c = 0. \quad (5.1)$$

A quadric is not necessarily a surface. For example, the quadric with equation $x^2+y^2+z^2 = 0$ is a single point, and that with equation $x^2+y^2 = 0$ is a straight line. A more interesting example is the quadric $xy = 0$, which is the union of the two intersecting planes $x = 0$ and $y = 0$, and is also not a surface. (Intuitively, it has a ‘corner’ along the line of intersection of the planes.) The following proposition shows that to understand all quadrics it is sufficient to consider quadrics whose equations take on a particularly simple form.

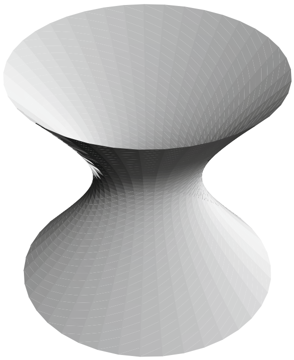
### Theorem 5.2.2

By applying a direct isometry of $\mathbb{R}^3$, every non-empty quadric (5.1) in which the coefficients are not all zero can be transformed into one whose Cartesian equation is one of the following:

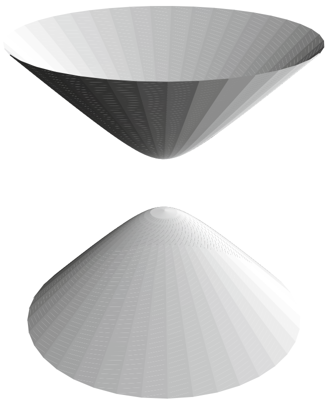
(i) Ellipsoid: $\frac{x^2}{p^2} + \frac{y^2}{q^2} + \frac{z^2}{r^2} = 1$.



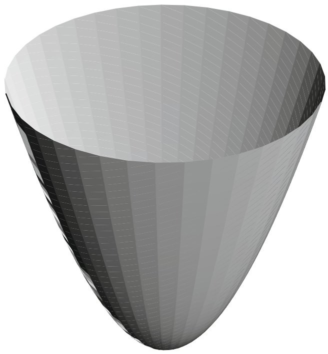
(ii) Hyperboloid of one sheet: $\frac{x^2}{p^2} + \frac{y^2}{q^2} - \frac{z^2}{r^2} = 1$.



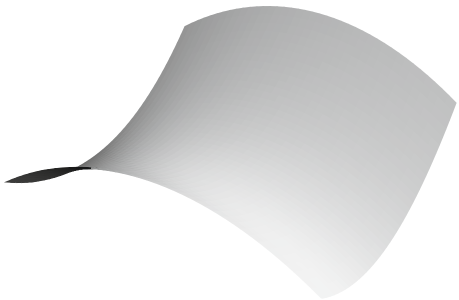
(iii) Hyperboloid of two sheets: $\frac{z^2}{r^2} - \frac{x^2}{p^2} - \frac{y^2}{q^2} = 1$.



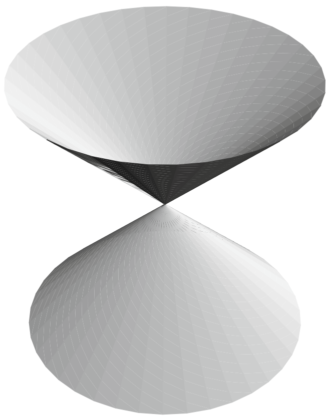
(iv) Elliptic paraboloid: $\frac{x^2}{p^2} + \frac{y^2}{q^2} = z$.



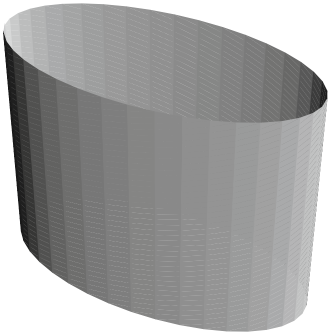
(v) Hyperbolic paraboloid: $\frac{x^2}{p^2} - \frac{y^2}{q^2} = z$.



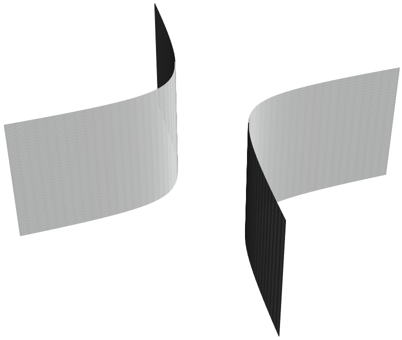
(vi) Quadric cone: $\frac{x^2}{p^2} + \frac{y^2}{q^2} - \frac{z^2}{r^2} = 0$.



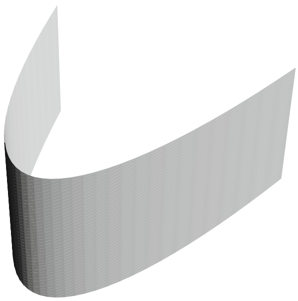
(vii) Elliptic cylinder: $\frac{x^2}{p^2} + \frac{y^2}{q^2} = 1$.



(viii) Hyperbolic cylinder: $\frac{x^2}{p^2} - \frac{y^2}{q^2} = 1$.



(ix) Parabolic cylinder: $\frac{x^2}{p^2} = y$.



(x) Plane: $x = 0$.

(xi) Two parallel planes: $x^2 = p^2$.

(xii) Two intersecting planes: $\frac{x^2}{p^2} - \frac{y^2}{q^2} = 0$.

(xiii) Straight line: $\frac{x^2}{p^2} + \frac{y^2}{q^2} = 0$.

(xiv) Single point: $\frac{x^2}{p^2} + \frac{y^2}{q^2} + \frac{z^2}{r^2} = 0$.

In each case, $p$, $q$ and $r$ are non-zero constants.

## Proof

By Theorem A.0.4, there is an orthogonal matrix $P$ such that $PAP^t$ is a *diagonal* matrix, say

$$A' = \begin{pmatrix} a_1' & 0 & 0 \\ 0 & a_2' & 0 \\ 0 & 0 & a_3' \end{pmatrix}$$

($P^t$ denotes the transpose of $P$ and $I$ denotes the identity matrix). Then, $\det(P) = \pm 1$, and by replacing $P$ by $-P$ if necessary, we can assume that $\det(P) = 1$. The diagonal entries of $A'$ are the eigenvalues of $A$, and the rows of $P$ are the corresponding eigenvectors. Define $\mathbf{v}' = (x', y', z')$, $\mathbf{b}' = (b_1', b_2', b_3')$, where

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = P \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \begin{pmatrix} b_1' \\ b_2' \\ b_3' \end{pmatrix} = P \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

Noting that

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = P^t \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}, \quad \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = P^t \begin{pmatrix} b_1' \\ b_2' \\ b_3' \end{pmatrix},$$

the quadric in Definition 5.2.1 becomes

$$(x'\, y'\, z')A'(x'\, y'\, z')^t + (b_1'\, b_2'\, b_3')(x'\, y'\, z')^t + c = 0,$$

$$\text{i.e.,} \qquad a_1'x'^2 + a_2'y'^2 + a_3'z'^2 + b_1'x' + b_2'y' + b_3'z' + c = 0.$$

This new quadric is obtained from the given one by applying the direct isometry $\mathbf{v} \mapsto P\mathbf{v}$ (see Appendix 1), so we might as well consider the quadric in (5.1), but assume that $a_4 = a_5 = a_6 = 0$, i.e.,

$$a_1x^2 + a_2y^2 + a_3z^2 + b_1x + b_2y + b_3z + c = 0. \tag{5.2}$$

Suppose now that, in Eq. 5.2, $a_1 \neq 0$. If we define $x' = x + b_1/2a_1$, corresponding to a translation of $\mathbb{R}^3$, the equation becomes

$$a_1x'^2 + a_2y^2 + a_3z^2 + b_2y + b_3z + c' = 0,$$

where $c'$ is a constant. In other words, if $a_1 \neq 0$, we can assume that $b_1 = 0$, and similarly for $a_2$ and $a_3$, of course.

If $a_1, a_2$ and $a_3$ in Eq. 5.2 are all non-zero, we may therefore reduce to the form

$$a_1x^2 + a_2y^2 + a_3z^2 + c = 0.$$

If $c \neq 0$, we get cases (i), (ii) and (iii), depending on the signs of $a_1, a_2, a_3$ and $c$, and if $c = 0$ we get cases (vi) and (xiv).

If exactly one of $a_1, a_2$ and $a_3$ is zero, say $a_3 = 0$, we are reduced to the form

$$a_1x^2 + a_2y^2 + b_3z + c = 0. \tag{5.3}$$

If $b_3 \neq 0$, we may define $z' = z + c/b_3$. Thus, by a translation (and by dividing by $b_3$), we are reduced to the case

$$a_1x^2 + a_2y^2 + z = 0.$$

This gives cases (iv) and (v).

If $b_3 = 0$ in Eq. 5.3, we have

$$a_1x^2 + a_2y^2 + c = 0.$$

If $c = 0$ we get cases (xii) and (xiii). If $c \neq 0$, dividing through by it leads to cases (vii) and (viii).

Suppose now that $a_2 = a_3 = 0$, but $a_1 \neq 0$. Then we have

$$a_1x^2 + b_2y + b_3z + c = 0. \tag{5.4}$$

If $b_2$ and $b_3$ are not both zero, by applying a rotation in the $xz$-plane that takes the $y$-axis to a line parallel to the vector $(b_2, b_3)$, we can arrive at the situation $b_2 \neq 0, b_3 = 0$, and then by a translation along the $y$-axis we can arrange that $c = 0$. This leads to the equation

$$a_1x^2 + y = 0,$$

which gives case (ix). If $b_2 = b_3 = 0$ in Eq. 5.4, then $c = 0$ gives case (x) and $c \neq 0$ gives case (xi).

Finally, if $a_1 = a_2 = a_3 = 0$, (5.6) is the equation of a plane, so after applying a Suitable composite of rotations and translations we are in case (x) again. □

## Corollary 5.2.3

Every non-empty quadric of types (i)–(x) is a surface (for type (vi) one must remove the vertex of the cone).

## Proof

This is easily verified using Exercise 4.2.8, Theorem 5.1.1 and the special form of the equations of quadrics in Theorem 5.2.2. □

## Example 5.2.4

Consider the quadric

$$x^2 + 2y^2 + 6x - 4y + 3z = 7.$$

Setting $x' = x + 3, y' = y - 1$ (a translation), we get

$$x'^2 + 2y'^2 + 3z = 18.$$

Setting $z' = z - 6$ (another translation) gives

$$x'^2 + 2y'^2 + 3z' = 0.$$

Finally, setting $x'' = x', y'' = -y', z'' = -z'$ (a rotation by $\pi$ about the $x$-axis) gives

$$\frac{1}{3}x''^2 + \frac{2}{3}y''^2 = z'',$$

which is an elliptic paraboloid. It can be parametrized by setting $x'' = u, y'' = v, z'' = \frac{1}{3}u^2 + \frac{2}{3}v^2$. This corresponds to $x = u - 3, y = 1 - v, z = 6 - \frac{1}{3}u^2 - \frac{2}{3}v^2$, and shows that the given quadric is a smooth surface with an atlas consisting of the single surface patch

$$\boldsymbol{\sigma}(u, v) = \left(u - 3, 1 - v, 6 - \frac{1}{3}u^2 - \frac{2}{3}v^2\right).$$

## *EXERCISES*

5.2.1 Write down parametrizations of each of the quadrics in parts (i)–(xi) of Theorem 5.2.2 (in case (vi) one must remove the origin).

5.2.2 Show that the quadric

$$x^2 + y^2 - 2z^2 - \frac{2}{3}xy + 4z = c$$

is a hyperboloid of one sheet if $c > 2$, and a hyperboloid of two sheets if $c < 2$. What if $c = 2$? (This exercise requires a knowledge of eigenvalues and eigenvectors.)

5.2.3 Show that, if a quadric contains three points on a straight line, it contains the whole line. Deduce that, if $L_1, L_2$ and $L_3$ are non-intersecting straight lines in $\mathbb{R}^3$, there is a quadric containing all three lines.

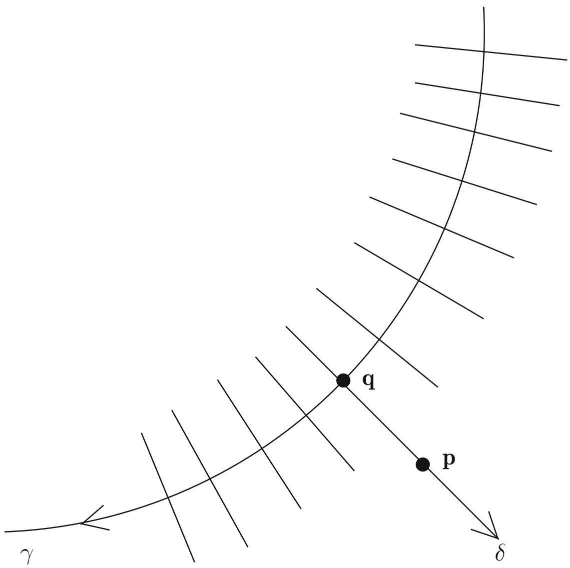
5.2.4 Use the preceding exercise to show that any doubly ruled surface is (part of) a quadric surface. (A surface is doubly ruled if it is the union of each of two families of straight lines such that no two lines of the same family intersect, but every line of the first family intersects every line of the second family, with at most a finite number of exceptions.) Which quadric surfaces are doubly ruled?

## 5.3 Ruled surfaces and surfaces of revolution

Level surfaces have an 'algebraic' origin, in that they arise from a function $f(x, y, z)$. On the other hand, the two classes of surfaces considered in this section arise from geometric constructions.

### Example 5.3.1

A *ruled surface* is a surface that is a union of straight lines, called the *rulings* (or sometimes the *generators*) of the surface.



Suppose that $\mathcal{C}$ is a curve in $\mathbb{R}^3$ that meets each of these lines. Any point $\mathbf{p}$ of the surface lies on one of the given straight lines, which intersects $\mathcal{C}$ at $\mathbf{q}$, say. If $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^3$ is a parametrization of $\mathcal{C}$ with $\boldsymbol{\gamma}(u) = \mathbf{q}$, and if $\boldsymbol{\delta}(u)$ is a non-zero vector in the direction of the line passing through $\boldsymbol{\gamma}(u)$, then

$$\mathbf{p} = \boldsymbol{\gamma}(u) + v\boldsymbol{\delta}(u),$$

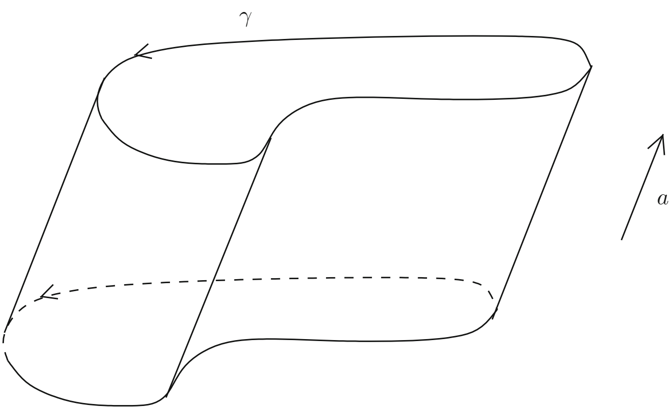
for some scalar $v$. Denoting the right-hand side by $\boldsymbol{\sigma}(u, v)$, it is clear that $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a smooth map, where $U = \{(u, v) \in \mathbb{R}^2 \mid \alpha < u < \beta\}$. Moreover, denoting $d/du$ by a dot,

$$\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}, \quad \boldsymbol{\sigma}_v = \boldsymbol{\delta}.$$

Thus, $\boldsymbol{\sigma}$ is regular if $\dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}$ are linearly independent. This will be true, for example, if $\dot{\boldsymbol{\gamma}}$ and $\boldsymbol{\delta}$ are linearly independent and $v$ is sufficiently small. Thus, to get a surface, the curve $\mathcal{C}$ must never be tangent to the rulings.

An important special case is that in which the rulings are all parallel to each other; the ruled surface $\mathcal{S}$ is then called a *generalized cylinder*. In the above notation, we can take $\boldsymbol{\delta}$ to be a constant unit vector, say $\mathbf{a}$, parallel to the rulings, and the parametrization becomes

$$\boldsymbol{\sigma}(u, v) = \boldsymbol{\gamma}(u) + v\mathbf{a}.$$




Since

$$\boldsymbol{\sigma}(u, v) = \boldsymbol{\sigma}(u', v') \iff \boldsymbol{\gamma}(u) - \boldsymbol{\gamma}(u') = (v' - v)\mathbf{a},$$

for $\boldsymbol{\sigma}$ to be a injective (and hence a surface patch), no straight line parallel to $\mathbf{a}$ should meet $\boldsymbol{\gamma}$ in more than one point. Finally, $\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}}$, $\boldsymbol{\sigma}_v = \mathbf{a}$, so $\boldsymbol{\sigma}$ is regular if and only if $\boldsymbol{\gamma}$ is never tangent to the rulings.

The parametrization is simplest when $\boldsymbol{\gamma}$ lies in a plane perpendicular to $\mathbf{a}$ (in fact, this can always be achieved by replacing $\boldsymbol{\gamma}$ by its perpendicular projection onto such a plane – see Exercise 5.3.3). The regularity condition is then clearly satisfied provided $\dot{\boldsymbol{\gamma}}$ is never zero, i.e., provided $\boldsymbol{\gamma}$ is regular. We might as well take the plane to be the $xy$-plane and $\mathbf{a} = (0, 0, 1)$ to be parallel to the $z$-axis. Then, $\boldsymbol{\gamma}(u) = (f(u), g(u), 0)$ for some smooth functions $f$ and $g$, and the parametrization becomes

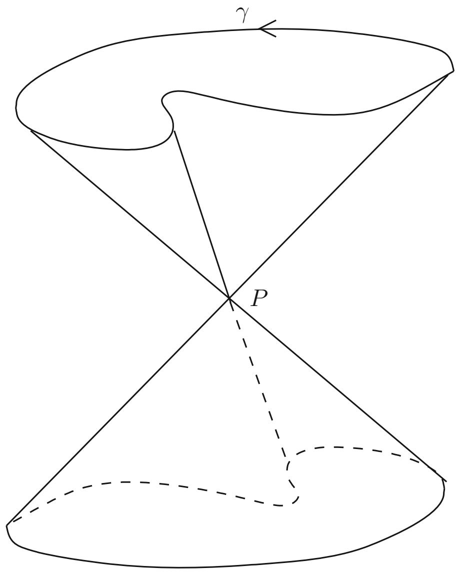
$$\boldsymbol{\sigma}(u, v) = (f(u), g(u), v).$$

For example, starting with a circle, we get a circular cylinder. Taking the circle to have centre the origin, radius 1 and to lie in the $xy$-plane, it can be parametrized by

$$\boldsymbol{\gamma}(u) = (\cos u, \sin u, 0),$$

defined for $0 < u < 2\pi$ and $-\pi < u < \pi$, say. This gives the atlas for the unit cylinder found in Example 4.1.3.

The second special case we shall consider is that in which the rulings all pass through a certain fixed point, say $\mathbf{v}$; then $\mathcal{S}$ is called a *generalized cone with vertex* $\mathbf{v}$.




We can take $\boldsymbol{\delta}(u) = \boldsymbol{\gamma}(u) - \mathbf{v}$, giving

$$\boldsymbol{\sigma}(u, v) = (1 + v)\boldsymbol{\gamma}(u) - v\mathbf{v}.$$

Now,

$$\boldsymbol{\sigma}(u, v) = \boldsymbol{\sigma}(u', v') \iff (1 + v)\boldsymbol{\gamma}(u) - (1 + v')\boldsymbol{\gamma}(u') + (v' - v)\mathbf{v} = \mathbf{0};$$

since $(1+v)-(1+v')+(v'-v) = 0$, the equation on the right-hand side means that the points $\mathbf{v}$, $\boldsymbol{\gamma}(u)$ and $\boldsymbol{\gamma}(u')$ are collinear. So, for $\boldsymbol{\sigma}$ to be a surface patch, no straight line passing through $\mathbf{v}$ should pass through more than one point of $\boldsymbol{\gamma}$ (in particular, $\boldsymbol{\gamma}$ should not pass through $\mathbf{v}$). Finally, we have $\boldsymbol{\sigma}_u = (1+v)\dot{\boldsymbol{\gamma}}$, $\boldsymbol{\sigma}_v = \boldsymbol{\gamma} - \mathbf{v}$, so $\boldsymbol{\sigma}$ is regular provided $v \neq -1$, i.e., the vertex of the cone is omitted (cf. Example 4.1.5), and none of the straight lines forming the cone is tangent to $\boldsymbol{\gamma}$.

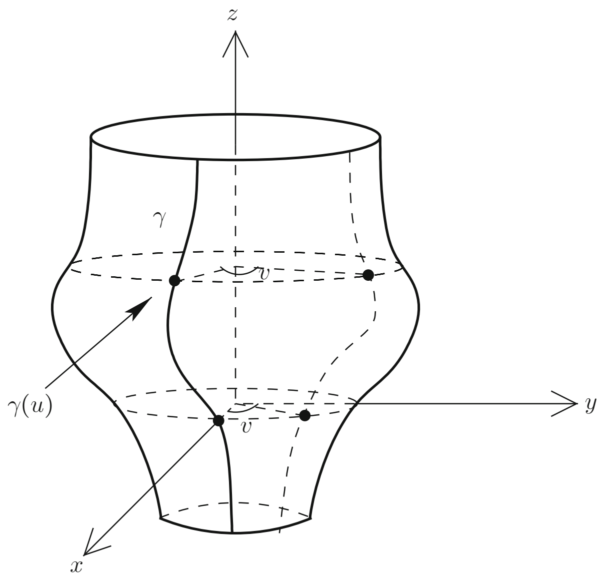
The parametrization is simplest when $\boldsymbol{\gamma}$ lies in a plane. If this plane contains $\mathbf{v}$, the cone is simply part of that plane. Otherwise, we can take $\mathbf{v}$ to be the origin and the plane to be $z = 1$. Then, $\boldsymbol{\gamma}(u) = (f(u), g(u), 1)$ for some smooth functions $f$ and $g$, and the parametrization takes the form

$$\boldsymbol{\sigma}(u, v) = v(f(u), g(u), 1),$$

after making the reparametrization $v \mapsto v - 1$.

## Example 5.3.2

A *surface of revolution* is the surface obtained by rotating a plane curve, called the *profile curve*, around a straight line in the plane. The circles obtained by rotating a fixed point on the profile curve around the axis of rotation are called the *parallels* of the surface, and the curves on the surface obtained by rotating the profile curve through a fixed angle are called its *meridians*. (This agrees with the use of these terms in geography, if we think of the earth as the surface obtained by rotating a circle passing through the poles about the polar axis and we take $u$ and $v$ to be latitude and longitude, respectively.)



Let us take the axis of rotation to be the $z$-axis and the plane to be the $xz$-plane. Any point $\mathbf{p}$ of the surface is obtained by rotating some point $\mathbf{q}$ of the profile curve through an angle $v$ (say) around the $z$-axis. If

$$\gamma(u) = (f(u), 0, g(u))$$

is a parametrization of the profile curve containing $\mathbf{q}$, $\mathbf{p}$ is of the form

$$\boldsymbol{\sigma}(u, v) = (f(u)\cos v, f(u)\sin v, g(u)).$$

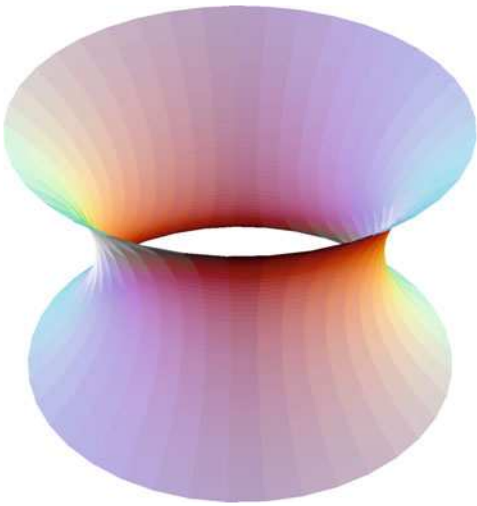
To check regularity, we compute (with a dot denoting $d/du$):

$$\begin{aligned} \boldsymbol{\sigma}_u &= (\dot f\cos v, \dot f\sin v, \dot g), \quad \boldsymbol{\sigma}_v = (-f\sin v, f\cos v, 0), \\ \therefore \quad \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= (f\dot g\cos v, -f\dot g\sin v, f\dot f), \\ \therefore \quad \|\, \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \,\|^2 &= f^2(\dot f^2 + \dot g^2). \end{aligned}$$

Thus, $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$ will be non-vanishing if $f(u)$ is never zero, i.e., if $\boldsymbol{\gamma}$ does not intersect the $z$-axis, and if $\dot f$ and $\dot g$ are never zero simultaneously, i.e., if $\boldsymbol{\gamma}$ is regular. In this case, we might as well assume that $f(u) > 0$, so that $f(u)$ is the distance of $\boldsymbol{\sigma}(u, v)$ from the axis of rotation. Then, $\boldsymbol{\sigma}$ is injective provided that $\boldsymbol{\gamma}$ does not self-intersect and the angle of rotation $v$ is restricted to lie in an open interval of length $\leq 2\pi$. Under these conditions, surface patches of the form $\boldsymbol{\sigma}$ give the surface of revolution the structure of a surface.

## *EXERCISES*

5.3.1 The surface obtained by rotating the curve $x = \cosh z$ in the $xz$-plane around the $z$-axis is called a *catenoid*. Describe an atlas for this surface.



5.3.2 Show that

$$\boldsymbol{\sigma}(u,v) = (\operatorname{sech} u \cos v, \operatorname{sech} u \sin v, \tanh u)$$

is a regular surface patch for $S^2$ (it is called *Mercator's projection*). Show that meridians and parallels on $S^2$ correspond under $\boldsymbol{\sigma}$ to perpendicular straight lines in the plane. (This patch is 'derived' in Exercise 6.3.3.)

5.3.3 Show that, if $\boldsymbol{\sigma}(u,v)$ is the (generalized) cylinder in Example 5.3.1:

(i) The curve $\tilde{\boldsymbol{\gamma}}(u) = \boldsymbol{\gamma}(u) - (\boldsymbol{\gamma}(u)\cdot\mathbf{a})\mathbf{a}$ is contained in a plane perpendicular to $\mathbf{a}$.

(ii) $\boldsymbol{\sigma}(u,v) = \tilde{\boldsymbol{\gamma}}(u) + \tilde{v}\mathbf{a}$, where $\tilde{v} = v + \boldsymbol{\gamma}(u)\cdot\mathbf{a}$.

(iii) $\tilde{\boldsymbol{\sigma}}(u,\tilde{v}) = \tilde{\boldsymbol{\gamma}}(u) + \tilde{v}\mathbf{a}$ is a reparametrization of $\boldsymbol{\sigma}(u,v)$.

This exercise shows that, when considering a generalized cylinder $\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\mathbf{a}$, we can always assume that the curve $\boldsymbol{\gamma}$ is contained in a plane perpendicular to the vector $\mathbf{a}$.

5.3.4 Consider the ruled surface

$$\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\boldsymbol{\delta}(u), \tag{5.5}$$

where $\| \boldsymbol{\delta}(u) \| = 1$ and $\dot{\boldsymbol{\delta}}(u) \neq \mathbf{0}$ for all values of $u$ (a dot denotes $d/du$). Show that there is a unique point $\boldsymbol{\Gamma}(u)$, say, on the ruling through $\boldsymbol{\gamma}(u)$ at which $\dot{\boldsymbol{\delta}}(u)$ is perpendicular to the surface. The curve $\boldsymbol{\Gamma}$ is called the *line of striction* of the ruled surface $\boldsymbol{\sigma}$ (of course, it need not be a straight line). Show that $\dot{\boldsymbol{\Gamma}}\cdot\dot{\boldsymbol{\delta}} = 0$. Let $\tilde{v} = v + \frac{\dot{\boldsymbol{\gamma}}\cdot\dot{\boldsymbol{\delta}}}{\|\dot{\boldsymbol{\delta}}\|^2}$, and let $\tilde{\boldsymbol{\sigma}}(u,\tilde{v})$ be the corresponding reparametrization of $\boldsymbol{\sigma}$. Then, $\tilde{\boldsymbol{\sigma}}(u,\tilde{v}) = \boldsymbol{\Gamma}(u) + \tilde{v}\boldsymbol{\delta}(u)$. This means that, when considering ruled surfaces as in (5.5), *we can always assume that* $\dot{\boldsymbol{\gamma}}\cdot\dot{\boldsymbol{\delta}} = 0$. We shall make use of this in Chapter 12.

## 5.4 Compact surfaces

A subset $X$ of $\mathbb{R}^3$ is called *compact* if it is *closed* (i.e., the set of points in $\mathbb{R}^3$ that are *not* in $X$ is open) and *bounded* (i.e., $X$ is contained in some open ball). On several occasions later in the book we shall be particularly interested in compact surfaces.

### Example 5.4.1

Any sphere is compact. Let us consider the unit sphere $S^2$ for simplicity. Obviously $S^2$ is bounded as it is contained in the open ball $D_2(\mathbf{0})$ (for example).

To show that $S^2$ is closed, let $\mathbf{p}$ be a point not in $S^2$, so that $\parallel \mathbf{p} \parallel \neq 1$. Suppose, for example, that $\parallel \mathbf{p} \parallel > 1$ (a similar argument applies if $\parallel \mathbf{p} \parallel < 1$). Let $\epsilon = \parallel \mathbf{p} \parallel -1$. Then the open ball $D_\epsilon(\mathbf{p})$ does not intersect $S^2$, for if $\mathbf{q} \in D_\epsilon(\mathbf{p})$ the triangle inequality $\parallel \mathbf{p} \parallel = \parallel (\mathbf{p} - \mathbf{q}) + \mathbf{q} \parallel \leq \parallel \mathbf{p} - \mathbf{q} \parallel + \parallel \mathbf{q} \parallel$ gives

$$\parallel \mathbf{q} \parallel \geq \parallel \mathbf{p} \parallel - \parallel \mathbf{p} - \mathbf{q} \parallel > \parallel \mathbf{p} \parallel - \epsilon = 1,$$

so $\parallel \mathbf{q} \parallel > 1$ It follows that the set of points of $\mathbb{R}^3$ that are not in $S^2$ is open.

## Example 5.4.2

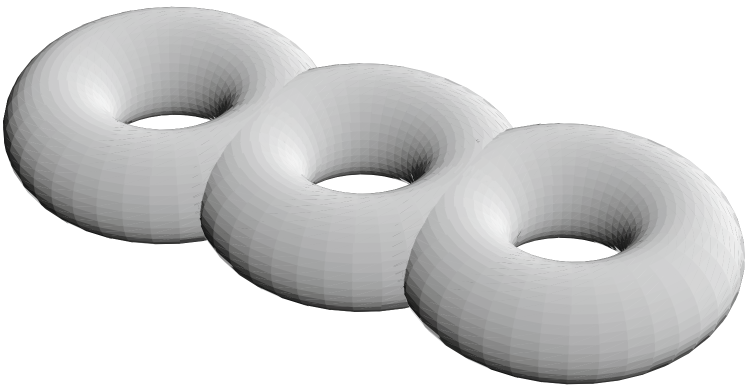
A plane is not compact since it is obviously unbounded.

## Example 5.4.3

The open disc

$$\mathcal{D} = \{(x, y, z) \in \mathbb{R}^2 \mid x^2 + y^2 < 1, \ z = 0\}$$

is a non-compact surface. It is obviously bounded (it is contained in $D_1(\mathbf{0})$); it is not closed, however, since the point $\mathbf{p} = (1, 0, 0)$ is not in $\mathcal{D}$ and for any $\epsilon > 0$ the open ball $D_\epsilon(\mathbf{p})$ contains the point $(1 - \frac{1}{2}\epsilon, 0, 0)$ which is in $\mathcal{D}$.



It is a surprising result that there are very few compact surfaces in $\mathbb{R}^3$ up to diffeomorphism, and they can all be described explicitly. We have already seen the simplest example, the sphere. The next simplest is the *torus* considered in Exercise 4.2.5. More generally, one can join such tori together (see above). This surface is denoted by $T_g$, where $g$ is the number of holes, called the *genus* of the surface (we take $g = 0$ for the sphere). We accept the following theorem without proof:

### Theorem 5.4.4

For any integer $g \geq 0$, $T_g$ has an atlas making it a smooth surface. Moreover, every compact surface is diffeomorphic to one of the $T_g$.

### Corollary 5.4.5

Every compact surface is orientable.

### Proof

Each of the surfaces $T_g$ obviously has an 'interior', which is bounded, and an 'exterior' which is unbounded. Hence, we can choose the unit normal at each point of the surface to point into the exterior region. This provides a smooth choice of unit normal at every point of the surface $T_g$, so $T_g$ is orientable. Since every compact surface is diffeomorphic to one of the surfaces $T_g$, the corollary follows from Exercise 4.5.2. □

### *EXERCISES*

5.4.1 One of the following surfaces is compact and one is not:

(i) $x^2 - y^2 + z^4 = 1$.

(ii) $x^2 + y^2 + z^4 = 1$.

Which is which, and why? Sketch the compact surface.

5.4.2 Explain, without giving a detailed proof, why the tube (Exercise 4.2.7) around a closed curve in $\mathbb{R}^3$ with no self-intersections is a compact surface diffeomorphic to a torus (provided the tube has sufficiently small radius).

## 5.5 Triply orthogonal systems

A *triply orthogonal system of surfaces* consists of three families of surfaces such that

(i) Exactly one surface of each family passes through each point of $\mathbb{R}^3$ (or of some open subset of $\mathbb{R}^3$).

(ii) Any two surfaces belonging to different families intersect orthogonally.

The simplest example of such a system is given by the families of planes parallel to the coordinates planes, namely

$$x = u,\ y = v,\ z = w.$$

Fixing the value of $u$ (say) determines a particular plane in the first family, and similarly for the other families. If $\mathbf{p} = (a, b, c) \in \mathbb{R}^3$, there is a unique plane from each family passing through $\mathbf{p}$, namely those corresponding to $u = a$, $v = b$ and $w = c$. The orthogonality property (ii) is obviously satisfied.

More generally, suppose that the three families are of the form

$$U(x, y, z) = u,\ V(x, y, z) = v,\ W(x, y, z) = w, \tag{5.6}$$

where $U$, $V$ and $W$ are smooth functions of $(x, y, z)$. By Theorem 5.1.1, these equations determine three families of smooth surfaces provided the vectors $\mathbf{\nabla} U$, $\mathbf{\nabla} V$ and $\mathbf{\nabla} W$ are non-zero everywhere. Assuming that this condition holds, by Exercise 5.1.2 the non-zero vector $\mathbf{\nabla} U$ is then perpendicular to the tangent plane of the surface $U(x, y, z) = u$ (and similarly for $V, W$), so condition (ii) in the definition of a triply orthogonal system becomes

$$\mathbf{\nabla} U \cdot \mathbf{\nabla} V = \mathbf{\nabla} V \cdot \mathbf{\nabla} W = \mathbf{\nabla} W \cdot \mathbf{\nabla} U = 0. \tag{5.7}$$

Now consider the smooth function

$$F(x, y, z) = (U(x, y, z), V(x, y, z), W(x, y, z)).$$

The Jacobian matrix of $F$ is

$$J(F) = \begin{pmatrix} U_x & U_y & U_z \\ V_x & V_y & V_z \\ W_x & W_y & W_z \end{pmatrix}$$

so the rows of $J(F)$ are the components of the non-zero vectors $\mathbf{\nabla} U$, $\mathbf{\nabla} V$ and $\mathbf{\nabla} W$. By Eq. 5.7, these vectors are orthogonal, and hence linearly independent, so the matrix $J(F)$ is invertible. By the inverse function theorem (see Section 5.6), Eq. 5.6 can be solved for $(x, y, z)$ in terms of $(u, v, w)$ (at least if $(u, v, w)$ is restricted to lie in a suitable open subset of $\mathbb{R}^3$), say

$$(x, y, z) = \mathbf{\Sigma}(u, v, w). \tag{5.8}$$

Then, setting $u$ equal to a constant $u_0$ (say) gives a parametrization $(v, w) \mapsto \mathbf{\Sigma}(u_0, v, w)$ of the surface $U(x, y, z) = u_0$ (and similarly for the other two families of surfaces).

Regarding $x, y, z$ as functions of $u, v, w$ via Eq. 5.8, we can differentiate both sides of the equation

$$U(x, y, z) = u$$

with respect to $u$, $v$ and $w$. This gives

$$\begin{aligned} U_x x_u + U_y y_u + U_z z_u &= 1 \\ U_x x_v + U_y y_v + U_z z_v &= 0 \\ U_x x_w + U_y y_w + U_z z_w &= 0. \end{aligned}$$

These three equations, together with the corresponding equations for $V$ and $W$, can be written in vector form as follows:

$$\begin{aligned} \boldsymbol{\nabla} U \cdot \boldsymbol{\Sigma}_u = 1,\ \boldsymbol{\nabla} U \cdot \boldsymbol{\Sigma}_v = 0,\ \boldsymbol{\nabla} U \cdot \boldsymbol{\Sigma}_w &= 0, \\ \boldsymbol{\nabla} V \cdot \boldsymbol{\Sigma}_u = 0,\ \boldsymbol{\nabla} V \cdot \boldsymbol{\Sigma}_v = 1,\ \boldsymbol{\nabla} V \cdot \boldsymbol{\Sigma}_w &= 0, \\ \boldsymbol{\nabla} W \cdot \boldsymbol{\Sigma}_u = 0,\ \boldsymbol{\nabla} W \cdot \boldsymbol{\Sigma}_v = 0,\ \boldsymbol{\nabla} W \cdot \boldsymbol{\Sigma}_w &= 1. \end{aligned} \tag{5.9}$$

By Eqs. 5.8 and 5.9, $\boldsymbol{\nabla} U$ and $\boldsymbol{\Sigma}_u$ are both perpendicular to $\boldsymbol{\nabla} V$ and $\boldsymbol{\nabla} W$, so they are parallel to each other. Thus, $\boldsymbol{\Sigma}_u$ is normal to the surface $U(x, y, z) = u$, and $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_w$ are tangent to it (the last statement is also obvious from the statement at the end of the preceding paragraph).

We shall have more to say about triply orthogonal systems later, but now we shall describe one of the most beautiful examples of such systems, which is provided by the theory of quadric surfaces. Let $p, q$ and $r$ be constants such that $0 < p^2 < q^2 < r^2$. For $(x, y, z) \in \mathbb{R}^3$, $t \neq p^2, q^2$ or $r^2$, let

$$F_t(x, y, z) = \frac{x^2}{p^2 - t} + \frac{y^2}{q^2 - t} + \frac{z^2}{r^2 - t}.$$

Fix a point $(a, b, c) \in \mathbb{R}^3$ with $a, b$ and $c$ all non-zero. The following properties are clear:

(i) $F_t(a, b, c)$ is a continuous function of $t$ in each of the open intervals $(-\infty, p^2)$, $(p^2, q^2)$, $(q^2, r^2)$ and $(r^2, \infty)$.

(ii) $F_t(a, b, c) \to 0$ as $t \to \pm\infty$.

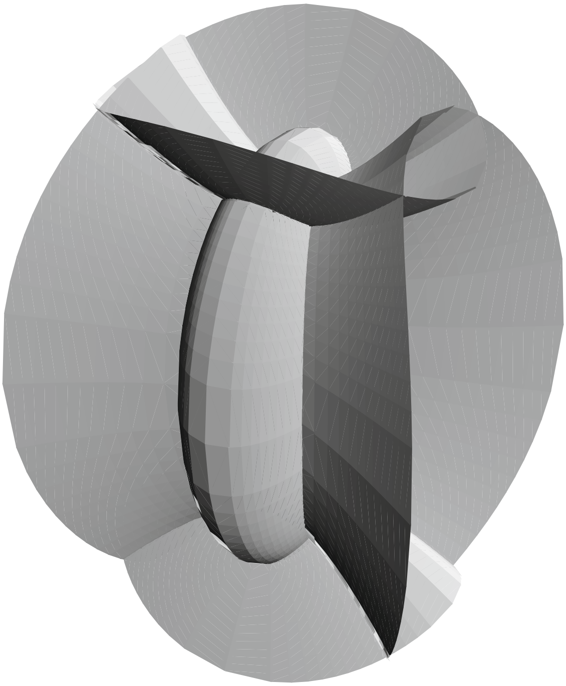
(iii) $F_t(a, b, c) \to \infty$ as $t$ approaches $p^2, q^2$ or $r^2$ from the left, and $F_t(a, b, c) \to -\infty$ as $t$ approaches $p^2, q^2$ or $r^2$ from the right.

It follows from these properties and the intermediate value theorem that there is at least one value of $t$ in each open interval $(-\infty, p^2)$, $(p^2, q^2)$ and $(q^2, r^2)$ such that $F_t(a, b, c) = 1$. On the other hand, the equation $F_t(a, b, c) = 1$ is equivalent to the cubic equation $G_t(a, b, c) = 0$, where

$$\begin{aligned} G_t(a, b, c) = a^2(q^2 - t)(r^2 - t) + b^2(p^2 - t)(r^2 - t) + c^2(p^2 - t)(q^2 - t) \\ -(p^2 - t)(q^2 - t)(r^2 - t), \end{aligned} \tag{5.10}$$

which has at most three real roots. It follows that there are unique numbers $u \in (-\infty, p^2)$, $v \in (p^2, q^2)$ and $w \in (q^2, r^2)$ (depending on $(a, b, c)$, of course) such that

$$F_u(a, b, c) = 1, \quad F_v(a, b, c) = 1, \quad F_w(a, b, c) = 1. \tag{5.11}$$



The three quadrics $F_u(x, y, z) = 1$, $F_v(x, y, z) = 1$ and $F_w(x, y, z) = 1$ are ellipsoids, hyperboloids of one sheet and hyperboloids of two sheets, respectively, and we have shown that there is one of each passing through each point $(a, b, c) \in \mathbb{R}^3$ that does not lie on any of the coordinate planes. We show that they form a triply orthogonal system.

Indeed, the vector

$$\left(\frac{x}{p^2 - t}, \frac{y}{q^2 - t}, \frac{z}{r^2 - t}\right)$$

is perpendicular to the tangent plane of the surface $F_t(x, y, z) = 1$ at $(x, y, z)$. Thus, to show that the first two surfaces in (5.11) are perpendicular at $(a, b, c)$, for example, we have to show that

$$\frac{a^2}{(p^2 - u)(p^2 - v)} + \frac{b^2}{(q^2 - u)(q^2 - v)} + \frac{c^2}{(r^2 - u)(r^2 - v)} = 0.$$

But the left-hand side of this equation is

$$\frac{F_u(a, b, c) - F_v(a, b, c)}{u - v} = \frac{1 - 1}{u - v} = 0.$$

We can also construct a simultaneous parametrization of the three families. Note that the cubic $G_t(a,b,c)$ in (5.10) is equal to $(t-u)(t-v)(t-w)$, since it is divisible by this product and the coefficients of $t^3$ agree. Putting $t = p^2, q^2$ and $r^2$ and solving the resulting equations for $a^2, b^2$ and $c^2$, we find that

$$\begin{aligned} a &= \pm\sqrt{\frac{(p^2-u)(p^2-v)(p^2-w)}{(r^2-p^2)(q^2-p^2)}}, \\ b &= \pm\sqrt{\frac{(q^2-u)(q^2-v)(q^2-w)}{(p^2-q^2)(r^2-q^2)}}, \\ c &= \pm\sqrt{\frac{(r^2-u)(r^2-v)(r^2-w)}{(p^2-r^2)(q^2-r^2)}}. \end{aligned} \tag{5.12}$$

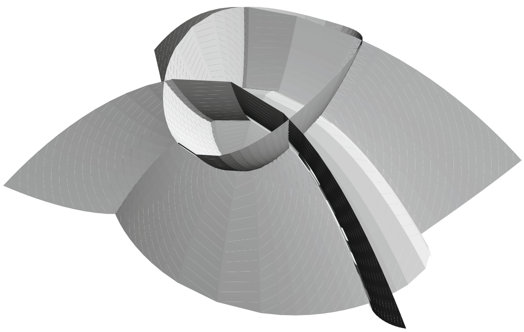
Define $\boldsymbol{\sigma}(u,v,w) = (x,y,z)$, where $x, y$ and $z$ are the right-hand sides of the three equations in (5.12), respectively, with any combination of signs. For fixed $u$ (resp. fixed $v$, fixed $w$), this gives eight surface patches for the corresponding ellipsoid $F_u(x,y,z) = 1$ (resp. hyperboloid of one sheet $F_v(x,y,z) = 1$, hyperboloid of two sheets $F_w(x,y,z) = 1$).

## *EXERCISES*

5.5.1 Show that the following are triply orthogonal systems:

(i) The spheres with centre the origin, the planes containing the $z$-axis and the circular cones with axis the $z$-axis.

(ii) The planes parallel to the $xy$-plane, the planes containing the $z$-axis and the circular cylinders with axis the $z$-axis.



5.5.2 By considering the quadric surface $F_t(x,y,z) = 0$, where

$$F_t(x,y,z) = \frac{x^2}{p^2-t} + \frac{y^2}{q^2-t} - 2z + t,$$

construct a triply orthogonal system (illustrated above) consisting of two families of elliptic paraboloids and one family of hyperbolic paraboloids. Find a parametrization of these surfaces analogous to (5.12).

## 5.6 Applications of the inverse function theorem

In this section we give the proofs of Propositions 4.2.6 and 4.4.6 and Theorem 5.1.1.

Suppose first that $f : U \to \mathbb{R}^n$ is a smooth map, where $U$ is an open subset of $\mathbb{R}^m$. If we write $(\tilde{u}_1, \ldots, \tilde{u}_n) = f(u_1, \ldots, u_m)$, the *Jacobian matrix* of $f$ is

$$J(f) = \begin{pmatrix} \frac{\partial \tilde{u}_1}{\partial u_1} & \frac{\partial \tilde{u}_1}{\partial u_2} & \cdots & \frac{\partial \tilde{u}_1}{\partial u_m} \\ \frac{\partial \tilde{u}_2}{\partial u_1} & \frac{\partial \tilde{u}_2}{\partial u_2} & \cdots & \frac{\partial \tilde{u}_2}{\partial u_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \tilde{u}_n}{\partial u_1} & \frac{\partial \tilde{u}_n}{\partial u_2} & \cdots & \frac{\partial \tilde{u}_n}{\partial u_m} \end{pmatrix}.$$

This has already been used in the case $m = n = 2$ in Section 4.2, but now we shall need it in other cases too.

The main tool that we use is the following theorem.

### Theorem 5.6.1 (Inverse Function Theorem)

Let $f : U \to \mathbb{R}^n$ be a smooth map defined on an open subset $U$ of $\mathbb{R}^n$ $(n \geq 1)$. Assume that, at some point $x_0 \in U$, the Jacobian matrix $J(f)$ is invertible. Then, there is an open subset $V$ of $\mathbb{R}^n$ and a smooth map $g : V \to \mathbb{R}^n$ such that

(i) $y_0 = f(x_0) \in V$

(ii) $g(y_0) = x_0$

(iii) $g(V) \subseteq U$

(iv) $g(V)$ is an open subset of $\mathbb{R}^n$

(v) $f(g(y)) = y$ for all $y \in V$

In particular, $g : V \to g(V)$ and $f : g(V) \to V$ are inverse bijections.

Thus, the inverse function theorem says that, if $J(f)$ is invertible at some point, then $f$ is bijective near that point and its inverse map is smooth. A proof of this theorem can be found in books on multivariable calculus.

As our first application of Theorem 5.6.1, we complete the proof of Proposition 4.4.6. Suppose then that $f : \mathcal{S} \to \tilde{\mathcal{S}}$ is a smooth map between surfaces $\mathcal{S}$ and $\tilde{\mathcal{S}}$, let $\mathbf{p} \in \mathcal{S}$ and assume that the linear map $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is invertible. Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a surface patch of $\mathcal{S}$ containing $\mathbf{p}$, say $\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$, and let $\tilde{\boldsymbol{\sigma}} : \tilde{U} \to \mathbb{R}^3$ be a surface patch of $\tilde{\mathcal{S}}$ containing $f(\mathbf{p})$. By shrinking $U$ if necessary, we can assume that $f$ maps $\boldsymbol{\sigma}(U)$ into $\tilde{\boldsymbol{\sigma}}(\tilde{U})$. Since $f$ is smooth, there are smooth functions $\alpha : U \to \mathbb{R}$ and $\beta : U \to \mathbb{R}$ such that

$$f(\boldsymbol{\sigma}(u, v)) = \tilde{\boldsymbol{\sigma}}(\alpha(u, v), \beta(u, v)).$$

From the remarks following Proposition 4.4.4, the matrix of $D_{\mathbf{p}}f$ with respect to the bases $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$ and $\{\tilde{\boldsymbol{\sigma}}_{\tilde{u}}, \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\}$ of $T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is the Jacobian matrix

$$\begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix}.$$

Since $D_{\mathbf{p}}f$ is invertible, so is this matrix. By the inverse function theorem, the smooth map $U \to \mathbb{R}^2$ given by $(u, v) \mapsto (\alpha(u, v), \beta(u, v))$ is a diffeomorphism from an open subset $V$ (say) of $U$ containing $(u_0, v_0)$ to an open subset $\tilde{V}$ (say) of $\tilde{U}$. Then $\mathcal{O} = \boldsymbol{\sigma}(V)$ and $\tilde{\mathcal{O}} = \tilde{\boldsymbol{\sigma}}(\tilde{V})$ are open subsets of $\mathcal{S}$ and $\tilde{\mathcal{S}}$, respectively, and $f$ is a diffeomorphism from $\mathcal{O}$ to $\tilde{\mathcal{O}}$. This proves that $f$ is a local diffeomorphism.

We now give the proof of Proposition 4.2.6. We want to show that, if $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ and $\tilde{\boldsymbol{\sigma}} : \tilde{U} \to \mathbb{R}^3$ are two regular patches in the atlas of a surface $\mathcal{S}$, the transition map from $\boldsymbol{\sigma}$ to $\tilde{\boldsymbol{\sigma}}$ is smooth where it is defined.

Suppose that a point $\mathbf{p}$ lies in both patches, say $\boldsymbol{\sigma}(u_0, v_0) = \tilde{\boldsymbol{\sigma}}(\tilde{u}_0, \tilde{v}_0) = \mathbf{p}$. Write

$$\boldsymbol{\sigma}(u, v) = (f(u, v), g(u, v), h(u, v)).$$

Since $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent, the Jacobian matrix

$$\begin{pmatrix} f_u & f_v \\ g_u & g_v \\ h_u & h_v \end{pmatrix}$$

of $\boldsymbol{\sigma}$ has rank 2 everywhere. Hence, at least one of its three $2 \times 2$ submatrices is invertible at each point. Suppose that the submatrix

$$\begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}$$

is invertible at $\mathbf{p}$. (The proof is similar in the other two cases.) By the inverse function theorem applied to the map $F : U \to \mathbb{R}^2$ given by

$$F(u, v) = (f(u, v), g(u, v)),$$

there is an open subset $V$ of $\mathbb{R}^2$ containing $F(u_0, v_0)$ and an open subset $W$ of $U$ containing $(u_0, v_0)$ such that $F : W \to V$ is bijective with a smooth inverse $F^{-1} : V \to W$. Since $\boldsymbol{\sigma} : W \to \boldsymbol{\sigma}(W)$ is bijective, the projection $\pi : \boldsymbol{\sigma}(W) \to V$ given by $\pi(x, y, z) = (x, y)$ is also bijective, since $\pi = F \circ \boldsymbol{\sigma}^{-1}$ on $\boldsymbol{\sigma}(W)$. It follows that $\tilde{W} = \tilde{\boldsymbol{\sigma}}^{-1}(\boldsymbol{\sigma}(W))$ is an open subset of $\tilde{U}$ and that

$$\boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}} = F^{-1} \circ \tilde{F}$$

on $\tilde{W}$, where $\tilde{F} = \pi \circ \tilde{\boldsymbol{\sigma}}$. Since $F^{-1}$ and $\tilde{F}$ are smooth on $\tilde{W}$, so is the transition map $\boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}}$. Since $\boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}}$ is smooth on an open set containing any point $(u_0, v_0)$ where it is defined, it is smooth.

Finally, we give the proof of Theorem 5.1.1. Let $\mathbf{p}$, $W$ and $f$ be as in the statement of the theorem, and suppose that $\mathbf{p} = (x_0, y_0, z_0)$ and that $f_z \neq 0$ at $\mathbf{p}$. (The proof is similar in the other two cases.) Consider the map $F : W \to \mathbb{R}^3$ defined by

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

The Jacobian matrix of $F$ is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ f_x & f_y & f_z \end{pmatrix},$$

and is clearly invertible at $\mathbf{p}$ since $f_z \neq 0$. By the inverse function theorem, there is an open subset $V$ of $\mathbb{R}^3$ containing $F(x_0, y_0, z_0) = (x_0, y_0, 0)$ and a smooth map $G : V \to W$ such that $\tilde{W} = G(V)$ is open and $F : \tilde{W} \to V$ and $G : V \to \tilde{W}$ are inverse bijections.

Since $V$ is open, there are open subsets $U_1$ of $\mathbb{R}^2$ containing $(x_0, y_0)$ and $U_2$ of $\mathbb{R}$ containing 0 such that $V$ contains the open set $U_1 \times U_2$ of all points $(x, y, w)$ with $(x, y) \in U_1$ and $w \in U_2$. Hence, we might as well assume that $V = U_1 \times U_2$. The fact that $F$ and $G$ are inverse bijections means that

$$G(x, y, w) = (x, y, g(x, y, w))$$

for some smooth map $g : U_1 \times U_2 \to \mathbb{R}$, and

$$f(x, y, g(x, y, w)) = w$$

for all $(x, y) \in U_1$, $w \in U_2$.

Define $\boldsymbol{\sigma} : U_1 \to \mathbb{R}^3$ by

$$\boldsymbol{\sigma}(x, y) = (x, y, g(x, y, 0)).$$

Then $\boldsymbol{\sigma}$ is a homeomorphism from $U_1$ to $\mathcal{S} \cap \tilde{W}$ (whose inverse is the restriction to $\mathcal{S} \cap \tilde{W}$ of the projection $\pi(x, y, z) = (x, y)$). It is obvious that $\boldsymbol{\sigma}$ is smooth, and it is regular because

$$\boldsymbol{\sigma}_x \times \boldsymbol{\sigma}_y = (-g_x, -g_y, 1)$$

is nowhere zero. So $\boldsymbol{\sigma}$ is a regular surface patch on $\mathcal{S}$ containing the given point $\mathbf{p}$. Since $\mathbf{p}$ was an arbitrary point of $\mathcal{S}$, we have constructed an atlas for $\mathcal{S}$ making it into a (smooth) surface.

## *EXERCISES*

5.6.1 Show that, if $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^3$ is a curve whose image is contained in a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$, then $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ for some smooth map $(\alpha, \beta) \to U$, $t \mapsto (u(t), v(t))$.

5.6.2 Prove Theorem 1.5.1 and its analogue for level curves in $\mathbb{R}^3$ (Exercise 1.5.1).

5.6.3 Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a smooth map such that $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \neq \mathbf{0}$ at some point $(u_0, v_0) \in U$. Show that there is an open subset $W$ of $U$ containing $(u_0, v_0)$ such that the restriction of $\boldsymbol{\sigma}$ to $W$ is injective. Note that, in the text, surface patches are injective by definition, but this exercise shows that injectivity near a given point is a consequence of regularity.

5.6.4 Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular surface patch, let $(u_0, v_0) \in U$ and let $\boldsymbol{\sigma}(u_0, v_0) = (x_0, y_0, z_0)$. Suppose that the unit normal $\mathbf{N}(u_0, v_0)$ is not parallel to the $xy$-plane. Show that there is an open set $V$ in $\mathbb{R}^2$ containing $(x_0, y_0)$, an open subset $W$ of $U$ containing $(u_0, v_0)$ and a smooth function $\varphi : V \to \mathbb{R}$ such that $\tilde{\boldsymbol{\sigma}}(x, y) = (x, y, \varphi(x, y))$ is a reparametrization of $\boldsymbol{\sigma} : W \to \mathbb{R}^3$. Thus, 'near' $\mathbf{p}$, the surface is part of the graph $z = \varphi(x, y)$.
What happens if $\mathbf{N}(u_0, v_0)$ is parallel to the $xy$-plane?

# 6
# The first fundamental form

Perhaps the first thing that a geometrically inclined bug living on a surface might wish to do is to measure the distance between two points of the surface. Of course, this will usually be different from the distance between these points as measured by an inhabitant of the ambient three-dimensional space, since the straight line segment which furnishes the shortest path between the points in $\mathbb{R}^3$ will generally not be contained in the surface. The object that allows one to compute lengths on a surface, and also angles and areas, is the *first fundamental form* of the surface.

## 6.1 Lengths of curves on surfaces

If our bug-geometer walks along a curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$, the distance he travels is

$$\int \parallel \dot{\boldsymbol{\gamma}}(t) \parallel \, dt$$

(see Definition 1.2.1). To compute this he would need to be able to find the length of *tangent vectors* to the surface, such as $\dot{\boldsymbol{\gamma}}$, which in turn can be computed from the object in the following definition.

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_6,

### Definition 6.1.1

Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$. The *first fundamental form* of $\mathcal{S}$ at $\mathbf{p}$ associates to tangent vectors $\mathbf{v}, \mathbf{w} \in T_\mathbf{p}\mathcal{S}$ the scalar

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p},\mathcal{S}} = \mathbf{v} \cdot \mathbf{w}.$$

Thus, $\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p},\mathcal{S}}$ is just the dot product, but restricted to tangent vectors to $\mathcal{S}$ at $\mathbf{p}$. We shall usually omit one or both of the subscripts unless there is some danger of confusion as to which point or surface is intended.

The first fundamental form $\langle\ ,\ \rangle$ is an example of an *inner product* (see Appendix 0): this follows immediately from the fact that the dot product defines an inner product on $\mathbb{R}^3$.

In traditional works on this subject, the first fundamental form looks slightly different. Suppose that $\boldsymbol{\sigma}(u, v)$ is a surface patch of $\mathcal{S}$. Then, any tangent vector to $\mathcal{S}$ at a point $\mathbf{p}$ in the image of $\boldsymbol{\sigma}$ can be expressed uniquely as a linear combination of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Define maps $du : T_\mathbf{p}\mathcal{S} \to \mathbb{R}$ and $dv : T_\mathbf{p}\mathcal{S} \to \mathbb{R}$ by

$$du(\mathbf{v}) = \lambda, \quad dv(\mathbf{v}) = \mu \qquad \text{if } \mathbf{v} = \lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v,$$

for some $\lambda, \mu \in \mathbb{R}$. It is easy to see that $du$ and $dv$ are *linear* maps. Then, using the fact that $\langle\ ,\ \rangle$ is a symmetric bilinear form, we have

$$\langle \mathbf{v}, \mathbf{v} \rangle = \lambda^2 \langle \boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u \rangle + 2\lambda\mu \langle \boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v \rangle + \mu^2 \langle \boldsymbol{\sigma}_v, \boldsymbol{\sigma}_v \rangle.$$

Writing

$$E = \parallel \boldsymbol{\sigma}_u \parallel^2, \quad F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v, \quad G = \parallel \boldsymbol{\sigma}_v \parallel^2,$$

this becomes

$$\langle \mathbf{v}, \mathbf{v} \rangle = E\lambda^2 + 2F\lambda\mu + G\mu^2 = Edu(\mathbf{v})^2 + 2Fdu(\mathbf{v})dv(\mathbf{v}) + Gdv(\mathbf{v})^2.$$

Traditionally, the expression

$$Edu^2 + 2Fdudv + Gdv^2$$

is called the first fundamental form of the surface patch $\boldsymbol{\sigma}(u, v)$. Note that the coefficients $E, F, G$ and the linear maps $du, dv$ depend on the choice of surface patch for $\mathcal{S}$ (see Exercise 6.1.4), but the first fundamental form itself depends only on $\mathcal{S}$ and $\mathbf{p}$.

If $\boldsymbol{\gamma}$ is a curve lying in the image of a surface patch $\boldsymbol{\sigma}$, we have

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$$

for some smooth functions $u(t)$ and $v(t)$. Then, denoting $d/dt$ by a dot, we have $\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$ by the chain rule, so

$$\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2,$$

and the length of $\boldsymbol{\gamma}$ is given by

$$\int (E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{1/2} dt. \tag{6.1}$$

## Example 6.1.2

For the plane

$$\boldsymbol{\sigma}(u,v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q}$$

(see Example 4.1.2) with $\mathbf{p}$ and $\mathbf{q}$ being perpendicular unit vectors, we have $\boldsymbol{\sigma}_u = \mathbf{p}$, $\boldsymbol{\sigma}_v = \mathbf{q}$, so $E = \parallel \boldsymbol{\sigma}_u \parallel^2 = \parallel \mathbf{p} \parallel^2 = 1$, $F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = \mathbf{p} \cdot \mathbf{q} = 0$, $G = \parallel \boldsymbol{\sigma}_v \parallel^2 = \parallel \mathbf{q} \parallel^2 = 1$, and the first fundamental form is simply

$$du^2 + dv^2.$$

## Example 6.1.3

Consider a surface of revolution

$$\boldsymbol{\sigma}(u,v) = (f(u)\cos v, f(u)\sin v, g(u)).$$

Recall from Example 5.3.2 that we can assume that $f(u) > 0$ for all values of $u$ and that the profile curve $u \mapsto (f(u), 0, g(u))$ is unit-speed, i.e., $\dot{f}^2 + \dot{g}^2 = 1$ (a dot denoting $d/du$). Then:

$$\boldsymbol{\sigma}_u = (\dot{f}\cos v, \dot{f}\sin v, \dot{g}), \quad \boldsymbol{\sigma}_v = (-f\sin v, f\cos v, 0),$$
$$\therefore \quad E = \parallel \boldsymbol{\sigma}_u \parallel^2 = \dot{f}^2 + \dot{g}^2 = 1,\ F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0,\ G = \parallel \boldsymbol{\sigma}_v \parallel^2 = f^2.$$

So the first fundamental form is

$$du^2 + f(u)^2 dv^2.$$

A special case is the unit sphere $S^2$ in latitude-longitude coordinates (Example 4.1.4). We take $u = \theta$, $v = \varphi$, $f(\theta) = \cos\theta$, $g(\theta) = \sin\theta$, giving the first fundamental form of $S^2$ as

$$d\theta^2 + \cos^2\theta\, d\varphi^2.$$

## Example 6.1.4

We consider a generalized cylinder

$$\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\mathbf{a}$$

defined in Example 5.3.1. As we saw in Exercise 5.3.3, we can assume that $\boldsymbol{\gamma}$ is unit-speed, $\mathbf{a}$ is a unit vector, and $\boldsymbol{\gamma}$ is contained in a plane perpendicular to $\mathbf{a}$. Then, denoting $d/du$ by a dot, $\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}}$, $\boldsymbol{\sigma}_v = \mathbf{a}$, so $E = \parallel \boldsymbol{\sigma}_u \parallel^2 = \parallel \dot{\boldsymbol{\gamma}} \parallel^2 = 1$,

$F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = \dot{\boldsymbol{\gamma}} \cdot \mathbf{a} = 0$, $G = \| \boldsymbol{\sigma}_v \|^2 = \| \mathbf{a} \|^2 = 1$, and the first fundamental form of $\boldsymbol{\sigma}$ is

$$du^2 + dv^2.$$

Note that this is the *same* as the first fundamental form of the plane (Example 6.1.2). The geometrical reason for this coincidence will be revealed in the next section.

### Example 6.1.5

We consider a generalized cone

$$\boldsymbol{\sigma}(u, v) = (1 + v)\boldsymbol{\gamma}(u) - v\mathbf{v}$$

(Example 5.3.1). Before computing its first fundamental form, we make some simplifications to $\boldsymbol{\sigma}$.

First, translating the surface by $\mathbf{v}$ (which does not change its first fundamental form – see Exercise 6.1.2), we get the surface patch $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma} - \mathbf{v} = (1 + v)(\boldsymbol{\gamma} - \mathbf{v})$, so if we replace $\boldsymbol{\gamma}$ by $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma} - \mathbf{v}$ we get $\boldsymbol{\sigma}_1 = (1 + v)\boldsymbol{\gamma}_1$. This means that we might as well assume that $\mathbf{v} = \mathbf{0}$ to begin with. Next, we saw in Example 5.3.1 that for $\boldsymbol{\sigma}$ to be a regular surface patch, $\boldsymbol{\gamma}$ must not pass through the origin, so we can define a new curve $\tilde{\boldsymbol{\gamma}}$ by $\tilde{\boldsymbol{\gamma}}(u) = \boldsymbol{\gamma}(u)/ \| \boldsymbol{\gamma}(u) \|$. Setting $\tilde{u} = u$, $\tilde{v} = (1 + v)/ \| \boldsymbol{\gamma}(u) \|$, we get a reparametrization $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v}) = \tilde{v}\tilde{\boldsymbol{\gamma}}(\tilde{u})$ of $\boldsymbol{\sigma}$ with $\| \tilde{\boldsymbol{\gamma}} \| = 1$. We can therefore assume to begin with that $\boldsymbol{\sigma}(u, v) = v\boldsymbol{\gamma}(u)$ with $\| \boldsymbol{\gamma}(u) \| = 1$ for all values of $u$ (geometrically, this means that we can replace $\boldsymbol{\gamma}$ by the intersection of the cone with $S^2$). Finally, reparametrizing again, we can assume that $\boldsymbol{\gamma}$ is unit-speed, for we saw in Example 5.3.1 that for $\boldsymbol{\sigma}$ to be regular, $\boldsymbol{\gamma}$ must be regular.

With these assumptions, and with a dot denoting $d/du$, we have $\boldsymbol{\sigma}_u = v\dot{\boldsymbol{\gamma}}$, $\boldsymbol{\sigma}_v = \boldsymbol{\gamma}$, so $E = \| v\dot{\boldsymbol{\gamma}} \|^2 = v^2 \| \dot{\boldsymbol{\gamma}} \|^2 = v^2$, $F = v\dot{\boldsymbol{\gamma}} \cdot \boldsymbol{\gamma} = 0$ (since $\| \boldsymbol{\gamma} \| = 1$), $G = \| \boldsymbol{\gamma} \|^2 = 1$, and the first fundamental form is

$$v^2 du^2 + dv^2.$$

Note that, as for the generalized cylinder in Example 6.1.4, the first fundamental form of the generalized cone does not depend on the curve $\boldsymbol{\gamma}$.

## *EXERCISES*

6.1.1 Calculate the first fundamental forms of the following surfaces:

(i) $\boldsymbol{\sigma}(u, v) = (\sinh u \sinh v, \sinh u \cosh v, \sinh u)$.

(ii) $\boldsymbol{\sigma}(u, v) = (u - v, u + v, u^2 + v^2)$.

(iii) $\boldsymbol{\sigma}(u, v) = (\cosh u, \sinh u, v)$.

(iv) $\boldsymbol{\sigma}(u, v) = (u, v, u^2 + v^2)$.

What kinds of surfaces are these?

6.1.2 Show that applying an isometry of $\mathbb{R}^3$ to a surface does not change its first fundamental form. What is the effect of a dilation (i.e., a map $\mathbb{R}^3 \to \mathbb{R}^3$ of the form $\mathbf{v} \mapsto a\mathbf{v}$ for some constant $a \neq 0$)?

6.1.3 Let $Edu^2 + 2Fdudv + Gdv^2$ be the first fundamental form of a surface patch $\boldsymbol{\sigma}(u, v)$ of a surface $\mathcal{S}$. Show that, if $\mathbf{p}$ is a point in the image of $\boldsymbol{\sigma}$ and $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$, then

$$\langle \mathbf{v}, \mathbf{w} \rangle = Ed\ \ (\mathbf{v})d\ \ (\mathbf{w}) + F(du(\mathbf{v})dv(\mathbf{w}) + du(\mathbf{w})dv(\mathbf{v})) + Gdv(\mathbf{w})dv(\mathbf{w}).$$

6.1.4 Suppose that a surface patch $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ is a reparametrization of a surface patch $\boldsymbol{\sigma}(u, v)$, and let

$$\tilde{E}d\tilde{u}^2 + 2\tilde{F}d\tilde{u}d\tilde{v} + \tilde{G}d\tilde{v}^2 \quad \text{and} \quad Edu^2 + 2Fdudv + Gdv^2$$

be their first fundamental forms. Show that:

(i) $du = \frac{\partial u}{\partial \tilde{u}}d\tilde{u} + \frac{\partial u}{\partial \tilde{v}}d\tilde{v}, \quad dv = \frac{\partial v}{\partial \tilde{u}}d\tilde{u} + \frac{\partial v}{\partial \tilde{v}}d\tilde{v}$.

(ii) If

$$J = \begin{pmatrix} \dfrac{\partial u}{\partial \tilde{u}} & \dfrac{\partial u}{\partial \tilde{v}} \\ \dfrac{\partial v}{\partial \tilde{u}} & \dfrac{\partial v}{\partial \tilde{v}} \end{pmatrix}$$

is the Jacobian matrix of the reparametrization map $(\tilde{u}, \tilde{v}) \mapsto (u, v)$, and $J^t$ is the transpose of $J$, then

$$\begin{pmatrix} \tilde{E} & \tilde{F} \\ \tilde{F} & \tilde{G} \end{pmatrix} = J^t \begin{pmatrix} E & F \\ F & G \end{pmatrix} J.$$

6.1.5 Show that the following are equivalent conditions on a surface patch $\boldsymbol{\sigma}(u, v)$ with first fundamental form $Edu^2 + 2Fdudv + Gdv^2$:

(i) $E_v = G_u = 0$.

(ii) $\boldsymbol{\sigma}_{uv}$ is parallel to the standard unit normal $\mathbf{N}$.

(iii) The opposite sides of any quadrilateral formed by parameter curves of $\boldsymbol{\sigma}$ have the same length (see the remarks following the proof of Proposition 4.4.2).

When these conditions are satisfied, the parameter curves of $\boldsymbol{\sigma}$ are said to form a *Chebyshev net*. Show that, in that case, $\boldsymbol{\sigma}$ has a reparametrization $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ with first fundamental form

$$d\tilde{u}^2 + 2\cos\theta\, d\tilde{u}d\tilde{v} + d\tilde{v}^2,$$

where $\theta$ is a smooth function of $(\tilde{u}, \tilde{v})$. Show that $\theta$ is the angle between the parameter curves of $\tilde{\boldsymbol{\sigma}}$. Show further that, if we put $\hat{u} = \tilde{u} + \tilde{v}$, $\hat{v} = \tilde{u} - \tilde{v}$, the resulting reparametrization $\hat{\boldsymbol{\sigma}}(\hat{u}, \hat{v})$ of $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ has first fundamental form

$$\cos^2\omega\, d\hat{u}^2 + \sin^2\omega\, d\hat{v}^2,$$

where $\omega = \theta/2$.

## 6.2 Isometries of surfaces

We observed in Example 6.1.4 that a plane and a generalized cylinder, when suitably parametrized, have the *same* first fundamental form. The geometric reason for this is not hard to see. A plane piece of paper can be 'wrapped' on a cylinder in the obvious way without crumpling the paper (see Example 4.3.2). If we draw a curve on the plane, then after wrapping it becomes a curve on the cylinder. Because there is no crumpling, the lengths of these two curves will be the same. Since the lengths are computed as the integral of (the square root of) the first fundamental form, it is plausible that the first fundamental forms of the two surfaces should be the same. Experiment suggests, on the other hand, that it is impossible to wrap a plane sheet of paper around a sphere without crumpling. Thus, we expect that a plane and a sphere do not have the same first fundamental form.

The following definition makes precise what it means to wrap one surface onto another without crumpling.

### Definition 6.2.1

If $\mathcal{S}_1$ and $\mathcal{S}_2$ are surfaces, a smooth map $f : \mathcal{S}_1 \to \mathcal{S}_2$ is called a *local isometry* if it takes any curve in $\mathcal{S}_1$ to a curve of *the same length* in $\mathcal{S}_2$. If a local isometry $f : \mathcal{S}_1 \to \mathcal{S}_2$ exists, we say that $\mathcal{S}_1$ and $\mathcal{S}_2$ are *locally isometric*.

We shall see that every local isometry is a local diffeomorphism; a local isometry that is a diffeomorphism is called an *isometry*. It is obvious that any composite of local isometries is a local isometry, and that the inverse of any isometry is an isometry.

To express the condition for a local isometry in a more useful form, we need the following construction. Let $f : \mathcal{S}_1 \to \mathcal{S}_2$ be a smooth map and let $\mathbf{p} \in \mathcal{S}_1$. For $\mathbf{v}, \mathbf{w} \in T_\mathbf{p}\mathcal{S}_1$, define

$$f^*\langle \mathbf{v}, \mathbf{w} \rangle_\mathbf{p} = \langle D_\mathbf{p} f(\mathbf{v}), D_\mathbf{p} f(\mathbf{w}) \rangle_{f(\mathbf{p})}.$$

Then, $f^*\langle\ ,\ \rangle_\mathbf{p}$ is a symmetric bilinear form on $T_\mathbf{p}\mathcal{S}_1$. Indeed, the symmetry is obvious and if $\lambda, \lambda' \in \mathbb{R}$, $\mathbf{v}, \mathbf{v}', \mathbf{w} \in T_\mathbf{p}$,

$$\begin{aligned}
f^*\langle \lambda\mathbf{v} + \lambda'\mathbf{v}', \mathbf{w} \rangle_\mathbf{p} &= \langle D_\mathbf{p} f(\lambda\mathbf{v} + \lambda'\mathbf{v}'), D_\mathbf{p} f(\mathbf{w}) \rangle_{f(\mathbf{p})} \\
&= \langle \lambda D_\mathbf{p} f(\mathbf{v}) + \lambda' D_\mathbf{p} f(\mathbf{v}'), D_\mathbf{p} f(\mathbf{w}) \rangle_{f(\mathbf{p})} \\
&= \lambda \langle D_\mathbf{p} f(\mathbf{v}), D_\mathbf{p} f(\mathbf{w}) \rangle_{f(\mathbf{p})} + \lambda' \langle D_\mathbf{p} f(\mathbf{v}'), D_\mathbf{p} f(\mathbf{w}) \rangle_{f(\mathbf{p})} \\
&= \lambda f^*\langle \mathbf{v}, \mathbf{w} \rangle_\mathbf{p} + \lambda' f^*\langle \mathbf{v}', \mathbf{w} \rangle_\mathbf{p}.
\end{aligned}$$

## Theorem 6.2.2

A smooth map $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a local isometry if and only if the symmetric bilinear forms $\langle\ ,\ \rangle_\mathbf{p}$ and $f^*\langle\ .\ \rangle_\mathbf{p}$ on $T_\mathbf{p}\mathcal{S}_1$ are equal for all $\mathbf{p} \in \mathcal{S}_1$.

## Proof

If $\boldsymbol{\gamma}_1$ is a curve on $\mathcal{S}_1$, the length of the part of $\boldsymbol{\gamma}_1$ with endpoints $\boldsymbol{\gamma}_1(t_0)$ and $\boldsymbol{\gamma}_1(t_1)$ is

$$\int_{t_0}^{t_1} \langle \dot{\boldsymbol{\gamma}}_1, \dot{\boldsymbol{\gamma}}_1 \rangle^{1/2} dt. \tag{6.2}$$

The length of the corresponding part of the curve $\boldsymbol{\gamma}_2 = f \circ \boldsymbol{\gamma}_1$ on $\mathcal{S}_2$ is

$$\int_{t_0}^{t_1} \langle \dot{\boldsymbol{\gamma}}_2, \dot{\boldsymbol{\gamma}}_2 \rangle^{1/2} dt = \int_{t_0}^{t_1} \langle Df(\dot{\boldsymbol{\gamma}}_1), Df(\dot{\boldsymbol{\gamma}}_1) \rangle^{1/2} dt = \int_{t_0}^{t_1} f^*\langle \dot{\boldsymbol{\gamma}}_1, \dot{\boldsymbol{\gamma}}_1 \rangle^{1/2} dt. \tag{6.3}$$

It is now obvious that, if the two symmetric bilinear forms in the statement of the theorem are equal, the curves $\boldsymbol{\gamma}_1$ and $f \circ \boldsymbol{\gamma}_1$ have the same length.

Conversely, suppose that the integrals in ( 6.2) and ( 6.3) are equal for all curves $\boldsymbol{\gamma}$ on $\mathcal{S}_1$. Then, the integrands must be the same for all $\boldsymbol{\gamma}$:

$$\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle = f^*\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle.$$

Since any tangent vector $\mathbf{v}$ to $\mathcal{S}_1$ is the tangent vector of a curve on $\mathcal{S}_1$, it follows that

$$\langle \mathbf{v}, \mathbf{v} \rangle = f^*\langle \mathbf{v}, \mathbf{v} \rangle \quad \text{for all } \mathbf{v}. \tag{6.4}$$

Since $\langle\ ,\ \rangle$ and $f^*\langle\ ,\ \rangle$ are symmetric bilinear forms, it follows from (6.4) that they are equal (see Appendix 0). $\square$

Thus, $f$ is a local isometry if and only if

$$\langle D_{\mathbf{p}}f(\mathbf{v}), D_{\mathbf{p}}f(\mathbf{w})\rangle_{f(\mathbf{p})} = \langle \mathbf{v}, \mathbf{w}\rangle_{\mathbf{p}}$$

for all $\mathbf{p} \in \mathcal{S}_1$ and all $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}_1$. This means that the linear map $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S}_1 \to T_{f(\mathbf{p})}\mathcal{S}_2$ is an *isometry*, i.e., it preserves lengths (see Appendix 1). In short, $f$ is a local isometry if and only if $D_{\mathbf{p}}f$ is an isometry for all $\mathbf{p} \in \mathcal{S}_1$.

It follows from this theorem that every local isometry is a local diffeomorphism. Indeed, let $f : \mathcal{S}_1 \to \mathcal{S}_2$ be a local isometry and let $\mathbf{p} \in \mathcal{S}_1$. If $D_{\mathbf{p}}f$ is not invertible, there is a non-zero tangent vector $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}_1$ such that $D_{\mathbf{p}}f(\mathbf{v}) = \mathbf{0}$. But this gives a contradiction: since $f$ is a local isometry,

$$0 \neq \langle \mathbf{v}, \mathbf{v}\rangle_{\mathbf{p}} = \langle D_{\mathbf{p}}f(\mathbf{v}), D_{\mathbf{p}}f(\mathbf{v})\rangle_{f(\mathbf{p})} = \langle \mathbf{0}, \mathbf{0}\rangle_{\mathbf{p}} = 0.$$

Hence, $D_{\mathbf{p}}f$ is invertible, and so $f$ is a local diffeomorphism (Proposition 4.4.6).

It will be useful to express Theorem 6.2.2 in terms of surface patches.

## Corollary 6.2.3

A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a local isometry if and only if, for any surface patch $\boldsymbol{\sigma}_1$ of $\mathcal{S}_1$, the patches $\boldsymbol{\sigma}_1$ and $f \circ \boldsymbol{\sigma}_1$ of $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, have the same first fundamental form.

## Proof

In view of the theorem, we have to show that the patches $\boldsymbol{\sigma}_1$ and $f \circ \boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$, say, have the same first fundamental form if and only if the symmetric bilinear forms $\langle\ ,\ \rangle_{\mathbf{p}}$ and $f^*\langle\ ,\ \rangle_{\mathbf{p}}$ are equal for all $\mathbf{p}$ in the image of $\boldsymbol{\sigma}_1$.

The first fundamental form of $\boldsymbol{\sigma}_i$ $(i = 1, 2)$ is $E_i du^2 + 2F_i du dv + G_i dv^2$, where $E_i = \langle (\boldsymbol{\sigma}_i)_u, (\boldsymbol{\sigma}_i)_u\rangle$, $F_i = \langle (\boldsymbol{\sigma}_i)_u, (\boldsymbol{\sigma}_i)_v\rangle$, $G_i = \langle (\boldsymbol{\sigma}_i)_v, (\boldsymbol{\sigma}_i)_v\rangle$. We compute

$$\langle (\boldsymbol{\sigma}_2)_u, (\boldsymbol{\sigma}_2)_u\rangle = \langle Df((\boldsymbol{\sigma}_1)_u), Df((\boldsymbol{\sigma}_1)_u)\rangle = f^*\langle (\boldsymbol{\sigma}_1)_u, (\boldsymbol{\sigma}_1)_u\rangle.$$

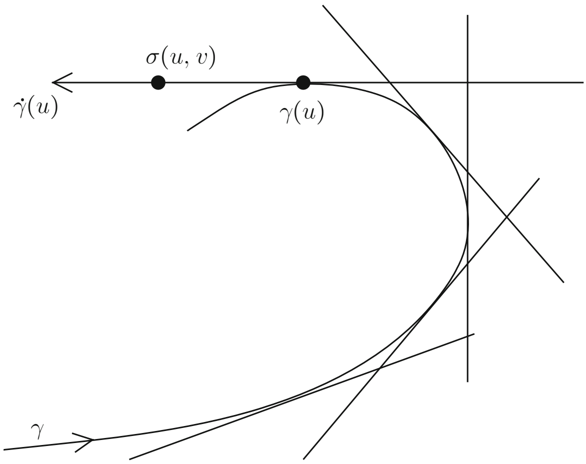
Thus, if $\langle\ ,\ \rangle = f^*\langle\ .\ \rangle$, then $E_1 = E_2$, and similarly $F_1 = F_2$ and $G_1 = G_2$. Conversely, if these last three equations hold, then $\langle \mathbf{v}, \mathbf{w}\rangle = f^*\langle \mathbf{v}, \mathbf{w}\rangle$ whenever the tangent vectors $\mathbf{v}, \mathbf{w}$ are of the form $(\boldsymbol{\sigma}_1)_u$ or $(\boldsymbol{\sigma}_1)_v$. The bilinearity property then implies that $\langle \mathbf{v}, \mathbf{w}\rangle = f^*\langle \mathbf{v}, \mathbf{w}\rangle$ for all $\mathbf{v}, \mathbf{w}$. $\square$

This proof actually shows that, if $\mathbf{p} \in \mathcal{S}_1$ is in the image of a surface patch $\boldsymbol{\sigma}_1$, then $\boldsymbol{\sigma}_1$ and $f \circ \boldsymbol{\sigma}_1$ have the same first fundamental form at $\mathbf{p}$ if and only if $D_{\mathbf{p}}f$ is an isometry; it follows that, if $\mathbf{p}$ is in the image of another surface patch $\boldsymbol{\sigma}_2$, then $\boldsymbol{\sigma}_1$ and $f \circ \boldsymbol{\sigma}_1$ have the same first fundamental form at $\mathbf{p}$ if and only if the same is true of $\boldsymbol{\sigma}_2$ and $f \circ \boldsymbol{\sigma}_2$.

## Example 6.2.4

The map $f$ from the $yz$-plane to the unit cylinder defined in Example 4.3.2 is a local isometry. For, if we use the surface patch $\boldsymbol{\sigma}_1(u,v) = (0,u,v)$ for the plane and $\boldsymbol{\sigma}_2(u,v) = (\cos u, \sin u, v)$ for the cylinder, then $f(\boldsymbol{\sigma}_1(u,v)) = \boldsymbol{\sigma}_2(u,v)$, and by Example 6.1.4 $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ have the same first fundamental form.

A similar argument shows that a generalized cone is locally isometric to a plane (see Example 6.2.1). It turns out that there is another class of surfaces that is locally isometric to a plane, called *tangent developables*. (In older works, a 'development' of one surface on another was the term used for a local isometry.) A tangent developable is the union of the tangent lines to a curve in $\mathbb{R}^3$ – the tangent line to a curve $\boldsymbol{\gamma}$ at a point $\boldsymbol{\gamma}(u)$ is the straight line passing through $\boldsymbol{\gamma}(u)$ and parallel to the tangent vector $\dot{\boldsymbol{\gamma}}(u)$.




We might as well assume that $\boldsymbol{\gamma}$ is unit-speed. The most general point on the tangent line at $\boldsymbol{\gamma}(u)$ is

$$\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\dot{\boldsymbol{\gamma}}(u),$$

for some scalar $v$. Now

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (\dot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\gamma}}) \times \dot{\boldsymbol{\gamma}} = v\ddot{\boldsymbol{\gamma}} \times \dot{\boldsymbol{\gamma}}.$$

For $\boldsymbol{\sigma}$ to be regular, it is thus necessary that $\ddot{\boldsymbol{\gamma}}$ is never zero, or in other words, the curvature $\kappa = \parallel \ddot{\boldsymbol{\gamma}} \parallel$ is $> 0$ at all points of $\boldsymbol{\gamma}$. Now, $\dot{\boldsymbol{\gamma}} = \mathbf{t}$, the unit tangent vector of $\boldsymbol{\gamma}$, and $\ddot{\boldsymbol{\gamma}} = \dot{\mathbf{t}} = \kappa\mathbf{n}$, where $\mathbf{n}$ is the principal normal to $\boldsymbol{\gamma}$, so

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \kappa v \mathbf{n} \times \mathbf{t} = -\kappa v \mathbf{b},$$

where $\mathbf{b}$ is the binormal of $\boldsymbol{\gamma}$. Thus, $\boldsymbol{\sigma}$ will be regular if $\kappa > 0$ everywhere and $v \neq 0$. The latter condition means that, for regularity, we must exclude the curve $\boldsymbol{\gamma}$ itself from the surface. Typically, the regions $v > 0$ and $v < 0$ of the tangent developable form two sheets which meet along a sharp edge formed by the curve $\boldsymbol{\gamma}$ where $v = 0$, as the following illustration of the tangent developable of a circular helix indicates (see Exercise 6.2.4):

Our interest in tangent developables stems from the following result.

## Proposition 6.2.5

Any tangent developable is locally isometric to a plane.

## Proof

We use the above notation, assuming that $\boldsymbol{\gamma}$ is unit-speed and that $\kappa > 0$. Now,

$$\begin{aligned} E &= \| \boldsymbol{\sigma}_u \|^2 = (\dot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\gamma}}) \cdot (\dot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\gamma}}) = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} + 2v\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} + v^2\ddot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 1 + v^2\kappa^2, \\ F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = (\dot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\gamma}}) \cdot \dot{\boldsymbol{\gamma}} = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 1, \\ G &= \| \boldsymbol{\sigma}_v \|^2 = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = 1, \end{aligned}$$

since $\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = 1$, $\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 0$, $\ddot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = \kappa^2$. So the first fundamental form of the tangent developable is

$$(1 + v^2\kappa^2)du^2 + 2dudv + dv^2. \tag{6.5}$$

We are going to show that an open subset of the plane can be parametrized so that it has the same first fundamental form. This will prove the proposition.

By Theorem 2.2.5, there is a *plane* unit-speed curve $\tilde{\boldsymbol{\gamma}}$ whose curvature is $\kappa$ (we can even assume that its signed curvature is $\kappa$). By the above calculations, the first fundamental form of the tangent developable of $\tilde{\boldsymbol{\gamma}}$ is also given by (6.5).

But since $\tilde{\gamma}$ is a plane curve, its tangent lines obviously fill out part of the plane in which $\tilde{\gamma}$ lies. □

There is a converse to Proposition 6.2.5: any sufficiently small open subset of a surface locally isometric to a plane *is* an open subset of a plane, a generalized cylinder, a generalized cone or a tangent developable. The proof of this will be given in Section 8.4.

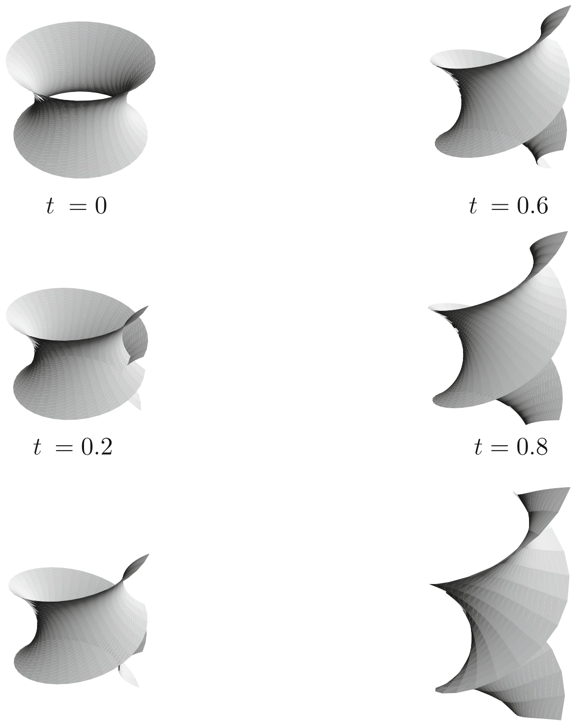
## EXERCISES

6.2.1 By thinking about how a circular cone can be 'unwrapped' onto the plane, write down an isometry from

$$\boldsymbol{\sigma}(u, v) = (u\cos v, u\sin v, u), \qquad u > 0,\ 0 < v < 2\pi,$$

(a circular half-cone with a straight line removed) to an open subset of the $xy$-plane.

6.2.2 Is the map from the circular half-cone $x^2 + y^2 = z^2$, $z > 0$, to the $xy$-plane given by $(x, y, z) \mapsto (x, y, 0)$ a local isometry?



$t = 0$ $t = 0.6$

$t = 0.2$ $t = 0.8$

$t = 0.4$ $t = 1$

6.2.3 Consider the surface patches

$$\boldsymbol{\sigma}(u,v) = (\cosh u \cos v, \cosh u \sin v, u), \quad \tilde{\boldsymbol{\sigma}}(u,v) = (u\cos v, u\sin v, v),$$

parametrizing the catenoid (Exercise 5.3.1) and the helicoid (Exercise 4.2.6), respectively. Show that the map from the catenoid to the helicoid that takes $\boldsymbol{\sigma}(u,v)$ to $\tilde{\boldsymbol{\sigma}}(\sinh u, v)$ is a local isometry. Which curves on the helicoid correspond under this isometry to the parallels and meridians of the catenoid?
In fact, there is an *isometric deformation* of the catenoid into a helicoid. Let

$$\hat{\boldsymbol{\sigma}}(u,v) = (-\sinh u \sin v, \sinh u \cos v, -v).$$

This is the result of reflecting the helicoid $\tilde{\boldsymbol{\sigma}}$ in the $xy$-plane and then translating it by $\pi/2$ parallel to the $z$-axis. Define

$$\boldsymbol{\sigma}^t(u,v) = \cos t\, \boldsymbol{\sigma}(u,v) + \sin t\, \hat{\boldsymbol{\sigma}}(u,v),$$

so that $\boldsymbol{\sigma}^0(u,v) = \boldsymbol{\sigma}(u,v)$ and $\boldsymbol{\sigma}^{\pi/2}(u,v) = \hat{\boldsymbol{\sigma}}(u,v)$. Show that, for all values of $t$, the map $\boldsymbol{\sigma}(u,v) \mapsto \boldsymbol{\sigma}^t(u,v)$ is a local isometry. Show also that the tangent plane of $\boldsymbol{\sigma}^t$ at the point $\boldsymbol{\sigma}^t(u,v)$ depends only on $u, v$ and not on $t$. The surfaces $\boldsymbol{\sigma}^t$ are shown above for several values of $t$. (The result of this exercise is 'explained' in Exercises 12.5.3 and 12.5.4.)

6.2.4 Show that the line of striction (Exercise 5.3.4) of the tangent developable of a unit-speed curve $\boldsymbol{\gamma}$ is $\boldsymbol{\gamma}$ itself. Show also that the intersection of this surface with the plane passing through a point $\boldsymbol{\gamma}(u_0)$ of the curve and perpendicular to it at that point is a curve of the form

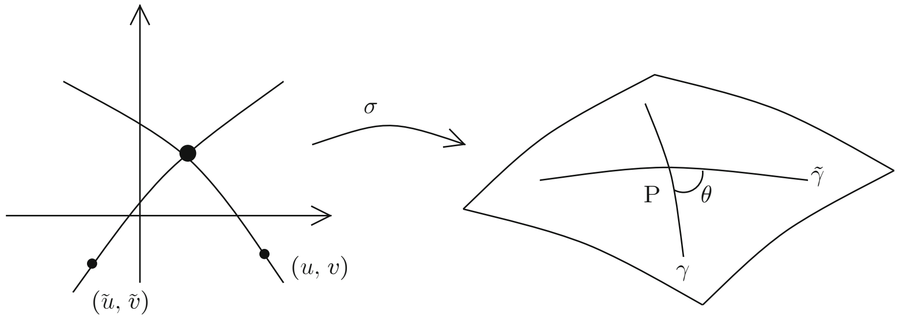
$$\boldsymbol{\Gamma}(v) = \boldsymbol{\gamma}(u_0) - \frac{1}{2}\kappa(u_0)v^2\mathbf{n}(u_0) + \frac{1}{3}\kappa(u_0)\tau(u_0)v^3\mathbf{b}(u_0)$$

if we neglect higher powers of $v$ (we assume that the curvature $\kappa(u_0)$ and the torsion $\tau(u_0)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(u_0)$ are both non-zero). Note that this curve has an ordinary cusp (Exercise 1.3.3) at $\boldsymbol{\gamma}(u_0)$, so the tangent developable has a sharp 'edge' where the two sheets $v > 0$ and $v < 0$ meet along $\boldsymbol{\gamma}$. This is evident for the tangent developable of a circular helix illustrated earlier in this section.

## 6.3 Conformal mappings of surfaces

Now that we understand how to measure lengths of curves on surfaces, it is natural to ask about angles. Suppose that two curves $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ on a surface $\mathcal{S}$ intersect at a point $\mathbf{p}$. The *angle* $\theta$ of intersection of $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ at $\mathbf{p}$ is defined to be the angle between the tangent vectors $\dot{\boldsymbol{\gamma}}$ and $\dot{\tilde{\boldsymbol{\gamma}}}$ (evaluated at $t = t_0$ and $t = \tilde{t}_0$, respectively). Using the dot product formula for the angle between vectors, we see that $\theta$ is given by

$$\cos\theta = \frac{\dot{\boldsymbol{\gamma}} \cdot \dot{\tilde{\boldsymbol{\gamma}}}}{\| \dot{\boldsymbol{\gamma}} \| \| \dot{\tilde{\boldsymbol{\gamma}}} \|} = \frac{\langle \dot{\boldsymbol{\gamma}}, \dot{\tilde{\boldsymbol{\gamma}}} \rangle}{\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle^{1/2} \langle \dot{\tilde{\boldsymbol{\gamma}}}, \dot{\boldsymbol{\gamma}} \rangle^{1/2}}. \tag{6.6}$$




As usual, it will be useful to have an expression for this in terms of a surface patch. Suppose then that $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ lie in a surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$, so that $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ and $\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}(\tilde{u}(t), \tilde{v}(t))$ for some smooth functions $u, v, \tilde{u}$ and $\tilde{v}$. If $Edu^2 + 2Fdudv + Gdv^2$ is the first fundamental form of $\boldsymbol{\sigma}$, then by ( 6.6) we have

$$\cos\theta = \frac{E\dot{u}\dot{\tilde{u}} + F(\dot{u}\dot{\tilde{v}} + \dot{\tilde{u}}\dot{v}) + G\dot{v}\dot{\tilde{v}}}{(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{1/2}(E\dot{\tilde{u}}^2 + 2F\dot{\tilde{u}}\dot{\tilde{v}} + G\dot{\tilde{v}}^2)^{1/2}}. \tag{6.7}$$

### Example 6.3.1

The *parameter curves* on a surface patch $\boldsymbol{\sigma}(u, v)$ can be parametrized by

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u_0, t), \quad \tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}(t, v_0),$$

respectively, where $u_0$ is the constant value of $u$ and $v_0$ is the constant value of $v$ in the two cases. Thus,

$$\begin{aligned} u(t) &= u_0, \quad v(t) = t, \qquad \tilde{u}(t) = t, \quad \tilde{v}(t) = v_0, \\ \therefore \quad \dot{u} &= 0, \quad \dot{v} = 1, \qquad\quad \dot{\tilde{u}} = 1, \quad \dot{\tilde{v}} = 0. \end{aligned}$$

These parameter curves intersect at the point $\boldsymbol{\sigma}(u_0, v_0)$ of the surface. By Eq. 6.7, their angle of intersection $\theta$ is given by

$$\cos\theta = \frac{F}{\sqrt{EG}},$$

where $E, F$ and $G$ are evaluated at $(u_0, v_0)$. In particular, the parameter curves are orthogonal if and only if $F = 0$.

Corresponding to the Definition 6.2.1 of a local isometry, we have the following definition.

## Definition 6.3.2

If $\mathcal{S}_1$ and $\mathcal{S}_2$ are surfaces, a *conformal map* $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a local diffeomorphism such that, if $\boldsymbol{\gamma}_1$ and $\tilde{\boldsymbol{\gamma}}_1$ are any two curves on $\mathcal{S}_1$ that intersect, say at a point $\mathbf{p} \in \mathcal{S}_1$, and if $\boldsymbol{\gamma}_2$ and $\tilde{\boldsymbol{\gamma}}_2$ are their images under $f$, the angle of intersection of $\boldsymbol{\gamma}_1$ and $\tilde{\boldsymbol{\gamma}}_1$ at $\mathbf{p}$ is equal to the angle of intersection of $\boldsymbol{\gamma}_2$ and $\tilde{\boldsymbol{\gamma}}_2$ at $f(\mathbf{p})$.

In short, *$f$ is conformal if and only if it preserves angles.* The reason this definition requires $f$ to be a local diffeomorphism is contained in Exercise 4.4.4 – note that the angle between two intersecting curves is well defined only when both curves are regular.

It is obvious that any composite of conformal maps is conformal, and that the inverse of any conformal diffeomorphism is conformal.

As a special case, if $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a surface, then $\boldsymbol{\sigma}$ may be viewed as a map from an open subset of the plane (namely $U$), parametrized by $(u, v)$ in the usual way, and the image $\mathcal{S}$ of $\boldsymbol{\sigma}$, and we say that $\boldsymbol{\sigma}$ is a *conformal parametrization* or a *conformal surface patch* of $\mathcal{S}$ if this map between surfaces is conformal.

## Theorem 6.3.3

A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is conformal if and only if there is a function $\lambda : \mathcal{S}_1 \to \mathbb{R}$ such that

$$f^*\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p}} = \lambda(\mathbf{p}) \langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{p}} \quad \text{for all } \mathbf{p} \in \mathcal{S}_1 \text{ and } \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}_1.$$

It is not hard to see that the function $\lambda$, if it exists, is necessarily smooth.

## Proof

Let $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ be two curves on $\mathcal{S}_1$ that intersect at a point $\mathbf{p} \in \mathcal{S}_1$. The angle $\theta$ of intersection of the curves is given by Eq. 6.6. The corresponding angle of intersection of the curves $f \circ \gamma$ and $f \circ \tilde{\gamma}$ on $\mathcal{S}_2$ is obtained from the expression on the right-hand side of Eq. 6.6 by replacing $\dot{\gamma}$ and $\dot{\tilde{\gamma}}$ with $(f \circ \gamma)^{\cdot}$ and $(f \circ \tilde{\gamma})^{\cdot}$, respectively. Now,

$$\langle (f \circ \gamma)^{\cdot}, (f \circ \tilde{\gamma})^{\cdot} \rangle_{f(\mathbf{p})} = \langle D_{\mathbf{p}} f(\dot{\gamma}), D_{\mathbf{p}} f(\dot{\tilde{\gamma}}) \rangle_{f(\mathbf{p})} = f^* \langle \dot{\gamma}, \dot{\tilde{\gamma}} \rangle_{\mathbf{p}},$$

with similar expressions for $\langle (f \circ \gamma)^{\cdot}, f \circ \gamma)^{\cdot} \rangle_{f(\mathbf{p})}$ and $\langle (f \circ \tilde{\gamma})^{\cdot}, f \circ \tilde{\gamma})^{\cdot} \rangle_{f(\mathbf{p})}$. Thus, to compute the angle of intersection of the curves $f \circ \gamma$ and $f \circ \tilde{\gamma}$ on $\mathcal{S}_2$, we must replace $\langle\ ,\ \rangle$ in the numerator and denominator of the expression on the right-hand side of Eq. 6.6 by $f^* \langle\ ,\ \rangle$. It is now clear that, if $f^* \langle\ ,\ \rangle = \lambda \langle\ ,\ \rangle$, this replacement leaves the expression in Eq. 6.6 unchanged (since the factor $\lambda$ cancels out) and so $f$ is conformal.

For the converse, we must show that if

$$\frac{\langle \dot{\gamma}, \dot{\tilde{\gamma}} \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2} \langle \dot{\tilde{\gamma}}, \dot{\tilde{\gamma}} \rangle^{1/2}} = \frac{f^* \langle \dot{\gamma}, \dot{\tilde{\gamma}} \rangle}{f^* \langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2} f^* \langle \dot{\tilde{\gamma}}, \dot{\tilde{\gamma}} \rangle^{1/2}} \tag{6.8}$$

for all pairs of intersecting curves $\gamma$ and $\tilde{\gamma}$ on $\mathcal{S}_1$, then $f^* \langle\ ,\ \rangle$ is proportional to $\langle\ ,\ \rangle$. Since every tangent vector to $\mathcal{S}_1$ is the tangent vector of a curve on $\mathcal{S}_1$, Eq. 6.8 implies that

$$\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle^{1/2} \langle \mathbf{w}, \mathbf{w} \rangle^{1/2}} = \frac{f^* \langle \mathbf{v}, \mathbf{w} \rangle}{f^* \langle \mathbf{v}, \mathbf{v} \rangle^{1/2} f^* \langle \mathbf{w}, \mathbf{w} \rangle^{1/2}} \tag{6.9}$$

for all tangent vectors $\mathbf{v}, \mathbf{w}$ to $\mathcal{S}_1$.

Choose an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ of the tangent plane to $\mathcal{S}_1$ with respect to its first fundamental form $\langle\ ,\ \rangle$. Let

$$\lambda = f^* \langle \mathbf{v}_1, \mathbf{v}_1 \rangle, \quad \mu = f^* \langle \mathbf{v}_1, \mathbf{v}_2 \rangle, \quad \nu = f^* \langle \mathbf{v}_2, \mathbf{v}_2 \rangle.$$

We apply Eq. 6.9 with $\mathbf{v} = \mathbf{v}_1$ and $\mathbf{w} = \cos\theta\, \mathbf{v}_1 + \sin\theta\, \mathbf{v}_2$, where $\theta \in \mathbb{R}$. This gives

$$\cos\theta = \frac{\lambda \cos\theta + \mu \sin\theta}{\sqrt{\lambda(\lambda \cos^2\theta + 2\mu \sin\theta \cos\theta + \nu \sin^2\theta)}}.$$

Taking $\theta = \pi/2$ gives $\mu = 0$, which implies that

$$\lambda = \lambda \cos^2\theta + \nu \sin^2\theta \quad \text{for all } \theta \in \mathbb{R}.$$

Hence, $\lambda = \nu$. This implies that $f^* \langle \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle$ whenever $\mathbf{v}$ and $\mathbf{w}$ are basis vectors. Since both sides are bilinear forms, it follows that $f^* \langle\ ,\ \rangle = \lambda \langle\ ,\ \rangle$. $\square$

Reinterpreting this result in terms of surface patches gives

## Corollary 6.3.4

A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is conformal if and only if, for any surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}_1$, the first fundamental forms of the patches $\boldsymbol{\sigma}$ of $\mathcal{S}_1$ and $f \circ \boldsymbol{\sigma}$ of $\mathcal{S}_2$ are proportional.

In particular, a surface patch $\boldsymbol{\sigma}(u, v)$ is conformal if and only if its first fundamental form is $\lambda(du^2 + dv^2)$ for some smooth function $\lambda(u, v)$.

## Example 6.3.5

We consider the unit sphere $S^2$. If $\mathbf{q}$ is any point of $S^2$ other than the north pole $\mathbf{n} = (0, 0, 1)$, the straight line joining $\mathbf{n}$ and $\mathbf{q}$ intersects the $xy$-plane at some point $\mathbf{p}$, say. The map that takes $\mathbf{q}$ to $\mathbf{p}$ is called *stereographic projection* from $S^2$ to the plane, and we denote it by $\Pi$. We are going to show that $\Pi$ is conformal.

Let $\mathbf{p} = (u, v, 0)$, $\mathbf{q} = (x, y, z)$. Since $\mathbf{p}, \mathbf{q}, \mathbf{n}$ lie on a straight line, there is a scalar $\rho$ such that

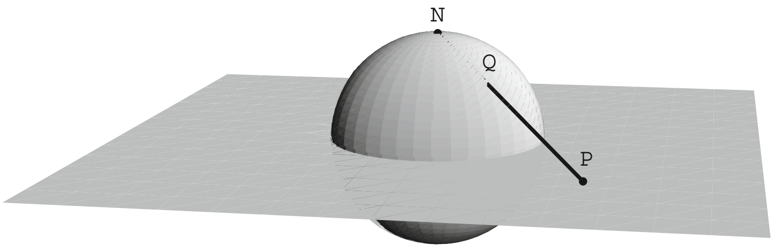
$$\mathbf{q} - \mathbf{n} = \rho(\mathbf{p} - \mathbf{n}),$$

and hence

$$(x, y, z) = (0, 0, 1) + \rho((u, v, 0) - (0, 0, 1)) = (\rho u, \rho v, 1 - \rho). \tag{6.10}$$

Hence, $\rho = 1 - z$, $u = x/(1 - z)$, $v = y/(1 - z)$ and we have

$$\Pi(x, y, z) = \left(\frac{x}{1-z}, \frac{y}{1-z}, 0\right).$$




On the other hand, from Eq. 6.10 and $x^2 + y^2 + z^2 = 1$ we get $\rho = 2/(u^2 + v^2 + 1)$ and hence

$$\mathbf{q} = \left(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, \frac{u^2+v^2-1}{u^2+v^2+1}\right).$$

If we denote the right-hand side by $\boldsymbol{\sigma}_1(u, v)$, then $\boldsymbol{\sigma}_1$ is a parametrization of $S^2$ with the north pole removed. Parametrizing the $xy$-plane by $\boldsymbol{\sigma}_2(u, v) = (u, v, 0)$, we then have

$$\Pi(\boldsymbol{\sigma}_1(u, v)) = \boldsymbol{\sigma}_2(u, v).$$

According to Corollary 6.3.4, to show that $\Pi$ is conformal we have to show that the first fundamental forms of $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are proportional. The first fundamental form of $\boldsymbol{\sigma}_2$ is $du^2 + dv^2$. As to $\boldsymbol{\sigma}_1$, we get

$$\begin{aligned}
(\boldsymbol{\sigma}_1)_u &= \left(\frac{2(v^2 - u^2 + 1)}{(u^2 + v^2 + 1)^2}, \frac{-4uv}{(u^2 + v^2 + 1)^2}, \frac{4u}{(u^2 + v^2 + 1)^2}\right), \\
(\boldsymbol{\sigma}_1)_v &= \left(\frac{-4uv}{(u^2 + v^2 + 1)^2}, \frac{2(u^2 - v^2 + 1)}{(u^2 + v^2 + 1)^2}, \frac{4v}{(u^2 + v^2 + 1)^2}\right).
\end{aligned}$$

This gives

$$E_1 = (\boldsymbol{\sigma}_1)_u \cdot (\boldsymbol{\sigma}_1)_u = \frac{4(v^2 - u^2 + 1)^2 + 16u^2v^2 + 16u^2}{(u^2 + v^2 + 1)^4} = \frac{4}{(u^2 + v^2 + 1)^2}.$$

Similarly, $F_1 = 0$, $G_1 = 4/(u^2 + v^2 + 1)^2$. Thus, the first fundamental form of $\boldsymbol{\sigma}_2$ is $\lambda$ times that of $\boldsymbol{\sigma}_1$, where $\lambda = \frac{1}{4}(u^2 + v^2 + 1)^2$.

It is often useful to think of $\Pi$ as a map to the complex numbers $\mathbb{C}$ rather than to the $xy$-plane, by identifying $u + iv \in \mathbb{C}$ with $(u, v, 0)$. Moreover, we can parametrize $S^2$ itself in a partly complex way by identifying $(x, y, z) \in S^2$ with $(x + iy, z)$. Then, $S^2$ becomes the set of pairs $(w, z)$ where $w \in \mathbb{C}$, $z \in \mathbb{R}$ and $|w|^2 + z^2 = 1$. Stereographic projection then takes the simple form

$$\Pi(w, z) = \frac{w}{1 - z},$$

and the surface patch $\boldsymbol{\sigma}_1$ is given by

$$\boldsymbol{\sigma}_1(w) = \left(\frac{2w}{|w|^2 + 1}, \frac{|w|^2 - 1}{|w|^2 + 1}\right).$$

The inconvenience of having to exclude the north pole from the domain of definition of $\Pi$ can be overcome by introducing a 'point at infinity' $\infty$ and defining the 'extended complex plane' $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$. If we agree that $\Pi$ maps the north pole to $\infty$, it defines a *bijection* $\Pi : S^2 \to \mathbb{C}_\infty$. Further discussion of this map is left to the exercises.

Returning now to the general case, it is natural to ask when there is a conformal map between two surfaces. The surprising answer is that this is *always the case locally*: if $\mathbf{p}_1$ and $\mathbf{p}_2$ are points of two surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, there are open subsets $\mathcal{O}_1$ of $\mathcal{S}_1$ containing $\mathbf{p}_1$ and $\mathcal{O}_2$ of $\mathcal{S}_2$ containing $\mathbf{p}_2$ and a conformal diffeomorphism $\mathcal{O}_1 \to \mathcal{O}_2$. This follows from the following theorem:

### Theorem 6.3.6

Every surface has an atlas consisting of conformal surface patches.

Indeed, if $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are conformal parametrizations of $\mathcal{S}_1$ and $\mathcal{S}_2$, the map $\boldsymbol{\sigma}_1(u,v) \mapsto \boldsymbol{\sigma}_2(u,v)$ will be conformal as it is the composite of the conformal diffeomorphism $\boldsymbol{\sigma}_2$ and the inverse of the conformal diffeomorphism $\boldsymbol{\sigma}_1$.

We shall prove a special case of Theorem 6.3.6 later (see Theorem 12.4.1), but the general case is beyond the scope of this book.

## *EXERCISES*

6.3.1 Show that every local isometry is conformal. Give an example of a conformal map that is not a local isometry.

6.3.2 Show that *Enneper's surface*

$$\boldsymbol{\sigma}(u,v) = \left(u - \frac{u^3}{3} + uv^2, v - \frac{v^3}{3} + vu^2, u^2 - v^2\right)$$

is conformally parametrized.

6.3.3 Recall from Example 6.1.3 that the first fundamental form of the latitude–longitude parametrization $\boldsymbol{\sigma}(\theta,\varphi)$ of $S^2$ is

$$d\theta^2 + \cos^2\theta\, d\varphi^2.$$

Find a smooth function $\psi$ such that the reparametrization $\tilde{\boldsymbol{\sigma}}(u,v) = \boldsymbol{\sigma}(\psi(u),v)$ is conformal. Verify that $\tilde{\boldsymbol{\sigma}}$ is, in fact, the Mercator parametrization in Exercise 5.3.2.

6.3.4 Let $\Phi : U \to V$ be a diffeomorphism between open subsets of $\mathbb{R}^2$. Write

$$\Phi(u,v) = (f(u,v), g(u,v)),$$

where $f$ and $g$ are smooth functions on the $uv$-plane. Show that $\Phi$ is conformal if and only if

$$\text{either} \quad (f_u = g_v \quad \text{and} \quad f_v = -g_u) \quad \text{or} \quad (f_u = -g_v \quad \text{and} \quad f_v = g_u). \tag{6.11}$$

Show that, if $J(\Phi)$ is the Jacobian matrix of $\Phi$, then $\det(J(\Phi)) > 0$ in the first case and $\det(J(\Phi)) < 0$ in the second case.

6.3.5 (This exercise requires a basic knowledge of complex analysis.) Recall that the transition map between two surface patches in an atlas for a surface $\mathcal{S}$ is a smooth map between open subsets of $\mathbb{R}^2$. Since

$\mathbb{R}^2$ is the 'same' as the complex numbers $\mathbb{C}$ (via $(u, v) \leftrightarrow u + iv$), we can ask whether such a transition map is *holomorphic*. One says that $\mathcal{S}$ is a *Riemann surface* if $\mathcal{S}$ has an atlas for which all the transition maps are holomorphic. Deduce from Theorem 6.3.6 and the preceding exercise that every orientable surface has an atlas making it a Riemann surface. (You will need to recall from complex analysis that a smooth function $\Phi$ as in the preceding exercise is holomorphic if and only if the first pair of equations in (6.11) hold – these are the *Cauchy–Riemann equations*. If the second pair of equations in (6.11) hold, $\Phi$ is said to be *anti-holomorphic*.)

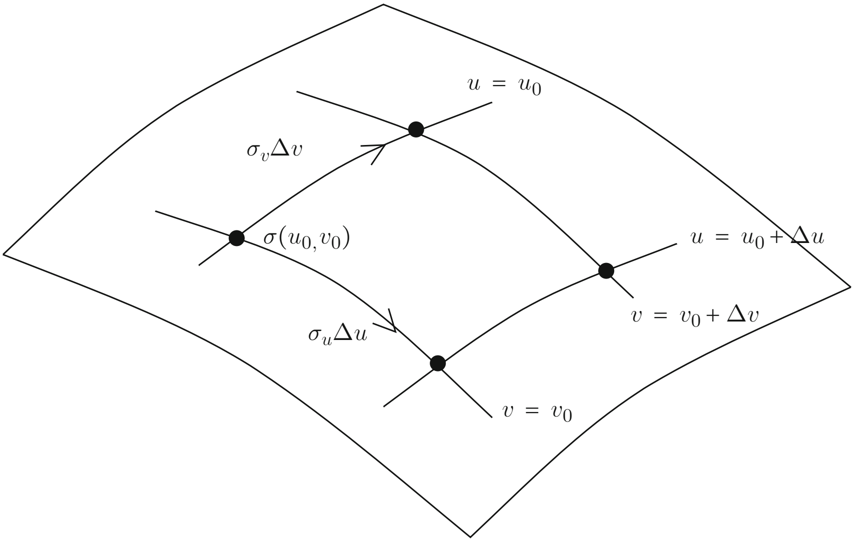
6.3.6 Define a map $\tilde{\Pi}$ similar to $\Pi$ by projecting from the *south* pole of $S^2$ onto the $xy$-plane. Show that this defines a second conformal surface patch $\tilde{\boldsymbol{\sigma}}_1$, which covers the whole of $S^2$ except the south pole. What is the transition map between these two patches? Why do the two patches $\boldsymbol{\sigma}_1$ and $\tilde{\boldsymbol{\sigma}}_1$ *not* give $S^2$ the structure of a Riemann surface? How can $\tilde{\boldsymbol{\sigma}}_1$ be modified to produce such a structure?

6.3.7 Show that the stereographic projection map $\Pi$ takes circles on $S^2$ to Circles in $\mathbb{C}_\infty$, and that every Circle arises in this way. (A circle on $S^2$ is the intersection of $S^2$ with a plane; a Circle in $\mathbb{C}_\infty$ is a line or a circle in $\mathbb{C}$ – see Appendix 2).

6.3.8 Show that, if $M$ is a Möbius transformation or a conjugate-Möbius transformation (see Appendix 2), the bijection $\Pi^{-1} \circ M \circ \Pi : S^2 \to S^2$ is a conformal diffeomorphism of $S^2$. It can be shown that every conformal diffeomorphism of $S^2$ is of this type.

## 6.4 Equiareal maps and a theorem of Archimedes

Suppose that $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a surface patch on a surface $\mathcal{S}$. The image of $\boldsymbol{\sigma}$ is covered by the two families of parameter curves obtained by setting $u =$ constant and $v =$ constant, respectively. Fix $(u_0, v_0) \in U$; since the change in $\boldsymbol{\sigma}(u, v)$ corresponding to a small change $\Delta u$ in $u$ is approximately $\boldsymbol{\sigma}_u \Delta u$ and that corresponding to a small change $\Delta v$ in $v$ is approximately $\boldsymbol{\sigma}_v \Delta v$, the part of the surface contained by the parameter curves on the surface corresponding to $u = u_0$, $u = u_0 + \Delta u$, $v = v_0$ and $v = v_0 + \Delta v$ is approximately a parallelogram in the plane with sides given by the vectors $\boldsymbol{\sigma}_u \Delta u$ and $\boldsymbol{\sigma}_v \Delta v$ (the derivatives being evaluated at $(u_0, v_0)$):




Recalling that the area of a parallelogram in the plane with sides $\mathbf{a}$ and $\mathbf{b}$ is $\parallel \mathbf{a} \times \mathbf{b} \parallel$, we see that the area of the parallelogram on the surface is approximately

$$\parallel \boldsymbol{\sigma}_u \Delta u \times \boldsymbol{\sigma}_v \Delta v \parallel = \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel \Delta u \Delta v.$$

This suggests the following definition.

## Definition 6.4.1

The *area* $\mathcal{A}_{\boldsymbol{\sigma}}(R)$ of the part $\boldsymbol{\sigma}(R)$ of a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ corresponding to a region $R \subseteq U$ is

$$\mathcal{A}_{\boldsymbol{\sigma}}(R) = \int_R \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel dudv.$$

Of course, this integral may be infinite – think of the area of a whole plane, for example. However, the integral will be finite if, say, $R$ is contained in a rectangle that is entirely contained, along with its boundary, in $U$.

The quantity $\parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel$ that appears in the definition of area is easily computed in terms of the first fundamental form $Edu^2 + 2Fdudv + Gdv^2$ of $\boldsymbol{\sigma}$:

## Proposition 6.4.2

$\parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel = (EG - F^2)^{1/2}.$

### Proof

We use a result from vector algebra: if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{d}$ are vectors in $\mathbb{R}^3$,

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}).$$

Applying this to $\| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \|^2 = (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) \cdot (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)$, we get

$$\| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \|^2 = (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u)(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v) - (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)^2 = EG - F^2. \qquad \square$$

Note that, for a regular surface, $EG - F^2 > 0$ everywhere, since for a regular surface $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$ is never zero.

Thus, our definition of area is

$$\mathcal{A}_{\boldsymbol{\sigma}}(R) = \int_R (EG - F^2)^{1/2} du dv. \tag{6.12}$$

We sometimes denote $(EG - F^2)^{1/2} du dv$ by $d\mathcal{A}_{\boldsymbol{\sigma}}$. But we have still to check that this definition is sensible, i.e., it is unchanged if $\boldsymbol{\sigma}$ is reparametrized. This is certainly not obvious, since $E$, $F$ and $G$ change under reparametrization (see Exercise 6.1.4).

### Proposition 6.4.3

The area of a surface patch is unchanged by reparametrization.

### Proof

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a surface patch and let $\tilde{\boldsymbol{\sigma}} : \tilde{U} \to \mathbb{R}^3$ be a reparametrization of $\boldsymbol{\sigma}$, with reparametrization map $\Phi : \tilde{U} \to U$. Thus, if $\Phi(\tilde{u}, \tilde{v}) = (u, v)$, we have

$$\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v}) = \boldsymbol{\sigma}(u, v).$$

Let $\tilde{R} \subseteq \tilde{U}$ be a region, and let $R = \Phi(\tilde{R}) \subseteq U$. We have to prove that

$$\int_R \| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \| \, du dv = \int_{\tilde{R}} \| \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \| \, d\tilde{u} d\tilde{v}.$$

We showed in the proof of Proposition 4.2.7 that

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} = \det(J(\Phi)) \, \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v,$$

where $J(\Phi)$ is the Jacobian matrix of $\Phi$. Hence,

$$\int_{\tilde{R}} \| \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \| \; d\tilde{u}d\tilde{v} = \int_{\tilde{R}} |\det(J(\Phi))| \, \| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \| \; d\tilde{u}d\tilde{v}.$$

By the change of variables formula for double integrals, the right-hand side of this equation is exactly

$$\int_{R} \| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \| \; dudv. \qquad \square$$

Now that we have a good definition of area, we can ask which maps between surfaces are area-preserving.

## Definition 6.4.4

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two surfaces. A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is said to be *equiareal* if it takes any region in $\mathcal{S}_1$ to a region of *the same area* in $\mathcal{S}_2$ (we assume that each of the regions is sufficiently small, so that it is contained in the image of some surface patch).

We have the following analogue of Theorem 6.2.2.

## Theorem 6.4.5

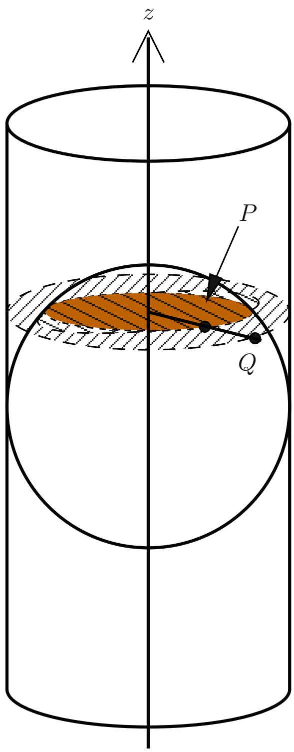
A local diffeomorphism $f : \mathcal{S}_1 \to \mathcal{S}_2$ is equiareal if and only if, for any surface patch $\boldsymbol{\sigma}(u, v)$ on $\mathcal{S}_1$, the first fundamental forms

$$E_1 du^2 + 2F_1 dudv + G_1 dv^2 \quad \text{and} \quad E_2 du^2 + 2F_2 dudv + G_2 dv^2$$

of the patches $\boldsymbol{\sigma}$ on $\mathcal{S}_1$ and $f \circ \boldsymbol{\sigma}$ on $\mathcal{S}_2$ satisfy

$$E_1 G_1 - F_1^2 = E_2 G_2 - F_2^2. \tag{6.13}$$

The proof is very similar to that of Theorem 6.2.2 and we leave it as Exercise 6.4.6. As with isometries and conformal maps, it is obvious that any composite of equaireal diffeomorphism is equiareal, and that the inverse of any equiareal difeomorphism is equaireal.




One of the most famous examples of an equiareal map was found by Archimedes. Legend has it that the discovery was inscribed onto his tombstone by the Roman general Marcellus who led the siege of Syracuse in which Archimedes perished. Naturally, since calculus was not available to him, Archimedes' proof of his theorem was quite different from ours.

Consider the unit sphere $x^2+y^2+z^2 = 1$ and the unit cylinder $x^2+y^2 = 1$. The sphere is contained inside the cylinder, and the two surfaces touch along the circle $x^2 + y^2 = 1$ in the $xy$-plane. For each point $\mathbf{p} \in S^2$ other than the poles $(0, 0, \pm 1)$, there is a unique straight line parallel to the $xy$-plane and passing through the point $\mathbf{p}$ and the $z$-axis. This line intersects the cylinder in two points, one of which, say $\mathbf{q}$, is closest to $\mathbf{p}$. Let $f$ be the map from $S^2$ (with the two poles removed) to the cylinder that takes $\mathbf{p}$ to $\mathbf{q}$.

To find a formula for $f$, let $(x, y, z)$ be the Cartesian coordinates of $\mathbf{p}$, and $(X, Y, Z)$ those of $\mathbf{q}$. Since the line through $\mathbf{p}$ and $\mathbf{q}$ is parallel to the $xy$-plane, we have $Z = z$ and $(X, Y) = \lambda(x, y)$ for some scalar $\lambda$. Since $(X, Y, Z)$ is on the cylinder,

$$1 = X^2 + Y^2 = \lambda^2(x^2 + y^2),$$

$$\therefore \quad \lambda = \pm(x^2+y^2)^{-1/2}.$$

Taking the $+$ sign gives the point $\mathbf{q}$, so we get

$$f(x, y, z) = \left( \frac{x}{(x^2+y^2)^{1/2}}, \frac{y}{(x^2+y^2)^{1/2}}, z \right).$$

We shall show in the proof of the next theorem that $f$ is a diffeomorphism.

## Theorem 6.4.6 (Archimedes' Theorem)

The map $f$ is an equiareal diffeomorphism.

## Proof

We take the atlas for the surface $\mathcal{S}_1$ consisting of the sphere minus the north and south poles with two patches, both given by the formula

$$\boldsymbol{\sigma}_1(\theta, \varphi) = (\cos\theta\cos\varphi, \cos\theta\sin\varphi, \sin\theta),$$

and defined on the open sets

$$\{-\pi/2 < \theta < \pi/2,\ 0 < \varphi < 2\pi\} \text{ and } \{-\pi/2 < \theta < \pi/2,\ -\pi < \varphi < \pi\}.$$

The image of $\boldsymbol{\sigma}_1(\theta, \varphi)$ under the map $f$ is the point

$$\boldsymbol{\sigma}_2(\theta, \varphi) = (\cos\varphi, \sin\varphi, \sin\theta) \tag{6.14}$$

of the cylinder. It is easy to check that this gives an atlas for the surface $\mathcal{S}_2$, consisting of the part of the cylinder between the planes $z = 1$ and $z = -1$, with two patches, both given by Eq. 6.14 and defined on the same two open sets as $\boldsymbol{\sigma}_1$. We have to show that Eq. 6.13 holds.

We computed the coefficients $E_1$, $F_1$ and $G_1$ of the first fundamental form of $\boldsymbol{\sigma}_1$ in Example 6.1.3:

$$E_1 = 1, \quad F_1 = 0, \quad G_1 = \cos^2\theta.$$

For $\boldsymbol{\sigma}_2$, we get $(\boldsymbol{\sigma}_2)_\theta = (0, 0, \cos\theta)$, $(\boldsymbol{\sigma}_2)_\varphi = (-\sin\varphi, \cos\varphi, 0)$, and so

$$E_2 = \cos^2\theta, \quad F_2 = 0, \quad G_2 = 1.$$

It is now clear that Eq. 6.13 holds.

Note that, since $f$ corresponds simply to the identity map $(\theta, \varphi) \mapsto (\theta, \varphi)$ in terms of the parametrizations $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ of the unit sphere and cylinder, respectively, it follows that $f$ is a diffeomorphism. $\square$

The following classical result provides a beautiful application of Archimedes' theorem. A *spherical triangle* is a triangle on a sphere whose sides are arcs of great circles.

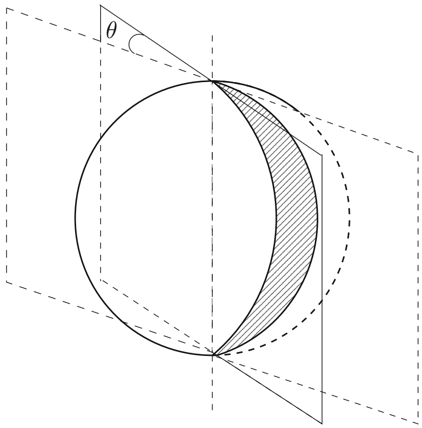
## Theorem 6.4.7

The area of a spherical triangle on the unit sphere $S^2$ with internal angles $\alpha$, $\beta$ and $\gamma$ is

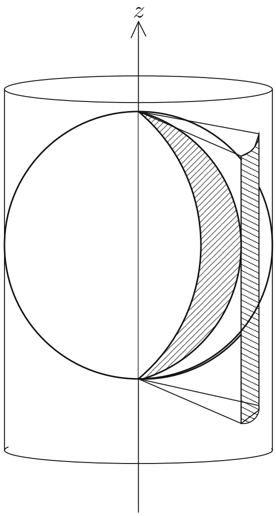
$$\alpha + \beta + \gamma - \pi.$$

## Proof

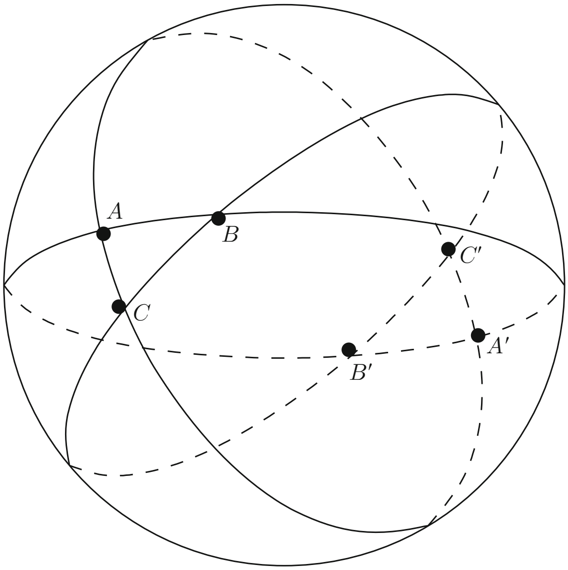
We begin by using Archimedes' Theorem 6.4.6 to compute the area of a 'lune', i.e., the area enclosed between two great circles:




We can assume that the great circles intersect at the poles, since this can be achieved by applying a rotation of $S^2$, and this does not change areas. If $\theta$ is the angle between them, the image of the lune under the map $f$ is a curved rectangle on the cylinder of width $\theta$ and height 2 (see next page). If we now apply the isometry which unwraps the cylinder on the plane, this curved rectangle on the cylinder will map to a genuine rectangle on the plane, with width $\theta$ and height 2. By Archimedes' theorem, the lune has the same area as the curved rectangle on the cylinder, and since every isometry is an equiareal map (see Exercise 6.4.6), this has the same area as the genuine rectangle in the plane, namely $2\theta$. Note that this correctly gives the area of the whole sphere to be $4\pi$.




Turning now to the proof of the theorem, let $A$, $B$ and $C$ be the vertices of the triangle (so that $\alpha$ is the angle at $A$, etc.). The three great circles, of which the sides of the triangle are arcs, divide $S^2$ into eight triangles, as shown in the following diagram (in which $A'$ is the antipodal point of $A$, etc.).




Note that the two triangles with vertices $A, B, C$ and $A', B, C$ together form a lune with angle $\alpha$, etc. Hence, denoting the triangle with vertices $A, B, C$ by $ABC$ and its area by $\mathcal{A}(ABC)$, etc., we have, by the preceding calculation,

$$
\begin{aligned}
\mathcal{A}(ABC) + \mathcal{A}(A'BC) &= 2\alpha, \\
\mathcal{A}(ABC) + \mathcal{A}(AB'C) &= 2\beta, \\
\mathcal{A}(ABC) + \mathcal{A}(ABC') &= 2\gamma.
\end{aligned}
$$

Adding these equations, we get

$$2\mathcal{A}(ABC) + \{\mathcal{A}(ABC) + \mathcal{A}(A'BC) + \mathcal{A}(AB'C) + \mathcal{A}(ABC')\} = 2\alpha + 2\beta + 2\gamma. \tag{6.15}$$

Now, the triangles $ABC$, $AB'C$, $AB'C'$ and $ABC'$ together make a hemisphere (namely, the hemisphere containing the vertex $A$ with boundary the great circle passing through $B$ and $C$), so

$$\mathcal{A}(ABC) + \mathcal{A}(AB'C) + \mathcal{A}(AB'C') + \mathcal{A}(ABC') = 2\pi. \tag{6.16}$$

Finally, since the map that takes each point of $S^2$ to its antipodal point is an isometry, and hence equiareal, we have

$$\mathcal{A}(A'BC) = \mathcal{A}(AB'C').$$

Inserting this into Eq. 6.16, we see that the term in $\{\ \}$ on the left-hand side of Eq. 6.15 is equal to $2\pi$. Rearranging now gives the result. $\square$

In Chapter 13, we shall obtain a far-reaching generalization of this result in which $S^2$ is replaced by an arbitrary surface, and great circles by arbitrary curves on the surface.

## *EXERCISES*

6.4.1 Determine the area of the part of the paraboloid $z = x^2 + y^2$ with $z \leq 1$ and compare with the area of the hemisphere $x^2 + y^2 + z^2 = 1$, $z \leq 0$.

6.4.2 A sailor circumnavigates Australia by a route consisting of a triangle whose sides are arcs of great circles. Prove that at least one interior angle of the triangle is $\geq \frac{\pi}{3} + \frac{10}{169}$ radians. (Take the Earth to be a sphere of radius 6,500km and assume that the area of Australia is 7.5 million square kilometres.)

6.4.3 A *spherical polygon* on $S^2$ is the region formed by the intersection of $n$ hemispheres of $S^2$, where $n$ is an integer $\geq 3$. Show that, if $\alpha_1, \ldots, \alpha_n$ are the interior angles of such a polygon, its area is equal to

$$\sum_{i=1}^{n} \alpha_i - (n-2)\pi.$$

6.4.4 Suppose that $S^2$ is covered by spherical polygons such that the intersection of any two polygons is either empty or a common edge or vertex of each polygon. Suppose that there are $F$ polygons, $E$ edges and $V$ vertices (a common edge or vertex of more than one polygon being counted only once). Show that the sum of the angles of all the polygons is $2\pi V$. By using the preceding exercise, deduce that $V - E + F = 2$. (This result is due to Euler; it will be generalized in Chapter 13.)

6.4.5 Show that:

(i) Every local isometry is an equiareal map.

(ii) A map that is both conformal and equiareal is a local isometry.

Give an example of an equiareal map that is not a local isometry.

6.4.6 Prove Theorem 6.4.5.

6.4.7 Let $\boldsymbol{\sigma}(u, v)$ be a surface patch with standard unit normal $\mathbf{N}$. Show that

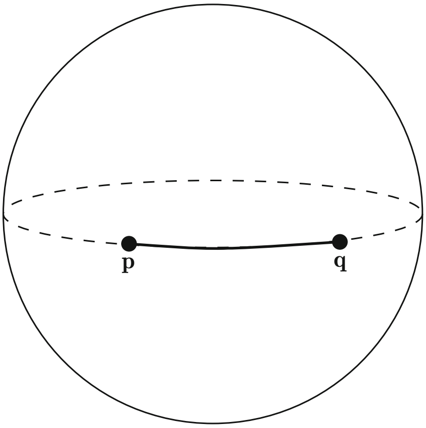
$$\mathbf{N} \times \boldsymbol{\sigma}_u = \frac{E\boldsymbol{\sigma}_v - F\boldsymbol{\sigma}_u}{\sqrt{EG - F^2}}, \quad \mathbf{N} \times \boldsymbol{\sigma}_v = \frac{F\boldsymbol{\sigma}_v - G\boldsymbol{\sigma}_v}{\sqrt{EG - F^2}}.$$

## 6.5 Spherical geometry

We conclude this chapter with a brief discussion of the simplest example of a geometry different from Euclid's, namely spherical geometry. The study of spherical geometry, like that of plane geometry, began in antiquity. Its importance was astronomical: to locate an object in the sky such as a star, one imagines a fixed large sphere centred on the observer; the straight line connecting the observer to the star intersects the sphere in a point whose position gives the direction in which the observer must look in order to see the star. Thus, the three-dimensional universe is projected onto the surface of a sphere. Of course, spherical geometry is also important because we live on the surface of a sphere, to a reasonably good approximation.

If we are to develop spherical geometry by analogy with Euclidean plane geometry, the first thing to do is to decide what should be the analogue of straight lines. Now straight lines are the shortest curves joining any two of their points (Exercise 1.2.4), so it is natural to ask what the corresponding shortest curves are on the sphere. We are going to show that these are arcs of *great circles*.

For simplicity, we work with the unit sphere $S^2$. If $\mathbf{p}$ and $\mathbf{q}$ are two distinct points of $S^2$, there is always at least one great circle passing through $\mathbf{p}$ and $\mathbf{q}$. To see this, note first that if $\mathbf{p}$ and $\mathbf{q}$ are *antipodal* points, i.e., if $\mathbf{p} = -\mathbf{q}$, the intersection of $S^2$ with any plane containing this diameter is a great circle through $\mathbf{p}$ and $\mathbf{q}$. If $\mathbf{p}$ and $\mathbf{q}$ are not antipodal points, the plane passing through the origin perpendicular to the (non-zero) vector $\mathbf{p} \times \mathbf{q}$ intersects $S^2$ in a great circle passing through $\mathbf{p}$ and $\mathbf{q}$. The argument shows, in fact, that if $\mathbf{p}$ and $\mathbf{q}$ are not antipodal there is a unique great circle passing through them both; in this case $\mathbf{p}$ and $\mathbf{q}$ divide this great circle into two circular arcs, one shorter than the other. If $\mathbf{p}$ and $\mathbf{q}$ are antipodal, there are infinitely many great circles passing through both points, each of which is divided by $\mathbf{p}$ and $\mathbf{q}$ into two semicircles (see below).




## Proposition 6.5.1

Let $\mathbf{p}$ and $\mathbf{q}$ be distinct points of $S^2$. If $\mathbf{p} \neq -\mathbf{q}$, the short great circle arc joining $\mathbf{p}$ and $\mathbf{q}$ is the unique curve of shortest length joining $\mathbf{p}$ and $\mathbf{q}$. If $\mathbf{p} = -\mathbf{q}$, any great semicircle joining $\mathbf{p}$ and $\mathbf{q}$ is a shortest curve joining these two points.

## Proof

By using a rotation of $S^2$ (which is an isometry of $S^2$ – see Exercise 6.1.2) we can assume that $\mathbf{p}$ is the north pole $(0, 0, 1)$, and by a further rotation about the $z$-axis we can assume in addition that $\mathbf{q}$ is a point on the great semicircle $\mathcal{C}$ passing through the north and south poles and the point $(1, 0, 0)$, say $(\cos\alpha, 0, \sin\alpha)$, where $-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$. Then the distance from $\mathbf{p}$ to $\mathbf{q}$ measured along the short great circle arc joining them is $\pi/2 - \alpha$.

The first fundamental form of the latitude-longitude parametrization $\boldsymbol{\sigma}(\theta, \varphi)$ is $d\theta^2 + \cos^2\theta\, d\varphi^2$ (Example 6.1.3) so the length of a curve $\boldsymbol{\gamma}(t)$ passing through $\mathbf{p}$ when $t = t_0$ and through $\mathbf{q}$ when $t = t_1$, say, is

$$\int_{t_0}^{t_1} \left(\dot{\theta}^2 + \cos^2\theta\, \dot{\varphi}^2\right)^{1/2} dt.$$

The integrand is not less than $|\dot{\theta}|$, so the length of the part of $\boldsymbol{\gamma}$ between $\mathbf{p}$ and $\mathbf{q}$ is not less than

$$\int_{t_0}^{t_1} |\dot{\theta}|\, dt = \int_{\alpha}^{\pi/2} d\theta = \pi/2 - \alpha,$$

which is the length of the short great circle arc passing through $\mathbf{p}$ and $\mathbf{q}$.

Conversely, if $\boldsymbol{\gamma}$ has exactly this length, we must have

$$\left(\dot{\theta}^2 + \cos^2\theta\, \dot{\varphi}^2\right)^{1/2} = |\dot{\theta}|,$$

and hence

$$\cos\theta\, \dot{\varphi} = 0$$

for all $t$ between $t_0$ and $t_1$. Since $\cos\theta = 0$ only at the north and south poles $(0, 0, \pm 1)$, we must therefore have $\dot{\varphi} = 0$ at all other points of $\boldsymbol{\gamma}$; this means that $\varphi$ is a constant, which must be zero since $\boldsymbol{\gamma}$ passes through $\mathbf{p}$, and so $\boldsymbol{\gamma}$ is part of $\mathcal{C}$. $\square$

Thus, great circles are the spherical analogues of straight lines in Euclidean geometry. One immediate difference between spherical and plane geometry is that *there are no parallel lines in spherical geometry*, for any two great circles intersect (the two planes containing the two great circles intersect in a diameter of $S^2$, the endpoints of which are the points of intersection of the two great circles).

The *spherical distance* $d_{S^2}(\mathbf{p}, \mathbf{q})$ between two points $\mathbf{p}, \mathbf{q} \in S^2$ is the length of the short great circle arc joining $\mathbf{p}$ and $\mathbf{q}$. This is simply the angle between the vectors $\mathbf{p}$ and $\mathbf{q}$ in the range $0 \leq d_{S^2}(\mathbf{p}, \mathbf{q}) \leq \pi$: in symbols,

$$\cos d_{S^2}(\mathbf{p}, \mathbf{q}) = \mathbf{p} \cdot \mathbf{q}.$$

There is a beautiful formula for the spherical distance in terms of the stereographic projection map $\Pi$ (see Example 6.3.5). Recall that $\Pi$ defines a bijection from $S^2$ to the extended complex plane $\mathbb{C}_\infty$; we write $d_{S^2}(\Pi^{-1}(w), \Pi^{-1}(z))$ simply as $d_{S^2}(w, z)$.

## Proposition 6.5.2

If $w, z \in \mathbb{C}$, the spherical distance $d_{S^2}(z, w)$ between the points of $S^2$ corresponding to $w$ and $z$ under stereographic projection is given by

$$\tan \frac{1}{2} d_{S^2}(w, z) = \frac{|w - z|}{|1 + \bar{w} z|}.$$

## Proof

From Example 6.3.5, the point of $S^2$ corresponding to $w \in \mathbb{C}$ is

$$\Pi^{-1}(w) = \left( \frac{w + \bar{w}}{|w|^2 + 1}, \frac{w - \bar{w}}{i(|w|^2 + 1)}, \frac{|w|^2 - 1}{|w|^2 + 1} \right).$$

Hence,

$$\begin{aligned} \cos d_{S^2}(w, z) &= \Pi^{-1}(w) \cdot \Pi^{-1}(z) \\ &= \frac{(w + \bar{w})(z + \bar{z}) - (w - \bar{w})(z - \bar{z}) + (|w|^2 - 1)(|z|^2 - 1)}{(|w|^2 + 1)(|z|^2 + 1)} \\ &= \frac{2(\bar{w} z + w \bar{z}) + (1 - |w|^2)(1 - |z|^2)}{(|w|^2 + 1)(|z|^2 + 1)}. \end{aligned} \tag{6.17}$$

On the other hand, let $t$ denote the right-hand side of the formula in the statement of the proposition. Then,

$$\begin{aligned} \frac{1 - t^2}{1 + t^2} &= \frac{|1 + \bar{w} z|^2 - |w - z|^2}{|1 + \bar{w} z|^2 + |w - z|^2} \\ &= \frac{(1 + \bar{w} z)(1 + w \bar{z}) - (w - z)(\bar{w} - \bar{z})}{(1 + \bar{w} z)(1 + w \bar{z}) + (w - z)(\bar{w} - \bar{z})} \\ &= \frac{2(\bar{w} z + w \bar{z}) + (1 - |w|^2)(1 - |z|^2)}{(|w|^2 + 1)(|z|^2 + 1)}. \end{aligned} \tag{6.18}$$

The proposition follows on comparing Eqs. 6.17 and 6.18 and recalling the identity

$$\cos \theta = \frac{1 - \tan^2 \frac{1}{2} \theta}{1 + \tan^2 \frac{1}{2} \theta}. \qquad \square$$

Much of Euclidean geometry deals with the properties of triangles. We shall always consider only spherical triangles with sides of length less than $\pi$.

### Proposition 6.5.3

Suppose that a spherical triangle has sides of length $A$, $B$ and $C$, and let $\alpha$, $\beta$ and $\gamma$ be its internal angles (so that $\alpha$ is the angle opposite the side of length $A$, etc., and $0 \leq \alpha, \beta, \gamma < \pi$). Then,

$$\text{(i)} \qquad \cos\gamma = \frac{\cos C - \cos A \cos B}{\sin A \sin B},$$

$$\text{(ii)} \qquad \frac{\sin\alpha}{\sin A} = \frac{\sin\beta}{\sin B} = \frac{\sin\gamma}{\sin C}.$$

Two formulas similar to that in (i) can, of course, be obtained by making the cyclic permutations $A \to B \to C \to A$, $\alpha \to \beta \to \gamma \to \alpha$.

Part (i) is called the 'cosine rule' for spherical triangles because it becomes the usual cosine rule when $A, B, C$ are small, in which case the spherical triangle is 'almost' a plane triangle: using the approximations $\cos A = 1 - \frac{1}{2}A^2$ and $\sin A = A$, etc. we get

$$C^2 = A^2 + B^2 - 2AB\cos\gamma.$$

Similarly (ii) reduces to the familiar sine rule for plane triangles when $A$, $B$, $C$ are small.

*Proof 6.5.3* Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be the vertices of the triangle, so that $\alpha$ is the angle at $\mathbf{a}$, etc. Since $A$ is the angle (measured in radians) between the unit vectors $\mathbf{b}$ and $\mathbf{c}$, etc., we have

$$\cos A = \mathbf{b}\cdot\mathbf{c}, \quad \cos B = \mathbf{c}\cdot\mathbf{a}, \quad \cos C = \mathbf{a}\cdot\mathbf{b}. \tag{6.19}$$

Next, the side of the triangle of length $C$ is an arc of the great circle that is the intersection of $S^2$ with the plane $\Pi_C$ through the origin and perpendicular to the vector $\mathbf{a}\times\mathbf{b}$ (and similarly for the other sides). Let $\Pi_{\mathbf{c}}$ be the plane passing through the vertex $\mathbf{c}$ parallel to the tangent plane of $S^2$ there. Then $\Pi_{\mathbf{c}}$ intersects the planes $\Pi_A$ and $\Pi_B$ in two straight lines that are tangent to the sides of the triangle passing through $\mathbf{c}$. It follows that $\gamma$ is the angle between these two lines, which in turn is equal to the angle between $\Pi_A$ and $\Pi_B$, i.e., the angle between $\mathbf{b}\times\mathbf{c}$ and $\mathbf{a}\times\mathbf{c}$:

$$\cos\gamma = \frac{(\mathbf{b}\times\mathbf{c})\cdot(\mathbf{a}\times\mathbf{c})}{\parallel \mathbf{b}\times\mathbf{c} \parallel \parallel \mathbf{a}\times\mathbf{c} \parallel}. \tag{6.20}$$

Of course, there are similar formulas for $\cos\alpha$ and $\cos\beta$.

Now

$$\| \mathbf{b} \times \mathbf{c} \| = \sin A, \quad \| \mathbf{a} \times \mathbf{c} \| = \sin B.$$

On the other hand, the triple product identity (see the proof of Proposition 6.4.2) gives

$$(\mathbf{b} \times \mathbf{c}) \cdot (\mathbf{a} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{b})(\mathbf{c} \cdot \mathbf{c}) - (\mathbf{b} \cdot \mathbf{c})(\mathbf{a} \cdot \mathbf{c}) = \cos C - \cos A \cos B,$$

using Eq. 6.19. Inserting these formulas in Eq. 6.20 gives formula (i).

For (ii), we have

$$\sin \alpha = \frac{\| (\mathbf{a} \times \mathbf{c}) \times (\mathbf{a} \times \mathbf{b}) \|}{\sin B \sin C} = \frac{\| ((\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b})\mathbf{a} - ((\mathbf{a} \times \mathbf{c}) \cdot \mathbf{a})\mathbf{b} \|}{\sin B \sin C} = \frac{|(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b}|}{\sin B \sin C}.$$

Hence,

$$\frac{\sin \alpha}{\sin A} = \frac{|(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b}|}{\sin A \sin B \sin C}. \tag{6.21}$$

Now, the scalar triple product $(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b}$ is unchanged, up to a sign, by any permutation of the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. It follows that the left-hand side of Eq. 6.21 is unchanged under any permutation of the vertices of the triangle. This gives formula (ii). □

As a special case, we have the spherical analogue of Pythagoras' theorem:

## Corollary 6.5.4

Suppose that a spherical triangle has sides of length $A$, $B$ and $C$ and that the angle opposite the side of length $C$ is a right angle. Then,

$$\cos C = \cos A \cos B.$$

The formal analogy between Eqs. 6.19 and 6.20 suggests that we should consider the spherical triangle with vertices

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\| \mathbf{b} \times \mathbf{c} \|}, \quad \mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{\| \mathbf{c} \times \mathbf{a} \|}, \quad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\| \mathbf{a} \times \mathbf{b} \|}.$$

Note that the cyclic order $\mathbf{a} \to \mathbf{b} \to \mathbf{c} \to \mathbf{a}$ of the vertices is preserved in these formulas; if the cyclic order was reversed the sign of all three vectors would change. The triangles with vertices $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ and $-\mathbf{a}^*, -\mathbf{b}^*, -\mathbf{c}^*$ are called the *dual triangles* of the triangle with vertices $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

Note that each of the two dual triangles is obtained from the other by applying the antipodal map $\mathbf{v} \mapsto -\mathbf{v}$ of $S^2$; since this is an isometry of $\mathbb{R}^3$ (see Appendix 1), it is also an isometry of $S^2$ (Exercise 6.1.2) so the two dual triangles have the same angles and sides of the same length. Geometrically, $\pm\mathbf{a}^*$ are the endpoints of the diameter of $S^2$ perpendicular to the plane that intersects $S^2$ in the great circle passing through $\mathbf{b}$ and $\mathbf{c}$: they are called the *poles* of this great circle (thus, the north and south poles of $S^2$ are the poles of the equator).

Note also that $\pm\mathbf{a}$ are the poles of the great circle through $\mathbf{b}^*$ and $\mathbf{c}^*$, since $\mathbf{a}$ is perpendicular to $\mathbf{b}^*$ and $\mathbf{c}^*$. It follows that the dual triangles of the triangle with vertices $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ are the original triangle with vertices $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and its image under the antipodal map. This can also be verified algebraically:

$$\begin{aligned} \mathbf{b}^* \times \mathbf{c}^* = \frac{(\mathbf{c} \times \mathbf{a}) \times (\mathbf{a} \times \mathbf{b})}{\parallel \mathbf{c} \times \mathbf{a} \parallel \parallel \mathbf{a} \times \mathbf{b} \parallel} &= \frac{((\mathbf{c} \times \mathbf{a}) \cdot \mathbf{b})\mathbf{a}}{\parallel \mathbf{c} \times \mathbf{a} \parallel \parallel \mathbf{a} \times \mathbf{b} \parallel}, \\ \therefore \quad \frac{\mathbf{b}^* \times \mathbf{c}^*}{\parallel \mathbf{b}^* \times \mathbf{c}^* \parallel} &= \pm\mathbf{a}, \end{aligned}$$

the sign being that of $(\mathbf{c} \times \mathbf{a}) \cdot \mathbf{b} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$. Thus, the dual triangle of the triangle with vertices $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ is the original triangle if $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) > 0$ and is its image under the antipodal map if $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) < 0$.

## Proposition 6.5.5

Let $\alpha, \beta, \gamma$ and $A, B, C$ be the angles and the lengths of the sides of a spherical triangle, so that $\alpha$ is the angle opposite the side of length $A$, etc. Let $\alpha^*, \beta^*, \gamma^*, A^*, B^*, C^*$ be the corresponding quantities for either of the dual triangles. Then,

$$\begin{aligned} \alpha^* = \pi - A, \quad \beta^* = \pi - B, \quad \gamma^* = \pi - C, \\ A^* = \pi - \alpha, \quad B^* = \pi - \beta, \quad C^* = \pi - \gamma. \end{aligned}$$

## Proof

Denoting the vertices of the triangle by $\mathbf{a}, \mathbf{b}, \mathbf{c}$ as above, Eq. 6.19 gives

$$\cos A^* = \mathbf{b}^* \cdot \mathbf{c}^* = \frac{(\mathbf{c} \times \mathbf{a}) \cdot (\mathbf{a} \times \mathbf{b})}{\parallel \mathbf{c} \times \mathbf{a} \parallel \parallel \mathbf{a} \times \mathbf{b} \parallel} = -\cos\alpha,$$

so, since both $\alpha$ and $A^*$ are between 0 and $\pi$,

$$A^* = \pi - \alpha. \tag{6.22}$$

The formula $\alpha^* = \pi - A$ is obtained by applying Eq. 6.22 to the dual triangles.

□

## Corollary 6.5.6

With the notation in Proposition 6.5.3, we have

$$\cos A = \frac{\cos\alpha + \cos\beta\cos\gamma}{\sin\beta\sin\gamma},$$

together with two similar formulas obtained by making the permutations $A \to B \to C \to A$, $\alpha \to \beta \to \gamma \to \alpha$.

## Proof

Just apply part (i) of Proposition 6.5.3 to the dual triangle and use Proposition 6.5.5. □

This formula is important because it shows that *the sides of a spherical triangle are determined by its angles*, unlike the situation in plane geometry in which there are 'similar' triangles with the same angles but possibly different sizes. The 'reason' for this is that in spherical geometry there is an absolute standard of length, namely the radius of the sphere.

Much of Euclidean geometry is concerned with the question of when two geometrical figures (such as triangles) are *congruent*, which means that one figure can be 'moved' so that it coincides with the other. The types of 'motions' that are allowed are those that do not change the size or shape of the triangles, namely the *isometries* of the plane (see Appendix 1). Hence, we need to determine the isometries of the sphere.

We know that any isometry of $\mathbb{R}^3$ that preserves $S^2$ will give an isometry of $S^2$ (see Exercise 6.1.2). The following proposition shows that we get all the isometries of $S^2$ this way (cf. Theorem A.1.5 and its proof).

## Proposition 6.5.7

Every isometry of $S^2$ is a composite of reflections in planes passing through the origin. In fact, at most three reflections are required.

## Proof

The first thing to observe is that isometries of $S^2$ must take great circles to great circles, since these are the curves of shortest length and isometries preserve length.

Let $F$ be any isometry of $S^2$, and let $\boldsymbol{e}_1 = (1,0,0)$, $\boldsymbol{e}_2 = (0,1,0)$ and $\boldsymbol{e}_3 = (0,0,1)$. If $F(\boldsymbol{e}_1) = \boldsymbol{e}_1$ let $G_1$ be the identity map. Otherwise, let $G_1$ be the reflection in the plane perpendicular to the line joining $\boldsymbol{e}_1$ to $F(\boldsymbol{e}_1)$ and passing through its mid-point; note that since $\| \boldsymbol{e}_1 \| = \| F(\boldsymbol{e}_1) \|$, this plane passes through the origin so $G_1$ is an isometry of $S^2$. Then $G_1 \circ F$ fixes $\boldsymbol{e}_1$. If $\boldsymbol{e}_2 = G_1(F(\boldsymbol{e}_2))$ let $G_2$ be the identity map. Otherwise, let $G_2$ be the reflection in the perpendicular bisector of the line joining $\boldsymbol{e}_2$ and $G_1(F(\boldsymbol{e}_2))$. Since $\| \boldsymbol{e}_2 \| = \| G_1(F(\boldsymbol{e}_2)) \|$ (because $F$ and $G_1$ are isometries), this plane passes through the origin so $G_2$ is an isometry of $S^2$, and since

$$\| \boldsymbol{e}_1 - G_1(F(\boldsymbol{e}_2)) \| = \| G_1(F(\boldsymbol{e}_1)) - G_1(F(\boldsymbol{e}_2)) \| = \| \boldsymbol{e}_1 - \boldsymbol{e}_2 \|,$$

$\boldsymbol{e}_1$ is fixed by $G_2$. Hence, $G_2 \circ G_1 \circ F$ fixes $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. Now the north and south poles $\pm\boldsymbol{e}_3$ are the only two points whose spherical distance from $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ is equal to $\pi/2$, so $G_2 \circ G_1 \circ F$ must either fix $\boldsymbol{e}_3$ or take it to $-\boldsymbol{e}_3$. In the first case let $G_3$ be the identity, in the second let $G_3$ be reflection in the $xy$-plane. Then, $H = G_3 \circ G_2 \circ G_1 \circ F$ is an isometry of $S^2$ that fixes $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$.

Since $H$ fixes $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ it must fix each point of the equator, since the equator is the unique great circle passing through these two points and any point on the equator is uniquely determined by its spherical distances from them. Similarly, $H$ must fix each point of the great circle passing through $\boldsymbol{e}_1$ and $\boldsymbol{e}_3$. If $\mathbf{a}$ is any point of $S^2$ other than the poles $\pm\boldsymbol{e}_3$, the unique great circle $\mathcal{C}$ passing through $\mathbf{a}$ and the poles intersects the equator at a point $\mathbf{b}$, say. Since $H$ fixes $\mathbf{b}$ and the poles, it fixes every point of $\mathcal{C}$ by the previous argument. In particular, $H$ fixes $\mathbf{a}$. Since $\mathbf{a}$ was an arbitrary point of the sphere, $H$ must be the identity map.

Hence, $F = G_1 \circ G_2 \circ G_3$ is a product of $\leq 3$ reflections. □

One of the most striking differences between Euclidean and spherical geometry is contained in the following result, which is strongly suggested by Corollary 6.5.6.

### Proposition 6.5.8

In spherical geometry, similar triangles are congruent.

This means that if two spherical triangles have vertices $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{a}', \mathbf{b}', \mathbf{c}'$, and if the angle of the first triangle at $\mathbf{a}$ is equal to that of the second triangle at $\mathbf{a}'$, and similarly for the other two angles, there is an isometry of $S^2$ that takes $\mathbf{a}$ to $\mathbf{a}'$, $\mathbf{b}$ to $\mathbf{b}'$ and $\mathbf{c}$ to $\mathbf{c}'$. We leave the proof to Exercise 6.5.2.

## *EXERCISES*

6.5.1 Find the angles and the lengths of the sides of an equilateral spherical triangle whose area is one quarter of the area of the sphere.

6.5.2 Show that similar spherical triangles are congruent.

6.5.3 The *spherical circle* of centre $\mathbf{p} \in S^2$ and radius $R$ is the set of points of $S^2$ that are a spherical distance $R$ from $\mathbf{p}$. Show that, if $0 \leq R \leq \pi/2$:

(i) A spherical circle of radius $R$ is a circle of radius $\sin R$.

(ii) The area inside a spherical circle of radius $R$ is $2\pi(1 - \cos R)$.

What if $R > \pi/2$?

6.5.4 This exercise describes the transformations of $\mathbb{C}_\infty$ corresponding to the isometries of $S^2$ under the stereographic projection map $\Pi : S^2 \to \mathbb{C}_\infty$ (Example 6.3.5). If $F$ is any isometry of $S^2$, let $F_\infty = \Pi \circ F \circ \Pi^{-1}$ be the corresponding bijection $C_\infty \to \mathbb{C}_\infty$.

(i) A Möbius transformation

$$M(w) = \frac{aw + b}{cw + d},$$

where $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$, is said to be *unitary* if $d = \bar{a}$ and $c = -\bar{b}$ (see Appendix 2). Show that the composite of two unitary Möbius transformations is unitary and that the inverse of a unitary Möbius transformation is unitary.

(ii) Show that if $F$ is the reflection in the plane passing through the origin and perpendicular to the unit vector $(a, b)$ (where $a \in \mathbb{C}$, $b \in \mathbb{R}$ – see Example 5.3.4), then

$$F_\infty(w) = \frac{-a\bar{w} + b}{b\bar{w} + \bar{a}}.$$

(iii) Deduce that if $F$ is any isometry of $S^2$ there is a unitary Möbius transformation $M$ such that either $F_\infty = M$ or $F_\infty = M \circ J$ where $J(w) = -\bar{w}$.

(iv) Show conversely that if $M$ is any unitary Möbius transformation, the bijections $\mathbb{C}_\infty \to \mathbb{C}_\infty$ given by $M$ and $M \circ J$ are both of the form $F_\infty$ for some isometry $F$ of $S^2$.

# 7 Curvature of surfaces

In this chapter, we discuss several approaches to the problem of measuring how 'curved' a surface is. Although they use quite different methods, we show that each of the approaches leads to the same geometric object: the *second fundamental form* of a surface. It turns out (see Theorem 10.1.3) that a surface is determined up to an isometry of $\mathbb{R}^3$ by its first and second fundamental forms, just as a unit-speed plane curve is determined up to an isometry of $\mathbb{R}^2$ by its signed curvature.

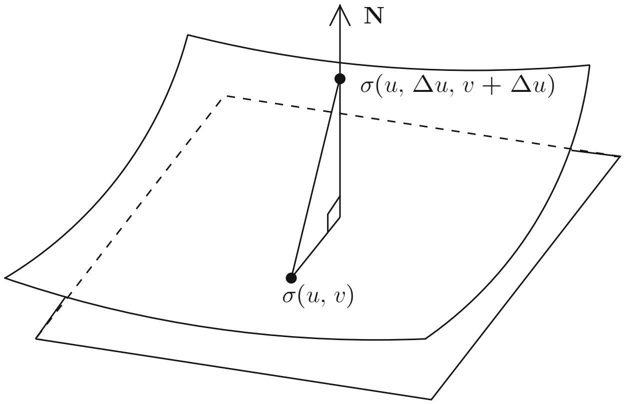
Throughout this chapter we shall work with *oriented* surfaces. Recall from Section 4.5 that every surface *patch* is oriented.

## 7.1 The second fundamental form

In our first attempt to define the curvature of a surface, we imitate the discussion at the beginning of Section 2.1, which leads to the definition of the curvature of a curve. Suppose then that $\boldsymbol{\sigma}$ is a surface patch in $\mathbb{R}^3$ with standard unit normal $\mathbf{N}$. As the parameters $(u, v)$ of $\boldsymbol{\sigma}$ change to $(u+\Delta u, v+\Delta v)$, the surface moves away from the plane through $\boldsymbol{\sigma}(u, v)$ parallel to the tangent plane by a distance

$$(\boldsymbol{\sigma}(u+\Delta u, v+\Delta v) - \boldsymbol{\sigma}(u, v)) \cdot \mathbf{N}.$$

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_7,




By the two variable form of Taylor's theorem, $\boldsymbol{\sigma}(u+\Delta u, v+\Delta v) - \boldsymbol{\sigma}(u, v)$ is equal to

$$\boldsymbol{\sigma}_u \Delta u + \boldsymbol{\sigma}_v \Delta v + \frac{1}{2}\left(\boldsymbol{\sigma}_{uu}(\Delta u)^2 + 2\boldsymbol{\sigma}_{uv}\Delta u \Delta v + \boldsymbol{\sigma}_{vv}(\Delta v)^2\right) + \text{ remainder},$$

where (remainder)$/((\Delta u)^2 + (\Delta v)^2)$ tends to zero as $(\Delta u)^2 + (\Delta v)^2$ tends to zero. Now $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are tangent to the surface, hence perpendicular to $\mathbf{N}$, so the deviation of $\boldsymbol{\sigma}$ from its tangent plane is

$$\frac{1}{2}\left(L(\Delta u)^2 + 2M\Delta u \Delta v + N(\Delta v)^2\right) + \text{remainder}, \tag{7.1}$$

where

$$L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N}, \quad M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N}, \quad N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N}. \tag{7.2}$$

Comparing Eq. 7.2 with Eq. 7.1, we see that the expression

$$L(\Delta u)^2 + 2M\Delta u \Delta v + N(\Delta v)^2$$

is the analogue for the surface of the curvature term $\kappa(\Delta t)^2$ in the case of a curve.

One calls the expression

$$Ldu^2 + 2Mdudv + Ndv^2 \tag{7.3}$$

the *second fundamental form* of the surface patch $\boldsymbol{\sigma}$. It clearly resembles the first fundamental form of a surface patch, and we can make sense of it in the same way, by interpreting $du$ and $dv$ as linear maps as in Section 6.1. Imitating the discussion there, we define a symmetric bilinear form on the tangent plane by

$$\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle = Ldu(\mathbf{v})du(\mathbf{w}) + M(du(\mathbf{v})dv(\mathbf{w}) + du(\mathbf{w})dv(\mathbf{v})) + Ndv(\mathbf{v})dv(\mathbf{w})$$

(cf. Exercise 6.1.3). In the next section, we shall give this form an appealing geometric interpretation, and extend its definition to an arbitrary oriented surface.

## Example 7.1.1

Consider the plane

$$\boldsymbol{\sigma}(u,v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q}$$

(see Example 4.1.2). Since $\boldsymbol{\sigma}_u = \mathbf{p}$ and $\boldsymbol{\sigma}_v = \mathbf{q}$ are constant vectors, we have $\boldsymbol{\sigma}_{uu} = \boldsymbol{\sigma}_{uv} = \boldsymbol{\sigma}_{vv} = \mathbf{0}$. Hence, the second fundamental form of a plane is zero.

## Example 7.1.2

Consider a surface of revolution (Example 5.3.2)

$$\boldsymbol{\sigma}(u,v) = (f(u)\cos v, f(u)\sin v, g(u));$$

we assume as usual that $f(u) > 0$ for all values of $u$ and that the profile curve $u \mapsto (f(u), 0, g(u))$ is unit-speed, i.e., $\dot{f}^2 + \dot{g}^2 = 1$ (a dot denoting $d/du$). Then:

$$\begin{aligned}
&\boldsymbol{\sigma}_u = (\dot{f}\cos v, \dot{f}\sin v, \dot{g}), \quad \boldsymbol{\sigma}_v = (-f\sin v, f\cos v, 0),\\
&\quad \therefore \quad \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (-f\dot{g}\cos v, -f\dot{g}\sin v, f\dot{f}),\\
&\quad \therefore \quad \| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \| = f \quad (\text{since } \dot{f}^2 + \dot{g}^2 = 1),\\
&\therefore \quad \mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \|} = (-\dot{g}\cos v, -\dot{g}\sin v, \dot{f}),\\
&\qquad\qquad \boldsymbol{\sigma}_{uu} = (\ddot{f}\cos v, \ddot{f}\sin v, \ddot{g}),\\
&\qquad\qquad \boldsymbol{\sigma}_{uv} = (-\dot{f}\sin v, \dot{f}\cos v, 0),\\
&\qquad\qquad \boldsymbol{\sigma}_{vv} = (-f\cos v, -f\sin v, 0),\\
&\therefore \quad L = \boldsymbol{\sigma}_{uu}\cdot\mathbf{N} = \dot{f}\ddot{g} - \ddot{f}\dot{g}, \;\; M = \boldsymbol{\sigma}_{uv}\cdot\mathbf{N} = 0, \;\; N = \boldsymbol{\sigma}_{vv}\cdot\mathbf{N} = f\dot{g},
\end{aligned}$$

so the second fundamental form is

$$(\dot{f}\ddot{g} - \ddot{f}\dot{g})du^2 + f\dot{g}dv^2.$$

For the unit sphere $S^2$ in latitude–longitude coordinates (Example 4.1.4), $u = \theta$, $v = \varphi$, $f(\theta) = \cos\theta$, $g(\theta) = \sin\theta$, giving the second fundamental form of $S^2$ as

$$d\theta^2 + \cos^2\theta\, d\varphi^2.$$

Note that this is the same as the *first* fundamental form of $S^2$ (see Example 6.1.3; the reason for this will appear in Section 8.2).

If the surface is the unit cylinder, we can take $f(u) = 1$, $g(u) = u$ (again, the conditions $f > 0$ and $\dot{f}^2 + \dot{g}^2 = 1$ are satisfied). This gives $L = M = 0$, $N = 1$, so the second fundamental form of the cylinder is $dv^2$.

## *EXERCISES*

7.1.1 Compute the second fundamental form of the elliptic paraboloid

$$\boldsymbol{\sigma}(u,v) = (u, v, u^2 + v^2).$$

7.1.2 Suppose that the second fundamental form of a surface patch $\boldsymbol{\sigma}$ is zero everywhere. Prove that $\boldsymbol{\sigma}$ is an open subset of a plane. This is the analogue for surfaces of the theorem that a curve with zero curvature everywhere is part of a straight line.

7.1.3 Let a surface patch $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ be a reparametrization of a surface patch $\boldsymbol{\sigma}(u, v)$ with reparametrization map $(u, v) = \Phi(\tilde{u}, \tilde{v})$. Prove that

$$\begin{pmatrix} \tilde{L} & \tilde{M} \\ \tilde{M} & \tilde{N} \end{pmatrix} = \pm J^t \begin{pmatrix} L & M \\ M & N \end{pmatrix} J,$$

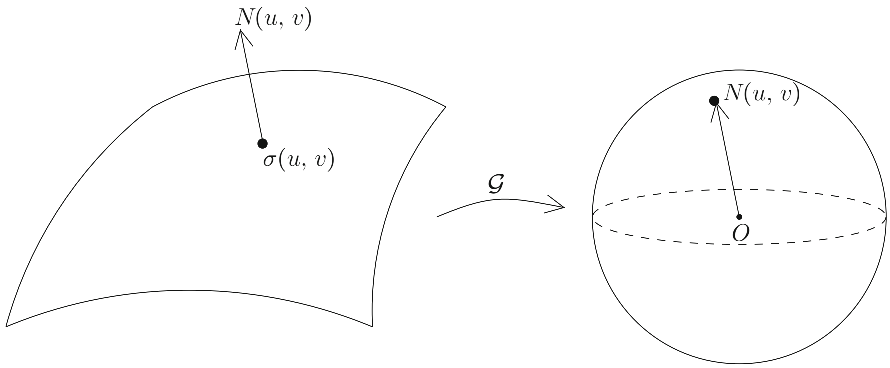
where $J$ is the Jacobian matrix of $\Phi$ and we take the plus sign if $\det(J) > 0$ and the minus sign if $\det(J) < 0$. Deduce from Exercise 6.1.4 that the second fundamental form of a surface patch is unchanged by a reparametrization of the patch which preserves its orientation.

7.1.4 What is the effect on the second fundamental form of a surface of applying an isometry of $\mathbb{R}^3$? Or a dilation?

# 7.2 The Gauss and Weingarten maps

Our second approach to defining the curvature of an oriented surface $\mathcal{S}$ is to consider its unit normal $\mathbf{N}$. The way that $\mathbf{N}$ varies clearly reflects the way in which $\mathcal{S}$ curves: $\mathbf{N}$ varies rapidly near a point at which the surface is highly curved and slowly where the surface is only slightly curved. If $\mathcal{S}$ is a plane, $\mathbf{N}$ is the same at all points of $\mathcal{S}$, i.e., $\mathbf{N}$ is a constant, and the curvature should be zero.

The values of $\mathbf{N}$ at the points of $\mathcal{S}$ are recorded by its *Gauss map* $\mathcal{G}_{\mathcal{S}}$ (or just $\mathcal{G}$ if there is no doubt as to which surface is intended). This is the map from $\mathcal{S}$ to the unit sphere $S^2$ that assigns to any point $\mathbf{p} \in \mathcal{S}$ the point $\mathbf{N}_{\mathbf{p}} \in S^2$, where $\mathbf{N}_{\mathbf{p}}$ is the unit normal of $\mathcal{S}$ at $\mathbf{p}$.




The rate at which $\mathbf{N}$ varies across $\mathcal{S}$ is measured by the derivative of $\mathcal{G}$:

$$D_{\mathbf{p}}\mathcal{G} : T_{\mathbf{p}}\mathcal{S} \to T_{\mathcal{G}}(\mathbf{p})S^2.$$

Now, the tangent plane at a point $\mathbf{q} \in S^2$ is the plane passing through the origin perpendicular to $\mathbf{q}$. Thus, $T_{\mathcal{G}}(\mathbf{p})S^2$ is the plane through the origin perpendicular to $\mathbf{N}_{\mathbf{p}}$, in other words, $T_{\mathbf{p}}\mathcal{S}$. Thus, the derivative of $\mathcal{G}$ is a linear map from the tangent plane of $\mathcal{S}$ to itself:

$$D_{\mathbf{p}}\mathcal{G} : T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S}.$$

## Definition 7.2.1

Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$. The *Weingarten map* $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ of $\mathcal{S}$ at $\mathbf{p}$ is defined by

$$\mathcal{W}_{\mathbf{p},\mathcal{S}} = -D_{\mathbf{p}}\mathcal{G}. \tag{7.4}$$

The *second fundamental form* of $\mathcal{S}$ at $\mathbf{p} \in \mathcal{S}$ is the bilinear form on $T_{\mathbf{p}}\mathcal{S}$ given by

$$\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle_{\mathbf{p},\mathcal{S}} = \langle \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}), \mathbf{w} \rangle_{\mathbf{p},\mathcal{S}}, \quad \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}.$$

The minus sign is introduced in Eq. 7.4 as a matter of convention (it will reduce the total number of minus signs later). We shall often omit the subscripts $\mathbf{p}$ and $\mathcal{S}$ from the Weingarten map if there is no danger of confusion.

The bilinearity asserted in this definition is easy to check. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$, $\lambda_1, \lambda_2 \in \mathbb{R}$. Then,

$$\begin{aligned}
\langle\langle \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2, \mathbf{w} \rangle\rangle &= \langle \mathcal{W}(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2), \mathbf{w} \rangle \\
&= \langle \lambda_1\mathcal{W}(\mathbf{v}_1) + \lambda_2\mathcal{W}(\mathbf{v}_2), \mathbf{w} \rangle \\
&= \lambda_1\langle \mathcal{W}(\mathbf{v}_1), \mathbf{w} \rangle + \lambda_2\langle \mathcal{W}(\mathbf{v}_2), \mathbf{w} \rangle \\
&= \lambda_1\langle\langle \mathbf{v}_1, \mathbf{w} \rangle\rangle + \lambda_2\langle\langle \mathbf{v}_2, \mathbf{w} \rangle\rangle,
\end{aligned}$$

where we used the linearity of $\mathcal{W}$ in passing from the first line to the second and the bilinearity of $\langle\ ,\ \rangle$ in passing from the second line to the third. This proves that $\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle$ is a linear function of $\mathbf{v}$ for each fixed $\mathbf{w}$. The proof that it is linear in $\mathbf{w}$ for each fixed $\mathbf{v}$ is similar but easier.

## Proposition 7.2.2

Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$, let $\boldsymbol{\sigma}(u,v)$ be a surface patch of $\mathcal{S}$ with $\mathbf{p}$ in its image, and let $Ldu^2 + 2Mdudv + Ndv^2$ be the second fundamental form of $\boldsymbol{\sigma}$ defined in Section 7.1. Then, for any $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$,

$$\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle = Ldu(\mathbf{v})du(\mathbf{w}) + M(du(\mathbf{v})dv(\mathbf{w}) + du(\mathbf{w})dv(\mathbf{v})) + Ndv(\mathbf{v})dv(\mathbf{w}). \tag{7.5}$$

To prove this we shall need the following lemma, which will also be used elsewhere.

## Lemma 7.2.3

Let $\boldsymbol{\sigma}(u,v)$ be a surface patch with standard unit normal $\mathbf{N}(u,v)$. Then,

$$\mathbf{N}_u \cdot \boldsymbol{\sigma}_u = -L, \quad \mathbf{N}_u \cdot \boldsymbol{\sigma}_v = \mathbf{N}_v \cdot \boldsymbol{\sigma}_u = -M, \quad \mathbf{N}_v \cdot \boldsymbol{\sigma}_v = -N,$$

where $L, M$ and $N$ are as defined in Eq. 7.2.

## Proof

Since $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are tangent vectors to the surface patch,

$$\mathbf{N} \cdot \boldsymbol{\sigma}_u = \mathbf{N} \cdot \boldsymbol{\sigma}_v = 0.$$

Differentiating these equations with respect to $u$ and $v$ gives

$$\begin{aligned} \mathbf{N}_u \cdot \boldsymbol{\sigma}_u &= -\mathbf{N} \cdot \boldsymbol{\sigma}_{uu} = -L, \quad & \mathbf{N}_v \cdot \boldsymbol{\sigma}_u &= -\mathbf{N} \cdot \boldsymbol{\sigma}_{uv} = -M, \\ \mathbf{N}_u \cdot \boldsymbol{\sigma}_v &= -\mathbf{N} \cdot \boldsymbol{\sigma}_{uv} = -M, \quad & \mathbf{N}_v \cdot \boldsymbol{\sigma}_v &= -\mathbf{N} \cdot \boldsymbol{\sigma}_{vv} = -N. \end{aligned} \qquad \square$$

*Proof 7.2.2* Since both sides of the equation in the statement of the proposition define bilinear forms on $T_{\mathbf{p}}\mathcal{S}$, it suffic  to verify that they agree when $\mathbf{v}$ and $\mathbf{w}$ are $\boldsymbol{\sigma}_u$ or $\boldsymbol{\sigma}_v$. Recalling that $du(\boldsymbol{\sigma}_u) = dv(\boldsymbol{\sigma}_v) = 1$, $du(\boldsymbol{\sigma}_v) = dv(\boldsymbol{\sigma}_u) = 0$, we have to prove that

$$\langle\langle \boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u \rangle\rangle = L, \quad \langle\langle \boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v \rangle\rangle = \langle\langle \boldsymbol{\sigma}_v, \boldsymbol{\sigma}_u \rangle\rangle = M, \quad \langle\langle \boldsymbol{\sigma}_v, \boldsymbol{\sigma}_v \rangle\rangle = N. \tag{7.6}$$

Let $\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$. Then, with the derivatives evaluated at $(u_0, v_0)$,

$$\mathcal{W}(\boldsymbol{\sigma}_u) = -\left.\frac{d}{du}\right|_{u=u_0} \mathcal{G}(\boldsymbol{\sigma}(u, v_0)) = -\left.\frac{d}{du}\right|_{u=u_0} \mathbf{N}(u, v_0) = -\mathbf{N}_u,$$

where $\mathbf{N}$ is the standard unit normal of $\boldsymbol{\sigma}$. Similarly, $\mathcal{W}(\boldsymbol{\sigma}_v) = -\mathbf{N}_v$. Hence,

$$\langle\langle \boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u \rangle\rangle = \langle \mathcal{W}(\boldsymbol{\sigma}_u), \boldsymbol{\sigma}_u \rangle = -\mathbf{N}_u \cdot \boldsymbol{\sigma}_u,$$

which is equal to $L$ by Lemma 7.2.3. The other equations in (7.6) are proved similarly. □

Since the formula (7.5) for $\langle\langle \mathbf{v}, \mathbf{w} \rangle\rangle$ is obviously unchanged when $\mathbf{v}$ and $\mathbf{w}$ are interchanged, we obtain

### Corollary 7.2.4

The second fundamental form is a symmetric bilinear form. Equivalently, the Weingarten map is self-adjoint.

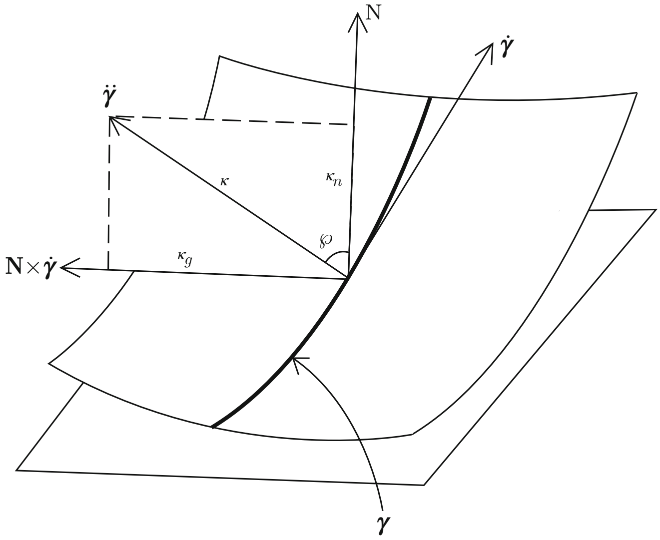
### *EXERCISES*

7.2.1 Calculate the Gauss map $\mathcal{G}$ of the paraboloid $\mathcal{S}$ with equation $z = x^2 + y^2$. What is the image of $\mathcal{G}$?

7.2.2 Show that the Weingarten map changes sign when the orientation of the surface changes.

## 7.3 Normal and geodesic curvatures

It is obvious that the shape of a surface influences the curvature of curves on the surface. For example, a curve on a plane or a cylinder can have zero curvature everywhere, but this is not possible for curves on a sphere since no segment of a straight line can lie on a sphere. Thus, another natural way to investigate how much a surface curves is to look at the curvature of curves on the surface. We shall see that this leads, once again, to the second fundamental form of the surface.




If $\gamma$ is a unit-speed curve on an oriented surface $\mathcal{S}$, then $\dot{\gamma}$ is a unit vector and is, by definition, a tangent vector to $\mathcal{S}$. Hence, $\dot{\gamma}$ is perpendicular to the unit normal $\mathbf{N}$ of $\mathcal{S}$, so $\dot{\gamma}$, $\mathbf{N}$ and $\mathbf{N} \times \dot{\gamma}$ are mutually perpendicular unit vectors. Again since $\gamma$ is unit-speed, $\ddot{\gamma}$ is perpendicular to $\dot{\gamma}$, and hence is a linear combination of $\mathbf{N}$ and $\mathbf{N} \times \dot{\gamma}$:

$$\ddot{\gamma} = \kappa_n \mathbf{N} + \kappa_g \mathbf{N} \times \dot{\gamma}. \tag{7.7}$$

## Definition 7.3.1

The scalars $\kappa_n$ and $\kappa_g$ in Eq. 7.7 are called the *normal curvature* and the *geodesic curvature* of $\gamma$, respectively.

Note that $\kappa_n$ and $\kappa_s$ both change sign when $\mathbf{N}$ is replaced by $-\mathbf{N}$, so on a general (not necessarily orientable) surface only the magnitudes of $\kappa_n$ and $\kappa_s$ are well defined.

## Proposition 7.3.2

With the above notation, we have

$$\kappa_n = \ddot{\gamma} \cdot \mathbf{N}, \quad \kappa_g = \ddot{\gamma} \cdot (\mathbf{N} \times \dot{\gamma}), \tag{7.8}$$

$$\kappa^2 = \kappa_n^2 + \kappa_g^2, \tag{7.9}$$

$$\kappa_n = \kappa \cos\psi, \kappa_g = \pm\kappa \sin\psi, \tag{7.10}$$

where $\kappa$ is the curvature of $\boldsymbol{\gamma}$ and $\psi$ is the angle between $\mathbf{N}$ and the principal normal $\mathbf{n}$ of $\boldsymbol{\gamma}$.

## Proof

Equations 7.8 and 7.9 follow from Eq. 7.7 and the fact that $\mathbf{N}$ and $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$ are perpendicular unit vectors. The first equation in (7.10) follows from $\ddot{\boldsymbol{\gamma}} = \kappa\mathbf{n}$, and the second then follows from Eq. 7.9. □

If $\boldsymbol{\gamma}$ is regular, but not necessarily unit-speed, we define the geodesic and normal curvatures of $\boldsymbol{\gamma}$ to be those of a unit-speed reparametrization of $\boldsymbol{\gamma}$ (see Exercise 7.3.1). When a unit-speed parameter $t$ is changed to another such parameter $\pm t + c$, where $c$ is a constant, it is clear that $\kappa_n \mapsto \kappa_n$ and $\kappa_g \mapsto \pm\kappa_g$, so $\kappa_n$ is well defined for any regular curve, while $\kappa_g$ is well defined up to sign. Equations 7.9 and 7.10 continue to hold if $\boldsymbol{\gamma}$ is any regular curve.

The following proposition is the most important single fact about the normal curvature, and reveals its relation to the second fundamental form $\langle\langle\ ,\ \rangle\rangle$.

## Proposition 7.3.3

If $\boldsymbol{\gamma}$ is a unit-speed curve on an oriented surface $\mathcal{S}$, its normal curvature is given by

$$\kappa_n = \langle\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle\rangle.$$

If $\boldsymbol{\sigma}$ is a surface patch of $\mathcal{S}$ and $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ is a curve in $\boldsymbol{\sigma}$,

$$\kappa_n = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2$$

in the notation of Section 7.1.

This result means that two curves which *touch* each other at a point $\mathbf{p}$ of a surface (i.e., which intersect at $\mathbf{p}$ and have parallel tangent vectors at $\mathbf{p}$) have the same normal curvature at $\mathbf{p}$.

## Proof

Since $\dot{\boldsymbol{\gamma}}$ is a tangent vector to $\mathcal{S}$, $\mathbf{N}.\dot{\boldsymbol{\gamma}} = 0$. Hence, $\mathbf{N} \cdot \ddot{\boldsymbol{\gamma}} = -\dot{\mathbf{N}} \cdot \dot{\boldsymbol{\gamma}}$ so

$$\kappa_n = \mathbf{N} \cdot \ddot{\boldsymbol{\gamma}} = -\dot{\mathbf{N}} \cdot \dot{\boldsymbol{\gamma}} = \langle \mathcal{W}(\dot{\boldsymbol{\gamma}}), \dot{\boldsymbol{\gamma}} \rangle = \langle\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle\rangle,$$

since

$$\dot{\mathbf{N}} = \frac{d}{dt}\mathcal{G}(\boldsymbol{\gamma}(t)) = -\mathcal{W}(\dot{\boldsymbol{\gamma}}).$$

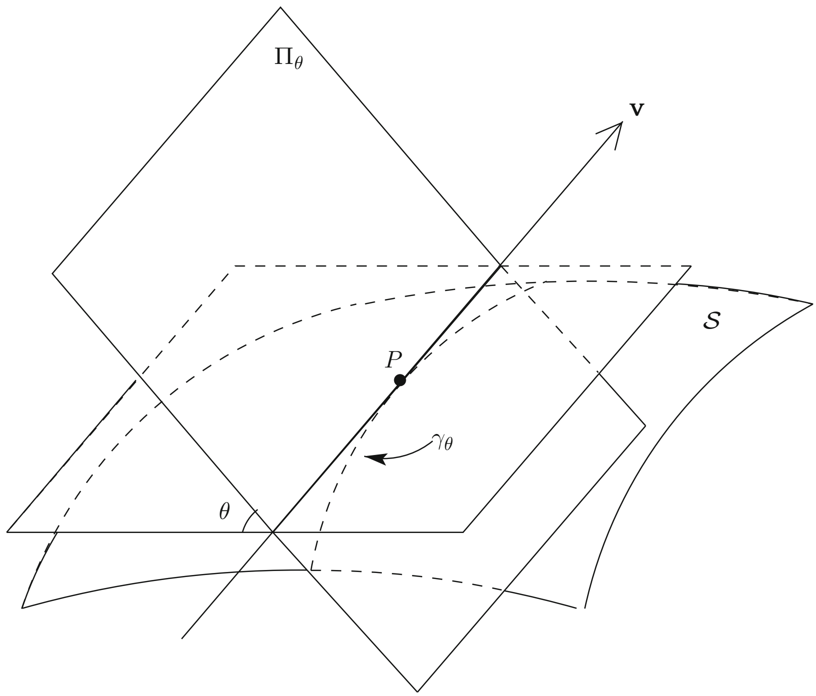
The second part follows from the first and Proposition 7.2.2. □

It turns out that, while the normal curvature depends on the second fundamental form of the surface, the geodesic curvature $\kappa_g$ depends only on its *first* fundamental form (see Exercise 7.3.4). But we leave further discussion of $\kappa_g$ to Chapter 9.

Here is a classical application of Proposition 7.3.3. It takes almost as long to state as to prove.

## Proposition 7.3.4 (Meusnier's Theorem)

Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$ and let $\mathbf{v}$ be a unit tangent vector to $\mathcal{S}$ at $\mathbf{p}$. Let $\Pi_\theta$ be the plane containing the line through $\mathbf{p}$ parallel to $\mathbf{v}$ and making an angle $\theta$ with the tangent plane $T_{\mathbf{p}}\mathcal{S}$, and assume that $\Pi_\theta$ is not parallel to $T_{\mathbf{p}}\mathcal{S}$. Suppose that $\Pi_\theta$ intersects $\mathcal{S}$ in a curve with curvature $\kappa_\theta$. Then, $\kappa_\theta \sin\theta$ is independent of $\theta$.




## Proof

Assume that $\boldsymbol{\gamma}_\theta$ is a unit-speed parametrization of the curve of intersection of $\Pi_\theta$ and $\mathcal{S}$. Then, at $\mathbf{p}$, $\dot{\boldsymbol{\gamma}}_\theta = \pm\mathbf{v}$, so $\ddot{\boldsymbol{\gamma}}_\theta$ is perpendicular to $\mathbf{v}$ and is parallel to $\Pi_\theta$. Thus, in the notation of Proposition 7.3.2, $\psi = \pi/2 - \theta$ and so Eq. 7.10 gives

$$\kappa_\theta \sin\theta = \kappa_n.$$

But $\kappa_n$ depends only on $\mathbf{p}$ and $\mathbf{v}$, and not on $\theta$. □

An important special case is that in which $\boldsymbol{\gamma}$ is a *normal section* of the surface, i.e., $\boldsymbol{\gamma}$ is the intersection of the surface with a plane $\Pi$ that is perpendicular to the tangent plane of the surface at every point of $\boldsymbol{\gamma}$.

## Corollary 7.3.5

The curvature $\kappa$, normal curvature $\kappa_n$ and geodesic curvature $\kappa_g$ of a normal section of a surface are related by

$$\kappa_n = \pm\kappa, \quad \kappa_g = 0.$$

## Proof

As in the proof of Proposition 7.3.3, $\kappa_n = \kappa \sin\theta$, where $\theta = \pm\pi/2$ for a normal section. This gives the first equation; the second follows from it and Eq. 7.9. □

## *EXERCISES*

7.3.1 Let $\boldsymbol{\gamma}$ be a regular, but not necessarily unit-speed, curve on a surface. Prove that (with the usual notation) the normal and geodesic curvatures of $\boldsymbol{\gamma}$ are

$$\kappa_n = \frac{\langle\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle\rangle}{\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle} \quad \text{and} \quad \kappa_g = \frac{\ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}})}{\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle^{3/2}}.$$

7.3.2 Show that the normal curvature of any curve on a sphere of radius $r$ is $\pm 1/r$.

7.3.3 Compute the geodesic curvature of any circle on a sphere (not necessarily a great circle).

7.3.4 Show that, if $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ is a unit-speed curve on a surface patch $\boldsymbol{\sigma}$ with first fundamental form $Edu^2 + 2Fdudv + Gdv^2$, the geodesic curvature of $\boldsymbol{\gamma}$ is

$$\kappa_g = (\ddot{v}\dot{u} - \dot{v}\ddot{u})\sqrt{EG - F^2} + A\dot{u}^3 + B\dot{u}^2\dot{v} + C\dot{u}\dot{v}^2 + D\dot{v}^3,$$

where $A$, $B$, $C$ and $D$ can be expressed in terms of $E$, $F$, $G$ and their derivatives. Find $A, B, C, D$ explicitly when $F = 0$.

7.3.5 Suppose that a unit-speed curve $\boldsymbol{\gamma}$ with curvature $\kappa > 0$ and principal normal $\mathbf{n}$ is a parametrization of the intersection of two oriented surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ with unit normals $\mathbf{N}_1$ and $\mathbf{N}_2$. Show that, if $\kappa_1$ and $\kappa_2$ are the normal curvatures of $\boldsymbol{\gamma}$ when viewed as a curve in $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, then

$$\kappa_1 \mathbf{N}_2 - \kappa_2 \mathbf{N}_1 = \kappa(\mathbf{N}_1 \times \mathbf{N}_2) \times \mathbf{n}.$$

Deduce that, if $\alpha$ is the angle between the two surfaces,

$$\kappa^2 \sin^2 \alpha = \kappa_1^2 + \kappa_2^2 - 2\kappa_1\kappa_2 \cos\alpha.$$

7.3.6 A curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$ is called *asymptotic* if its normal curvature is everywhere zero. Show that any straight line on a surface is an asymptotic curve. Show also that a curve $\boldsymbol{\gamma}$ with positive curvature is asymptotic if and only if its binormal $\mathbf{b}$ is parallel to the unit normal of $\mathcal{S}$ at all points of $\boldsymbol{\gamma}$.

7.3.7 Prove that the asymptotic curves on the surface

$$\boldsymbol{\sigma}(u, v) = (u\cos v, u \sin v, \ln u)$$

are given by

$$\ln u = \pm(v + c),$$

where $c$ is an arbitrary constant.

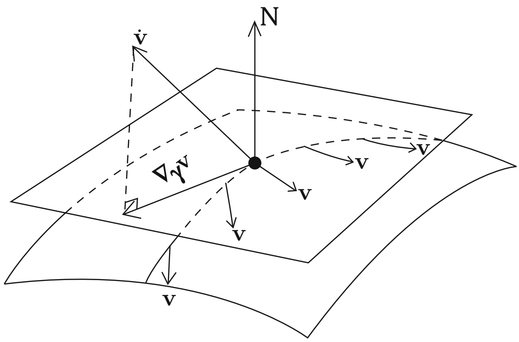
## 7.4 Parallel transport and covariant derivative

Imagine that there is a road that runs along the Earth's equator and that you are driving along this road at constant speed. The road would appear perfectly straight – you would not have to turn to the right or left to continue along the road. Thus, you would perceive your *velocity* (and not just your speed) as being *constant*. On the other hand, an observer in space would see that your velocity is not constant as you are travelling in a circle rather than in a straight line. The resolution of this apparent paradox is that an observer restricted to the surface of the Earth perceives only the component of the acceleration tangential to the surface. From the point of view of the observer in space, the acceleration vector points towards the centre of the Earth, and so has zero tangential component.

In general, suppose that $\boldsymbol{\gamma}$ is a curve on a surface $\mathcal{S}$ and let $\mathbf{v}$ be a *tangent vector field* along $\boldsymbol{\gamma}$, i.e., a smooth map from an open interval $(\alpha, \beta)$ to $\mathbb{R}^3$ such that $\mathbf{v}(t) \in T_{\boldsymbol{\gamma}(t)}\mathcal{S}$ for all $t \in (\alpha, \beta)$. To an observer moving along the curve $\boldsymbol{\gamma}$, the perceived rate of change of $\mathbf{v}$ is the tangential component of $\dot{\mathbf{v}}$, i.e., the orthogonal projection of $\dot{\mathbf{v}} = d\mathbf{v}/dt$ onto $T_{\boldsymbol{\gamma}(t)}\mathcal{S}$. If $\mathbf{N}$ is a unit normal to $\boldsymbol{\sigma}$, the component of $\dot{\mathbf{v}}$ perpendicular to the surface is $(\dot{\mathbf{v}} \cdot \mathbf{N})\mathbf{N}$, so the tangential component is

$$\nabla_{\boldsymbol{\gamma}}\mathbf{v} = \dot{\mathbf{v}} - (\dot{\mathbf{v}} \cdot \mathbf{N})\mathbf{N}. \tag{7.11}$$

Note that this is unchanged if $\mathbf{N}$ is replaced by $-\mathbf{N}$, so $\nabla_{\boldsymbol{\gamma}}\mathbf{v}$ is well defined on any surface, orientable or not.




## Definition 7.4.1

Let $\boldsymbol{\gamma}$ be a curve on a surface $\mathcal{S}$ and let $\mathbf{v}$ be a tangent vector field along $\boldsymbol{\gamma}$. The *covariant derivative* of $\mathbf{v}$ along $\boldsymbol{\gamma}$ is the orthogonal projection $\nabla_{\boldsymbol{\gamma}}\mathbf{v}$ of $d\mathbf{v}/dt$ onto the tangent plane $T_{\boldsymbol{\gamma}(t)}\mathcal{S}$ at a point $\boldsymbol{\gamma}(t)$.

In particular, an inhabitant of $\mathcal{S}$ would perceive $\mathbf{v}$ as being constant along $\boldsymbol{\gamma}$ if $\nabla_{\boldsymbol{\gamma}}\mathbf{v} = \mathbf{0}$. In this case, $\mathbf{v}$ is sometimes said to be *covariant constant*, but the usual terminology is contained in

## Definition 7.4.2

With the notation in Definition 7.4.1, $\mathbf{v}$ is said to be *parallel along* $\boldsymbol{\gamma}$ if $\nabla_{\boldsymbol{\gamma}}\mathbf{v} = \mathbf{0}$ at every point of $\boldsymbol{\gamma}$.

## Proposition 7.4.3

A tangent vector field $\mathbf{v}$ is parallel along a curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$ if and only if $\dot{\mathbf{v}}$ is perpendicular to the tangent plane of $\mathcal{S}$ at all points of $\boldsymbol{\gamma}$.

## Proof

This is clear from the definitions: if the right-hand side of Eq. 7.11 is zero, then $\dot{\mathbf{v}}$ is obviously parallel to $\mathbf{N}$. Conversely, if $\dot{\mathbf{v}} = \lambda\mathbf{N}$ for some scalar $\lambda$, then

$$\nabla_{\boldsymbol{\gamma}}\mathbf{v} = \dot{\mathbf{v}} - (\lambda\mathbf{N}\cdot\mathbf{N})\mathbf{N} = \dot{\mathbf{v}} - \lambda\mathbf{N} = \mathbf{0}. \qquad \square$$

To establish the existence of parallel tangent vector fields, we shall express the covariant derivative in terms of the parameters $u, v$ of a parametrization $\boldsymbol{\sigma}$ of the surface. To do this, we shall need the following calculation, which will also be used later.

## Proposition 7.4.4 (Gauss Equations)

Let $\boldsymbol{\sigma}(u, v)$ be a surface patch with first and second fundamental forms

$$Edu^2 + 2Fdudv + Gdv^2 \quad \text{and} \quad Ldu^2 + 2Mdudv + Ndv^2.$$

Then,

$$\begin{aligned}
\boldsymbol{\sigma}_{uu} &= \Gamma^1_{11}\boldsymbol{\sigma}_u + \Gamma^2_{11}\boldsymbol{\sigma}_v + L\mathbf{N},\\
\boldsymbol{\sigma}_{uv} &= \Gamma^1_{12}\boldsymbol{\sigma}_u + \Gamma^2_{12}\boldsymbol{\sigma}_v + M\mathbf{N},\\
\boldsymbol{\sigma}_{vv} &= \Gamma^1_{22}\boldsymbol{\sigma}_u + \Gamma^2_{22}\boldsymbol{\sigma}_v + N\mathbf{N},
\end{aligned}$$

where

$$\begin{aligned}
\Gamma^1_{11} &= \frac{GE_u - 2FF_u + FE_v}{2(EG-F^2)}, & \Gamma^2_{11} &= \frac{2EF_u - EE_v - FE_u}{2(EG-F^2)},\\
\Gamma^1_{12} &= \frac{GE_v - FG_u}{2(EG-F^2)}, & \Gamma^2_{12} &= \frac{EG_u - FE_v}{2(EG-F^2)},\\
\Gamma^1_{22} &= \frac{2GF_v - GG_u - FG_v}{2(EG-F^2)}, & \Gamma^2_{22} &= \frac{EG_v - 2FF_v + FG_u}{2(EG-F^2)}.
\end{aligned}$$

The six $\Gamma$ coefficients in these formulas are called *Christoffel symbols*. Note that they depend only on the *first* fundamental form of $\boldsymbol{\sigma}$.

## Proof

Since $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v, \mathbf{N}\}$ is a basis of $\mathbb{R}^3$, scalar functions $\alpha_1, \ldots, \gamma_3$ satisfying

$$\begin{aligned}
\boldsymbol{\sigma}_{uu} &= \alpha_1\boldsymbol{\sigma}_u + \alpha_2\boldsymbol{\sigma}_v + \alpha_3\mathbf{N},\\
\boldsymbol{\sigma}_{uv} &= \beta_1\boldsymbol{\sigma}_u + \beta_2\boldsymbol{\sigma}_v + \beta_3\mathbf{N},\\
\boldsymbol{\sigma}_{vv} &= \gamma_1\boldsymbol{\sigma}_u + \gamma_2\boldsymbol{\sigma}_v + \gamma_3\mathbf{N},
\end{aligned} \tag{7.12}$$

certainly exist. Taking the dot product of each equation with $\mathbf{N}$ gives

$$\alpha_3 = L, \quad \beta_3 = M, \quad \gamma_3 = N.$$

Now we take the dot product of each equation in (7.12) with $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. This gives six scalar equations from which we determine the remaining six coefficients. For example, taking the dot product of the first equation in (7.12) with $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ gives the two equations

$$E\alpha_1 + F\alpha_2 = \boldsymbol{\sigma}_{uu} \cdot \boldsymbol{\sigma}_u = \frac{1}{2}E_u,$$

$$F\alpha_1 + G\alpha_2 = \boldsymbol{\sigma}_{uu} \cdot \boldsymbol{\sigma}_v = (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)_u - \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uv} = F_u - \frac{1}{2}E_v.$$

Solving these equations gives $\alpha_1 = \Gamma^1_{11}, \alpha_2 = \Gamma^2_{11}$; similarly for the other four coefficients in Eq. 7.12. □

We can now establish the conditions for a tangent vector field $\mathbf{v}$ to be parallel along a curve $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t),\, v(t))$ on a surface patch $\boldsymbol{\sigma}(u, v)$. Since the tangent plane of $\boldsymbol{\sigma}$ is spanned by the vectors $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$, there are smooth scalar functions $\alpha$ and $\beta$ such that

$$\mathbf{v}(t) = \alpha(t)\boldsymbol{\sigma}_u + \beta(t)\boldsymbol{\sigma}_v,$$

the derivatives of $\boldsymbol{\sigma}$ being evaluated at $\boldsymbol{\sigma}(u(t), v(t))$.

## Proposition 7.4.5

Let $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ be a curve on a surface patch $\boldsymbol{\sigma}$, and let $\mathbf{v}(t) = \alpha(t)\boldsymbol{\sigma}_u + \beta(t)\boldsymbol{\sigma}_v$ be a tangent vector field along $\boldsymbol{\gamma}$, where $\alpha$ and $\beta$ are smooth functions of $t$. Then, $\mathbf{v}$ is parallel along $\boldsymbol{\gamma}$ if and only if the following equations are satisfied:

$$\begin{aligned} \dot{\alpha} + (\Gamma^1_{11}\dot{u} + \Gamma^1_{12}\dot{v})\alpha + (\Gamma^1_{12}\dot{u} + \Gamma^1_{22}\dot{v})\beta = 0 \\ \dot{\beta} + (\Gamma^2_{11}\dot{u} + \Gamma^2_{12}\dot{v})\alpha + (\Gamma^2_{12}\dot{u} + \Gamma^2_{22}\dot{v})\beta = 0. \end{aligned} \tag{7.13}$$

Note that these equations involve only the *first* fundamental form of $\boldsymbol{\sigma}$.

## Proof

Using the Gauss equations, we have

$$\begin{aligned} \dot{\mathbf{v}} = {} & \dot{\alpha}\boldsymbol{\sigma}_u + \dot{\beta}\boldsymbol{\sigma}_v + \alpha\dot{u}(\Gamma^1_{11}\boldsymbol{\sigma}_u + \Gamma^2_{11}\boldsymbol{\sigma}_v + L\mathbf{N}) \\ & + (\alpha\dot{v} + \beta\dot{u})(\Gamma^1_{12}\boldsymbol{\sigma}_u + \Gamma^2_{12}\boldsymbol{\sigma}_v + M\mathbf{N}) + \beta\dot{v}(\Gamma^1_{22}\boldsymbol{\sigma}_u + \Gamma^2_{22}\boldsymbol{\sigma}_v + N\mathbf{N}). \end{aligned} \tag{7.14}$$

By Proposition 7.4.3, $\mathbf{v}$ is parallel along $\boldsymbol{\gamma}$ if and only if $\dot{\mathbf{v}}$ is parallel to $\mathbf{N}$, which means that the coefficients of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ on the right-hand side of Eq. 7.14 must both be zero. But these coefficients are the left-hand sides of the two equations in (7.13). □

Eqs. (7.13) are of the form

$$\dot{\alpha} = f(\alpha, \beta, t), \quad \dot{\beta} = g(\alpha, \beta, t), \tag{7.15}$$

where $f$ and $g$ are smooth functions of three variables. It is proved in the theory of ordinary differential equations that such equations have a unique solution for any given set of *initial conditions*, i.e., if $t_0$ is some particular value of $t$, and $\alpha_0$, $\beta_0 \in \mathbb{R}$, there are unique smooth functions $\alpha(t)$ and $\beta(t)$, defined on an open interval containing $t_0$, that satisfy Eq. 7.16 and are such that

$$\alpha(t_0) = \alpha_0, \quad \beta(t_0) = \beta_0. \tag{7.16}$$

In the situation considered in Proposition 7.4.5, the initial conditions (7.16) are equivalent to

$$\mathbf{v}(t_0) = \alpha_0 \boldsymbol{\sigma}_u + \beta_0 \boldsymbol{\sigma}_v.$$

So we obtain

## Corollary 7.4.6

Let $\boldsymbol{\gamma}$ be a curve on a surface $\mathcal{S}$ and let $\mathbf{v}_0$ be a tangent vector of $\mathcal{S}$ at the point $\boldsymbol{\gamma}(t_0)$. Then, there is exactly one tangent vector field $\mathbf{v}$ that is parallel along $\boldsymbol{\gamma}$ and is such that $\mathbf{v}(t_0) = \mathbf{v}_0$.

## Example 7.4.7

Take $\boldsymbol{\gamma}$ to be a circle of latitude $\theta = \theta_0$ $(-\pi/2 < \theta_0 < \pi/2)$ on the unit sphere with the latitude–longitude parametrization $\boldsymbol{\sigma}(\theta, \varphi)$ (Example 4.1.4); thus, $\boldsymbol{\gamma}(\varphi) = \boldsymbol{\sigma}(\theta_0, \varphi)$. The first fundamental form of $\boldsymbol{\sigma}$ is $d\theta^2 + \cos^2\theta d\varphi^2$ (Example 6.1.3) from which we find

$$\Gamma^1_{11} = \Gamma^2_{11} = \Gamma^2_{22} = \Gamma^1_{12} = 0, \quad \Gamma^2_{12} = -\tan\theta, \quad \Gamma^1_{22} = \sin\theta\cos\theta.$$

The differential equations (7.13) become

$$\dot{\alpha} = -\beta \sin\theta_0 \cos\theta_0, \quad \dot{\beta} = \alpha \tan\theta_0. \tag{7.17}$$

If $\theta_0 = 0$, then $\alpha$ and $\beta$ are constant. If $\theta_0 \neq 0$, eliminating $\beta$ gives

$$\ddot{\alpha} + \alpha \sin^2\theta_0 = 0,$$

which has the general solution

$$\alpha(\varphi) = A\cos(\varphi\sin\theta_0) + B\sin(\varphi\sin\theta_0),$$

where $A$ and $B$ are constants; the second equation in (7.17) now gives

$$\beta = A\frac{\sin(\varphi\sin\theta_0)}{\cos\theta_0} - B\frac{\cos(\varphi\sin\theta_0)}{\sin\theta_0}.$$

Let us consider the special case in which $\mathbf{v} = \boldsymbol{\sigma}_\varphi$ is tangent to $\boldsymbol{\gamma}$ when $\varphi = 0$. Then, $\alpha = 0$, $\beta = 1$ when $\varphi = 0$, which gives $A = 0$, $B = -\sin\theta_0$ and hence

$$\mathbf{v}(\varphi) = -\sin\theta_0\sin(\varphi\sin\theta_0)\boldsymbol{\sigma}_\theta + \cos(\varphi\sin\theta_0)\boldsymbol{\sigma}_\varphi$$

(this solution is also correct when $\theta_0 = 0$). Note that $\mathbf{v}(\varphi)$ is not tangent to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(\varphi)$ in general. However, if $\theta_0 = 0$ then $\mathbf{v}$ is tangent to $\boldsymbol{\gamma}$ for all $\varphi$. Thus, *the tangent vector of $\boldsymbol{\gamma}$ is parallel along $\boldsymbol{\gamma}$ if and only if $\boldsymbol{\gamma}$ is a great circle.* We shall see the reason for this in Section 9.1.

If $\mathbf{p}$ and $\mathbf{q}$ are two points on a curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$, the covariant derivative enables us to associate to any vector in the tangent plane $T_\mathbf{p}\mathcal{S}$ a vector in the tangent plane $T_\mathbf{q}\mathcal{S}$. Indeed, suppose that $\mathbf{p}$ and $\mathbf{q}$ correspond to the parameter values $t_0$ and $t_1$, let $\mathbf{v}_0 \in T_\mathbf{p}\mathcal{S}$, let $\mathbf{v}(t)$ be the unique parallel vector field along $\boldsymbol{\gamma}$ such that $\mathbf{v}(t_0) = \mathbf{v}_0$ (see Corollary 7.4.6), and let $\mathbf{v}_1 = \mathbf{v}(t_1) \in T_\mathbf{q}\mathcal{S}$.

## Definition 7.4.8

With the above notation, the map $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}} : T_\mathbf{p}\mathcal{S} \to T_\mathbf{q}\mathcal{S}$ that takes $\mathbf{v}_0 \in T_\mathbf{p}\mathcal{S}$ to $\mathbf{v}_1 \in T_\mathbf{q}\mathcal{S}$ is called *parallel transport* from $\mathbf{p}$ to $\mathbf{q}$ along $\boldsymbol{\gamma}$.

## Proposition 7.4.9

With the notation in Definition 7.4.8,

(i) $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}} : T_\mathbf{p}\mathcal{S} \to T_\mathbf{q}\mathcal{S}$ is a linear map

(ii) $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}$ is an isometry, i.e., it preserves lengths and angles.

## Proof

Let $\mathbf{v}_0$, $\mathbf{w}_0 \in T_\mathbf{p}\mathcal{S}$ and let $\lambda, \mu \in \mathbb{R}$. Let $\mathbf{v}(t)$, $\mathbf{w}(t)$ be the parallel vector fields along $\boldsymbol{\gamma}$ such that $\mathbf{v}(t_0) = \mathbf{v}_0$, $\mathbf{w}(t_0) = \mathbf{w}_0$. If $\mathbf{V} = \lambda\mathbf{v} + \mu\mathbf{w}$, then $\dot{\mathbf{V}} = \lambda\dot{\mathbf{v}} + \mu\dot{\mathbf{w}}$ is parallel to the unit normal $\mathbf{N}$ of $\mathcal{S}$ because $\dot{\mathbf{v}}$ and $\dot{\mathbf{w}}$ are parallel to $\mathbf{N}$, so $\mathbf{V}$ is parallel along $\boldsymbol{\gamma}$. Hence,

$$\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}(\lambda\mathbf{v}_0 + \mu\mathbf{w}_0) = \Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}(\mathbf{V}(t_0)) = \mathbf{V}(t_1) = \lambda\mathbf{v}_1 + \mu\mathbf{w}_1 = \lambda\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}(\mathbf{v}_0) + \mu\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}(\mathbf{w}_0),$$

which proves that $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}$ is linear.

For (ii), note that

$$\frac{d}{dt}(\mathbf{v}\cdot\mathbf{w}) = \dot{\mathbf{v}}\cdot\mathbf{w} + \mathbf{v}\cdot\dot{\mathbf{w}} = ((\dot{\mathbf{v}}\cdot\mathbf{N})\mathbf{N})\cdot\mathbf{w} + \mathbf{v}\cdot((\dot{\mathbf{w}}\cdot\mathbf{N})\mathbf{N}) = 0,$$

since $\mathbf{v}\cdot\mathbf{N} = \mathbf{w}\cdot\mathbf{N} = 0$ ($\mathbf{v}$ and $\mathbf{w}$ are tangent to the surface). Hence,

$$\mathbf{v}_0\cdot\mathbf{w}_0 = \mathbf{v}_1\cdot\mathbf{w}_1.$$

Thus, $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}$ preserves dot products of vectors. Since lengths and angles are expressible in terms of dot products, part (ii) is proved. □

## Example 7.4.10

Take $\boldsymbol{\gamma}$ to be the equator of $S^2$. We saw in Example 7.4.7 that the tangent vector $\boldsymbol{\sigma}_\varphi$ of $\boldsymbol{\gamma}$ is parallel along $\boldsymbol{\gamma}$. Now, at points of the equator, $\boldsymbol{\sigma}_\theta$ is a unit vector perpendicular to $\boldsymbol{\sigma}_\varphi$. By Proposition 7.4.9, the parallel vector field along $\boldsymbol{\gamma}$ equal to $\boldsymbol{\sigma}_\theta$ when $\varphi = 0$ has the same property. It must therefore be equal to $\boldsymbol{\sigma}_\theta$; in other words, $\boldsymbol{\sigma}_\theta$ is also parallel along $\boldsymbol{\gamma}$ (this can also be checked directly from the formulas in Example 7.4.7). Since parallel transport is a linear map, it follows that, for any two points $\mathbf{p}$ and $\mathbf{q}$ on the equator, and any $\lambda, \mu \in \mathbb{R}$,

$$\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}(\lambda\boldsymbol{\sigma}_\theta + \mu\boldsymbol{\sigma}_\varphi) = \lambda\boldsymbol{\sigma}_\theta + \mu\boldsymbol{\sigma}_\varphi. \tag{7.18}$$

(Note, however, that $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}$ is not the identity map unless $\mathbf{p}$ and $\mathbf{q}$ coincide! If $\mathbf{p} \neq \mathbf{q}$ the derivatives $\boldsymbol{\sigma}_\theta$ and $\boldsymbol{\sigma}_\varphi$ on the two sides of (7.18) are being evaluated at different points of $S^2$.)

## *EXERCISES*

7.4.1 Let $\tilde{\boldsymbol{\gamma}}$ be a reparametrization of $\boldsymbol{\gamma}$, so that $\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}(\varphi(t))$ for some smooth function $\varphi$ with $d\varphi/dt \neq 0$ for all values of $t$. If $\mathbf{v}$ is a tangent vector field along $\boldsymbol{\gamma}$, show that $\tilde{\mathbf{v}}(t) = \mathbf{v}(\varphi(t))$ is one along $\tilde{\boldsymbol{\gamma}}$. Prove that

$$\nabla_{\tilde{\boldsymbol{\gamma}}}\tilde{\mathbf{v}} = \frac{d\varphi}{dt}\nabla_{\boldsymbol{\gamma}}\mathbf{v},$$

and deduce that $\mathbf{v}$ is parallel along $\boldsymbol{\gamma}$ if and only if $\tilde{\mathbf{v}}$ is parallel along $\tilde{\boldsymbol{\gamma}}$.

7.4.2 Show that the parallel transport map $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}} : T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{q}}\mathcal{S}$ is invertible. What is its inverse?

7.4.3 Suppose that a triangle on the unit sphere whose sides are arcs of great circles has vertices $\mathbf{p}$, $\mathbf{q}$, $\mathbf{r}$. Let $\mathbf{v}_0$ be a non-zero tangent vector to the arc $\overline{\mathbf{pq}}$ through $\mathbf{p}$ and $\mathbf{q}$ at $\mathbf{p}$. Show that, if we parallel transport $\mathbf{v}_0$ along $\overline{\mathbf{pq}}$, then along $\overline{\mathbf{qr}}$ and then along $\overline{\mathbf{rp}}$, the result is to rotate $\mathbf{v}_0$ through an angle $2\pi - \mathcal{A}$, where $\mathcal{A}$ is the area of the triangle. For an analogous result see Theorem 13.6.4.

# 8
# Gaussian, mean and principal curvatures

In this chapter, we show how to extract geometric information from the second fundamental form of a surface or, equivalently, from its Weingarten map.

## 8.1 Gaussian and mean curvatures

We start by defining two new measures of the curvature of a surface.

### Definition 8.1.1

Let $\mathcal{W}$ be the Weingarten map of an oriented surface $\mathcal{S}$ at a point $\mathbf{p} \in \mathcal{S}$. The Gaussian curvature $K$ and mean curvature $H$ of $\mathcal{S}$ at $\mathbf{p}$ are defined by

$$K = \det(\mathcal{W}), \quad H = \frac{1}{2}\mathrm{trace}(\mathcal{W}).$$

Recall that the determinant and trace of a linear map (such as $\mathcal{W}$) can be computed as the determinant and the sum of the diagonal entries of the matrix of the linear map with respect to any basis (in this case of the tangent plane), and that they depend only on the linear map and not on the choice of basis.

When the sign of the unit normal of $\mathcal{S}$ is changed, the Weingarten map also changes sign (Exercise 7.2.2), thus leaving $K$ unchanged. This implies that the Gaussian curvature is defined for *any* surface $\mathcal{S}$, orientable or not: to define $K$ at a point $\mathbf{p} \in \mathcal{S}$, choose a surface patch $\boldsymbol{\sigma}$ with $\mathbf{p}$ in its image; this is an

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_8,

oriented surface, which may be used to define $K$, and the result is independent of the choice of $\boldsymbol{\sigma}$. On the other hand, on a surface that is not necessarily orientable, $H$ is in general only well defined up to sign.

To get explicit formulas for $H$ and $K$, we work in a surface patch of $\mathcal{S}$. Let $\boldsymbol{\sigma}(u,v)$ be a surface patch with first and second fundamental forms

$$Edu^2 + 2Fdudv + Gdv^2 \quad \text{and} \quad Ldu^2 + 2Mdudv + Ndv^2,$$

respectively. Define symmetric $2 \times 2$ matrices $\mathcal{F}_I$ and $\mathcal{F}_{II}$ by

$$\mathcal{F}_I = \begin{pmatrix} E & F \\ F & G \end{pmatrix}, \quad \mathcal{F}_{II} = \begin{pmatrix} L & M \\ M & N \end{pmatrix}.$$

## Proposition 8.1.2

Let $\boldsymbol{\sigma}$ be a surface patch of an oriented surface $\mathcal{S}$. Then, with the above notation, the matrix of $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$ is $\mathcal{F}_I^{-1}\mathcal{F}_{II}$.

## Proof

By the proof of Proposition 7.2.2, $\mathcal{W}(\boldsymbol{\sigma}_u) = -\mathbf{N}_u$ and $\mathcal{W}(\boldsymbol{\sigma}_v) = -\mathbf{N}_v$, so the matrix of $\mathcal{W}$ is $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$, where

$$-\mathbf{N}_u = a\boldsymbol{\sigma}_u + b\boldsymbol{\sigma}_v, \quad -\mathbf{N}_v = c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v.$$

Take the dot product of each of these equations with $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ and use Lemma 7.2.3; this gives

$$\begin{aligned} L &= aE + bF, \qquad & M &= cE + dF, \\ M &= aF + bG, \qquad & N &= cF + dG. \end{aligned}$$

These four scalar equations are equivalent to the single matrix equation

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

$$\text{i.e.,} \quad \mathcal{F}_{II} = \mathcal{F}_I \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Hence, the matrix of $\mathcal{W}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \mathcal{F}_I^{-1}\mathcal{F}_{II}.$$

□

## Corollary 8.1.3

We have

$$H = \frac{LG - 2MF + NE}{2(EG - F^2)}, \quad K = \frac{LN - M^2}{EG - F^2}.$$

## Proof

By Definition 8.1.1,

$$K = \det\left(\mathcal{F}_I^{-1}\mathcal{F}_{II}\right) = \frac{\det(\mathcal{F}_{II})}{\det(\mathcal{F}_I)} = \frac{LN - M^2}{EG - F^2}.$$

To compute $H$, we need the trace of the matrix

$$\begin{aligned} \mathcal{F}_I^{-1}\mathcal{F}_{II} &= \frac{1}{EG - F^2}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}\begin{pmatrix} L & M \\ M & N \end{pmatrix} \\ &= \frac{1}{EG - F^2}\begin{pmatrix} LG - MF & MG - NF \\ ME - LF & NE - MF \end{pmatrix}. \end{aligned}$$

Thus,

$$2H = \operatorname{trace}\left(\mathcal{F}_I^{-1}\mathcal{F}_{II}\right) = \frac{LG - 2MF + NE}{EG - F^2}. \qquad \square$$

## Example 8.1.4

In Examples 6.1.3 and 7.1.2 we considered the surface of revolution

$$\boldsymbol{\sigma}(u, v) = (f(u)\cos v, f(u)\sin v, g(u)),$$

where we can assume that $f > 0$ and $\dot{f}^2 + \dot{g}^2 = 1$ everywhere (a dot denoting $d/du$). We found that

$$E = 1,\ F = 0,\ G = f^2, \quad L = \dot{f}\ddot{g} - \ddot{f}\dot{g},\ M = 0,\ N = f\dot{g}.$$

By Corollary 8.1.3, the Gaussian curvature is

$$K = \frac{LN - M^2}{EG - F^2} = \frac{(\dot{f}\ddot{g} - \ddot{f}\dot{g})f\dot{g}}{f^2}.$$

We can simplify this formula by noting that $\dot{f}^2 + \dot{g}^2 = 1$ implies (by differentiating with respect to $u$) that $\dot{f}\ddot{f} + \dot{g}\ddot{g} = 0$,

$$\begin{aligned} \therefore \quad (\dot{f}\ddot{g} - \ddot{f}\dot{g})\dot{g} &= -\dot{f}^2\ddot{f} - \ddot{f}\dot{g}^2 = -\ddot{f}(\dot{f}^2 + \dot{g}^2) = -\ddot{f}, \\ \therefore \quad K &= -\frac{\ddot{f}f}{f^2} = -\frac{\ddot{f}}{f}. \end{aligned}$$

We consider some special cases. If $\boldsymbol{\gamma}(u) = (u, 0, 0)$ is the $x$-axis, the corresponding surface of revolution is the $xy$-plane; since $f(u) = u$, we have $\dot{f} = 1$, $\ddot{f} = 0$, so $K = 0$. If $\boldsymbol{\gamma}(u) = (1, 0, u)$ is a straight line parallel to the $z$-axis, the corresponding surface is the unit cylinder; since $f(u) = 1$, $\ddot{f} = 0$, so $K = 0$. Finally, if $\boldsymbol{\gamma}(u) = (\cos u, 0, \sin u)$ is a circle of radius 1, the corresponding surface is the unit sphere; since $f(u) = \cos u$, $\dot{f} = -\sin u$, $\ddot{f} = -\cos u$ so $K = -\ddot{f}/f = -(-\cos u)/\cos u = 1$. Note that in each of these examples the curve $\boldsymbol{\gamma}$ is unit-speed.

## Example 8.1.5

For a ruled surface, take a patch

$$\boldsymbol{\sigma}(u, v) = \boldsymbol{\gamma}(u) + v\boldsymbol{\delta}(u),$$

(see Example 5.3.1). Denoting $d/du$ by a dot, we have $\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}$, $\boldsymbol{\sigma}_v = \boldsymbol{\delta}$, so

$$\boldsymbol{\sigma}_{uv} = \dot{\boldsymbol{\delta}}, \quad \boldsymbol{\sigma}_{vv} = \mathbf{0}.$$

Hence, if $\mathbf{N} = (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)/ \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel$ is the standard unit normal of $\boldsymbol{\sigma}$, then $M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} = \dot{\boldsymbol{\delta}} \cdot \mathbf{N}$ and $N = 0$. So

$$K = \frac{LN - M^2}{EG - F^2} = \frac{-(\dot{\boldsymbol{\delta}} \cdot \mathbf{N})^2}{EG - F^2} \leq 0,$$

i.e., *the Gaussian curvature of a ruled surface is negative or zero.*

Gauss discovered a way to obtain the Gaussian curvature from the Gauss map itself, rather than from its derivative, the Weingarten map. His result is an analogue of Proposition 2.2.3, which shows that, if $\boldsymbol{\gamma}$ is a unit-speed plane curve, its signed curvature $\kappa_s = \dot{\varphi}$, where $\varphi$ is the angle between its tangent vector $\dot{\boldsymbol{\gamma}}$ and a fixed direction, i.e., the (signed) *curvature is the rate of change of direction of the tangent vector of* $\boldsymbol{\gamma}$ *per unit length.* The 'direction' of the tangent plane to an oriented surface $\mathcal{S}$ is measured by its unit normal $\mathbf{N}$, so we might expect that a measure of the curvature of $\boldsymbol{\sigma}$ is the 'rate of change of $\mathbf{N}$ per unit area'. The values of $\mathbf{N}$ at points of $\mathcal{S}$ are recorded by the Gauss map $\mathcal{G}$, so if $R$ is a small region on $\mathcal{S}$ containing a point $\mathbf{p}$, we should look at the ratio

$$\frac{\text{Area}(\mathcal{G}(R))}{\text{Area}(R)}$$

in the limit as the region $R$ shrinks down to the point $\mathbf{p}$.

To make this idea precise, we work in a surface patch.

## Theorem 8.1.6

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a surface patch, let $(u_0, v_0) \in U$, and let $\delta > 0$ be such that the closed disc

$$R_\delta = \{(u, v) \in \mathbb{R}^2 \mid (u - u_0)^2 + (v - v_0)^2 \le \delta^2\}$$

with centre $(u_0, v_0)$ and radius $\delta$ is contained in $U$. Then,

$$\lim_{\delta \to 0} \frac{\mathcal{A}_{\mathbf{N}}(R_\delta)}{\mathcal{A}_{\boldsymbol{\sigma}}(R_\delta)} = |K|,$$

where $K$ is the Gaussian curvature of $\boldsymbol{\sigma}$ at $\boldsymbol{\sigma}(u_0, v_0)$.

Note that a $\delta$ with the properties in the statement of the theorem exists because $U$ is open.

## Proof

By Definition 6.4.1,

$$\frac{\mathcal{A}_{\mathbf{N}}(R_\delta)}{\mathcal{A}_{\boldsymbol{\sigma}}(R_\delta)} = \frac{\int_{R_\delta} \parallel \mathbf{N}_u \times \mathbf{N}_v \parallel \, dudv}{\int_{R_\delta} \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel \, dudv}. \tag{8.1}$$

In the notation of the proof of Proposition 8.1.2,

$$\begin{aligned}
\mathbf{N}_u \times \mathbf{N}_v &= (a\boldsymbol{\sigma}_u + b\boldsymbol{\sigma}_v) \times (c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v) \\
&= (ad - bc)\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \\
&= \det(\mathcal{F}_I^{-1}\mathcal{F}_{II})\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \\
&= \frac{\det(\mathcal{F}_{II})}{\det(\mathcal{F}_I)}\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \\
&= \frac{LN - M^2}{EG - F^2}\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \\
&= K\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \quad \text{(by Corollary 8.1.3).}
\end{aligned} \tag{8.2}$$

Substituting in Eq. 8.1, we get

$$\frac{\mathcal{A}_{\mathbf{N}}(R_\delta)}{\mathcal{A}_{\boldsymbol{\sigma}}(R_\delta)} = \frac{\int_{R_\delta} |K| \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel \, dudv}{\int_{R_\delta} \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel \, dudv}.$$

Let $\epsilon$ be any positive number. Since $K(u, v)$ is a continuous function of $(u, v)$ (see Exercise 8.1.3), we can choose $\delta > 0$ so small that

$$|K(u, v) - K(u_0, v_0)| < \epsilon$$

if $(u,v) \in R_\delta$. Since, for any real numbers $a, b$, $|a-b| \geq ||a| - |b||$, it follows that $||K(u,v)| - |K(u_0,v_0)|| < \epsilon$ if $(u,v) \in R_\delta$, i.e.,

$$|K(u_0,v_0)| - \epsilon < |K(u,v)| < |K(u_0,v_0)| + \epsilon$$

if $(u,v) \in R_\delta$. Multiplying through by $\parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel$ and integrating over $R_\delta$, we get

$$\begin{aligned}
(|K(u_0,v_0)| - \epsilon)\int \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel dudv &< \int |K(u,v)| \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel dudv \\
&< (|K(u_0,v_0)| + \epsilon)\int \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel dudv, \\
\therefore \quad |K(u_0,v_0)| - \epsilon < \frac{\mathcal{A}_{\mathbf{N}}(R_\delta)}{\mathcal{A}_{\boldsymbol{\sigma}}(R_\delta)} &< |K(u_0,v_0)| + \epsilon \quad \text{(using Eq. 8.1)} \\
\therefore \quad \left| \frac{\mathcal{A}_{\mathbf{N}}(R_\delta)}{\mathcal{A}_{\boldsymbol{\sigma}}(R_\delta)} - |K(u_0,v_0)| \right| &< \epsilon.
\end{aligned}$$

This proves the theorem. □

Although this proposition only gives the absolute value of the Gaussian curvature $K$, the sign can be recovered from the Gauss map if we define the *signed area* of $\mathcal{G}(R)$ to be $\pm\mathcal{A}_{\mathbf{N}}(R)$, where the sign is $+$ or $-$ according to whether $\mathbf{N}_u \times \mathbf{N}_v$ points in the same or the opposite direction as $\mathbf{N}$. By Eq. 8.3, this sign is that of $K$, so $K$ is the limit of the ratio

$$\frac{\text{Signed area}(\mathcal{G}(R))}{\text{Area}(R)}$$

as the region $R$ shrinks to the point $\mathbf{p}$.

As the following examples show, Theorem 8.1.6 sometimes allows one to find the Gaussian curvature of a surface with no calculation.

## Example 8.1.7

For a plane, the unit normal is constant. Thus, for any $R$, $\mathcal{G}(R)$ is a single point, and thus has zero area. By the theorem, a plane has Gaussian curvature zero everywhere.

For a generalized cylinder, the unit normal is clearly always perpendicular to the rulings of the cylinder, so the image of the Gauss map is contained in the great circle on $S^2$ formed by intersecting $S^2$ with the plane passing through its centre perpendicular to the rulings of the cylinder. Any great circle obviously has zero area, so the cylinder has zero Gaussian curvature too.

Finally, for the unit sphere $S^2$ itself, the unit normal at a point $\mathbf{p}$ is clearly parallel to the radius vector from the centre of the sphere to $\mathbf{p}$. In other words, the Gauss map is the identity map or the antipodal map (depending on the choice of orientation). Both of these maps are obviously equiareal, so the absolute value of the Gaussian curvature of $S^2$ is 1. In fact, if $\boldsymbol{\sigma}$ is any surface patch of $S^2$, we have $\mathbf{N} = \pm\boldsymbol{\sigma}$ so with either choice of sign $\mathbf{N}_u \times \mathbf{N}_v = \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$ is a positive multiple of $\mathbf{N}$ and the Gaussian curvature is $+1$.

## *EXERCISES*

8.1.1 Show that the Gaussian and mean curvatures of the surface $z = f(x, y)$, where $f$ is a smooth function, are

$$K = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}, \quad H = \frac{(1 + f_y^2)f_{xx} - 2f_x f_y f_{xy} + (1 + f_x^2)f_{yy}}{2(1 + f_x^2 + f_y^2)^{3/2}}.$$

8.1.2 Calculate the Gaussian curvature of the helicoid and catenoid (Exercises 4.2.6 and 5.3.1).

8.1.3 Show that the Gaussian and mean curvatures of a surface $\mathcal{S}$ are smooth functions on $\mathcal{S}$.

8.1.4 In the notation of Example 8.1.5, show that if $\boldsymbol{\delta}$ is the principal normal $\mathbf{n}$ of $\boldsymbol{\gamma}$ or its binormal $\mathbf{b}$, then $K = 0$ if and only if $\boldsymbol{\gamma}$ is planar.

8.1.5 What is the effect on the Gaussian and mean curvatures of a surface $\mathcal{S}$ if we apply a dilation of $\mathbb{R}^3$ to $\mathcal{S}$?

8.1.6 Show that the Weingarten map $\mathcal{W}$ of a surface satisfies the quadratic equation

$$\mathcal{W}^2 - 2H\mathcal{W} + K = 0,$$

in the usual notation.

8.1.7 Show that the image of the Gauss map of a generalized cone is a *curve* on $S^2$, and deduce that the cone has zero Gaussian curvature.

8.1.8 Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a patch of a surface $\mathcal{S}$. Show that the image under the Gauss map of the part $\boldsymbol{\sigma}(R)$ of $\mathcal{S}$ corresponding to a region $R \subseteq U$ has area

$$\int_R |K| d\mathcal{A}_{\boldsymbol{\sigma}},$$

where $K$ is the Gaussian curvature of $\mathcal{S}$.

8.1.9 Let $\mathcal{S}$ be the torus in Exercise 4.2.5. Describe the parts $\mathcal{S}^+$ and $\mathcal{S}^-$ of $\mathcal{S}$ where the Gaussian curvature $K$ of $\mathcal{S}$ is positive and negative, respectively. Show, without calculation, that

$$\int_{\mathcal{S}^+} K\, d\mathcal{A} = -\int_{\mathcal{S}^-} K\, d\mathcal{A} = 4\pi.$$

It follows that $\int_{\mathcal{S}} K\, d\mathcal{A} = 0$, a result that will be 'explained' in Section 13.4.

8.1.10 Let $\mathbf{w}(u, v)$ be a *smooth tangent vector field on a surface patch* $\boldsymbol{\sigma}(u, v)$. This means that

$$\mathbf{w}(u, v) = \alpha(u, v)\boldsymbol{\sigma}_u + \beta(u, v)\boldsymbol{\sigma}_v$$

where $\alpha$ and $\beta$ are smooth functions of $(u, v)$. Then, if $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ is any curve on $\boldsymbol{\sigma}$, $\mathbf{w}$ gives rise to the tangent vector field $\mathbf{w}|_{\boldsymbol{\gamma}}(t) = \mathbf{w}(u(t), v(t))$ along $\boldsymbol{\gamma}$. Let $\nabla_u \mathbf{w}$ be the covariant derivative of $\mathbf{w}|_{\boldsymbol{\gamma}}$ along a parameter curve $v =$ constant, and define $\nabla_v \mathbf{w}$ similarly. (Note that if $\boldsymbol{\sigma}$ is the $uv$-plane, then $\nabla_u$ and $\nabla_v$ become $\partial/\partial u$ and $\partial/\partial v$). Show that

$$\nabla_v(\nabla_u \mathbf{w}) - \nabla_u(\nabla_v \mathbf{w}) = (\mathbf{w}_v \cdot \mathbf{N})\mathbf{N}_u - (\mathbf{w}_u \cdot \mathbf{N})\mathbf{N}_v,$$

where $\mathbf{N}$ is the unit normal of $\boldsymbol{\sigma}$. Deduce that, if $\lambda(u, v)$ is a smooth function of $(u, v)$, then

$$\nabla_v(\nabla_u(\lambda \mathbf{w})) - \nabla_u(\nabla_v(\lambda \mathbf{w})) = \lambda \left(\nabla_v(\nabla_u \mathbf{w}) - \nabla_u(\nabla_v \mathbf{w})\right).$$

Use Proposition 8.1.2 to show that

$$\nabla_v(\nabla_u \boldsymbol{\sigma}_u) - \nabla_u(\nabla_v \boldsymbol{\sigma}_u) = K(-F\boldsymbol{\sigma}_u + E\boldsymbol{\sigma}_v),$$

where

$$K = \frac{LN - M^2}{EG - F^2},$$

and find a similar expression for $\nabla_v(\nabla_u \boldsymbol{\sigma}_v) - \nabla_u(\nabla_v \boldsymbol{\sigma}_v)$. Deduce that

$$\nabla_v(\nabla_u \mathbf{w}) = \nabla_u(\nabla_v \mathbf{w})$$

for all tangent vector fields $\mathbf{w}$ if and only if $K = 0$ everywhere on the surface. (Note that this holds for the plane: $\mathbf{w}_{uv} = \mathbf{w}_{vu}$.) We shall see the significance of the condition $K = 0$ in Section 8.4.

## 8.2 Principal curvatures of a surface

We now examine the Weingarten map $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ of a surface $\mathcal{S}$ at a point $\mathbf{p} \in \mathcal{S}$ in a little more detail (we shall usually omit the subscripts). The crucial point is that $\mathcal{W}$ is *self-adjoint* (Corollary 7.2.4). From Theorem A.0.3 we deduce the following proposition.

### Proposition 8.2.1

Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$. There are scalars $\kappa_1, \kappa_2$ and a basis $\{\mathbf{t}_1, \mathbf{t}_2\}$ of the tangent plane $T_{\mathbf{p}}\mathcal{S}$ such that

$$\mathcal{W}(\mathbf{t}_1) = \kappa_1 \mathbf{t}_1, \quad \mathcal{W}(\mathbf{t}_2) = \kappa_2 \mathbf{t}_2.$$

Moreover, if $\kappa_1 \neq \kappa_2$, then $\langle \mathbf{t}_1, \mathbf{t}_2 \rangle = 0$.

The real numbers $\kappa_1$ and $\kappa_2$ are the eigenvalues of $\mathcal{W}$, and $\mathbf{t}_1$ and $\mathbf{t}_2$ are corresponding eigenvectors. But in this situation, we adopt a special terminology: $\kappa_1$ and $\kappa_2$ are called the *principal curvatures* of $\mathcal{S}$, and $\mathbf{t}_1$ and $\mathbf{t}_2$ are called *principal vectors* corresponding to $\kappa_1$ and $\kappa_2$.

Points of the surface at which the two principal curvatures are equal (to $\kappa$, say) are called *umbilics*. At an umbilic, the equations $\mathcal{W}(\mathbf{t}_1) = \kappa \mathbf{t}_1$ and $\mathcal{W}(\mathbf{t}_2) = \kappa \mathbf{t}_2$ imply that $\mathcal{W}(\mathbf{t}) = \kappa \mathbf{t}$ if $\mathbf{t}$ is any linear combination of $\mathbf{t}_1$ and $\mathbf{t}_2$. Thus, $\mathbf{p}$ *is an umbilic if and only if* $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ *is a scalar multiple of the identity map*, and in that case every tangent vector is principal. On the other hand, if $\mathbf{p} \in \mathcal{S}$ is not an umbilic, Proposition 8.2.1 tells us that principal vectors corresponding to the two principal curvatures are necessarily orthogonal (Theorem A.0.3). Thus, whether or not $\mathbf{p}$ is an umbilic we can always find two orthogonal principal vectors in $T_{\mathbf{p}}\mathcal{S}$, and we obtain:

### Corollary 8.2.2

If $\mathbf{p}$ is a point of a surface $\mathcal{S}$, there is an orthonormal basis of the tangent plane $T_{\mathbf{p}}\mathcal{S}$ consisting of principal vectors.

The principal curvatures are related in a simple way to the mean and Gaussian curvatures:

## Proposition 8.2.3

If $\kappa_1$ and $\kappa_2$ are the principal curvatures of a surface, the mean and Gaussian curvatures are given by

$$H = \frac{1}{2}(\kappa_1 + \kappa_2), \quad K = \kappa_1 \kappa_2.$$

## Proof

The determinant and trace of the Weingarten map $\mathcal{W}$ can be computed using the matrix of $\mathcal{W}$ with respect to *any* basis of the tangent plane. Using the basis formed by the principal vectors, the matrix is

$$\begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix}.$$

The proposition now follows immediately from Definition 8.1.1. □

One reason for introducing the principal curvatures and principal vectors is contained in the following result, which shows that, if we know the principal curvatures and principal vectors of a surface, it is easy to calculate the normal curvature of any curve on the surface:

## Euler's Theorem 8.2.4

Let $\boldsymbol{\gamma}$ be a curve on an oriented surface $\mathcal{S}$, and let $\kappa_1$ and $\kappa_2$ be the principal curvatures of $\boldsymbol{\sigma}$, with non-zero principal vectors $\mathbf{t}_1$ and $\mathbf{t}_2$. Then, the normal curvature of $\boldsymbol{\gamma}$ is

$$\kappa_n = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta,$$

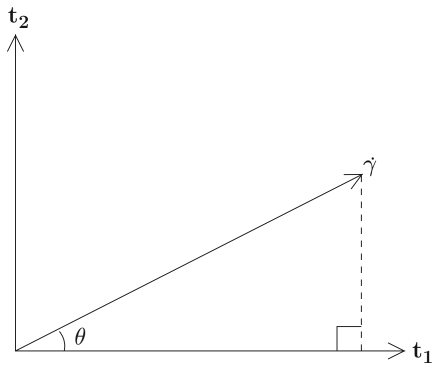
where $\theta$ is the oriented angle $\widehat{\mathbf{t}_1\dot{\boldsymbol{\gamma}}}$.

## Proof

Let $\mathbf{p} \in \mathcal{S}$, let $\kappa_1$ and $\kappa_2$ be the principal curvatures of $\mathcal{S}$ at $\mathbf{p}$, and let $\mathbf{t}_1$ and $\mathbf{t}_2$ be corresponding principal vectors. By Corollary 8.2.2, we can assume that $\{\mathbf{t}_1, \mathbf{t}_2\}$ is an orthonormal basis of $T_\mathbf{p}\mathcal{S}$. Moreover, by replacing $\mathbf{t}_2$ by $-\mathbf{t}_2$ if necessary, we can assume that the oriented angle $\widehat{\mathbf{t}_1\mathbf{t}_2} = +\pi/2$.

With these assumptions, we have

$$\dot{\boldsymbol{\gamma}} = \cos\theta\mathbf{t}_1 + \sin\theta\mathbf{t}_2.$$




By Proposition 7.3.5,

$$\kappa_n = \langle\langle \dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}} \rangle\rangle = \cos^2\theta \langle\langle \mathbf{t}_1, \mathbf{t}_1 \rangle\rangle + 2\sin\theta\cos\theta \langle\langle \mathbf{t}_1, \mathbf{t}_2 \rangle\rangle + \sin^2\theta \langle\langle \mathbf{t}_2, \mathbf{t}_2 \rangle\rangle.$$

Now, for $i, j = 1, 2$,

$$\langle\langle \mathbf{t}_i, \mathbf{t}_j \rangle\rangle = \langle \mathcal{W}(\mathbf{t}_i), \mathbf{t}_j \rangle = \langle \kappa_i \mathbf{t}_i, \mathbf{t}_j \rangle = \begin{cases} \kappa_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

Hence the result. □

## Corollary 8.2.5

The principal curvatures at a point of a surface are the maximum and minimum values of the normal curvature of all curves on the surface that pass through the point. Moreover, the principal vectors are the tangent vectors of the curves giving these maximum and minimum values.

## Proof

If the principal curvatures $\kappa_1$ and $\kappa_2$ are different, we might as well suppose that $\kappa_1 > \kappa_2$. Let $\kappa_n$ be the normal curvature of a curve $\boldsymbol{\gamma}$ on the surface. Then, since

$$\kappa_n = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta = \kappa_1 - (\kappa_1 - \kappa_2)\sin^2\theta,$$

it is clear that $\kappa_n \leq \kappa_1$ with equality if and only if $\theta = 0$ or $\pi$, i.e., if and only if the tangent vector $\dot{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ is parallel to the principal vector $\mathbf{t}_1$. Similarly, one shows that $\kappa_n \geq \kappa_2$ with equality if and only if $\dot{\boldsymbol{\gamma}}$ is parallel to $\mathbf{t}_2$.

If $\kappa_1 = \kappa_2$, the normal curvature of every curve is equal to $\kappa_1$ by Euler's Theorem and every tangent vector to the surface is a principal vector. □

To compute the principal curvatures, we work in a surface patch $\boldsymbol{\sigma}(u,v)$; let

$$Edu^2 + 2Fdudv + Gdv^2 \quad \text{and} \quad Ldu^2 + 2Mdudv + Ndv^2$$

be its first and second fundamental forms. In the notation of Section 8.1, the matrix of the Weingarten map $\mathcal{W}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of the tangent plane is $\mathcal{F}_I^{-1}\mathcal{F}_{II}$. Hence, the principal curvatures are the roots $\kappa$ of the equation

$$\det(\mathcal{F}_I^{-1}\mathcal{F}_{II} - \kappa I) = 0,$$

and a tangent vector $\mathbf{t} = \xi\boldsymbol{\sigma}_u + \eta\boldsymbol{\sigma}_v$ is a principal vector if

$$(\mathcal{F}_I^{-1}\mathcal{F}_{II} - \kappa I)\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Writing $\mathcal{F}_I^{-1}\mathcal{F}_{II} - \kappa I$ as $\mathcal{F}_I^{-1}(\mathcal{F}_{II} - \kappa\mathcal{F}_I)$, we obtain the following.

## Proposition 8.2.6

In the above notation, the principal curvatures are the roots of the equation

$$\begin{vmatrix} L-\kappa E & M-\kappa F \\ M-\kappa F & N-\kappa G \end{vmatrix} = 0,$$

and the principal vectors corresponding to the principal curvature $\kappa$ are the tangent vectors $\mathbf{t} = \xi\boldsymbol{\sigma}_u + \eta\boldsymbol{\sigma}_v$ such that

$$\begin{pmatrix} L-\kappa E & M-\kappa F \\ M-\kappa F & N-\kappa G \end{pmatrix}\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

## Example 8.2.7

It is intuitively clear that a sphere curves the same amount in every direction, and at every point of the sphere. Thus, we expect that the principal curvatures of a sphere are equal to each other at every point, and are constant over the sphere. To confirm this by calculation, we work with the unit sphere $S^2$ and use the latitude longitude parametrization as usual. We found in Example 6.1.3 that $E = 1, F = 0, G = \cos^2\theta$ and in Example 7.1.2 that $L = 1, M = 0, N = \cos^2\theta$. So the principal curvatures are the roots of

$$\begin{vmatrix} 1-\kappa & 0 \\ 0 & \cos^2\theta - \kappa\cos^2\theta \end{vmatrix} = 0,$$

i.e., $\kappa = 1$ (repeated root), as we expected. Every tangent vector is a principal vector.

## Example 8.2.8

We consider the unit cylinder parametrized in the usual way:

$$\boldsymbol{\sigma}(u, v) = (\cos v, \sin v, u).$$

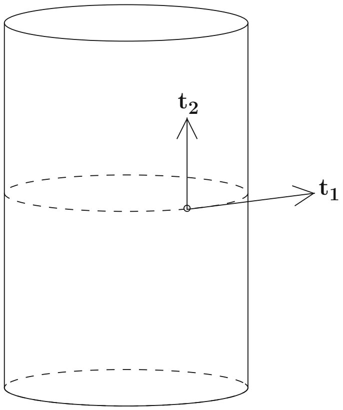
We found in Example 6.1.4 that $E = 1, F = 0, G = 1$ and in Example 7.1.2 that $L = 0, M = 0, N = 1$. So the principal curvatures are the roots of

$$\begin{vmatrix} 0 - \kappa & 0 \\ 0 & 1 - \kappa \end{vmatrix} = 0,$$

i.e., $\kappa = 0$ or 1. Any principal vector $\mathbf{t}_1$ corresponding to $\kappa_1 (= 1)$ satisfies

$$\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \eta_1 \end{pmatrix} = 0,$$

so $\xi_1 = 0$ and $\mathbf{t}_1$ is a multiple of $\boldsymbol{\sigma}_v = (-\sin v, \cos v, 0)$. Similarly, one finds that any principal vector corresponding to $\kappa_2$ $(= 0)$ is a multiple of $\boldsymbol{\sigma}_u = (0, 0, 1)$.




Example 8.2.7 proves the intuitively obvious fact that on a sphere every point is an umbilic. The same is clearly true for a plane, since in that case both principal curvatures are zero everywhere. Remarkably, there are no other surfaces with this property:

## Proposition 8.2.9

Let $\mathcal{S}$ be a (connected) surface of which every point is an umbilic. Then, $\mathcal{S}$ is an open subset of a plane or a sphere.

## Proof

For every tangent vector $\mathbf{t}$, we have $\mathcal{W}(\mathbf{t}) = \kappa\mathbf{t}$ where $\kappa$ is the principal curvature. Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a surface patch of $\mathcal{S}$ with $U$ a (connected) open subset of $\mathbb{R}^2$. Taking $\mathbf{t} = \boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ and recalling from the proof of Proposition 7.2.2 that $\mathcal{W}(\boldsymbol{\sigma}_u) = -\mathbf{N}_u, \mathcal{W}(\boldsymbol{\sigma}_v) = -\mathbf{N}_v$, we get

$$\mathbf{N}_u = -\kappa\boldsymbol{\sigma}_u, \quad \mathbf{N}_v = -\kappa\boldsymbol{\sigma}_v. \tag{8.3}$$

Hence,

$$(\kappa\boldsymbol{\sigma}_u)_v = -(\mathbf{N}_u)_v = -(\mathbf{N}_v)_u = (\kappa\boldsymbol{\sigma}_v)_u,$$

so

$$\kappa_v\boldsymbol{\sigma}_u = \kappa_u\boldsymbol{\sigma}_v.$$

Since $\boldsymbol{\sigma}$ is regular, $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent, so the last equation implies that $\kappa_u = \kappa_v = 0$. Thus, $\kappa$ is constant.

There are now two cases to consider. If $\kappa = 0$, Eqs. 8.3 show that $\mathbf{N}$ is constant. Then,

$$(\mathbf{N}\cdot\boldsymbol{\sigma})_u = \mathbf{N}\cdot\boldsymbol{\sigma}_u = 0, \quad (\mathbf{N}\cdot\boldsymbol{\sigma})_v = \mathbf{N}\cdot\boldsymbol{\sigma}_v = 0,$$

so $\mathbf{N}\cdot\boldsymbol{\sigma}$ is a constant, say $c$. Then $\boldsymbol{\sigma}(U)$ is an open subset of the plane $\mathbf{v}\cdot\mathbf{N} = c$.

If $\kappa \neq 0$, Eq. 8.3 shows that

$$\mathbf{N} = -\kappa\boldsymbol{\sigma} + \mathbf{a},$$

where $\mathbf{a}$ is a constant vector. Hence,

$$\left\|\boldsymbol{\sigma} - \frac{1}{\kappa}\mathbf{a}\right\|^2 = \left\|-\frac{1}{\kappa}\mathbf{N}\right\|^2 = \frac{1}{\kappa^2},$$

so $\boldsymbol{\sigma}(U)$ is an open subset of the sphere with centre $\kappa^{-1}\mathbf{a}$ and radius $\kappa^{-1}$.

We have now proved the proposition when $\mathcal{S}$ is covered by a single surface patch. For an arbitrary surface $\mathcal{S}$, the preceding argument shows that each patch in the atlas of $\mathcal{S}$ is contained in a plane or a sphere. But if the images of two patches intersect they must clearly be part of the same plane or the same sphere. It follows that the whole of $\mathcal{S}$ is contained in a plane or a sphere. $\square$

Note that this proposition is an analogue for surfaces of Example 2.2.7, which tells us that a plane curve with constant curvature is part of a circle.

We conclude this section by showing how the values of the principal curvatures at a point $\mathbf{p}$ of a surface $\mathcal{S}$ provide information about the shape of $\mathcal{S}$ near $\mathbf{p}$. To simplify the situation, we assume that $\mathbf{p}$ is the origin and that $T_{\mathbf{p}}\mathcal{S}$ is the $xy$-plane: this can be arranged by applying a suitable isometry of $\mathbb{R}^3$ to $\mathcal{S}$ (which does not change its shape). By a further rotation around the $z$-axis, we can also assume that the tangent vectors $\mathbf{t}_1 = (1, 0, 0)$ and $\mathbf{t}_2 = (0, 1, 0)$ are principal, and correspond to principal curvatures $\kappa_1$ and $\kappa_2$. Finally, by reflecting in the $xy$-plane if necessary, we can assume that the unit normal of $\mathcal{S}$ at $\mathbf{p}$ is $\mathbf{N} = (0, 0, 1)$.

Let $\boldsymbol{\sigma}$ be a surface patch of $\mathcal{S}$ with $\boldsymbol{\sigma}(0, 0) = \mathbf{0}$. For any $x, y \in \mathbb{R}$, there are unique $s, t \in \mathbb{R}$ such that

$$(x, y, 0) = s\boldsymbol{\sigma}_u + t\boldsymbol{\sigma}_v$$

(here and below, the derivatives of $\boldsymbol{\sigma}$ are evaluated at $(0, 0)$). By Taylor's theorem,

$$\boldsymbol{\sigma}(s, t) = \boldsymbol{\sigma}(0, 0) + s\boldsymbol{\sigma}_u + t\boldsymbol{\sigma}_v + \frac{1}{2}(s^2\boldsymbol{\sigma}_{uu} + 2st\boldsymbol{\sigma}_{uv} + t^2\boldsymbol{\sigma}_{vv})$$

if we neglect terms involving higher powers of $s$ and $t$. Hence, if $x$ and $y$ (and hence $s$ and $t$) are small, we have $\boldsymbol{\sigma}(s, t) = (x, y, z)$, where

$$z = \frac{1}{2}(s^2\boldsymbol{\sigma}_{uu} + 2st\boldsymbol{\sigma}_{uv} + t^2\boldsymbol{\sigma}_{vv}) \cdot \mathbf{N} = \frac{1}{2}(Ls^2 + 2Mst + Nt^2)$$

approximately, where $Ldu^2 + 2Mdudv + Ndv^2$ is the second fundamental form of $\boldsymbol{\sigma}$ at the origin. If $\mathbf{t} = s\boldsymbol{\sigma}_u + t\boldsymbol{\sigma}_v$, then by Proposition 7.3.3,

$$Ls^2 + 2Mst + Nt^2 = \langle\langle \mathbf{t}, \mathbf{t} \rangle\rangle = \langle \mathcal{W}(\mathbf{t}), \mathbf{t} \rangle.$$

Now, $\mathbf{t} = x\mathbf{t}_1 + y\mathbf{t}_2$ so

$$\mathcal{W}(\mathbf{t}) = x\mathcal{W}(\mathbf{t}_1) + y\mathcal{W}(\mathbf{t}_2) = \kappa_1 x\mathbf{t}_1 + \kappa_2 y\mathbf{t}_2 = (\kappa_1 x, \kappa_2 y, 0).$$

Hence,

$$Ls^2 + 2Mst + Nt^2 = (\kappa_1 x, \kappa_2 y, 0) \cdot (x, y, 0) = \kappa_1 x^2 + \kappa_2 y^2.$$

Hence, near the point $\mathbf{p}$, $\mathcal{S}$ is approximated by the quadric surface

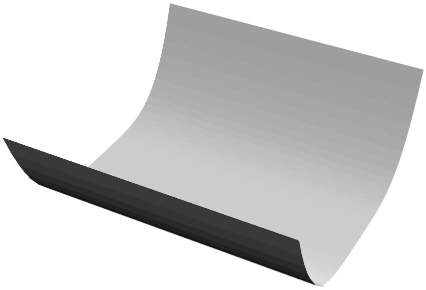
$$z = \frac{1}{2}(\kappa_1 x^2 + \kappa_2 y^2). \tag{8.4}$$

We distinguish four cases:

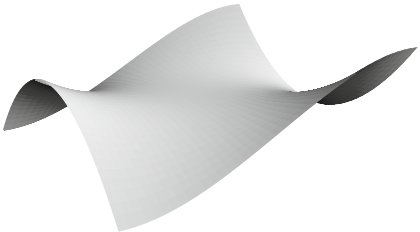
(i) *$\kappa_1$ and $\kappa_2$ are both $> 0$ or both $< 0$.* Then, (8.4) is the equation of an elliptic paraboloid (see Theorem 5.2.2) and one says that $\mathbf{p}$ is an *elliptic point* of the surface.

(ii) *$\kappa_1$ and $\kappa_2$ are of opposite sign (both non-zero).* Then, (8.4) is the equation of a hyperbolic paraboloid and one says that $\mathbf{p}$ is a *hyperbolic point* of the surface.

(iii) *One of* $\kappa_1$ *and* $\kappa_2$ *is zero, the other is non-zero.* Then, (8.4) is the equation of a parabolic cylinder and one says that $\mathbf{p}$ is a *parabolic point* of the surface.



$z = y^4$



$z = x^3 - 3xy^2$

(iv) *Both principal curvatures are zero at* $\boldsymbol{p}$. Then, (8.4) is the equation of a plane, and one says that $\mathbf{p}$ is a *planar point* of the surface. In this case, one cannot determine the shape of the surface near $\mathbf{p}$ without examining derivatives of order higher than the second (in the non-planar case, these terms are small compared to $\kappa_1 x^2 + \kappa_2 y^2$ when $x$ and $y$ are small). For example, the surfaces above both have the origin as a planar point, but they have quite different shapes. (The surface on the right is called the *monkey saddle* as it is the right shape for the saddle on a bicycle ridden by a monkey: two ways down for the two legs and a third for the tail.)

The classification of points of a surface as elliptic, hyperbolic, parabolic or planar is independent of the surface patch $\boldsymbol{\sigma}$, since reparametrizing either leaves the principal curvatures unchanged or changes the sign of both of them (Exercise 8.2.8).

## Example 8.2.10

On $S^2$, $\kappa_1 = \kappa_2 = \pm 1$ (the sign depending on the parametrization) so all points are elliptic (and umbilics). On a circular cylinder, $\kappa_1 = \pm 1, \kappa_2 = 0$, so every point is parabolic (and there are no umbilics). On a plane, $\kappa_1 = \kappa_2 = 0$ so all points are planar (!) (and umbilics).

## Example 8.2.11

For the torus $\boldsymbol{\sigma}(\theta, \varphi) = ((a+b\cos\theta)\cos\varphi, (a+b\cos\theta)\sin\varphi, b\sin\theta)$ (see Exercise 4.2.5), we find that the first and second fundamental forms are

$$b^2 d\theta^2 + (a + b\cos\theta)^2 d\varphi^2 \quad \text{and} \quad b\, d\theta^2 + (a + b\cos\theta)\cos\theta\, d\varphi^2,$$

respectively, so the principal curvatures are

$$\kappa_1 = \frac{1}{b}, \quad \kappa_2 = \frac{\cos\theta}{a + b\cos\theta}.$$

Since $\kappa_1 > 0$ (everywhere), the point $\boldsymbol{\sigma}(\theta, \varphi)$ of the torus is elliptic, parabolic or hyperbolic according to $\kappa_2$ is $> 0$, $= 0$ or $< 0$, respectively; from the formula for $\kappa_2$, these are the regions of the torus given by $-\pi/2 < \theta < \pi/2$, $\theta = \pm\pi/2$ and $\pi/2 < \theta < 3\pi/2$, respectively. Pictures of the elliptic and hyperbolic regions can be found in the solution to Exercise 8.1.9 (where they are labelled $\mathcal{S}^+$ and $\mathcal{S}^-$, respectively); the parabolic region consists of two circles of radius $a$ centred on the $z$-axis.

## *EXERCISES*

8.2.1 Calculate the principal curvatures of the helicoid and the catenoid, defined in Exercises 4.2.6 and 5.3.1, respectively.

8.2.2 A curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$ is called a *line of curvature* if the tangent vector of $\boldsymbol{\gamma}$ is a principal vector of $\mathcal{S}$ at all points of $\boldsymbol{\gamma}$ (a 'line' of curvature need not be a straight line!). Show that $\boldsymbol{\gamma}$ is a line of curvature if and only if

$$\dot{\mathbf{N}} = -\lambda\dot{\boldsymbol{\gamma}},$$

for some scalar $\lambda$, where $\mathbf{N}$ is the standard unit normal of $\boldsymbol{\sigma}$, and that in this case the corresponding principal curvature is $\lambda$. (This is called *Rodrigues' formula*.)

8.2.3 Show that a curve $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ on a surface patch $\boldsymbol{\sigma}$ is a line of curvature if and only if (in the usual notation)

$$(EM - FL)\dot{u}^2 + (EN - GL)\dot{u}\dot{v} + (FN - GM)\dot{v}^2 = 0.$$

Deduce that all parameter curves are lines of curvature if and only if either

(i) the second fundamental form of $\boldsymbol{\sigma}$ is proportional to its first fundamental form, or

(ii) $F = M = 0$.

For which surfaces does (i) hold? Show that the meridians and parallels of a surface of revolution are lines of curvature.

8.2.4 In the notation of Example 8.1.5, show that if $\boldsymbol{\gamma}$ is a curve on a surface $\mathcal{S}$ and $\boldsymbol{\delta}$ is the unit normal of $\mathcal{S}$, then $K = 0$ if and only if $\boldsymbol{\gamma}$ is a line of curvature of $\mathcal{S}$.

8.2.5 Suppose that two surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ intersect in a curve $\mathcal{C}$ that is a line of curvature of $\mathcal{S}_1$. Show that $\mathcal{C}$ is a line of curvature of $\mathcal{S}_2$ if and only if the angle between the tangent planes of $\mathcal{S}_1$ and $\mathcal{S}_2$ is constant along $\mathcal{C}$.

8.2.6 Let $\mathbf{\Sigma} : W \to \mathbb{R}^3$ be a smooth function defined on an open subset $W$ of $\mathbb{R}^3$ such that, for each fixed value of $u$ (resp. $v$, $w$), $\mathbf{\Sigma}(u, v, w)$ is a (regular) surface patch. Assume also that

$$\mathbf{\Sigma}_u \cdot \mathbf{\Sigma}_v = \mathbf{\Sigma}_v \cdot \mathbf{\Sigma}_w = \mathbf{\Sigma}_w \cdot \mathbf{\Sigma}_u = 0. \tag{8.5}$$

This means that the three families of surfaces formed by fixing the values of $u$, $v$ or $w$ constitute a triply orthogonal system (see Section 5.5).

(i) Show that $\mathbf{\Sigma}_u \cdot \mathbf{\Sigma}_{vw} = \mathbf{\Sigma}_v \cdot \mathbf{\Sigma}_{uw} = \mathbf{\Sigma}_w \cdot \mathbf{\Sigma}_{uv} = 0$.

(ii) Show that, for each of the surfaces in the triply orthogonal system, the matrices $\mathcal{F}_I$ and $\mathcal{F}_{II}$ are diagonal.

(iii) Deduce that the intersection of any surface from one family of the triply orthogonal system with any surface from another family is a line of curvature on both surfaces. (This is called *Dupin's Theorem.*)

8.2.7 Show that, if $p$, $q$ and $r$ are distinct positive numbers, there are exactly four umbilics on the ellipsoid

$$\frac{x^2}{p^2} + \frac{y^2}{q^2} + \frac{z^2}{r^2} = 1.$$

What happens if $p, q$ and $r$ are not distinct?

8.2.8 Show that the principal curvatures of a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ are smooth functions on $U$ provided that $\boldsymbol{\sigma}$ has no umbilics. Show also that the principal curvatures either stay the same or both change sign when $\boldsymbol{\sigma}$ is reparametrized.

## 8.3 Surfaces of constant Gaussian curvature

We have seen in the examples in Section 8.1 some surfaces of zero and constant positive curvature. For an example of a surface with *constant negative* Gaussian curvature, however, we have to construct a new surface. To this end, we examine again the surface of revolution

$$\boldsymbol{\sigma}(u, v) = (f(u)\cos v, f(u)\sin v, g(u))$$

obtained by rotating the unit-speed curve $u \mapsto (f(u), 0, g(u))$ in the $xz$-plane around the $z$-axis. We found in Example 8.1.4 that its Gaussian curvature is

$$K = -\frac{\ddot{f}}{f}. \tag{8.6}$$

*Suppose first that* $K = 0$ *everywhere.* Then, Eq. 8.6 gives $\ddot{f} = 0$, so $f(u) = au + b$ for some constants $a$ and $b$. Since $\dot{f}^2 + \dot{g}^2 = 1$, we get $\dot{g} = \pm\sqrt{1 - a^2}$ (so we must have $|a| \leq 1$) and hence $g(u) = \pm\sqrt{1 - a^2}u + c$, where $c$ is another constant. By applying a translation along the $z$-axis we can assume that $c = 0$, and by applying a rotation by $\pi$ about the $x$-axis, if necessary, we can assume that the sign is $+$. This gives the ruled surface

$$\boldsymbol{\sigma}(u, v) = (b\cos v, b\sin v, 0) + u(a\cos v, a\sin v, \sqrt{1 - a^2}).$$

If $a = 0$ this is a circular cylinder; if $|a| = 1$ it is the $xy$-plane; and if $0 < |a| < 1$ it is a circular cone (to see this, put $\tilde{u} = au + b$).

*Now suppose that* $K > 0$*, say* $K = 1/R^2$*, where* $R > 0$ *is a constant.* Then, Eq. 8.6 becomes

$$\ddot{f} + \frac{f}{R^2} = 0,$$

which has the general solution

$$f(u) = a\cos\left(\frac{u}{R} + b\right),$$

where $a$ and $b$ are constants. We can assume that $b = 0$ by performing a reparametrization $\tilde{u} = u + Rb, \tilde{v} = v$. Then, up to a change of sign and adding a constant,

$$g(u) = \int \sqrt{1 - \frac{a^2}{R^2}\sin^2\frac{u}{R}}\, du.$$

The integral in the formula for $g(u)$ can be evaluated in terms of 'elementary' functions only when $a = 0$ or $\pm R$. The case $a = 0$ does not give a surface, and if $a = R$ then $f(u) = R\cos\frac{u}{R}$, $g(u) = R\sin\frac{u}{R}$, and we have a sphere of radius $R$ (the case $a = -R$ can be reduced to this by rotating the surface by $\pi$ around the $z$-axis).

*Suppose finally that* $K < 0$. We can restrict ourselves to the case $K = -1$, as the general case can be obtained from this by applying a dilation of $\mathbb{R}^3$ (see Exercise 8.1.5). In view of the preceding case, we can think of a surface with $K = -1$ as a 'sphere of imaginary radius' $\sqrt{-1}$, or a 'pseudosphere'.

When $K = -1$ the general solution of Eq. 8.6 is

$$f(u) = ae^u + be^{-u},$$

where $a$ and $b$ are arbitrary constants. The function $g(u)$ can be expressed in terms of elementary functions only if one of $a$ or $b$ is zero. If $b = 0$ we can assume that $a = 1$ by a reparametrization $u \mapsto u + \text{constant}$, and the case in which $a = 0$ can be reduced to the case $b = 0$ by the reparametrization $u \mapsto -u$. Suppose then that $a = 1$ and $b = 0$; then, $f(u) = e^u$ and we can take

$$g(u) = \int \sqrt{1 - e^{2u}}\, du. \tag{8.7}$$

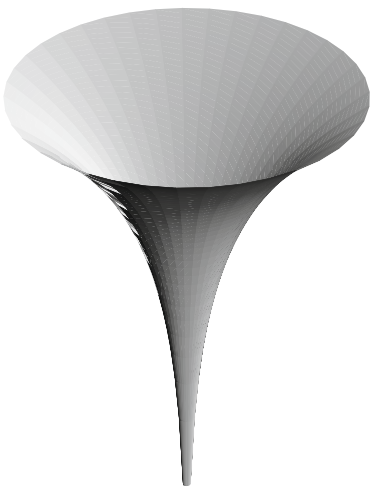
Note that we must have $u \leq 0$ for the integral in Eq. 8.7 to make sense, since otherwise $1 - e^{2u}$ would be negative. The integral can be evaluated by putting $\cos\theta = e^u$. Then,

$$\begin{aligned}\int \sqrt{1 - e^{2u}}\, du &= -\int \frac{\sin^2\theta}{\cos\theta}\, d\theta = \sin\theta - \ln(\sec\theta + \tan\theta) \\ &= \sqrt{1 - e^{2u}} - \ln(e^{-u} + \sqrt{e^{-2u} - 1}).\end{aligned}$$

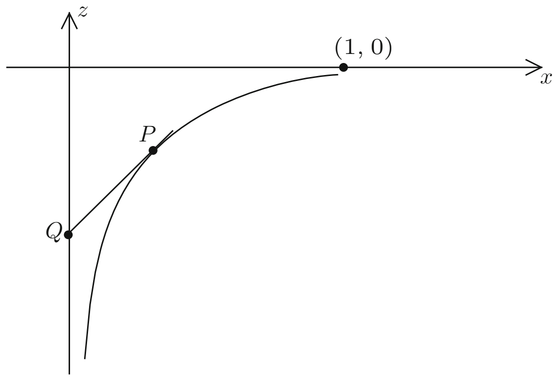
We have omitted the arbitrary constant, but we can take it to be zero by a suitable translation of the surface parallel to the $z$-axis. Putting $x = f(u)$, $z = g(u)$, and noting that $\cosh^{-1}(v) = \ln(v + \sqrt{v^2 - 1})$, we see that the profile curve in the $xz$-plane has equation

$$z = \sqrt{1 - x^2} - \cosh^{-1}\left(\frac{1}{x}\right). \tag{8.8}$$

Rotating this curve around the $z$ -axis thus gives a surface which has Gaussian curvature $-1$ everywhere. Note that, since $u \leq 0$, $x = e^u$ is restricted to the range $0 < x \leq 1$.



The curve defined by Eq. 8.8 is called the *tractrix*, and it has an interesting geometrical property. Consider the tangent line at a point $P$ of its graph, and suppose that it intersects the $z$-axis at the point $Q$. Let us compute the distance from $P$ to $Q$.




Suppose that $P$ is the point $(x_0, z_0)$. Either by a direct calculation or by inspecting the calculation of the integral (8.7), one finds that

$$\frac{dz}{dx} = \frac{\sqrt{1-x^2}}{x}.$$

Hence, the tangent line at $P$ has equation

$$z - z_0 = \frac{\sqrt{1-x_0^2}}{x_0}(x - x_0).$$

This meets the $z$-axis at the point $(0, z_1)$, where

$$z_1 - z_0 = \frac{\sqrt{1-x_0^2}}{x_0}(0 - x_0) = -\sqrt{1-x_0^2}.$$

Hence, the square of the distance from $P$ to $Q$ is

$$x_0^2 + (z_1 - z_0)^2 = x_0^2 + 1 - x_0^2 = 1,$$

so the distance from $P$ to $Q$ is *constant* and equal to 1.

This means that the tractrix has the following description. Let a donkey pull a box of stones by a rope of length 1. Suppose that the donkey is initially at $(0, 0)$, the box is initially at $(1, 0)$, and let the donkey walk slowly along the negative $z$-axis. Then, the box of stones moves along the tractrix.

## *EXERCISES*

8.3.1 Show that:

(i) Setting $w = e^{-u}$ gives a reparametrization $\boldsymbol{\sigma}_1(v, w)$ of the pseudosphere with first fundamental form

$$\frac{dv^2 + dw^2}{w^2}$$

(called the *upper half-plane model*).

(ii) Setting

$$V = \frac{v^2 + w^2 - 1}{v^2 + (w+1)^2}, \quad W = \frac{-2v}{v^2 + (w+1)^2}$$

defines a reparametrization $\boldsymbol{\sigma}_2(V, W)$ of the pseudosphere with first fundamental form

$$\frac{4(dV^2 + dW^2)}{(1 - V^2 - W^2)^2}$$

(called the *Poincaré disc model*: the region $w > 0$ of the $vw$-plane corresponds to the disc $V^2 + W^2 < 1$ in the $VW$-plane).

(iii) Setting

$$\bar{V} = \frac{2V}{V^2 + W^2 + 1}, \quad \bar{W} = \frac{2W}{V^2 + W^2 + 1}$$

defines a reparametrization $\boldsymbol{\sigma}_2(\bar{V}, \bar{W})$ of the pseudosphere with first fundamental form

$$\frac{(1 - \bar{W}^2)d\bar{V}^2 + 2\bar{V}\bar{W}d\bar{V}d\bar{W} + (1 - \bar{V}^2)d\bar{W}^2}{(1 - \bar{V}^2 - \bar{W}^2)^2}$$

(called the *Beltrami-Klein model*: the region $w > 0$ of the $vw$-plane again corresponds to the disc $\bar{V}^2 + \bar{W}^2 < 1$ in the $\bar{V}\bar{W}$-plane).

In cases (i) and (ii), find the open subsets of the $vw$- and $VW$-plane, respectively, corresponding to the open set

$$\{(u, v) \mid u < 0, -\pi < v < \pi\}$$

in the parametrization of the pseudosphere given in the text.

These models are discussed in much more detail in Chapter 11.

## 8.4 Flat surfaces

In Section 8.3, we gave some examples of surfaces of constant Gaussian curvature $K$, but this certainly falls well short of a complete classification of such surfaces. It is possible, however, to give a fairly complete description of *flat surfaces*, i.e., surfaces for which $K = 0$ everywhere. To do so, we shall make use of a special parametrization, valid for any surface, described in the following proposition.

### Proposition 8.4.1

Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$, and suppose that $\mathbf{p}$ is not an umbilic. Then, there is a surface patch $\boldsymbol{\sigma}(u,v)$ of $\mathcal{S}$ containing $\mathbf{p}$ whose first and second fundamental forms are

$$Edu^2 + Gdv^2 \quad \text{and} \quad Ldu^2 + Ndv^2,$$

respectively, for some smooth functions $E$,$G$,$L$ and $N$.

We recall that a point $\mathbf{p}$ of a surface $\mathcal{S}$ is an umbilic if the two principal curvatures of $\mathcal{S}$ at $\mathbf{p}$ are equal. From Section 8.2, we see that for the patch $\boldsymbol{\sigma}$ in the statement of the proposition, $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are principal vectors with corresponding principal curvatures $L/E$ and $N/G$. We call $\boldsymbol{\sigma}$ a *principal patch*.

We assume Proposition 8.4.1 for the moment, and use it to give the proof of

### Proposition 8.4.2

Let $\mathbf{p}$ be a point of a flat surface $\mathcal{S}$, and assume that $\mathbf{p}$ is not an umbilic. Then, there is a patch of $\mathcal{S}$ containing $\mathbf{p}$ that is a ruled surface.

### Proof

We take a principal patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ containing $\mathbf{p}$ as in Proposition 8.4.1, say $\mathbf{p} = \boldsymbol{\sigma}(u_0, v_0)$. By Corollary 8.1.3, the Gaussian curvature $K = LN/EG$. Since the Gaussian curvature is zero everywhere, either $L = 0$ or $N = 0$ at each point of $U$, and since $\mathbf{p}$ is not an umbilic $L$ and $N$ are not both zero. Suppose that $L(u_0, v_0) \neq 0$, say. Then, $L(u,v) \neq 0$ for $(u,v)$ in some open subset of $U$ containing $(u_0, v_0)$. Hence, by shrinking $U$ if necessary, we can assume that $L \neq 0$ at every point of $U$. Then, $N = 0$ everywhere, and the second fundamental form of $\boldsymbol{\sigma}$ is $Ldu^2$.

We shall prove that the parameter curves $u =$ constant are straight lines. Such a curve can be parametrized by $v \mapsto \boldsymbol{\sigma}(u_0, v)$, where $u_0$ is the constant value of $u$. A unit tangent vector to this curve is $\mathbf{t} = \boldsymbol{\sigma}_v/G^{1/2}$, so by Proposition 1.1.6 what we have to prove is that $\mathbf{t}_v = \mathbf{0}$.

By the proof of Proposition 8.1.2, the derivatives of the unit normal are

$$\mathbf{N}_u = -E^{-1}L\boldsymbol{\sigma}_u, \quad \mathbf{N}_v = \mathbf{0}. \tag{8.9}$$

Hence, $\mathbf{t}_v \cdot \boldsymbol{\sigma}_u = -EL^{-1}\mathbf{t}_v \cdot \mathbf{N}_u$. Now, $\mathbf{t} \cdot \mathbf{N}_u = 0$ and $\mathbf{N}_{uv} = \mathbf{0}$ by Eq. 8.9, so $\mathbf{t}_v \cdot \mathbf{N}_u = -\mathbf{t} \cdot \mathbf{N}_{uv} = 0$. Hence, $\mathbf{t}_v \cdot \boldsymbol{\sigma}_u = 0$. Next, $\mathbf{t}_v \cdot \mathbf{t} = 0$ since $\mathbf{t}$ is a unit vector by construction, so $\mathbf{t}_v \cdot \boldsymbol{\sigma}_v = 0$. Finally, $\mathbf{t}_v \cdot \mathbf{N} = -\mathbf{t} \cdot \mathbf{N}_v = 0$ by Eq. 8.9 again. Since the vectors $\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v$ and $\mathbf{N}$ form a basis of $\mathbb{R}^3$, we have proved that $\mathbf{t}_v = \mathbf{0}$. □

Our task, then, is to describe the structure of flat ruled surfaces. We parametrize the ruled surface as in Example 8.1.5:

$$\boldsymbol{\sigma}(u, v) = \boldsymbol{\gamma}(u) + v\boldsymbol{\delta}(u).$$

We found there that $\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}$, $\boldsymbol{\sigma}_v = \boldsymbol{\delta}$, the dot denoting $d/du$, and that the Gaussian curvature of $\boldsymbol{\sigma}$ is zero if and only if

$$\dot{\boldsymbol{\delta}} \cdot (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) = 0.$$

Since

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \dot{\boldsymbol{\gamma}} \times \boldsymbol{\delta} + v\dot{\boldsymbol{\delta}} \times \boldsymbol{\delta},$$

and $\dot{\boldsymbol{\delta}} \cdot (\dot{\boldsymbol{\delta}} \times \boldsymbol{\delta}) = 0$,

$$K = 0 \quad \text{if and only if} \quad \dot{\boldsymbol{\delta}} \cdot (\dot{\boldsymbol{\gamma}} \times \boldsymbol{\delta}) = 0. \tag{8.10}$$

Thus, $K = 0$ *if and only if* $\dot{\boldsymbol{\gamma}}$, $\boldsymbol{\delta}$ *and* $\dot{\boldsymbol{\delta}}$ *are everywhere linearly dependent.*

To proceed further, let us assume, as we may, that $\boldsymbol{\delta}(u)$ is a unit vector for all values of $u$. Then, $\boldsymbol{\delta} \cdot \dot{\boldsymbol{\delta}} = 0$. Suppose first that $\dot{\boldsymbol{\delta}}(u) = \mathbf{0}$ for all values of $u$. Then, $\boldsymbol{\delta}$ is a constant vector and $\boldsymbol{\sigma}$ is a *generalized cylinder.*

Suppose now that $\dot{\boldsymbol{\delta}}$ is never zero. Then, $\boldsymbol{\delta}$ and $\dot{\boldsymbol{\delta}}$ are linearly independent as they are non-zero and perpendicular, so if $\dot{\boldsymbol{\gamma}}, \boldsymbol{\delta}$ and $\dot{\boldsymbol{\delta}}$ are linearly dependent, then

$$\dot{\boldsymbol{\gamma}}(u) = f(u)\boldsymbol{\delta}(u) + g(u)\dot{\boldsymbol{\delta}}(u)$$

for some smooth functions $f$ and $g$. Assume first that $f = \dot{g}$ everywhere. Then, $\dot{\boldsymbol{\gamma}} = (g\boldsymbol{\delta})^{\cdot}$ and so $\boldsymbol{\gamma} = g\boldsymbol{\delta} + \mathbf{a}$, where $\mathbf{a}$ is a constant vector; hence,

$$\boldsymbol{\sigma}(u, v) = \mathbf{a} + (v + g(u))\boldsymbol{\delta}(u).$$

Putting $\tilde{u} = u$, $\tilde{v} = v + g(u)$, we see that this is a reparametrization of a *generalized cone.*

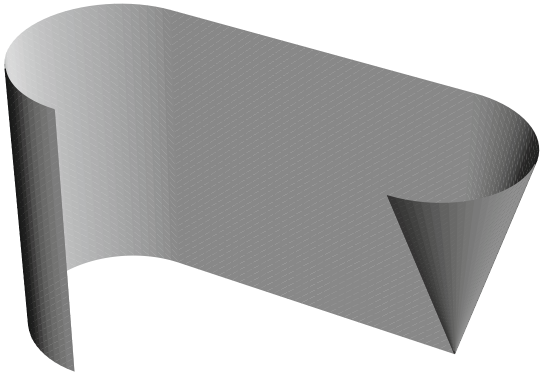
Suppose finally that $\dot{\boldsymbol{\delta}}$ and $f - \dot{g}$ are both nowhere zero. If we define

$$\tilde{\boldsymbol{\gamma}}(u) = \boldsymbol{\gamma}(u) - g(u)\boldsymbol{\delta}(u), \quad \tilde{v} = \frac{v + g(u)}{f(u) - \dot{g}(u)},$$

a short calculation gives

$$\boldsymbol{\sigma}(u, v) = \tilde{\boldsymbol{\gamma}}(u) + \tilde{v}\dot{\tilde{\boldsymbol{\gamma}}}(u),$$

so $\boldsymbol{\sigma}$ is a reparametrization of an open subset of the *tangent developable* of $\tilde{\boldsymbol{\gamma}}$.



Of course, it could be that none of the conditions on $\boldsymbol{\delta}$, $f$ and $g$ considered above are satisfied. In fact, we have only shown that certain open subsets of the surface are parts of generalized cylinders, generalized cones or tangent developables. It is not true that the whole surface must be one of these three types, since flat surfaces of different types can be joined together to make a smooth surface, as shown in the diagram above. It can be shown that the most general flat surface is a patchwork consisting of pieces of generalized cylinders, generalized cones and tangent developables, joined together along segments of straight lines.

The remainder of this section is devoted to the proof of Proposition 8.4.1 and can safely be omitted by readers who are uncomfortable with the use of the inverse function theorem. In fact, we can prove a more general result with no additional effort:

## Proposition 8.4.3

Let $\tilde{\boldsymbol{\sigma}} : \tilde{U} \to \mathbb{R}^3$ be a surface patch, and suppose that for all $(\tilde{u}, \tilde{v}) \in \tilde{U}$ we are given tangent vectors

$$\boldsymbol{e}_1(\tilde{u}, \tilde{v}) = a(\tilde{u}, \tilde{v})\tilde{\boldsymbol{\sigma}}_{\tilde{u}} + b(\tilde{u}, \tilde{v})\tilde{\boldsymbol{\sigma}}_{\tilde{v}}, \quad \boldsymbol{e}_2(\tilde{u}, \tilde{v}) = c(\tilde{u}, \tilde{v})\tilde{\boldsymbol{\sigma}}_{\tilde{u}} + d(\tilde{u}, \tilde{v})\tilde{\boldsymbol{\sigma}}_{\tilde{v}},$$

whose components $a, b, c, d$ are smooth functions of $(\tilde{u}, \tilde{v})$. Assume that, at some point $(\tilde{u}_0, \tilde{v}_0) \in \tilde{U}$, the vectors $\boldsymbol{e}_1(\tilde{u}_0, \tilde{v}_0)$ and $\boldsymbol{e}_2(\tilde{u}_0, \tilde{v}_0)$ are linearly independent. Then, there is an open subset $\tilde{V}$ of $\tilde{U}$ containing $(\tilde{u}_0, \tilde{v}_0)$ and a reparametrization $\boldsymbol{\sigma}(u, v)$ of $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$, for $(\tilde{u}, \tilde{v}) \in \tilde{V}$, such that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are parallel to $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, respectively.

Proposition 8.4.1 is a special case of Proposition 8.4.3. In fact, let $\tilde{\boldsymbol{\sigma}}$ be any surface patch of $\mathcal{S}$ containing $\mathbf{p}$, and let $\mathbf{p} = \tilde{\boldsymbol{\sigma}}(\tilde{u}_0, \tilde{v}_0)$. Since the principal curvatures $\kappa_1$ and $\kappa_2$ of $\tilde{\boldsymbol{\sigma}}$ are distinct at $\mathbf{p}$, and are continuous functions by Exercise 8.2.8, they remain distinct for $(\tilde{u}, \tilde{v})$ in some open set $\tilde{U}$ containing $(\tilde{u}_0, \tilde{v}_0)$ on which $\tilde{\boldsymbol{\sigma}}$ is defined. Let

$$\boldsymbol{e}_1 = \xi_1 \tilde{\boldsymbol{\sigma}}_{\tilde{u}} + \eta_1 \tilde{\boldsymbol{\sigma}}_{\tilde{v}}, \quad \boldsymbol{e}_2 = \xi_2 \tilde{\boldsymbol{\sigma}}_{\tilde{u}} + \eta_2 \tilde{\boldsymbol{\sigma}}_{\tilde{v}}$$

be unit principal vectors corresponding to $\kappa_1$ and $\kappa_2$; they are perpendicular by Proposition 8.2.1. Let $\boldsymbol{\sigma}(u, v)$ be a reparametrization of $\tilde{\boldsymbol{\sigma}}$ as in Proposition 8.4.3. Then, $\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0$ because $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are perpendicular, so the first fundamental form of $\boldsymbol{\sigma}$ is of the form $Edu^2 + Gdv^2$. Also, $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are principal vectors corresponding to $\kappa_1$ and $\kappa_2$, so we have

$$(\mathcal{F}_{II} - \kappa_1 \mathcal{F}_I) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (\mathcal{F}_{II} - \kappa_2 \mathcal{F}_I) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where $\mathcal{F}_I$ and $\mathcal{F}_{II}$ are the matrices associated to the first and second fundamental forms of $\boldsymbol{\sigma}$. Since $\mathcal{F}_I = \begin{pmatrix} E & 0 \\ 0 & G \end{pmatrix}$, these equations imply that $\mathcal{F}_{II} = \begin{pmatrix} \kappa_1 E & 0 \\ 0 & \kappa_2 G \end{pmatrix}$, so the second fundamental form of $\boldsymbol{\sigma}$ is $Ldu^2 + Ndv^2$, where $L = \kappa_1 E$ and $N = \kappa_2 G$.

We are thus left with the proof of Proposition 8.4.3. To begin, we observe that, if

$$\boldsymbol{e} = A \tilde{\boldsymbol{\sigma}}_{\tilde{u}} + B \tilde{\boldsymbol{\sigma}}_{\tilde{v}},$$

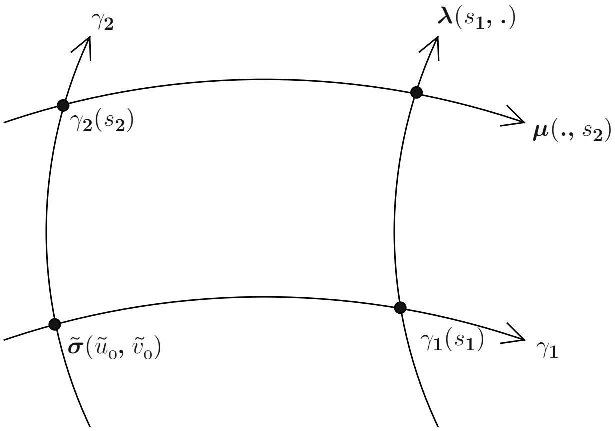
where $A$ and $B$ are any given smooth functions of $(\tilde{u}, \tilde{v}) \in \tilde{U}$, we can find a curve $\boldsymbol{\gamma}$ in $\tilde{\boldsymbol{\sigma}}$ with $\dot{\boldsymbol{\gamma}} = \boldsymbol{e}$ and with any given point $\mathbf{q} = \tilde{\boldsymbol{\sigma}}(\alpha, \beta)$ as starting point $\boldsymbol{\gamma}(0)$. For, finding such a curve $\boldsymbol{\gamma}(t) = \tilde{\boldsymbol{\sigma}}(\tilde{u}(t), \tilde{v}(t))$ is equivalent to solving the pair of ordinary differential equations

$$\dot{\tilde{u}} = A(\tilde{u}, \tilde{v}), \quad \dot{\tilde{v}} = B(\tilde{u}, \tilde{v})$$

with initial conditions $\tilde{u}(0) = \alpha$, $\tilde{v}(0) = \beta$. It is proved in the theory of ordinary differential equations that this problem has a unique solution $\tilde{u}(t), \tilde{v}(t)$ defined on some open interval containing $t = 0$. Moreover, $\tilde{u}$ and $\tilde{v}$ are smooth functions of the three variables $t, \alpha$ and $\beta$.

Applying this observation to $\boldsymbol{e} = \boldsymbol{e}_1$, we can find a curve $\boldsymbol{\gamma}_1(s_1)$ in $\tilde{\boldsymbol{\sigma}}$ with $\boldsymbol{\gamma}_1(0) = \tilde{\boldsymbol{\sigma}}(\tilde{u}_0, \tilde{v}_0)$ and $d\boldsymbol{\gamma}_1/ds_1 = \boldsymbol{e}_1$. Now applying the same observation to $\boldsymbol{e} = \boldsymbol{e}_2$, we can find, for each value of $s_1$ close to 0, a curve $s_2 \mapsto \boldsymbol{\lambda}(s_1, s_2)$ in $\tilde{\boldsymbol{\sigma}}$ with $\partial \boldsymbol{\lambda}/\partial s_2 = \boldsymbol{e}_2$ and $\boldsymbol{\lambda}(s_1, 0) = \boldsymbol{\gamma}_1(s_1)$. Define $(\tilde{u}, \tilde{v})$ as functions of $(s_1, s_2)$ by

$$\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v}) = \boldsymbol{\lambda}(s_1, s_2). \tag{8.11}$$




Differentiating with respect to $s_1$ and $s_2$ gives

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \frac{\partial \tilde{u}}{\partial s_1} + \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \frac{\partial \tilde{v}}{\partial s_1} = \boldsymbol{\lambda}_{s_1}, \quad \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \frac{\partial \tilde{u}}{\partial s_2} + \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \frac{\partial \tilde{v}}{\partial s_2} = \boldsymbol{\lambda}_{s_2}.$$

We have

$$\boldsymbol{\lambda}_{s_1}|_{s_2=0} = \frac{d}{ds_1}\boldsymbol{\lambda}(s_1, 0) = \frac{d\boldsymbol{\gamma}_1}{ds_1} = \boldsymbol{e}_1, \quad \boldsymbol{\lambda}_{s_2} = \frac{\partial \lambda}{\partial s_2} = \boldsymbol{e}_2. \tag{8.12}$$

Equating coefficients of $\tilde{\boldsymbol{\sigma}}_{\tilde{u}}$ and $\tilde{\boldsymbol{\sigma}}_{\tilde{v}}$, we see from the last two sets of equations that, at the point $\tilde{\boldsymbol{\sigma}}(\tilde{u}_0, \tilde{v}_0)$, where $s_1 = s_2 = 0$, the Jacobian matrix

$$\begin{pmatrix} \frac{\partial \tilde{u}}{\partial s_1} & \frac{\partial \tilde{u}}{\partial s_2} \\ \frac{\partial \tilde{v}}{\partial s_1} & \frac{\partial \tilde{v}}{\partial s_2} \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}. \tag{8.13}$$

Since $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are linearly independent at $(\tilde{u}_0, \tilde{v}_0)$, this matrix is invertible. By the Inverse Function Theorem 5.6.1, Eq. 8.11 can be solved for $(s_1, s_2)$ as smooth functions of $(\tilde{u}, \tilde{v})$ when $(\tilde{u}, \tilde{v})$ is in some open set $\tilde{W}$ of $\tilde{U}$ containing $(\tilde{u}_0, \tilde{v}_0)$. Thus, $\boldsymbol{\lambda}$ is an allowable surface patch; by Eq. 8.12, it has the property that $\boldsymbol{\lambda}_{s_1} = \boldsymbol{e}_1$ when $s_2 = 0$, and $\boldsymbol{\lambda}_{s_2} = \boldsymbol{e}_2$ everywhere.

We now repeat the procedure, this time starting with a curve $\boldsymbol{\gamma}_2(t_2)$ with $d\boldsymbol{\gamma}_2/dt_2 = \boldsymbol{e}_2$ and $\boldsymbol{\gamma}_2(0) = \tilde{\boldsymbol{\sigma}}(\tilde{u}_0, \tilde{v}_0)$, and then taking a curve $t_1 \mapsto \boldsymbol{\mu}(t_1, t_2)$ with $\partial \boldsymbol{\mu}/\partial t_1 = \boldsymbol{e}_1$ and $\boldsymbol{\mu}(0, t_2) = \boldsymbol{\gamma}_2(t_2)$. This gives an allowable patch $\boldsymbol{\mu}(t_1, t_2)$ such that

$$\boldsymbol{\mu}(t_1, t_2) = \tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$$

for $(\tilde{u}, \tilde{v})$ in some open subset $\tilde{Z}$ of $\tilde{U}$ containing $(\tilde{u}_0, \tilde{v}_0)$. This patch has the property that $\boldsymbol{\mu}_{t_1} = \boldsymbol{e}_1$ everywhere and $\boldsymbol{\mu}_{t_2} = \boldsymbol{e}_2$ when $t_1 = 0$.

The parametrization we want is $\boldsymbol{\sigma}(u,v)$, where $\boldsymbol{\sigma}(u,v)$ is the intersection of the curve $s_2 \mapsto \boldsymbol{\lambda}(u,s_2)$ with the curve $t_1 \mapsto \boldsymbol{\mu}(t_1,v)$. Thus, we consider the equations

$$\tilde{\boldsymbol{\sigma}}(\tilde{u},\tilde{v}) = \boldsymbol{\lambda}(u,s_2) = \boldsymbol{\mu}(t_1,v).$$

From Eq. 8.13,

$$\frac{\partial \tilde{u}}{\partial u} = a, \quad \frac{\partial \tilde{v}}{\partial u} = b,$$

and similarly

$$\frac{\partial \tilde{u}}{\partial v} = c, \quad \frac{\partial \tilde{v}}{\partial v} = d.$$

Hence, the Jacobian matrix

$$\begin{pmatrix} \frac{\partial \tilde{u}}{\partial u} & \frac{\partial \tilde{u}}{\partial v} \\ \frac{\partial \tilde{v}}{\partial u} & \frac{\partial \tilde{v}}{\partial v} \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

As usual, the fact that this matrix is invertible means that $(u,v)$ can be expressed as smooth functions of $(\tilde{u},\tilde{v})$, for $(\tilde{u},\tilde{v})$ in some open subset $\tilde{V}$ of $\tilde{W} \cap \tilde{Z}$ containing $(\tilde{u}_0,\tilde{v}_0)$, and we get a reparametrization $\boldsymbol{\sigma}(u,v)$ of $\tilde{\boldsymbol{\sigma}}(\tilde{u},\tilde{v})$. Finally, the equation $\boldsymbol{\sigma}(u,v) = \boldsymbol{\mu}(t_1,v)$ implies that

$$\boldsymbol{\sigma}_u = \frac{\partial t_1}{\partial u}\boldsymbol{\mu}_{t_1} = \frac{\partial t_1}{\partial u}\boldsymbol{e}_1,$$

and similarly

$$\boldsymbol{\sigma}_v = \frac{\partial s_2}{\partial v}\boldsymbol{e}_2,$$

so $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are parallel to $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ everywhere. □

### *EXERCISES*

8.4.1 Let $\mathbf{p}$ be a hyperbolic point of a surface $\mathcal{S}$ (see Section 8.2). Show that there is a patch of $\mathcal{S}$ containing $\mathbf{p}$ whose parameter curves are asymptotic curves (see Exercise 7.3.6). Show that the second fundamental form of such a patch is of the form $2Mdudv$.

## 8.5 Surfaces of constant mean curvature

We now consider surfaces whose mean curvature $H$ is constant. Such surfaces have an interesting physical interpretation: we shall show in Section 12.1 that soap bubbles always adopt the form of a surface of constant mean curvature. In this section we give two simple constructions of surfaces of constant non-zero mean curvature; the case in which $H = 0$ is treated in much more detail in Chapter 12.

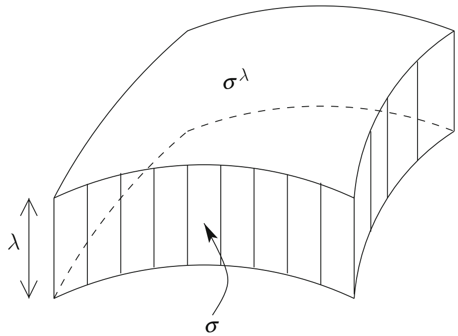
The first of these gives a correspondence between surfaces of constant non-zero mean curvature and surfaces of constant positive Gaussian curvature.

### Definition 8.5.1

Let $\mathcal{S}$ be an oriented surface and let $\lambda \in \mathbb{R}$. The *parallel surface* $\mathcal{S}^\lambda$ of $\mathcal{S}$ is

$$\mathcal{S}^\lambda = \{\mathbf{p} + \lambda \mathbf{N_p} \,|\, \mathbf{p} \in \mathcal{S}\},$$

where $\mathbf{N_p}$ is the unit normal of $\mathcal{S}$ at the point $\mathbf{p}$.




Roughly speaking, $\mathcal{S}^\lambda$ is obtained by translating the surface $\mathcal{S}$ at a distance $\lambda$ perpendicular to itself (but this will not be a genuine translation since $\mathbf{N_p}$ will in general depend on $\mathbf{p}$).

### Proposition 8.5.2

Let $\kappa_1$ and $\kappa_2$ be the principal curvatures of an oriented surface $\mathcal{S}$, let $\lambda \in \mathbb{R}$ and let $\mathcal{S}^\lambda$ be the corresponding parallel surface of $\mathcal{S}$. Assume that neither $\kappa_1$ nor $\kappa_2$ is equal to $1/\lambda$ at any point of $\mathcal{S}$. Then,

(i) $\mathcal{S}^\lambda$ is a (smooth) oriented surface, the unit normal of $\mathcal{S}^\lambda$ at $\mathbf{p} + \lambda\mathbf{N_p}$ being equal to $\epsilon\mathbf{N_p}$, where $\epsilon$ is the sign of $(1 - \lambda\kappa_1)(1 - \lambda\kappa_2)$.

(ii) The principal curvatures of $\mathcal{S}^\lambda$ are $\epsilon\kappa_1/(1 - \lambda\kappa_1)$ and $\epsilon\kappa_2/(1 - \lambda\kappa_2)$, and the corresponding principal vectors are the same as those of $\mathcal{S}$ for the principal curvatures $\kappa_1$ and $\kappa_2$, respectively.

(iii) The Gaussian and mean curvatures of $\mathcal{S}^\lambda$ are

$$\frac{K}{1 - 2\lambda H + \lambda^2 K} \quad \text{and} \quad \frac{\epsilon(H - \lambda K)}{1 - 2\lambda H + \lambda^2 K},$$

respectively, where $K$ and $H$ are the Gaussian and mean curvatures of $\mathcal{S}$.

## Proof

Let $\boldsymbol{\sigma}(u,v)$ be a surface patch of $\mathcal{S}$ with standard unit normal $\mathbf{N}(u,v)$. Define

$$\boldsymbol{\sigma}^\lambda(u,v) = \boldsymbol{\sigma}(u,v) + \lambda\mathbf{N}(u,v).$$

By Proposition 8.1.2,

$$\begin{aligned}
\boldsymbol{\sigma}^\lambda_u &= \boldsymbol{\sigma}_u + \lambda\mathbf{N}_u = (1-\lambda a)\,\boldsymbol{\sigma}_u - \lambda b\,\boldsymbol{\sigma}_v,\\
\boldsymbol{\sigma}^\lambda_v &= \boldsymbol{\sigma}_v + \lambda\mathbf{N}_v = -\lambda c\,\boldsymbol{\sigma}_u + (1-\lambda d)\,\boldsymbol{\sigma}_v,
\end{aligned} \tag{8.14}$$

where

$$\mathcal{W}_{\boldsymbol{\sigma}} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

is the matrix of the Weingarten map of $\mathcal{S}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of the tangent plane. Hence,

$$\boldsymbol{\sigma}^\lambda_u \times \boldsymbol{\sigma}^\lambda_v = (1 - \lambda(a+d) + \lambda^2(ad-bc))\,\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v.$$

Since $\kappa_1$ and $\kappa_2$ are the eigenvalues of $\mathcal{W}_{\boldsymbol{\sigma}}$ (see Section 8.2), and since the sum and product of the eigenvalues of a matrix are equal to the trace and the determinant of the matrix, respectively,

$$\kappa_1 + \kappa_2 = a + d, \quad \kappa_1\kappa_2 = ad - bc.$$

Hence,

$$\boldsymbol{\sigma}^\lambda_u \times \boldsymbol{\sigma}^\lambda_v = (1-\lambda\kappa_1)(1-\lambda\kappa_2)\,\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v. \tag{8.15}$$

The assertions in part (i) follow from this equation.

The principal curvatures of $\mathcal{S}^\lambda$ are the eigenvalues of the matrix $\mathcal{W}_{\boldsymbol{\sigma}^\lambda}$ of the Weingarten map of $\mathcal{S}^\lambda$ with respect to the basis $\{\boldsymbol{\sigma}^\lambda_u, \boldsymbol{\sigma}^\lambda_v\}$. By the proof of Proposition 8.1.2, this is the negative of the matrix expressing $\mathbf{N}^\lambda_u$ and $\mathbf{N}^\lambda_v$ in terms of $\boldsymbol{\sigma}^\lambda_u$ and $\boldsymbol{\sigma}^\lambda_v$, where $\mathbf{N}^\lambda$ is the standard unit normal of $\boldsymbol{\sigma}^\lambda$. Equation 8.14 says that the matrix expressing $\boldsymbol{\sigma}^\lambda_u$ and $\boldsymbol{\sigma}^\lambda_v$ in terms of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ is $I - \lambda\mathcal{W}_{\boldsymbol{\sigma}}$, and the fact that $\mathbf{N}^\lambda = \epsilon\mathbf{N}$ implies that $-\epsilon\mathcal{W}_{\boldsymbol{\sigma}}$ is the matrix expressing $\mathbf{N}^\lambda_u$ and $\mathbf{N}^\lambda_v$ in terms of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Combining these two observations we get

$$\mathcal{W}_{\boldsymbol{\sigma}^\lambda} = \epsilon(I - \lambda\mathcal{W}_{\boldsymbol{\sigma}})^{-1}\mathcal{W}_{\boldsymbol{\sigma}}.$$

If $T$ is an eigenvector of $\mathcal{W}_{\boldsymbol{\sigma}}$ with eigenvalue $\kappa$, then $T$ is also an eigenvector of $\mathcal{W}_{\boldsymbol{\sigma}^\lambda}$ with eigenvalue $\epsilon\kappa/(1-\lambda\kappa)$. The assertions in part (ii) follows from this.

Part (iii) follows from part (ii) by straightforward algebra. $\square$

## Corollary 8.5.3

If $\mathcal{S}$ has constant Gaussian curvature $1/R^2$, the parallel surfaces $\mathcal{S}^{\pm R}$ have constant mean curvature $1/2R$. Conversely, if $\mathcal{S}$ has constant mean curvature $1/2R$, the parallel surface $\mathcal{S}^R$ has constant Gaussian curvature $1/R^2$.

## Proof

This follows from part (iii) of the proposition by straightforward algebra. For example, if $H = 1/2R$ the Gaussian curvature of $\mathcal{S}^R$ is

$$\frac{K}{1 - 2RH + R^2K} = \frac{K}{R^2K} = \frac{1}{R^2}. \qquad \square$$

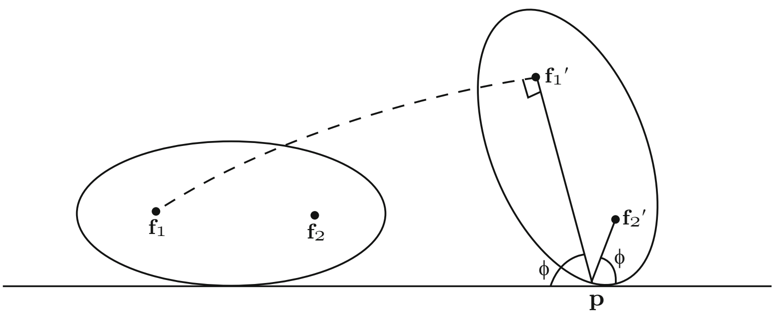
The next construction gives a beautiful geometric description of the surfaces of revolution which have constant non-zero mean curvature in terms of the curve traced out by the focus of an ellipse that rolls without slipping along a straight line (cf. Exercise 2.2.10). Take the ellipse to be

$$\frac{x^2}{p^2} + \frac{(y-q)^2}{q^2} = 1,$$

where $p > q > 0$ are constants. Thus, the ellipse is tangent to the $x$-axis at the origin. The foci of the ellipse are the points $\mathbf{f}_1 = (-\epsilon p, q)$ and $\mathbf{f}_2 = (\epsilon p, q)$, where the eccentricity $\epsilon = \sqrt{1 - \frac{q^2}{p^2}}$.

## Proposition 8.5.4

With the above notation, let $\mathcal{C}$ be the curve traced out by one of the foci of the ellipse as it rolls without slipping along the $x$-axis. Let $\mathcal{S}$ be the surface obtained by rotating $\mathcal{C}$ around the $x$-axis. Then, $\mathcal{S}$ has constant non-zero mean curvature.



## Proof

We consider a situation in which the ellipse has rolled along the $x$-axis so that its point of contact with the $x$-axis is at a point $\mathbf{p}$, the focus $\mathbf{f}_1$ has moved to a point $\mathbf{f}_1{}' = (x, y)$ on $\mathcal{C}$, and the focus $\mathbf{f}_2$ has moved to $\mathbf{f}_2{}' = (X, Y)$, say. Let $\varphi$ be the angle between $\mathbf{p} - \mathbf{f}_1{}'$ and the $x$-axis; then $\varphi$ is also the angle between $\mathbf{p} - \mathbf{f}_2{}'$ and the $x$-axis by Exercise 1.1.6(iii). Hence,

$$y = \| \mathbf{p} - \mathbf{f}_1{}' \| \sin\varphi, \quad Y = \| \mathbf{p} - \mathbf{f}_2{}' \| \sin\varphi$$

and so

$$y + Y = 2p\sin\varphi$$

by Exercise 1.1.6(i). But Exercise 1.1.6(ii) gives $yY = q^2$ so

$$y + \frac{q^2}{y} = 2p\sin\varphi. \tag{8.16}$$

Now, since the ellipse rolls without slipping, the point of contact of the ellipse with the $x$-axis is stationary. This implies that the point $\mathbf{f}_1{}'$ moves as if rotating instantaneously about $\mathbf{p}$, so that the tangent vector to $\mathcal{C}$ at $\mathbf{f}_1{}'$ is perpendicular to $\mathbf{p} - \mathbf{f}_1{}'$. (If this heuristic argument is unconvincing, an analytical proof can be found in Exercise 2.2.10.) It follows that

$$\frac{dy}{dx} = \cot\varphi. \tag{8.17}$$

Eliminating $\varphi$ between Eqs. 8.16 and 8.17 gives

$$y^2 + q^2 = \frac{2py}{\sqrt{1 + \left(\frac{dy}{dx}\right)^2}}. \tag{8.18}$$

The surface $\mathcal{S}$ obtained by rotating $\mathcal{C}$ around the $x$-axis can be parametrized by

$$\boldsymbol{\sigma}(x, \theta) = (x, y\cos\theta, y\sin\theta)$$

where $\theta$ is the angle of rotation. The first and second fundamental forms of $\boldsymbol{\sigma}$ are

$$\left(1 + \left(\frac{dy}{dx}\right)^2\right) dx^2 + y^2 d\theta^2 \quad \text{and} \quad -\frac{d^2y}{dx^2}dx^2 + \frac{y}{\sqrt{1 + \left(\frac{dy}{dx}\right)^2}}d\theta^2,$$

respectively. Using the formula in Corollary 8.1.3, the mean curvature is found to be

$$H = \frac{1}{2y\sqrt{1 + \left(\frac{dy}{dx}\right)^2}} - \frac{\frac{d^2y}{dx^2}}{2\left(1 + \left(\frac{dy}{dx}\right)^2\right)^{3/2}}. \tag{8.19}$$

Differentiating both sides of Eq. 8.18 we get

$$2y\frac{dy}{dx} = \frac{2p\frac{dy}{dx}}{\sqrt{1+\left(\frac{dy}{dx}\right)^2}} - \frac{2py\frac{dy}{dx}\frac{d^2y}{dx^2}}{\left(1+\left(\frac{dy}{dx}\right)^2\right)^{3/2}}.$$

Dividing both sides by $4py\frac{dy}{dx}$ and comparing with Eq. 8.19 shows that the surface $\mathcal{S}$ has mean curvature $1/2p$. □

## *EXERCISES*

8.5.1 Suppose that the first fundamental form of a surface patch $\boldsymbol{\sigma}(u, v)$ is of the form $E(du^2 + dv^2)$. Prove that $\boldsymbol{\sigma}_{uu} + \boldsymbol{\sigma}_{vv}$ is perpendicular to $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Deduce that the mean curvature $H = 0$ everywhere if and only if the Laplacian

$$\boldsymbol{\sigma}_{uu} + \boldsymbol{\sigma}_{vv} = \mathbf{0}.$$

Show that the surface patch

$$\boldsymbol{\sigma}(u, v) = \left(u - \frac{u^3}{3} + uv^2, v - \frac{v^3}{3} + u^2v, u^2 - v^2\right)$$

has $H = 0$ everywhere. (A picture of this surface can be found in Section 12.2.)

8.5.2 Prove that $H = 0$ for the surface

$$z = \ln\left(\frac{\cos y}{\cos x}\right).$$

(A picture of this surface can also be found in Section 12.2.)

8.5.3 Let $\boldsymbol{\sigma}(u, v)$ be a surface with first and second fundamental forms $Edu^2 + Gdv^2$ and $Ldu^2 + Ndv^2$, respectively (cf. Proposition 8.4.1). Define

$$\boldsymbol{\Sigma}(u, v, w) = \boldsymbol{\sigma}(u, v) + w\mathbf{N}(u, v),$$

where $\mathbf{N}$ is the standard unit normal of $\boldsymbol{\sigma}$. Show that the three families of surfaces obtained by fixing the values of $u$, $v$ or $w$ in $\boldsymbol{\Sigma}$ form a triply orthogonal system (see Section 5.5). The surfaces $w =$ constant are parallel surfaces of $\boldsymbol{\sigma}$. Show that the surfaces $u =$ constant and $v =$ constant are flat ruled surfaces.

## 8.6 Gaussian curvature of compact surfaces

We have seen in Section 8.2 how the relative signs of the principal curvatures at a point $\mathbf{p}$ of a surface $\mathcal{S}$ determine the shape of $\mathcal{S}$ near $\mathbf{p}$. In fact, since the Gaussian curvature $K$ of $\mathcal{S}$ is the product of its principal curvatures, the discussion there shows that

(i) If $K > 0$ at $\mathbf{p}$, then $\mathbf{p}$ is an elliptic point.

(ii) If $K < 0$ at $\mathbf{p}$, then $\mathbf{p}$ is a hyperbolic point.

(iii) If $K = 0$ at $\mathbf{p}$, then $\mathbf{p}$ is either a parabolic point or a planar point.

In this section, we give a result which shows how the Gaussian curvature influences the *global* shape of a surface. We shall give another result of a similar nature in Section 13.4.

### Proposition 8.6.1

If $\mathcal{S}$ is a compact surface, there is a point of $\mathcal{S}$ at which its Gaussian curvature $K$ is $> 0$.

In the proof, we shall make use of the following fact about compact sets: if $X$ is a compact subset of $\mathbb{R}^3$ and $f : \mathbb{R}^3 \to \mathbb{R}$ is a continuous function, then there are points $\mathbf{p}, \mathbf{q} \in X$ such that $f(\mathbf{q}) \leq f(\mathbf{r}) \leq f(\mathbf{p})$ for all points $\mathbf{r} \in X$, so that $f$ attains its maximum value on $X$ at $\mathbf{p}$ and its minimum at $\mathbf{q}$.

### Proof

Define $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(\mathbf{v}) = \parallel \mathbf{v} \parallel^2$. Then, $f$ is continuous, so the fact that $\mathcal{S}$ is compact implies that there is a point $\mathbf{p} \in \mathcal{S}$ where $f$ attains its maximum value. Then $\mathcal{S}$ is contained inside the closed ball of radius $\parallel \mathbf{p} \parallel$ and centre the origin, and $\mathcal{S}$ intersects its boundary sphere at $\mathbf{p}$. The idea is that $\mathcal{S}$ is at least as curved as the sphere at $\mathbf{p}$, so its Gaussian curvature should be at least that of the sphere at $\mathbf{p}$, i.e., at least $1/ \parallel \mathbf{p} \parallel^2$.

To make this argument precise, let $\boldsymbol{\gamma}(t)$ be any unit-speed curve in $\mathcal{S}$ passing through $\mathbf{p}$ when $t = 0$. Then, $f(\boldsymbol{\gamma}(t))$ has a local maximum at $t = 0$, so

$$\frac{d}{dt} f(\boldsymbol{\gamma}(t)) = 0, \quad \frac{d^2}{dt^2} f(\boldsymbol{\gamma}(t)) \leq 0$$

at $t = 0$, i.e.,

$$\boldsymbol{\gamma}(0) \cdot \dot{\boldsymbol{\gamma}}(0) = 0, \quad \boldsymbol{\gamma}(0) \cdot \ddot{\boldsymbol{\gamma}}(0) + 1 \leq 0. \tag{8.20}$$

The equation in (8.20) shows that $\mathbf{p} = \boldsymbol{\gamma}(0)$ is perpendicular to every unit tangent vector to $\mathcal{S}$ at $\mathbf{p}$, and hence is perpendicular to the tangent plane $T_{\mathbf{p}}\mathcal{S}$.

Choose a surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$ containing $\mathbf{p}$, and let $\mathbf{N}$ be its standard unit normal. By the preceding remark,

$$\mathbf{N} = \pm \frac{\mathbf{p}}{\| \mathbf{p} \|}. \tag{8.21}$$

The inequality in (8.20) implies that the normal curvature $\kappa_n = \ddot{\boldsymbol{\gamma}}(0) \cdot \mathbf{N}$ of $\boldsymbol{\gamma}$ at $\mathbf{p}$ (computed in the patch $\boldsymbol{\sigma}$) is $\leq -1/ \| \mathbf{p} \|$ or $\geq 1/ \| \mathbf{p} \|$, according to whether the sign in Eq. 8.21 is $+$ or $-$, respectively. By Corollary 8.2.5, the principal curvatures of $\boldsymbol{\sigma}$ at $\mathbf{p}$ are either both $\leq -1/ \| \mathbf{p} \|$ or both $\geq 1/ \| \mathbf{p} \|$. In each case, $K \geq 1/ \| \mathbf{p} \|^2 > 0$ at $\mathbf{p}$. $\square$

# 9 Geodesics

Geodesics are the curves in a surface that a bug living in the surface would perceive to be straight. For example, the shortest path between two points in a surface is always a geodesic. We shall actually begin by giving a quite different definition of geodesics, since this definition is easier to work with. We give various methods of finding geodesics on surfaces, before finally making contact with the idea of shortest paths towards the end of the chapter.

## 9.1 Definition and basic properties

If we drive along a 'straight' road, we do not have to turn the wheel of our car to the right or left (this is what we mean by 'straight'!). However, the road is not, in fact, a straight line as the surface of the earth is, to a good approximation, a sphere and there can be no straight line on the surface of a sphere. If the road is represented by a curve $\boldsymbol{\gamma}$, its acceleration $\ddot{\boldsymbol{\gamma}}$ will be non-zero, but we perceive the curve as being straight because the *tangential component* of $\ddot{\boldsymbol{\gamma}}$ is zero, in other words because $\ddot{\boldsymbol{\gamma}}$ is perpendicular to the surface. This suggests

### Definition 9.1.1

A curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$ is called a *geodesic* if $\ddot{\boldsymbol{\gamma}}(t)$ is zero or perpendicular to the tangent plane of the surface at the point $\boldsymbol{\gamma}(t)$, i.e., parallel to its unit normal, for all values of the parameter $t$.

Andrew Pressley, *Elementary Differential Geometry: Second Edition*, Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_9,

Equivalently, $\gamma$ is a geodesic if and only if its tangent vector $\dot{\gamma}$ is *parallel* along $\gamma$ (see Section 7.4).

Note that this definition makes sense for any surface, orientable or not.

There is an interesting mechanical interpretation of geodesics: a particle moving on the surface, and subject to no forces except a force acting perpendicular to the surface that keeps the particle on the surface, would move along a geodesic. This is because Newton's second law of motion states that the force on the particle is parallel to its acceleration $\ddot{\gamma}$, which would therefore be perpendicular to the surface.

We begin our study of geodesics by noting that there is essentially no choice in their parametrization.

## Proposition 9.1.2

Any geodesic has constant speed.

## Proof

Let $\gamma(t)$ be a geodesic on a surface $\mathcal{S}$. Then, denoting $d/dt$ by a dot,

$$\frac{d}{dt} \parallel \dot{\gamma} \parallel^2 = \frac{d}{dt}(\dot{\gamma} \cdot \dot{\gamma}) = 2\ddot{\gamma} \cdot \dot{\gamma}.$$

Since $\gamma$ is a geodesic, $\ddot{\gamma}$ is perpendicular to the tangent plane and is therefore perpendicular to the tangent vector $\dot{\gamma}$. So $\ddot{\gamma} \cdot \dot{\gamma} = 0$ and the last equation shows that $\parallel \dot{\gamma} \parallel$ is constant. $\square$

It follows from this proposition that a unit-speed reparametrization of a geodesic $\gamma$ is still a geodesic. For, if $\parallel \dot{\gamma} \parallel = \lambda$, then $\tilde{\gamma}(t) = \gamma(t/\lambda)$ is a unit-speed reparametrization of $\gamma$ and $\frac{d^2\tilde{\gamma}}{dt^2} = \frac{1}{\lambda^2}\frac{d^2\gamma}{dt^2}$ is parallel to $\ddot{\gamma}$, and hence is also perpendicular to the surface. Thus, we can always restrict to unit-speed geodesics if we wish. In general, however, a reparametrization of a geodesic will not be a geodesic (see Exercise 9.1.2).

We observe next that there is an equivalent definition of a geodesic expressed in terms of the geodesic curvature $\kappa_g$ (see Section 7.3). Of course, this is why $\kappa_g$ is called the geodesic curvature!

## Proposition 9.1.3

A unit-speed curve on a surface is a geodesic if and only if its geodesic curvature is zero everywhere.

## Proof

Let $\boldsymbol{\gamma}$ be a unit-speed curve on the surface $\mathcal{S}$, and let $\mathbf{p} \in \mathcal{S}$. Let $\boldsymbol{\sigma}$ be a surface patch of $\mathcal{S}$ with $\mathbf{p}$ in its image, and let $\mathbf{N}$ be the standard unit normal of $\boldsymbol{\sigma}$, so that

$$\kappa_g = \ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \tag{9.1}$$

(changing $\boldsymbol{\sigma}$ may change the sign of $\mathbf{N}$, and hence that of $\kappa_g$, but that is not relevant to the present discussion). If $\ddot{\boldsymbol{\gamma}}$ is parallel to $\mathbf{N}$, it is obviously perpendicular to $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$, so by Eq. 9.1, $\kappa_g = 0$.

Conversely, suppose that $\kappa_g = 0$. Then, $\ddot{\boldsymbol{\gamma}}$ is perpendicular to $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$. But then, since $\dot{\boldsymbol{\gamma}}$, $\mathbf{N}$ and $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$ are perpendicular unit vectors in $\mathbb{R}^3$ (see the discussion in Section 7.3), and since $\ddot{\boldsymbol{\gamma}}$ is perpendicular to $\dot{\boldsymbol{\gamma}}$, it follows that $\ddot{\boldsymbol{\gamma}}$ is parallel to $\mathbf{N}$. □

The following result gives the simplest examples of geodesics.

## Proposition 9.1.4

Any (part of a) straight line on a surface is a geodesic.

By this, we mean that every straight line can be parametrized so that it is a geodesic. A similar remark applies to other geodesics we consider and whose parametrization is not specified (see Exercise 9.1.2).

## Proof

This is obvious, for any straight line has a (constant speed) parametrization of the form

$$\boldsymbol{\gamma}(t) = \mathbf{a} + \mathbf{b}t,$$

where $\mathbf{a}$ and $\mathbf{b}$ are constant vectors, and clearly $\ddot{\boldsymbol{\gamma}} = \mathbf{0}$. □

## Example 9.1.5

All straight lines in the plane are geodesics, as are the rulings of any ruled surface, such as those of a (generalized) cylinder or a (generalized) cone, or the straight lines on a hyperboloid of one sheet.

The next result is almost as simple:

## Proposition 9.1.6

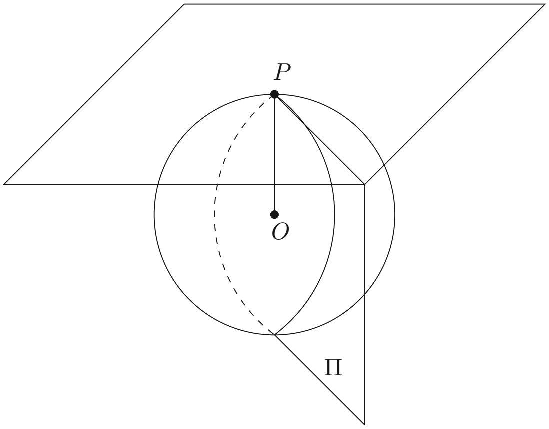
Any normal section of a surface is a geodesic.

## Proof

Recall from Section 7.3 that a normal section of a surface $\mathcal{S}$ is the intersection $\mathcal{C}$ of $\mathcal{S}$ with a plane $\Pi$, such that $\Pi$ is perpendicular to the surface at each point of $\mathcal{C}$. We showed in Corollary 7.3.4 that $\kappa_g = 0$ for a normal section, and so the result follows from Proposition 9.1.3. □

## Example 9.1.7

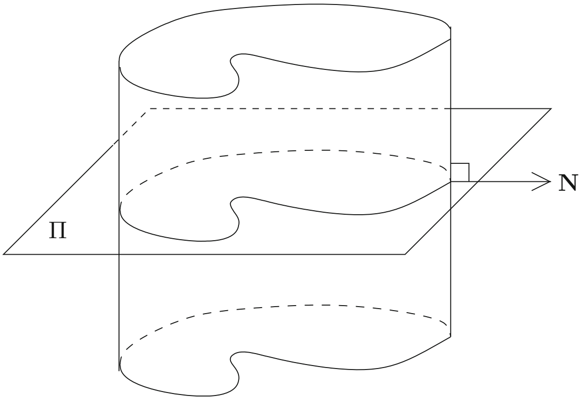
All great circles on a sphere are geodesics. For a great circle is the intersection of




the sphere with a plane $\Pi$ passing through the centre $O$ of the sphere, and so if $P$ is a point of the great circle, the straight line through $O$ and $P$ lies in $\Pi$ and is perpendicular to the tangent plane of the sphere at $P$. Hence, $\Pi$ is perpendicular to the tangent plane at $P$.

## Example 9.1.8

The intersection of a generalized cylinder with a plane $\Pi$ perpendicular to the rulings of the cylinder is a geodesic. For it is clear that the unit normal $\mathbf{N}$ is perpendicular to the rulings. It follows that $\mathbf{N}$ is parallel to $\Pi$, and hence that $\Pi$ is perpendicular to the tangent plane.




## EXERCISES

9.1.1 Describe four different geodesics on the hyperboloid of one sheet

$$x^2 + y^2 - z^2 = 1$$

passing through the point $(1, 0, 0)$.

9.1.2 A (regular) curve $\boldsymbol{\gamma}$ with nowhere vanishing curvature on a surface $\mathcal{S}$ is called a *pre-geodesic* on $\mathcal{S}$ if some reparametrization of $\boldsymbol{\gamma}$ is a geodesic on $\mathcal{S}$ (recall that a reparametrization of a geodesic is not usually a geodesic). Show that:

(i) A curve $\boldsymbol{\gamma}$ is a pre-geodesic if and only if $\ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) = 0$ everywhere on $\boldsymbol{\gamma}$ (in the notation of the proof of Proposition 9.1.3).

(ii) Any reparametrization of a pre-geodesic is a pre-geodesic.

(iii) Any constant speed reparametrization of a pre-geodesic is a geodesic.

(iv) A pre-geodesic is a geodesic if and only if it has constant speed.

9.1.3 Consider the tube of radius $a > 0$ around a unit-speed curve $\boldsymbol{\gamma}$ in $\mathbb{R}^3$ defined in Exercise 4.2.7:

$$\boldsymbol{\sigma}(s, \theta) = \boldsymbol{\gamma}(s) + a(\cos\theta\, \mathbf{n}(s) + \sin\theta\, \mathbf{b}(s)).$$

Show that the parameter curves on the tube obtained by fixing the value of $s$ are circular geodesics on $\boldsymbol{\sigma}$.

9.1.4 Let $\boldsymbol{\gamma}(t)$ be a geodesic on an ellipsoid $\mathcal{S}$ (see Theorem 5.2.2(i)). Let $2R(t)$ be the length of the diameter of $\mathcal{S}$ parallel to $\dot{\boldsymbol{\gamma}}(t)$, and let $S(t)$ be the distance from the centre of $\mathcal{S}$ to the tangent plane $T_{\boldsymbol{\gamma}(t)}\mathcal{S}$. Show that the curvature of $\boldsymbol{\gamma}$ is $S(t)/R(t)^2$, and that the product $R(t)S(t)$ is independent of $t$.

9.1.5 Show that a geodesic with nowhere vanishing curvature is a plane curve if and only if it is a line of curvature.

9.1.6 Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two surfaces that intersect in a curve $\mathcal{C}$, and let $\boldsymbol{\gamma}$ be a unit-speed parametrization of $\mathcal{C}$.

(i) Show that if $\boldsymbol{\gamma}$ is a geodesic on *both* $\mathcal{S}_1$ and $\mathcal{S}_2$ and if the curvature of $\boldsymbol{\gamma}$ is nowhere zero, then $\mathcal{S}_1$ ad $\mathcal{S}_2$ touch along $\boldsymbol{\gamma}$ (i.e., they have the same tangent plane at each point of $\mathcal{C}$). Give an example of this situation.

(ii) Show that if $\mathcal{S}_1$ and $\mathcal{S}_2$ intersect orthogonally at each point of $\mathcal{C}$, then $\boldsymbol{\gamma}$ is a geodesic on $\mathcal{S}_1$ if and only if $\dot{\mathbf{N}}_2$ is parallel to $\mathbf{N}_1$ at each point of $\mathcal{C}$ (where $\mathbf{N}_1$ and $\mathbf{N}_2$ are unit normals of $\mathcal{S}_1$ and $\mathcal{S}_2$). Show also that, in this case, $\boldsymbol{\gamma}$ is a geodesic on *both* $\mathcal{S}_1$ and $\mathcal{S}_2$ if and only if $\mathcal{C}$ is part of a straight line.

## 9.2 Geodesic equations

Unfortunately, Propositions 9.1.4 and 9.1.6 are not usually sufficient to determine all the geodesics on a given surface. For that, we need the following result:

### Theorem 9.2.1

A curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$ is a geodesic if and only if, for any part $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ of $\boldsymbol{\gamma}$ contained in a surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$, the following two equations are satisfied:

$$\begin{aligned} \frac{d}{dt}(E\dot{u} + F\dot{v}) &= \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2), \\ \frac{d}{dt}(F\dot{u} + G\dot{v}) &= \frac{1}{2}(E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2), \end{aligned} \tag{9.2}$$

where $Edu^2 + 2Fdudv + Gdv^2$ is the first fundamental form of $\boldsymbol{\sigma}$.

The differential equations (9.2) are called the *geodesic equations*.

## Proof

Since $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis of the tangent plane of $\boldsymbol{\sigma}$, $\boldsymbol{\gamma}$ is a geodesic if and only if $\ddot{\boldsymbol{\gamma}}$ is perpendicular to $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Since $\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$, this is equivalent to

$$\left(\frac{d}{dt}(\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v)\right) \cdot \boldsymbol{\sigma}_u = 0 \quad \text{and} \quad \left(\frac{d}{dt}(\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v)\right) \cdot \boldsymbol{\sigma}_v = 0. \tag{9.3}$$

We show that these two equations are equivalent to the two geodesic equations.

The left-hand side of the first equation in (9.3) is equal to

$$\begin{aligned} &\frac{d}{dt}\left((\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \cdot \boldsymbol{\sigma}_u\right) - (\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \cdot \frac{d\boldsymbol{\sigma}_u}{dt} \\ &= \frac{d}{dt}(E\dot{u} + F\dot{v}) - (\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \cdot (\dot{u}\boldsymbol{\sigma}_{uu} + \dot{v}\boldsymbol{\sigma}_{uv}) \qquad (9.4) \\ &= \frac{d}{dt}(E\dot{u} + F\dot{v}) - (\dot{u}^2(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu}) + \dot{u}\dot{v}(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uv} + \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uu}) + \dot{v}^2(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uv})). \end{aligned}$$

Now,

$$E_u = (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u)_u = \boldsymbol{\sigma}_{uu} \cdot \boldsymbol{\sigma}_u + \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu} = 2\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu},$$

so $\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu} = \frac{1}{2}E_u$. Similarly, $\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uv} = \frac{1}{2}G_u$. Finally,

$$\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uv} + \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uu} = (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)_u = F_u.$$

Substituting these values into (9.4) gives

$$\left(\frac{d}{dt}(\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v)\right) \cdot \boldsymbol{\sigma}_u = \frac{d}{dt}(E\dot{u} + F\dot{v}) - \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2).$$

This shows that the first equation in (9.3) is equivalent to the first geodesic equation in (9.2). Similarly for the other equations. □

The geodesic equations are non-linear differential equations, and are usually difficult or impossible to solve explicitly. The following example is one case in which this can be done. Another is given in Exercise 9.2.3.

## Example 9.2.2

We determine the geodesics on the unit sphere $S^2$ by solving the geodesic equations. For the usual parametrization by latitude $\theta$ and longitude $\varphi$,

$$\boldsymbol{\sigma}(\theta, \varphi) = (\cos\theta\cos\varphi, \cos\theta\sin\varphi, \sin\theta),$$

we found in Example 6.1.3 that the first fundamental form is

$$d\theta^2 + \cos^2\theta\, d\varphi^2.$$

We might as well restrict ourselves to unit-speed curves $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(\theta(t), \varphi(t))$, so that

$$\dot{\theta}^2 + \dot{\varphi}^2 \cos^2\theta = 1,$$

and if $\boldsymbol{\gamma}$ is a geodesic the second equation in (9.2) gives

$$\frac{d}{dt}(\dot{\varphi}\cos^2\theta) = 0,$$

so that

$$\dot{\varphi}\cos^2\theta = \Omega,$$

where $\Omega$ is a constant. If $\Omega = 0$, then $\dot{\varphi} = 0$ and so $\varphi$ is constant and $\boldsymbol{\gamma}$ is part of a meridian. We assume that $\dot{\varphi} \neq 0$ from now on.

The unit-speed condition gives

$$\dot{\theta}^2 = 1 - \frac{\Omega^2}{\cos^2\theta},$$

so along the geodesic we have

$$\left(\frac{d\theta}{d\varphi}\right)^2 = \frac{\dot{\theta}^2}{\dot{\varphi}^2} = \cos^2\theta(\Omega^{-2}\cos^2\theta - 1),$$

and hence

$$\pm(\varphi - \varphi_0) = \int \frac{d\theta}{\cos\theta\sqrt{\Omega^{-2}\cos^2\theta - 1}},$$

where $\varphi_0$ is a constant. The integral can be evaluated by making the substitution $u = \tan\theta$. This gives

$$\pm(\varphi - \varphi_0) = \int \frac{du}{\sqrt{\Omega^{-2} - 1 - u^2}} = \sin^{-1}\left(\frac{u}{\sqrt{\Omega^{-2} - 1}}\right),$$

and hence

$$\tan\theta = \pm\sqrt{\Omega^{-2} - 1}\,\sin(\varphi - \varphi_0).$$

This implies that the coordinates $x = \cos\theta\cos\varphi$, $y = \cos\theta\sin\varphi$ and $z = \sin\theta$ of $\boldsymbol{\gamma}(t)$ satisfy the equation

$$z = ax + by,$$

where $a = \mp\sqrt{\Omega^{-2} - 1}\,\sin\varphi_0$, and $b = \pm\sqrt{\Omega^{-2} - 1}\,\cos\varphi_0$. This shows that $\boldsymbol{\gamma}$ is contained in the intersection of $S^2$ with a plane passing through the origin.

Hence, in all cases, $\boldsymbol{\gamma}$ is part of a great circle.

The geodesic equations can be expressed in a different, but equivalent, form which is sometimes more useful than that in Theorem 9.2.1.

## Proposition 9.2.3

A curve $\boldsymbol{\gamma}$ on a surface $\mathcal{S}$ is a geodesic if and only if, for any part $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ of $\boldsymbol{\gamma}$ contained in a surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$, the following two equations are satisfied:

$$\begin{aligned}
\ddot{u} + \Gamma_{11}^{1}\dot{u}^2 + 2\Gamma_{12}^{1}\dot{u}\dot{v} + \Gamma_{22}^{1}\dot{v}^2 &= 0\\
\ddot{v} + \Gamma_{11}^{2}\dot{u}^2 + 2\Gamma_{12}^{2}\dot{u}\dot{v} + \Gamma_{22}^{2}\dot{v}^2 &= 0.
\end{aligned}$$

## Proof

As we noted after Definition 9.1.1, $\boldsymbol{\gamma}$ is a geodesic if and only if $\dot{\boldsymbol{\gamma}}$ is parallel along $\boldsymbol{\gamma}$. Since $\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$, the equations in the statement of the proposition follow from Proposition 7.4.5. □

It can of course be verified directly that the differential equations in Proposition 9.2.3 are equivalent to those in Theorem 9.2.1 (see Exercise 9.2.6).

Proposition 9.2.3 makes it obvious that the geodesic equations are second-order ordinary differential equations for the functions $u(t)$ and $v(t)$. Even though we may be unable in many situations to solve these equations explicitly, the general theory of ordinary differential equations provides valuable information about their solutions. This leads to the following result, which tells us exactly 'how many' geodesics there are.

## Proposition 9.2.4

Let $\mathbf{p}$ be a point of a surface $\mathcal{S}$, and let $\mathbf{t}$ be a unit tangent vector to $\mathcal{S}$ at $\mathbf{p}$. Then, there exists a unique unit-speed geodesic $\boldsymbol{\gamma}$ on $\mathcal{S}$ which passes through $\mathbf{p}$ and has tangent vector $\mathbf{t}$ there.

In short, *there is a unique geodesic through any given point of a surface in any given tangent direction.*

## Proof

The geodesic equations in Proposition 9.2.3 are of the form

$$\ddot{u} = f(u, v, \dot{u}, \dot{v}), \quad \ddot{v} = g(u, v, \dot{u}, \dot{v}), \tag{9.5}$$

where $f$ and $g$ are smooth functions of the four variables $u, v, \dot{u}$ and $\dot{v}$. It is proved in the theory of ordinary differential equations that, for any given constants $a, b, c$, and $d$, and any value $t_0$ of $t$, there is a solution of Eqs. 9.5 such that

$$u(t_0) = a,\ v(t_0) = b,\ \dot{u}(t_0) = c,\ \dot{v}(t_0) = d, \tag{9.6}$$

and such that $u(t)$ and $v(t)$ are defined and smooth for all $t$ satisfying $|t-t_0| < \epsilon$, where $\epsilon$ is some positive number. Moreover, any two solutions of Eqs. 9.5 satisfying (9.6) agree for all values of $t$ such that $|t-t_0| < \epsilon'$, where $\epsilon'$ is some positive number $\leq \epsilon$.

We now apply these facts to the geodesic equations. Suppose that $\mathbf{p}$ lies in a patch $\boldsymbol{\sigma}(u,v)$ of $\mathcal{S}$, say $\mathbf{p} = \boldsymbol{\sigma}(a,b)$, and that $\mathbf{t} = c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v$, where $a, b, c$, and $d$ are scalars and the derivatives are evaluated at $u = a$, $v = b$. A unit-speed curve $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ passes through $\mathbf{p}$ at $t = t_0$ if and only if $u(t_0) = a$, $v(t_0) = b$, and has tangent vector $\mathbf{t}$ there if and only if

$$c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v = \mathbf{t} = \dot{\boldsymbol{\gamma}}(t_0) = \dot{u}(t_0)\boldsymbol{\sigma}_u + \dot{v}(t_0)\boldsymbol{\sigma}_v,$$

i.e., $\dot{u}(t_0) = c$, $\dot{v}(t_0) = d$. Thus, finding a (unit-speed) geodesic $\boldsymbol{\gamma}$ passing through $\mathbf{p}$ at $t = t_0$ and having tangent vector $\mathbf{t}$ is equivalent to solving the geodesic equations subject to the initial conditions (9.6). But we have said above that this problem has a unique solution. $\square$

## Example 9.2.5

We already know that all straight lines in a plane are geodesics. Since there is a straight line in the plane through any given point of the plane in any given direction parallel to the plane, it follows from Proposition 9.2.4 that there are no other geodesics.

## Example 9.2.6

Similarly, on a sphere, the great circles are the only geodesics, for there is clearly a great circle passing through any given point of the sphere in any given direction tangent to the sphere. (If $\mathbf{p}$ is the point and $\mathbf{t}$ the tangent direction, let $\Pi$ be the plane passing through the origin parallel to $\mathbf{p}$ and $\mathbf{t}$ (i.e., with normal $\mathbf{p} \times \mathbf{t}$); then take the intersection of the sphere with $\Pi$.)

The following consequence of Theorem 9.2.1 can also be used in some cases to find geodesics without solving the geodesic equations.

## Corollary 9.2.7

Any local isometry between two surfaces takes the geodesics of one surface to the geodesics of the other.

## Proof

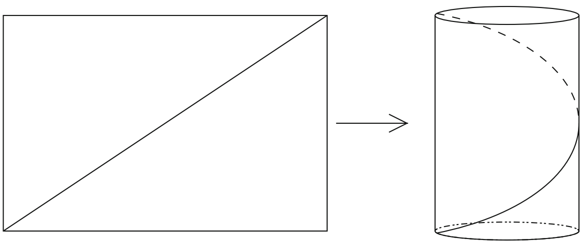
Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be the two surfaces, let $f : \mathcal{S}_1 \to \mathcal{S}_2$ be the local isometry, and let $\boldsymbol{\gamma}_1$ be a geodesic on $\mathcal{S}_1$. Let $\mathbf{p}$ be a point on $\boldsymbol{\gamma}_1$ and let $\boldsymbol{\sigma}(u, v)$ be a surface patch of $\mathcal{S}_1$ with $\mathbf{p}$ in its image. Then, the part of $\boldsymbol{\gamma}_1$ lying in the patch $\boldsymbol{\sigma}$ is of the form $\boldsymbol{\gamma}_1(t) = \boldsymbol{\sigma}(u(t), v(t))$ with $a < t < b$, say, where the smooth functions $u$ and $v$ satisfy the geodesic equations (9.2), with $E$, $F$ and $G$ being the coefficients of the first fundamental form of $\boldsymbol{\sigma}$. By Corollary6.2.3, $f \circ \boldsymbol{\sigma}$ is a patch of $\mathcal{S}_2$ with the *same* first fundamental form as $\boldsymbol{\sigma}$. Hence, by Theorem 9.2.1, $\boldsymbol{\gamma}_2(t) = f(\boldsymbol{\sigma}(u(t), v(t)))$, with $a < t < b$, is a geodesic on $\mathcal{S}_2$. This implies that $\ddot{\boldsymbol{\gamma}}_2$ is perpendicular to $\mathcal{S}_2$ at $f(\mathbf{p})$. As this is true for all $\mathbf{p}$, $\boldsymbol{\gamma}_2$ is a geodesic on $\mathcal{S}_2$. $\square$

## Example 9.2.8

On the unit cylinder $\mathcal{S}$ given by $x^2 + y^2 = 1$, we know that the circles obtained by intersecting $\mathcal{S}$ with planes parallel to the $xy$-plane are geodesics (since they are normal sections). We also know that the straight lines on $\mathcal{S}$ parallel to the $z$-axis are geodesics. However, these are certainly not the only geodesics, for there is only one geodesic of each of the two types passing through each point of $\mathcal{S}$ (whereas we know that there is a geodesic passing through each point in *any given tangent direction*).

To find the missing geodesics, we recall that $\mathcal{S}$ is locally isometric to the plane (see Example 6.2.4). In fact, the local isometry takes the point $(u, v, 0)$ of the $xy$-plane to the point $(\cos u, \sin u, v) \in \mathcal{S}$. By Corollary 9.2.7, this map takes geodesics on the plane (i.e., straight lines) to geodesics on $\mathcal{S}$, and vice versa. So to find all the geodesics on $\mathcal{S}$, we have only to find the images under the local isometry of all the straight lines in the plane. Any line not parallel to the $y$-axis has equation $y = mx + c$, where $m$ and $c$ are constants. Parametrizing this line by $x = u$, $y = mu + c$, we see that its image is the curve

$$\boldsymbol{\gamma}(u) = (\cos u, \sin u, mu + c)$$



on $\mathcal{S}$. Comparing with Example 2.1.3, we see that this is a *circular helix* of radius one and pitch $2\pi|m|$ (adding $c$ to the $z$-coordinate just translates the helix vertically). Note that if $m = 0$, we get the circular geodesics that we already know. Finally, any straight line in the $xy$-plane parallel to the $y$-axis is mapped by the local isometry to a straight line on $\mathcal{S}$ parallel to the $z$-axis, giving the other family of geodesics that we already know.

## *EXERCISES*

9.2.1 Show that, if $\mathbf{p}$ and $\mathbf{q}$ are distinct points of the unit cylinder, there are either two or infinitely many geodesics on the cylinder with end-points $\mathbf{p}$ and $\mathbf{q}$ (and which do not otherwise pass through $\mathbf{p}$ or $\mathbf{q}$). Which pairs $\mathbf{p}, \mathbf{q}$ have the former property?

9.2.2 Use Corollary 9.2.7 to find all the geodesics on a circular cone.

9.2.3 Find the geodesics on the unit cylinder by solving the geodesic equations.

9.2.4 Consider the following three properties that a curve $\boldsymbol{\gamma}$ on a surface may have:

(i) $\boldsymbol{\gamma}$ has constant speed.

(ii) $\boldsymbol{\gamma}$ satisfies the first of the geodesic equations (9.2).

(iii) $\boldsymbol{\gamma}$ satisfies the second of the geodesic equations (9.2).

Show, without using Theorem 9.2.1, that (ii) and (iii) together imply (i). Show also that if (i) holds and if $\boldsymbol{\gamma}$ is not a parameter curve, then (ii) and (iii) are equivalent.

9.2.5 Let $\boldsymbol{\gamma}(t)$ be a unit-speed curve on the helicoid

$$\boldsymbol{\sigma}(u, v) = (u \cos v, u \sin v, v)$$

(Exercise 4.2.6). Show that

$$\dot{u}^2 + (1 + u^2)\dot{v}^2 = 1$$

(a dot denotes $d/dt$). Show also that, if $\boldsymbol{\gamma}$ is a geodesic on $\boldsymbol{\sigma}$, then

$$\dot{v} = \frac{a}{1 + u^2},$$

where $a$ is a constant. Find the geodesics corresponding to $a = 0$ and $a = 1$.

Suppose that a geodesic $\boldsymbol{\gamma}$ on $\boldsymbol{\sigma}$ intersects a ruling at a point $\mathbf{p}$ a distance $D > 0$ from the $z$-axis, and that the angle between $\boldsymbol{\gamma}$ and the ruling at $\mathbf{p}$ is $\alpha$, where $0 < \alpha < \pi/2$. Show that the geodesic intersects the $z$-axis if $D > \cot\alpha$, but that if $D < \cot\alpha$ its smallest distance from the $z$-axis is $\sqrt{D^2\sin^2\alpha - \cos^2\alpha}$. Find the equation of the geodesic if $D = \cot\alpha$.

9.2.6 Verify directly that the differential equations in Proposition 9.2.3 are equivalent to the geodesic equations in Theorem 9.2.1.

## 9.3 Geodesics on surfaces of revolution

It turns out that, although the geodesic equations for a surface of revolution cannot usually be solved explicitly, they can be used to get a good *qualitative* understanding of the geodesics on such a surface.

We parametrize the surface of revolution in the usual way

$$\boldsymbol{\sigma}(u,v) = (f(u)\cos v, f(u)\sin v, g(u)),$$

where we assume that $f > 0$ and $\left(\frac{df}{du}\right)^2 + \left(\frac{dg}{du}\right)^2 = 1$ (see Example 5.3.2 – we used a dot there to denote $d/du$, but now a dot is reserved for $d/dt$, where $t$ is the parameter along a geodesic). We found in Example 6.1.3 that the first fundamental form of $\boldsymbol{\sigma}$ is $du^2 + f(u)^2dv^2$. Referring to Eq. 9.2,

$$\ddot{u} = f(u)\frac{df}{du}\dot{v}^2, \quad \frac{d}{dt}(f(u)^2\dot{v}) = 0. \tag{9.7}$$

We might as well consider unit-speed geodesics, so that

$$\dot{u}^2 + f(u)^2\dot{v}^2 = 1. \tag{9.8}$$

From this, we make the following easy deductions:

### Proposition 9.3.1

On the surface of revolution

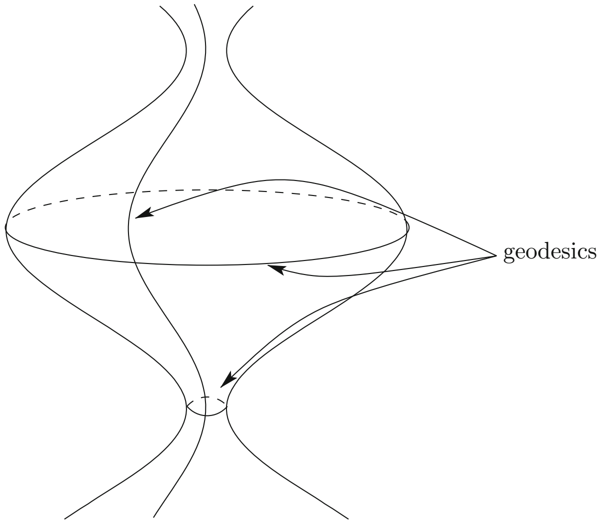
$$\boldsymbol{\sigma}(u,v) = (f(u)\cos v, f(u)\sin v, g(u)),$$

(i) Every meridian is a geodesic.

(ii) A parallel $u = u_0$ (say) is a geodesic if and only if $df/du = 0$ when $u = u_0$, i.e., $u_0$ is a stationary point of $f$.

## Proof

On a meridian, we have $v =$ constant so the second equation in (9.7) is obviously satisfied. Equation 9.8 gives $\dot{u} = \pm 1$, so $\dot{u}$ is constant and the first equation in (9.7) is also satisfied.




For (ii), note that if $u = u_0$ is constant, then by Eq. 9.8, $\dot{v} = \pm 1/f(u_0)$ is non-zero, and so the first equation in (9.7) holds only if $df/du = 0$. Conversely, if $df/du = 0$ when $u = u_0$, the first equation in (9.7) obviously holds, and the second holds because $\dot{v} = \pm 1/f(u_0)$ and $f(u) = f(u_0)$ are constant. $\square$

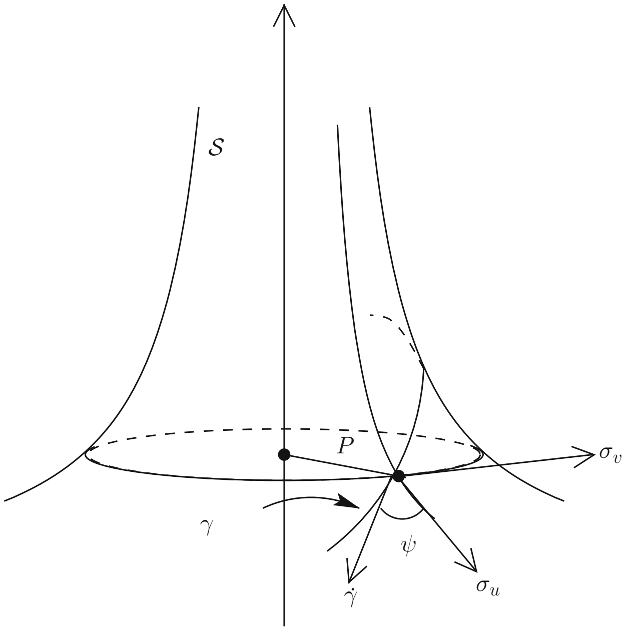
Of course, this proposition only gives some of the geodesics on a surface of revolution. The following result is very helpful in understanding the remaining geodesics.

## Proposition 9.3.2 (Clairaut's Theorem)

Let $\boldsymbol{\gamma}$ be a unit-speed curve on a surface of revolution $\mathcal{S}$, let $\rho : \mathcal{S} \to \mathbb{R}$ be the distance of a point of $\mathcal{S}$ from the axis of rotation, and let $\psi$ be the angle between $\dot{\boldsymbol{\gamma}}$ and the meridians of $\mathcal{S}$. If $\boldsymbol{\gamma}$ is a geodesic, then $\rho \sin\psi$ is constant along $\boldsymbol{\gamma}$. Conversely, if $\rho \sin\psi$ is constant along $\boldsymbol{\gamma}$, and if no part of $\boldsymbol{\gamma}$ is part of some parallel of $\mathcal{S}$, then $\boldsymbol{\gamma}$ is a geodesic.

By a 'part' of $\boldsymbol{\gamma}$ we mean $\boldsymbol{\gamma}(J)$, where $J$ is an open interval. The hypothesis there cannot be relaxed, for on a parallel $\psi = \pi/2$, and so $\rho \sin\psi$

is certainly constant. But parallels are not geodesics in general, as Proposition 9.3.1(ii) shows.




## Proof

Parametrizing $\mathcal{S}$ as in Proposition 9.3.1, we have $\rho = f(u)$. Note that $\boldsymbol{\sigma}_u$ and $\rho^{-1}\boldsymbol{\sigma}_v$ are unit vectors tangent to the meridians and parallels, respectively, and that they are perpendicular since $F = 0$. Assuming that $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ is unit-speed, we have

$$\dot{\boldsymbol{\gamma}} = \cos\psi\,\boldsymbol{\sigma}_u + \rho^{-1}\sin\psi\,\boldsymbol{\sigma}_v$$

(this equation actually serves to define the sign of $\psi$, which is left ambiguous in the statement of Clairaut's Theorem). Hence,

$$\boldsymbol{\sigma}_u \times \dot{\boldsymbol{\gamma}} = \rho^{-1}\sin\psi\,\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v.$$

Since $\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$, this gives

$$\dot{v}\,\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \rho^{-1}\sin\psi\,\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v.$$

Hence, $\rho\dot{v} = \sin\psi$ and so

$$\rho\sin\psi = \rho^2\dot{v}.$$

But the second equation in (9.7) shows that this is a constant, say $\Omega$, along the geodesic.

For the converse, if $\rho \sin\psi$ is a constant $\Omega$ along a unit-speed curve $\boldsymbol{\gamma}$ in $\mathcal{S}$, the above argument shows that the second equation in (9.7) is satisfied, and we must show that the first equation in (9.7) is satisfied too. Since

$$\dot{v} = \frac{\sin\psi}{\rho} = \frac{\Omega}{\rho^2}, \tag{9.9}$$

Eq. 9.8 gives

$$\dot{u}^2 = 1 - \frac{\Omega^2}{\rho^2}. \tag{9.10}$$

Differentiating both sides with respect to $t$ gives

$$\begin{aligned} 2\dot{u}\ddot{u} &= \frac{2\Omega^2}{\rho^3}\dot{\rho} = \frac{2\Omega^2}{\rho^3}\frac{d\rho}{du}\dot{u}, \\ \therefore \qquad \dot{u}\left(\ddot{u} - \rho\frac{d\rho}{du}\dot{v}^2\right) &= 0. \end{aligned}$$

If the term in brackets does not vanish at some point of the curve, say at $\boldsymbol{\gamma}(t_0) = \boldsymbol{\sigma}(u_0, v_0)$, there will be a number $\epsilon > 0$ such that it does not vanish for $|t - t_0| < \epsilon$. But then $\dot{u} = 0$ for $|t - t_0| < \epsilon$, and so $\boldsymbol{\gamma}$ coincides with the parallel $u = u_0$ when $|t - t_0| < \epsilon$, contrary to our assumption. Hence, the term in brackets must vanish everywhere on $\boldsymbol{\gamma}$, i.e.,

$$\ddot{u} = \rho\frac{d\rho}{du}\dot{v}^2,$$

showing that the first equation in (9.7) is indeed satisfied. $\square$

Clairaut's Theorem has a simple mechanical interpretation. Recall that the geodesics on a surface $\mathcal{S}$ are the curves traced on $\mathcal{S}$ by a particle subject to no forces except a force normal to $\mathcal{S}$ that constrains it to move on $\mathcal{S}$. When $\mathcal{S}$ is a surface of revolution, the force at a point $\mathbf{p} \in \mathcal{S}$ lies in the plane containing the axis of revolution and $\mathbf{p}$, and so has no moment about the axis. It follows that the angular momentum $\Omega$ of the particle about the axis is constant. But, if the particle moves along a unit-speed geodesic, the component of its velocity along the parallel through $\mathbf{p}$ is $\sin\psi$, so its angular momentum about the axis is proportional to $\rho \sin\psi$.

### Example 9.3.3

We use Clairaut's theorem to determine the geodesics on the pseudosphere:

$$\boldsymbol{\sigma}(u, v) = (e^u \cos v, e^u \sin v, \sqrt{1 - e^{2u}} - \cosh^{-1}(e^{-u})).$$

We found in Section 8.3 that its first fundamental form is

$$du^2 + e^{2u} dv^2.$$

It is convenient to reparametrize by setting $w = e^{-u}$. The reparametrized surface is

$$\tilde{\boldsymbol{\sigma}}(v, w) = \left( \frac{1}{w} \cos v, \frac{1}{w} \sin v, \sqrt{1 - \frac{1}{w^2}} - \cosh^{-1} w \right),$$

and its first fundamental form is

$$\frac{dv^2 + dw^2}{w^2}. \tag{9.11}$$

We must have $w > 1$ for $\tilde{\boldsymbol{\sigma}}$ to be well defined and smooth.

If $\boldsymbol{\gamma}(t) = \tilde{\boldsymbol{\sigma}}(v(t), w(t))$ is a unit-speed geodesic, the unit-speed condition gives

$$\dot{v}^2 + \dot{w}^2 = w^2, \tag{9.12}$$

and Clairaut's theorem gives

$$\frac{1}{w} \sin \psi = \frac{1}{w^2} \dot{v} = \Omega, \tag{9.13}$$

where $\Omega$ is a constant, since $\rho = 1/w$. Thus, $\dot{v} = \Omega w^2$. If $\Omega = 0$, we get a meridian $v =$ constant. Assuming now that $\Omega \neq 0$ and substituting in Eq. 9.12 gives

$$\dot{w} = \pm w \sqrt{1 - \Omega^2 w^2}.$$

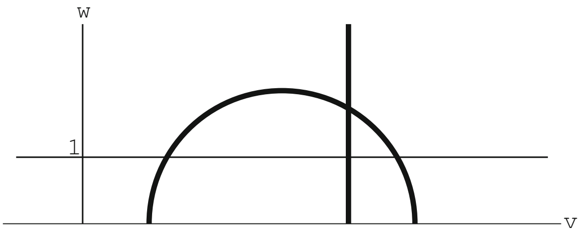
Hence, along the geodesic,

$$\frac{dv}{dw} = \frac{\dot{v}}{\dot{w}} = \pm \frac{\Omega w}{\sqrt{1 - \Omega^2 w^2}},$$

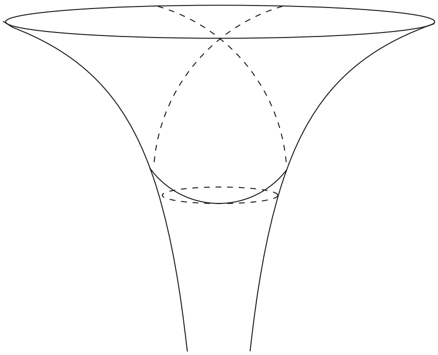
$$\therefore \quad (v - v_0) = \mp \frac{1}{\Omega} \sqrt{1 - \Omega^2 w^2}, \tag{9.14}$$

$$\therefore \quad (v - v_0)^2 + w^2 = \frac{1}{\Omega^2},$$

where $v_0$ is a constant. So the geodesics are the images under $\tilde{\boldsymbol{\sigma}}$ of the parts of the circles in the $vw$-plane given by Eq. 9.14 and lying in the region $w > 1$. Note that these circles all have centre on the $v$-axis, and so intersect the $v$-axis perpendicularly. The meridians correspond to straight lines perpendicular to the $v$-axis.




The corresponding geodesics on the pseudosphere itself are shown below. Note that the geodesics cannot be extended indefinitely, in one direction in the case of the meridians and in both directions for the others. This is because the geodesics 'run into' the circular edge of the pseudosphere in the $xy$-plane. A bug walking at constant speed along such a geodesic would reach the edge in a finite time, and thus would suffer the fate feared by ancient mariners of falling off the edge of the world. In terms of the $vw$-plane, the reason for this is that the line $w = 1$ is a boundary of the region that corresponds to the pseudosphere and the straight lines and semicircles that correspond to the geodesics cross this line.



Clairaut's theorem can often be used to determine the *qualitative* behaviour of the geodesics on a surface $\mathcal{S}$, when solving the geodesic differential equations explicitly may be difficult or impossible. Note first that, in general, there are two geodesics passing through any given point $\mathbf{p} \in \mathcal{S}$ with a given angular momentum $\Omega$, for $\dot{v}$ is determined by Eq. 9.9 and $\dot{u}$ up to sign by Eq. 9.10. In fact, one geodesic is obtained from the other by reflecting in the plane through $\mathbf{p}$ containing the of rotation (which changes $\Omega$ to $-\Omega$) followed by changing the parameter $t$ of the geodesic to $-t$ (which changes the angular momentum back to $\Omega$ again).

The discussion in the preceding paragraph shows that we may as well assume that $\Omega > 0$, which we do from now on. Then, Eq. 9.10 shows that *the geodesic is confined to the part of $\mathcal{S}$ which is at a distance $\geq \Omega$ from the axis.*

If all of $\mathcal{S}$ is a distance $> \Omega$ from the axis, the geodesic will cross every parallel of $\mathcal{S}$. For otherwise, $u$ would be bounded above or below on $\mathcal{S}$, say the former. Let $u_0$ be the least upper bound of $u$ on the geodesic, and let $\Omega + 2\epsilon$, where $\epsilon > 0$, be the radius of the parallel $u = u_0$. If $u$ is sufficiently close to $u_0$, the radius of the corresponding parallel will be $\geq \Omega + \epsilon$, and on the part of the geodesic lying in this region we shall have

$$|\dot{u}| \geq \sqrt{1 - \left(\frac{\Omega}{\Omega + \epsilon}\right)^2} > 0$$

by Eq. 9.10. But this clearly implies that the geodesic will cross $u = u_0$, contradicting our assumption.

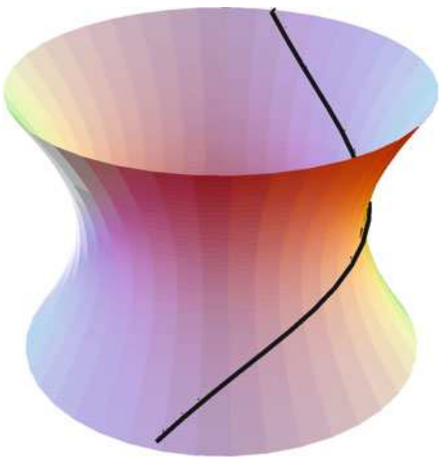
Thus, the interesting case is that in which part of $\mathcal{S}$ is within a distance $\Omega$ of the axis. The discussion of this case will be clearer if we consider a concrete example whose geodesics nevertheless exhibit essentially all possible forms of behaviour.

## Example 9.3.4

We consider the hyperboloid of one sheet obtained by rotating the hyperbola

$$x^2 - z^2 = 1, \quad x > 0,$$

in the $xz$-plane around the $z$-axis. Since all of the surface is at a distance $\geq 1$ from the $z$-axis, we have seen above that, if $0 \leq \Omega < 1$, a geodesic with angular momentum $\Omega$ crosses every parallel of the hyperboloid and so extends from $z = -\infty$ to $z = \infty$.



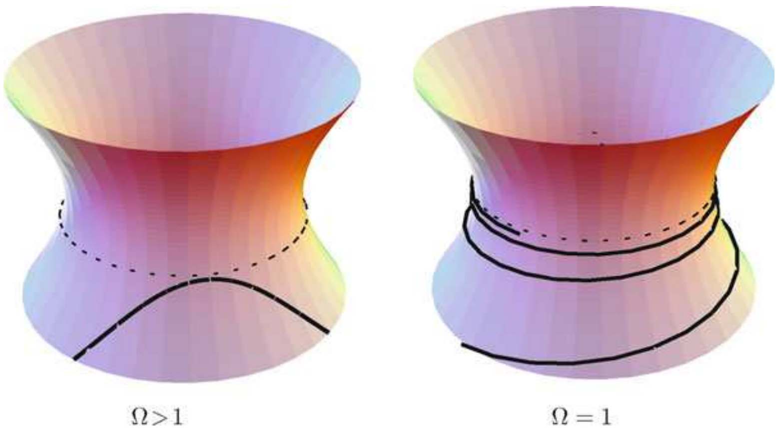
$0 < \Omega < 1$

*Suppose now that* $\Omega > 1$. Then the geodesic is confined to one of the two regions

$$z \geq \sqrt{\Omega^2 - 1}, \quad z \leq -\sqrt{\Omega^2 - 1},$$

which are bounded by circles $\Gamma^+$ and $\Gamma^-$, respectively, of radius $\Omega$. Let $\mathbf{p}$ be a point on $\Gamma^-$, and consider the geodesic $\mathcal{C}$ that passes through $\mathbf{p}$ and is tangent to $\Gamma^-$ there. Then, $\psi = \pi/2$ and $\rho = \Omega$ at $\mathbf{p}$, so $\mathcal{C}$ has angular momentum $\Omega$. Now $\mathcal{C}$ cannot be contained in $\Gamma^-$, since $\Gamma^-$ is not a geodesic (by Proposition 9.3.1(ii)), so $\mathcal{C}$ must head into the region below $\Gamma^-$ as it leaves $\mathbf{p}$. Moreover, $\mathcal{C}$ must be symmetric about $\mathbf{p}$, since reflection in the plane through $\mathbf{p}$ containing the $z$-axis takes $\mathcal{C}$ to another geodesic that also passes through $\mathbf{p}$ and is tangent to $\Gamma^-$ there, and so must coincide with $\mathcal{C}$ by the uniqueness part of Corollary 9.2.4. Since $\dot{u} \neq 0$ in the region below $\Gamma^-$ by Eq. 9.10, the geodesic crosses every parallel below $\Gamma^-$ and $z \to -\infty$ as $t \to \pm\infty$.

Suppose now that $\tilde{\mathcal{C}}$ is *any* geodesic with angular momentum $\Omega > 1$ in the region below $\Gamma^-$. Then a suitable rotation around the $z$-axis will cause $\tilde{\mathcal{C}}$ to intersect $\mathcal{C}$, say at $\mathbf{q}$, and so to coincide with it (possibly after reflecting in the plane through $\mathbf{q}$ containing the $z$-axis and changing $t$ to $-t$). We have therefore described the behaviour of every geodesic with angular momentum $\Omega > 1$ that is confined to the region below $\Gamma^-$. Of course, the geodesics with angular momentum $\Omega > 1$ in the region above $\Gamma^+$ are obtained by reflecting those below $\Gamma^-$ in the $xy$-plane.




*Suppose finally that* $\Omega = 1$. Let $\mathcal{C}$ be a geodesic with angular momentum 1 passing through a point $\mathbf{p}$. If $\mathbf{p}$ is on the waist $\Gamma$ of the hyperboloid (i.e., the unit circle in the $xy$-plane), which is a geodesic by Proposition 9.3.1(ii), then $\rho = 1$ at $\mathbf{p}$ and so $\psi = \pi/2$ and $\mathcal{C}$ is tangent to $\Gamma$ at $\mathbf{p}$. It must therefore coincide with $\Gamma$. If, on the other hand, $\mathbf{p}$ is in the region below $\Gamma$, then $0 < \psi < \pi/2$

at $\mathbf{p}$, so as it leaves $\mathbf{p}$ in one direction, $\mathcal{C}$ approaches $\Gamma$. It must in fact get arbitrarily close to $\Gamma$. For if it were to stay always below a parallel $\tilde{\Gamma}$ of radius $1+\epsilon$, say (with $\epsilon > 0$), then we would have

$$|\dot{u}| \geq \sqrt{1 - \left(\frac{1}{1+\epsilon}\right)^2}$$

everywhere along $\mathcal{C}$ by Eq. 9.10, which clearly implies that $\mathcal{C}$ must cross every parallel, contradicting our assumption. So, if $\Omega = 1$, the geodesic spirals around the hyperboloid approaching, and getting arbitrarily close to, $\Gamma$ but never quite reaching it.

### *EXERCISES*

9.3.1 There is another way to see that all the meridians, and the parallels corresponding to the stationary points of $f$, are geodesics on a surface of revolution considered in this section. What is it?

9.3.2 Describe qualitatively the geodesics on:

(i) A spheroid, obtained by rotating an ellipse around one of its axes.

(ii) A torus (Exercise 4.2.5).

9.3.3 Show that a geodesic on the pseudosphere with non-zero angular momentum $\Omega$ intersects itself if and only if $\Omega < (1+\pi^2)^{-1/2}$. How many self-intersections are there in that case?

9.3.4 Show that if we reparametrize the pseudosphere as in Exercise 8.3.1(ii), the geodesics on the pseudosphere correspond to segments of straight lines and circles in the parameter plane that intersect the boundary of the disc orthogonally. Deduce that, in the parametrization of Exercise 8.3.1(iii), the geodesics correspond to segments of *straight lines* in the parameter plane. We shall see in Section 10.4 that there are very few surfaces that have parametrizations with this property.

## 9.4 Geodesics as shortest paths

Everyone knows that the straight line segment joining two points $\mathbf{p}$ and $\mathbf{q}$ in a plane is the shortest path between $\mathbf{p}$ and $\mathbf{q}$ (see Exercise 1.2.4). It is almost as well known that great circles are the shortest paths on a sphere (Proposition 6.5.1). And we have seen that the straight lines are the geodesics in a plane, and the great circles are the geodesics on a sphere.

To see the connection between geodesics and shortest paths on an arbitrary surface $\mathcal{S}$, we consider a unit-speed curve $\boldsymbol{\gamma}$ on $\mathcal{S}$ passing through two fixed points $\mathbf{p}, \mathbf{q} \in \mathcal{S}$. If $\boldsymbol{\gamma}$ is a shortest path on $\mathcal{S}$ from $\mathbf{p}$ to $\mathbf{q}$, then the part of $\boldsymbol{\gamma}$ contained in any surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$ must be the shortest path between any two of its points. For if $\mathbf{p}'$ and $\mathbf{q}'$ are any two points of $\boldsymbol{\gamma}$ in (the image of) $\boldsymbol{\sigma}$, and if there were a shorter path in $\boldsymbol{\sigma}$ from $\mathbf{p}'$ to $\mathbf{q}'$ than $\boldsymbol{\gamma}$, we could replace the part of $\boldsymbol{\gamma}$ between $\mathbf{p}'$ and $\mathbf{q}'$ by this shorter path, thus giving a shorter path from $\mathbf{p}$ to $\mathbf{q}$ in $\mathcal{S}$.

We may therefore consider a path $\boldsymbol{\gamma}$ entirely contained in a surface patch $\boldsymbol{\sigma}$. To test whether $\boldsymbol{\gamma}$ has smaller length than any other path in $\boldsymbol{\sigma}$ passing through two fixed points $\mathbf{p}, \mathbf{q}$ on $\boldsymbol{\sigma}$; we embed $\boldsymbol{\gamma}$ in a smooth family of curves on $\boldsymbol{\sigma}$ passing through $\mathbf{p}$ and $\mathbf{q}$. By such a family, we mean a curve $\boldsymbol{\gamma}^\tau$ on $\boldsymbol{\sigma}$, for each $\tau$ in an open interval $(-\delta, \delta)$, such that

(i) there is an $\epsilon > 0$ such that $\boldsymbol{\gamma}^\tau(t)$ is defined for all $t \in (-\epsilon, \epsilon)$ and all $\tau \in (-\delta, \delta)$;

(ii) for some $a, b$ with $-\epsilon < a < b < \epsilon$, we have

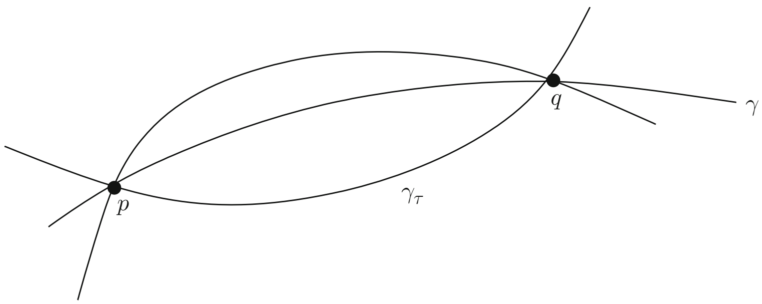
$$\boldsymbol{\gamma}^\tau(a) = \mathbf{p} \quad \text{and} \quad \boldsymbol{\gamma}^\tau(b) = \mathbf{q} \quad \text{for all} \quad \tau \in (-\delta, \delta);$$

(iii) the map from the rectangle $(-\delta, \delta) \times (-\epsilon, \epsilon)$ into $\mathbb{R}^3$ given by

$$(\tau, t) \mapsto \boldsymbol{\gamma}^\tau(t)$$

is smooth;

(iv) $\boldsymbol{\gamma}^0 = \boldsymbol{\gamma}$.




The length of the part of $\boldsymbol{\gamma}^\tau$ between $\mathbf{p}$ and $\mathbf{q}$ is

$$\mathcal{L}(\tau) = \int_a^b \parallel \dot{\boldsymbol{\gamma}}^\tau \parallel \, dt,$$

where a dot denotes $d/dt$.

### Theorem 9.4.1

With the above notation, the unit-speed curve $\boldsymbol{\gamma}$ is a geodesic if and only if

$$\frac{d}{d\tau}\mathcal{L}(\tau) = 0 \quad \text{when } \tau = 0$$

for all families of curves $\boldsymbol{\gamma}^\tau$ with $\boldsymbol{\gamma}^0 = \boldsymbol{\gamma}$.

Note that although we assumed that $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0$ is unit-speed, we *cannot assume that* $\boldsymbol{\gamma}^\tau$ *is unit-speed if* $\tau \neq 0$, as this would imply that the length of the segment of $\boldsymbol{\gamma}^\tau$ corresponding to $a \leq t \leq b$ is independent of $\tau$.

### Proof

We use the formula for 'differentiating under the integral sign': if $f(\tau, t)$ is smooth,

$$\frac{d}{d\tau}\int f(\tau, t)dt = \int \frac{\partial f}{\partial \tau}dt.$$

Thus,

$$\begin{aligned}
\frac{d}{d\tau}\mathcal{L}(\tau) &= \frac{d}{d\tau}\int_a^b \parallel \dot{\boldsymbol{\gamma}}^\tau \parallel \, dt \\
&= \frac{d}{d\tau}\int_a^b (E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{1/2}\, dt \\
&= \int_a^b \frac{\partial}{\partial \tau}(g(\tau, t)^{1/2})\, dt \qquad (9.15)\\
&= \frac{1}{2}\int_a^b g(\tau, t)^{-1/2}\frac{\partial g}{\partial \tau}\, dt,
\end{aligned}$$

where

$$g(\tau, t) = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2$$

and a dot denotes $d/dt$. Now,

$$\begin{aligned}
\frac{\partial g}{\partial \tau} &= \frac{\partial E}{\partial \tau}\dot{u}^2 + 2\frac{\partial F}{\partial \tau}\dot{u}\dot{v} + \frac{\partial G}{\partial \tau}\dot{v}^2 + 2E\dot{u}\frac{\partial \dot{u}}{\partial \tau} + 2F\left(\frac{\partial \dot{u}}{\partial \tau}\dot{v} + \dot{u}\frac{\partial \dot{v}}{\partial \tau}\right) + 2G\dot{v}\frac{\partial \dot{v}}{\partial \tau} \\
&= \left(E_u\frac{\partial u}{\partial \tau} + E_v\frac{\partial v}{\partial \tau}\right)\dot{u}^2 + 2\left(F_u\frac{\partial u}{\partial \tau} + F_v\frac{\partial v}{\partial \tau}\right)\dot{u}\dot{v} + \left(G_u\frac{\partial u}{\partial \tau} + G_v\frac{\partial v}{\partial \tau}\right)\dot{v}^2 \\
&\quad + 2E\dot{u}\frac{\partial^2 u}{\partial \tau \partial t} + 2F\left(\frac{\partial^2 u}{\partial \tau \partial t}\dot{v} + \dot{u}\frac{\partial^2 v}{\partial \tau \partial t}\right) + 2G\dot{v}\frac{\partial^2 v}{\partial \tau \partial t} \\
&= (E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2)\frac{\partial u}{\partial \tau} + (E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2)\frac{\partial v}{\partial \tau} \\
&\quad + 2(E\dot{u} + F\dot{v})\frac{\partial^2 u}{\partial \tau \partial t} + 2(F\dot{u} + G\dot{v})\frac{\partial^2 v}{\partial \tau \partial t}.
\end{aligned}$$

The contribution to the integral in Eq. 9.15 coming from the terms involving the second partial derivatives is

$$\begin{aligned}
&\int_a^b g^{-1/2}\left\{(E\dot{u}+F\dot{v})\frac{\partial^2 u}{\partial\tau\partial t} +(F\dot{u}+G\dot{v})\frac{\partial^2 v}{\partial\tau\partial t}\right\} dt \\
&\quad = g^{-1/2}\left\{(E\dot{u}+F\dot{v})\frac{\partial u}{\partial\tau}+(F\dot{u}+G\dot{v})\frac{\partial v}{\partial\tau}\right\}\Bigg|_{t=a}^{t=b} \\
&\qquad -\int_a^b\left(\frac{\partial}{\partial t}\left\{g^{-1/2}(E\dot{u}+F\dot{v})\right\}\frac{\partial u}{\partial\tau}+\frac{\partial}{\partial t}\left\{g^{-1/2}(F\dot{u}+G\dot{v})\right\}\frac{\partial v}{\partial\tau}\right) dt,
\end{aligned} \tag{9.16}$$

using integration by parts. Now, since $\boldsymbol{\gamma}^\tau(a)$ and $\boldsymbol{\gamma}^\tau(b)$ are independent of $\tau$ (being equal to $\mathbf{p}$ and $\mathbf{q}$, respectively), we have

$$\frac{\partial\boldsymbol{\gamma}^\tau}{\partial\tau}=\mathbf{0} \quad \text{when } t=a \text{ or } b.$$

Since

$$\frac{\partial\boldsymbol{\gamma}^\tau}{\partial\tau}=\frac{\partial u}{\partial\tau}\boldsymbol{\sigma}_u+\frac{\partial v}{\partial\tau}\boldsymbol{\sigma}_v,$$

we see that

$$\frac{\partial u}{\partial\tau}=\frac{\partial v}{\partial\tau}=0 \quad \text{when } t=a \text{ or } b.$$

Hence, the first term on the right-hand side of Eq. 9.16 is zero. Inserting the remaining terms in Eq. 9.16 back into Eq. 9.15, we get

$$\frac{d}{d\tau}\mathcal{L}(\tau)=\int_a^b\left(U\frac{\partial u}{\partial\tau}+V\frac{\partial v}{\partial\tau}\right) dt, \tag{9.17}$$

where

$$\begin{aligned}
U(\tau,t)&=\frac{1}{2}g^{-1/2}(E_u\dot{u}^2+2F_u\dot{u}\dot{v}+G_u\dot{v}^2)-\frac{d}{dt}\left\{g^{-1/2}(E\dot{u}+F\dot{v})\right\},\\
V(\tau,t)&=\frac{1}{2}g^{-1/2}(E_v\dot{u}^2+2F_v\dot{u}\dot{v}+G_v\dot{v}^2)-\frac{d}{dt}\left\{g^{-1/2}(F\dot{u}+G\dot{v})\right\}.
\end{aligned} \tag{9.18}$$

Now $\boldsymbol{\gamma}^0=\boldsymbol{\gamma}$ is unit-speed, so since $\parallel \dot{\boldsymbol{\gamma}}^\tau \parallel^2= g(\tau,t)$, we have $g(\tau,t)=1$ for all $t$ when $\tau=0$. Comparing Eq. 9.18 with the geodesic equations in (9.2), we see that, if $\boldsymbol{\gamma}$ is a geodesic, then $U=V=0$ when $\tau=0$, and hence by Eq. 9.17,

$$\frac{d}{d\tau}\mathcal{L}(\tau)=0 \quad \text{when } \tau=0.$$

For the converse, we have to show that, if

$$\int_a^b\left(U\frac{\partial u}{\partial\tau}+V\frac{\partial v}{\partial\tau}\right) dt=0 \quad \text{when } \tau=0 \tag{9.19}$$

for *all* families of curves $\boldsymbol{\gamma}^\tau$, then $U = V = 0$ when $\tau = 0$ (since this will prove that $\boldsymbol{\gamma}$ satisfies the geodesic equations). Assume, then, that condition (9.19) holds, and suppose, for example, that $U \neq 0$ when $\tau = 0$. We will show that this leads to a contradiction.

Since $U \neq 0$ when $\tau = 0$, there is some $t_0 \in (a, b)$ such that $U(0, t_0) \neq 0$, say $U(0, t_0) > 0$. Since $U$ is a continuous function, there exists $\eta > 0$ such that

$$U(0, t) > 0 \quad \text{if } t \in (t_0 - \eta, t_0 + \eta).$$

Let $\phi$ be a smooth function such that

$$\phi(t) > 0 \quad \text{if } t \in (t_0 - \eta, t_0 + \eta) \quad \text{and} \quad \phi(t) = 0 \quad \text{if } t \notin (t_0 - \eta, t_0 + \eta). \tag{9.20}$$

(The construction of such a function $\phi$ is outlined in Exercise 9.4.3.) Suppose that $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$, and consider the family of curves $\boldsymbol{\gamma}^\tau(t) = \boldsymbol{\sigma}(u(\tau, t), v(\tau, t))$, where

$$u(\tau, t) = u(t) + \tau\phi(t), \quad v(\tau, t) = v(t).$$

Then, $\partial u/\partial\tau = \phi$ and $\partial v/\partial\tau = 0$ for all $\tau$ and $t$, so Eq. 9.19 gives

$$0 = \int_a^b \left( U\frac{\partial u}{\partial \tau} + V\frac{\partial v}{\partial \tau} \right)\bigg|_{\tau=0} dt = \int_{t_0-\eta}^{t_0+\eta} U(0, t)\phi(t)\, dt. \tag{9.21}$$

But $U(0, t)$ and $\phi(t)$ are both $> 0$ for all $t \in (t_0 - \eta, t_0 + \eta)$, so the integral on the right-hand side of Eq. 9.21 is $> 0$. This contradiction proves that we must have $U(0, t) = 0$ for all $t \in (a, b)$. One proves similarly that $V(0, t) = 0$ for all $t \in (a, b)$. Together, these results prove that $\boldsymbol{\gamma}$ satisfies the geodesic equations. $\square$

It is worth making several comments on Theorem 9.4.1 to be clear about what it implies, and also what it does not imply.

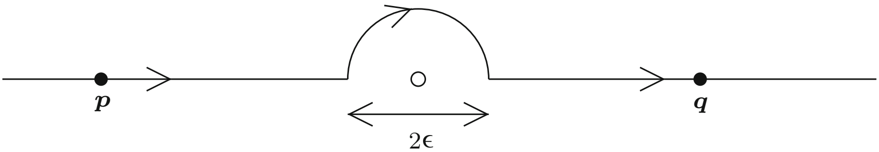
First, if $\boldsymbol{\gamma}$ is a shortest path on $\boldsymbol{\sigma}$ from $\mathbf{p}$ to $\mathbf{q}$, then $\mathcal{L}(\tau)$ must have an absolute minimum when $\tau = 0$. This implies that $\frac{d}{d\tau}\mathcal{L}(\tau) = 0$ when $\tau = 0$, and hence by Theorem 9.4.1 that $\boldsymbol{\gamma}$ is a geodesic.

Second, if $\boldsymbol{\gamma}$ is a geodesic on $\boldsymbol{\sigma}$ passing through $\mathbf{p}$ and $\mathbf{q}$, then $\mathcal{L}(\tau)$ has a stationary point (extremum) when $\tau = 0$, but this need not be an absolute minimum, or even a local minimum, so $\boldsymbol{\gamma}$ need not be a shortest path from $\mathbf{p}$ to $\mathbf{q}$. For example, if $\mathbf{p}$ and $\mathbf{q}$ are two nearby points on a sphere, the short great circle arc joining $\mathbf{p}$ and $\mathbf{q}$ is the shortest path from $\mathbf{p}$ to $\mathbf{q}$ (this is not quite obvious – see below), but the long great circle arc joining $\mathbf{p}$ and $\mathbf{q}$ is also a geodesic – see the diagram preceding Proposition 6.5.1.

Third, in general, a shortest path joining two points on a surface may not exist. For example, consider the surface $\mathcal{S}$ consisting of the $xy$-plane with the origin removed. This is a perfectly good surface, but there is *no* shortest path on the surface from the point $\mathbf{p} = (-1, 0)$ to the point $\mathbf{q} = (1, 0)$. Of course, the shortest path should be the straight line segment joining the two points, but this does not lie entirely on the surface, since it passes through the origin which is not part of the surface. For a 'real life' analogy, imagine trying to walk from $\mathbf{p}$ to $\mathbf{q}$ but finding that there is a deep hole in the ground at the origin. The solution might be to walk in a straight line as long as possible, and then skirt around the hole at the last minute, say taking something like the route shown below. This path consists of two straight line segments of length $1 - \epsilon$, together with a semicircle of radius $\epsilon$, so its total length is

$$2(1 - \epsilon) + \pi\epsilon = 2 + (\pi - 2)\epsilon.$$

Of course, this is greater than the straight line distance 2, but it can be made as close as we like to 2 by taking $\epsilon$ sufficiently small. In the language of real analysis, the greatest lower bound of the lengths of curves on the surface joining $\mathbf{p}$ and $\mathbf{q}$ is 2, but there is no curve from $\mathbf{p}$ to $\mathbf{q}$ *in the surface* whose length is equal to this lower bound.




Finally, it can be proved that if a surface $\mathcal{S}$ is a *closed* subset of $\mathbb{R}^3$ (i.e., if the set of points of $\mathbb{R}^3$ that are *not* in $\mathcal{S}$ is an open subset of $\mathbb{R}^3$), and if there is *some* path in $\mathcal{S}$ joining any two points of $\mathcal{S}$, then there is always a shortest path joining any two points of $\mathcal{S}$. For example, a plane is a closed subset of $\mathbb{R}^3$, and so there is a shortest path joining any two points. This path must be a straight line, for by the first remark above it is a geodesic, and we know that the only geodesics on a plane are the straight lines. Similarly, a sphere is a closed subset of $\mathbb{R}^3$, and it follows that the short great circle arc joining two points on the sphere is the shortest path joining them. But the surface $\mathcal{S}$ considered above is *not* a closed subset of $\mathbb{R}^3$, for $(0, 0) \notin \mathcal{S}$, but any open ball containing $(0, 0)$ must clearly contain points of $\mathcal{S}$, and so the set of points not in $\mathcal{S}$ is not open.

Another property of surfaces that are closed subsets of $\mathbb{R}^3$ (that we shall also not prove) is that geodesics on such surfaces can be extended indefinitely, i.e., they can be defined on the whole of $\mathbb{R}$. This is clear for straight lines in the plane, for example, and for great circles on the sphere (although in the latter case the geodesics 'close up' after an increment in the unit-speed parameter equal to the circumference of the sphere). But, for the straight line $\gamma(t) = (t - 1, 0)$ on the surface $\mathcal{S}$ defined above, which passes through $\mathbf{p}$ when $t = 0$, the largest interval containing $t = 0$ on which it is defined as a curve *in the surface* is $(-\infty, 1)$. We encountered a less artificial example of this 'incompleteness' in

Example 9.3.3: the pseudosphere considered there fails to be a closed subset of $\mathbb{R}^3$ because the points of its boundary circle in the $xy$-plane are not in the surface.

## EXERCISES

9.4.1 The geodesics on a circular (half) cone were determined in Exercise 9.2.2. Interpreting 'line' as 'geodesic', which of the following (true) statements in plane Euclidean geometry are true for the cone?

(i) There is a line passing through any two points.

(ii) There is a unique line passing through any two distinct points.

(iii) Any two distinct lines intersect in at most one point.

(iv) There are lines that do not intersect each other.

(v) Any line can be continued indefinitely.

(vi) A line defines the shortest distance between any two of its points.

(vii) A line cannot intersect itself transversely (i.e., with two non-parallel tangent vectors at the point of intersection).

9.4.2 Show that the long great circle arc on $S^2$ joining the points $\mathbf{p} = (1, 0, 0)$ and $\mathbf{q} = (0, 1, 0)$ is not even a *local* minimum of the length function $\mathcal{L}$ (see the remarks following the proof of Theorem 9.4.1).

9.4.3 Construct a smooth function with the properties in (9.20) in the following steps:

(i) Show that, for all integers $n$ (positive and negative), $t^n e^{-1/t^2}$ tends to 0 as $t$ tends to 0.

(ii) Deduce from (i) that the function

$$\theta(t) = \begin{cases} e^{-1/t^2} & \text{if } t \geq 0, \\ 0 & \text{if } t \leq 0 \end{cases}$$

is smooth everywhere.

(iii) Show that the function

$$\psi(t) = \theta(1 + t)\theta(1 - t)$$

is smooth everywhere, that $\psi(t) > 0$ if $-1 < t < 1$, and that $\psi(t) = 0$ otherwise.

(iv) Show that the function

$$\phi(t) = \psi\left(\frac{t - t_0}{\eta}\right)$$

has the properties we want.

## 9.5 Geodesic coordinates

The existence of geodesics on a surface $\mathcal{S}$ allows us to construct a very useful atlas for $\mathcal{S}$. For this, let $\mathbf{p} \in \mathcal{S}$ and let $\boldsymbol{\gamma}$, with parameter $v$ say, be a unit-speed geodesic on $\mathcal{S}$ with $\boldsymbol{\gamma}(0) = \mathbf{p}$. For any value of $v$, let $\tilde{\boldsymbol{\gamma}}^v$, with parameter $u$, say, be a unit-speed geodesic such that $\tilde{\boldsymbol{\gamma}}^v(0) = \boldsymbol{\gamma}(v)$ and which is perpendicular to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(v)$ ($\tilde{\boldsymbol{\gamma}}^v$ is unique up to the reparametrization $u \mapsto -u$). We define $\boldsymbol{\sigma}(u, v) = \tilde{\boldsymbol{\gamma}}^v(u)$.

### Proposition 9.5.1

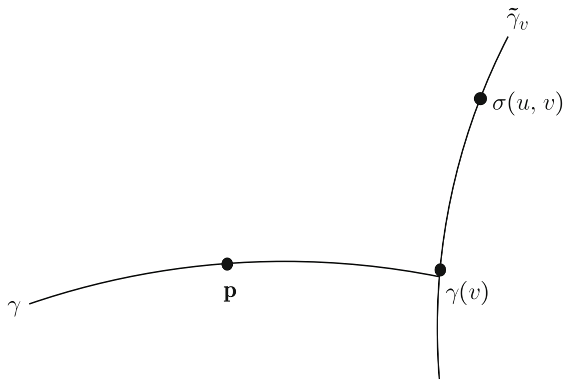
With the above notation, there is an open subset $U$ of $\mathbb{R}^2$ containing $(0, 0)$ such that $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is an allowable surface patch of $\mathcal{S}$. Moreover, the first fundamental form of $\boldsymbol{\sigma}$ is

$$du^2 + G(u, v)dv^2,$$

where $G$ is a smooth function on $U$ such that

$$G(0, v) = 1, \quad G_u(0, v) = 0,$$

whenever $(0, v) \in U$.



## Proof

The proof that $\boldsymbol{\sigma}$ is (for a suitable open set $U$) an allowable surface patch makes use of the inverse function theorem (see Section 5.6).

Note first that, for any value of $v$,

$$\boldsymbol{\sigma}_u(0,v) = \frac{d}{du}\tilde{\boldsymbol{\gamma}}^v(u)\bigg|_{u=0}, \quad \boldsymbol{\sigma}_v(0,v) = \frac{d}{dv}\tilde{\boldsymbol{\gamma}}^v(0) = \frac{d}{dv}\boldsymbol{\gamma}(v),$$

and that these are perpendicular unit vectors by construction. If

$$\boldsymbol{\sigma}(u,v) = (f(u,v), g(u,v), h(u,v)),$$

it follows that the Jacobian matrix

$$\begin{pmatrix} f_u & f_v \\ g_u & g_v \\ h_u & h_v \end{pmatrix}$$

has rank 2 when $u = v = 0$. Hence, at least one of its three $2 \times 2$ submatrices is invertible at $(0,0)$, say

$$\begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}. \tag{9.22}$$

By the Inverse Function Theorem 5.6.1, there is an open subset $U$ of $\mathbb{R}^2$ such that the map given by

$$F(u,v) = (f(u,v), g(u,v))$$

is a bijection from $U$ to an open subset $F(U)$ of $\mathbb{R}^2$, and such that its inverse map $F(U) \to U$ is also smooth. The matrix (9.22) is then invertible for all $(u,v) \in U$, and so $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent for $(u,v) \in U$. It follows that $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a surface patch.

As to the first fundamental form of $\boldsymbol{\sigma}$, note first that

$$E = \| \boldsymbol{\sigma}_u \|^2 = \left\| \frac{d}{du}\tilde{\boldsymbol{\gamma}}^v(u) \right\|^2 = 1$$

because $\tilde{\boldsymbol{\gamma}}^v$ is a unit-speed curve. Next, we apply the second of the geodesic equations (9.2) to $\tilde{\boldsymbol{\gamma}}^v$. The unit-speed parameter is $u$ and $v$ is constant, so we get $F_u = 0$. But when $u = 0$, we have already seen that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are perpendicular, so $F = 0$. It follows that $F = 0$ everywhere. Hence, the first fundamental form of $\boldsymbol{\sigma}$ is

$$du^2 + G(u,v)dv^2.$$

We have

$$G(0,v) = \| \boldsymbol{\sigma}_v(0,v) \|^2 = \left\| \frac{d\boldsymbol{\gamma}}{dv} \right\|^2 = 1$$

because $\boldsymbol{\gamma}$ is unit-speed. Finally, from the first geodesic equation in (9.2) applied to the geodesic $\boldsymbol{\gamma}$, for which $u = 0$ and $v$ is the unit-speed parameter, we get $G_u(0, v) = 0$. □

A surface patch $\boldsymbol{\sigma}$ constructed as above is called a *geodesic patch*, and $u$ and $v$ are called *geodesic coordinates*.

### Example 9.5.2

If $\mathbf{p}$ is a point on the equator of the unit sphere $S^2$, take $\boldsymbol{\gamma}$ to be the equator with parameter the longitude $\varphi$, and let $\tilde{\boldsymbol{\gamma}}^{\varphi}$ be the meridian parametrized by latitude $\theta$ and passing through the point on the equator with longitude $\varphi$. The corresponding geodesic patch is the usual latitude-longitude patch, for which the first fundamental form is

$$d\theta^2 + \cos^2\theta\, d\varphi^2,$$

in accordance with Proposition 9.5.1.

We give an application of geodesic coordinates in the proof of Theorem 10.3.1.

## EXERCISES

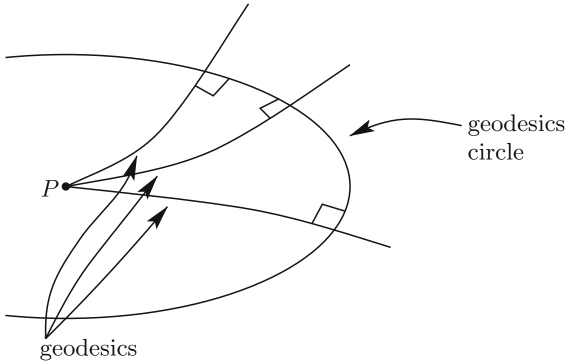
9.5.1 Let $P$ be a point of a surface $\mathcal{S}$ and let $\mathbf{v}$ be a unit tangent vector to $\mathcal{S}$ at $P$. Let $\boldsymbol{\gamma}^{\theta}(r)$ be the unit-speed geodesic on $\mathcal{S}$ passing through $P$ when $r = 0$ and such that the oriented angle $\widehat{\mathbf{v}\,\frac{d\boldsymbol{\gamma}^{\theta}}{dr}} = \theta$. It can be shown that $\boldsymbol{\sigma}(r, \theta) = \boldsymbol{\gamma}^{\theta}(r)$ is smooth for $-\epsilon < r < \epsilon$ and all values of $\theta$, where $\epsilon$ is some positive number, and that it is an allowable surface patch for $\mathcal{S}$ defined for $0 < r < \epsilon$ and for $\theta$ in any open interval of length $\leq 2\pi$. This is called a *geodesic polar patch* on $\mathcal{S}$.

Show that, if $0 < R < \epsilon$,

$$\int_0^R \left\| \frac{d\boldsymbol{\gamma}^{\theta}}{dr} \right\|^2 dr = R.$$

By differentiating both sides with respect to $\theta$, prove that

$$\boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_{\theta} = 0.$$




This is called *Gauss' Lemma* – geometrically, it means that the parameter curve $r = R$, called the *geodesic circle* with centre $P$ and radius $R$, is perpendicular to each of its radii, i.e., the geodesics passing through $P$. Deduce that the first fundamental form of $\boldsymbol{\sigma}$ is

$$dr^2 + G(r, \theta)d\theta^2,$$

for some smooth function $G(r, \theta)$.

9.5.2 Let $P$ and $Q$ be two points on a surface $\mathcal{S}$, and assume that there is a geodesic polar patch with centre $P$ as in Exercise 9.5.1 that also contains $Q$; suppose that $Q$ is the point $\boldsymbol{\sigma}(R, \alpha)$, where $0 < R < \epsilon$, $0 \leq \alpha < 2\pi$. Show in the following steps that the geodesic $\boldsymbol{\gamma}^{\alpha}(t) = \boldsymbol{\sigma}(t, \alpha)$ is (up to reparametrization) the unique *shortest* curve on $\mathcal{S}$ joining $P$ and $Q$.

(i) Let $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(f(t), g(t))$ be any curve in the patch $\boldsymbol{\sigma}$ joining $P$ and $Q$. We assume that $\boldsymbol{\gamma}$ passes through $P$ when $t = 0$ and through $Q$ when $t = R$ (this can always be achieved by a suitable reparametrization). Show that the length of the part of $\boldsymbol{\gamma}$ between $P$ and $Q$ is $\geq R$, and that $R$ is the length of the part of $\boldsymbol{\gamma}^{\alpha}$ between $P$ and $Q$.

(ii) Show that, if $\boldsymbol{\gamma}$ is *any* curve on $\mathcal{S}$ joining $P$ and $Q$ (not necessarily staying inside the patch $\boldsymbol{\sigma}$), the length of the part of $\boldsymbol{\gamma}$ between $P$ and $Q$ is $\geq R$.

(iii) Show that, if the part of a curve $\boldsymbol{\gamma}$ on $\mathcal{S}$ joining $P$ to $Q$ has length $R$, then $\boldsymbol{\gamma}$ is a reparametrization of $\boldsymbol{\gamma}^{\alpha}$.

# 10
# Gauss' Theorema Egregium

One of Gauss' most important discoveries about surfaces is that the Gaussian curvature is unchanged when the surface is bent without stretching. Gauss called this result 'egregium', and the Latin word for 'remarkable' has remained attached to his theorem ever since. We shall deduce the Theorema Egregium from two results which relate the first and second fundamental forms of a surface, and which have other important consequences.

## 10.1 The Gauss and Codazzi–Mainardi equations

It is natural to ask if there are any relations between the first and second fundamental forms of a surface. Note that, by Examples 6.1.2, 6.1.4, 7.1.1 and 7.1.2, the plane and the unit cylinder, when suitably parametrized, have the same first fundamental form but different second fundamental forms, and so the second fundamental form certainly cannot be 'deduced' from the first. Nevertheless, there are some nontrivial relations between the two forms.

### Proposition 10.1.1 (Codazzi–Mainardi Equations)

Let

$$Edu^2 + 2Fdudv + Gdv^2 \quad \text{and} \quad Ldu^2 + 2Mdudv + Ndv^2$$

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_10,

be the first and second fundamental forms of a surface patch $\boldsymbol{\sigma}(u,v)$, and define the Christoffel symbols as in Proposition 7.4.4. Then,

$$\begin{aligned} L_v - M_u &= L\Gamma_{12}^1 + M(\Gamma_{12}^2 - \Gamma_{11}^1) - N\Gamma_{11}^2, \\ M_v - N_u &= L\Gamma_{22}^1 + M(\Gamma_{22}^2 - \Gamma_{12}^1) - N\Gamma_{12}^2. \end{aligned} \tag{10.1}$$

## Proposition 10.1.2 (Gauss Equations)

If $K$ is the Gaussian curvature of the surface patch $\boldsymbol{\sigma}(u,v)$ in the preceding proposition, then

$$\begin{aligned} EK &= (\Gamma_{11}^2)_v - (\Gamma_{12}^2)_u + \Gamma_{11}^1\Gamma_{12}^2 + \Gamma_{11}^2\Gamma_{22}^2 - \Gamma_{12}^1\Gamma_{11}^2 - (\Gamma_{12}^2)^2 \\ FK &= (\Gamma_{12}^1)_u - (\Gamma_{11}^1)_v + \Gamma_{12}^2\Gamma_{12}^1 - \Gamma_{11}^2\Gamma_{22}^1 \\ FK &= (\Gamma_{12}^2)_v - (\Gamma_{22}^2)_u + \Gamma_{12}^1\Gamma_{12}^2 - \Gamma_{22}^1\Gamma_{11}^2 \\ GK &= (\Gamma_{22}^1)_u - (\Gamma_{12}^1)_v + \Gamma_{22}^1\Gamma_{11}^1 + \Gamma_{22}^2\Gamma_{12}^1 - (\Gamma_{12}^1)^2 - \Gamma_{12}^2\Gamma_{22}^1. \end{aligned}$$

## Proof

We prove both propositions simultaneously. Write down the equation $(\boldsymbol{\sigma}_{uu})_v = (\boldsymbol{\sigma}_{uv})_u$, using Proposition 7.4.4 for $\boldsymbol{\sigma}_{uu}$ and $\boldsymbol{\sigma}_{uv}$:

$$\begin{aligned} &(\Gamma_{11}^1\boldsymbol{\sigma}_u + \Gamma_{11}^2\boldsymbol{\sigma}_v + L\mathbf{N})_v = (\Gamma_{12}^1\boldsymbol{\sigma}_u + \Gamma_{12}^2\boldsymbol{\sigma}_v + M\mathbf{N})_u, \\ \therefore\quad &\left(\frac{\partial\Gamma_{11}^1}{\partial v} - \frac{\partial\Gamma_{12}^1}{\partial u}\right)\boldsymbol{\sigma}_u + \left(\frac{\partial\Gamma_{11}^2}{\partial v} - \frac{\partial\Gamma_{12}^2}{\partial u}\right)\boldsymbol{\sigma}_v + (L_v - M_u)\mathbf{N} \\ &= \Gamma_{12}^1\boldsymbol{\sigma}_{uu} + (\Gamma_{12}^2 - \Gamma_{11}^1)\boldsymbol{\sigma}_{uv} - \Gamma_{11}^2\boldsymbol{\sigma}_{vv} - L\mathbf{N}_v + M\mathbf{N}_u \qquad (10.2) \\ &= \Gamma_{12}^1(\Gamma_{11}^1\boldsymbol{\sigma}_u + \Gamma_{11}^2\boldsymbol{\sigma}_v + L\mathbf{N}) + (\Gamma_{12}^2 - \Gamma_{11}^1)(\Gamma_{12}^1\boldsymbol{\sigma}_u + \Gamma_{12}^2\boldsymbol{\sigma}_v + M\mathbf{N}) \\ &\quad -\Gamma_{11}^2(\Gamma_{22}^1\boldsymbol{\sigma}_u + \Gamma_{22}^2\boldsymbol{\sigma}_v + N\mathbf{N}) - L\mathbf{N}_v + M\mathbf{N}_u, \end{aligned}$$

using Proposition 7.4.4 again. Now, $\mathbf{N}_u$ and $\mathbf{N}_v$ are perpendicular to $\mathbf{N}$, and so are linear combinations of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Hence, equating $\mathbf{N}$ components on both sides of the last equation gives

$$L_v - M_u = L\Gamma_{12}^1 + M(\Gamma_{12}^2 - \Gamma_{11}^1) - N\Gamma_{11}^2,$$

which is the first of the Codazzi–Mainardi equations (10.1). The other equation follows in a similar way by equating coefficients of $\mathbf{N}$ in the equation $(\boldsymbol{\sigma}_{uv})_v = (\boldsymbol{\sigma}_{vv})_u$.

Now we use the formulas in the proof of Proposition 8.1.2 to express $\mathbf{N}_u$ and $\mathbf{N}_v$ in terms of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Equating coefficients of $\boldsymbol{\sigma}_u$ in Eq. 10.2 then gives

$$(\Gamma_{11}^1)_v - (\Gamma_{12}^1)_u = \Gamma_{12}^2\Gamma_{12}^1 - \Gamma_{11}^2\Gamma_{22}^1 + Lc - Ma \tag{10.3}$$

in the notation of the proof of Proposition 8.1.2, from which we find that

$$a = \frac{LG - MF}{EG - F^2}, \quad c = \frac{MG - NF}{EG - F^2}$$

and hence

$$Lc - Ma = \frac{L(MG - NF) - M(LG - MF)}{EG - F^2} = -\frac{F(LN - M^2)}{EG - F^2} = -FK$$

by Corollary 8.1.3. Substituting in Eq. 10.3 and rearranging gives the second of the Gauss equations. The other three are proved in the same way, equating coefficients of $\boldsymbol{\sigma}_v$ in $(\boldsymbol{\sigma}_{uu})_v = (\boldsymbol{\sigma}_{uv})_u$ and those of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ in $(\boldsymbol{\sigma}_{uv})_v = (\boldsymbol{\sigma}_{vv})_u$. □

The following theorem tells us that there are no other relations between the first and second fundamental forms other than those in Propositions 10.1.1 and 10.1.2.

## Theorem 10.1.3

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ and $\tilde{\boldsymbol{\sigma}} : U \to \mathbb{R}^3$ be surface patches with the same first and second fundamental forms. Then, there is a direct isometry $M$ of $\mathbb{R}^3$ such that $\tilde{\boldsymbol{\sigma}} = M(\boldsymbol{\sigma})$.

Moreover, let $V$ be an open subset of $\mathbb{R}^3$ and let $E, F, G, L, M$ and $N$ be smooth functions on $V$. Assume that $E > 0, G > 0, EG - F^2 > 0$ and that the equations in Propositions 10.1.1 and 10.1.2 hold, with $K = \frac{LN - M^2}{EG - F^2}$ and the Christoffel symbols defined as in Proposition 7.4.4. Then, if $(u_0, v_0) \in V$, there is an open set $U$ contained in $V$ and containing $(u_0, v_0)$, and a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$, such that $Edu^2 + 2Fdudv + Gdv^2$ and $Ldu^2 + 2Mdudv + Ndv^2$ are the first and second fundamental forms of $\boldsymbol{\sigma}$, respectively.

This theorem is an analogue for surfaces of Theorem 2.2.5, which shows that unit-speed plane curves are determined up to a direct isometry of $\mathbb{R}^3$ by their signed curvature. We shall not prove Theorem 10.1.3 here. The first part depends on uniqueness theorems for the solution of systems of ordinary differential equations, and is not particularly difficult. The second part is more sophisticated and depends on existence theorems for the solution of certain *partial* differential equations. The following example illustrates what is involved.

## Example 10.1.4

Consider the first and second fundamental forms $du^2 + dv^2$ and $-du^2$, respectively. Let us first see whether a surface patch with these first and second fundamental forms exists. Since all the coefficients of these forms are constant, all the Christoffel symbols are zero and the Codazzi–Mainardi equations are obviously satisfied. The first formula in Proposition 10.1.2 gives $K = 0$, so the only other condition to be checked is $LN - M^2 = 0$, and this clearly holds since $M = N = 0$. Theorem 10.1.3 therefore tells us that a surface patch with the given first and second fundamental forms exists.

To find it, we note that the Gauss equations give

$$\boldsymbol{\sigma}_{uu} = -\mathbf{N}, \quad \boldsymbol{\sigma}_{uv} = \mathbf{0}, \quad \boldsymbol{\sigma}_{vv} = \mathbf{0}.$$

The last two equations tell us that $\boldsymbol{\sigma}_v$ is a constant vector, say $\mathbf{a}$, so

$$\boldsymbol{\sigma}(u, v) = \mathbf{b}(u) + \mathbf{a}v, \tag{10.4}$$

where $\mathbf{b}$ is a function of $u$ only. The first equation then gives $\mathbf{N} = -\mathbf{b}''$ (a dash denoting $d/du$). We now need to use the expressions for $\mathbf{N}_u$ and $\mathbf{N}_v$ in terms of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ in the proof of Proposition 8.1.2. The matrix of the Weingarten map with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of the tangent plane is

$$\mathcal{F}_I^{-1}\mathcal{F}_{II} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix},$$

so Proposition 8.1.2 gives

$$\mathbf{N}_u = \boldsymbol{\sigma}_u, \quad \mathbf{N}_v = \mathbf{0}.$$

The second equation tells us nothing new, since we already know that $\mathbf{N} = -\mathbf{b}''$ depends only on $u$. The first equation gives

$$\mathbf{b}''' + \mathbf{b}' = \mathbf{0}.$$

Hence, $\mathbf{b}'' + \mathbf{b}$ is a constant vector, which we can take to be zero by applying a translation to $\boldsymbol{\sigma}$ (see Eq. 10.4). Then,

$$\mathbf{b}(u) = \mathbf{c}\cos u + \mathbf{d}\sin u,$$

where $\mathbf{c}$ and $\mathbf{d}$ are constant vectors, and $\mathbf{N} = -\mathbf{b}'' = \mathbf{b}$. This must be a unit vector for all values of $u$. It is easy to see that this is possible only if $\mathbf{c}$ and $\mathbf{d}$ are perpendicular unit vectors, in which case we can arrange that $\mathbf{c} = (1, 0, 0)$ and $\mathbf{d} = (0, 1, 0)$ by applying an isometry of $\mathbb{R}^3$, giving $\mathbf{b}(u) = (\cos u, \sin u, 0)$. Finally, $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \lambda\mathbf{N}$ for some non-zero scalar $\lambda$, so $\mathbf{b}' \times \mathbf{a} = \lambda\mathbf{b}$. This forces $\mathbf{a} = (0, 0, \lambda)$, and the patch is given by

$$\boldsymbol{\sigma}(u, v) = (\cos u, \sin u, \lambda v),$$

a parametrization of the unit cylinder (which the reader had probably guessed some time ago).

## EXERCISES

10.1.1 A surface patch has first and second fundamental forms

$$\cos^2 v\,du^2 + dv^2 \quad \text{and} \quad -\cos^2 v\,du^2 - dv^2,$$

respectively. Show that the surface is an open subset of a sphere of radius one. Write down a parametrization of $S^2$ with these first and second fundamental forms.

10.1.2 Show that there is no surface patch whose first and second fundamental forms are

$$du^2 + \cos^2 u\,dv^2 \quad \text{and} \quad \cos^2 u\,du^2 + dv^2,$$

respectively.

10.1.3 Suppose that a surface patch $\boldsymbol{\sigma}(v, w)$ has first and second fundamental forms

$$\frac{dv^2 + dw^2}{w^2} \quad \text{and} \quad Ldv^2 + Ndw^2,$$

respectively, where $w > 0$. Prove that $L$ and $N$ do not depend on $v$, that $LN = -1/w^4$ and that

$$Lw^5 \frac{dL}{dw} = 1 - L^2 w^4.$$

Solve this equation for $L$ and deduce that $\boldsymbol{\sigma}$ cannot be defined in the whole of the half-plane $w > 0$. Compare the discussion of the pseudosphere in Example 9.3.3.

10.1.4 Suppose that the first and second fundamental forms of a surface patch are $Edu^2 + Gdv^2$ and $Ldu^2 + Ndv^2$, respectively. Show that the Codazzi–Mainardi equations reduce to

$$L_v = \frac{1}{2}E_v \left(\frac{L}{E} + \frac{N}{G}\right), \quad N_u = \frac{1}{2}G_u \left(\frac{L}{E} + \frac{N}{G}\right).$$

Deduce that the principal curvatures $\kappa_1 = L/E$ and $\kappa_2 = N/G$ satisfy the equations

$$(\kappa_1)_v = \frac{E_v}{2E}(\kappa_2 - \kappa_1), \quad (\kappa_2)_u = \frac{G_u}{2G}(\kappa_1 - \kappa_2).$$

## 10.2 Gauss' remarkable theorem

We noted after Proposition 7.4.4 that the Christoffel symbols depend only on the coefficients of the *first* fundamental form. It follows from the formulas in Proposition 10.1.2 that this is also true of the Gaussian curvature $K$. In view of Corollary 6.2.3, we obtain:

### Theorem 10.2.1 (Gauss' Theorema Egregium)

The Gaussian curvature of a surface is preserved by local isometries.

This means, more explicitly, that if $\mathcal{S}_1$ and $\mathcal{S}_2$ are two surfaces and $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a local isometry, then for any point $\mathbf{p} \in \mathcal{S}_1$ the Gaussian curvature of $\mathcal{S}_1$ at $\mathbf{p}$ is equal to that of $\mathcal{S}_2$ at $f(\mathbf{p})$. The theorem is sometimes expressed by saying that the Gaussian curvature is an *intrinsic* property of a surface, for it implies that the Gaussian curvature could be measured by a bug living in the surface.

By substituting into the equations in Proposition 10.1.2 the expressions for the Christoffel symbols in Proposition 7.4.4, we can of course find an explicit expression for $K$ in terms of $E, F$ and $G$. (It appears that we could get four such formulas, one from each of the equations in Proposition 10.1.2, but these turn out to be the same.) Here is the result.

### Corollary 10.2.2

The Gaussian curvature is given by

$$K = \frac{\begin{vmatrix} -\frac{1}{2}E_{vv} + F_{uv} - \frac{1}{2}G_{uu} & \frac{1}{2}E_u & F_u - \frac{1}{2}E_v \\ F_v - \frac{1}{2}G_u & E & F \\ \frac{1}{2}G_v & F & G \end{vmatrix} - \begin{vmatrix} 0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & F \\ \frac{1}{2}G_u & F & G \end{vmatrix}}{(EG - F^2)^2}.$$

We shall not go through the details of this calculation, partly because the algebra is tedious, and partly because the following special cases are often all that is needed.

### Corollary 10.2.3

(i) If $F = 0$, we have

$$K = -\frac{1}{2\sqrt{EG}} \left\{ \frac{\partial}{\partial u} \left( \frac{G_u}{\sqrt{EG}} \right) + \frac{\partial}{\partial v} \left( \frac{E_v}{\sqrt{EG}} \right) \right\}.$$

(ii) If $E = 1$ and $F = 0$, we have

$$K = -\frac{1}{\sqrt{G}} \frac{\partial^2 \sqrt{G}}{\partial u^2}.$$

## Proof

If $F = 0$ Proposition 7.4.4 gives

$$\Gamma^1_{11} = \frac{E_u}{2E},\ \Gamma^2_{11} = -\frac{E_v}{2G},\ \Gamma^1_{12} = \frac{E_v}{2E},\ \Gamma^2_{12} = \frac{G_u}{2G},\ \Gamma^1_{22} = -\frac{G_u}{2E},\ \Gamma^2_{22} = \frac{G_v}{2G}.$$

Substituting into the first formula in Proposition 10.1.2 gives

$$\begin{aligned} EK &= -\left(\frac{E_v}{2G}\right)_v - \left(\frac{G_u}{2G}\right)_u + \frac{E_u G_u}{4EG} - \frac{E_v G_v}{4G^2} + \frac{E_v^2}{4EG} - \frac{G_u^2}{4G^2} \\ \text{i.e.,} \quad -2K\sqrt{EG} &= \frac{E_{vv} + G_{uu}}{(EG)^{1/2}} - \frac{E_v(EG_v + E_v G)}{2(EG)^{3/2}} - \frac{G_u(E_u G + EG_u)}{2(EG)^{3/2}} \\ &= \frac{E_{vv}}{(EG)^{1/2}} - \frac{1}{2}\frac{E_v(EG)_v}{(EG)^{3/2}} + \frac{G_{uu}}{(EG)^{1/2}} - \frac{1}{2}\frac{G_u(EG)_u}{(EG)^{3/2}} \\ &= \left(\frac{E_v}{(EG)^{1/2}}\right)_v + \left(\frac{G_u}{(EG)^{1/2}}\right)_u, \end{aligned}$$

proving the equation in (i). If, in addition, $E = 1$, the second term on the right-hand side of the formula in (i) vanishes, so

$$K = -\frac{1}{2\sqrt{G}} \frac{\partial}{\partial u}\left(\frac{G_u}{\sqrt{G}}\right) = -\frac{1}{\sqrt{G}} \frac{\partial^2 \sqrt{G}}{\partial u^2}. \qquad \square$$

## Example 10.2.4

For the surface of revolution

$$\boldsymbol{\sigma}(u, v) = (f(u)\cos v, f(u)\sin v, g(u)),$$

where $f > 0$ and $\dot{f}^2 + \dot{g}^2 = 1$ (a dot denoting $d/du$), we found in Example 6.1.3 that $E = 1$, $F = 0$, and $G = f(u)^2$. Hence, Corollary 10.2.3(ii) applies and gives

$$K = -\frac{1}{\sqrt{G}} \frac{\partial^2 \sqrt{G}}{\partial u^2} = -\frac{\ddot{f}}{f},$$

in agreement with Example 8.1.4.

The Theorema Egregium provides a necessary condition for the existence of a local isometry between surfaces: if such a local isometry exists, the Gaussian curvature must be the same at corresponding points of the two surfaces. We give two examples of this idea; others can be found in the exercises.

Our first result shows that it is impossible to draw a 'perfect' map of the Earth (which is why cartography is an interesting subject).

## Proposition 10.2.5

Any map of any region of the earth's surface must distort distances.

## Proof

A map of a region of the earth's surface which did *not* distort distances would be a diffeomorphism from this region of a sphere to a region in a plane (the map) which multiplied all distances by the same constant factor, say $C$. We might as well assume that the plane passes through the origin. Then, by composing this map with the map $\mathbf{r} \mapsto C^{-1}\mathbf{r}$ from the plane to itself, we would get an *isometry* between this region of the sphere and some region of a plane. This would imply, by the Theorema Egregium, that these regions of the sphere and the plane have the same Gaussian curvature. But we know that a plane has Gaussian curvature zero everywhere, and a sphere has constant positive Gaussian curvature everywhere (if the sphere has radius $R$, the Gaussian curvature is $1/R^2$). So no such isometry can exist. □

Note, on the other hand, that it *is* possible to draw a map of the Earth that correctly represents angles, and a map that correctly represents areas, for we saw in Example 6.3.5 that the stereographic projection is conformal, and Archimedes' Theorem 6.4.6 shows that there is a map that correctly represents areas. However, it is not possible to represent both angles and areas correctly with the same map (Exercise 6.4.5(ii)).

Our next example shows how the Theorema Egregium can sometimes be used to determine all the isometries of a surface.

## Proposition 10.2.6

The only isometries of a helicoid (Exercise 4.2.6)

$$\boldsymbol{\sigma}(u,v) = (u\cos v, u\sin v, v)$$

are $S_\lambda$, $R_x \circ S_\lambda$, $R_y \circ S_\lambda$ and $R_z \circ S_\lambda$ for some value of $\lambda$, where $S_\lambda$ is the screwing motion $\boldsymbol{\sigma}(u,v) \mapsto \boldsymbol{\sigma}(u, v+\lambda)$, and $R_x$, $R_y$ and $R_z$ are rotations by $\pi$ around the $x$-, $y$- and $z$-axes.

## Proof

Suppose that an isometry of the helicoid takes $\boldsymbol{\sigma}(u,v)$ to $\boldsymbol{\sigma}(\tilde{u},\tilde{v})$, where $\tilde{u}$ and $\tilde{v}$ are smooth functions of $u$ and $v$. Since the Gaussian curvature at $\boldsymbol{\sigma}(u,v)$ is $-1/(1+u^2)^2$ (see Exercise 8.1.2), the Theorema Egregium tells us that

$$\frac{-1}{(1+u^2)^2} = \frac{-1}{(1+\tilde{u}^2)^2},$$

so $\tilde{u} = \pm u$. Applying a rotation $R_z$ by $\pi$ around the $z$-axis changes $u$ to $-u$ (and fixes $v$), so we assume that $\tilde{u} = u$. Let $\tilde{v} = f(u,v)$. By Corollary 6.2.3, the patches $\boldsymbol{\sigma}(u,v)$ and $\tilde{\boldsymbol{\sigma}}(u,v) = \boldsymbol{\sigma}(u,f(u,v))$ have the same first fundamental form. That of $\boldsymbol{\sigma}$ is $du^2 + (1+u^2)dv^2$, and that of $\tilde{\boldsymbol{\sigma}}$ is found to be

$$(1+(1+u^2)f_u^2)du^2 + 2(1+u^2)f_u f_v du dv + (1+u^2)f_v^2 dv^2.$$

Equating these, we find that $f_u = 0$ and $f_v = \pm 1$. Hence,

$$\tilde{v} = f(u,v) = \pm v + \lambda,$$

where $\lambda$ is a constant. A rotation $R_x$ by $\pi$ around the $x$-axis changes $v$ to $-v$ (and fixes $u$), so we take the $+$ sign. This gives the isometry

$$S_\lambda : \boldsymbol{\sigma}(u,v) \mapsto \boldsymbol{\sigma}(u, v+\lambda).$$

This proves the proposition (the isometry $R_y \circ S_\lambda$ arises because $R_y = R_x \circ R_z$). □

## *EXERCISES*

10.2.1 Show that if a surface patch has first fundamental form $e^\lambda(du^2+dv^2)$, where $\lambda$ is a smooth function of $u$ and $v$, its Gaussian curvature $K$ satisfies

$$\Delta\lambda + 2Ke^\lambda = 0,$$

where $\Delta$ denotes the Laplacian $\partial^2/\partial u^2 + \partial^2/\partial v^2$.

10.2.2 With the notation of Exercise 9.5.1, define $u = r\cos\theta$, $v = r\sin\theta$, and let $\tilde{\boldsymbol{\sigma}}(u,v)$ be the corresponding reparametrization of $\boldsymbol{\sigma}$. It can be shown that $\tilde{\boldsymbol{\sigma}}$ is an allowable surface patch for $\mathcal{S}$ defined on the open set $u^2+v^2 < \epsilon^2$. (Note that this is not quite obvious because $\boldsymbol{\sigma}$ is not allowable when $r = 0$.)

(i) Show that the first fundamental form of $\tilde{\boldsymbol{\sigma}}$ is $\tilde{E}du^2 + 2\tilde{F}dudv + \tilde{G}dv^2$, where

$$\tilde{E} = \frac{u^2}{r^2} + \frac{Gv^2}{r^4}, \quad \tilde{F} = \left(1 - \frac{G}{r^2}\right)\frac{uv}{r^2}, \quad \tilde{G} = \frac{v^2}{r^2} + \frac{Gu^2}{r^4}.$$

(ii) Show that $u^2(\tilde{E}-1) = v^2(\tilde{G}-1)$, and by considering the Taylor expansions of $\tilde{E}$ and $\tilde{G}$ about $u = v = 0$, deduce that

$$G(r,\theta) = r^2 + kr^4 + \text{remainder}$$

for some constant $k$, where remainder$/r^4$ tends to zero as $r$ tends to zero.

(iii) Show that $k = -K(P)/3$, where $K(P)$ is the value of the Gaussian curvature of $\mathcal{S}$ at $P$.

10.2.3 With the notation of Exercises 9.5.1 and 10.2.2, show that:

(i) The circumference of the geodesic circle with centre $P$ and radius $R$ is

$$C_R = 2\pi R\left(1 - \frac{K(P)}{6}R^2 + \text{remainder}\right),$$

where remainder$/R^2$ tends to zero as $R$ tends to zero.

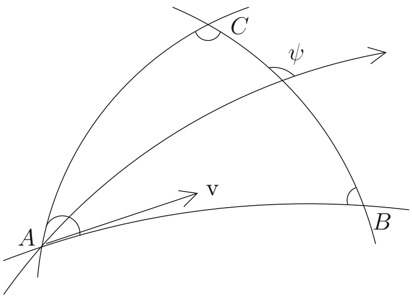
(ii) The area inside the geodesic circle in (i) is

$$A_R = \pi R^2\left(1 - \frac{K(P)}{12}R^2 + \text{remainder}\right),$$

where the remainder satisfies the same condition as in (i).

Verify that these formulas are consistent with those in spherical geometry obtained in Exercise 6.5.3.

10.2.4 Let $A, B$ and $C$ be the vertices of a triangle $\mathcal{T}$ on a surface $\mathcal{S}$ whose sides are arcs of geodesics, and let $\alpha, \beta$ and $\gamma$ be its internal angles (so that $\alpha$ is the angle at $A$, etc.). Assume that the triangle is contained in a geodesic patch $\boldsymbol{\sigma}$ as in Exercise 9.5.1 with $P = A$. Thus, with the notation in that exercise, if we take $\mathbf{v}$ to be tangent at $A$ to the side passing through $A$ and $B$, then the sides meeting at $A$ are the parameter curves $\theta = 0$ and $\theta = \alpha$, and the remaining side can be parametrized by $\boldsymbol{\gamma}(\theta) = \boldsymbol{\sigma}(f(\theta), \theta)$ for some smooth function $f$ and $0 \leq \theta \leq \alpha$.



(i) Use the geodesic equations (9.2) to show that

$$f'' - \frac{f'\lambda'}{\lambda^2} = \frac{1}{2}\frac{\partial G}{\partial r},$$

where a dash denotes $d/d\theta$ and $\lambda = \parallel \boldsymbol{\gamma}' \parallel$.

(ii) Show that, if $\psi(\theta)$ is the angle between $\boldsymbol{\sigma}_r$ and the tangent vector to the side opposite $A$ at $\boldsymbol{\gamma}(\theta)$, then

$$\psi'(\theta) = -\frac{\partial\sqrt{G}}{\partial r}(f(\theta), \theta).$$

(iii) Show that, if $K$ is the Gaussian curvature of $\mathcal{S}$,

$$\int_{\mathcal{T}} K\, d\mathcal{A}_{\boldsymbol{\sigma}} = \alpha + \beta + \gamma - \pi.$$

This result will be generalized in Corollary 13.2.3.

10.2.5 Show that the Gaussian curvature of the Möbius band in Example 4.5.3 is equal to $-1/4$ everywhere along its median circle. Deduce that this Möbius band *cannot* be constructed by taking a strip of paper and joining the ends together with a half-twist. (The analytic description of the 'cut and paste' Möbius band is more complicated than the version in Example 4.5.3.)

10.2.6 Show that the only isometries from the catenoid to itself are products of rotations around its axis, reflections in planes containing the axis, and reflection in the plane containing the waist of the catenoid.

## 10.3 Surfaces of constant Gaussian curvature

The results of the preceding two sections enable us to gain a good understanding of surfaces of constant Gaussian curvature. We begin with the following local description.

### Theorem 10.3.1

Any point of a surface of constant Gaussian curvature is contained in a patch that is isometric to an open subset of a plane, a sphere or a pseudosphere.

## Proof

Let $\mathbf{p}$ be a point on a surface $\mathcal{S}$ with constant Gaussian curvature $K$. By applying a dilation of $\mathbb{R}^3$ (see Exercise 8.1.5), we need only consider the cases $K = 0, 1$ and $-1$.

We take a geodesic patch $\boldsymbol{\sigma}(u, v)$ with $\boldsymbol{\sigma}(0, 0) = \mathbf{p}$. Writing $g = \sqrt{G}$, the first fundamental form is

$$du^2 + g(u, v)^2 dv^2.$$

By Corollary 10.2.3(ii),

$$\frac{\partial^2 g}{\partial u^2} + Kg = 0. \tag{10.5}$$

Note that

$$g(0, v) = 1, \quad g_u(0, v) = 0, \tag{10.6}$$

by Proposition 9.5.1.

If $K = 0$, the solution of Eq. 10.5 is $g(u, v) = \alpha u + \beta$, where $\alpha$ and $\beta$ are smooth functions of $v$ only. The initial conditions (10.6) give $\alpha = 0$, $\beta = 1$, so $g = 1$ and the first fundamental form of $\boldsymbol{\sigma}$ is $du^2 + dv^2$. This is the same as the first fundamental form of the usual parametrization of the plane (see Example 6.1.2), and Corollary 6.2.3 now shows that $\boldsymbol{\sigma}$ is isometric to an open subset of the plane.

If $K = 1$, the general solution of Eq. 10.5 is $g = \alpha \cos u + \beta \sin u$, where $\alpha$ and $\beta$ only depend on $v$. The boundary conditions (10.6) give $\alpha = 1$, $\beta = 0$, and the first fundamental form of $\boldsymbol{\sigma}$ is $du^2 + \cos^2 u \, dv^2$. This is the first fundamental form of $S^2$, with $u$ and $v$ being latitude and longitude, respectively (see Example 6.1.3). Hence, $\boldsymbol{\sigma}$ is isometric to an open subset of $S^2$.

Finally, if $K = -1$, we find in the same way that the first fundamental form of $\boldsymbol{\sigma}$ is

$$du^2 + \cosh^2 u \, dv^2.$$

We have not encountered this first fundamental form before. However, let us reparametrize $\boldsymbol{\sigma}$ by defining

$$V = e^v \tanh u, \quad W = e^v \operatorname{sech} u.$$

We then find, using the formulas in Exercise 6.1.4, that the first fundamental form becomes

$$\frac{dV^2 + dW^2}{W^2}.$$

Comparing with Exercise 8.3.1, we see that this is the first fundamental form of a parametrization of the pseudosphere. $\square$

The next result gives a much more precise, though still local, description of surfaces of constant negative Gaussian curvature.

## Proposition 10.3.2

Let $\mathcal{S}$ be a surface of constant Gaussian curvature $-1$. If $\mathbf{p} \in \mathcal{S}$, there is a surface patch of $\mathcal{S}$ containing $\mathbf{p}$ whose first and second fundamental forms are

$$du^2 + 2\cos\theta\, dudv + dv^2 \quad \text{and} \quad 2\sin\theta\, dudv, \tag{10.7}$$

respectively, where $\theta(u, v)$ is a smooth function such that $0 < \theta < \pi$ for all $u, v$.

Thus, the parameter curves of this parametrization form a *Chebyshev net* (see Exercises 6.1.9 and 7.4.6).

## Proof

Since $K$ is negative the surface $\mathcal{S}$ has no umbilics. Hence, there is a surface patch $\boldsymbol{\sigma}(U, V)$ containing $\mathbf{p}$ whose first and second fundamental forms are

$$EdU^2 + GdV^2 \quad \text{and} \quad LdU^2 + NdV^2,$$

respectively (Proposition 8.4.1). The principal curvatures are

$$\kappa_1 = \frac{L}{E}, \quad \kappa_2 = \frac{N}{G}. \tag{10.8}$$

Since $\kappa_1\kappa_2 = -1$, there are two cases:

(i) $\kappa_1 > 0$ and $\kappa_2 < 0$ for all $U, V$

(ii) $\kappa_1 < 0$ and $\kappa_2 > 0$ for all $U, V$

We shall concentrate on case (i), indicating briefly the modifications necessary for case (ii).

In case (i), there is a smooth function $\omega(U, V)$ such that $0 < \omega < \pi/2$ and

$$\kappa_1 = \tan\omega, \quad \kappa_2 = -\cot\omega. \tag{10.9}$$

(In case (ii), we put $\kappa_1 = -\cot\omega$, $\kappa_2 = \tan\omega$.) The first Codazzi-Mainardi equation is

$$L_V = \frac{1}{2}E_V\left(\frac{L}{E} + \frac{N}{G}\right)$$

(cf. Exercise 10.1.4). From Eqs. 10.8 and 10.9, we get

$$\begin{aligned}(E\tan\omega)_V &= \frac{1}{2}E_V(\tan\omega - \cot\omega),\\ \text{i.e.,}\quad \omega_V &= -\frac{E_V}{2E}\cot\omega. \end{aligned}\tag{10.10}$$

Then,

$$\left(\frac{\cos\omega}{\sqrt{E}}\right)_V = -\frac{\omega_V \sin\omega}{\sqrt{E}} - \frac{E_V\cos\omega}{2E^{3/2}} = -\frac{\sin\omega}{\sqrt{E}}\left(\omega_V + \frac{E_V}{2E}\cot\omega\right) = 0$$

by Eq. 10.10. Hence, there is a positive smooth function $e(U)$ such that

$$E = e(U)\cos^2\omega.$$

Similarly, the second Codazzi-Mainardi equation implies that there is a positive smooth function $g(V)$ such that

$$G = g(V)\sin^2\omega.$$

(In case (ii), we get $E = e(U)\sin^2\omega$, $G = g(V)\cos^2\omega$.)

Let $\tilde{U}(U)$ and $\tilde{V}(V)$ be smooth functions such that

$$\frac{d\tilde{U}}{dU} = \sqrt{e(U)}, \quad \frac{d\tilde{V}}{dV} = \sqrt{g(V)},$$

and let $\tilde{\boldsymbol{\sigma}}(\tilde{U},\tilde{V})$ be the corresponding reparametrization of $\boldsymbol{\sigma}(U,V)$. Then, the first fundamental form of $\tilde{\boldsymbol{\sigma}}$ is

$$\cos^2\omega\, d\tilde{U}^2 + \sin^2\omega\, d\tilde{V}^2$$

and its second fundamental form is

$$\frac{L}{e}d\tilde{U}^2 + \frac{N}{g}d\tilde{V}^2 = \sin\omega\cos\omega(d\tilde{U}^2 - d\tilde{V}^2).$$

Setting $u = (\tilde{U}+\tilde{V})/2$, $v = (\tilde{U}-\tilde{V})/2$ gives a reparametrization of $\tilde{\boldsymbol{\sigma}}$ whose first and second fundamental forms are as stated in the proposition, with $\theta = 2\omega$. (In case (ii) we set $u = (\tilde{U}+\tilde{V})/2$, $v = (\tilde{V}-\tilde{U})/2$.) $\square$

The function $\theta(u,v)$ appearing in (10.7) is not arbitrary, as the Gauss equations (Proposition 10.1.2) must still be satisfied. As we remarked earlier, these four equations are equivalent to each other, so we need only consider one of them. The Christoffel symbols are found to be

$$\begin{aligned}\Gamma^1_{11} &= \theta_u\cot\theta, & \Gamma^1_{12} &= 0, & \Gamma^1_{22} &= -\theta_v\operatorname{cosec}\theta,\\ \Gamma^2_{11} &= -\theta_u\operatorname{cosec}\theta, & \Gamma^2_{12} &= 0, & \Gamma^2_{22} &= \theta_v\cot\theta.\end{aligned}$$

Substituting in the first of the Gauss equations gives

$$\begin{aligned}-1 &= (-\theta_u\operatorname{cosec}\theta)_v - \theta_u\theta_v\operatorname{cosec}\theta\cot\theta,\\ \text{i.e.,} \qquad \theta_{uv} &= \sin\theta. \end{aligned} \tag{10.11}$$

This is called the *sine-Gordon equation*; it arises in fluid mechanics and various other areas of physics.

In view of Theorem 10.1.3, we can summarize our discussion as follows.

## Theorem 10.3.3

Let $\theta(u, v)$ be a smooth solution of the sine-Gordon equation (10.11) such that $0 < \theta < \pi$. Then there exists a surface, unique up to a direct isometry of $\mathbb{R}^3$, with constant Gaussian curvature $-1$ and first and second fundamental forms given by (10.7). Conversely, any surface of constant Gaussian curvature $-1$ has a parametrization with first and second fundamental forms (10.7) in which $\theta$ is a solution of the sine-Gordon equation.

We conclude this section with a beautiful global result that characterizes the *compact* surfaces of constant Gaussian curvature. Note that, by Proposition 8.6.1, the value of the constant Gaussian curvature in this case must be $> 0$.

## Theorem 10.3.4

Every connected compact surface whose Gaussian curvature is constant is a sphere.

The proof of this theorem depends on the following lemma.

## Lemma 10.3.5

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a surface patch containing a point $\mathbf{p}$ that is not an umbilic. Let $\kappa_1 \geq \kappa_2$ be the principal curvatures of $\boldsymbol{\sigma}$ and suppose that $\kappa_1$ has a local maximum at $\mathbf{p}$ and $\kappa_2$ has a local minimum there. Then, the Gaussian curvature of $\boldsymbol{\sigma}$ at $\mathbf{p}$ is $\leq 0$.

*Proof 10.3.5* Since $\mathbf{p}$ is not an umbilic, $\kappa_1 > \kappa_2$ at $\mathbf{p}$, so by shrinking $U$ if necessary, we may assume that $\kappa_1 > \kappa_2$ everywhere.

By Proposition 8.4.1, we can assume that the first and second fundamental forms of $\boldsymbol{\sigma}$ are

$$Edu^2 + Gdv^2 \quad \text{and} \quad Ldu^2 + Ndv^2,$$

respectively. By Exercise 10.1.4,

$$E_v = -\frac{2E}{\kappa_1 - \kappa_2}(\kappa_1)_v, \quad G_u = \frac{2G}{\kappa_1 - \kappa_2}(\kappa_2)_u,$$

and by Corollary 10.2.3(i), the Gaussian curvature

$$K = -\frac{1}{2\sqrt{EG}}\left(\frac{\partial}{\partial u}\left(\frac{G_u}{\sqrt{EG}}\right) + \frac{\partial}{\partial v}\left(\frac{E_v}{\sqrt{EG}}\right)\right).$$

Since $\mathbf{p}$ is a stationary point of $\kappa_1$ and $\kappa_2$, we have $(\kappa_1)_v = (\kappa_2)_u = 0$, and hence $E_v = G_u = 0$, at $\mathbf{p}$. Hence, at $\mathbf{p}$,

$$K = -\frac{1}{2EG}(G_{uu} + E_{vv}) = -\frac{1}{2EG}\left(\frac{2G}{\kappa_1 - \kappa_2}(\kappa_2)_{uu} - \frac{2E}{\kappa_1 - \kappa_2}(\kappa_1)_{vv}\right)$$

(again dropping terms involving $E_v$, $G_u$ and the first derivatives of $\kappa_1$ and $\kappa_2$). Since $\kappa_1$ has a local maximum at $\mathbf{p}$, $(\kappa_1)_{vv} \leq 0$ there, and since $\kappa_2$ has a local minimum at $\mathbf{p}$, $(\kappa_2)_{uu} \geq 0$ there. Hence, the last equation shows that $K \leq 0$ at $\mathbf{p}$. □

*Proof 10.3.4* The proof of Theorem 10.3.4 uses a little point set topology. We consider the continuous function on the surface $\mathcal{S}$ given by $J = (\kappa_1 - \kappa_2)^2$, where $\kappa_1$ and $\kappa_2$ are the principal curvatures. Note that this function is well defined even though $\kappa_1$ and $\kappa_2$ are not, partly because we do not know which principal curvature is to be called $\kappa_1$ and which $\kappa_2$, and partly because the principal curvatures are only well defined up to sign unless $\mathcal{S}$ has a definite orientation. We shall prove that this function is identically zero on $\mathcal{S}$, so that every point of $\mathcal{S}$ is an umbilic. Since the Gaussian curvature $K > 0$, it follows from Proposition 8.2.9 that $\mathcal{S}$ is an open subset of a sphere, say $\mathbf{S}$. In fact, $\mathcal{S}$ must be the whole of $\mathbf{S}$. For, since $\mathcal{S}$ is compact, it is necessarily a closed subset of $\mathbb{R}^3$, and hence a *closed* subset of $\mathbf{S}$. But since $\mathbf{S}$ is connected, the only non-empty subset of $\mathbf{S}$ that is both open and closed is $\mathbf{S}$ itself.

Suppose then, to get a contradiction, that $J$ is not identically zero on $\mathcal{S}$. Since $\mathcal{S}$ is compact, $J$ must attain its maximum value at some point $\mathbf{p} \in \mathcal{S}$, and this maximum value is $> 0$. Choose a patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ of $\mathcal{S}$ containing $\mathbf{p}$, and let $\kappa_1$ and $\kappa_2$ be its principal curvatures. Since $\kappa_1\kappa_2 > 0$, by reparametrizing if necessary, we can assume that $\kappa_1$ and $\kappa_2$ are both $> 0$ (see Exercise 8.2.8). Suppose that $\kappa_1 > \kappa_2$ at $\mathbf{p}$; since $\kappa_1$ and $\kappa_2$ are continuous (in fact, smooth) functions on $U$ (Exercise 8.2.8 again), by shrinking $U$ if necessary, we can assume that $\kappa_1 > \kappa_2$ everywhere on $U$. Since $K$ is a constant $> 0$, the function $\left(x - \frac{K}{x}\right)^2$ increases with $x$ provided that $x > K/x > 0$. Since $\kappa_1 > K/\kappa_1 = \kappa_2 > 0$, this function is increasing at $x = \kappa_1$, so $\kappa_1$ must have a local maximum at $\mathbf{p}$, and then $\kappa_2 = K/\kappa_1$ must have a local minimum there. By Lemma 10.3.5, $K \leq 0$ at $\mathbf{p}$. This contradicts the assumption that $K > 0$. □

## *EXERCISES*

10.3.1 Show that a compact surface with Gaussian curvature $> 0$ everywhere and constant *mean* curvature is a sphere.

10.3.2 Show that the solution of the sine-Gordon equation corresponding to the pseudosphere constructed in Section 8.3 is

$$\theta(u, v) = 2\tan^{-1}(\sinh(u - v + c)),$$

where $c$ is a constant.

## 10.4 Geodesic mappings

We considered in Chapter 6 local diffeomorphisms between surfaces that preserve lengths, angles or areas. In the same way, it is natural to consider local diffeomorphisms that preserve geodesics. In fact, to obtain interesting results we have to consider local diffeomorphisms $F : \mathcal{S} \to \tilde{\mathcal{S}}$ from a surface $\mathcal{S}$ to a surface $\tilde{\mathcal{S}}$ such that, if $\boldsymbol{\gamma}$ is a geodesic on $\mathcal{S}$, then $F \circ \boldsymbol{\gamma}$ is a *reparametrization* of a geodesic on $\tilde{\mathcal{S}}$ (cf. Exercise 10.4.1). We recall from Exercise 9.1.2 that a curve that is a reparametrization of a geodesic is called a pre-geodesic. Thus, we are led to

### Definition 10.4.1

If $\mathcal{S}$ and $\tilde{\mathcal{S}}$ are surfaces, a local diffeomorphism $F : \mathcal{S} \to \tilde{\mathcal{S}}$ is said to be *geodesic* if $f$ takes every pre-geodesic on $\mathcal{S}$ to a pre-geodesic on $\tilde{\mathcal{S}}$.

The following result provides some examples of geodesic local diffeomorphisms.

### Proposition 10.4.2

The following are geodesic local diffeomorphisms:

(i) Every local isometry.

(ii) Every dilation of $\mathbb{R}^3$.

(iii) Every composite of geodesic local diffeomorphisms.

### Proof

(i) If $\boldsymbol{\gamma}$ is a pre-geodesic on a surface patch $\boldsymbol{\sigma}$ of a surface $\mathcal{S}$, then $\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}_1(\varphi(t))$ for some geodesic $\boldsymbol{\gamma}_1$ and some reparametrization map $\varphi$ (Exercise 9.1.2). If $F : \mathcal{S} \to \tilde{\mathcal{S}}$ is an isometry, then $\tilde{\boldsymbol{\gamma}}_1 = F \circ \boldsymbol{\gamma}_1$ is a geodesic on the surface patch $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$ of $\tilde{\mathcal{S}}$ by Corollary 9.2.7. Hence, $F \circ \boldsymbol{\gamma} = F \circ \boldsymbol{\gamma}_1 \circ \varphi = \tilde{\boldsymbol{\gamma}}_1 \circ \varphi$ is a reparametrization of the geodesic $\tilde{\boldsymbol{\gamma}}_1$, and so is a pre-geodesic.

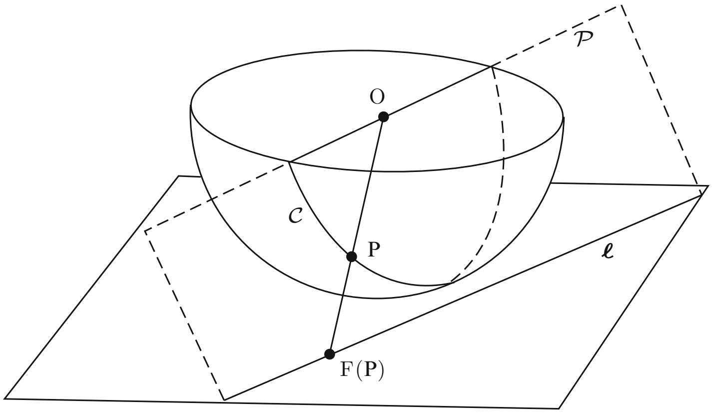
(ii) The dilation $(x, y, z) \mapsto a(x, y, z)$, where $a$ is a non-zero constant, multiplies the first fundamental form by $a^2$ and hence leaves the Christoffel symbols unchanged (see Proposition 7.4.4). By Proposition 9.2.3, applying the dilation leaves the equations determining the geodesics unchanged. It follows that the dilation takes geodesics to geodesics. By the argument in part (i), it also takes pre-geodesics to pre-geodesics.

(iii) This is obvious. □

As the preceding proof made clear, local isometries and dilations take geodesics to geodesics. The following is an example of a geodesic local diffeomorphism that *does not take geodesics to geodesics*. By the preceding proposition, it cannot be a composite of local isometries and dilations.

## Example 10.4.3

We consider the map $F$ from the lower hemisphere of the unit sphere $S^2$ to the plane $z = -1$ given by projection from the centre of the sphere.




Every pre-geodesic $\mathcal{C}$ on $S^2$ is a great circle formed by the intersection of $S^2$ with a plane $\mathcal{P}$ passing through the origin. Central projection takes the part of $\mathcal{C}$ in the lower hemisphere to the straight line $l$ that is the intersection of $\mathcal{P}$ with the plane $z = -1$. Hence, $F$ is a geodesic diffeomorphism.

As a particular case, let $\mathcal{P}$ be the $xz$-plane, so that $\mathcal{C}$ can be parametrized by $\boldsymbol{\gamma}(t) = (\cos t, 0, \sin t)$; note that $\boldsymbol{\gamma}$ is unit-speed and so is a geodesic. It is easy to show that $F(\boldsymbol{\gamma}(t)) = (-\cot t, 0, -1)$: this curve is a parametrization of the straight line $y = 0$, $z = -1$ (hence it is a pre-geodesic), but it is *not* a geodesic as it does not have constant speed (Proposition 9.1.2).

The inverse of the diffeomorphism $F$ in Example 10.4.3 provides a parametrization of (an open subset of) $S^2$ with the property that geodesics on $S^2$ correspond to straight lines in the parameter plane. Beltrami proved the following beautiful theorem which determines exactly which surfaces admit parametrizations with this property. It motivated him to carry out a detailed investigation of surfaces of constant Gaussian curvature, which in turn led to his work on hyperbolic geometry that we describe in Chapter 11.

## Theorem 10.4.4

Let $\mathcal{S}$ be a connected surface. If there is a geodesic local diffeomorphism from $\mathcal{S}$ to a plane, then $\mathcal{S}$ has constant Gaussian curvature. Conversely, if $\mathcal{S}$ has constant Gaussian curvature, then for any point $\mathbf{p} \in \mathcal{S}$ there is a surface patch $\boldsymbol{\sigma} : U \to \mathcal{S}$ of $\mathcal{S}$ such that $\mathbf{p} \in \boldsymbol{\sigma}(U)$ and $\boldsymbol{\sigma}^{-1} : \boldsymbol{\sigma}(U) \to U$ is a geodesic diffeomorphism.

## Proof

If $F$ is a geodesic local diffeomorphism from $\mathcal{S}$ to the plane, $\boldsymbol{\sigma} = F^{-1}$ is a parametrization of an open subset of $\mathcal{S}$ with the property that $\boldsymbol{\sigma}$ takes every straight line in the plane to a pre-geodesic on $\mathcal{S}$. Denoting the parameters by $u, v$, we can take the line to have equation

$$v = \lambda u + \mu, \tag{10.12}$$

where $\lambda$ and $\mu$ are constants. By Exercise 9.1.2, there is a parametrization $t \mapsto (u(t), v(t))$ of this line such that $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ is a unit-speed geodesic on $\mathcal{S}$. By Proposition 9.2.3, we have

$$\begin{aligned} \ddot{u} + (\Gamma_{11}^1 + 2\lambda\Gamma_{12}^1 + \lambda^2\Gamma_{22}^1)\dot{u}^2 &= 0, \\ \lambda\ddot{u} + (\Gamma_{11}^2 + 2\lambda\Gamma_{12}^2 + \lambda^2\Gamma_{22}^2)\dot{u}^2 &= 0. \end{aligned}$$

Eliminating $\ddot{u}$ and noting that $\dot{u}$ cannot be zero (for then we would have $\dot{v} = 0$ too, since $\dot{v} = \lambda\dot{u}$ by Eq. 10.12, contradicting the fact that $\boldsymbol{\gamma}$ is unit-speed), we get

$$\lambda(\Gamma_{11}^1 + 2\lambda\Gamma_{12}^1 + \lambda^2\Gamma_{22}^1) = \Gamma_{11}^2 + 2\lambda\Gamma_{12}^2 + \lambda^2\Gamma_{22}^2.$$

Since *every* straight line in the $uv$-plane corresponds to a geodesic $\boldsymbol{\gamma}$ in this way, the preceding equation must hold for all values of the constant $\lambda$. This implies that

$$\Gamma_{11}^2 = \Gamma_{22}^1 = 0, \quad \Gamma_{11}^1 = 2\Gamma_{12}^2, \quad \Gamma_{22}^2 = 2\Gamma_{12}^1.$$

The Gauss equations (Proposition 10.1.2) now become

$$\begin{aligned}
KE &= (\Gamma^2_{12})^2 - (\Gamma^2_{12})_u, \\
KF &= (\Gamma^1_{12})_u - 2(\Gamma^2_{12})_v + \Gamma^2_{12}\Gamma^1_{12}, \\
KF &= (\Gamma^2_{12})_v - 2(\Gamma^1_{12})_u + \Gamma^1_{12}\Gamma^2_{12}, \\
KG &= (\Gamma^1_{12})^2 - (\Gamma^1_{12})_v.
\end{aligned}$$

Subtracting the middle two equations gives $(\Gamma^1_{12})_u = (\Gamma^2_{12})_v$, and using this the four equations become

$$\begin{aligned}
KE &= (\Gamma^2_{12})^2 - (\Gamma^2_{12})_u, && (10.13)\\
KF &= -(\Gamma^2_{12})_v + \Gamma^1_{12}\Gamma^2_{12}, && (10.14)\\
KF &= -(\Gamma^1_{12})_u + \Gamma^1_{12}\Gamma^2_{12}, && (10.15)\\
KG &= (\Gamma^1_{12})^2 - (\Gamma^1_{12})_v. && (10.16)
\end{aligned}$$

Differentiating Eq. 10.13 with respect to $v$ and Eq. 10.14 with respect to $u$, and then equating the expressions for $(\Gamma^2_{12})_{uv} = (\Gamma^2_{12})_{vu}$ gives

$$\begin{aligned}
EK_v - FK_u &= K(F_u - E_v) + 2\Gamma^2_{12}(\Gamma^2_{12})_v - \Gamma^2_{12}(\Gamma^1_{12})_u - \Gamma^1_{12}(\Gamma^2_{12})_u \\
&= K(F_u - E_v) + 2\Gamma^2_{12}(\Gamma^1_{12}\Gamma^2_{12} - KF) - \Gamma^2_{12}(\Gamma^1_{12}\Gamma^2_{12} - KF) \\
&\quad - \Gamma^1_{12}((\Gamma^2_{12})^2 - KE) \\
&= K(F_u - E_v) + K(E\Gamma^1_{12} - F\Gamma^2_{12}), && (10.17)
\end{aligned}$$

where we used Eqs. 10.13–10.15 in passing from the first line to the second.

Using the definition of the Christoffel symbols (Proposition 7.4.4), we find that

$$E\Gamma^1_{12} - F\Gamma^2_{12} = \frac{(EG+F^2)E_v - 2EFG_u}{2(EG-F^2)}. \tag{10.18}$$

Now, the equations $\Gamma^2_{11} = 0$ and $\Gamma^1_{11} = 2\Gamma^2_{12}$ give

$$EE_v = 2EF_u - FE_u, \quad 3FE_v = 2EG_u + 2FF_u - GE_u.$$

Hence,

$$\begin{aligned}
(EG+F^2)E_v - 2EFG_u &= 2(EG-F^2)E_v + (3F^2 - EG)E_v - 2EFG_u \\
&= 2(EG-F^2)E_v + F(2EG_u + 2FF_u - GE_u) \\
&\quad - G(2EF_u - FE_u) - 2EFG_u \\
&= 2(EG-F^2)(E_v - F_u).
\end{aligned}$$

Inserting this result into Eq. 10.18 gives

$$E\Gamma^1_{12} - F\Gamma^2_{12} = E_v - F_u,$$

and then Eq. 10.17 gives

$$EK_v - FK_u = 0.$$

A similar calculation starting with Eqs. 10.15 and 10.16 leads to

$$FK_v - GK_u = 0.$$

Hence,

$$(EG - F^2)K_v = GFK_u - FGK_u = 0,$$

so $K_v = 0$ (since $EG - F^2 > 0$). Similarly $K_u = 0$. Hence, $K$ is constant.

For the converse, assume that $\mathcal{S}$ has constant Gaussian curvature. By Theorem 10.3.1, each point of $\mathcal{S}$ is contained in a surface patch that is isometric to an open subset of a plane, a sphere or a pseudosphere. Since local isometries and dilations are geodesic local diffeomorphisms (Proposition 10.4.2), we have only to show that there are geodesic local diffeomorphisms from $\mathcal{S}$ to the plane in three cases:

(i) $\mathcal{S}$ is a plane, when there is nothing to prove.

(ii) $\mathcal{S}$ is the unit sphere, which is Example 10.4.3.

(iii) $\mathcal{S}$ is the pseudosphere in Section 8.3, which is Exercise 9.3.4.

The proof of Beltrami's theorem is complete. □

## *EXERCISES*

10.4.1 Show that a local diffeomorphism between surfaces that takes unit-speed geodesics to unit-speed geodesics must be a local isometry.

10.4.2 Show that a local diffeomorphism between surfaces that is the composite of a dilation and a local isometry takes geodesics to geodesics. Is the converse of this statement true?

10.4.3 This exercise shows that a geodesic local diffeomorphism $F$ from a surface $\mathcal{S}$ to a surface $\tilde{\mathcal{S}}$ that is also conformal is the composite of a dilation and a local isometry.

(i) Let $\mathbf{p} \in \mathcal{S}$ and let $\boldsymbol{\sigma}$ be a geodesic patch containing $\mathbf{p}$ as in Proposition 9.5.1, with first fundamental form $du^2 + Gdv^2$. Show that $\tilde{\boldsymbol{\sigma}} = F \circ \boldsymbol{\sigma}$ is a patch of $\tilde{\boldsymbol{\sigma}}$ containing $F(\mathbf{p})$ with first fundamental form $\lambda(du^2 + Gdv^2)$ for some smooth function $\lambda(u, v)$.

(ii) Show that the parameter curves $v =$ constant are pre-geodesics on $\tilde{\boldsymbol{\sigma}}$ and deduce that $\lambda$ is independent of $v$.

(iii) Show that if $\boldsymbol{\gamma}$ is a geodesic on $\boldsymbol{\sigma}$ and $\theta$ is the oriented angle between $\boldsymbol{\gamma}$ and the parameter curves $v =$ constant,

$$\frac{d\theta}{dv} + \frac{G_u}{2\sqrt{G}} = 0. \tag{10.19}$$

(iv) Show that

$$\frac{d\theta}{dv} + \frac{(\lambda G)_u}{2\lambda\sqrt{G}} = 0. \tag{10.20}$$

(v) Deduce from Eqs. 10.19 and 10.20 that $\lambda$ is constant.

(vi) Show that $F : \boldsymbol{\sigma} \to \tilde{\boldsymbol{\sigma}}$ is the composite of a dilation and a local isometry.

# 11
# Hyperbolic geometry

One of the most remarkable discoveries of nineteenth century mathematics is that the pseudosphere discussed in Section 8.3 has a geometry that closely resembles Euclidean geometry, with geodesics playing the role of straight lines. In fact, the closest correspondence with Euclidean geometry is obtained by 'embedding' the pseudosphere in a larger geometry, which is called *hyperbolic* or *non-Euclidean* geometry. When this is done, we find that all the axioms of Euclidean geometry hold in hyperbolic geometry, except the so-called 'parallel axiom': this states that if $\mathbf{p}$ is a point that is not on a straight line $l$, there is a unique straight line passing through $\mathbf{p}$ that does not intersect $l$ (i.e., which is 'parallel' to $l$ in the usual sense).

Hyperbolic geometry was discovered independently and almost simultaneously by the Hungarian mathematician Janos Bolyai and the Russian Nicolai Lobachevsky, although the formulations of it that we shall describe in this chapter are due to Eugenio Beltrami, Felix Klein and Henri Poincaré. David Hilbert, one of the greatest mathematicians of the twentieth century, wrote that the discovery of non-Euclidean geometry was 'one of the two most suggestive and notable achievements of the last century'. It ended centuries of attempts by Greek, Arab and later Western mathematicians to deduce the parallel axiom from the other axioms of Euclidean geometry, and it profoundly changed our view of what geometry is.

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_11,

## 11.1 Upper half-plane model

We saw in Example 9.3.3 that if the pseudosphere is parametrized as

$$\tilde{\boldsymbol{\sigma}}(v,w) = \left(\frac{1}{w}\cos v, \frac{1}{w}\sin v, \sqrt{1-\frac{1}{w^2}} - \cosh^{-1} w\right),$$

where we must have $w > 1$ for $\tilde{\boldsymbol{\sigma}}$ to be well defined and smooth, its geodesics correspond to arcs of circles and straight lines in the $vw$-plane that intersect the $v$-axis perpendicularly. The line $w = 1$ appears to be a rather artificial boundary in the $vw$-plane, since the geodesics are well defined in the entire region $w > 0$. On the other hand, the line $w = 0$ is a 'real' boundary since the first fundamental form

$$\frac{dv^2 + dw^2}{w^2} \tag{11.1}$$

of the pseudosphere is undefined when $w = 0$. It is therefore natural to ask if there is a surface that corresponds to the whole of the half-plane $w > 0$ with this first fundamental form. In fact, there is no such surface for a celebrated theorem of Hilbert shows that there is no surface with constant negative Gaussian curvature that is 'geodesically complete', i.e., a surface for which all geodesics can be extended indefinitely in both directions (see Exercise 10.1.3).

One possible response to Hilbert's theorem is essentially to ignore it: all those properties of surfaces that depend only on the first fundamental form (lengths, angles, areas, geodesics, local isometries, ...) can still be studied for the half-plane

$$\mathcal{H} = \{(v,w) \in \mathbb{R}^2 \mid w > 0\}$$

equipped with the first fundamental form (11.1). They will then be called hyperbolic lengths, hyperbolic angles, etc. (We shall see the reason for the adjective 'hyperbolic' later.)

It will often be convenient to identify $\mathbb{R}^2$ with the set of complex numbers $\mathbb{C}$ via $(v,w) \leftrightarrow v + iw$, so that

$$\mathcal{H} = \{z \in \mathbb{C} \mid \mathfrak{Im}(z) > 0\}$$

is the set of complex numbers with positive imaginary part.

The first thing to observe is that $\mathcal{H}$ is a 'conformal model':

### Proposition 11.1.1

Hyperbolic angles in $\mathcal{H}$ are the same as Euclidean angles.

## Proof

This is just because the first fundamental form (11.1) of $\mathcal{H}$ is a multiple of $dv^2 + dw^2$ (see Corollary 6.3.4). $\square$

The 'hyperbolic lines' are the geodesics in $\mathcal{H}$, which were determined in Example 9.3.3.

## Proposition 11.1.2

The geodesics in $\mathcal{H}$ are the half-lines parallel to the imaginary axis and the semicircles with centres on the real axis.

Here are some simple properties of hyperbolic lines.

## Proposition 11.1.3

(i) There is a unique hyperbolic line passing through any two distinct points of $\mathcal{H}$.

(ii) The parallel axiom does not hold in $\mathcal{H}$.

In the following proof, and later in this chapter, 'lines' and 'circles' will mean Euclidean lines and circles ('hyperbolic line' means 'geodesic'). On the other hand, 'lengths' and 'angles' will always mean hyperbolic lengths and angles, unless explicitly stated otherwise.

## Proof

(i) Let $a, b \in \mathcal{H}$, $a \neq b$. If the line passing through $a$ and $b$ is parallel to the imaginary axis, the unique hyperbolic line passing through the points $a$ and $b$ is the half-line containing them. If the line through $a$ and $b$ is not parallel to the imaginary axis, its perpendicular bisector intersects the real axis at some point $c$, say, and the unique hyperbolic line passing through $a$ and $b$ is the semicircle with centre $c$ and radius $|a - c| = |b - c|$.
(ii) Take $l$ to be the imaginary axis and let $a \in \mathcal{H}$ be any point not on $l$. For definiteness, assume that the real part $\mathfrak{Re}(a) > 0$. Then, the perpendicular bisector of the line joining $a$ to the origin intersects the real axis at some point $b > 0$. Let $c$ be a real number greater than $b$; then the semicircle with centre $c$ passing through $a$ is a hyperbolic line in $\mathcal{H}$ that does not intersect $l$. Of course,

the half-line through $a$ parallel to the imaginary axis is another hyperbolic line with the same property (it can be regarded as the limiting case when $c \to \infty$). $\square$

Since there is a unique hyperbolic line passing through any two points $a, b \in \mathcal{H}$, it makes sense to define the *hyperbolic distance* $d_{\mathcal{H}}(a, b)$ between $a$ and $b$ to be the length of the hyperbolic line segment joining them. It is shown in Exercise 11.2.1 that this is actually the hyperbolic length of the shortest curve joining $a$ and $b$.

## Proposition 11.1.4

The hyperbolic distance between two points $a, b \in \mathcal{H}$ is

$$d_{\mathcal{H}}(a, b) = 2 \tanh^{-1} \frac{|b-a|}{|b-\bar{a}|}.$$

In this formula, $\bar{a}$ denotes the complex conjugate of the complex number $a$. The appearance of the hyperbolic tangent gives an indication of the reason why the geometry of $\mathcal{H}$ is called 'hyperbolic geometry'.

## Proof

There are two cases, depending on whether the hyperbolic line joining $a$ and $b$ is a semicircle or a half-line. We shall deal with the semicircle case, leaving the simpler case of the half-line to Exercise 11.1.2.

Suppose then that $a$ and $b$ lie on the semicircle with centre $c$ on the real axis and radius $r$. The semicircle can be parametrized by

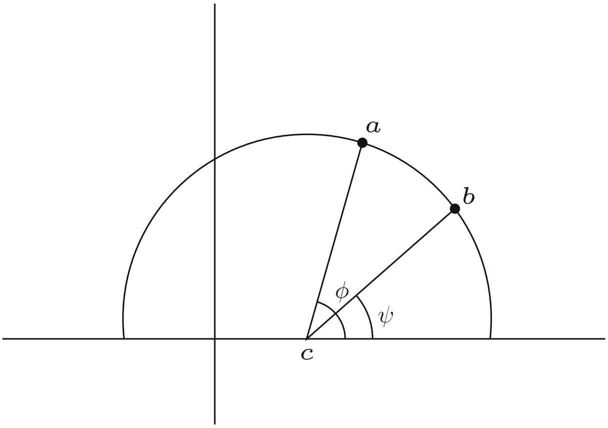
$$v = c + r\cos\theta, \quad w = r\sin\theta.$$

Writing $d$ for $d_{\mathcal{H}}(a, b)$ and denoting $d/d\theta$ by a dot, we have

$$d = \int_{\varphi}^{\psi} \sqrt{\frac{\dot{v}^2 + \dot{w}^2}{w^2}}\, d\theta = \int_{\varphi}^{\psi} \sqrt{\frac{r^2\sin^2\theta + r^2\cos^2\theta}{r^2\sin^2\theta}}\, d\theta = \int_{\varphi}^{\psi} \frac{d\theta}{\sin\theta},$$

where $\varphi = \arg(a-c)$, $\psi = \arg(b-c)$ (note that $d$ is independent of the radius of the semicircle). Hence,

$$d = \ln \frac{\tan\frac{\psi}{2}}{\tan\frac{\varphi}{2}}.$$




Now,

$$\tanh\frac{d}{2} = \frac{e^d - 1}{e^d + 1} = \frac{\tan\frac{\psi}{2} - \tan\frac{\varphi}{2}}{\tan\frac{\psi}{2} + \tan\frac{\varphi}{2}} = \frac{\sin\frac{\psi}{2}\cos\frac{\varphi}{2} - \cos\frac{\psi}{2}\sin\frac{\varphi}{2}}{\sin\frac{\psi}{2}\cos\frac{\varphi}{2} + \cos\frac{\psi}{2}\sin\varphi 2} = \frac{\sin\frac{\psi-\varphi}{2}}{\sin\frac{\psi+\varphi}{2}}. \tag{11.2}$$

On the other hand,

$$\begin{aligned} |b-a|^2 &= r^2((\cos\psi - \cos\varphi)^2 + (\sin\psi - \sin\varphi)^2) \\ &= 2r^2(1 - \cos(\psi - \varphi)) = 4r^2\sin^2\frac{\psi-\varphi}{2}, \end{aligned}$$

and similarly

$$|b-\bar{a}|^2 = 4r^2\sin^2\frac{\psi+\varphi}{2}.$$

Combining the last two equations with Eq. 11.2 gives

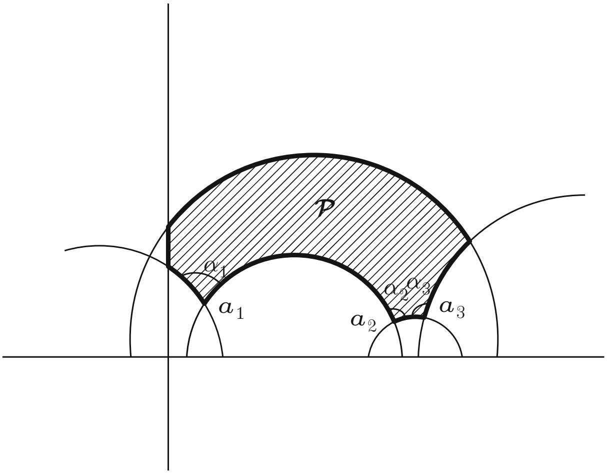
$$\tanh\frac{d}{2} = \frac{|b-a|}{|b-\bar{a}|}. \qquad \square$$

We conclude this section with another beautiful formula, this time for the area of a hyperbolic polygon, i.e., a polygon whose sides are hyperbolic lines.

## Theorem 11.1.5

Let $\mathcal{P}$ be a $n$-sided hyperbolic polygon in $\mathcal{H}$ with internal angles $\alpha_1, \alpha_2, \ldots, \alpha_n$. Then, the hyperbolic area of the polygon is given by

$$\mathcal{A}(\mathcal{P}) = (n-2)\pi - \alpha_1 - \alpha_2 - \cdots - \alpha_n.$$




In particular, for a triangle with angles $\alpha, \beta, \gamma$, the area is

$$\pi - \alpha - \beta - \gamma.$$

This should be compared with the well-known formula

$$\alpha + \beta + \gamma = \pi$$

for the sum of the angles of a Euclidean triangle with straight line sides, and the formula

$$\alpha + \beta + \gamma - \pi$$

for the area of a triangle on the unit sphere with geodesic (i.e., great circle) sides (Theorem 6.4.7).

*Proof 11.1.5* Let $a_1, \ldots, a_n$ be the vertices of $\mathcal{P}$ and $\mathcal{C}$ its boundary, consisting of $n$ hyperbolic line segments $a_1a_2, a_2a_3, \ldots, a_na_1$ (we assume that $\alpha_1$ is the internal angle of $\mathcal{P}$ at the vertex $a_1$, etc.). Since the first fundamental form is $(dv^2 + dw^2)/w^2$, the area of $\mathcal{P}$ is

$$\int_{\mathcal{P}} \frac{dvdw}{w^2}.$$

We evaluate this integral by using Green's theorem (Section 3.2):

$$\int_{\mathcal{C}} pdv + qdw = \int_{\mathcal{P}} \left( \frac{\partial q}{\partial v} - \frac{\partial p}{\partial w} \right) dvdw,$$

where $p$ and $q$ are smooth functions of $(v, w)$. Taking $p = 1/w$ and $q = 0$ gives

$$\int_{\mathcal{P}} \frac{dvdw}{w^2} = \int_{\mathcal{C}} \frac{dv}{w}. \tag{11.3}$$

To evaluate this integral we first prove the following lemma.

## Lemma 11.1.6

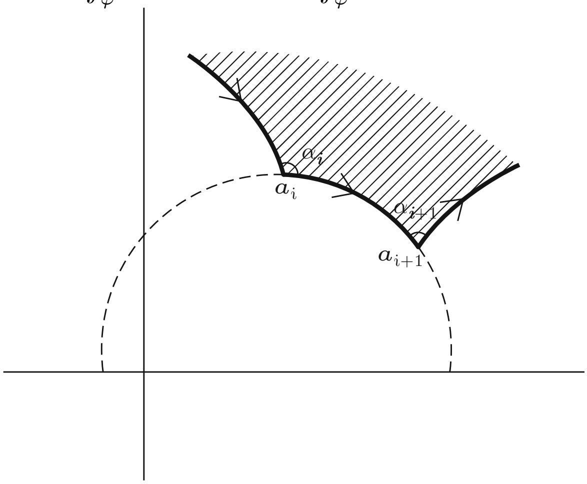
Let $a$ and $b$ be the endpoints of a segment $l$ of a hyperbolic line in $\mathcal{H}$ that forms part of a semicircle with centre $p$ on the real axis, and suppose that the radius vectors joining $p$ to $a$ and $p$ to $b$ make angles $\varphi$ and $\psi$, respectively, with the positive real axis (see the diagram in the proof of Proposition 11.1.4). Then,

$$\int_l \frac{dv}{w} = \varphi - \psi.$$

Note that the integral is independent of the radius of the semicircle, and that the formula is correct even if the hyperbolic line is part of a half-line, for in that case the integral vanishes since $v$ is constant along the hyperbolic line.

*Proof 11.1.6* We parametrize the hyperbolic line by $v = r\cos\theta$, $w = r\sin\theta$, where $r$ is the radius of the semicircle. Then, the integral is

$$\int_{\varphi}^{\psi} \frac{-r\sin\theta\, d\theta}{r\sin\theta} = -\int_{\varphi}^{\psi} d\theta = \varphi - \psi. \qquad \square$$



Returning to the proof of Theorem 11.1.5, let $\varphi_i$ and $\psi_i$ be the angles defined in the lemma corresponding to the side with endpoints $a_i$ and $a_{i+1}$, for $i = 1, \ldots, n$ (it is understood that $a_{n+1}$ means $a_1$). By Eq. 11.3 and the lemma,

$$\int_{\mathcal{P}} \frac{dvdw}{w^2} = \sum_{i=1}^{n}(\varphi_i - \psi_i). \tag{11.4}$$

We can simplify this sum by considering the change in direction of the outward-pointing normal of $\mathcal{P}$ as we traverse its boundary in an anticlockwise direction. As we traverse the side with endpoints $a_i$ and $a_{i+1}$, the outward normal rotates anticlockwise through an angle $\psi_i - \varphi_i$, while at the vertex $a_i$ it rotates by $\pi - \alpha_i$. Hence, as we traverse the boundary of $\mathcal{P}$, the outward normal rotates through an angle

$$n\pi + \sum_{i=1}^{n}(\psi_i - \varphi_i - \alpha_i).$$

But this angle of rotation is $2\pi$ (cf. Theorem 3.1.4), so we have the equation

$$2\pi = n\pi + \sum_{i=1}^{n}(\psi_i - \varphi_i - \alpha_i).$$

Rearranging, we get

$$\sum_{i=1}^{n}(\varphi_i - \psi_i) = (n-2)\pi - \sum_{i=1}^{n}\alpha_i.$$

By Eq. 11.4, this is the desired area. □

Note that the area of a hyperbolic triangle, i.e., a triangle whose sides are hyperbolic lines, depends only on its angles. We found in Proposition 6.5.8 that the same result holds in spherical geometry, but as we noted there this is completely different to the Euclidean situation, where we can change the size of a triangle (and hence its area) without changing its angles. In fact, we shall show in the next section that, as in spherical geometry (Exercise 6.5.2), two hyperbolic triangles with the same angles are *congruent*. But first we must discuss what congruence means in hyperbolic geometry.

## *EXERCISES*

11.1.1 Show that, if $l$ is a half-line geodesic in $\mathcal{H}$ and $a$ is a point not on $l$, there are infinitely many hyperbolic lines passing through $a$ that do not intersect $l$.

11.1.2 Complete the proof of Proposition 11.1.4 by dealing with the case in which the hyperbolic line passing through $a$ and $b$ is a half-line.

11.1.3 Show that for any $a \in \mathcal{H}$ there is a unique hyperbolic line passing through $a$ that intersects the hyperbolic line $l$ given by $v = 0$ perpendicularly. If $b$ is the point of intersection, one calls $d_{\mathcal{H}}(a,b)$ the *hyperbolic distance of* $a$ *from* $l$.

11.1.4 The *hyperbolic circle* $\mathcal{C}_{a,R}$ with centre $a \in \mathcal{H}$ and radius $R > 0$ is the set of points of $\mathcal{H}$ which are a hyperbolic distance $R$ from $a$:

$$\mathcal{C}_{a,R} = \{z \in \mathcal{H} \mid d_{\mathcal{H}}(z,a) = R\}.$$

Show that $\mathcal{C}_{a,R}$ is a Euclidean circle.

Show that the Euclidean centre of $\mathcal{C}_{ic,R}$, where $c > 0$, is $ib$ and that its Euclidean radius is $r$, where

$$c = \sqrt{b^2 - r^2}, \quad R = \frac{1}{2} \ln \frac{b+r}{b-r}.$$

Deduce that the hyperbolic length of the circumference of $\mathcal{C}_{ic,R}$ is $2\pi \sinh R$ and that the hyperbolic area inside it is $2\pi(\cosh R - 1)$. Note that these do not depend on $c$; in fact, it follows from the results of the next section that the circumference and area of $\mathcal{C}_{a,R}$ depend only on $R$ and not on $a$ (see the remarks preceding Theorem 11.2.4).

Compare these formulas with the case of a spherical circle in Exercise 6.5.3, and verify that they are consistent with Exercise 10.2.3.

## 11.2 Isometries of $\mathcal{H}$

In Euclidean plane geometry, two triangles are said to be *congruent* if one triangle can be moved until it coincides with the other. The types of motion that are allowed are combinations of rotations, translations, and reflections, i.e., the *isometries* of the plane (see Appendix 1). Similarly, to discuss congruence in spherical geometry, it was necessary in Section 6.5 to determine the isometries of the sphere.

It is easy to identify some isometries of $\mathcal{H}$:

(i) *Translations parallel to the real axis*, given by

$$T_a(z) = z + a, \qquad a \in \mathbb{R}.$$

(ii) *Reflections in lines parallel to the imaginary axis*, given by

$$R_a(z) = 2a - \bar{z}, \qquad a \in \mathbb{R}.$$

$R_a(z)$ is the 'reflection' of $z$ in the line $\mathfrak{Re}(z) = a$, thought of as a mirror; each point of this line is fixed by $R_a$.

(iii) *Dilations by a factor $a > 0$*, given by

$$D_a(z) = az.$$

In terms of the parameters $(v, w)$, these maps are given by $(v, w) \mapsto (v + a, w)$, $(v, w) \mapsto (2a - v, w)$ and $(v, w) \mapsto (av, aw)$, respectively, each of which obviously takes $\mathcal{H}$ to $\mathcal{H}$ and preserves the first fundamental form (11.1). But there is also a fourth type of isometry that is not quite as obvious:

(iv) *Inversions in circles with centres on the real axis.* The inversion in the circle with centre $a \in \mathbb{R}$ and radius $r > 0$ is

$$\mathcal{I}_{a,r}(z) = a + \frac{r^2}{\bar{z} - a}$$

(see Appendix 2).

To see that $\mathcal{I}_{a,r}$ is an isometry of $\mathcal{H}$, we consider first the case $a = 0$, $r = 1$, and denote $\mathcal{I}_{0,1}$ by $\mathcal{I}$. Then,

$$\mathcal{I}(v + iw) = \frac{v + iw}{v^2 + w^2},$$

which makes it clear that $\mathcal{I}$ takes any point in $\mathcal{H}$ to another point of $\mathcal{H}$ and any point on the real axis to another point on the real axis. To see that $\mathcal{I}$ is indeed an isometry of $\mathcal{H}$, we use the result of Exercise 6.1.4: if $\tilde{v} = \frac{v}{v^2+w^2}$ and $\tilde{w} = \frac{w}{v^2+w^2}$, then

$$d\tilde{v} = \frac{(w^2 - v^2)dv - 2vwdw}{(v^2 + w^2)^2}, \quad d\tilde{w} = \frac{-2vwdv + (v^2 - w^2)dw}{(v^2 + w^2)^2},$$

and hence

$$\begin{aligned}
\frac{d\tilde{v}^2 + d\tilde{w}^2}{\tilde{w}^2} &= \frac{1}{w^2(v^2 + w^2)^2}\Big\{\big((w^2 - v^2)dv - 2vwdw\big)^2 \\
&\qquad + \big(-2vwdv + (v^2 - w^2)dw\big)^2\Big\} \\
&= \frac{(w^2 - v^2)^2 + 4v^2w^2}{w^2(v^2 + w^2)^2}(dv^2 + dw^2) = \frac{dv^2 + dw^2}{w^2}.
\end{aligned}$$

Returning to the general case, we note that

$$\begin{aligned}\mathcal{I}_{a,r}(z) = T_a\left(\frac{r^2}{\bar{z}-a}\right) &= T_a D_{r^2}\left(\frac{1}{\bar{z}-a}\right)\\ &= T_a D_{r^2}\mathcal{I}(z-a) = T_a D_{r^2}\mathcal{I}T_{-a}(z),\end{aligned}$$

so $\mathcal{I}_{a,r}$ is a composite $T_a \circ D_{r^2} \circ \mathcal{I} \circ T_{-a}$ of maps that are already known to be isometries of $\mathcal{H}$. Since any composite of isometries is an isometry, it follows that $\mathcal{I}_{a,r}$ is an isometry of $\mathcal{H}$.

We summarize our observations as follows:

## Proposition 11.2.1

Any composite of a finite number of maps of the types (i)–(iv) defined above is an isometry of $\mathcal{H}$.

We shall call an isometry of one of the types (i)–(iv) an *elementary isometry* of $\mathcal{H}$. In fact, every isometry of $\mathcal{H}$ is a composite of a finite number of elementary isometries, but since we shall not make use of this result we leave its proof to the exercises.

Since isometries take geodesics to geodesics (Corollary 9.2.7), we know that the elementary isometries take half-lines and semicircles perpendicular to the real axis to other half-lines and semicircles perpendicular to the real axis. In fact, it is clear that translations, dilations and reflections take half-lines to half-lines and semicircles to semicircles, but the situation for inversions is a little more complicated:

## Proposition 11.2.2

The inversion $\mathcal{I}_{a,r}$ in the circle with centre $a \in \mathbb{R}$ and radius $r > 0$ takes hyperbolic lines that intersect the real axis perpendicularly at $a$ to half-lines, and all other hyperbolic lines to semicircles.

See Appendix 2 for the proof. The result is intuitively clear, since if a point of $\mathcal{H}$ "tends to" a point $a$ its image under $\mathcal{I}_{a,r}$ "tends to infinity" (both limits in the Euclidean sense) and so cannot lie on a semicircle geodesic.

Isometries can be used to simplify the solution of many problems in hyperbolic geometry, by reducing the problem to a 'standard' situation. The basic result needed for this is

## Proposition 11.2.3

Let $l_1$ and $l_2$ be hyperbolic lines in $\mathcal{H}$, and let $z_1$ and $z_2$ be points on $l_1$ and $l_2$, respectively. Then, there is an isometry of $\mathcal{H}$ that takes $l_1$ to $l_2$ and $z_1$ to $z_2$.

## Proof

We observe first that it is enough to prove this result in the special case in which $l_2$ is the half-line $l$ passing through the origin and $z_2 = i$. For if the proposition has been proved in this case, there is an isometry $F_1$ that takes $l_1$ to $l$ and $z_1$ to $i$, and an isometry $F_2$ that takes $l_2$ to $l$ and $z_2$ to $i$. Then, $F_2^{-1} \circ F_1$ is an isometry that takes $l_1$ to $l_2$ and $z_1$ to $z_2$.

There are now two cases depending on whether $l_1$ is a half-line or a semi-circle. If $l_1$ is the half-line $v = a$, say, the translation $T_{-a}$ takes $l_1$ to $l$ and $z_1$ to some point $ib$, say, on $l$, where $b > 0$. Then, the dilation $D_{b^{-1}}$ takes $l$ to itself and $ib$ to $i$, and so the isometry we want is $D_{b^{-1}} \circ T_{-a}$.

Finally, suppose that $l_1$ is a semicircle, and let $a$ be one of the two points in which it intersects the real axis. By Proposition 11.2.2, the inversion $\mathcal{I}_{a,1}$ takes $l$ to a half-line geodesic $l'$, say, and $z_1$ to some point $z'$ on $l'$. By the preceding case, there is an isometry $F$ that takes $l'$ to $l$ and $z'$ to $i$, so the isometry we want is $F \circ \mathcal{I}_{a,1}$. □

As a simple application, we can now complete Exercise 11.1.4. If $a, b \in \mathcal{H}$, there is an isometry $F$ of $\mathcal{H}$ that takes $a$ to $b$. Then, $F$ will clearly take the hyperbolic circle $\mathcal{C}_{a,R}$ to $\mathcal{C}_{b,R}$ for all $R > 0$. It follows that these hyperbolic circles have the same circumference and area.

Here is a more important application.

## Theorem 11.2.4

In hyperbolic geometry, similar triangles are congruent.

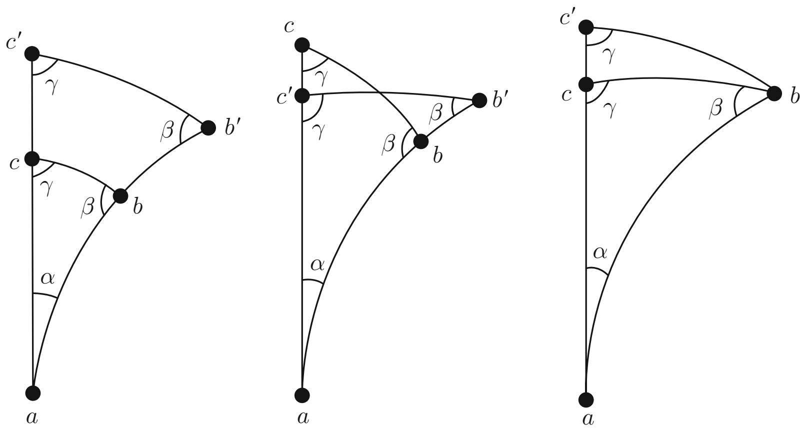
## Proof

We have to prove that if we have two triangles $T$ and $T'$ with vertices $a, b, c$ and $a', b', c'$, and if the angle $\alpha$ of $T$ at $a$ is equal to that of $T'$ at $a'$, and similarly for the angles $\beta$ at $b$ and $b'$ and for the angles $\gamma$ at $c$ and $c'$, then there is an isometry $F$ of $\mathcal{H}$ such that $F(a) = a'$, $F(b) = b'$ and $F(c) = c'$.

Let $l$, $m$, $n$ and $l'$, $m'$, $n'$ be the sides of $T$ and $T'$ (so that $l$ is the side opposite the vertex $a$, etc.). It is enough to prove the theorem in the special case in which $a = a' = i$ and $m = m'$ is the imaginary axis. For by Proposition 11.2.3, there is an isometry $G$ that takes $a$ to $i$ and $m$ to the imaginary axis, and an isometry $G'$ that takes $a'$ to $i$ and $m'$ to the imaginary axis. If $F$ is the desired isometry in the special case, then $(G')^{-1} \circ F \circ G$ is the desired isometry in the general case.

Assume then that $a = a' = i$ and $m = m'$ is the imaginary axis. By applying the reflection in the imaginary axis if necessary, we can further assume that $b$ and $b'$ are on the same side of the imaginary axis. Then either the hyperbolic lines $n$ and $n'$ coincide, or one is obtained from the other by applying the inversion $\mathcal{I}_{0,1}$ (which fixes $m$ and the vertex $i$). Hence, we can assume that $n = n'$.

If now $b = b'$ and $c = c'$ the theorem is proved. If not, then we must be in one of the three situations shown below. By making use of Theorem 11.1.5, we shall prove that each of these situations is impossible.




In the first case, the angle sum of the quadrilateral with vertices $b, c, c', b'$ is

$$(\pi - \beta) + (\pi - \gamma) + \gamma + \beta = 2\pi,$$

whereas by Theorem 11.1.5 the angle sum must be $< 2\pi$.

In the second case, the angle sum of the triangle with vertices $d, b', b$ is

$$\delta + (\pi - \beta) + \beta,$$

where $\delta$ is the angle between $l$ and $l'$ at their intersection point $d$. This is $> \pi$, again contradicting Theorem 11.1.5.

Finally, in the third case the triangle with vertices $b, c, c'$ has angle sum

$$\delta + (\pi - \gamma) + \gamma > \pi,$$

where $\delta$ is as in the preceding case (if $c$ and $c'$ are interchanged the argument is the same). $\square$

It follows from this theorem that there must be a formula for the lengths of the sides of a triangle in $\mathcal{H}$ in terms of its angles. Although we could prove such a formula now, it is slightly easier to establish it in a different model of hyperbolic geometry, and this is what we consider next.

## *EXERCISES*

11.2.1 Show that if $a, b \in \mathcal{H}$, the hyperbolic distance $d_{\mathcal{H}}(a, b)$ is the length of the shortest curve in $\mathcal{H}$ joining $a$ and $b$.

11.2.2 Show that, if $l$ is any hyperbolic line in $\mathcal{H}$ and $a$ is a point not on $l$, there are infinitely many hyperbolic lines passing through $a$ that do not intersect $l$.

11.2.3 Let $a$ be a point of $\mathcal{H}$ that is not on a hyperbolic line $l$. Show that there is a unique hyperbolic line $m$ passing through $a$ that intersects $l$ perpendicularly. If $b$ is the point of intersection of $l$ and $m$, and $c$ is any other point of $l$, prove that

$$d_{\mathcal{H}}(a, b) < d_{\mathcal{H}}(a, c).$$

Thus, $b$ is the unique point of $l$ that is closest to $a$.

11.2.4 This exercise and the next determine all the isometries of $\mathcal{H}$.

(i) Let $F$ be an isometry of $\mathcal{H}$ that fixes each point of the imaginary axis $l$ and each point of the semicircle geodesic $m$ at the centre of the origin and radius 1. Show that $F$ is the identity map.

(ii) Let $F$ be an isometry of $\mathcal{H}$ such that $F(l) = l$ and $F(m) = m$, where $l$ and $m$ are as in (i). Prove that $F$ is the identity map, the reflection $R_0$, the inversion $\mathcal{I}_{0,1}$ or the composite $\mathcal{I}_{0,1} \circ R_0$ (in the notation at the beginning of this section).

(iii) Show that every isometry of $\mathcal{H}$ is a composite of elementary isometries.

(iv) Show that every isometry of $\mathcal{H}$ is a composite of reflections and inversions in lines and circles perpendicular to the real axis.

11.2.5 A Möbius transformation (see Appendix 2) is said to be *real* if it is of the form

$$M(z) = \frac{az + b}{cz + d},$$

where $a, b, c, d \in \mathbb{R}$. Show that:

(i) Any composite of real Möbius transformations is a real Möbius transformation, and the inverse of any real Möbius transformation is a real Möbius transformation.

(ii) The Möbius transformations that take $\mathcal{H}$ to itself are exactly the real Möbius transformations such that $ad - bc > 0$.

(iii) Every real Möbius transformation is a composite of elementary isometries of $\mathcal{H}$, and hence is an isometry of $\mathcal{H}$.

(iv) If $J(z) = -\bar{z}$ and $M$ is a real Möbius transformation, $M \circ J$ is an isometry of $\mathcal{H}$.

(v) If we call an isometry of type (iii) or (iv) a *Möbius isometry*, any composite of Möbius isometries is a Möbius isometry;

(vi) Every isometry of $\mathcal{H}$ is a Möbius isometry.

## 11.3 Poincaré disc model

We now consider a model of hyperbolic geometry based on the unit disc in the complex plane. Poincaré used this model to bring hyperbolic geometry into the mainstream of mathematics by establishing its connections with other areas, notably complex analysis and number theory.

We consider the transformation

$$\mathcal{P}(z) = \frac{z - i}{z + i}.$$

It defines a bijection between the complex plane with the point $-i$ removed and the complex plane with the point 1 removed, its inverse being

$$\mathcal{P}^{-1}(z) = \frac{z + 1}{i(z - 1)}.$$

In particular, $\mathcal{P}$ is well defined at all points of $\mathcal{H}$ and its boundary the real axis.

Let us determine the image of $\mathcal{H}$ under $\mathcal{P}$. We have,

$$\mathcal{P}(v + iw) = \frac{v + i(w - 1)}{v + i(w + 1)},$$

so

$$|\mathcal{P}(v+iw)| = \left(\frac{v^2+w^2+1-2w}{v^2+w^2+1+2w}\right)^{1/2}.$$

Hence, $|\mathcal{P}(v+iw)|$ is $< 1$ if $w > 0$, is $= 1$ if $w = 0$ and is $> 1$ if $w < 0$. Thus, $\mathcal{P}$ takes $\mathcal{H}$ to the unit disc

$$\mathcal{D} = \{z \in \mathbb{C} \mid |z| < 1\},$$

and the real axis to the boundary of $\mathcal{D}$, i.e., the unit circle $\mathcal{C}$ given by $|z| = 1$.

## Definition 11.3.1

The Poincaré disc model $\mathcal{D}_P$ of hyperbolic geometry is the disc $\mathcal{D}$ equipped with the first fundamental form for which $\mathcal{P} : \mathcal{H} \to \mathcal{D}_P$ is an isometry.

## Proposition 11.3.2

The first fundamental form of $\mathcal{D}_P$ is

$$\frac{4(dv^2+dw^2)}{(1-v^2-w^2)^2}.$$

In particular, $\mathcal{D}_P$ is a conformal model of hyperbolic geometry.

## Proof

If $\tilde{v} + i\tilde{w} = \mathcal{P}^{-1}(v+iw)$, we find that

$$\tilde{v} = \frac{-2w}{(v-1)^2+w^2}, \quad \tilde{w} = \frac{1-v^2-w^2}{(v-1)^2+w^2},$$

which gives

$$d\tilde{v} = \frac{4(v-1)wdv - 2((v-1)^2-w^2)dw}{((v-1)^2+w^2)^2},$$
$$d\tilde{w} = \frac{2((v-1)^2-w^2)dv + 4(v-1)wdw}{((v-1)^2+w^2)^2}.$$

Hence,

$$d\tilde{v}^2 + d\tilde{w}^2 = \frac{16(v-1)^2w^2 + 4((v-1)^2-w^2)^2}{((v-1)^2+w^2)^4}(dv^2+dw^2) = \frac{4(dv^2+dw^2)}{((v-1)^2+w^2)^2}$$

and so

$$\frac{d\tilde{v}^2 + d\tilde{w}^2}{\tilde{w}^2} = \frac{4(dv^2 + dw^2)}{(1 - v^2 - w^2)^2}.$$

Since the first fundamental form of $\mathcal{D}_P$ is a multiple of $du^2 + dv^2$, $\mathcal{D}_P$ is a conformal model. $\square$

Since $\mathcal{P} : \mathcal{H} \to \mathcal{D}_P$ is an isometry, it follows that the isometries of $\mathcal{D}_P$ are exactly the maps

$$\mathcal{P} \circ F \circ \mathcal{P}^{-1},$$

where $F$ is any isometry of $\mathcal{H}$. Indeed, since any composite of isometries is an isometry, $\mathcal{P} \circ F \circ \mathcal{P}^{-1}$ is an isometry of $\mathcal{D}_P$ if $F$ is an isometry of $\mathcal{H}$; conversely, if $G$ is any isometry of $\mathcal{D}_P$, then $F = \mathcal{P}^{-1} \circ G \circ \mathcal{P}$ is an isometry of $\mathcal{H}$, and $G = \mathcal{P} \circ F \circ \mathcal{P}^{-1}$.

Here is a simple application of this observation:

## Proposition 11.3.3

(i) Let $\Gamma$ be a circle that intersects $\mathcal{C}$ perpendicularly. Then, inversion in $\Gamma$ is an isometry of $\mathcal{D}_P$.

(ii) Let $l$ be a line passing through the origin (and so perpendicular to $\mathcal{C}$). Then, (Euclidean) reflection in $l$ is an isometry of $\mathcal{D}_P$.

## Proof

For (i), let $\Gamma$ have centre $a \in \mathbb{C}$ and radius $r > 0$; then, the inversion in $\Gamma$ is given by

$$\mathcal{I}_{a,r} = a + \frac{r^2}{\bar{z} - \bar{a}}.$$

By Proposition A.2.8, $\mathcal{I}_{a,r}$ takes $\mathcal{D}_P$ to itself. We have to show that $\mathcal{P}^{-1} \circ \mathcal{I}_{a,r} \circ \mathcal{P}$ is an isometry of $\mathcal{H}$. We find that

$$\mathcal{I}_{a,r}(\mathcal{P}(z)) = \frac{(a - |a|^2 + r^2)\bar{z} + i(a + |a|^2 - r^2)}{(1 - \bar{a})\bar{z} + i(1 + \bar{a})}.$$

Now, since $\Gamma$ intersects $\mathcal{C}$ at right angles, $|a|^2 = r^2 + 1$, so

$$\mathcal{I}_{a,r}(\mathcal{P}(z)) = \frac{(a - 1)\bar{z} + i(a + 1)}{(1 - \bar{a})\bar{z} + i(1 + \bar{a})}.$$

This leads to

$$\mathcal{P}^{-1}(\mathcal{I}_{a,r}(\mathcal{P}(z))) = \frac{i(a-\bar{a})\bar{z} - (2+a+\bar{a})}{(2-a-\bar{a})\bar{z} - i(a-\bar{a})}.$$

This is a real Möbius transformation (Exercise 11.2.5) and so is an isometry of $\mathcal{H}$.

For (ii), let $l$ make an angle $\theta$ with the real axis, so that reflection in $l$ is the map $\mathcal{R}(z) = e^{2i\theta}\bar{z}$. We find that

$$\mathcal{P}^{-1}(\mathcal{R}(\mathcal{P}(z))) = \frac{z\cos\theta + \sin\theta}{-z\sin\theta + \cos\theta},$$

which is again a real Möbius transformation. □

Note that simple isometries in one model may not correspond to simple isometries in the other. For example, it is clear from Proposition 11.3.2 that any rotation about the origin is an isometry of $\mathcal{D}_P$ (because such a rotation is an isometry of the Euclidean plane, and hence preserves $dv^2 + dw^2$ and $v^2 + w^2$), but the corresponding isometry of $\mathcal{H}$ is quite complicated (it is not an elementary isometry, for example).

Since $\mathcal{P}$ is an isometry, the geodesics (i.e., the hyperbolic lines) in $\mathcal{D}_P$ are the images under $\mathcal{P}$ of the geodesics in $\mathcal{H}$. Hence, the properties of the hyperbolic lines in $\mathcal{H}$ can be transferred to $\mathcal{D}_P$. For example, if $a$ and $b$ are two distinct points of $\mathcal{D}_P$, then by Proposition 11.1.3, there is a unique hyperbolic line $l$ in $\mathcal{H}$ passing through the distinct points $\mathcal{P}^{-1}(a)$ and $\mathcal{P}^{-1}(b)$, so $\mathcal{P}(l)$ is the unique hyperbolic line in $\mathcal{D}_P$ passing through $a$ and $b$. Similarly, Proposition 11.2.3 holds as stated with $\mathcal{H}$ replaced by $\mathcal{D}_P$.

The distance between two points of $\mathcal{D}_P$ is given by

$$d_{\mathcal{D}_P}(a,b) = d_{\mathcal{H}}(\mathcal{P}^{-1}(a), \mathcal{P}^{-1}(b)), \qquad a,b \in \mathcal{D}_P.$$

Using the formula in Proposition 11.1.4, it is straightforward (see Exercise 11.3.1) to prove

## Proposition 11.3.4

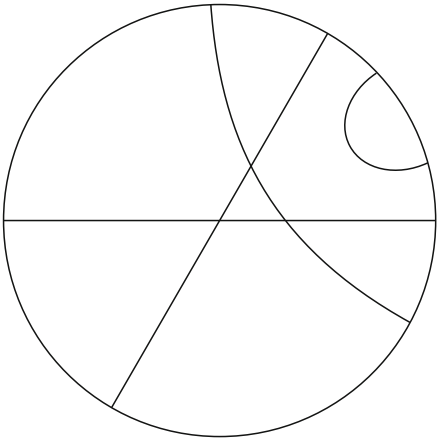
For $a, b \in \mathcal{D}_P$, we have

$$d_{\mathcal{D}_P}(a,b) = 2\tanh^{-1}\frac{|b-a|}{|1-\bar{a}b|}.$$

The explicit form of the hyperbolic lines in $\mathcal{D}_P$ can, of course, be determined from the first fundamental form in Proposition 11.3.2. But it is easier to make use of some simple properties of the map $\mathcal{P}$.

## Proposition 11.3.5

The hyperbolic lines in $\mathcal{D}_P$ are the lines and circles that intersect $\mathcal{C}$ perpendicularly (see the diagram below).



## Proof

This follows from Proposition 11.1.2 and the fact that $\mathcal{P}$ takes the boundary of $\mathcal{H}$ to that of $\mathcal{D}_P$ and, being a Möbius transformation, preserves (Euclidean) angles and takes lines and circles to lines and circles (see Appendix 2). □

Note that Proposition 11.3.3 tells us that 'reflection' in any hyperbolic line in $\mathcal{D}_P$ is an isometry of $\mathcal{D}_P$ – 'reflection' in a circle being interpreted as inversion (and Exercise 11.3.5 shows that every isometry of $\mathcal{D}_P$ is a composite of such reflections).

We shall now establish some new properties of hyperbolic geometry to which the Poincaré model is particularly well suited, starting with the basic result in hyperbolic trigonometry.

## Theorem 11.3.6

Consider a hyperbolic triangle with angles $\alpha, \beta, \gamma$ and sides of length $A, B, C$ (so that $A$ is the length of the side opposite $\alpha$, etc.). Then,

$$\cosh C = \cosh A \cosh B - \sinh A \sinh B \cos \gamma,$$

and two analogous formulas can be obtained by applying the cyclic permutations $\alpha \to \beta \to \gamma \to \alpha$ and $A \to B \to C \to A$.

This formula is called the 'hyperbolic cosine rule' because it becomes the usual cosine rule when $A, B$, and $C$ are small: using the approximations $\cosh A = 1 + \frac{1}{2}A^2$ and $\sinh A = A$, etc. we get

$$C^2 = A^2 + B^2 - 2AB\cos\gamma$$

(compare the spherical case treated in Proposition 6.5.3(i)).

## Proof

Let $a, b$, and $c$ be the vertices of the triangle, so that $\alpha$ is the angle at $a$, etc. By applying an isometry of $\mathcal{D}_P$ that takes $c$ to the origin followed by a suitable rotation about the origin (i.e. another isometry), we can assume that $c = 0 \in \mathcal{D}_P$ and that $a > 0$. By Proposition 11.3.4,

$$a = \tanh\frac{1}{2}B, \quad b = e^{i\gamma}\tanh\frac{1}{2}A.$$

Now

$$\cosh A = \cosh^2\frac{1}{2}A + \sinh^2\frac{1}{2}A = \frac{1+\tanh^2\frac{1}{2}A}{\operatorname{sech}^2\frac{1}{2}A} = \frac{1+\tanh^2\frac{1}{2}A}{1-\tanh^2\frac{1}{2}A} = \frac{1+|a|^2}{1-|a|^2}$$

and by Proposition 11.3.4 again

$$\tanh\frac{1}{2}C = \frac{|b-a|}{|1-\bar{a}b|},$$

so

$$\begin{aligned}
\cosh C &= \frac{1+\tanh^2\frac{1}{2}C}{1-\tanh^2\frac{1}{2}C} \\
&= \frac{|1-\bar{a}b|^2 + |b-a|^2}{|1-\bar{a}b|^2 - |b-a|^2} \\
&= \frac{(1-\bar{a}b)(1-a\bar{b}) + (b-a)(\bar{b}-\bar{a})}{(1-\bar{a}b)(1-a\bar{b}) - (b-a)(\bar{b}-\bar{a})} \\
&= \frac{1+|a|^2+|b|^2+|a|^2|b|^2 - 2(\bar{a}b + a\bar{b})}{1-|a|^2-|b|^2+|a|^2|b|^2} \\
&= \frac{(1+|a|^2)(1+|b|^2) - 2(\bar{a}b+a\bar{b})}{(1-|a|^2)(1-|b|^2)} \\
&= \cosh A\cosh B - 4\cos\gamma\frac{\tanh\frac{1}{2}A\tanh\frac{1}{2}B}{(1-\tanh^2\frac{1}{2}A)(1-\tanh^2\frac{1}{2}B)} \\
&= \cosh A\cosh B - \sinh A\sinh B\cos\gamma,
\end{aligned}$$

using $\sinh A = 2\sinh\frac{1}{2}A\cosh\frac{1}{2}A$. □

In particular, we have the hyperbolic analogue of Pythagoras' theorem:

## Corollary 11.3.7

Suppose that a hyperbolic triangle has sides of lengths $A, B$, and $C$ and that the angle opposite the side of length $C$ is a right angle. Then,

$$\cosh C = \cosh A \cosh B.$$

Further results in hyperbolic trigonometry can be found in the exercises.

## *EXERCISES*

11.3.1 Prove Proposition 11.3.4.

11.3.2 Let $l$ and $m$ be hyperbolic lines in $\mathcal{D}_P$ that intersect at right angles. Prove that there is an isometry of $\mathcal{D}_P$ that takes $l$ to the real axis and $m$ to the imaginary axis. How many such isometries are there?

11.3.3 Show that the Möbius transformations that take $\mathcal{D}_P$ to itself are those of the form

$$z \mapsto \frac{az+b}{\bar{b}z+\bar{a}}, \quad |a| > |b|.$$

Recall (Exercise 6.5.4) that these are unitary Möbius transformations.

11.3.4 Show that the isometries of $\mathcal{D}_P$ are the transformations of the following two types:

$$z \mapsto \frac{az+b}{\bar{b}z+\bar{a}}, \qquad z \mapsto \frac{a\bar{z}+b}{\bar{b}\bar{z}+\bar{a}},$$

where $a$ and $b$ are complex numbers such that $|a| > |b|$. Note that this and the preceding exercise show that the isometries of $\mathcal{D}_P$ are exactly the Möbius and conjugate-Möbius transformations that take $\mathcal{D}_P$ to itself.

11.3.5 Prove that every isometry of $\mathcal{D}_P$ is the composite of finitely many isometries of the two types in Proposition 11.3.3.

11.3.6 Consider a hyperbolic triangle with vertices $a$, $b$, and $c$, sides of length $A, B$, and $C$ and angles $\alpha, \beta$, and $\gamma$ (so that $A$ is the length of the side opposite $a$ and $\alpha$ is the angle at $a$, etc.). Prove the *hyperbolic sine rule*

$$\frac{\sin\alpha}{\sinh A} = \frac{\sin\beta}{\sinh B} = \frac{\sin\gamma}{\sinh C}.$$

11.3.7 With the notation in the preceding exercise, suppose that $\gamma = \pi/2$. Prove that:

(i) $\cos\alpha = \frac{\sinh B \cosh A}{\sinh C}$.

(ii) $\cosh A = \frac{\cos\alpha}{\sin\beta}$.

(iii) $\sinh A = \frac{\tanh B}{\tan\beta}$.

11.3.8 With the notation in Exercise 11.3.6, prove that

$$\cosh A = \frac{\cos\alpha + \cos\beta\cos\gamma}{\sin\beta\sin\gamma}.$$

This is the formula we promised at the end of Section 11.2 for the lengths of the sides of a hyperbolic triangle in terms of its angles.

11.3.9 Show that if $\mathbb{R}^2$ is provided with the first fundamental form

$$\frac{4(du^2 + dv^2)}{(1 + u^2 + v^2)^2},$$

the stereographic projection map $\Pi : S^2 \backslash \{\text{north pole}\} \to \mathbb{R}^2$ defined in Example 6.3.5 is an isometry. Note the similarity between this formula and that in Proposition 11.3.2: the plane with this first fundamental form provides a 'model' for the sphere in the same way as the half-plane with the first fundamental form in Proposition 11.3.2 is a 'model' for the pseudosphere.

## 11.4 Hyperbolic parallels

In Euclidean plane geometry, there are many equivalent criteria for two lines $l$ and $m$ to be parallel. For example:

(i) $l$ and $m$ do not intersect.

(ii) $l$ and $m$ have a common perpendicular line.

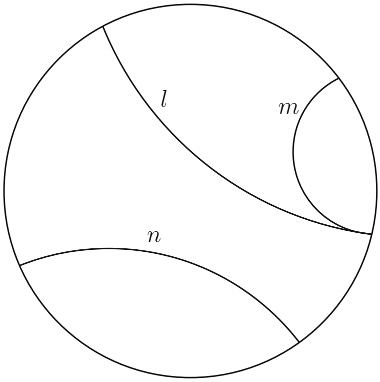
(iii) $l$ and $m$ are a constant distance apart.

(A fourth criterion is considered in Exercise 11.4.3.) In hyperbolic geometry, these conditions are *not* equivalent. In fact, two distinct hyperbolic lines are *never* a constant distance apart (see Exercise 11.4.2), so (iii) is not relevant to the discussion of parallels in hyperbolic geometry. Further, it is clear that in hyperbolic geometry (i) does not imply (ii) (consider two half-line geodesics in $\mathcal{H}$, for example), so we must distinguish two cases:

## Definition 11.4.1

Let $l$ and $m$ be hyperbolic lines in $\mathcal{D}_P$ that do not intersect at any point of $\mathcal{D}_P$. If $l$ and $m$ intersect at a point of the boundary of $\mathcal{D}_P$ they are said to be *parallel*; otherwise they are said to be *ultra-parallel.*

In the diagram below, $l$ and $m$ are parallel, and $l$ and $n$ are ultra-parallel.




We have already noted (Proposition 11.1.3(ii)) that the parallel axiom does not hold in hyperbolic geometry. In fact, if $a$ is a point that is not on a hyperbolic line $l$, there are infinitely many hyperbolic lines through $a$ that do not intersect $l$ (see Exercise 11.1.1). The following result shows that exactly two of these hyperbolic lines are parallel to $l$.

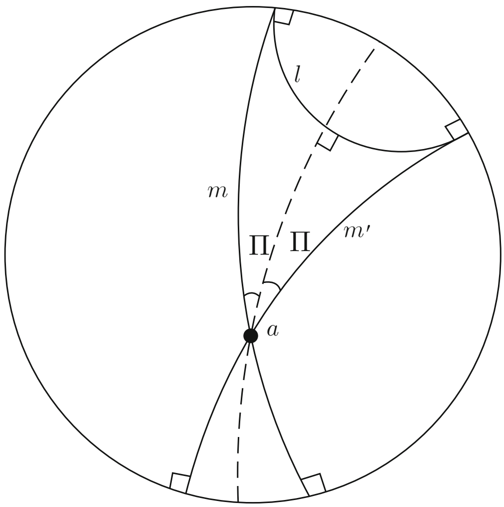
## Proposition 11.4.2

Suppose that $a \in \mathcal{D}_P$ is a point not on a hyperbolic line $l$. Then, there are exactly two hyperbolic lines, say $m$ and $m'$, passing through $a$ that are parallel to $l$. The angle between $m$ and $m'$ at $a$ is $2\Pi$, where

$$\sin \Pi = \operatorname{sech} d,$$

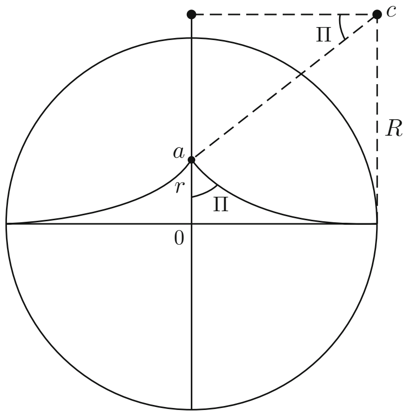
and $d$ is the hyperbolic distance of $a$ from $l$ (Exercise 11.1.3). Moreover, a hyperbolic line through $a$ intersects $l$ if and only if it lies between $m$ and $m'$ on the same side of $a$ as $l$, and the hyperbolic line through $a$ perpendicular to $l$ bisects the angle between $m$ and $m'$.

The angle $\Pi$ is called the *angle of parallelism*.




## Proof

We first show that there is an isometry of $\mathcal{D}_P$ that takes $l$ to the real axis and $a$ to a point on the imaginary axis. In that case, all the assertions made in the proposition are clear, except for the formula for $\Pi$.




Let $\tilde{a} = \mathcal{P}^{-1}(a)$, $\tilde{l} = \mathcal{P}^{-1}(l)$. There is an isometry $F$ of $\mathcal{H}$ that takes $\tilde{l}$ to the imaginary axis; let $b = F(\tilde{a})$. The isometry $D_{1/|b|}$ takes $b$ to a point on the unit circle $v^2 + w^2 = 1$ and fixes the imaginary axis. Now note that $\mathcal{P}$ takes the imaginary axis in $\mathcal{H}$ to the real axis in $\mathcal{D}_P$ and the unit circle in $\mathcal{H}$ to the imaginary axis in $\mathcal{D}_P$.

We can therefore assume that $l$ is the real axis and that $a = ir$ where $r = \tanh\frac{1}{2}d$ by Proposition 11.3.4. The circle $m$ through $a$ that touches the real axis at 1 has centre $c = 1 + iR$ and radius $R$ for some $R > 0$, and so has equation

$$|z - 1 - iR| = R.$$

Since $m$ passes through $ir$, we have $|-1 + i(r - R)| = R$, which gives

$$R = \frac{1 + r^2}{2r}.$$

In the right-angled (Euclidean) triangle with vertices $a$, $iR$ and $c$, the hypotenuse is perpendicular to $m$, so the angle of the triangle at $a$ is $\pi/2 - \Pi$ (see the diagram above). Hence, by Euclidean trigonometry,

$$R \sin \Pi = R - r$$

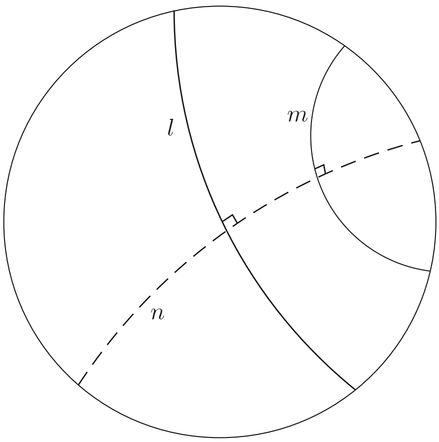
and we get

$$\sin \Pi = 1 - \frac{r}{R} = 1 - \frac{2r^2}{1 + r^2} = \frac{1 - r^2}{1 + r^2} = \frac{1 - \tanh^2 \frac{1}{2}d}{1 + \tanh^2 \frac{1}{2}d} = \frac{1}{\cosh d}. \qquad \square$$

As we mentioned above, one characterization of parallel lines in Euclidean plane geometry is that such lines have a common perpendicular. In hyperbolic geometry, this property characterizes ultra-parallels:

## Proposition 11.4.3

Two hyperbolic lines in $\mathcal{D}_P$ are ultra-parallel if and only if they have a common perpendicular (i.e., a hyperbolic line that intersects them both at right-angles). In that case the common perpendicular is unique.



In Euclidean plane geometry, of course, two parallel lines have infinitely many common perpendiculars.

## Proof

Suppose first that $l$ and $m$ are hyperbolic lines in $\mathcal{D}_P$ that have a common perpendicular $n$ which intersects them at the points $a$ and $b$. We can assume that $l$ and $n$ are the real and imaginary axes, respectively, and that $a$ is the origin (see Exercise 11.3.2). Then $m$ is part of a circle with centre at some point $iR$ on the imaginary axis, where $|R| > 1$. Since $m$ intersects $\mathcal{C}$ at right angles, the radius $r$ of $\mathcal{C}$ satisfies

$$R^2 = r^2 + 1.$$

In particular, $|R| > r$, so $m$ does not intersect the real axis. Hence, $l$ and $m$ are ultra-parallel.

Conversely, suppose that $l$ and $m$ are ultra-parallel. As before, we can assume that $l$ is the real axis. Suppose that $m$ is the circle with centre $a$ and radius $r$; then, as above,

$$|a|^2 = r^2 + 1. \tag{11.5}$$

We claim that

$$-1 < \mathfrak{Re}(a) < 1. \tag{11.6}$$

Indeed, $m$ intersects the real axis at a point $v$ if and only if

$$|v - a| = r.$$

In view of (11.5), this gives

$$v^2 - 2v\,\mathfrak{Re}(a) + 1 = 0. \tag{11.7}$$

If $|\mathfrak{Re}(a)| > 1$, Eq. 11.7 has two distinct real roots whose product is equal to 1, hence one root $v$ satisfies $-1 < v < 1$. This means that $l$ and $m$ intersect in $\mathcal{D}_P$, contrary to assumption. Similarly, if $|\mathfrak{Re}(a)| = 1$, Eq. 11.10 has $\pm 1$ as a repeated root, so $l$ touches $m$ at 1 or $-1$ on the boundary of $\mathcal{D}_P$, again contrary to assumption. Hence, (11.6) must hold.

We now consider a circle with centre $b$ on the real axis and radius $s$. This intersects both $m$ and $\mathcal{C}$ at right angles if and only if

$$b^2 = s^2 + 1 \quad \text{and} \quad |b - a|^2 = r^2 + s^2.$$

If $\mathfrak{Re}(a) \neq 0$, these equations have the unique solution

$$b = \frac{1}{\mathfrak{Re}(a)}, \quad s = \sqrt{(\mathfrak{Re}(a)^{-2} - 1},$$

and the corresponding circle $n$ is the unique common perpendicular to $l$ and $m$. If $\mathfrak{Re}(a) = 0$, it is clear that the imaginary axis is the unique common perpendicular. $\square$

## *EXERCISES*

11.4.1 Which pairs of hyperbolic lines in $\mathcal{H}$ are parallel? Ultra-parallel?

11.4.2 Let $l$ be the imaginary axis in $\mathcal{H}$. Show that, for any $R > 0$, the set of points that are a distance $R$ from $l$ is the union of two half-lines passing through the origin, but that these half-lines are *not* hyperbolic lines. This contrasts with the situation in Euclidean geometry, in which the set of points at a fixed distance from a line is a pair of lines.

11.4.3 Let $a$ and $b$ be two distinct points in $\mathcal{D}_P$, and let $0 < \mathcal{A} < \pi$. Show that the set of points $c \in \mathcal{D}_P$ such that the hyperbolic triangle with vertices $a, b$ and $c$ has area $\mathcal{A}$ is the union of two segments of lines or circles, but that these are not hyperbolic lines. Note that this equal-area property could be used to characterize lines in Euclidean geometry.

# 11.5 Beltrami–Klein model

The final model of non-Euclidean geometry that we shall discuss was actually the first to be introduced by Beltrami, but it was Klein who realised that this model could be used to unify non-Euclidean geometry with *projective geometry*, a subject that has been studied since antiquity. (We do not assume that the reader is familiar with projective geometry.)

The model is constructed by using two projections of the unit sphere $S^2$. We recall the stereographic projection map $\Pi$ (Example 6.3.5) from $S^2$ to the $xy$-plane. This map defines a diffeomorphism from the 'southern hemisphere'

$$S^2_- = \{(x, y, z) \in S^2 \mid z < 0\}$$

to the unit disc

$$\mathcal{D} = \{(x, y, 0) \in \mathbb{R}^3 \mid x^2 + y^2 < 1\}.$$

We shall also need the 'vertical' projection of $\mathbb{R}^3$ onto the $xy$-plane:

$$\mathrm{pr}(x, y, z) = (x, y, 0).$$

This also defines a diffeomorphism from $S^2_-$ to $\mathcal{D}$. Hence, the composite map

$$\mathcal{K} = \mathrm{pr} \circ \Pi^{-1} : \mathcal{D} \to \mathcal{D}$$

is a diffeomorphism. It is easy to see (Exercise 11.5.1) that, if we identify the $xy$-plane with $\mathbb{C}$ by $(x, y, 0) \mapsto x + iy$ as usual, then

$$\mathcal{K}(z) = \frac{2z}{|z|^2 + 1}, \qquad z \in \mathcal{D}. \tag{11.8}$$

## Definition 11.5.1

The Beltrami-Klein model $\mathcal{D}_K$ of non-Euclidean geometry is the disc $\mathcal{D}$ equipped with the first fundamental form for which the diffeomorphism

$$\mathcal{K} : \mathcal{D}_P \to \mathcal{D}_K$$

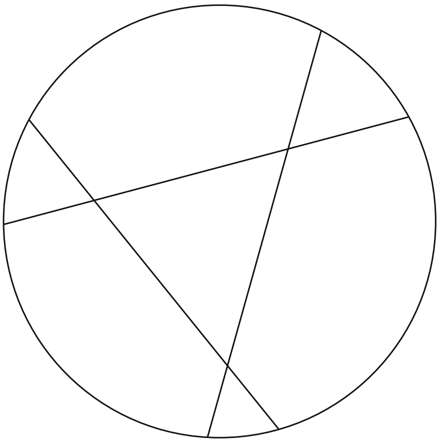
is an isometry.

We shall not need to know the first fundamental form of $\mathcal{D}_K$ explicitly (it was actually computed in Exercise 8.3.1(iii)).

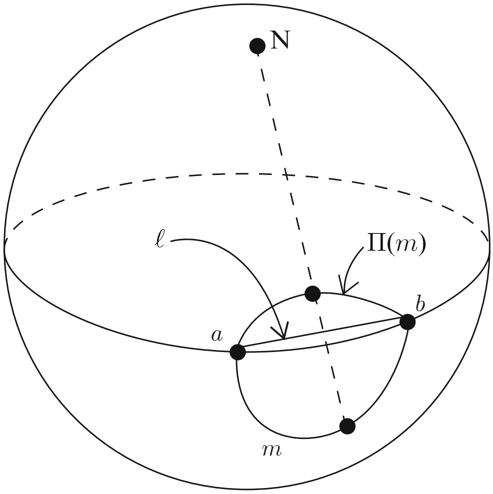
The Beltrami-Klein model has the following remarkable property.

## Proposition 11.5.2

The hyperbolic lines in the Beltrami-Klein model are the (Euclidean) straight line segments contained in the disc $\mathcal{D}_K$.



*Proof 11.5.2* Let $l$ be the line segment joining points $a$ and $b$ on $\mathcal{C}$. The curve on $S^2_-$ that corresponds to $\ell$ under the projection pr is the intersection of $S^2$ with the plane perpendicular to the $xy$-plane that contains $\ell$. This is a semicircle $m$, say, that intersects $\mathcal{C}$ at right angles at $a$ and $b$.




Since $\Pi$ is a conformal map that takes circles on $S^2$ to lines and circles in the $xy$-plane (see Example 6.3.5 and Exercise 6.3.7), $\Pi\,(m)$ is an arc of a circle in $\mathcal{D}$ that intersects the boundary of $\mathcal{D}$ at right angles, in other words a hyperbolic line in $\mathcal{D}_P$. It follows that every line segment in $\mathcal{D}_K$ is a hyperbolic line. Since there is a line segment passing through any given point of $\mathcal{D}_K$ in any given direction, these must be all of the hyperbolic lines in $\mathcal{D}_K$ (see Proposition 9.2.4). $\square$

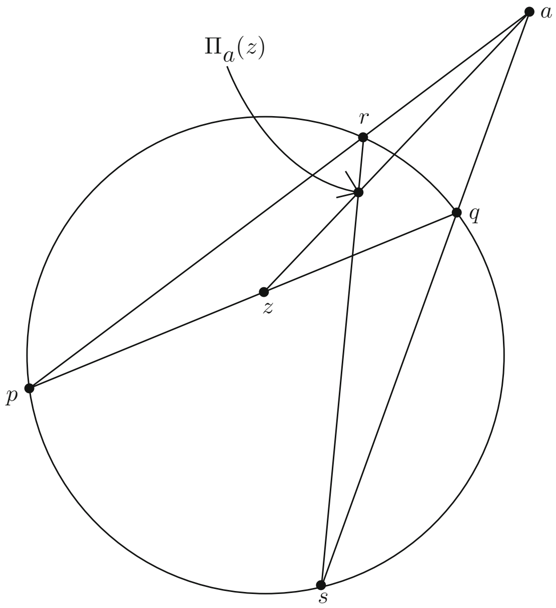
## Corollary 11.5.3

$\mathcal{D}_K$ is not a conformal model of hyperbolic geometry.

## Proof

Consider a hyperbolic triangle in $\mathcal{D}_K$. By Proposition 11.5.2 this is also a Euclidean triangle, so the sum of its internal Euclidean angles is $\pi$. But, by Theorem 11.1.5, the sum of its internal hyperbolic angles is $< \pi$. $\square$

The isometries of $\mathcal{D}_K$ can, of course, be deduced from those of $\mathcal{D}_P$ by using the isometry $\mathcal{K}$. For example, any rotation about the origin is an isometry of $\mathcal{D}_K$. For, if $\rho_\theta$ is such a rotation by an angle $\theta$, so that $\rho_\theta(z) = e^{i\theta}z$, it is clear from Eq. 11.8 that $\mathcal{K} \circ \rho_\theta \circ \mathcal{K}^{-1} = \rho_\theta$ and we know that $\rho_\theta$ is an isometry of $\mathcal{D}_P$ (see the remarks following the proof of Proposition 11.3.3). But to proceed further, it is more instructive to take a different, and more geometric, approach.




If $a \in \mathbb{C}$ and $|a| > 1$, define the *perspectivity*

$$\Pi_a : \mathcal{D}_K \to \mathcal{D}_K$$

*with centre* $a$ as follows. Let $z \in \mathcal{D}_K$ and let $l$ be any hyperbolic line in $\mathcal{D}_K$ passing through $z$. Thus, $l$ is a (Euclidean) line segment that intersects $\mathcal{C}$ at two points, say $p$ and $q$. Let the lines through $a$ and $p$ and through $a$ and $q$ intersect $\mathcal{C}$ again at $r$ and $s$, respectively (if the line through $a$ and $p$ happens to be tangent to $\mathcal{C}$ at $p$, then $r = p$; and similarly for the line through $a$ and $q$). Then, $\Pi_a(z)$ is defined to be the point of intersection of the line through $a$ and $z$ with the line through $r$ and $s$ (see the diagram above).

Of course, it is not obvious that this definition makes sense, i.e., that $\Pi_a(z)$ depends only on $z$ (and $a$) and not on the choice of the line $l$, but this follows from

## Proposition 11.5.4

With the above notation,

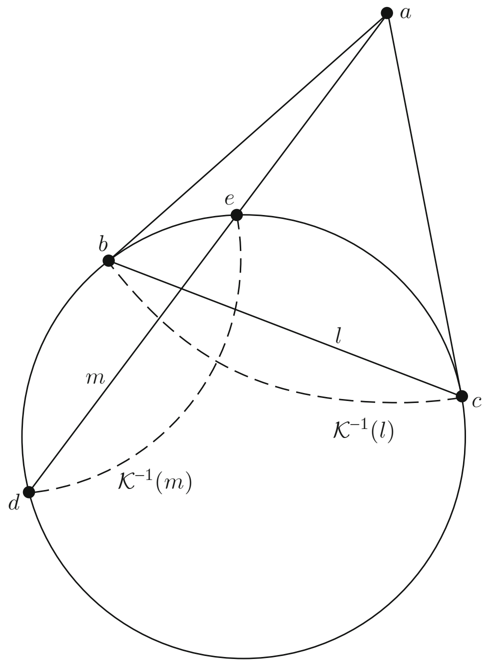
$$\Pi_a = \mathcal{K} \circ \mathcal{I}_{a,r} \circ \mathcal{K}^{-1},$$

where $r = \sqrt{|a|^2 - 1}$. In particular, $\Pi_a$ is an isometry of $\mathcal{D}_K$.

To prove this we need

## Lemma 11.5.5

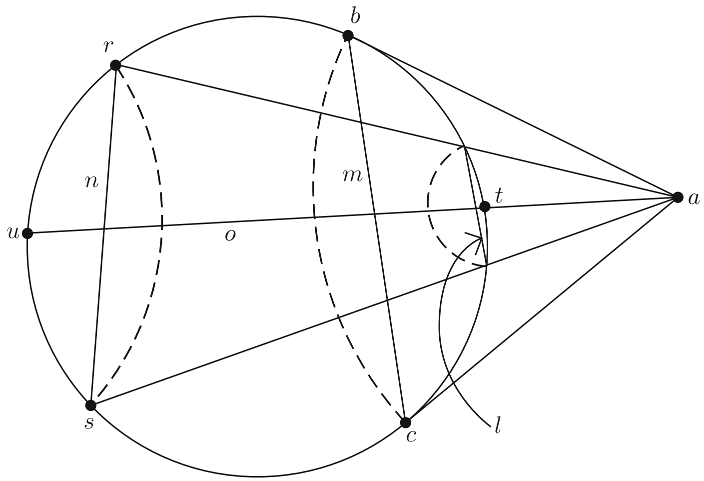
Let $l$ and $m$ be hyperbolic lines in $\mathcal{D}_K$ and suppose that these lines intersect $\mathcal{C}$ at the points $b, c$ and $d, e$, respectively. Suppose that the tangents to $\mathcal{C}$ at $b$ and $c$ intersect at $a$, and that the extension of $m$ passes through $a$. Then, $l$ and $m$ intersect at right angles in the hyperbolic sense.




*Proof 11.5.5* The hyperbolic lines $\mathcal{K}^{-1}(l)$ and $\mathcal{K}^{-1}(m)$ in $\mathcal{D}_P$ corresponding to $l$ and $m$ are circular arcs that intersect $\mathcal{C}$ at right angles at the points $b, c$ and $d, e$, respectively. Let $\mathcal{I}$ be the inversion in the circle of which $\mathcal{K}^{-1}(l)$ is an arc, so that $\mathcal{I}$ is an isometry of $\mathcal{D}_P$ (see Appendix 2, especially Proposition A.2.8). Now $\mathcal{I}$ takes $\mathcal{K}^{-1}(m)$ to a circular arc that intersects $\mathcal{C}$ at right angles (Corollary A.2.7), and it obviously interchanges the points $d$ and $e$. It follows that $\mathcal{I}$ preserves $\mathcal{K}^{-1}(m)$. This implies that $\mathcal{K}^{-1}(l)$ and $\mathcal{K}^{-1}(m)$ are perpendicular in the Euclidean sense (Proposition A.2.8), and hence in the hyperbolic sense since $\mathcal{D}_P$ is a conformal model. Since $\mathcal{K} : \mathcal{D}_P \to \mathcal{D}_K$ is an isometry, $l$ and $m$ are perpendicular in the hyperbolic sense. □

*Proof 11.5.4* Let the tangents from $a$ to $\mathcal{C}$ touch it at $b$ and $c$, let $m$ be the line segment with endpoints $b, c$ and let the line through $a$ and $z$ intersect $\mathcal{C}$ at $t$ and $u$. Let $l$ be any line passing through $z$ and let $l$ intersect $\mathcal{C}$ at $p, q$. Let $r, s$

be the points of $\mathcal{C}$ such that the lines through $p$ and $r$ and through $q$ and $s$ pass through $a$, let $n$ be the line segment with endpoints $r, s$ and let $u$ be the point of intersection of $\mathcal{C}$ with the line $o$ passing through $a$ and $z$. Since $z$ is the intersection of $l$ and $o$, $\mathcal{K}^{-1}(z)$ is the intersection of $\mathcal{K}^{-1}(l)$ and $\mathcal{K}^{-1}(o)$; similarly, $\mathcal{K}^{-1}(\Pi_a(z))$ is the intersection of $\mathcal{K}^{-1}(n)$ and $\mathcal{K}^{-1}(o)$.




By Lemma 11.5.5, $m$ and $o$ are perpendicular in $\mathcal{D}_K$, so $\mathcal{K}^{-1}(m)$ and $\mathcal{K}^{-1}(o)$ are perpendicular in $\mathcal{D}_P$. It follows that $\mathcal{I}_{a,r}$ fixes $\mathcal{K}^{-1}(o)$. Since $\mathcal{I}_{a,r}$ takes $p$ to $r$ and $q$ to $s$, it takes $\mathcal{K}^{-1}(l)$ to $\mathcal{K}^{-1}(n)$. Hence, $\mathcal{I}_{a,r}$ takes $\mathcal{K}^{-1}(z)$ to $\mathcal{K}^{-1}(\Pi_a(z))$:

$$\mathcal{I}_{a,r}(\mathcal{K}^{-1}(z)) = \mathcal{K}^{-1}(\Pi_a(z)).$$

This is what we wanted to prove. □

Now that we have the isometries $\Pi_a$ at our disposal, we can prove a beautiful formula for the distance between two points of $\mathcal{D}_K$. For this, we shall need the following concept from projective geometry.

### Definition 11.5.6

If $a, b, c$, and $d$ are distinct complex numbers, their *cross-ratio* is

$$(a, b; c, d) = \frac{(a-c)(b-d)}{(a-d)(b-c)}.$$

### Proposition 11.5.7

Suppose that the points $a, b, c$, and $d$ lie on a line and that $a$ and $b$ are between $c$ and $d$. Then, $(a, b; c, d) > 0$. Moreover, if $p$ is a point distinct from $a, b, c$, and $d$ and if the lines through $p$ and each of the points $a, b, c$, and $d$ intersect another line at $a', b', c'$, and $d'$, then

$$(a, b; c, d) = (a', b'; c', d').$$

This result is expressed by saying that the cross-ratio is a 'projective invariant': the cross-ratio of four points on a line is unchanged when they are 'projected' from some point $p$ onto another line.

## Proof

Let $l$ be the line containing $a, b, c$, and $d$. Since $a$ and $b$ are on the 'same side' of $l$ relative to $c$, $\arg(a-c) = \arg(b-c)$, so

$$\frac{a-c}{b-c} = \frac{|a-c|}{|b-c|}.$$

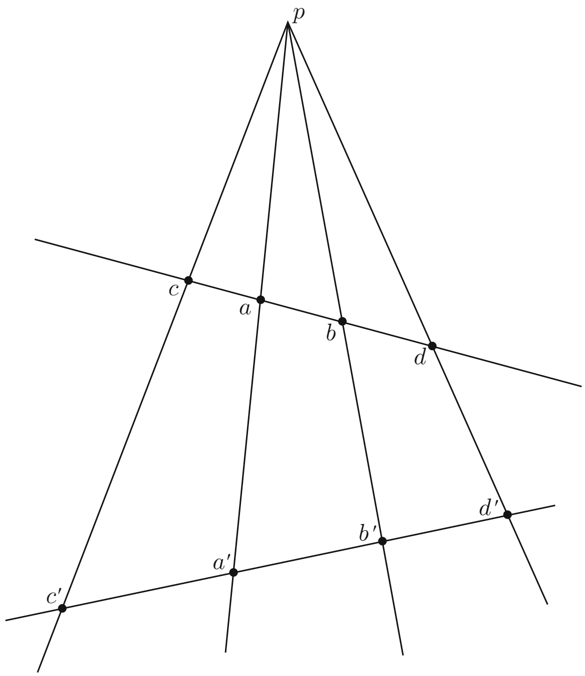
Similarly,

$$\frac{b-d}{a-d} = \frac{|b-d|}{|a-d|}.$$

Hence,

$$(a, b; c, d) = \frac{|a-c||b-d|}{|a-d||b-c|}.$$

In particular, this cross-ratio is a positive number.



Let $\angle apb$ be the angle between the lines through $p$ and $a$ and through $p$ and $b$, etc. By the Euclidean sine rule,

$$\frac{|a-c|}{\sin\angle apc} = \frac{|p-c|}{\sin\angle pac}, \quad \frac{|a-d|}{\sin\angle apd} = \frac{|p-d|}{\sin\angle pad},$$
$$\frac{|b-c|}{\sin\angle bpc} = \frac{|p-c|}{\sin\angle pbc}, \quad \frac{|b-d|}{\sin\angle bpd} = \frac{|p-d|}{\sin\angle pbd}.$$

Hence,

$$(a,b;c,d) = \frac{\sin\angle apc \sin\angle bpd}{\sin\angle apd \sin\angle bpc}.$$

But obviously $\angle a'p'c' = \angle apc$, etc., hence the result. $\square$

In particular, the cross-ratio $(a,b;c,d)$, with $a,b,c,d \in \mathcal{D}_K$, is unchanged if $a,b,c$, and $d$ are subjected to any perspectivity. Note that the cross-ratio is also unchanged if $a,b,c$, and $d$ are subjected to any rotation about the origin, since this amounts to multiplying each of $a,b,c$, and $d$ by a non-zero complex number.

## Theorem 11.5.8

Let $a,b \in \mathcal{D}_K$ and let $c,d$ be the points of intersection of the line through $a,b$ with $\mathcal{C}$. Then, the Beltrami-Klein distance between $a$ and $b$ is

$$d_{\mathcal{D}_K}(a,b) = \frac{1}{2}|\ln(a,b;c,d)|.$$

## Proof

We use a suitable isometry of $\mathcal{D}_K$ to reduce to the case in which $a$ and $b$ are real. Let $l$ be the line through $c$ and 1, and $m$ the line through $d$ and $-1$. We consider two cases, according to whether $l$ and $m$ are parallel (in the Euclidean sense) or not.

If $l$ and $m$ are parallel, the line joining $c$ and $d$ passes through the origin, and a suitable rotation about the origin will take $c$ to 1, $d$ to $-1$ and $a,b$ to points $a',b'$ on the real axis. Such a rotation is an isometry of $\mathcal{D}_K$ by the remarks following Corollary 11.5.3.

Suppose, on the other hand, that $l$ and $m$ intersect at a point $p$, say. If $|p| > 1$, the perspectivity $\Pi_p$ takes $c$ to 1, $d$ to $-1$ and $a,b$ to points $a',b'$ on the line joining $-1$ and 1, i.e., the real axis. If $|p| < 1$, the lines $l'$ joining $c$ and $-1$ and $m'$ joining $d$ and 1 intersect at a point $p'$ with $|p'| > 1$ and the perspectivity $\Pi_{p'}$ takes $c$ to $-1$, $d$ to 1 and $a,b$ to points $a',b'$ on the real axis.

We compute the distance $d_{\mathcal{D}_K}(a,b) = d_{\mathcal{D}_K}(a',b')$ by transferring to $\mathcal{D}_P$ using the isometry $\mathcal{K} : \mathcal{D}_P \to \mathcal{D}_K$, so that $d_{\mathcal{D}_K}(a',b') = d_{\mathcal{D}_P}(\mathcal{K}^{-1}(a'), \mathcal{K}^{-1}(b'))$. Using Proposition 11.3.2, this gives

$$d_{\mathcal{D}_K}(a',b') = \int_{\mathcal{K}^{-1}(a')}^{\mathcal{K}^{-1}(b')} \frac{2dv}{1-v^2} = \ln \frac{(1+\mathcal{K}^{-1}(b'))(1-\mathcal{K}^{-1}(a'))}{(1+\mathcal{K}^{-1}(a'))(1-\mathcal{K}^{-1}(a'))}. \tag{11.9}$$

Using the formula (11.8) for $\mathcal{K}$, we find that

$$\mathcal{K}^{-1}(\lambda) = \frac{1}{\lambda}(1-\sqrt{1-\lambda^2}), \quad \lambda \in \mathcal{D},$$

which implies that

$$\frac{1+\mathcal{K}^{-1}(\lambda)}{1-\mathcal{K}^{-1}(\lambda)} = \sqrt{\frac{1+\lambda}{1-\lambda}}.$$

Using this, (11.9) becomes

$$d_{\mathcal{D}_K}(a',b') = \frac{1}{2} \ln \frac{(1+b')(1-a')}{(1-b')(1+a')}. \tag{11.10}$$

On the other hand, we have seen that there is a perspectivity or a rotation about the origin that takes $(a,b,c,d)$ to $(a',b',1,-1)$ or $(a',b',-1,1)$ with $a',b' \in \mathbb{R}$, and that these transformations of $\mathcal{D}_K$ leave the cross-ratio unchanged (see the remarks following the proof of Proposition 11.5.7). In the first case,

$$(a,b;c,d) = (a',b';1,-1) = \frac{(1-a')(1+b')}{(1+a')(1-b')},$$

and in the second case,

$$(a,b;c,d) = (a',b';-1,1) = \frac{(1+a')(1-b')}{(1-a')(1+b')},$$

so in both cases

$$d_{\mathcal{D}_K}(a,b) = d_{\mathcal{D}_K}(a',b') = \frac{1}{2}|\ln(a,b;c,d)|. \qquad \square$$

## *EXERCISES*

11.5.1 Prove Eq. 11.8.

11.5.2 Extend the definition of cross-ratio in the obvious way to include the possibility that one of the points is equal to $\infty$, e.g., $(\infty, b; c, d) = (b-d)/(b-c)$. Show that, if $M : \mathbb{C}_\infty \to \mathbb{C}_\infty$ is a Möbius transformation, then

$$(M(a), M(b); M(c), M(d)) = (a,b;c,d) \text{ for all distinct points } a,b,c,d \in \mathbb{C}_\infty. \tag{11.11}$$

Show, conversely, that if $M : \mathbb{C}_\infty \to \mathbb{C}_\infty$ is a bijection satisfying this condition, then $M$ is a Möbius transformation.

11.5.3 Use the preceding exercise to show that, if $(a, b, c)$ and $(a', b', c')$ are two triples of distinct points of $\mathbb{C}_\infty$, there is a unique Möbius transformation $M$ such that $M(a) = a'$, $M(b) = b'$ and $M(c) = c'$.

11.5.4 Let $a, b \in \mathbb{C}_\infty$ and let $d$ be the spherical distance between the points of $S^2$ that correspond to $a, b$ under the stereographic projection map $\Pi$ (Example 6.3.5). Show that

$$-\tan^2 \frac{1}{2} d = \left(a, -\frac{1}{\bar{a}}; b, -\frac{1}{\bar{b}}\right).$$

11.5.5 Show that, if $\mathcal{R}$ is the reflection in a line passing through the origin, then $\mathcal{K}\mathcal{R} = \mathcal{R}\mathcal{K}$. Deduce that $\mathcal{R}$ is an isometry of $\mathcal{D}_K$.

11.5.6 Show that the isometries of $\mathcal{D}_K$ are precisely the composites of (finitely many) perspectivities and reflections in lines passing through the origin.

# 12
# Minimal surfaces

In Section 9.4 we considered the problem of finding the shortest paths between two points on a surface. We now consider the analogous problem in one higher dimension, that of finding a surface of minimal area with a fixed curve as its boundary. This is called *Plateau's Problem*. The solutions to Plateau's problem turn out to be surfaces whose mean curvature vanishes everywhere. The study of these so-called minimal surfaces was initiated by Euler and Lagrange in the mid-eighteenth century, but new examples of minimal surfaces are still being discovered.

## 12.1 Plateau's problem

In Section 9.4, we found the condition for a curve on a surface to minimize distance between its endpoints by embedding the given curve in a family of curves passing through the same two points, and studying how the length of the curve varies as the curve varies through the family. Accordingly, we shall now study a family of surface patches $\boldsymbol{\sigma}^\tau : U \to \mathbb{R}^3$, where $U$ is an open subset of $\mathbb{R}^2$ independent of $\tau$, and $\tau$ lies in some open interval $(-\delta, \delta)$, for some $\delta > 0$. Let $\boldsymbol{\sigma} = \boldsymbol{\sigma}^0$. The family is required to be *smooth*, in the sense that the map $(u, v, \tau) \mapsto \boldsymbol{\sigma}^\tau(u, v)$ from the open subset $\{(u, v, \tau) \mid (u, v) \in U,\ \tau \in (-\delta, \delta)\}$ of $\mathbb{R}^3$ to $\mathbb{R}^3$ is smooth. The *surface variation* of the family is the function $\boldsymbol{\varphi} : U \to \mathbb{R}^3$ given by

$$\boldsymbol{\varphi} = \dot{\boldsymbol{\sigma}}^\tau|_{\tau=0}\,,$$

where here and elsewhere in this section, a dot denotes $d/d\tau$.

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_12,

Let $\boldsymbol{\pi}$ be a simple closed curve that is contained, along with its interior $\mathrm{int}(\boldsymbol{\pi})$, in $U$ (see Section 3.1). Then $\boldsymbol{\pi}$ corresponds to a closed curve $\boldsymbol{\gamma}^\tau = \boldsymbol{\sigma}^\tau \circ \boldsymbol{\pi}$ in the surface patch $\boldsymbol{\sigma}^\tau$, and we define the area function $\mathcal{A}(\tau)$ to be the area of the part of $\boldsymbol{\sigma}^\tau$ inside $\boldsymbol{\gamma}^\tau$:

$$\mathcal{A}(\tau) = \int_{\mathrm{int}(\boldsymbol{\pi})} d\mathcal{A}_{\boldsymbol{\sigma}^\tau}.$$

Note that, if we are considering a family of surfaces with a *fixed* boundary curve $\boldsymbol{\gamma}$, then $\boldsymbol{\gamma}^\tau = \boldsymbol{\gamma}$ for all $\tau$, and hence $\boldsymbol{\varphi}^\tau(u, v) = \mathbf{0}$ when $(u, v)$ is a point on the curve $\boldsymbol{\pi}$.

## Theorem 12.1.1

With the above notation, assume that the surface variation $\boldsymbol{\varphi}^\tau$ vanishes along the boundary curve $\boldsymbol{\pi}$. Then,

$$\dot{\mathcal{A}}(0) = -2\int_{\mathrm{int}(\boldsymbol{\pi})} H(EG - F^2)^{1/2}\alpha\, dudv, \tag{12.1}$$

where $H$ is the mean curvature of $\boldsymbol{\sigma}$, $E, F$ and $G$ are the coefficients of its first fundamental form, and $\alpha = \boldsymbol{\varphi} \cdot \mathbf{N}$ where $\mathbf{N}$ is the standard unit normal of $\boldsymbol{\sigma}$.

We defer the proof of this theorem to the end of this section.

If $\boldsymbol{\sigma}$ has the smallest area among all surfaces with the given boundary curve $\boldsymbol{\gamma}$, then $\mathcal{A}$ must have an absolute minimum at $\tau = 0$, so $\dot{\mathcal{A}}(0) = 0$ for *all* smooth families of surfaces as above. This means that the integral in (12.1) must vanish for all smooth functions $\alpha : U \to \mathbb{R}$. As in the proof of Theorem 9.4.1, this can happen only if the term that multiplies $\alpha$ in the integrand vanishes, in other words only if $H = 0$. This suggests the following definition.

## Definition 12.1.2

A *minimal surface* is a surface whose mean curvature is zero everywhere.

Theorem 12.1.1 and the preceding discussion then give

## Corollary 12.1.3

If a surface $\mathcal{S}$ has least area among all surfaces with the same boundary curve, then $\mathcal{S}$ is a minimal surface.

Minimal surfaces have an interesting physical interpretation as the shapes taken up by soap films. A soap film has energy by virtue of surface tension, and this energy is proportional to its area. A soap film spanning a wire in the shape of a curve $\mathcal{C}$ should therefore adopt the shape of a surface of least area with boundary $\mathcal{C}$. By Corollary 12.1.3, this will be a minimal surface.

More generally, if the soap film separates two regions of different pressure, the film will adopt the shape of a surface of *constant* mean curvature. This is the case for a soap bubble, for example, for which the air pressure inside the bubble is greater than the pressure outside. To see this, we apply the principle of 'virtual work'. This tells us that, if the soap film is in equilibrium, and we imagine a ('virtual') change in the surface, the change in the energy of the film must be the same as the work done by the film against the air pressure. If $p$ is the pressure difference, the force exerted by the air on a small piece of the surface of area $\Delta\mathcal{A}$ is $p\Delta\mathcal{A}$, and so the work done when it moves a small distance $\alpha$ perpendicular to itself is $\alpha p\Delta\mathcal{A}$. On the other hand, the formula in Theorem 12.1.1 shows that the change in area of the surface is proportional to $\alpha H\Delta\mathcal{A}$ (note that $\alpha$ is the component of the variation $\boldsymbol{\varphi}$ perpendicular to the surface). So $p$ is proportional to $H$. Since the pressure difference must be the same across the whole surface, so must the mean curvature $H$. Surfaces of constant non-zero mean curvature were discussed in Section 8.5.

For the moment, we give only one example of a minimal surface; others will be given in the next section. This example already shows, however, that the converse of Corollary 12.1.3 is false.

## Example 12.1.4

The surface obtained by revolving the curve $x = \cosh z$ in the $xz$-plane around the $z$-axis is called a *catenoid* (a picture of a catenoid can be found in Exercise 5.3.1). The catenoid can be parametrized by

$$\boldsymbol{\sigma}(u,v) = (\cosh u\cos v, \cosh u\sin v, u).$$

Then,

$$\begin{gathered}
\boldsymbol{\sigma}_u = (\sinh u\cos v, \sinh u\sin v, 1), \boldsymbol{\sigma}_v = (-\cosh u\sin v, \cosh u\cos v, 0),\\
\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (-\cosh u\cos v, -\cosh u\sin v, \sinh u\cosh u),\\
\mathbf{N} = (-\mathrm{sech}\, u\cos v, -\mathrm{sech}\, u\sin v, \mathrm{tanh} u),\\
\boldsymbol{\sigma}_{uu} = (\cosh u\cos v, \cosh u\sin v, 0),\\
\boldsymbol{\sigma}_{uv} = (-\sinh u\sin v, \sinh u\cos v, 0),\\
\boldsymbol{\sigma}_{vv} = (-\cosh u\cos v, -\cosh u\sin v, 0).
\end{gathered}$$

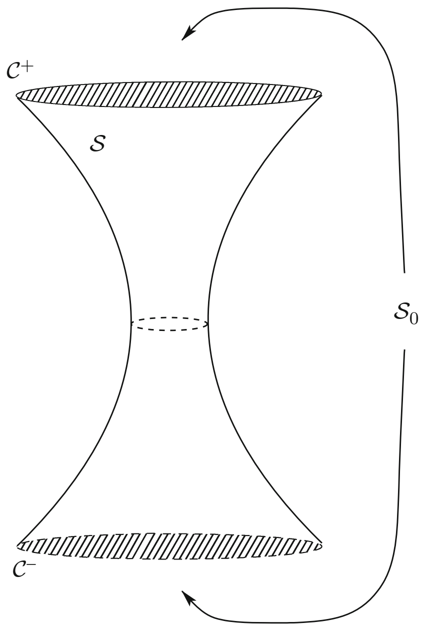
This gives the coefficients of the first and second fundamental forms of $\boldsymbol{\sigma}$ as

$$E = G = \cosh^2 u,\ F = 0, \quad L = -1,\ M = 0,\ N = 1.$$

The first three of these equations show that the parametrization $\boldsymbol{\sigma}$ is conformal, and Corollary 8.1.3 gives

$$H = \frac{LG - 2MF + NE}{2(EG - F^2)} = \frac{-\cosh^2 u + \cosh^2 u}{2\cosh^4 u} = 0,$$

showing that the catenoid is a minimal surface.

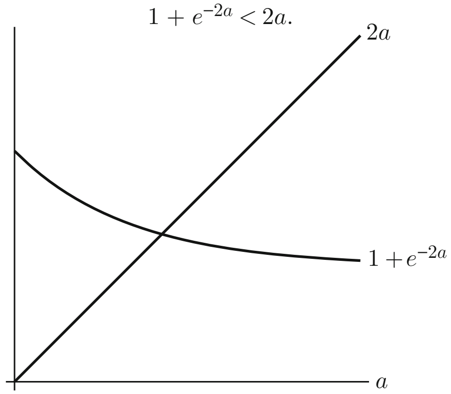



Fix $a > 0$, and let $b = \cosh a$. The surface $\mathcal{S}$ consisting of the part of the catenoid with $|z| < a$ has the two circles $\mathcal{C}^{\pm}$ of radius $b$ in the planes $z = \pm a$ with centres on the $z$-axis as boundary. Another surface spanning the same two circles is, of course, the surface $\mathcal{S}_0$ consisting of the two discs $x^2 + y^2 \leq b^2$ in the planes $z = \pm a$. The area of $\mathcal{S}$ is, by Proposition 6.4.2,

$$\int_0^{2\pi} \int_{-a}^{a} (EG - F^2)^{1/2} du dv = \int_0^{2\pi} \int_{-a}^{a} \cosh^2 u \, du dv = 2\pi(a + \sinh a \cosh a).$$

The area of $\mathcal{S}_0$ is, of course, $2\pi b^2 = 2\pi \cosh^2 a$. So the minimal surface $\mathcal{S}$ will *not* minimize the area among all surfaces with boundary the two circles $\mathcal{C}^{\pm}$ if $\cosh^2 a < a + \sinh a \cosh a$, i.e., if

$$1 + e^{-2a} < 2a. \tag{12.2}$$




The graphs of $1 + e^{-2a}$ and $2a$ as functions of $a$ clearly intersect in exactly one point $a = a_0$, say, and the inequality (12.2) holds if $a > a_0$. If this condition is satisfied, the catenoid is not area minimizing.

It can be shown that if $a < a_0$ the catenoid does have least area among all surfaces spanning the circles $\mathcal{C}^+$ and $\mathcal{C}^-$.

It is time to prove Theorem 12.1.1.

## Proof

Let $\boldsymbol{\varphi}^\tau = \dot{\boldsymbol{\sigma}}^\tau$, so that $\boldsymbol{\varphi}^0 = \boldsymbol{\varphi}$, and let $\mathbf{N}^\tau$ be the standard unit normal of $\boldsymbol{\sigma}^\tau$. There are smooth functions $\alpha^\tau, \beta^\tau$ and $\gamma^\tau$ of $(u, v, \tau)$ such that

$$\boldsymbol{\varphi}^\tau = \alpha^\tau \mathbf{N}^\tau + \beta^\tau \boldsymbol{\sigma}_u^\tau + \gamma^\tau \boldsymbol{\sigma}_v^\tau,$$

so that $\alpha = \alpha^0$. To simplify the notation, we drop the superscript $\tau$ for the rest of the proof; at the end of the proof we put $\tau = 0$.

We have

$$\mathcal{A}(\tau) = \int_{\mathrm{int}(\boldsymbol{\pi})} \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel \, dudv = \int_{\mathrm{int}(\boldsymbol{\pi})} \mathbf{N} \cdot (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)\, dudv,$$

so

$$\dot{\mathcal{A}} = \int_{\mathrm{int}(\boldsymbol{\pi})} \frac{\partial}{\partial \tau} \left(\mathbf{N} \cdot (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)\right) dudv. \tag{12.3}$$

Now,

$$\frac{\partial}{\partial \tau}(\mathbf{N} \cdot (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)) = \dot{\mathbf{N}} \cdot (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) + \mathbf{N} \cdot (\dot{\boldsymbol{\sigma}}_u \times \boldsymbol{\sigma}_v) + \mathbf{N} \cdot (\boldsymbol{\sigma}_u \times \dot{\boldsymbol{\sigma}}_v). \tag{12.4}$$

Since $\mathbf{N}$ is a unit vector,

$$\dot{\mathbf{N}} \cdot (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) = \dot{\mathbf{N}} \cdot \mathbf{N} \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel = 0.$$

On the other hand,

$$\begin{aligned}
\mathbf{N}\cdot(\dot{\boldsymbol{\sigma}}_u \times \boldsymbol{\sigma}_v) &= \frac{(\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)\cdot(\dot{\boldsymbol{\sigma}}_u \times \boldsymbol{\sigma}_v)}{\parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel} \\
&= \frac{(\boldsymbol{\sigma}_u \cdot \dot{\boldsymbol{\sigma}}_u)(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v) - (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)(\boldsymbol{\sigma}_v \cdot \dot{\boldsymbol{\sigma}}_u)}{\parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel} \\
&= \frac{G(\boldsymbol{\sigma}_u \cdot \dot{\boldsymbol{\sigma}}_u) - F(\boldsymbol{\sigma}_v \cdot \dot{\boldsymbol{\sigma}}_u)}{(EG - F^2)^{1/2}},
\end{aligned}$$

using Proposition 6.4.2. Similarly,

$$\mathbf{N}\cdot(\boldsymbol{\sigma}_u \times \dot{\boldsymbol{\sigma}}_v) = \frac{E(\boldsymbol{\sigma}_v \cdot \dot{\boldsymbol{\sigma}}_v) - F(\boldsymbol{\sigma}_u \cdot \dot{\boldsymbol{\sigma}}_v)}{(EG - F^2)^{1/2}}.$$

Substituting these results into Eq. 12.4, we get

$$\frac{\partial}{\partial \tau}(\mathbf{N}\cdot(\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)) = \frac{E(\boldsymbol{\sigma}_v \cdot \dot{\boldsymbol{\sigma}}_v) - F(\dot{\boldsymbol{\sigma}}_u \cdot \boldsymbol{\sigma}_v + \boldsymbol{\sigma}_u \cdot \dot{\boldsymbol{\sigma}}_v) + G(\boldsymbol{\sigma}_u \cdot \dot{\boldsymbol{\sigma}}_u)}{(EG - F^2)^{1/2}}. \quad (12.5)$$

Now

$$\begin{aligned}
\dot{\boldsymbol{\sigma}}_u &= \boldsymbol{\varphi}_u = \alpha_u \mathbf{N} + \beta_u \boldsymbol{\sigma}_u + \gamma_u \boldsymbol{\sigma}_v + \alpha \mathbf{N}_u + \beta \boldsymbol{\sigma}_{uu} + \gamma \boldsymbol{\sigma}_{uv}, \\
\therefore \quad \boldsymbol{\sigma}_u \cdot \dot{\boldsymbol{\sigma}}_u &= E\beta_u + F\gamma_u + (\boldsymbol{\sigma}_u \cdot \mathbf{N}_u)\alpha + (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu})\beta + (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uv})\gamma.
\end{aligned}$$

Since $\boldsymbol{\sigma}_u \cdot \mathbf{N}_u = -\boldsymbol{\sigma}_{uu} \cdot \mathbf{N} = -L$, $\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uu} = \frac{1}{2}E_u$ and $\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uv} = \frac{1}{2}E_v$, we get

$$\boldsymbol{\sigma}_u \cdot \dot{\boldsymbol{\sigma}}_u = E\beta_u + F\gamma_u - L\alpha + \frac{1}{2}E_u\beta + \frac{1}{2}E_v\gamma.$$

Similarly,

$$\begin{aligned}
\boldsymbol{\sigma}_v \cdot \dot{\boldsymbol{\sigma}}_u &= F\beta_u + G\gamma_u - M\alpha + (F_u - \frac{1}{2}E_v)\beta + \frac{1}{2}G_u\gamma, \\
\boldsymbol{\sigma}_u \cdot \dot{\boldsymbol{\sigma}}_v &= E\beta_v + F\gamma_v - M\alpha + \frac{1}{2}E_v\beta + (F_v - \frac{1}{2}G_u)\gamma, \\
\boldsymbol{\sigma}_v \cdot \dot{\boldsymbol{\sigma}}_v &= F\beta_v + G\gamma_v - N\alpha + \frac{1}{2}G_u\beta + \frac{1}{2}G_v\gamma.
\end{aligned}$$

Substituting these last four equations into the right-hand side of Eq. 12.5, simplifying, and using the formula for $H$ in Corollary 8.1.3, we find that

$$\begin{aligned}
\frac{\partial}{\partial \tau}(\mathbf{N}\cdot(\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)) = \Big(\beta(EG - F^2)^{1/2}\Big)_u &+ \Big(\gamma(EG - F^2)^{1/2}\Big)_v \\
&-2\alpha H(EG - F^2)^{1/2}.
\end{aligned} \quad (12.6)$$

Comparing with Eq. 12.3, and reinstating the superscripts, we see that we must prove that

$$\int_{\mathrm{int}(\boldsymbol{\pi})} \left\{ \Big(\beta^0(EG - F^2)^{1/2}\Big)_u + \Big(\gamma^0(EG - F^2)^{1/2}\Big)_v \right\} du dv = 0. \quad (12.7)$$

But by Green's theorem (see Section 3.2), this integral is equal to

$$\int_{\boldsymbol{\pi}} (EG - F^2)^{1/2}(\beta^0 dv - \gamma^0 du),$$

and this obviously vanishes because $\beta^0 = \gamma^0 = 0$ along the boundary curve $\boldsymbol{\pi}$. This completes the proof of Theorem 12.1.1. □

Note that we did not quite use the full force of the assumptions in Theorem 12.1.1, since they imply that $\alpha^0$ $(= \alpha)$ vanishes along the boundary curve, and this was not used in the proof. So Eq. 12.1 holds provided the surface variation $\boldsymbol{\varphi}$ is normal to the surface along the boundary curve.

Note also that Theorem 12.1.1 is intuitively obvious for variations $\boldsymbol{\varphi}$ that are parallel to the surface, i.e., those for which $\alpha = 0$ everywhere on the surface, since such a parallel variation causes the surface to slide along itself and will not change the shape, and in particular the area, of the surface. Thus, the main point is to prove Theorem 12.1.1 for normal variations, i.e., those for which $\beta = \gamma = 0$ everywhere on the surface. Making this restriction simplifies the above proof considerably.

## *EXERCISES*

12.1.1 Show that the Gaussian curvature of a minimal surface is $\leq 0$ everywhere, and that it is zero everywhere if and only if the surface is an open subset of a plane. We shall obtain a much more precise result in Corollary 12.5.6.

12.1.2 Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a minimal surface patch, and assume that $\mathcal{A}_{\boldsymbol{\sigma}}(\mathcal{U}) < \infty$ (see Definition 6.4.1). Let $\lambda \neq 0$ and assume that the principal curvatures $\kappa$ of $\boldsymbol{\sigma}$ satisfy $|\lambda\kappa| < 1$ everywhere, so that the parallel surface $\boldsymbol{\sigma}^\lambda$ of $\boldsymbol{\sigma}$ (Definition 8.5.1) is a regular surface patch. Prove that

$$\mathcal{A}_{\boldsymbol{\sigma}^\lambda}(\mathcal{U}) \leq \mathcal{A}_{\boldsymbol{\sigma}}(\mathcal{U})$$

and that equality holds for some $\lambda \neq 0$ if and only if $\boldsymbol{\sigma}(U)$ is an open subset of a plane. (Thus, any minimal surface is area-minimizing among its family of parallel surfaces.)

12.1.3 Show that there is no compact minimal surface.

12.1.4 Show that applying a dilation or an isometry of $\mathbb{R}^3$ to a minimal surface gives another minimal surface. Can there be a *local* isometry between a minimal surface and a non-minimal surface?

## 12.2 Examples of minimal surfaces

The simplest minimal surface is, of course, the plane, for which both principal curvatures are zero everywhere. Apart from this, the first minimal surfaces to be discovered were those in the following two examples.

### Example 12.2.1

If $a$ is a non-zero constant, the surface $\mathcal{S}_a$ obtained by rotating the curve $x = \frac{1}{a}\cosh az$ in the $xz$-plane around the $z$-axis is called a *catenoid*. More generally, a catenoid is a surface obtained by applying an isometry of $\mathbb{R}^3$ to a surface $\mathcal{S}_a$. Any catenoid $\mathcal{S}$ is a minimal surface, since $\mathcal{S}$ can be obtained from the special catenoid $\mathcal{S}_1$ in Example 12.1.4 by applying an isometry and a dilation (Exercise 12.1.4). A picture of a catenoid can be found in Exercise 5.3.1.

Catenoids are surfaces of revolution. In fact, apart from the plane, they are the only minimal surfaces of revolution:

### Proposition 12.2.2

Any minimal surface of revolution $\mathcal{S}$ is an open subset of a plane or a catenoid.

### Proof

By applying an isometry of $\mathbb{R}^3$, we can assume that the axis of the surface $\mathcal{S}$ is the $z$-axis and the profile curve lies in the $xz$-plane. We parametrize $\mathcal{S}$ in the usual way (see Example 5.3.2):

$$\boldsymbol{\sigma}(u, v) = (f(u)\cos v, f(u)\sin v, g(u)),$$

where the profile curve $u \mapsto (f(u), 0, g(u))$ is assumed to be unit-speed and $f > 0$. From Examples 6.1.3 and 7.1.2, the first and second fundamental forms are

$$du^2 + f(u)^2 dv^2 \quad \text{and} \quad (\dot{f}\ddot{g} - \ddot{f}\dot{g})du^2 + f\dot{g}dv^2,$$

respectively, a dot denoting $d/du$. By Corollary 8.1.3, the mean curvature is

$$H = \frac{1}{2}\left(\dot{f}\ddot{g} - \ddot{f}\dot{g} + \frac{\dot{g}}{f}\right).$$

We suppose now that, for some value of $u$, say $u = u_0$, we have $\dot{g}(u_0) \neq 0$. Since $\dot{g}$ is continuous (because $g$ is smooth), we shall then have $\dot{g}(u) \neq 0$ for $u$

in some open interval containing $u_0$. Let $(\alpha, \beta)$ be the largest such interval. Supposing now that $u \in (\alpha, \beta)$, the unit-speed condition $\dot{f}^2 + \dot{g}^2 = 1$ gives (as in Example 8.1.4)

$$\dot{f}\ddot{g} - \ddot{f}\dot{g} = -\frac{\ddot{f}}{\dot{g}},$$

and so we get

$$H = \frac{1}{2}\left(\frac{\dot{g}}{f} - \frac{\ddot{f}}{\dot{g}}\right).$$

Since $\dot{g}^2 = 1 - \dot{f}^2$, $\mathcal{S}$ is minimal if and only if

$$f\ddot{f} = 1 - \dot{f}^2. \tag{12.8}$$

To solve the differential equation (12.8), put $h = \dot{f}$, and note that

$$\ddot{f} = \frac{dh}{dt} = \frac{dh}{df}\frac{df}{dt} = h\frac{dh}{df}.$$

Hence, Eq. 12.8 becomes

$$fh\frac{dh}{df} = 1 - h^2.$$

Note that, since $\dot{g} \neq 0$, we have $h^2 \neq 1$, and so we can integrate this equation as follows:

$$\int \frac{hdh}{1-h^2} = \int \frac{df}{f},$$

$$\therefore \quad h = \frac{\sqrt{a^2f^2 - 1}}{af},$$

where $a$ is a non-zero constant. (We have omitted a $\pm$, but the sign can be changed by replacing $u$ by $-u$ if necessary.) Writing $h = df/du$ and integrating again,

$$\int \frac{af\,df}{\sqrt{a^2f^2-1}} = \int du,$$

$$\therefore \quad f = \frac{1}{a}\sqrt{1 + a^2(u+b)^2},$$

where $b$ is a constant. By a change of parameter $u \mapsto u + b$, we can assume that $b = 0$. So,

$$f = \frac{1}{a}\sqrt{1 + a^2u^2}.$$

To compute $g$, we have

$$\begin{aligned}
\dot{g}^2 &= 1 - \dot{f}^2 = 1 - h^2 = \frac{1}{a^2 f^2}, \\
\therefore \quad \frac{dg}{du} &= \pm \frac{1}{\sqrt{1 + a^2 u^2}}, \\
\therefore \quad g &= \pm \frac{1}{a} \sinh^{-1}(au) + c \qquad \text{(where } c \text{ is a constant)}, \\
\therefore \quad au &= \pm \sinh(a(g - c)), \\
\therefore \quad f &= \frac{1}{a} \cosh(a(g - c)).
\end{aligned}$$

Thus, the profile curve of $\mathcal{S}$ is

$$x = \frac{1}{a} \cosh(a(z - c)).$$

This surface is obtained by applying to the catenoid $\mathcal{S}_a$ a translation along the $z$-axis.

We are not quite finished, however. So far, we have only shown that the open subset of $\mathcal{S}$ corresponding to $u \in (\alpha, \beta)$ is part of the catenoid, for in the proof we used in an essential way that $\dot{g} \neq 0$. This is why the proof has so far excluded the possibility that $\mathcal{S}$ is a plane. To complete the proof, we argue as follows. Suppose that $\beta < \infty$. Then, if the profile curve is defined for values of $u \geq \beta$, we must have $\dot{g}(\beta) = 0$, for otherwise $\dot{g}$ would be non-zero on an open interval containing $\beta$, which would contradict our assumption that $(\alpha, \beta)$ is the *largest* open interval containing $u_0$ on which $\dot{g} \neq 0$. But the formulas above show that

$$\dot{g}^2 = \frac{1}{1 + a^2 u^2} \quad \text{if } u \in (\alpha, \beta),$$

and so, since $\dot{g}$ is a continuous function of $u$, $\dot{g}(\beta) = \pm(1 + a^2\beta^2)^{-1/2} \neq 0$. This contradiction shows that the profile curve is not defined for values of $u \geq \beta$. Of course, this also holds trivially if $\beta = \infty$. A similar argument applies to $\alpha$, and shows that $(\alpha, \beta)$ is the entire domain of definition of the profile curve. Hence, the whole of $\mathcal{S}$ is an open subset of a catenoid.

The only remaining case to consider is that in which $\dot{g}(u) = 0$ for all values of $u$ for which the profile curve is defined. But then $g(u)$ is a constant, say $d$, and $\mathcal{S}$ is an open subset of the plane $z = d$. $\square$

## Example 12.2.3

A *helicoid* is a ruled surface swept out by a straight line that rotates at constant speed about an axis perpendicular to the line while simultaneously moving at constant speed along the axis. By applying an isometry of $\mathbb{R}^3$ we can take the axis to be the $z$-axis. Let $\omega$ be the angular velocity of the rotating line and $\alpha$ its speed along the $z$-axis. If the line starts along the $x$-axis, at time $v$ the centre of the line is at $(0,0,\alpha v)$ and it has rotated by an angle $\omega v$. Hence, the point of the line initially at $(u,0,0)$ is now at the point

$$\boldsymbol{\sigma}(u,v) = (u\cos\omega v, u\sin\omega v, \alpha v).$$

We leave it to Exercise 12.2.1 to check that this is a minimal surface. (A picture of a helicoid can be found in Exercise 4.2.6.)

We have the following analogue of Proposition 12.2.2.

## Proposition 12.2.4

Any ruled minimal surface is an open subset of a plane or a helicoid.

## Proof

We take the usual parametrization

$$\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\boldsymbol{\delta}(u)$$

(see Example 5.3.3), where $\boldsymbol{\gamma}$ is a curve that meets each of the rulings and $\boldsymbol{\delta}(u)$ is a vector parallel to the ruling through $\boldsymbol{\gamma}(u)$. We begin the proof by making some simplifications to the parametrization.

First, we can certainly assume that $\parallel \boldsymbol{\delta}(u) \parallel = 1$ for all values of $u$. We assume also that $\dot{\boldsymbol{\delta}}$ *is never zero*, where the dot denotes $d/du$. (We shall consider later what happens if $\dot{\boldsymbol{\delta}}(u) = \mathbf{0}$ for some values of $u$.) We can then assume that $\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\delta}} = 0$ (see Exercise 5.3.4).

We have $\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}$, $\boldsymbol{\sigma}_v = \boldsymbol{\delta}$, so

$$E = \parallel \dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}} \parallel^2, \quad F = (\dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}) \cdot \boldsymbol{\delta} = \dot{\boldsymbol{\gamma}} \cdot \boldsymbol{\delta}, \quad G = 1.$$

Let $A = \sqrt{EG - F^2}$. Then,

$$\mathbf{N} = A^{-1}(\dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}) \times \boldsymbol{\delta}.$$

Next, we have $\boldsymbol{\sigma}_{uu} = \ddot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\delta}}$, $\boldsymbol{\sigma}_{uv} = \dot{\boldsymbol{\delta}}$, $\boldsymbol{\sigma}_{vv} = \mathbf{0}$, so

$$\begin{aligned} L &= A^{-1}(\ddot{\boldsymbol{\gamma}} + v\ddot{\boldsymbol{\delta}}) \cdot ((\dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}) \times \boldsymbol{\delta}), \\ M &= A^{-1}\dot{\boldsymbol{\delta}} \cdot ((\dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}) \times \boldsymbol{\delta}) = A^{-1}\dot{\boldsymbol{\delta}}.(\dot{\boldsymbol{\gamma}} \times \boldsymbol{\delta}), \\ N &= 0. \end{aligned}$$

Hence, the minimal surface condition

$$H = \frac{LG - 2MF + NE}{2A^2} = 0$$

gives

$$(\ddot{\gamma} + v\ddot{\boldsymbol{\delta}}) \cdot ((\dot{\gamma} + v\dot{\boldsymbol{\delta}}) \times \boldsymbol{\delta}) = 2(\boldsymbol{\delta} \cdot \dot{\gamma})(\dot{\boldsymbol{\delta}} \cdot (\dot{\gamma} \times \boldsymbol{\delta})).$$

This equation must hold for all values of $(u, v)$. Equating coefficients of powers of $v$ gives

$$\ddot{\gamma} \cdot (\dot{\gamma} \times \boldsymbol{\delta}) = 2(\boldsymbol{\delta} \cdot \dot{\gamma})(\dot{\boldsymbol{\delta}} \cdot (\dot{\gamma} \times \boldsymbol{\delta})), \quad (12.9)$$

$$\ddot{\gamma} \cdot (\dot{\boldsymbol{\delta}} \times \boldsymbol{\delta}) + \ddot{\boldsymbol{\delta}} \cdot (\dot{\gamma} \times \boldsymbol{\delta}) = 0, \quad (12.10)$$

$$\ddot{\boldsymbol{\delta}} \cdot (\dot{\boldsymbol{\delta}} \times \boldsymbol{\delta}) = 0. \quad (12.11)$$

Equation 12.11 shows that $\boldsymbol{\delta}, \dot{\boldsymbol{\delta}}$ and $\ddot{\boldsymbol{\delta}}$ are linearly dependent. Since $\boldsymbol{\delta}$ and $\dot{\boldsymbol{\delta}}$ are perpendicular unit vectors, there are smooth functions $\alpha(u)$ and $\beta(u)$ such that

$$\ddot{\boldsymbol{\delta}} = \alpha\boldsymbol{\delta} + \beta\dot{\boldsymbol{\delta}}.$$

But, since $\boldsymbol{\delta}$ is unit-speed, $\dot{\boldsymbol{\delta}} \cdot \ddot{\boldsymbol{\delta}} = 0$. Also, differentiating $\boldsymbol{\delta} \cdot \dot{\boldsymbol{\delta}} = 0$ gives $\boldsymbol{\delta} \cdot \ddot{\boldsymbol{\delta}} = -\dot{\boldsymbol{\delta}} \cdot \dot{\boldsymbol{\delta}} = -1$. Hence, $\alpha = -1$ and $\beta = 0$, so

$$\ddot{\boldsymbol{\delta}} = -\boldsymbol{\delta}. \quad (12.12)$$

Equation 12.12 shows that the curvature of the curve $\boldsymbol{\delta}$ is 1, and that its principal normal is $-\boldsymbol{\delta}$. Hence, its binormal is $\dot{\boldsymbol{\delta}} \times (-\boldsymbol{\delta})$, and since

$$\frac{d}{du}(\dot{\boldsymbol{\delta}} \times \boldsymbol{\delta}) = \ddot{\boldsymbol{\delta}} \times \boldsymbol{\delta} + \dot{\boldsymbol{\delta}} \times \dot{\boldsymbol{\delta}} = -\boldsymbol{\delta} \times \boldsymbol{\delta} = \mathbf{0},$$

it follows that the torsion of $\boldsymbol{\delta}$ is zero. Hence, $\boldsymbol{\delta}$ parametrizes a circle of radius 1 (see Proposition 2.3.5). By applying an isometry of $\mathbb{R}^3$, we can assume that $\boldsymbol{\delta}$ is the circle with radius 1 and centre the origin in the $xy$-plane, so that

$$\boldsymbol{\delta}(u) = (\cos u, \sin u, 0).$$

From Eq. 12.12, we get $\ddot{\boldsymbol{\delta}} \cdot (\dot{\gamma} \times \boldsymbol{\delta}) = -\boldsymbol{\delta} \cdot (\dot{\gamma} \times \boldsymbol{\delta}) = 0$, so by Eq. 12.10,

$$\ddot{\gamma} \cdot (\dot{\boldsymbol{\delta}} \times \boldsymbol{\delta}) = 0.$$

It follows that $\ddot{\gamma}$ is parallel to the $xy$-plane, and hence that

$$\boldsymbol{\gamma}(u) = (f(u), g(u), au + b),$$

where $f$ and $g$ are smooth functions and $a$ and $b$ are constants. If $a = 0$, the surface is an open subset of the plane $z = b$. Otherwise, Eq. 12.9 gives

$$\ddot{g}\cos u - \ddot{f}\sin u = 2(\dot{f}\cos u + \dot{g}\sin u). \quad (12.13)$$

We finally make use of the condition $\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\delta}} = 0$, which gives

$$\dot{f} \sin u = \dot{g} \cos u. \tag{12.14}$$

Differentiating this gives

$$\ddot{f} \sin u + \dot{f} \cos u = \ddot{g} \cos u - \dot{g} \sin u. \tag{12.15}$$

Equations 12.13 and 12.15 together give

$$\dot{f} \cos u + \dot{g} \sin u = 0$$

and using Eq. 12.14 we get $\dot{f} = \dot{g} = 0$. Thus, $f$ and $g$ are constants. By a translation of the surface, we can assume that the constants $f$, $g$ and $b$ are zero, so that $\boldsymbol{\gamma}(u) = (0, 0, au)$ and

$$\boldsymbol{\sigma}(u, v) = (v \cos u, v \sin u, au),$$

which is a helicoid.

We assumed at the beginning that $\dot{\boldsymbol{\delta}}$ is never zero. If $\dot{\boldsymbol{\delta}}$ is *always* zero, then $\boldsymbol{\delta}$ is a constant vector and the surface is a generalized cylinder. But in fact a generalized cylinder is a minimal surface only if the cylinder is an open subset of a plane (Exercise 12.2.3). The proof is now completed by an argument similar to that used at the end of the proof of Proposition 12.2.2, which shows that the whole surface is an open subset of a plane or a helicoid. □

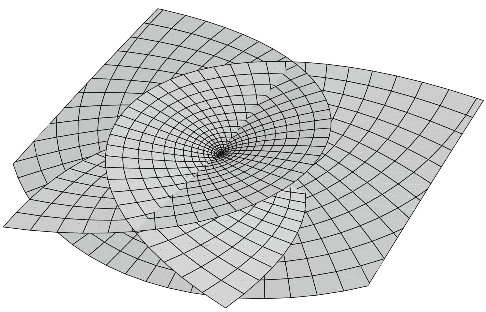
After the catenoid and helicoid, the next minimal surfaces to be discovered were the following two.

## Example 12.2.5

*Enneper's surface* is

$$\boldsymbol{\sigma}(u, v) = \left( u - \frac{u^3}{3} + uv^2, v - \frac{v^3}{3} + u^2 v, u^2 - v^2 \right).$$

It was shown in Exercise 8.5.1 that this is a minimal surface.



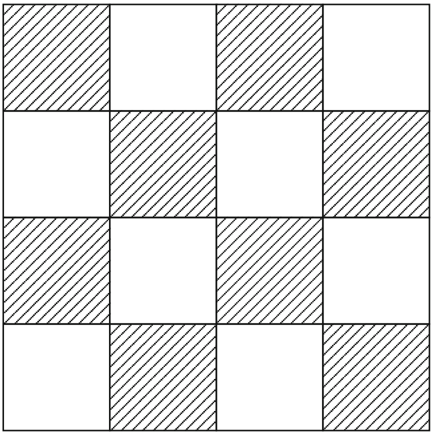
Strictly speaking, this is not a surface patch as it is not injective. The self-intersections are clearly visible in the picture above. However, if we restrict $(u, v)$ to lie in sufficiently small open sets, $\boldsymbol{\sigma}$ will be injective (see Exercise 5.6.3).

### Example 12.2.6

*Scherk's surface* is the surface with Cartesian equation

$$z = \ln\left(\frac{\cos y}{\cos x}\right).$$

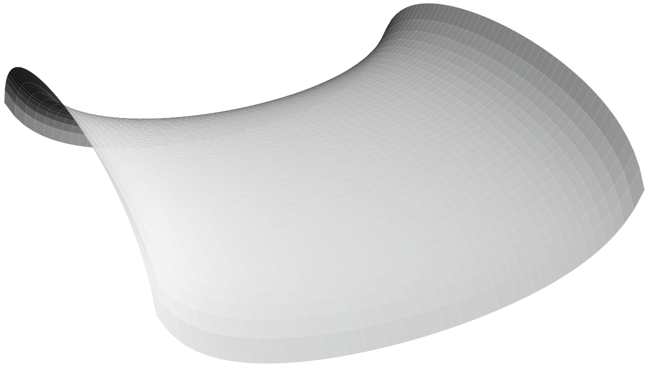
It was shown in Exercise 8.5.2 that this is a minimal surface. Note that the surface exists only when $\cos x$ and $\cos y$ are both $> 0$ or both $< 0$, in other words in the interiors of the white squares of the following chess board pattern, in which the squares have vertices at the points $(\pi/2 + m\pi, \pi/2 + n\pi)$, where $m$ and $n$ are integers, no two squares with a common edge have the same colour, and the square containing the origin is white:



The white squares have centres of the form $(m\pi, n\pi)$, where $m$ and $n$ are integers with $m + n$ even. Since, for such $m, n$,

$$\frac{\cos(y + n\pi)}{\cos(x + m\pi)} = \frac{\cos y}{\cos x},$$

it follows that the part of the surface over the square with centre $(m\pi, n\pi)$ is obtained from the part over the square with centre $(0, 0)$ by the translation $(x, y, z) \mapsto (x + m\pi, y + n\pi, z)$. So it suffices to exhibit the part of the surface over a single square (see below).



## *EXERCISES*

12.2.1 Show that every helicoid is a minimal surface.

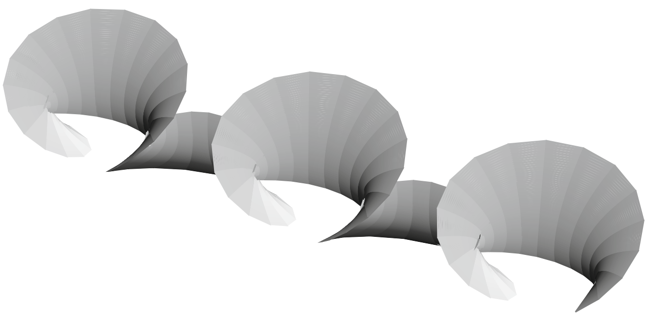
12.2.2 Show that the surfaces $\boldsymbol{\sigma}^t$ in the isometric deformation of a helicoid into a catenoid given in Exercise 6.2.2 are minimal surfaces. (This is 'explained' in Exercise 12.5.4.)

12.2.3 Show that a generalized cylinder is a minimal surface only when the cylinder is an open subset of a plane.

12.2.4 Verify that *Catalan's surface*

$$\boldsymbol{\sigma}(u,v) = \left(u - \sin u \cosh v, 1 - \cos u \cosh v, -4 \sin \frac{u}{2} \sinh \frac{v}{2}\right)$$

is a conformally parametrized minimal surface. (As in the case of Enneper's surface, Catalan's surface has self-intersections, so it is only a surface if we restrict $(u, v)$ to sufficiently small open sets.)



Show that:

(i) The parameter curve on the surface given by $u = 0$ is a straight line.

(ii) The parameter curve $u = \pi$ is a parabola.

(iii) The parameter curve $v = 0$ is a cycloid (see Exercise 1.1.7).

Show also that each of these curves, when suitably parametrized, is a geodesic on Catalan's surface. (There is a sense in which Catalan's surface is 'designed' to have a cycloidal geodesic – see Exercise 12.5.5.)

## 12.3 Gauss map of a minimal surface

Recall from Section 7.2 that the Gauss map $\mathcal{G}$ of an oriented surface $\mathcal{S}$ associates to each point $\mathbf{p} \in \mathcal{S}$ the unit normal $\mathbf{N}_{\mathbf{p}}$ of $\mathcal{S}$ at $\mathbf{p}$ regarded as a point of the unit sphere $S^2$. We begin with the following 'local' result:

### Proposition 12.3.1

With the above notation, suppose that the Gaussian curvature of $\mathcal{S}$ is non-zero at the point $\mathbf{p}$. Then, there is an open subset $V$ of $\mathcal{S}$ containing $\mathbf{p}$ such that the restriction of $\mathcal{G}$ to $V$ is injective.

This result (and its proof) implies that, if the Gaussian curvature of $\mathcal{S}$ is nowhere zero, the Gauss map of $\mathcal{S}$ is a *local diffeomorphism*.

### Proof

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a surface patch of $\mathcal{S}$ containing $\mathbf{p}$, say $\mathbf{p} = \boldsymbol{\sigma}(u_0, v_0)$, and let $\mathbf{N} : U \to \mathbb{R}^3$ be the standard unit normal of $\boldsymbol{\sigma}$. By Eq. 8.2,

$$\mathbf{N}_u \times \mathbf{N}_v = K\, \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v,$$

where $K$ is the Gaussian curvature of $\mathcal{S}$, so by Exercise 5.6.3 there is an open subset $W$ of $U$ containing $(u_0, v_0)$ such that the restriction of the map $\mathbf{N}$ to $W$ is injective. Then, $\boldsymbol{\sigma}(W)$ is an open subset of $\mathcal{S}$ containing $\mathbf{p}$ and the restriction of $\mathcal{G}$ to $\boldsymbol{\sigma}(W)$ is injective. $\square$

### Theorem 12.3.2

Let $\mathcal{S}$ be a minimal surface with nowhere vanishing Gaussian curvature. Then, the Gauss map is a conformal map from $\mathcal{S}$ to $S^2$.

### Proof

By Theorem 6.3.3, we have to show that the bilinear forms $\langle\ ,\ \rangle$ and $\mathcal{G}^*\langle\ ,\ \rangle$ are proportional. Now, if $\mathbf{p} \in \mathcal{S}$ and $\mathbf{v}, \mathbf{w} \in T_\mathbf{p}\mathcal{S}$,

$$\mathcal{G}^*\langle \mathbf{v}, \mathbf{w}\rangle = \langle \mathcal{D}_\mathbf{p}\mathcal{G}(\mathbf{v}), \mathcal{D}_\mathbf{p}\mathcal{G}(\mathbf{w})\rangle = \langle -\mathcal{W}(\mathbf{v}), -\mathcal{W}(\mathbf{w})\rangle = \langle \mathcal{W}^2(\mathbf{v}), \mathbf{w}\rangle,$$

where $\mathcal{W}$ is the Weingarten map; the last equation follows from the fact that $\mathcal{W}$ is self-adjoint (Corollary 7.2.4). But, by Exercise 8.1.6 and the fact that the mean curvature $H$ is zero, we have

$$\mathcal{W}^2 = -K,$$

the Gaussian curvature of $\mathcal{S}$. It follows that

$$\mathcal{G}^*\langle\ ,\ \rangle = -K\langle\ ,\ \rangle,$$

as we want. □

We saw in Exercise 6.3.4 that a conformal parametrization of the plane is necessarily holomorphic or anti-holomorphic, so this proposition strongly suggests a connection between minimal surfaces and holomorphic functions. This connection turns out to be very extensive, and we shall give an introduction to it in Section 12.5.

## *EXERCISES*

12.3.1 Let $\mathcal{S}$ be a connected surface whose Gauss map is conformal.

(i) Show that, if $\mathbf{p} \in \mathcal{S}$ and if the mean curvature $H$ of $\mathcal{S}$ at $\mathbf{p}$ is non-zero, there is an open subset of $\mathcal{S}$ containing $\mathbf{p}$ that is part of a sphere.

(ii) Deduce that, if $H$ is non-zero at $\mathbf{p}$, there is an open subset of $\mathcal{S}$ containing $\mathbf{p}$ on which $H$ is constant.

(iii) Deduce that $\mathcal{S}$ is either a minimal surface or an open subset of a sphere.

12.3.2 Show that:

(i) The Gauss map of a catenoid is injective and its image is the whole of $S^2$ except for the north and south poles.

(ii) The image of the Gauss map of a helicoid is the same as that of a catenoid, but that infinitely many points on the helicoid are sent by the Gauss map to any given point in its image.

(The fact that the Gauss maps of a catenoid and a helicoid have the same image is 'explained' in Exercise 12.5.3 (ii).)

## 12.4 Conformal parametrization of minimal surfaces

Our goal in this section is to prove the following theorem.

### Theorem 12.4.1

Let $\mathcal{S}$ be a minimal surface and let $\mathbf{p} \in \mathcal{S}$. Then, there is a surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$ containing $\mathbf{p}$ that is conformal.

Recall from Section 6.3 that this means that the first fundamental form of $\boldsymbol{\sigma}(u, v)$ is of the form $E(du^2 + dv^2)$ for some smooth function $E(u, v)$.

### Proof

Let $\mathbf{p} = (x_0, y_0, z_0)$. By Exercise 5.6.4, if the tangent plane of $\mathcal{S}$ at $\mathbf{p}$ does not contain the $z$-axis, there is an open set $U$ in $\mathbb{R}^2$ containing $(x_0, y_0)$ and a smooth function $f : V \to \mathbb{R}$ such that an open subset of $\mathcal{S}$ consisting of the points $(x, y, z)$ with $(x, y) \in V$ coincides with the graph of the function $f$. (If the tangent plane at $\mathbf{p}$ does contain the $z$-axis, then $\mathcal{S}$ will be a graph of the form $x = f(y, z)$ or $y = f(x, z)$ near $\mathbf{p}$.) We can also assume that $V$ is an open disc

$$D = \{(x, y) \mid (x - x_0)^2 + (y - y_0)^2 < r^2\},$$

for some $r > 0$, since any open set in $\mathbb{R}^2$ containing $(x_0, y_0)$ contains such a disc. We must therefore show that the surface patch

$$\tilde{\boldsymbol{\sigma}}(x, y) = (x, y, f(x, y)), \quad (x, y) \in D,$$

has a conformal reparametrization.

The coefficients of the first fundamental form of $\tilde{\boldsymbol{\sigma}}$ are

$$E = 1 + f_x^2, \quad F = f_x f_y, \quad G = 1 + f_y^2.$$

We show first that

$$\left(\frac{F}{A}\right)_x = \left(\frac{E}{A}\right)_y, \quad \left(\frac{G}{A}\right)_x = \left(\frac{F}{A}\right)_y, \tag{12.16}$$

where $A = \sqrt{EG - F^2}$. Indeed,

$$\begin{aligned}
\left(\frac{F}{A}\right)_x - \left(\frac{E}{A}\right)_y &= \frac{(1 + f_x^2 + f_y^2)(f_{xx}f_y + f_x f_{xy}) - f_x f_y(f_x f_{xx} + f_y f_{xy})}{(1 + f_x^2 + f_y^2)^{3/2}} \\
&\quad - \frac{2(1 + f_x^2 + f_y^2) f_x f_{xy} - (1 + f_x^2)(f_x f_{xy} + f_y f_{yy})}{(1 + f_x^2 + f_y^2)^{3/2}} \\
&= \frac{f_y((1 + f_y^2) f_{xx} - 2 f_x f_y f_{xy} + (1 + f_x^2) f_{yy})}{(1 + f_x^2 + f_y^2)^{3/2}} \\
&= 0,
\end{aligned}$$

by Exercise 8.1.1. The second equation in (12.16) is proved similarly.

From advanced calculus, we know that Eqs. 12.16 imply the existence of smooth functions $\varphi, \psi : D \to \mathbb{R}$ such that

$$\varphi_x = \frac{E}{A}, \quad \varphi_y = \frac{F}{A}, \quad \psi_x = \frac{F}{A}, \quad \psi_y = \frac{G}{A}.$$

In fact, we can just define

$$\varphi(x, y) = \int_0^1 \frac{xE((1-t)\mathbf{r}_0 + t\mathbf{r}) + yF((1-t)\mathbf{r}_0 + t\mathbf{r})}{A((1-t)\mathbf{r}_0 + t\mathbf{r})} \, dt,$$

where $\mathbf{r} = (x, y)$, $\mathbf{r}_0 = (x_0, y_0)$; and similarly for $\psi$.

The reparametrization map we want is

$$u(x, y) = x + \varphi(x, y), \quad v(x, y) = y + \psi(x, y).$$

Note that

$$\begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} = \begin{pmatrix} 1 + \varphi_x & \varphi_y \\ \psi_x & 1 + \psi_y \end{pmatrix} = \begin{pmatrix} 1 + \frac{E}{A} & \frac{F}{A} \\ \frac{F}{A} & 1 + \frac{G}{A} \end{pmatrix} \tag{12.17}$$

so

$$\begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = \left(1 + \frac{E}{A}\right)\left(1 + \frac{G}{A}\right) - \frac{F^2}{A^2} = 2 + \frac{E + G}{A} > 0.$$

By the Inverse Function Theorem 5.6.1, the function $F : D \to \mathbb{R}^2$ given by $F(x, y) = (u(x, y), v(x, y))$ has a smooth inverse function $F^{-1}$ (we may have to replace $D$ by a smaller open disc with centre $(x_0, y_0)$). Let

$$F^{-1}(u, v) = (x(u, v), y(u, v)).$$

We shall show that the reparametrization

$$\boldsymbol{\sigma}(u,v) = (x(u,v), y(u,v), f(x(u,v), y(u,v)))$$

of $\tilde{\boldsymbol{\sigma}}$ is conformal.

By the chain rule,

$$\begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix} \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} = I,$$

so

$$\begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix} = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}^{-1} = \frac{1}{E+G+2A} \begin{pmatrix} G+A & -F \\ -F & E+A \end{pmatrix}$$

by Eq. 12.17. Letting $z(u,v) = f(x(u,v), y(u,v))$, we get (again using the chain rule)

$$\begin{aligned} z_u &= f_x x_u + f_y y_u = \frac{f_x(G+A) - f_y F}{E+G+2A}, \\ z_v &= f_x x_v + f_y y_v = \frac{f_y(E+A) - f_x F}{E+G+2A}. \end{aligned}$$

Hence,

$$\begin{aligned} \boldsymbol{\sigma}_u.\boldsymbol{\sigma}_u &= x_u^2 + y_u^2 + z_u^2 \\ &= \frac{(G+A)^2 + F^2 + (f_x(G+A) - f_y F)^2}{(E+G+2A)^2} \\ &= \frac{(G+A)^2 + F^2 + (E-1)(G+A)^2 + (G-1)F^2 - 2(G+A)F^2}{(E+G+2A)^2} \\ &= \frac{E(G+A)^2 + GF^2 - 2(G+A)F^2}{(E+G+2A)^2} \\ &= \frac{EA^2 + 2A(EG-F^2) + G(EG-F^2)}{(E+G+2A)^2} \\ &= \frac{A^2}{E+G+2A}, \end{aligned}$$

using $f_x^2 = E-1$, $f_y^2 = G-1$ to pass from the second line to the third and $A^2 = EG - F^2$ to pass from the fifth line to the sixth. Similar calculations show that

$$\boldsymbol{\sigma}_v.\boldsymbol{\sigma}_v = \frac{A^2}{E+G+2A}, \quad \boldsymbol{\sigma}_u.\boldsymbol{\sigma}_v = 0. \qquad \square$$

## *EXERCISES*

12.4.1 Use Proposition 12.3.2 to give another proof of Theorem 12.4.1 for surfaces $\mathcal{S}$ with nowhere-vanishing Gaussian curvature.

# 12.5 Minimal surfaces and holomorphic functions

In this section, we shall make use of certain elementary properties of holomorphic functions. Readers without the necessary background in complex analysis may safely omit this section, the results of which are not used anywhere else in the book.

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a conformal surface patch. We introduce complex coordinates in the plane of which $U$ is an open subset by setting

$$\zeta = u + iv, \qquad (u, v) \in U,$$

and we define

$$\boldsymbol{\varphi}(\zeta) = \boldsymbol{\sigma}_u - i\boldsymbol{\sigma}_v. \tag{12.18}$$

Thus, $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$ has three components, each of which is a complex-valued function of $(u, v)$, i.e., of $\zeta$. The basic result which establishes the connection between minimal surfaces and holomorphic functions is the following proposition.

### Proposition 12.5.1

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a conformal surface patch. Then $\boldsymbol{\sigma}$ is minimal if and only if the function $\boldsymbol{\varphi}$ defined in Eq. 12.18 is holomorphic on $U$.

To say that $\boldsymbol{\varphi}$ is holomorphic means that each of its components $\varphi_1, \varphi_2$ and $\varphi_3$ is holomorphic.

### Proof

Let $\varphi(u, v)$ be a complex-valued smooth function, and let $\alpha$ and $\beta$ be its real and imaginary parts, so that $\varphi = \alpha + i\beta$. The Cauchy–Riemann equations

$$\alpha_u = \beta_v \quad \text{and} \quad \alpha_v = -\beta_u$$

are the necessary and sufficient conditions for $\varphi$ to be holomorphic. Applying this to each of the components of $\boldsymbol{\varphi}$, we see that $\boldsymbol{\varphi}$ is holomorphic if and only if

$$(\boldsymbol{\sigma}_u)_u = (-\boldsymbol{\sigma}_v)_v \quad \text{and} \quad (\boldsymbol{\sigma}_u)_v = -(-\boldsymbol{\sigma}_v)_u. \tag{12.19}$$

The second equation imposes no condition on $\boldsymbol{\sigma}$, and the first is equivalent to $\boldsymbol{\sigma}_{uu} + \boldsymbol{\sigma}_{vv} = \mathbf{0}$. But it was shown in Exercise 8.5.1 that a conformal surface patch $\boldsymbol{\sigma}$ is minimal if and only if $\boldsymbol{\sigma}_{uu} + \boldsymbol{\sigma}_{vv}$ is zero. □

The holomorphic function $\boldsymbol{\varphi}$ associated to a minimal surface $\boldsymbol{\sigma}$ is not arbitrary:

## Theorem 12.5.2

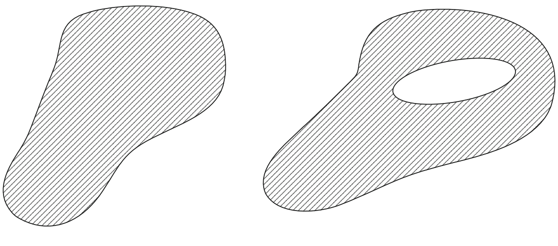
If $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a conformally parametrized minimal surface, the vector-valued holomorphic function $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$ defined in Eq. 12.18 satisfies the following conditions:

(i) $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = 0$.

(ii) $\boldsymbol{\varphi}$ is nowhere zero.

Conversely, if $U$ is simply-connected, and if $\varphi_1, \varphi_2$ and $\varphi_3$ are holomorphic functions on $U$ satisfying conditions (i) and (ii) above, there is a conformally parametrized minimal surface $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ such that $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$ satisfies Eq. 12.18. Moreover, $\boldsymbol{\sigma}$ is uniquely determined by $\varphi_1, \varphi_2$ and $\varphi_3$ up to a translation.

An open subset $U$ of $\mathbb{R}^2$ is said to be *simply-connected* if every simple closed curve in $U$ can be shrunk to a point staying inside $U$. Intuitively, this means that $U$ has no 'holes'.



Simply-connected          Not simply-connected

In the course of the following proof, and in the proof of Proposition 12.5.5 below, we shall need to recall that, if $F$ is a holomorphic function of $\zeta = u + iv$, then

$$F_u = F', \quad F_v = iF', \quad (\overline{F})_u = \overline{F'}, \quad (\overline{F})_v = -i\overline{F'},$$

where $F' = dF/d\zeta$ is the complex derivative of $F$, and the bar denotes complex-conjugate.

## Proof

Suppose first that $\boldsymbol{\sigma} = (\sigma^1, \sigma^2, \sigma^3)$ is minimal, where $\sigma^k : U \to \mathbb{R}$ for $k = 1, 2, 3$. We have to show that $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$ satisfies conditions (i) and (ii). Since $\varphi_k = \sigma_u^k - i\sigma_v^k$ for $k = 1, 2, 3$,

$$\sum_{k=1}^{3} \varphi_k^2 = \sum_{k=1}^{3} \left((\sigma_u^k)^2 - (\sigma_v^k)^2 - 2i\sigma_u^k\sigma_v^k\right) = \| \boldsymbol{\sigma}_u \|^2 - \| \boldsymbol{\sigma}_v \|^2 - 2i\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v, \tag{12.20}$$

which vanishes since $\boldsymbol{\sigma}$ is conformal. Finally, $\boldsymbol{\varphi} = \mathbf{0}$ if and only if $\boldsymbol{\sigma}_u = \boldsymbol{\sigma}_v = \mathbf{0}$, and this is impossible since $\boldsymbol{\sigma}$ is regular.

For the converse, take $\boldsymbol{\varphi}$ satisfying conditions (i) and (ii). We must show that $\boldsymbol{\varphi}$ arises from a minimal surface as above, and that this minimal surface is unique up to a translation of $\mathbb{R}^3$. Fix $(u_0, v_0) \in U$ and define $\boldsymbol{\sigma}$ as the real part of a complex line integral:

$$\boldsymbol{\sigma}(u, v) = \mathfrak{Re} \int_{\boldsymbol{\pi}} \boldsymbol{\varphi}(\xi) d\xi,$$

where $\boldsymbol{\pi}$ is any curve in $U$ from $(u_0, v_0)$ to $(u, v) \in U$. The fact that $U$ is simply-connected implies, by virtue of Cauchy's Theorem, that $\int_{\boldsymbol{\pi}} \boldsymbol{\varphi}(\xi) d\xi$ is independent of the path $\boldsymbol{\pi}$ chosen, and hence so is $\boldsymbol{\sigma}(u, v)$. Now, $\boldsymbol{\Phi}(\zeta) = \int_{\boldsymbol{\pi}} \boldsymbol{\varphi}(\xi) d\xi$ is a holomorphic function of $\zeta = u + iv$, and $\boldsymbol{\Phi}'(\zeta) = \boldsymbol{\varphi}(\zeta)$. Hence, by the facts stated just before the beginning of the proof,

$$\begin{aligned} \boldsymbol{\sigma}_u &= \mathfrak{Re}(\boldsymbol{\Phi}_u) = \mathfrak{Re}(\boldsymbol{\Phi}') = \mathfrak{Re}(\boldsymbol{\varphi}), \\ \boldsymbol{\sigma}_v &= \mathfrak{Re}(\boldsymbol{\Phi}_v) = \mathfrak{Re}(i\boldsymbol{\Phi}') = -\mathfrak{Im}(\boldsymbol{\varphi}), \end{aligned} \tag{12.21}$$

so $\boldsymbol{\varphi} = \boldsymbol{\sigma}_u - i\boldsymbol{\sigma}_v$.

To complete the proof, we have to show that $\boldsymbol{\sigma}$ is a conformal surface patch. But, condition (ii) and Eqs. 12.21 show that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are not both zero. By condition (i) and Eq. 12.20, $\| \boldsymbol{\sigma}_u \| = \| \boldsymbol{\sigma}_v \|$ and $\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0$. Since $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are not both zero, this proves that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are both non-zero and perpendicular, hence linearly independent, so that $\boldsymbol{\sigma}$ is a regular surface patch; it also proves that $\boldsymbol{\sigma}$ is conformal.

If another conformal minimal surface $\tilde{\boldsymbol{\sigma}}$ corresponds to the same holomorphic function $\boldsymbol{\varphi}$ as $\boldsymbol{\sigma}$, then $\tilde{\boldsymbol{\sigma}}_u = \boldsymbol{\sigma}_u$ and $\tilde{\boldsymbol{\sigma}}_v = \boldsymbol{\sigma}_v$ everywhere on $U$, which implies that $\tilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma}$ is a constant, say $\mathbf{a}$, so that $\tilde{\boldsymbol{\sigma}}$ is obtained from $\boldsymbol{\sigma}$ by translating by the vector $\mathbf{a}$. □

Before giving some examples, we observe that, if a holomorphic function $\boldsymbol{\varphi}$ satisfies the conditions in Theorem 12.5.2, so does $i\boldsymbol{\varphi}$. If $\boldsymbol{\varphi}$ is the holomorphic function corresponding to a minimal surface $\mathcal{S}$, the minimal surface to which $i\boldsymbol{\varphi}$ corresponds is called the *conjugate* of $\mathcal{S}$. It is well defined by $\mathcal{S}$ up to a translation.

## Example 12.5.3

The parametrization

$$\boldsymbol{\sigma}(u,v) = (\cosh u \cos v, \cosh u \sin v, u)$$

of the catenoid is conformal (see the solution of Exercise 6.2.3). The associated holomorphic function is

$$\begin{aligned}
\boldsymbol{\varphi}(\zeta) &= \boldsymbol{\sigma}_u - i\boldsymbol{\sigma}_v \\
&= (\sinh u \cos v + i \cosh u \sin v, \sinh u \sin v - i \cosh u \cos v, 1) \\
&= (\sinh(u+iv), -i\cosh(u+iv), 1) \\
&= (\sinh\zeta, -i\cosh\zeta, 1).
\end{aligned}$$

Note that conditions (i) and (ii) in Theorem 12.5.2 are satisfied, since $\boldsymbol{\varphi}$ is clearly never zero and the sum of the squares of its components is

$$\sinh^2\zeta - \cosh^2\zeta + 1 = 0.$$

Let us determine the conjugate minimal surface $\tilde{\boldsymbol{\sigma}}$ of the catenoid. From the proof of Theorem 12.5.2,

$$\begin{aligned}
\tilde{\boldsymbol{\sigma}}(u,v) &= \mathfrak{Re}\int_{\boldsymbol{\pi}} (i\sinh\xi, \cosh\xi, i)\, d\xi \\
&= \mathfrak{Re}(i\cosh\zeta, \sinh\zeta, i\zeta) \\
&= (-\sinh u \sin v, \sinh u \cos v, -v),
\end{aligned}$$

up to a translation. If we reparametrize by defining $\tilde{u} = \sinh u, \tilde{v} = v + \pi/2$, we get the surface

$$(\tilde{u}, \tilde{v}) \mapsto (\tilde{u}\cos\tilde{v}, \tilde{u}\sin\tilde{v}, -\tilde{v}),$$

after translating by $(0, 0, -\pi/2)$, which is obtained from the helicoid in Exercise 4.2.6 by reflecting in the $z$-axis. Note that the parametrization of the helicoid given in Exercise 4.2.6 is not conformal, so the constructions in this section cannot be applied to it.

It is actually possible to 'solve' the conditions on $\boldsymbol{\varphi}$ in Theorem 12.5.2.

## Proposition 12.5.4

Let $f(\zeta)$ be a holomorphic function on an open set $U$ in the complex plane, not identically zero, and let $g(\zeta)$ be a meromorphic function on $U$ such that, if $\zeta_0 \in U$ is a pole of $g$ of order $m \geq 1$, say, then $\zeta_0$ is also a zero of $f$ of order $\geq 2m$. Then,

$$\boldsymbol{\varphi} = \left(\frac{1}{2}f(1-g^2), \frac{i}{2}f(1+g^2), fg\right) \tag{12.22}$$

satisfies conditions (i) and (ii) in Theorem 12.5.2, and conversely every holomorphic function $\boldsymbol{\varphi}$ satisfying these conditions arises in this way.

The correspondence given by Theorem 12.5.2 and Proposition 12.5.4 between pairs of functions $f$ and $g$ and minimal surfaces is called *Weierstrass' representation*.

## Proof

Suppose that $f$ and $g$ are as in the statement of the proposition. If $g$ has a pole of order $m \geq 1$ at $\zeta_0 \in U$, and $f$ has a zero of order $n \geq 2m$ at $\zeta_0$, then the Laurent expansions of $f$ and $g$ about $\zeta_0$ are of the form

$$f(\zeta) = a(\zeta - \zeta_0)^n + \cdots \quad \text{and} \quad g(\zeta) = \frac{b}{(\zeta - \zeta_0)^m} + \cdots ,$$

where $a$ and $b$ are non-zero complex numbers and the $\cdots$ indicates terms involving higher powers of $\zeta - \zeta_0$. Then,

$$f(1 \pm g^2) = \pm ab^2(\zeta - \zeta_0)^{n-2m} + \cdots \quad \text{and} \quad fg = ab(\zeta - \zeta_0)^{n-m} + \cdots$$

involve only non-negative powers of $\zeta - \zeta_0$, so $\boldsymbol{\varphi}$ is holomorphic near $\zeta_0$. Since it is clear that $\boldsymbol{\varphi}$ is holomorphic wherever $g$ is holomorphic, it follows that the function $\boldsymbol{\varphi}$ defined by Eq. 12.22 is holomorphic everywhere on $U$. It is clear that $\boldsymbol{\varphi}$ is identically zero only if $f$ is identically zero, and simple algebra shows that $\boldsymbol{\varphi}$ satisfies condition (i) in Theorem 12.5.2.

Conversely, suppose that $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$ is a holomorphic function satisfying conditions (i) and (ii) in Theorem 12.5.2. If $\varphi_1 - i\varphi_2$ is not identically zero, define

$$f = \varphi_1 - i\varphi_2, \quad g = \frac{\varphi_3}{\varphi_1 - i\varphi_2}. \tag{12.23}$$

Since $\boldsymbol{\varphi}$ is holomorphic, $f$ is holomorphic and $g$ is meromorphic. Condition (i) implies that $(\varphi_1 + i\varphi_2)(\varphi_1 - i\varphi_2) = -\varphi_3^2$, and hence that

$$\varphi_1 + i\varphi_2 = -fg^2. \tag{12.24}$$

Simple algebra shows that Eqs. 12.23 and 12.24 imply Eq. 12.22. Equation 12.24 implies that $fg^2$ is holomorphic, and the argument with Laurent expansions in the first part of the proof now gives the condition on the zeros and poles of $f$ and $g$. Finally, if $\varphi_1 - i\varphi_2 = 0$, we repeat the above argument replacing $\varphi_1 \pm i\varphi_2$ by $\varphi_1 \mp i\varphi_2$ (note that $\varphi_1 - i\varphi_2$ and $\varphi_1 + i\varphi_2$ cannot both be zero, for if they were we would have $\varphi_1 = \varphi_2 = 0$, hence $\varphi_3 = 0$ by condition (i), and this would violate condition (ii)). $\square$

We give only one application of Weierstrass' representation.

## Proposition 12.5.5

The Gaussian curvature of the minimal surface corresponding to the functions $f$ and $g$ in Weierstrass' representation is

$$K = \frac{-16|dg/d\zeta|^2}{|f|^2(1+|g|^2)^4}.$$

## Proof

This is a straightforward, if tedious, computation, and we shall omit many of the details. Define $\overline{\boldsymbol{\varphi}}$ by taking the complex-conjugate of each component of $\boldsymbol{\varphi}$. Then, $\boldsymbol{\sigma}_u = \frac{1}{2}(\boldsymbol{\varphi} + \overline{\boldsymbol{\varphi}})$, $\boldsymbol{\sigma}_v = \frac{1}{2i}(\overline{\boldsymbol{\varphi}} - \boldsymbol{\varphi})$. Since $\boldsymbol{\varphi} \cdot \boldsymbol{\varphi} = \overline{\boldsymbol{\varphi}} \cdot \overline{\boldsymbol{\varphi}} = 0$, the first fundamental form is $\frac{1}{2}\boldsymbol{\varphi} \cdot \overline{\boldsymbol{\varphi}}(du^2 + dv^2)$. Substituting the formula for $\boldsymbol{\varphi}$ from Eq. 12.22 and simplifying, we find that the first fundamental form is

$$\frac{1}{4}|f|^2(1+|g|^2)^2(du^2+dv^2). \tag{12.25}$$

Next,

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \frac{1}{4i}(\boldsymbol{\varphi} + \overline{\boldsymbol{\varphi}}) \times (\overline{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \quad \frac{1}{2i}\boldsymbol{\varphi} \times \overline{\boldsymbol{\varphi}},$$

$$\therefore \quad \| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \|^2 \quad = \quad -\frac{1}{4}(\boldsymbol{\varphi} \times \overline{\boldsymbol{\varphi}}) \cdot (\boldsymbol{\varphi} \times \overline{\boldsymbol{\varphi}}) \quad -\frac{1}{4}( \ \boldsymbol{\varphi} \cdot \boldsymbol{\varphi}) \ \overline{\boldsymbol{\varphi}} \cdot \overline{\boldsymbol{\varphi}}) - (\boldsymbol{\varphi} \cdot \overline{\boldsymbol{\varphi}})^2) \quad \frac{1}{4}(\boldsymbol{\varphi} \cdot \overline{\boldsymbol{\varphi}})^2,$$

$$\therefore \quad \mathbf{N} = i\frac{\overline{\boldsymbol{\varphi}} \times \boldsymbol{\varphi}}{\boldsymbol{\varphi} \cdot \overline{\boldsymbol{\varphi}}}.$$

In terms of $f$ and $g$, this becomes

$$\mathbf{N} = \frac{1}{1+|g|^2}\left(g + \overline{g}, -i(g - \overline{g}), |g|^2 - 1\right). \tag{12.26}$$

Using the remarks preceding the proof of Theorem 12.5.2 and the formulas

$$L = -\boldsymbol{\sigma}_u \cdot \mathbf{N}_u, \quad M = -\boldsymbol{\sigma}_u \cdot \mathbf{N}_v, \quad N = -\boldsymbol{\sigma}_v \cdot \mathbf{N}_v$$

(which follow by differentiating $\boldsymbol{\sigma}_u \cdot \mathbf{N} = \boldsymbol{\sigma}_v \cdot \mathbf{N} = 0$), we find that the second fundamental form is

$$-\frac{1}{2}\left((fg' + \overline{fg'})(du^2+dv^2) + 2i(fg' - \overline{fg'})du dv\right). \tag{12.27}$$

Combining Eqs. 12.25–12.27, and using the formula for the Gaussian curvature $K$ in Corollary 8.1.3, we finally obtain the formula in the statement of the proposition. $\square$

## Corollary 12.5.6

Let $\mathcal{S}$ be a minimal surface that is not part of a plane. Then, the zeros of the Gaussian curvature of $\mathcal{S}$ are isolated.

This means that, if the Gaussian curvature $K$ vanishes at a point $\mathbf{p} \in \mathcal{S}$, then $K$ does not vanish at any other point of $\mathcal{S}$ sufficiently near to $\mathbf{p}$. More precisely, if $\mathbf{p}$ lies in a surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$, say $\mathbf{p} = \boldsymbol{\sigma}(u_0, v_0)$, there is a number $\epsilon > 0$ such that $K$ does not vanish at the point $\boldsymbol{\sigma}(u, v) \in \mathcal{S}$ if $0 < (u - u_0)^2 + (v - v_0)^2 < \epsilon^2$.

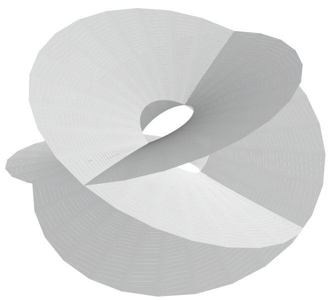
## Proof

From the formula for $K$ in Proposition 12.5.5, $K$ vanishes exactly where the meromorphic function $g'$ vanishes. If $g'$ is zero everywhere, so is $K$ and $\mathcal{S}$ is an open subset of a plane (this was shown in Proposition 8.2.9, but follows immediately from Eq. 12.26 which shows that $\mathbf{N}$ is constant if $g$ is constant). But it is a standard result of complex analysis that the zeros of a non-zero meromorphic function are isolated, so if $K$ is not identically zero its zeros must be isolated. $\square$

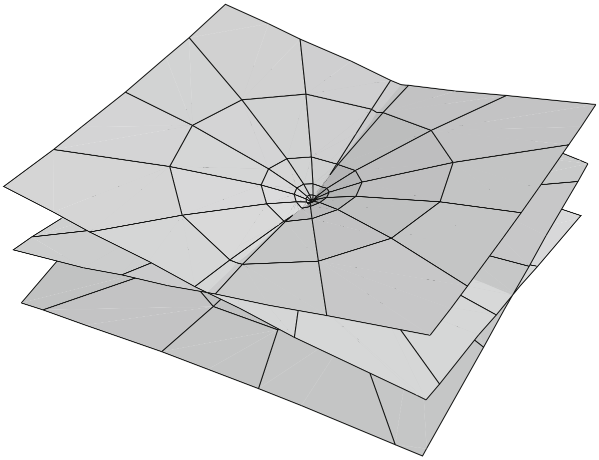
## *EXERCISES*

12.5.1 Find the holomorphic function $\boldsymbol{\varphi}$ corresponding to Enneper's minimal surface given in Example 12.2.5. Show that its conjugate minimal surface coincides with a reparametrization of the same surface rotated by $\pi/4$ around the $z$-axis.

12.5.2 Find a parametrization of *Henneberg's surface*, the minimal surface corresponding to the functions $f(\zeta) = 1 - \zeta^{-4}$, $g(\zeta) = \zeta$ in Weierstrass' representation. The following are a 'close up' view and a 'large scale' view of this surface.



Henneberg: close up



Henneberg: Large scale

12.5.3 Show that, if $\boldsymbol{\varphi}$ satisfies the conditions in Theorem 12.5.2, so does $a\boldsymbol{\varphi}$ for any non-zero constant $a \in \mathbb{C}$; let $\boldsymbol{\sigma}^a$ be the minimal surface patch corresponding to $a\boldsymbol{\varphi}$, and let $\boldsymbol{\sigma}^1 = \boldsymbol{\sigma}$ be that corresponding to $\boldsymbol{\varphi}$. Show that:

(i) If $a \in \mathbb{R}$, then $\boldsymbol{\sigma}^a$ is obtained from $\boldsymbol{\sigma}$ by applying a dilation and a translation.

(ii) If $|a| = 1$, the map $\boldsymbol{\sigma}(u,v) \mapsto \boldsymbol{\sigma}^a(u,v)$ is an isometry, and the tangent planes of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ at corresponding points are parallel (in particular, the images of the Gauss maps of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^a$ are the same).

12.5.4 Show that if the function $\boldsymbol{\varphi}$ in the preceding exercise is that corresponding to the catenoid (see Example 12.5.3), the surface $\boldsymbol{\sigma}^{e^{it}}$ coincides with the surface denoted by $\boldsymbol{\sigma}^t$ in Exercise 6.2.3.

12.5.5 Let $\boldsymbol{\gamma} : (\alpha, \beta) \to \mathbb{R}^3$ be a (regular) curve in the $xy$-plane, say

$$\boldsymbol{\gamma}(u) = (f(u), g(u), 0),$$

and assume that there are holomorphic functions $F$ and $G$ defined on a rectangle

$$\mathcal{U} = \{u + iv \in \mathbb{C} \,|\, \alpha < u < \beta,\ -\epsilon < v < \epsilon\},$$

for some $\epsilon > 0$, and such that $F(u) = f(u)$ and $G(u) = g(u)$ if $u$ is real and $\alpha < u < \beta$. Note that (with a dash denoting $d/dz$ as usual),

$$F'(z)^2 + G'(z)^2 \neq 0 \qquad \text{if } \mathfrak{Im}(z) = 0,$$

so by shrinking $\epsilon$ if necessary we can assume that $F'(z)^2+G'(z)^2 \neq 0$ for all $z \in \mathcal{U}$. Show that:

(i) The vector-valued holomorphic function

$$\boldsymbol{\varphi} = (F', G', i(F'^2 + G'^2)^{1/2})$$

satisfies the conditions of Theorem 12.5.2 and therefore defines a minimal surface $\boldsymbol{\sigma}(u, v)$.

(ii) Up to a translation, $\boldsymbol{\sigma}(u, 0) = \boldsymbol{\gamma}(u)$ for $\alpha < u < \beta$.

(iii) $\boldsymbol{\gamma}$ is a pre-geodesic on $\boldsymbol{\sigma}$ (see Exercise 9.1.2).

(iv) If we start with the cycloid

$$\boldsymbol{\gamma}(u) = (u - \sin u, 1 - \cos u, 0),$$

the resulting surface $\boldsymbol{\sigma}$ is, up to a translation, Catalan's surface and we have 'explained' why Catalan's surface has a cycloidal geodesic – see Exercise 12.2.4.

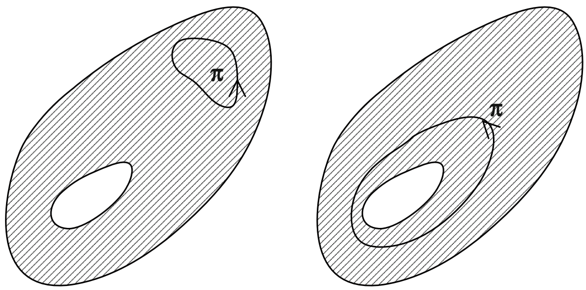
# 13
# The Gauss–Bonnet theorem

The Gauss–Bonnet theorem is the most beautiful and profound result in the theory of surfaces. Its most important version relates the average of the Gaussian curvature to a property of the surface called its 'Euler number' which is 'topological', i.e., it is unchanged by any diffeomorphism of the surface. Such diffeomorphisms will in general change the value of the Gaussian curvature, but the theorem says that its average over the surface does *not* change. The real importance of the Gauss–Bonnet theorem is as a prototype of analogous results which apply in higher dimensional situations, and which relate *geometrical* properties to *topological* ones. The study of such relations was one of the most important themes of twentieth century mathematics, and continues to be actively studied today.

## 13.1 Gauss–Bonnet for simple closed curves

The simplest version of the Gauss–Bonnet Theorem involves *simple closed curves* on a surface. In the special case when the surface is a plane, these curves have been discussed in Section 3.1. For a general surface, we make the following definition.

Andrew Pressley, *Elementary Differential Geometry: Second Edition*,
Springer Undergraduate Mathematics Series, DOI 10.1007/978-1-84882-891-9_13,




allowed not allowed

## Definition 13.1.1

A curve $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ on a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is called a *simple closed curve* with *period* $T$ if $\boldsymbol{\pi}(t) = (u(t), v(t))$ is a simple closed curve in $\mathbb{R}^2$ with period $T$ such that the region $\operatorname{int}(\boldsymbol{\pi})$ of $\mathbb{R}^2$ enclosed by $\boldsymbol{\pi}$ is entirely contained in $U$ (see the diagrams above). The curve $\boldsymbol{\gamma}$ is said to be *positively-oriented* if $\boldsymbol{\pi}$ is positively-oriented. Finally, the image of $\operatorname{int}(\boldsymbol{\pi})$ under the map $\boldsymbol{\sigma}$ is defined to be the *interior* $\operatorname{int}(\boldsymbol{\gamma})$ of $\boldsymbol{\gamma}$.

We can now state the first version of the Gauss–Bonnet Theorem.

## Theorem 13.1.2

Let $\boldsymbol{\gamma}(s)$ be a unit-speed simple closed curve on a surface patch $\boldsymbol{\sigma}$ of length $\ell(\boldsymbol{\gamma})$, and assume that $\boldsymbol{\gamma}$ is positively-oriented. Then,

$$\int_0^{\ell(\boldsymbol{\gamma})} \kappa_g ds = 2\pi - \int_{\operatorname{int}(\boldsymbol{\gamma})} K d\mathcal{A}_{\boldsymbol{\sigma}},$$

where $\kappa_g$ is the geodesic curvature of $\boldsymbol{\gamma}$, $K$ is the Gaussian curvature of $\boldsymbol{\sigma}$ and $d\mathcal{A}_{\boldsymbol{\sigma}}$ is the area element of $\boldsymbol{\sigma}$ (see Section 6.4).

We use $s$ to denote the parameter of $\boldsymbol{\gamma}$ to emphasize that $\boldsymbol{\gamma}$ is unit-speed. The double integral on the right-hand side of the equation in Theorem 13.1.2 is called the *total curvature* of the region $\operatorname{int}(\boldsymbol{\gamma})$.

## Proof

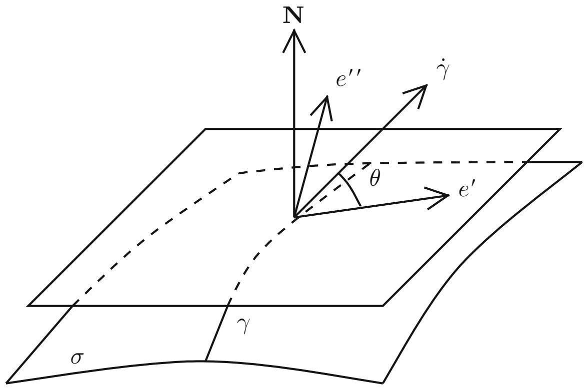
We start by computing the geodesic curvature of $\boldsymbol{\gamma}$. For this, we shall make use of a smooth *orthonormal* basis $\{\boldsymbol{e}', \boldsymbol{e}''\}$ of the tangent plane at each point of the surface patch, where 'smooth' means that $\boldsymbol{e}'$ and $\boldsymbol{e}''$ are smooth functions of the surface parameters $(u, v)$. Then, $\{\boldsymbol{e}', \boldsymbol{e}'', \mathbf{N}\}$ is an orthonormal basis of $\mathbb{R}^3$ ($\mathbf{N}$ being the standard unit normal of $\boldsymbol{\sigma}$), and we shall assume that it is *right-handed*, i.e., that $\mathbf{N} = \boldsymbol{e}' \times \boldsymbol{e}''$. This can always be achieved by interchanging $\boldsymbol{e}'$ and $\boldsymbol{e}''$ if necessary. Note that the dashes on $\boldsymbol{e}'$ and $\boldsymbol{e}''$ have nothing to do with derivatives.

Let $\theta(s)$ be the oriented angle $\widehat{\boldsymbol{e}'\dot{\boldsymbol{\gamma}}}$ between the unit tangent vector $\dot{\boldsymbol{\gamma}}(s)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(s)$ and the unit vector $\boldsymbol{e}'$ at the same point. Thus,

$$\dot{\boldsymbol{\gamma}} = \cos\theta \boldsymbol{e}' + \sin\theta \boldsymbol{e}''. \tag{13.1}$$

Then,

$$\mathbf{N} \times \dot{\boldsymbol{\gamma}} = -\sin\theta \boldsymbol{e}' + \cos\theta \boldsymbol{e}''. \tag{13.2}$$




Now, by Eq. 13.1,

$$\ddot{\boldsymbol{\gamma}} = \cos\theta \dot{\boldsymbol{e}}' + \sin\theta \dot{\boldsymbol{e}}'' + \dot{\theta}(-\sin\theta \boldsymbol{e}' + \cos\theta \boldsymbol{e}''), \tag{13.3}$$

so by Eqs. 13.2 and 13.3 the geodesic curvature of $\boldsymbol{\gamma}$ is

$$\begin{aligned}
\kappa_g &= (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \cdot \ddot{\boldsymbol{\gamma}} \quad \text{(see Section 7.3)} \\
&= \dot{\theta}(-\sin\theta \boldsymbol{e}' + \cos\theta \boldsymbol{e}'') \cdot (-\sin\theta \boldsymbol{e}' + \cos\theta \boldsymbol{e}'') \\
&\qquad + (-\sin\theta \boldsymbol{e}' + \cos\theta \boldsymbol{e}'') \cdot (\cos\theta \dot{\boldsymbol{e}}' + \sin\theta \dot{\boldsymbol{e}}'') \\
&= \dot{\theta} + \cos^2\theta(\dot{\boldsymbol{e}}' \cdot \boldsymbol{e}'') - \sin^2\theta(\dot{\boldsymbol{e}}'' \cdot \boldsymbol{e}') \\
&\qquad + \sin\theta\cos\theta(\dot{\boldsymbol{e}}'' \cdot \boldsymbol{e}'' - \dot{\boldsymbol{e}}' \cdot \boldsymbol{e}') \quad \text{(by Eqs. 13.1 and 13.2)}.
\end{aligned}$$

Since $\boldsymbol{e}'$ and $\boldsymbol{e}''$ are perpendicular unit vectors,

$$\boldsymbol{e}' \cdot \dot{\boldsymbol{e}}' = \boldsymbol{e}'' \cdot \dot{\boldsymbol{e}}'' = 0, \quad \dot{\boldsymbol{e}}' \cdot \boldsymbol{e}'' = -\boldsymbol{e}' \cdot \dot{\boldsymbol{e}}''.$$

Hence,

$$\kappa_g = \dot{\theta} - \boldsymbol{e}' \cdot \dot{\boldsymbol{e}}''. \tag{13.4}$$

Thus, to compute the left-hand side of the equation in Theorem 13.1.2, we must compute the integrals of $\dot{\theta}$ and of $\boldsymbol{e}' \cdot \dot{\boldsymbol{e}}''$ around the curve $\boldsymbol{\gamma}$. We begin with the latter, for which we shall need the following lemma.

## Lemma 13.1.3

Let

$$Edu^2 + 2Fdudv + Gdv^2 \quad \text{and} \quad Ldu^2 + 2Mdudv + Ndv^2$$

be the first and second fundamental forms of $\boldsymbol{\sigma}$, respectively. Then, with the above notation, we have

$$\mathbf{e}'_u \cdot \mathbf{e}''_v - \mathbf{e}''_u \cdot \mathbf{e}'_v = \frac{LN - M^2}{(EG - F^2)^{1/2}}. \tag{13.5}$$

Assuming this for a moment, we compute

$$\int_0^{\ell(\boldsymbol{\gamma})} \boldsymbol{e}' \cdot \dot{\boldsymbol{e}}'' ds = \int_0^{\ell(\boldsymbol{\gamma})} \boldsymbol{e}' \cdot (\dot{u}\boldsymbol{e}''_u + \dot{v}\boldsymbol{e}''_v) ds = \int_{\boldsymbol{\pi}} (\boldsymbol{e}' \cdot \boldsymbol{e}''_u) du + (\boldsymbol{e}' \cdot \boldsymbol{e}''_v) dv.$$

By Green's theorem (see Section 3.2), this can be rewritten as a double integral:

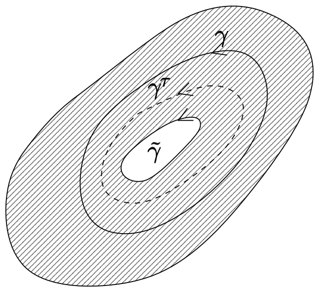
$$\begin{aligned}
\int_0^{\ell(\boldsymbol{\gamma})} \boldsymbol{e}' \cdot \dot{\boldsymbol{e}}'' ds &= \int_{\operatorname{int}(\boldsymbol{\pi})} \{(\boldsymbol{e}' \cdot \boldsymbol{e}''_v)_u - (\boldsymbol{e}' \cdot \boldsymbol{e}''_u)_v\} dudv \\
&= \int_{\operatorname{int}(\boldsymbol{\pi})} \{(\boldsymbol{e}'_u \cdot \boldsymbol{e}''_v) - (\boldsymbol{e}'_v \cdot \boldsymbol{e}''_u)\} dudv \\
&= \int_{\operatorname{int}(\boldsymbol{\pi})} \frac{LN - M^2}{(EG - F^2)^{1/2}} dudv \quad \text{(by Lemma 13.1.3)} \\
&= \int_{\operatorname{int}(\boldsymbol{\pi})} \frac{LN - M^2}{EG - F^2} (EG - F^2)^{1/2} dudv \\
&= \int_{\operatorname{int}(\boldsymbol{\pi})} K d\mathcal{A}_{\boldsymbol{\sigma}} \quad \text{(see Section 6.4).}
\end{aligned} \tag{13.6}$$

If we combine Eqs. 13.4 and 13.6 and compare with the statement of the theorem, we see that what remains is to prove that

$$\int_0^{\ell(\boldsymbol{\gamma})} \dot{\theta}\, ds = 2\pi.$$

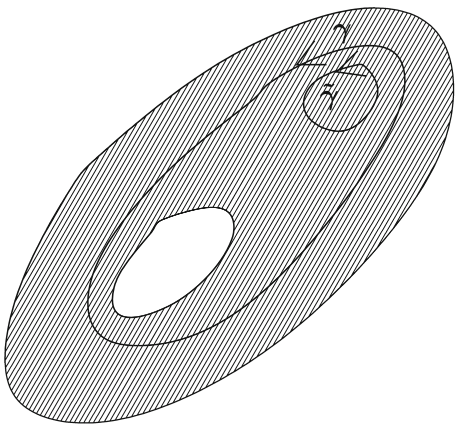
This is a version of Hopf's Umlaufsatz (cf. Theorem 3.1.4). We cannot give a fully satisfactory proof of it here but we offer the following heuristic argument.

The main observation is that, if $\tilde{\boldsymbol{\gamma}}$ is any other simple closed curve contained in the interior of $\boldsymbol{\gamma}$, there is a smooth family of simple closed curves $\boldsymbol{\gamma}^\tau$, defined for $0 \le \tau \le 1$, say, with $\boldsymbol{\gamma}^0 = \boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^1 = \tilde{\boldsymbol{\gamma}}$ (see Section 9.4 for the notion of a smooth family of curves). The existence of such a family is supposed to be 'intuitively obvious'.




Note, however, that it is crucial that the interior of $\boldsymbol{\pi}$ is entirely contained in $U$, otherwise such a family will not exist, in general (see the diagram below).

Observe next that the integral $\int_0^{\ell(\boldsymbol{\gamma}^\tau)} \dot{\theta}\, ds$ should depend continuously on $\tau$. Further, since $\boldsymbol{\gamma}^\tau$ and $\boldsymbol{e}'$ return to their original values as one goes once round $\boldsymbol{\gamma}^\tau$, the integral is always an integer multiple of $2\pi$. These two facts imply that the integral must be independent of $\tau$ – for by the intermediate value theorem a continuous variable cannot change from one integer to a different integer




without passing through some non-integer value. To compute $\int_0^{\ell(\boldsymbol{\gamma})} \dot{\theta}\, ds$, we can therefore replace $\boldsymbol{\gamma}$ by any other simple closed curve $\tilde{\boldsymbol{\gamma}}$ in the interior of $\boldsymbol{\gamma}$, since this will not change the value of the integral. We take $\tilde{\boldsymbol{\gamma}}$ to be the image under $\boldsymbol{\sigma}$ of a small circle in the interior of $\boldsymbol{\pi}$. It is 'intuitively clear' that

$$\int_0^{\ell(\tilde{\boldsymbol{\gamma}})} \dot{\theta}\, ds = 2\pi,$$

because

(i) $\boldsymbol{e}'$ is essentially constant at all points of $\tilde{\boldsymbol{\gamma}}$ (because the circle is very small), and

(ii) the tangent vector to $\tilde{\boldsymbol{\gamma}}$ rotates by $2\pi$ on going once round $\tilde{\boldsymbol{\gamma}}$ because the interior of $\tilde{\boldsymbol{\gamma}}$ can be considered to be essentially part of a plane, and it is 'intuitively clear' that the tangent vector of a simple closed curve in the plane rotates by $2\pi$ on going once round the curve (indeed, this is the content of Theorem 3.1.4).

This completes the 'proof' of Hopf's Umlaufsatz. To complete the proof of Theorem 13.1.2, all that remains is to prove Lemma 13.1.3.

*Proof 13.1.3* We can express the partial derivatives of $\boldsymbol{e}'$ and $\boldsymbol{e}''$ with respect to $u$ and $v$ in terms of the orthonormal basis $\{\boldsymbol{e}', \boldsymbol{e}'', \mathbf{N}\}$. Since both partial derivatives of $\boldsymbol{e}'$ are perpendicular to $\boldsymbol{e}'$, the $\boldsymbol{e}'$ components of $\boldsymbol{e}'_u$ and $\boldsymbol{e}'_v$ are zero (and similarly for $\boldsymbol{e}''$). Thus,

$$\begin{aligned}
\boldsymbol{e}'_u &= \qquad\quad \alpha\boldsymbol{e}'' + \lambda'\mathbf{N},\\
\boldsymbol{e}'_v &= \qquad\quad \beta\boldsymbol{e}'' + \mu'\mathbf{N},\\
\boldsymbol{e}''_u &= -\alpha'\boldsymbol{e}' \qquad\quad + \lambda''\mathbf{N},\\
\boldsymbol{e}''_v &= -\beta'\boldsymbol{e}' \qquad\quad + \mu''\mathbf{N},
\end{aligned}$$

for some scalars $\alpha, \beta, \alpha', \beta', \lambda', \mu', \lambda'', \mu''$ (which may depend on $u$ and $v$). Moreover, by differentiating the equation $\boldsymbol{e}' \cdot \boldsymbol{e}'' = 0$ with respect to $u$, we see that $\boldsymbol{e}'_u \cdot \boldsymbol{e}'' = -\boldsymbol{e}' \cdot \boldsymbol{e}''_u$. i.e., $\alpha' = \alpha$ (and similarly $\beta' = \beta$). Thus,

$$\begin{aligned}
\boldsymbol{e}'_u &= \qquad\quad \alpha\boldsymbol{e}'' + \lambda'\mathbf{N},\\
\boldsymbol{e}'_v &= \qquad\quad \beta\boldsymbol{e}'' + \mu'\mathbf{N},\\
\boldsymbol{e}''_u &= -\alpha\boldsymbol{e}' \qquad\quad + \lambda''\mathbf{N},\\
\boldsymbol{e}''_v &= -\beta\boldsymbol{e}' \qquad\quad + \mu''\mathbf{N}.
\end{aligned} \tag{13.7}$$

It follows that

$$\boldsymbol{e}'_u \cdot \boldsymbol{e}''_v - \boldsymbol{e}''_u \cdot \boldsymbol{e}'_v = \lambda'\mu'' - \lambda''\mu'. \tag{13.8}$$

On the other hand, combining the formula

$$\mathbf{N}_u \times \mathbf{N}_v = K\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$$

(see Eq. 8.2) with the formulas

$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \|}, \quad \| \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \| = (EG - F^2)^{1/2}$$

(see Proposition 6.4.2), we get

$$\mathbf{N}_u \times \mathbf{N}_v = \frac{LN - M^2}{(EG - F^2)^{1/2}}\mathbf{N},$$

and hence

$$(\mathbf{N}_u \times \mathbf{N}_v) \cdot \mathbf{N} = \frac{LN - M^2}{(EG - F^2)^{1/2}}. \tag{13.9}$$

Since $\mathbf{N} = \boldsymbol{e}' \times \boldsymbol{e}''$,

$$\begin{aligned}(\mathbf{N}_u \times \mathbf{N}_v) \cdot \mathbf{N} &= (\mathbf{N}_u \times \mathbf{N}_v) \cdot (\boldsymbol{e}' \times \boldsymbol{e}''), \\ &= (\mathbf{N}_u \cdot \boldsymbol{e}')(\mathbf{N}_v \cdot \boldsymbol{e}'') - (\mathbf{N}_u \cdot \boldsymbol{e}'')(\mathbf{N}_v \cdot \boldsymbol{e}') \\ &= (\mathbf{N} \cdot \boldsymbol{e}'_u)(\mathbf{N} \cdot \boldsymbol{e}''_v) - (\mathbf{N} \cdot \boldsymbol{e}''_u)(\mathbf{N} \cdot \boldsymbol{e}'_v) \qquad (13.10) \\ &= \lambda'\mu'' - \lambda''\mu' \quad \text{by Eq. 13.7},\end{aligned}$$

where in passing from the second line to the third we used the equations

$$\begin{aligned}\mathbf{N}_u \cdot \boldsymbol{e}' &= -\mathbf{N} \cdot \boldsymbol{e}'_u, & \mathbf{N}_u \cdot \boldsymbol{e}'' &= -\mathbf{N} \cdot \boldsymbol{e}''_u, \\ \mathbf{N}_v \cdot \boldsymbol{e}' &= -\mathbf{N} \cdot \boldsymbol{e}'_v, & \mathbf{N}_v \cdot \boldsymbol{e}'' &= -\mathbf{N} \cdot \boldsymbol{e}''_v,\end{aligned}$$

which follow by differentiating $\mathbf{N} \cdot \boldsymbol{e}' = 0 = \mathbf{N} \cdot \boldsymbol{e}''$ with respect to $u$ and $v$. Putting Eqs. 13.9 and 13.10 together shows that the right-hand sides of Eqs. 13.5 and 13.8 are equal. Since Eq. 13.8 has already been established, this proves Eq.13.5. □

## *EXERCISES*

13.1.1 Suppose that a surface patch $\boldsymbol{\sigma}$ has Gaussian curvature $\leq 0$ everywhere. Prove that there are no simple closed geodesics on $\boldsymbol{\sigma}$. How do you reconcile this with the fact that the parallels of a circular cylinder are geodesics?

13.1.2 Let $\boldsymbol{\gamma}$ be a unit-speed curve in $\mathbb{R}^3$ with nowhere vanishing curvature. Let $\mathbf{n}$ be the principal normal of $\boldsymbol{\gamma}$, viewed as a curve on $S^2$, and let $s$ be the arc-length of $\mathbf{n}$. Show that the geodesic curvature of $\mathbf{n}$ is, up to a sign,

$$\frac{d}{ds}\left(\tan^{-1}\frac{\tau}{\kappa}\right),$$

where $\kappa$ and $\tau$ are the curvature and torsion of $\boldsymbol{\gamma}$. Show also that, if $\mathbf{n}$ is a *simple closed* curve on $S^2$, the interior and exterior of $\mathbf{n}$ are regions of equal area (*Jacobi's Theorem*).

## 13.2 Gauss–Bonnet for curvilinear polygons

For the next version of Gauss–Bonnet, we shall have to generalize our notion of a curve by allowing the possibility of ‘ corners’. More precisely, we make the following definition.

### Definition 13.2.1

A *curvilinear polygon* in $\mathbb{R}^2$ is a continuous map $\boldsymbol{\pi} : \mathbb{R} \to \mathbb{R}^2$ such that, for some real number $T$ and some points $0 = t_0 < t_1 < \cdots < t_n = T$:

(i) $\boldsymbol{\pi}(t) = \boldsymbol{\pi}(t')$ if and only if $t' - t$ is an integer multiple of $T$.

(ii) $\boldsymbol{\pi}$ is smooth on each of the open intervals $(t_0, t_1), (t_1, t_2), \ldots, (t_{n-1}, t_n)$.

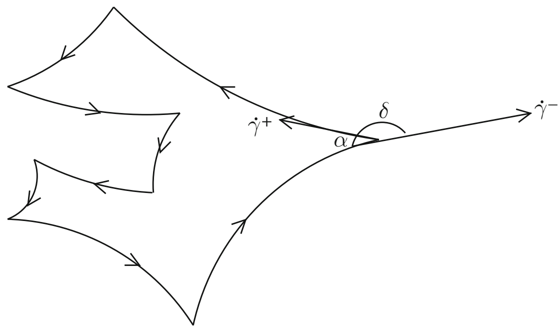
(iii) The one-sided derivatives

$$\dot{\boldsymbol{\pi}}^-(t_i) = \lim_{t \uparrow t_i} \frac{\boldsymbol{\pi}(t) - \boldsymbol{\pi}(t_i)}{t - t_i}, \quad \dot{\boldsymbol{\pi}}^+(t_i) = \lim_{t \downarrow t_i} \frac{\boldsymbol{\pi}(t) - \boldsymbol{\pi}(t_i)}{t - t_i} \tag{13.11}$$

exist for $i = 1, \ldots, n$ and are non-zero and not parallel.

The points $\boldsymbol{\gamma}(t_i)$ for $i = 1, \ldots, n$ are called the *vertices* of the curvilinear polygon $\boldsymbol{\pi}$, and the segments of it corresponding to the open intervals $(t_{i-1}, t_i)$ are called its *edges*.

It makes sense to say that a curvilinear polygon $\boldsymbol{\pi}$ is positively-oriented: for all $t$ such that $\boldsymbol{\pi}(t)$ is not a vertex, the vector $\mathbf{n}_s$ obtained by rotating $\dot{\boldsymbol{\pi}}$ anticlockwise by $\pi/2$ should point into $\mathrm{int}(\boldsymbol{\pi})$. (The region $\mathrm{int}(\boldsymbol{\pi})$ enclosed by $\boldsymbol{\pi}$ makes sense because the Jordan Curve Theorem applies to curvilinear polygons in the plane.)




Now let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a surface patch and let $\boldsymbol{\pi} : \mathbb{R} \to U$ be a curvilinear polygon in $U$, as in Definition 13.2.1. Then, $\boldsymbol{\gamma} = \boldsymbol{\sigma} \circ \boldsymbol{\pi}$ is called a curvilinear polygon on the surface patch $\boldsymbol{\sigma}$, $\mathrm{int}(\boldsymbol{\gamma})$ is the image under $\boldsymbol{\sigma}$ of $\mathrm{int}(\boldsymbol{\pi})$, the vertices of $\boldsymbol{\gamma}$ are the points $\boldsymbol{\gamma}(t_i)$ for $i = 1, \ldots, n$, and the edges of $\boldsymbol{\gamma}$ are the segments of it corresponding to the open intervals $(t_{i-1}, t_i)$. Since $\boldsymbol{\sigma}$ is allowable, the one-sided derivatives

$$\dot{\boldsymbol{\gamma}}^-(t_i) = \lim_{t \uparrow t_i} \frac{\boldsymbol{\gamma}(t) - \boldsymbol{\gamma}(t_i)}{t - t_i}, \quad \dot{\boldsymbol{\gamma}}^+(t_i) = \lim_{t \downarrow t_i} \frac{\boldsymbol{\gamma}(t) - \boldsymbol{\gamma}(t_i)}{t - t_i}$$

exist and are not parallel.

Let $\theta_i^\pm$ be the angles between $\dot{\boldsymbol{\gamma}}^\pm(t_i)$ and $\boldsymbol{e}'$, defined as in Eq. 13.1, let $\delta_i = \theta_i^+ - \theta_i^-$ be the external angle at the vertex $\boldsymbol{\gamma}(t_i)$, and let $\alpha_i = \pi - \delta_i$ be the internal angle. Since the tangent vectors $\dot{\boldsymbol{\gamma}}^+(t_i)$ and $\dot{\boldsymbol{\gamma}}^-(t_i)$ are not parallel, the angle $\delta_i$ is not a multiple of $\pi$. Note that all of these angles are well defined only up to multiples of $2\pi$. We assume from now on that $0 < \alpha_i < 2\pi$ for $i = 1, \ldots, n$.

A curvilinear polygon $\boldsymbol{\gamma}$ is said to be unit-speed if $\| \dot{\boldsymbol{\gamma}} \| = 1$ whenever $\dot{\boldsymbol{\gamma}}$ is defined, i.e., for all $t$ such that $\boldsymbol{\gamma}(t)$ is not a vertex of $\boldsymbol{\gamma}$. We denote the parameter of $\boldsymbol{\gamma}$ by $s$ if $\boldsymbol{\gamma}$ is unit-speed. The period of $\boldsymbol{\gamma}$ is then equal to its length $\ell(\boldsymbol{\gamma})$, which is the sum of the lengths of the edges of $\boldsymbol{\gamma}$.

## Theorem 13.2.2

Let $\boldsymbol{\gamma}$ be a positively-oriented unit-speed curvilinear polygon with $n$ edges on a surface $\boldsymbol{\sigma}$, and let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the interior angles at its vertices. Then,

$$\int_0^{\ell(\boldsymbol{\gamma})} \kappa_g ds = \sum_{i=1}^n \alpha_i - (n-2)\pi - \int_{\mathrm{int}(\boldsymbol{\gamma})} K d\mathcal{A}_{\boldsymbol{\sigma}}.$$

## Proof

Exactly the same argument as in the proof of Theorem 13.1.2 shows that

$$\int_0^{\ell(\boldsymbol{\gamma})} \kappa_g ds = \int_0^{\ell(\boldsymbol{\gamma})} \dot{\theta} ds - \int_{\mathrm{int}(\boldsymbol{\gamma})} K d\mathcal{A}_{\boldsymbol{\sigma}}.$$

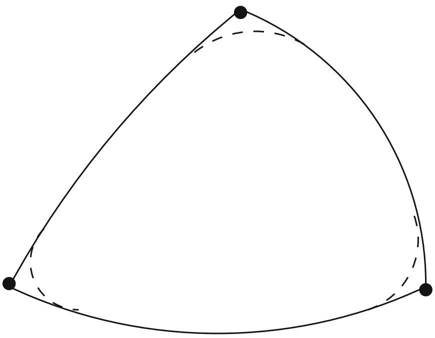
We shall prove that

$$\int_0^{\ell(\boldsymbol{\gamma})} \dot{\theta} ds = 2\pi - \sum_{i=1}^n \delta_i. \tag{13.12}$$

Assuming this, we get

$$\begin{aligned} \int_0^{\ell(\boldsymbol{\gamma})} \kappa_g ds &= 2\pi - \sum_{i=1}^n \delta_i - \int_{\mathrm{int}(\boldsymbol{\gamma})} K d\mathcal{A}_{\boldsymbol{\sigma}} \\ &= 2\pi - \sum_{i=1}^n (\pi - \alpha_i) - \int_{\mathrm{int}(\boldsymbol{\gamma})} K d\mathcal{A}_{\boldsymbol{\sigma}} \\ &= \sum_{i=1}^n \alpha_i - (n-2)\pi - \int_{\mathrm{int}(\boldsymbol{\gamma})} K d\mathcal{A}_{\boldsymbol{\sigma}}. \end{aligned}$$

To establish Eq. 13.12, we imagine 'smoothing' each vertex of $\boldsymbol{\gamma}$ as shown in the following diagram.



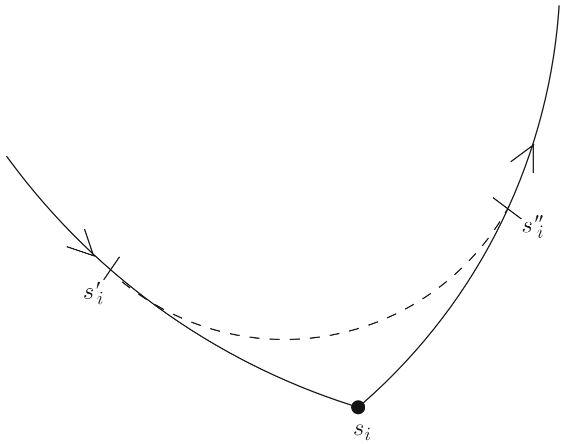
If the 'smoothed' curve $\tilde{\boldsymbol{\gamma}}$ is smooth (!), then, in an obvious notation,

$$\int_0^{\ell(\tilde{\boldsymbol{\gamma}})} \dot{\tilde{\theta}} ds = 2\pi. \tag{13.13}$$

Since $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ are the same except near the vertices of $\boldsymbol{\gamma}$, the difference

$$\int_0^{\ell(\tilde{\boldsymbol{\gamma}})} \dot{\tilde{\theta}} ds - \int_0^{\ell(\boldsymbol{\gamma})} \dot{\theta} ds \tag{13.14}$$

is a sum of $n$ contributions, one from near each vertex. Near $\boldsymbol{\gamma}(s_i)$, the picture is




i.e., $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ agree except when $s$ belongs to a small interval $(s_i', s_i'')$, say, containing $s_i$, so the contribution from the $i$th vertex is

$$\int_{s_i'}^{s_i''} \dot{\tilde{\theta}} ds - \int_{s_i'}^{s_i} \dot{\theta} ds - \int_{s_i}^{s_i''} \dot{\theta} ds.$$

The first integral is the angle between $\dot{\tilde{\boldsymbol{\gamma}}}(s_i'')$ and $\dot{\tilde{\boldsymbol{\gamma}}}(s_i')$, which as $s_i'$ and $s_i''$ tend to $s_i$ becomes the angle between $\dot{\boldsymbol{\gamma}}^+(s_i)$ and $\dot{\boldsymbol{\gamma}}^-(s_i)$, i.e., $\delta_i$. On the other hand, since $\boldsymbol{\gamma}(s)$ is smooth on each of the intervals $(s_i', s_i)$ and $(s_i, s_i'')$, the last two integrals go to zero as $s_i'$ and $s_i''$ tend to $s_i$. Thus, the contribution to the expression (13.14) from the $i$th vertex tends to $\delta_i$ as $s_i'$ and $s_i''$ tend to $s_i$ . Summing over all the vertices, we get

$$\int_0^{\ell(\tilde{\boldsymbol{\gamma}})} \dot{\tilde{\theta}} ds - \int_0^{\ell(\boldsymbol{\gamma})} \dot{\theta} ds = \sum_{i=1}^n \delta_i.$$

Equation 13.12 now follows from this and Eq. 13.13. □

## Corollary 13.2.3

If $\boldsymbol{\gamma}$ is a curvilinear polygon with $n$ edges each of which is an arc of a geodesic, then the internal angles $\alpha_1, \alpha_2, \ldots, \alpha_n$ of the polygon satisfy the equation

$$\sum_{i=1}^n \alpha_i = (n-2)\pi + \int_{\mathrm{int}(\boldsymbol{\gamma})} K d\mathcal{A}_{\boldsymbol{\sigma}}.$$

## Proof

This is immediate from Theorem 13.2.2, since $\kappa_g = 0$ along a geodesic. □

As a special case of Corollary 13.2.3, consider an $n$-gon in the plane with straight edges. Since $K = 0$ for the plane, Corollary 13.2.3 gives

$$\sum_{i=1}^n \alpha_i = (n-2)\pi,$$

a well-known result of elementary geometry.

For a curvilinear $n$-gon on $S^2$ whose sides are arcs of great circles, we have $K = 1$ so $\sum \alpha_i$ exceeds the plane value $(n-2)\pi$ by the area $\int d\mathcal{A}_{\boldsymbol{\sigma}}$ of the polygon. Taking $n = 3$, we get for the area of a spherical triangle with angles $\alpha, \beta, \gamma$ whose edges are arcs of great circles,

$$\mathcal{A} = \alpha + \beta + \gamma - \pi.$$

This is just Theorem 6.4.7, which is therefore a special case of Gauss–Bonnet.

Finally, for a geodesic $n$-gon on the pseudosphere (see Section 8.3), for which $K = -1$, we see that $\sum \alpha_i$ is *less* than $(n-2)\pi$ by the area of the polygon:

$$\mathcal{A} = (n-2)\pi - \alpha_1 - \alpha_2 - \cdots - \alpha_n,$$

which is Theorem 11.1.5.

## EXERCISES

13.2.1 Consider the surface of revolution

$$\boldsymbol{\sigma}(u,v) = (f(u)\cos v, f(u)\sin v, g(u)),$$

where $\boldsymbol{\gamma}(u) = (f(u), 0, g(u))$ is a unit-speed curve in the $xz$-plane. Let $u_1 < u_2$ be constants, let $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ be the two parallels $u = u_1$ and $u = u_2$ on $\boldsymbol{\sigma}$, and let $R$ be the region of the $uv$-plane given by

$$u_1 \leq u \leq u_2, \quad 0 < v < 2\pi.$$

Compute

$$\int_0^{\ell(\gamma_1)} \kappa_g ds, \quad \int_0^{\ell(\gamma_2)} \kappa_g ds \quad \text{and} \quad \int_R K d\mathcal{A}_{\boldsymbol{\sigma}},$$

and explain your result on the basis of the Gauss–Bonnet theorem.

# 13.3 Integration on compact surfaces

The most important version of the Gauss–Bonnet theorem applies to compact surfaces. Before we prove it (in the next section), we must first discuss the integration of functions on such surfaces. Our approach is a little non-standard, and we shall not attempt to give complete proofs. This section can be omitted by readers willing to accept the existence of an integral on $\mathcal{S}$ that has reasonable properties.

Suppose that $f : \mathcal{S} \to \mathbb{R}$ is a smooth function on a compact surface $\mathcal{S}$. If $f$ vanishes outside some set that is contained in a single surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$, we can define the integral of $f$ by using the formulas from Section 6.4:

$$\int_{\mathcal{S}} f \, d\mathcal{A} = \int_U f(\boldsymbol{\sigma}(u,v)) \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel dudv. \tag{13.15}$$

However, if $f$ does not have this property, another method of defining the integral of $f$ must be found.

The idea is to 'thicken' the surface $\mathcal{S}$ to obtain an open subset $V$ of $\mathbb{R}^3$ that contains $\mathcal{S}$. The function $f$ is extended to a function on $V$ by defining the extension to be constant in directions perpendicular to the surface. Then, the volume integral

$$\iiint_V f(x,y,x) \, dxdydz$$

should be approximately the 'thickness' of $V$ multiplied by the integral of $f$ over the surface $\mathcal{S}$.

Recall first that $\mathcal{S}$ is orientable (Corollary 5.4.5), so there is a smooth choice of unit normal $\mathbf{N_p}$ at every point $\mathbf{p} \in \mathcal{S}$. Let $\epsilon > 0$ and define $\mathcal{S}^{(-\epsilon,\epsilon)}$ to be the set of points of $\mathbb{R}^3$ of the form

$$\mathbf{p} + t\mathbf{N_p}, \qquad \mathbf{p} \in \mathcal{S},\ t \in (-\epsilon, \epsilon),$$

so that $\mathcal{S}^{(-\epsilon,\epsilon)}$ has 'thickness' $2\epsilon$. (Thus, $\mathcal{S}^{(-\epsilon,\epsilon)}$ is the region of $\mathbb{R}^3$ contained between two parallel surfaces of $\mathcal{S}$ – see Definition 8.5.1.) It can be shown that:

(i) $\mathcal{S}^{(-\epsilon,\epsilon)}$ is an open subset of $\mathbb{R}^3$.

(ii) If $\epsilon$ is sufficiently small, the map $\boldsymbol{\sigma}^{(-\epsilon,\epsilon)} : (\mathbf{p}, t) \mapsto \mathbf{p} + t\mathbf{N_p}$ is injective ($\mathbf{p} \in \mathcal{S}$, $t \in (-\epsilon, \epsilon)$).

By (ii), we can define a function $f^{(-\epsilon,\epsilon)} : \mathcal{S}^{(-\epsilon,\epsilon)} \to \mathbb{R}$ by

$$f^{(-\epsilon,\epsilon)}(\mathbf{p} + t\mathbf{N_p}) = f(\mathbf{p}).$$

Note that $f^{(-\epsilon,\epsilon)}$ agrees with $f$ at points of $\mathcal{S}$, and $f^{(-\epsilon,\epsilon)}$ is constant along each line perpendicular to $\mathcal{S}$. Then, we define the integral of $f$ over the surface $\mathcal{S}$ to be

$$\int_{\mathcal{S}} f\, d\mathcal{A} = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{\mathcal{S}^{(-\epsilon,\epsilon)}} f^{(-\epsilon,\epsilon)}(x, y, z)\, dxdydz. \tag{13.16}$$

We must now do two things: show that this definition agrees with the expected formula (13.15) when $f$ vanishes outside some surface patch, and show that the limit in the definition (13.16) exists for any $f$.

Suppose first that $f$ vanishes outside some set that is contained in a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ of $\mathcal{S}$. Then, by the change of variable formula for triple integrals,

$$\begin{aligned} &\int_{\boldsymbol{\sigma}(U)^{(-\epsilon,\epsilon)}} f^{(-\epsilon,\epsilon)}(x, y, z)\, dxdydz \\ &\quad = \int_{U \times (-\epsilon,\epsilon)} f^{(-\epsilon,\epsilon)}(\boldsymbol{\sigma}(u, v) + t\mathbf{N}(u, v))\, |D(u, v, t)|\, dudvdt, \end{aligned} \tag{13.17}$$

where $D$ is the determinant of the Jacobian matrix $J$ of $\boldsymbol{\sigma}^{(-\epsilon,\epsilon)}$. The rows of $J$ are the components of the vectors $\boldsymbol{\sigma}_u + t\mathbf{N}_u$, $\boldsymbol{\sigma}_v + t\mathbf{N}_v$ and $\mathbf{N}$, so

$$D = \mathbf{N} \cdot ((\boldsymbol{\sigma}_u + t\mathbf{N}_u) \times (\boldsymbol{\sigma}_v + t\mathbf{N}_v))$$

(see Exercise 13.3.1). Hence,

$$D = \mathbf{N} \cdot (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) + tg(u, v) + t^2 h(u, v) = \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel + tg(u, v) + t^2 h(u, v),$$

where $g$ and $h$ are smooth functions. If $\epsilon$ (and hence $t$) are very small, the integrand on the right-hand side of Eq. 13.17 is very nearly equal to

$$f^{(-\epsilon,\epsilon)}(\boldsymbol{\sigma}(u,v)) \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel = f(\boldsymbol{\sigma}(u,v)) \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel .$$

Thus, the right-hand side of Eq. 13.17 becomes

$$2\epsilon \int_U f(\boldsymbol{\sigma}(u,v)) \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel \, dudv.$$

Dividing by $2\epsilon$ and letting $\epsilon$ tend to zero (which makes the approximations we have made into equalities), we obtain the expression in Eq. 13.15.

This argument makes plausible that the limit in Eq. 13.16 exists, and has the expected value, when the function $f$ vanishes outside some surface patch. To deal with the case of an arbitrary smooth function $f : \mathcal{S} \to \mathbb{R}$, one makes use of a 'partition of unity'. This is a finite set of smooth functions $\varphi_1, \varphi_2, \ldots, \varphi_N$ on $\mathcal{S}$ such that:

(i) Each $\varphi_k$ is $\geq 0$ everywhere and vanishes outside some surface patch.

(ii) $\varphi_1 + \varphi_2 + \cdots + \varphi_N = 1$.

We shall not describe how to construct such functions here, but the idea is indicated in Exercise 13.3.2 which treats a one-dimensional analogue. Assuming that we have a partition of unity, we can write $f$ as a finite sum

$$f = f\varphi_1 + f\varphi_2 + \cdots + f\varphi_N$$

of smooth functions, each of which vanishes outside some surface patch. Then,

$$\begin{aligned}\lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{\mathcal{S}^{(-\epsilon,\epsilon)}} & f^{(-\epsilon,\epsilon)}(x,y,z)\, dxdydz \\ &= \sum_{k=1}^{N} \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{\mathcal{S}^{(-\epsilon,\epsilon)}} (f\varphi_k)^{(-\epsilon,\epsilon)}(x,y,z)\, dxdydz.\end{aligned}$$

Since each of the limits in the sum on the right-hand side is already known to exist, so does the limit on the left-hand side.

## *EXERCISES*

13.3.1 Show that, if a $3 \times 3$ matrix $A$ has the vectors $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ as rows, then

$$\det(A) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}).$$

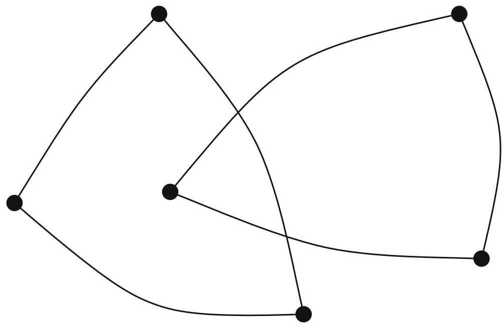
13.3.2 Let $n$ be a positive integer. Show that there are smooth functions $\varphi_1, \varphi_2, \dots, \varphi_{n-1}$ such that

(i) $\varphi_k(t) > 0$ for $\frac{k-1}{n} < t < \frac{k+1}{n}$ and $\varphi_k(t) = 0$ otherwise;

(ii) $\varphi_1(t) + \varphi_2(t) + \cdots + \varphi_{n-1}(t) = 1$ for all $0 < t < 1$.

## 13.4 Gauss–Bonnet for compact surfaces

The most important version of the Gauss–Bonnet theorem applies to a compact surface $\mathcal{S}$. It is obtained by covering a compact surface $\mathcal{S}$ with curvilinear polygons that fit together properly, applying Theorem 13.2.2 to each one, and adding up the results. We begin to make this more precise with the following definition.



### Definition 13.4.1

Let $\mathcal{S}$ be a surface, with atlas consisting of the patches $\boldsymbol{\sigma}_i : U_i \to \mathbb{R}^3$. A *triangulation* of $\mathcal{S}$ is a collection of curvilinear polygons, each of which is contained, together with its interior, in one of the $\boldsymbol{\sigma}_i(U_i)$, such that:

(i) Every point of $\mathcal{S}$ is in at least one of the curvilinear polygons.

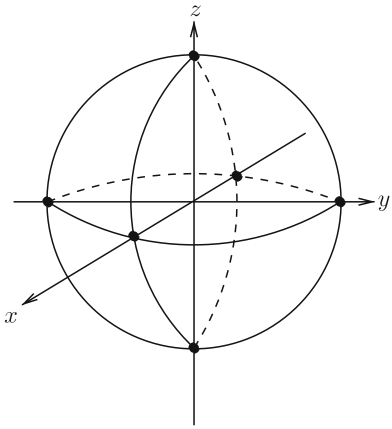
(ii) Two curvilinear polygons are either disjoint, or their intersection is a common edge or a common vertex.

(iii) Each edge is an edge of exactly two polygons.

Thus, situations like that shown above are not allowed.

## Example 13.4.2

A triangulation of $S^2$ with eight polygons is obtained by intersecting it with the three coordinate planes:




We state without proof:

## Theorem 13.4.3

Every compact surface has a triangulation with finitely many polygons.

We introduce the following number associated to any triangulation:

## Definition 13.4.4

The *Euler number* $\chi$ of a triangulation of a compact surface $\mathcal{S}$ with finitely many polygons is

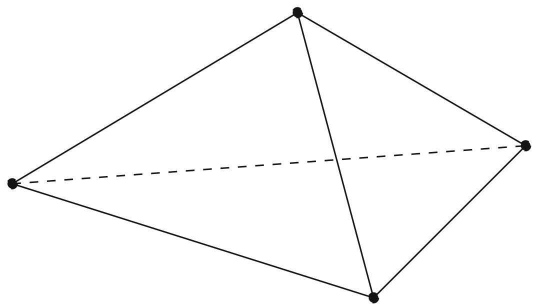
$$\chi = V - E + F,$$

where

$$\begin{aligned} V &= \text{ the total number of vertices of the triangulation,} \\ E &= \text{ the total number of edges of the triangulation,} \\ F &= \text{ the total number of polygons of the triangulation.} \end{aligned}$$

For the triangulation of the sphere given above, $V = 6$, $E = 12$ and $F = 8$, and so $\chi = 6 - 12 + 8 = 2$.

The importance of the Euler number is that, although different triangulations of a given surface will in general have different numbers of vertices, edges and polygons, $\chi$ is actually *independent* of the triangulation and depends only on the surface. For example, we can get another triangulation of $S^2$ by 'inflating' a regular tetrahedron:



This time, $V = 4$, $E = 6$ and $F = 4$, and so $\chi = 4 - 6 + 4 = 2$, the same as before. This property of $\chi$ is a consequence of the following theorem.

### Theorem 13.4.5

Let $\mathcal{S}$ be a compact surface. Then, for any triangulation of $\mathcal{S}$,

$$\int_{\mathcal{S}} K d\mathcal{A} = 2\pi\chi,$$

where $\chi$ is the Euler number of the triangulation.

Since $\int_{\mathcal{S}} K d\mathcal{A}$ is independent of the triangulation, Theorem 13.4.5 implies

### Corollary 13.4.6

The Euler number $\chi$ of a triangulation of a compact surface $\mathcal{S}$ depends only on $\mathcal{S}$ and not on the choice of triangulation.

We now give the proof of Theorem 13.4.5.

### Proof

By Corollary 5.4.5, there is a smooth choice of unit normal $\mathbf{N}$ at every point of $\mathcal{S}$. As above, we fix a triangulation of $\mathcal{S}$ with polygons $P_i$, say, each of which is contained in the image of some patch $\boldsymbol{\sigma}_i : U_i \to \mathbb{R}^3$ in the atlas of $\mathcal{S}$, say $P_i = \boldsymbol{\sigma}_i(R_i)$, where $R_i \subseteq U_i$; we can assume that $\mathbf{N}$ is the standard unit normal of each $\boldsymbol{\sigma}_i$. By Theorem 13.2.2,

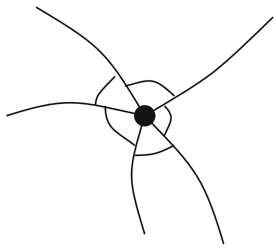
$$\int_{R_i} K d\mathcal{A}_{\boldsymbol{\sigma}_i} = \angle_i - (n_i - 2)\pi + \int_0^{\ell(\boldsymbol{\gamma}_i)} \kappa_g \, ds, \tag{13.18}$$

where $n_i$ is the number of vertices of $P_i$, $\boldsymbol{\gamma}_i$ is the curvilinear polygon that forms the boundary of $P_i$, $\ell(\boldsymbol{\gamma}_i)$ is its length, and $\angle_i$ is the sum of its interior angles. We must therefore sum the contributions of each of the three terms on the right-hand side of Eq. 13.18 over all the polygons $P_i$ in the triangulation.

First, $\sum_i \angle_i$ is the sum of all the internal angles of all the polygons. At each vertex, several polygons meet, but the sum of the angles at the vertex is obviously $2\pi$, so

$$\sum_i \angle_i = 2\pi V, \tag{13.19}$$

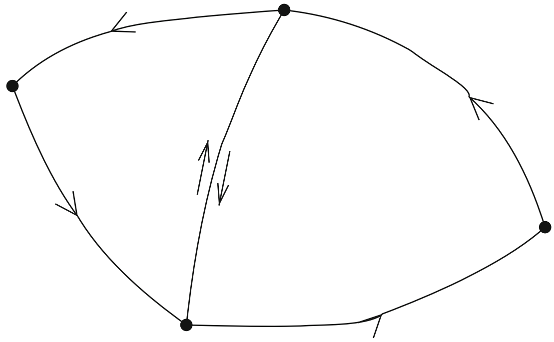
where $V$ is the total number of vertices.



Next,

$$\sum_i (n_i - 2)\pi = \left(\sum_i n_i\right)\pi - 2\pi F = 2\pi E - 2\pi F, \tag{13.20}$$

where $F$ is the total number of polygons and $E$ the total number of edges, since in the sum $\sum_i n_i$ each edge is counted twice (as each edge is an edge of exactly two polygons).



We now claim that

$$\sum_i \int_0^{\ell(\boldsymbol{\gamma}_i)} \kappa_g ds = 0. \tag{13.21}$$

Indeed, note that in the sum in Eq. 13.21, we integrate twice along each edge, once in each direction (see the diagram above). By Proposition 7.3.2, the geodesic curvature $\kappa_g = \ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}})$ of a curve $\boldsymbol{\gamma}$ changes sign when we traverse $\boldsymbol{\gamma}$ in the opposite direction (for this amounts to changing the sign of the curve parameter which changes the sign of $\dot{\boldsymbol{\gamma}}$ but leaves $\ddot{\boldsymbol{\gamma}}$ unchanged). Hence, the two integrals in (13.21) along any given edge cancel out. The various contributions to the sum in Eq. 13.21 therefore cancel out in pairs, thus proving Eq. 13.21.

Finally, the integral of the Gaussian curvature $K$ over the whole surface is the sum of its integrals over each polygon. This is because the polygons overlap, if at all, in an edge or a vertex, so that the overlap has zero area and thus does not contribute to the integral. Hence, putting Eqs. 13.18–13.21 together gives

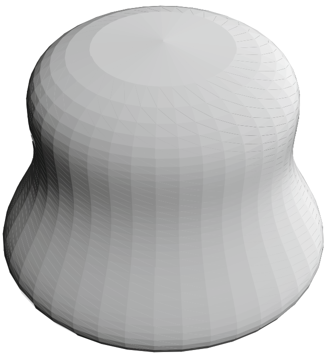
$$\begin{aligned}\int_{\mathcal{S}} K d\mathcal{A} &= \sum_i \int_{R_i} K d\mathcal{A}_{\boldsymbol{\sigma}_i} \\ &= \sum_i \angle_i - \sum_i (n_i - 2)\pi + \sum_i \int_0^{\ell(\boldsymbol{\gamma}_i)} \kappa_g \, ds \\ &= 2\pi V - (2\pi E - 2\pi F) + 0 \\ &= 2\pi\chi,\end{aligned}$$

proving Theorem 13.4.5. □

To see why Theorem 13.4.5 is so remarkable, let us apply it to the unit sphere $S^2$. Then, $\chi = 2$ and so we get

$$\int_{S^2} K d\mathcal{A} = 4\pi. \tag{13.22}$$

Of course, this result is not remarkable at all because $K = 1$ so the left-hand side of Eq. 13.22 is just the area of $S^2$. But now suppose that we deform $S^2$, i.e., we think of $S^2$ as being a rubber sheet and we pull and stretch it in any way we like, but without tearing, producing a new surface $\mathcal{S}$:



The Gaussian curvature $K$ of $\mathcal{S}$ will not be constant and the direct computation of the integral $\int_{\mathcal{S}} K d\mathcal{A}$ will be difficult. But if we start with a triangulation of $S^2$, then after deformation we shall have a triangulation of $\mathcal{S}$ *with the same* *number of vertices, edges and polygons as the original triangulation.* It follows that the Euler number of $\mathcal{S}$ is *the same* as that of $S^2$, i.e., 2, so by Theorem 13.4.5, $\int_{\mathcal{S}} K d\mathcal{A} = 4\pi$. (More generally, this discussion shows that *any* two diffeomorphic compact surfaces have the same Euler number.)

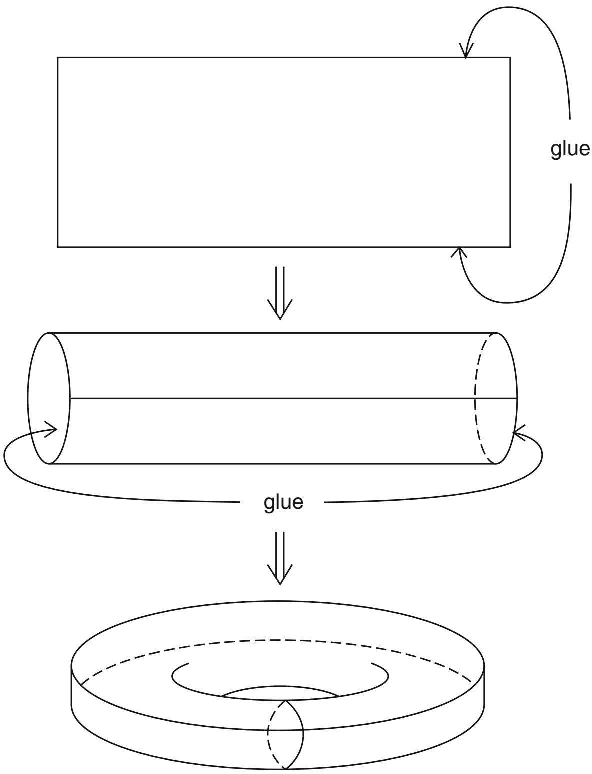
We complete the picture by determining the Euler numbers of all the compact surfaces, which were described in Section 5.4.

## Theorem 13.4.7

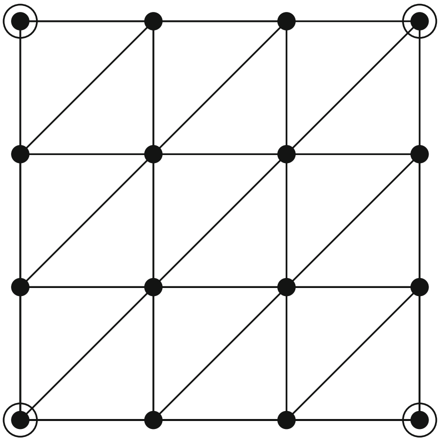
The Euler number of the compact surface $T_g$ of genus $g$ is $2 - 2g$.

## Proof

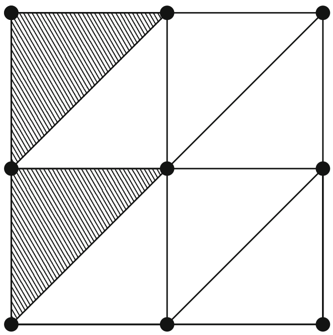
The formula is correct when $g = 0$, since we know that $\chi = 2$ for a sphere. We now prove it for the torus $T_1$. To find a triangulation of the torus, we use the fact that it can be obtained from a square in the plane by gluing opposite edges:



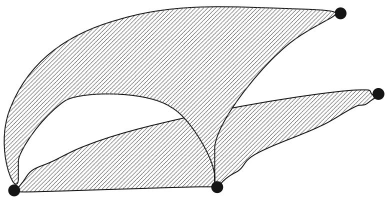
We subdivide the square into triangles as shown:



This leads to a triangulation of $T_1$ with $V = 9$, $E = 27$ and $F = 18$. One must count carefully: for example, the four circled vertices of the square correspond to a single vertex on the torus. Note also that not just any subdivision of the square into triangles is acceptable. For example, the subdivision



is *not* acceptable, since after gluing, the two shaded triangles intersect in *two* vertices, which is not allowed:

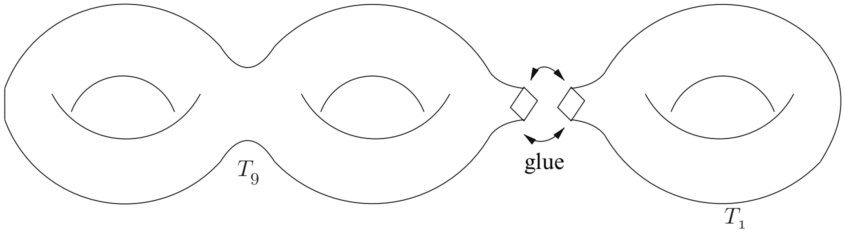


But the finer subdivision above does work, and gives

$$\chi = 9 - 27 + 18 = 0 = 2 - 2 \times 1,$$

proving the theorem when $g = 1$.

We now complete the proof by induction on $g$, using the fact that $T_{g+1}$ can be obtained from $T_g$ by 'gluing on' a copy of $T_1$:




Suppose we carry out the gluing by removing a curvilinear $n$-gon from $T_g$ and $T_1$ and gluing corresponding edges (having fixed suitable triangulations of $T_g$ and $T_1$). If $V'$, $E'$ and $F'$ are the numbers of vertices, edges and polygons in the triangulation of $T_g$, and $V''$, $E''$ and $F''$ are those for $T_1$, the numbers $V$, $E$ and $F$ for $T_{g+1}$ are given by

$$\begin{aligned} V &= V' - n + V'' - n + n = V' + V'' - n, \\ E &= E' - n + E'' - n + n = E' + E'' - n, \\ F &= F' - 1 + F'' - 1 = F' + F'' - 2. \end{aligned}$$

Indeed, $V$ is the number $V'$ of vertices in $T_g$ plus the number $V''$ in $T_1$, except that the $n$ vertices of the polygon along which $T_1$ and $T_g$ are glued have been counted twice, so $V = V' + V'' - n$; a similar argument applies to the edges; and $F$ is as stated because the polygon along which $T_1$ and $T_g$ are glued is not part of the triangulation of $T_{g+1}$. Hence,

$$\begin{aligned} \chi(T_{g+1}) &= V - E + F \\ &= (V' + V'' - n) - (E' + E'' - n) + (F' + F'' - 2) \\ &= V' - E' + F' + V'' - E'' + F'' - 2 \\ &= \chi(T_g) + \chi(T_1) - 2 \\ &= 2 - 2g + 0 - 2 \quad \text{(by the induction hypothesis)} \\ &= 2 - 2(g+1), \end{aligned}$$

proving the result for genus $g + 1$. $\square$

### Corollary 13.4.8

We have

$$\int_{T_g} K d\mathcal{A} = 4\pi(1-g).$$

### Proof

Just combine Theorems 13.4.5 and 13.4.7. □

### *EXERCISES*

13.4.1 Show that, if a compact surface $\mathcal{S}$ is diffeomorphic to the torus $T_1$, then

$$\int_{\mathcal{S}} K\, d\mathcal{A} = 0$$

(cf. Exercise 8.1.8). Can such a surface $\mathcal{S}$ have $K = 0$ everywhere?

13.4.2 Suppose that $\mathcal{S}$ is a compact surface whose Gaussian curvature $K$ is $> 0$ everywhere. Show that $\mathcal{S}$ is diffeomorphic to a sphere. Is the converse of this statement true?

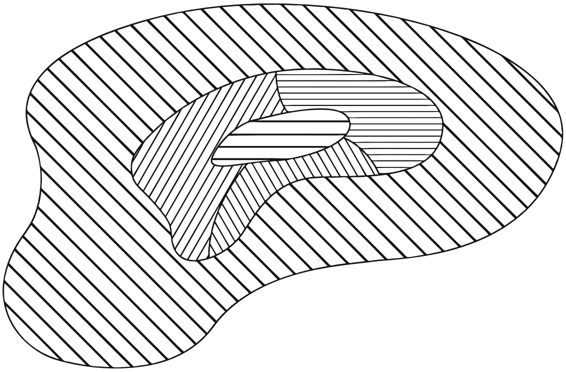
## 13.5 Map colouring

In the remainder of this book we shall give some applications of the Gauss–Bonnet theorem. The first is actually an application of Euler's formula

$$V - E + F = \chi = 2 - 2g$$

relating the number of vertices $V$, edges $E$ and polygons $F$ in any triangulation of a compact surface of Euler number $\chi$ and genus $g$. This is, of course, an immediate consequence of Corollary 13.4.6 and Theorem 13.4.7.

The application we have in mind is the problem of *map colouring*. In this context, a triangulation of $\mathcal{S}$ is called a 'map' (in the usual geographical sense) and the polygons are called 'countries'. Two countries are 'neighbours' if they have a common edge (not just a common vertex). If $n$ is a positive integer, an *$n$-colouring* of such a map is an assignment of one of $n$ different colours to each country in such a way that neighbouring countries never have the same colour.

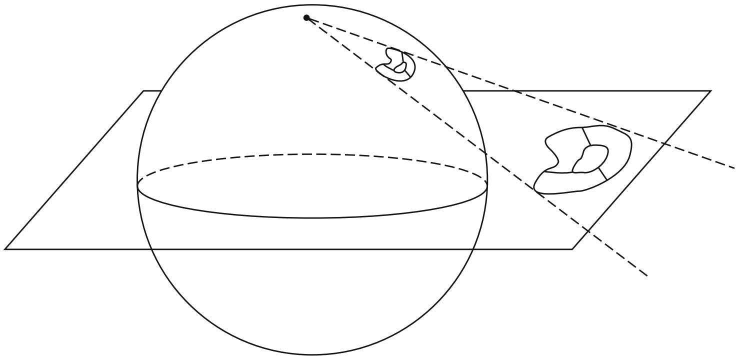


**Map Colour ng Pr** *For a given compact surface $\mathcal{S}$, what is the smallest positive integer $n$ such that every map on $\mathcal{S}$ can be $n$-coloured?*

This smallest integer $n$ is called the *chromatic number* of $\mathcal{S}$. (It is not obvious that this number even exists – why should we not be able to draw maps on a given surface requiring an arbitrarily large number of colours? – but it does, as we shall show later in this section.) It is clear that diffeomorphic surfaces have the same chromatic number.

The problem originated in a letter sent by the English mathematician Augustus de Morgan to the Irish mathematician Sir William Rowan Hamilton (the inventor of quaternions) on 23 October 1852. As De Morgan wrote, "A student of mine asked me today to give him a reason for a fact which I did not know was a fact – and do not yet. He says that if a figure be anyhow divided and the compartments differently coloured so that figures with any portion of common boundary line are differently coloured – four colours may be wanted, but not more ..."

The problem formulated by De Morgan, or his student (Peter Guthrie), was actually in the *plane* – the map was supposed to have finitely many countries and to occupy a certain bounded region. Of course, the plane is not a compact surface, but the problem described by De Morgan is equivalent to the map colouring problem for a *sphere*. This is easily seen using the stereographic projection in Example 6.3.5 (we need only consider $S^2$ as any two spheres are diffeomorphic). Suppose that we have a map on $S^2$, and assume that the north pole is not on an edge – this can always be achieved by a suitable rotation of $S^2$. Then stereographic projection takes the map on $S^2$ to one on the plane and vice versa; note that the country on $S^2$ containing the north pole goes to the (unbounded) exterior region of the map on the plane, which therefore has one fewer country than that on $S^2$.



Thus, De Morgan (or Guthrie) had made

**The Four Colour Conjecture** *The chromatic number of a sphere is* 4.

For surfaces other than a sphere, the corresponding result was conjectured by Percy Heawood in 1890. For a compact surface of Euler number $\chi$, let

$$N(\chi) = \frac{1}{2}(7 + \sqrt{49 - 24\chi}).$$

Note that, since the possible values of $\chi$ are $2, 0, -2, -4, \ldots$, $N(\chi)$ is a positive real number. Let $h(\chi)$ be the largest integer $\leq N(\chi)$.

**Heawood's Conjecture** *The chromatic number of a compact surface of Euler number $\chi \leq 0$ is $h(\chi)$.*

Since $h(2) = 4$ the formula also gives the conjectured answer for a sphere.

The proofs of these conjectures turned out to be very difficult and, rather surprisingly, the problem for the 'simplest' surface, the sphere, is *more* difficult than that for higher genus surfaces: Heawood's conjecture was proved in 1967 and the four colour conjecture 10 years later (the latter proof requires substantial computer assistance).

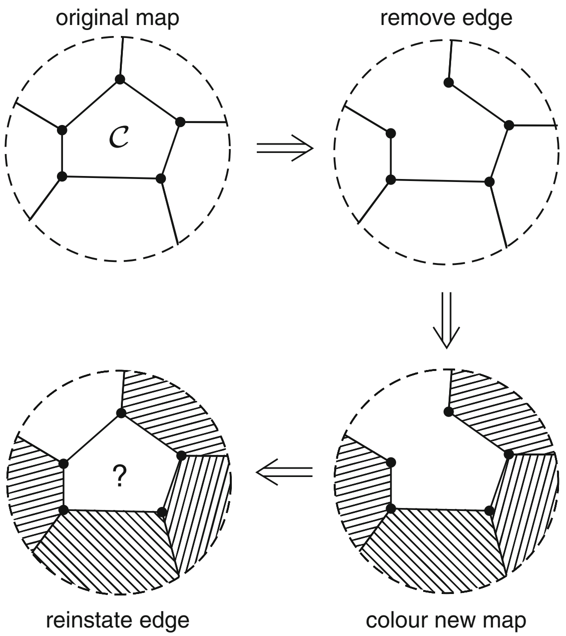
Euler's formula leads to a proof of 'half' of Heawood's conjecture:

## Theorem 13.5.1

Any compact surface of Euler number $\chi \leq 0$ can be $h(\chi)$-coloured.

## Proof

We can assume that at least three edges meet at each vertex, since a vertex at which only two edges meet can be removed without affecting the colouring. We can also assume that the number of countries $F$ is greater than $N(\chi)$, for if $F \leq N(\chi)$ then $F \leq h(\chi)$ and then $h(\chi)$ colours obviously suffice.




To prove the theorem, it is enough to prove that *at least one country* $\mathcal{C}$ *has* $\leq h(\chi) - 1$ *edges*. Indeed, if we assume this we can prove the theorem by induction on $F$. Let us remove an edge from $\mathcal{C}$, thus merging $\mathcal{C}$ with one of its neighbours. This creates a new map with $F-1$ countries. By the inductive assumption, this map can be $h(\chi)$-coloured. Choose such a colouring and then reinstate $\mathcal{C}$ by replacing the removed edge. Since $\mathcal{C}$ has at most $h(\chi)-1$ neighbours, we can colour $\mathcal{C}$ with a colour different from any of those used for its neighbours. This gives an $h(\chi)$-colouring of our original map and completes the induction.

We prove the assertion by contradiction. Suppose then that every country has $\geq h(\chi)$ edges. Since every edge is an edge of exactly two countries, we have

$$E \geq \frac{1}{2}h(\chi)F, \tag{13.23}$$

and since at least three edges meet at each vertex,

$$2E \geq 3V.$$

From this last inequality,

$$E \leq 3(E-V) = 3(F-\chi). \tag{13.24}$$

From (13.23) and (13.24),

$$h(\chi) \leq 6\left(1 - \frac{\chi}{F}\right).$$

Since $F > N(\chi)$ and $\chi \leq 0$, we get

$$h(\chi) \leq 6\left(1 - \frac{\chi}{N(\chi)}\right). \tag{13.25}$$

But $N(\chi)$ is a root of the quadratic equation

$$N^2 - 7N + 6\chi = 0,$$

from which we deduce that

$$6\left(1 - \frac{\chi}{N(\chi)}\right) = N(\chi) - 1.$$

Thus, the inequality (13.25) is impossible.

This contradiction completes the proof of Theorem 13.5.1. □

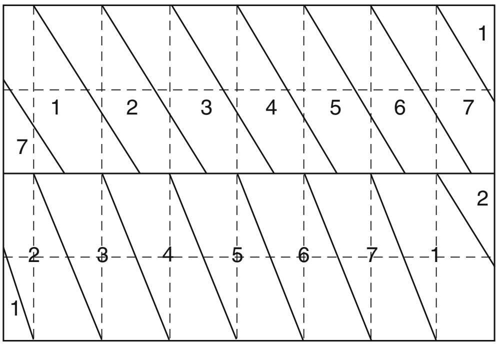
To complete the proof of Heawood's conjecture, we would have to exhibit, for every $\chi \leq 0$, a map on a surface of Euler number $\chi$ that cannot be coloured with fewer than $h(\chi)$ colours. Heawood apparently thought that this would be straightforward, but in fact it is the most difficult part of the proof. He did, however, give such a map on the torus, and so proved

## Theorem 13.5.2

The chromatic number of a torus is 7.

## Proof

Consider the map on the torus shown below, in which we use the description of the torus as a rectangle with opposite edges glued together (see the proof of Theorem 13.4.7):




Each of the seven countries of this map has all the others as neighbours, so it obviously requires seven colours. □

The proof of Theorem 13.5.1 breaks down if $\chi = 2$, and so does not help with the proof of the four colour conjecture. However, the same method can be used to prove

### Proposition 13.5.3 (Five Neighbours Theorem)

Every map on a sphere has at least one country with five or fewer neighbours.

See Exercise 13.5.1. If we take a map on a sphere, we can remove and then reinstate a country $\mathcal{C}$ which has no more than five neighbours, as in the proof of Theorem 13.4.1, but then we need at least six colours to be available to be sure of being able to colour $\mathcal{C}$ differently from its neighbours. Thus, Proposition 13.5.3 implies the following theorem.

### Corollary 13.5.4 (Six Colour Theorem)

Every map on a sphere can be six-coloured.

### *EXERCISES*

13.5.1 Prove Proposition 13.5.3.

13.5.2 Show that every triangulation of a compact surface of Euler number $\chi$ by curvilinear triangles has at least $N(\chi)$ vertices.

## 13.6 Holonomy and Gaussian curvature

Theorem 13.1.2 allows us to establish a connection between parallel transport along curves on a surface (see Section 7.4) and the Gaussian curvature of the surface. For this, we need

### Proposition 13.6.1

Let $\boldsymbol{\gamma}$ be a unit-speed curve on a surface patch $\boldsymbol{\sigma}$ and let $\mathbf{v}$ be a non-zero parallel vector field along $\boldsymbol{\gamma}$. Let $\varphi$ be the oriented angle $\widehat{\dot{\boldsymbol{\gamma}}\mathbf{v}}$ from $\dot{\boldsymbol{\gamma}}$ to $\mathbf{v}$. Then, the geodesic curvature of $\boldsymbol{\gamma}$ is

$$\kappa_g = -\frac{d\varphi}{ds}.$$

Note that, if $\mathbf{v}$ were replaced by another non-zero parallel vector field $\mathbf{w}$, then since the angle between $\mathbf{v}$ and $\mathbf{w}$ is constant by Proposition 7.4.9(ii), $\varphi$ would change by the addition of a constant and $d\varphi/ds$ would be unchanged.

## Proof

By Proposition 7.4.9(ii), the length of $\mathbf{v}$ is constant so by multiplying by a non-zero constant we can assume that $\parallel \mathbf{v} \parallel = 1$. Then,

$$\mathbf{v} = \cos\varphi\,\mathbf{t} + \sin\varphi\,\mathbf{t}_1,$$

where $\mathbf{t} = \dot{\boldsymbol{\gamma}}$ and $\mathbf{t}_1 = \mathbf{N} \times \mathbf{t}$, $\mathbf{N}$ being the unit normal of $\boldsymbol{\sigma}$. Since $\mathbf{v}$ is parallel along $\boldsymbol{\gamma}$, $\dot{\mathbf{v}}$ is parallel to $\mathbf{N}$ (Proposition 7.4.3). Hence,

$$0 = \mathbf{t} \cdot \dot{\mathbf{v}} = \mathbf{t} \cdot ((\cos\varphi\,\dot{\mathbf{t}} + \sin\varphi\,\dot{\mathbf{t}}_1) + \dot{\varphi}(-\sin\varphi\,\mathbf{t} + \cos\varphi\,\mathbf{t}_1)) = \sin\varphi(\mathbf{t} \cdot \dot{\mathbf{t}}_1 - \dot{\varphi}), \tag{13.26}$$

since $\mathbf{t} \cdot \dot{\mathbf{t}} = \mathbf{t} \cdot \mathbf{t}_1 = 0$. Similarly, $\mathbf{t}_1 \cdot \dot{\mathbf{v}} = 0$ leads to

$$0 = \cos\varphi(\mathbf{t}_1 \cdot \dot{\mathbf{t}} + \dot{\varphi}). \tag{13.27}$$

Now,

$$\kappa_g = \ddot{\gamma} \cdot (\mathbf{N} \times \dot{\gamma}) = \dot{\mathbf{t}} \cdot (\mathbf{N} \times \mathbf{t}) = \dot{\mathbf{t}} \cdot \mathbf{t}_1,$$

and since $\mathbf{t} \cdot \mathbf{t}_1 = 0$,

$$\mathbf{t} \cdot \dot{\mathbf{t}}_1 = -\dot{\mathbf{t}} \cdot \mathbf{t}_1 = -\kappa_g.$$

Hence, Eqs. 13.26 and 13.27 become

$$(\dot{\varphi} + \kappa_g)\sin\varphi = 0 = (\dot{\varphi} + \kappa_g)\cos\varphi,$$

and hence the result. Note that we really need both of these equations, since we could have $\cos\varphi = 0$ or $\sin\varphi = 0$ (but not both, of course). $\square$

Suppose now that $\boldsymbol{\gamma}$ is a (unit-speed) *closed* curve. On going once around $\boldsymbol{\gamma}$, $\varphi$ increases by

$$\int_0^{\ell(\boldsymbol{\gamma})} \frac{d\varphi}{ds} ds = -\int_0^{\ell(\boldsymbol{\gamma})} \kappa_g\, ds,$$

where $\ell(\boldsymbol{\gamma})$ is the length of $\boldsymbol{\gamma}$. If $\boldsymbol{\gamma}$ is actually a *simple* closed curve, the tangent vector $\dot{\boldsymbol{\gamma}}$ also rotates by $2\pi$ on going once around $\boldsymbol{\gamma}$ by the Umlaufsatz (Section 13.1). Hence, we obtain

## Proposition 13.6.2

Let $\boldsymbol{\gamma}$ be a positively-oriented unit-speed simple closed curve on a surface $\boldsymbol{\sigma}$, let $\kappa_g$ be the geodesic curvature of $\boldsymbol{\gamma}$, and let $\mathbf{v}$ be a non-zero parallel vector field along $\boldsymbol{\gamma}$. Then, on going once around $\boldsymbol{\gamma}$, $\mathbf{v}$ rotates through an angle

$$2\pi - \int_0^{\ell(\boldsymbol{\gamma})} \kappa_g \, ds. \tag{13.28}$$

## Definition 13.6.3

If $\boldsymbol{\gamma}$ is a unit-speed closed curve on a surface $\mathcal{S}$, the angle in Proposition 13.6.2 is called the *holonomy* around $\boldsymbol{\gamma}$, and is denoted by $h_{\boldsymbol{\gamma}}$.

The result we have been working towards is

## Theorem 13.6.4

Let $\boldsymbol{\gamma}$ be a positively-oriented simple closed curve on a surface patch $\boldsymbol{\sigma}$, let $h_{\boldsymbol{\gamma}}$ be the holonomy around $\boldsymbol{\gamma}$, and let $K$ be the Gaussian curvature of $\boldsymbol{\sigma}$. Then,

$$h_{\boldsymbol{\gamma}} = \int_{\mathrm{int}(\boldsymbol{\gamma})} K \, d\mathcal{A}_{\boldsymbol{\sigma}}.$$

## Proof

Just combine Theorem 13.1.2, Proposition 13.6.2 and Definition 13.6.3. $\square$

We can turn this theorem into a way of finding the Gaussian curvature at a point $\mathbf{p}$ of a surface $\mathcal{S}$: if $\boldsymbol{\gamma}$ is a small positively-oriented simple closed curve on the surface containing $\mathbf{p}$ in its interior, the Gaussian curvature of $\mathcal{S}$ at $\mathbf{p}$ will be approximately

$$\frac{h_{\boldsymbol{\gamma}}}{\mathrm{Area}(\mathrm{int}(\boldsymbol{\gamma}))}.$$

Using this idea, we can prove

## Proposition 13.6.5

Suppose that a surface $\mathcal{S}$ has the property that, for any two points $\mathbf{p}, \mathbf{q} \in \mathcal{S}$, the parallel transport $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}$ is independent of the curve $\boldsymbol{\gamma}$ joining $\mathbf{p}$ and $\mathbf{q}$. Then, $\mathcal{S}$ is flat.

### Proof

Suppose that parallel transport is independent of the curve joining any two given points. Then the holonomy around any *closed* curve on $\mathcal{S}$ must be zero. Indeed, if $\boldsymbol{\gamma}$ is such a curve, and if $\mathbf{p}$ is any point on $\boldsymbol{\gamma}$, then the parallel transport $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pp}}$ from $\mathbf{p}$ to $\mathbf{p}$ along $\boldsymbol{\gamma}$ must be the same as that from $\mathbf{p}$ to $\mathbf{p}$ along the constant curve at $\mathbf{p}$; but the latter is obviously the identity map. By Theorem 13.6.4, the integral of the Gaussian curvature $K$ over the interior of any simple closed curve $\boldsymbol{\gamma}$ on $\mathcal{S}$ must be zero. This implies that $K = 0$ everywhere: if $K \neq 0$ at some point $\mathbf{p} \in \mathcal{S}$, say $K(\mathbf{p}) > 0$, then $K > 0$ at all points in some open subset $\mathcal{O}$ of $\mathcal{S}$ containing $\mathbf{p}$; but then the integral of $K$ over the interior of a simple closed curve $\boldsymbol{\gamma}$ in $\mathcal{O}$ would be $> 0$, a contradiction. □

The converse of this proposition is not true without further assumptions (see Exercise 13.6.2). However, it is true if we assume that the surface $\mathcal{S}$ is *simply-connected*: this means that, for any two points $\mathbf{p}, \mathbf{q} \in \mathcal{S}$, and any two curves $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ joining $\mathbf{p}$ and $\mathbf{q}$, there is a family of curves $\boldsymbol{\gamma}^\tau$ joining $\mathbf{p}$ and $\mathbf{q}$ such that $\boldsymbol{\gamma}^0 = \boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^1 = \tilde{\boldsymbol{\gamma}}$ (see Section 9.4). For example, spheres are simply-connected but circular cylinders are not. We shall not prove these assertions as a proper discussion of simple-connectedness would take us too far into the realm of topology.

### *EXERCISES*

13.6.1 Let $\boldsymbol{\sigma}(\theta, \varphi)$ be the parametrization of the torus in Exercise 4.2.5. Show that the holonomy around a circle $\theta = \theta_0$ is $2\pi(1 - \sin\theta_0)$. Why is it obvious that the holonomy around a circle $\varphi =$ constant is $2\pi$? Note that these circles are *not* simple closed curves on the torus.

13.6.2 Calculate the holonomy around the parameter circle $v = 1$ on the cone $\boldsymbol{\sigma}(u, v) = (v\cos u, v\sin u, v)$, and conclude that the converse of Proposition 13.6.5 is false.

## 13.7 Singularities of vector fields

Suppose that $\mathcal{S}$ is a surface and that $\mathbf{V}$ is a *smooth tangent vector field* on $\mathcal{S}$. This means that, if $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a patch of $\mathcal{S}$ and $(u, v)$ are coordinates on $U$, then

$$\mathbf{V} = \alpha(u, v)\boldsymbol{\sigma}_u + \beta(u, v)\boldsymbol{\sigma}_v,$$

where $\alpha$ and $\beta$ are smooth functions on $U$. It is easy to see that this smoothness condition is independent of the choice of patch $\boldsymbol{\sigma}$ (see Exercise 13.7.2).

### Definition 13.7.1

If $\mathbf{V}$ is a smooth tangent vector field on a surface $\mathcal{S}$, a point $\mathbf{p} \in \mathcal{S}$ at which $\mathbf{V} = \mathbf{0}$ is called a *stationary point* of $\mathbf{V}$.

The reason for this terminology is as follows. We saw in the proof of Proposition 8.4.3 that, if $\mathbf{p} \in \mathcal{S}$, there is a unique curve $\boldsymbol{\gamma}(t)$ on $\mathcal{S}$ such that $\dot{\boldsymbol{\gamma}} = \mathbf{V}$ and $\boldsymbol{\gamma}(0) = \mathbf{p}$; $\boldsymbol{\gamma}$ is called an *integral curve* of $\mathbf{V}$. We can think of $\boldsymbol{\gamma}$ as the path followed by a particle of some fluid that is flowing over the surface. If $\mathbf{V} = \mathbf{0}$ at $\mathbf{p}$, the velocity $\dot{\boldsymbol{\gamma}}$ of the flow is zero at $\mathbf{p}$, so the fluid is stationary there.

We are going to prove a theorem which says that the number of stationary points of any smooth tangent vector field on a compact surface $\mathcal{S}$, counted with the appropriate multiplicity, is equal to the Euler number of $\mathcal{S}$. To define this multiplicity, let $\mathbf{p}$ be a stationary point of $\mathbf{V}$ contained in a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ of $\mathcal{S}$, say, with $\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$. Assume that $\mathbf{p}$ is the only stationary point of $\mathbf{V}$ in the region $\boldsymbol{\sigma}(U)$ of $\mathcal{S}$. Let $\boldsymbol{\xi}$ be a nowhere-vanishing smooth tangent vector field on $\boldsymbol{\sigma}(U)$ (e.g. we may choose $\boldsymbol{\xi} = \boldsymbol{\sigma}_u$ or $\boldsymbol{\sigma}_v$).

### Definition 13.7.2

With the above notation and assumption, the *multiplicity* of the stationary point $\mathbf{p}$ of the tangent vector field $\boldsymbol{V}$ is

$$\mu(\mathbf{p}) = \frac{1}{2\pi} \int_0^{\ell(\boldsymbol{\gamma})} \frac{d\psi}{ds}\, ds,$$

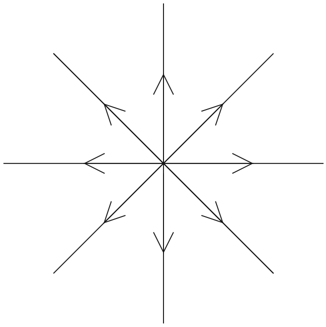
where $\boldsymbol{\gamma}(s)$ is any positively-oriented unit-speed simple closed curve of length $\ell(\boldsymbol{\gamma})$ in $\boldsymbol{\sigma}(U)$ with $\mathbf{p}$ in its interior, and $\psi(s)$ is the oriented angle $\widehat{\boldsymbol{\xi}\,\boldsymbol{V}}$ between $\boldsymbol{\xi}$ and $\boldsymbol{V}$ at the point $\boldsymbol{\gamma}(s)$.

It is clear that $\mu(\mathbf{p})$ is an integer, and an argument similar to our heuristic proof of Hopf's Umlaufsatz in Section 13.1 shows that $\mu(\mathbf{p})$ does not depend on the choice of simple closed curve $\boldsymbol{\gamma}$. It is also easy to see that it is independent of the choice of 'reference' vector field $\boldsymbol{\xi}$ (see Exercise 13.7.3). Finally, an argument similar to that used to prove Proposition 2.2.1 shows that the angle $\psi(s)$, which is only determined up to adding an integer multiple of $2\pi$, can be chosen to be a smooth function of $s$.

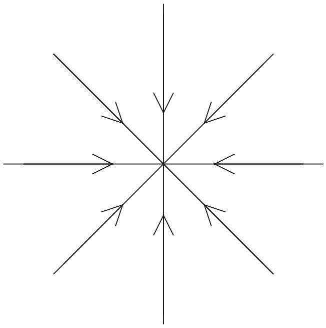
## Example 13.7.3

The following smooth tangent vector fields in the plane have stationary points of the indicated multiplicity at the origin (we have shown the integral curves of the vector fields for the sake of clarity):

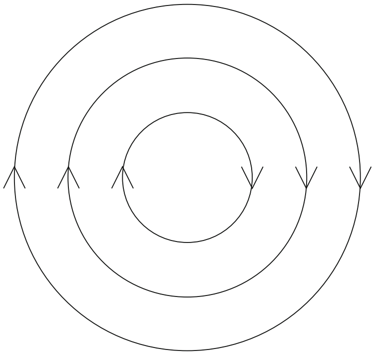
(i) $\mathbf{V}(x,y) = (x,y)$; $\mu = +1$.



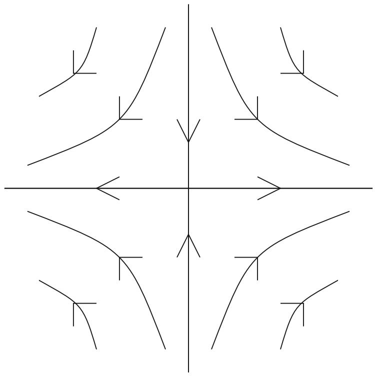
(ii) $\mathbf{V}(x,y) = (-x,-y)$; $\mu = +1$



(iii) $\mathbf{V}(x,y) = (y,-x)$; $\mu = +1$



(iv) $\mathbf{V}(x, y) = (x, -y)$; $\mu = -1$



The stationary point in examples (i), (ii), (iii) and (iv) is called a *source*, *sink*, *vortex* and *bifurcation*, respectively.

Let us verify the multiplicity in case (iv), for example. Take the 'reference' tangent vector field to be the constant vector field $\boldsymbol{\xi} = (1, 0)$. Then, the angle $\psi$ is given by

$$(\cos\psi, \sin\psi) = \frac{\mathbf{V}}{\parallel \mathbf{V} \parallel} = \left(\frac{x}{\sqrt{x^2+y^2}}, -\frac{y}{\sqrt{x^2+y^2}}\right).$$

Taking $\boldsymbol{\gamma}(s) = (\cos s, \sin s)$ to be the unit circle, at $\boldsymbol{\gamma}(s)$ the angle $\psi$ satisfies

$$(\cos\psi, \sin\psi) = (\cos s, -\sin s),$$

so $\psi = 2\pi - s$. Hence,

$$\mu(0,0) = \frac{1}{2\pi}\int_0^{2\pi} \frac{d}{ds}(2\pi - s)\, ds = -1.$$

## Theorem 13.7.4

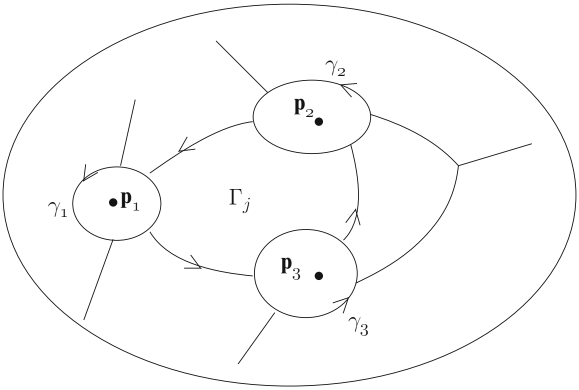
Let $\mathbf{V}$ be a smooth tangent vector field on a compact surface $\mathcal{S}$ which has only finitely many stationary points, say $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$. Then,

$$\sum_{i=1}^{n} \mu(\mathbf{p}_i) = \chi,$$

the Euler number of $\mathcal{S}$.

## Proof

Let $\boldsymbol{\gamma}_i$ be a positively-oriented unit-speed simple closed curve contained in a patch $\boldsymbol{\sigma}_i$ of $\mathcal{S}$ with $\mathbf{p}_i$ in the interior of $\boldsymbol{\gamma}_i$. Assume that the $\boldsymbol{\gamma}_i$ are chosen so small that their interiors are disjoint. Choose a triangulation of the part $\mathcal{S}'$ of $\mathcal{S}$ outside $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_n$ by curvilinear polygons $\boldsymbol{\Gamma}_j$. Note that the edges of some of these curvilinear polygons will be segments of the curves $\boldsymbol{\gamma}_i$.




Note also that, when these polygons are positively-oriented, the induced orientation of the $\boldsymbol{\gamma}_i$ is *opposite* to their positive orientation (see the diagram above, in which the arrows indicate the sense of positive orientation).

We can regard the curvilinear polygons in $\mathcal{S}'$, together with the simple closed curves $\boldsymbol{\gamma}_i$ and their interiors, as a triangulation of $\mathcal{S}$, so by Theorem 13.4.5,

$$\int_{\mathcal{S}'} K\, d\mathcal{A} + \sum_{i=1}^{n} \int_{\mathrm{int}(\boldsymbol{\gamma}_i)} K\, d\mathcal{A} = 2\pi\chi, \tag{13.29}$$

where $\chi$ is the Euler number of $\mathcal{S}$. On $\mathcal{S}'$, we choose an orthonormal basis $\{\boldsymbol{e}', \boldsymbol{e}''\}$ of the tangent plane of $\mathcal{S}$ at each point so that $\boldsymbol{e}'$ is parallel to the tangent vector field $\mathbf{V}$. Arguing as in the proof of Theorem 13.1.2, we see that

$$\int_{\mathcal{S}'} K\, d\mathcal{A} = \sum_j \int_0^{\ell(\boldsymbol{\Gamma}_j)} \boldsymbol{e}' \cdot \dot{\boldsymbol{e}}''\, ds, \tag{13.30}$$

where $s$ is arc-length on $\boldsymbol{\Gamma}_j$ and $\ell(\boldsymbol{\Gamma}_j)$ is its length. Any common edge of two of the curvilinear polygons $\boldsymbol{\Gamma}_j$ is traversed once in each direction and so their contributions to the sum in Eq. 13.30 cancel out. What remains is the integral along the segments of the curves $\boldsymbol{\gamma}_i$ that are part of the polygons $\boldsymbol{\Gamma}_j$. In view of the remark about orientations above, we get

$$\int_{\mathcal{S}'} K\, d\mathcal{A} = -\sum_{i=1}^{n} \int_0^{\ell(\boldsymbol{\gamma}_i)} \boldsymbol{e}' \cdot \dot{\boldsymbol{e}}''\, ds, \tag{13.31}$$

where $s$ is arc-length along $\boldsymbol{\gamma}_i$ and $\ell(\boldsymbol{\gamma}_i)$ is its length.

Now choose an orthonormal basis $\{\mathbf{f}', \mathbf{f}''\}$ of the tangent plane of $\mathcal{S}$ on each patch $\boldsymbol{\sigma}_i$. By the proof of Theorem 13.1.2,

$$\int_{\mathrm{int}(\boldsymbol{\gamma}_i)} K \, d\mathcal{A} = \int_0^{\ell(\boldsymbol{\gamma}_i)} \mathbf{f}' \cdot \dot{\mathbf{f}}'' \, ds. \tag{13.32}$$

Combining Eqs. 13.29, 13.31 and 13.32, we get

$$\sum_{i=1}^{n} \int_0^{\ell(\boldsymbol{\gamma}_i)} (\mathbf{f}' \cdot \dot{\mathbf{f}}'' - \boldsymbol{e}' \cdot \dot{\boldsymbol{e}}'') \, ds = 2\pi\chi. \tag{13.33}$$

But, from the proof of Theorem 13.1.2,

$$\boldsymbol{e}' \cdot \dot{\boldsymbol{e}}'' = \dot{\theta} - \kappa_g, \quad \mathbf{f}' \cdot \dot{\mathbf{f}}'' = \dot{\varphi} - \kappa_g,$$

where $\kappa_g$ is the geodesic curvature of $\boldsymbol{\gamma}_i$ and $\theta$ and $\varphi$ are the oriented angles $\widehat{\boldsymbol{e}'\dot{\boldsymbol{\gamma}}_i}$ and $\widehat{\mathbf{f}'\dot{\boldsymbol{\gamma}}_i}$, respectively. Then, $\psi = \varphi - \theta$ is the oriented angle $\widehat{\mathbf{f}'\boldsymbol{e}'}$, i.e., the oriented angle $\widehat{\mathbf{f}'\mathbf{V}}$ between $\mathbf{V}$ and the 'reference' tangent vector field $\mathbf{f}'$ on $\boldsymbol{\sigma}_i$. So the left-hand side of Eq. 13.33 is

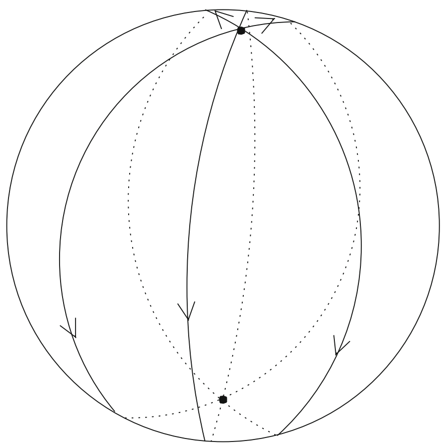
$$\sum_{i=1}^{n} \int_0^{\ell(\gamma_i)} \frac{d\psi}{ds} \, ds = 2\pi \sum_{i=1}^{n} \mu(\mathbf{p}_i),$$

as we want. □

We now give some simple examples of vector fields on surfaces (we show their integral curves for clarity).

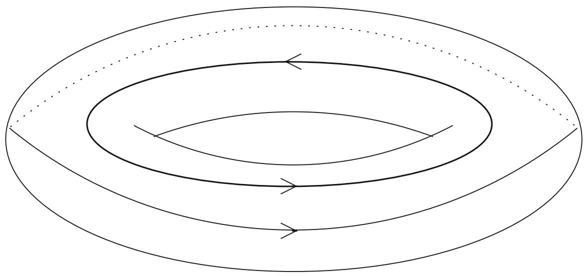
## Example 13.7.5

A vector field on the sphere with one source and one sink: $\chi = 2$



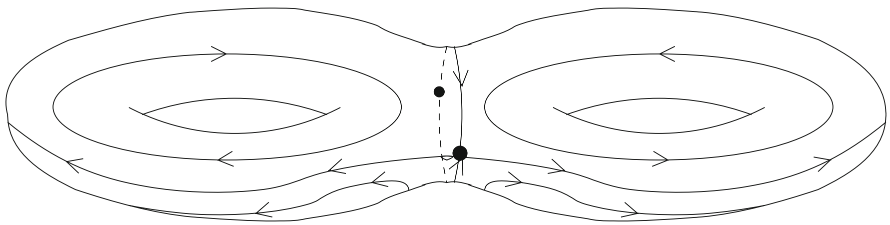
## Example 13.7.6

A vector field on the torus with no stationary points: $\chi = 0$



## Example 13.7.7

A vector field on the double torus $T_2$ with two bifurcations: $\chi = -2$



## *EXERCISES*

13.7.1 Let $k$ be a non-zero integer and let $\mathbf{V}(x, y) = (\alpha, \beta)$ be the vector field on the plane given by

$$\alpha + i\beta = \begin{cases} (x+iy)^k & \text{if } k > 0, \\ (x-iy)^{-k} & \text{if } k < 0. \end{cases}$$

Show that the origin is a stationary point of $\mathbf{V}$ of multiplicity $k$.

13.7.2 Show that the definition of a smooth tangent vector field is independent of the choice of surface patch. Show also that a tangent vector field $\mathbf{V}$ on $\mathcal{S}$ is smooth if and only if, for any surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$, the three components of $\mathbf{V}$ at the point $\boldsymbol{\sigma}(u, v)$ are smooth functions of $(u, v)$.

13.7.3 Show that the Definition 13.7.2 of the multiplicity of a stationary point of a tangent vector field $\mathbf{V}$ is independent of the'reference' vector field $\boldsymbol{\xi}$.

## 13.8 Critical points

If $f(u, v)$ is a smooth function defined on an open subset $U$ of $\mathbb{R}^2$, we say that a point $(u_0, v_0)$ is a *critical point* of $f$ if $\partial f/\partial u$ and $\partial f/\partial v$ both vanish at $(u_0, v_0)$. If now $F : \mathcal{S} \to \mathbb{R}$ is a smooth function on a surface $\mathcal{S}$ (see Exercise 4.3.1), and if $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a surface patch of $\mathcal{S}$, then $f = F \circ \boldsymbol{\sigma}$ is a smooth function on the open subset $U$ of $\mathbb{R}^2$. This suggests

### Definition 13.8.1

Let $\mathcal{S}$ be a surface and $F : \mathcal{S} \to \mathbb{R}$ a smooth function on $\mathcal{S}$. A point $\mathbf{p} \in \mathcal{S}$ is a *critical point* of $F$ if there is a surface patch $\boldsymbol{\sigma}$ of $\mathcal{S}$, with $\mathbf{p} = \boldsymbol{\sigma}(u_0, v_0)$, say, such that $f = F \circ \boldsymbol{\sigma}$ has a critical point at $(u_0, v_0)$.

It is easy to check directly that the definition of a critical point is independent of the choice of patch $\boldsymbol{\sigma}$ (see Exercise 13.8.1), but this will follow immediately from another characterization of critical points that we shall now give and which is independent of any arbitrary choices.

### Proposition 13.8.2

If $F$ is a smooth function on a surface $\mathcal{S}$, there is a unique smooth tangent vector field $\boldsymbol{\nabla}_{\mathcal{S}} F$ on $\mathcal{S}$ such that, if $\mathbf{p} \in \mathcal{S}$ and $\boldsymbol{\gamma}(t)$ is a curve in $\mathcal{S}$ which passes through $\mathbf{p}$ when $t = t_0$, we have

$$(\boldsymbol{\nabla}_{\mathcal{S}} F) \cdot \dot{\boldsymbol{\gamma}}(t_0) = \left.\frac{d}{dt}\right|_{t=t_0} F(\boldsymbol{\gamma}(t)). \tag{13.34}$$

Moreover, $\mathbf{p}$ is a critical point of $F$ if and only if $\boldsymbol{\nabla}_{\mathcal{S}} F = \mathbf{0}$ at $\mathbf{p}$.

### Proof

We showed in Exercise 4.4.3 that if $F$ is the restriction to $\mathcal{S}$ of a smooth function $F : \mathbb{R}^3 \to \mathbb{R}$, then Eq. 13.34 is satisfied if we take $\boldsymbol{\nabla}_{\mathcal{S}} F$ to be the orthogonal projection of the gradient $\boldsymbol{\nabla} F$ of $F$ onto the tangent plane of $\mathcal{S}$. But if $F : \mathcal{S} \to \mathbb{R}$ is not given to us as the restriction of a smooth function defined on $\mathbb{R}^3$, we must proceed differently, since $\nabla F$ only makes sense if $F$ is defined on $\mathbb{R}^3$, or at least on an open subset of $\mathbb{R}^3$.

We observe first that the tangent vector field $\nabla_{\mathcal{S}} F$ is obviously unique, if it exists. Indeed, every tangent vector to $\mathcal{S}$ at $\mathbf{p}$ is of the form $\dot{\boldsymbol{\gamma}}(t_0)$ for some $t_0$ and some curve $\boldsymbol{\gamma}$ on $\mathcal{S}$ with $\boldsymbol{\gamma}(t_0) = \mathbf{p}$, so any two choices of $\nabla_{\mathcal{S}} F$ at $\mathbf{p}$ would differ by a vector perpendicular to every tangent vector to $\mathcal{S}$ at $\mathbf{p}$, which must be zero.

To see that $\nabla_{\mathcal{S}} F$ exists, choose a surface patch $\boldsymbol{\sigma}(u,v)$ for $\mathcal{S}$ with $\boldsymbol{\sigma}(u_0,v_0) = \mathbf{p}$, say, and let $f = F \circ \boldsymbol{\sigma}$. Let $\{\boldsymbol{e}', \boldsymbol{e}''\}$ be the basis of the tangent plane of $\mathcal{S}$ at $\mathbf{p}$ such that

$$\boldsymbol{e}' \cdot \boldsymbol{\sigma}_u = \boldsymbol{e}'' \cdot \boldsymbol{\sigma}_v = 1, \quad \boldsymbol{e}' \cdot \boldsymbol{\sigma}_v = \boldsymbol{e}'' \cdot \boldsymbol{\sigma}_u = 0. \tag{13.35}$$

Explicitly,

$$\boldsymbol{e}' = \frac{G\boldsymbol{\sigma}_u - F\boldsymbol{\sigma}_v}{EG - F^2}, \quad \boldsymbol{e}'' = \frac{E\boldsymbol{\sigma}_v - F\boldsymbol{\sigma}_u}{EG - F^2}, \tag{13.36}$$

where $Edu^2 + 2Fdudv + Gdv^2$ is the first fundamental form of $\boldsymbol{\sigma}$. We take

$$\nabla_{\mathcal{S}} F = f_u \boldsymbol{e}' + f_v \boldsymbol{e}'', \tag{13.37}$$

where the derivatives are evaluated at $(u_0, v_0)$. If $\boldsymbol{\gamma}$ is as in the statement of the proposition, say $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$, then (with $d/dt$ denoted by a dot),

$$\begin{aligned}(\nabla_{\mathcal{S}} F) \cdot \dot{\boldsymbol{\gamma}}(0) &= (f_u \boldsymbol{e}' + f_v \boldsymbol{e}'') \cdot (\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \\ &= f_u \dot{u} + f_v \dot{v} \quad \text{(by Eq. 13.35)} \\ &= \dot{f},\end{aligned}$$

and the derivatives with respect to $t$ are being evaluated at $t = t_0$. It is clear from Eqs. 13.36 and 13.37 that $\nabla_{\mathcal{S}} F$ is smooth and that $\mathbf{p}$ is a critical point of $F$ if and only if $\nabla_{\mathcal{S}} F = \mathbf{0}$ at $\mathbf{p}$. $\square$

Since $\nabla_{\mathcal{S}} F$ is a smooth tangent vector field on $\mathcal{S}$, we can apply Theorem 13.7.4 to it. To do so, we must compute the multiplicity of the stationary points of $\nabla_{\mathcal{S}} F$. For this, we shall make an additional assumption about $F$, contained in the following.

## Definition 13.8.3

A critical point $\mathbf{p}$ of a smooth function $F$ on a surface $\mathcal{S}$ is said to be *non-degenerate* if, whenever $\boldsymbol{\sigma}(u,v)$ is a patch of $\mathcal{S}$ with $\mathbf{p} = \boldsymbol{\sigma}(u_0, v_0)$, say, the matrix

$$\mathcal{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial u^2} & \frac{\partial^2 f}{\partial u \partial v} \\ \frac{\partial^2 f}{\partial u \partial v} & \frac{\partial^2 f}{\partial v^2} \end{pmatrix}$$

is invertible, where $f = F \circ \boldsymbol{\sigma}$ and the derivatives are evaluated at $(u_0, v_0)$. In this case, the point $\mathbf{p}$ is called a *local maximum*, a *saddle point* or a *local minimum* if $\mathcal{H}$ has $2, 1$ or $0$ negative eigenvalues, respectively.

It is not difficult to show that this definition is sensible, i.e., independent of the choice of patch $\boldsymbol{\sigma}$ (see Exercise 13.8.1). Note that the matrix $\mathcal{H}$ is real and symmetric, so it always has two real eigenvalues (not necessarily distinct) – see Appendix 0.

## Proposition 13.8.4

Let $\mathbf{p}$ be a critical point of a smooth function $F$ on a surface $\mathcal{S}$. Then, the multiplicity of $\mathbf{p}$ as a stationary point of $\boldsymbol{\nabla}_{\mathcal{S}} F$ is

$$\mu(\mathbf{p}) = \begin{cases} 1 & \text{if } \mathbf{p} \text{ is a local maximum or a local minimum,} \\ -1 & \text{if } \mathbf{p} \text{ is a saddle point.} \end{cases}$$

## Example 13.8.5

The function on the plane given by $F(u, v) = -u^2 - v^2$ (resp. $u^2 - v^2$, $u^2 + v^2$) has a local maximum (respectively saddle point, local minimum) at the origin.

We shall not give a complete proof of Proposition 13.8.4 here. But the following argument should convince the reader of its truth. Let us assume that $(u_0, v_0) = (0, 0)$ for simplicity, and write the matrix in Definition 13.8.3 at $u = v = 0$ as

$$\mathcal{H} = \begin{pmatrix} \lambda & \mu \\ \mu & \nu \end{pmatrix}.$$

Then, Taylor's theorem tells us that

$$f(u, v) = \frac{1}{2}(\lambda u^2 + 2\mu uv + \nu v^2) + r(u, v),$$

where $r(u, v)/(u^2 + v^2)$ tends to zero as $u$ and $v$ tend to zero. It is plausible, then, that the behaviour of $\boldsymbol{\nabla}_{\mathcal{S}} F$ near $\mathbf{p}$ is the same as that of $\boldsymbol{\nabla}_{\mathcal{S}} \tilde{F}$, where

$$\tilde{F}(\boldsymbol{\sigma}(u, v)) = \frac{1}{2}(\lambda u^2 + 2\mu uv + \nu v^2).$$

In particular, $F$ and $\tilde{F}$ should have the same type of critical point at $\mathbf{p}$.

But the multiplicity of $\mathbf{p}$ as a critical point of $\tilde{F}$ is easy to compute. To do so, note first that there is an orthogonal matrix $P$ such that

$$P^t \begin{pmatrix} \lambda & \mu \\ \mu & \nu \end{pmatrix} P = \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix}$$

is a diagonal matrix. This means that, by applying an isometry in the $uv$-plane (i.e., a reparametrization of $\boldsymbol{\sigma}$), we can assume that

$$\tilde{F}(\boldsymbol{\sigma}(u,v)) = \frac{1}{2}(\epsilon_1 u^2 + \epsilon_2 v^2).$$

We can also assume that the patch $\boldsymbol{\sigma}$ is *conformal*, so that angles measured in the $uv$-plane are the same as those measured on the surface (see Section 6.3).

Taking our 'reference' tangent vector field to be $\boldsymbol{\sigma}_u$, the angle $\psi$ between $\boldsymbol{\nabla}_{\mathcal{S}}\tilde{F}$ and $\boldsymbol{\sigma}_u$ is given by

$$(\cos\psi, \sin\psi) = \frac{(\epsilon_1 u, \epsilon_2 v)}{\sqrt{\epsilon_1^2 u^2 + \epsilon_2^2 v^2}}.$$

We take the simple closed curve in the $uv$-plane given by the ellipse

$$\epsilon_1^2 u^2 + \epsilon_2^2 v^2 = r^2,$$

where $r$ is a small positive number, which can be parametrized by

$$u = \frac{r}{|\epsilon_1|}\cos t, \quad v\frac{r}{|\epsilon_2|}\sin t.$$

Hence,

$$\cos\psi = \frac{\epsilon_1}{|\epsilon_1|}\cos t, \quad \sin\psi = \frac{\epsilon_2}{|\epsilon_2|}\sin t,$$

and so

$$\psi = \begin{cases} t & \text{if } \epsilon_1 > 0 \text{ and } \epsilon_2 > 0, \\ 2\pi - t & \text{if } \epsilon_1 > 0 \text{ and } \epsilon_2 < 0, \\ \pi - t & \text{if } \epsilon_1 < 0 \text{ and } \epsilon_2 > 0, \\ \pi + t & \text{if } \epsilon_1 < 0 \text{ and } \epsilon_2 < 0. \end{cases}$$

This gives the multiplicity of $\mathbf{p}$ as a stationary point of $\boldsymbol{\nabla}_{\mathcal{S}}\tilde{F}$ as

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{d\psi}{dt}\,dt = \begin{cases} 1 & \text{if } \epsilon_1 \text{ and } \epsilon_2 \text{ have the same sign}, \\ -1 & \text{otherwise}, \end{cases}$$

in accordance with Proposition 13.8.4.

If we accept this heuristic argument, we can combine Theorem 13.7.4 and Proposition 13.8.4 to give

## Theorem 13.8.6

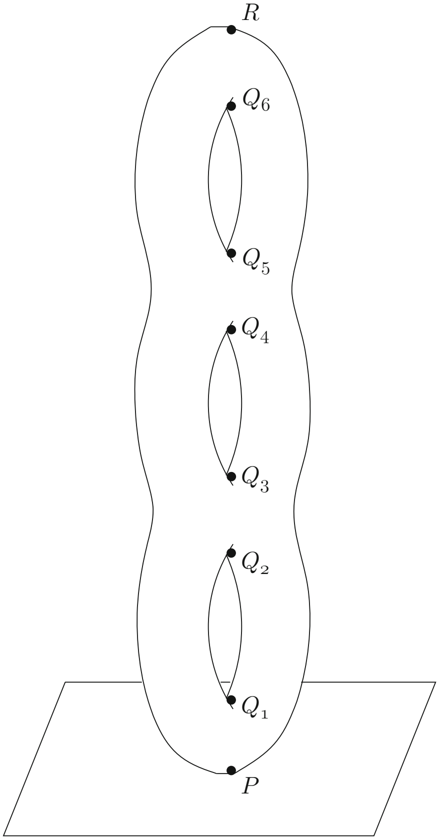
Let $F : \mathcal{S} \to \mathbb{R}$ be a smooth function on a compact surface $\mathcal{S}$ with only finitely many critical points, all non-degenerate. Then,

$$\begin{pmatrix}\text{number of local} \\ \text{maxima of } F\end{pmatrix} - \begin{pmatrix}\text{number of saddle} \\ \text{points of } F\end{pmatrix} + \begin{pmatrix}\text{number of local} \\ \text{minima of } F\end{pmatrix} = \chi,$$

the Euler number of $\mathcal{S}$.

### Example 13.8.7

If we take the surface $T_g$ of genus $g$ described in Section 5.4 and stand it upright on the $xy$-plane, the height above the plane is a smooth function $F$ on $T_g$. The critical points of $F$ are as shown in the following diagram, and they are all non-degenerate (cf. Exercise 13.8.3).




There is a unique local minimum $P$, $2g$ saddle points $Q_1, Q_2, \ldots, Q_{2g}$, and a unique local maximum $R$. Hence, Theorem 13.8.6 gives the Euler number of $T_g$ as

$$1 - 2g + 1 = 2 - 2g,$$

in accordance with Theorem 13.4.7.

## EXERCISES

13.8.1 Show directly that the definitions of a critical point (13.8.1), and whether it is non-degenerate (13.8.2), are independent of the choice of surface patch. Show that the classification of non-degenerate critical points into local maxima, local minima and saddle points is also independent of this choice.

13.8.2 For which of the following functions on the plane is the origin a non-degenerate critical point? In the non-degenerate case(s), classify the origin as a local maximum, local minimum or saddle point.

(i) $x^2 - 2xy + 4y^2$.

(ii) $x^2 + 4xy$.

(iii) $x^3 - 3xy^2$.

13.8.3 Let $\mathcal{S}$ be the torus obtained by rotating the circle $(x-2)^2+z^2 = 1$ in the $xz$-plane around the $z$-axis, and let $F : \mathcal{S} \to \mathbb{R}$ be the distance from the plane $x = -3$. Show that $F$ has four critical points, all non-degenerate, and classify them as local maxima, saddle points or local minima. (See Exercise 4.2.5 for a parametrization of $\mathcal{S}$.)

# Appendix 0
# *Inner product spaces and self-adjoint linear maps*

Throughout this appendix, $V$ denotes a vector space of finite dimension $n$ over $\mathbb{R}$. Proofs of all the results in this appendix can be found in standard books on linear algebra.

A map $V \times V \to \mathbb{R}$, denoted $(\mathbf{v}, \mathbf{w}) \mapsto \langle \mathbf{v}, \mathbf{w} \rangle$, is called a *bilinear form* if, for all $\lambda_1, \lambda_2 \in \mathbb{R}$, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in V$, we have

$$\langle \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2, \mathbf{w} \rangle = \lambda_1 \langle \mathbf{w}, \mathbf{v}_1 \rangle + \lambda_2 \langle \mathbf{w}, \mathbf{v}_2 \rangle$$
$$\langle \mathbf{w}, \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 \rangle = \lambda_1 \langle \mathbf{w}, \mathbf{v}_1 \rangle + \lambda_2 \langle \mathbf{w}, \mathbf{v}_2 \rangle.$$

Thus, $\langle \mathbf{v}, \mathbf{w} \rangle$ is a linear function of $\mathbf{v}$ for each fixed $\mathbf{w}$, and a linear function of $\mathbf{w}$ for each fixed $\mathbf{v}$.

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of $V$, any bilinear form $\langle\ ,\ \rangle$ on $V$ is determined by the $n \times n$ matrix whose $(i,j)$-entry is $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ for $i, j = 1, \ldots, n$. For, if

$$\mathbf{v} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i, \quad \mathbf{w} = \sum_{i=1}^{n} \mu_i \mathbf{v}_i$$

are any two vectors in $V$, then

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i,j=1}^{n} \lambda_i \mu_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle.$$

A bilinear form $\langle\ ,\ \rangle$ is called *symmetric* if

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle \quad \text{for all } \mathbf{v}, \mathbf{w} \in V.$$

Equivalently, the matrix of $\langle\ ,\ \rangle$ with respect to any basis of $V$ is symmetric.

Any symmetric bilinear form $\langle\ ,\ \rangle$ is uniquely determined by its associated *quadratic form* $q : V \to \mathbb{R}$ is given by

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

Indeed, for any $\mathbf{v}, \mathbf{w} \in V$,

$$\langle \mathbf{v}, \mathbf{w} \rangle = \frac{1}{2}(q(\mathbf{v}+\mathbf{w}) - q(\mathbf{v}) - q(\mathbf{w})).$$

An *inner product* on $V$ is a symmetric bilinear form $\langle\ ,\ \rangle$ on $V$ such that

$$\langle \mathbf{v}, \mathbf{v} \rangle > 0 \quad \text{for all non-zero } \mathbf{v} \in V.$$

The *length* of a vector $\mathbf{v}$ is then defined to be $\| \mathbf{v} \| = \langle \mathbf{v}, \mathbf{v} \rangle^{1/2}$, and two vectors $\mathbf{v}, \mathbf{w}$ are said to be *perpendicular*, or *orthogonal*, if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

## Proposition A.0.1

Let $\langle\ ,\ \rangle$ be an inner product on $V$. Then, $V$ has an *orthonormal basis*, i.e., a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ such that

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

If $V$ has an inner product $\langle\ ,\ \rangle$, a linear map $L : V \to V$ is called *self-adjoint* if

$$\langle L(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, L(\mathbf{w}) \rangle \quad \text{for all } \mathbf{v}, \mathbf{w} \in V.$$

Equivalently, the bilinear form

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

should be symmetric.

## Proposition A.0.2

Let $L : V \to V$ be a self-adjoint linear map. Then, the matrix of $L$ with respect to any orthonormal basis of $V$ is symmetric.

If $L : V \to V$ is a linear map, a real number $\lambda$ is called an *eigenvalue* of $L$ if $L(\mathbf{v}) = \lambda \mathbf{v}$ for some non-zero vector $\mathbf{v} \in V$. In that case, $\mathbf{v}$ is called an *eigenvector* of $L$ corresponding to the eigenvalue $\lambda$.

## Theorem A.0.3

Let $L : V \to V$ be a self-adjoint linear map. Then, $V$ has a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ consisting of eigenvectors of $L$. Moreover, if $\mathbf{v}_i$ and $\mathbf{v}_j$ are eigenvectors corresponding to distinct eigenvalues, then $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$.

The last sentence in this theorem implies that the basis of eigenvectors of $L$ can be chosen to be orthonormal.

This theorem is sometimes referred to by saying that $L$ is *diagonalizable*: the matrix of $L$ with respect to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ in the theorem is the diagonal matrix

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where $\lambda_i$ is the eigenvalue corresponding to the eigenvector $\mathbf{v}_i$.

In fact, all of the above results have matrix versions. If $A$ is an $n \times n$ matrix, an *eigenvalue* of $A$ is a real number $\lambda$ such that

$$\det(A - \lambda I) = 0,$$

where $I$ denotes the $n \times n$ identity matrix. An *eigenvector* of $A$ with eigenvalue $\lambda$ is an $n \times 1$ column matrix $\mathbf{v}$ such that $A\mathbf{v} = \lambda \mathbf{v}$. Theorem A.0.3 is equivalent to

## Theorem A.0.4

Let $S$ be a real symmetric $n \times n$ matrix. Then, there is an orthogonal matrix $P$ such that $PSP^t$ is a diagonal matrix.

The correspondence is as follows. Let $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be any orthonormal basis of $V$. Then, the linear map $L : V \to V$ such that

$$L(\mathbf{w}_j) = \sum_{i=1}^{n} s_{ij} \mathbf{w}_i$$

(where $S = (s_{ij})$) is self-adjoint. The matrix $P = (p_{ij})$ such that

$$\mathbf{w}_j = \sum_{i=1}^{n} p_{ij} \mathbf{v}_i$$

is orthogonal and $PSP^t$ is the diagonal matrix associated to $L$ by Theorem A.0.3.

# Appendix 1
## *Isometries of Euclidean spaces*

In this appendix, we collect some basic results about isometries of $\mathbb{R}^n$ that are used in various places in the book.

### Definition A.1.1

Let $n \geq 1$. An *isometry* of $\mathbb{R}^n$ is a map $F : \mathbb{R}^n \to \mathbb{R}^n$ that preserves the distance between any two points of $\mathbb{R}^n$:

$$\| F(\mathbf{v}) - F(\mathbf{w}) \| = \| \mathbf{v} - \mathbf{w} \| \quad \text{for all } \mathbf{v}, \mathbf{w} \in \mathbb{R}^n.$$

It is obvious that any composite of isometries is an isometry.

For $i = 1, \ldots, n$, let $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (with a 1 in the $i$th place).

### Proposition A.1.2

Let $F$ be an isometry of $\mathbb{R}^n$ such that $F(\mathbf{0}) = \mathbf{0}$ and $F(\boldsymbol{e}_i) = \boldsymbol{e}_i$ for $i = 1, \ldots, n$. Then, $F$ is the identity map.

### Proof

If $\mathbf{v} \in \mathbb{R}^n$, the assumptions on $F$ imply that $\| \mathbf{v} \| = \| F(\mathbf{v}) \|$ and $\| \mathbf{v} - \boldsymbol{e}_i \| = \| F(\mathbf{v}) - \boldsymbol{e}_i \|$ for $i = 1, \ldots, n$. These equations in turn imply that $\mathbf{v} \cdot \boldsymbol{e}_i = F(\mathbf{v}) \cdot \boldsymbol{e}_i$ for $i = 1, \ldots, n$, and hence that $\mathbf{v} = F(\mathbf{v})$. $\square$

It will be convenient to associate to any vector $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$ the column matrix

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

which we also denote by $\mathbf{v}$. The dot product of vectors can then be expressed in matrix form:

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^t \mathbf{w}.$$

Recall that an invertible $n \times n$ matrix $P$ is said to be *orthogonal* if $P^t = P^{-1}$, where $P^t$ denotes the transpose of $P$. An equivalent condition is that the columns of $P$ are perpendicular unit vectors.

## Theorem A.1.3

Let $P$ be an $n \times n$ orthogonal matrix and let $\mathbf{a} \in \mathbb{R}^n$. Then, the map $F : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$F(\mathbf{v}) = P\mathbf{v} + \mathbf{a} \tag{A.1.1}$$

is an isometry of $\mathbb{R}^n$. Moreover, every isometry of $\mathbb{R}^n$ is obtained in this way. In particular, every isometry is a bijective map and the inverse of any isometry is an isometry.

## Proof

If $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $F$ is as defined in Eq. A.1.1 with $P$ orthogonal, we have

$$\begin{aligned} \| F(\mathbf{v}) - F(\mathbf{w}) \|^2 &= (F(\mathbf{v}) - F(\mathbf{w})) \cdot (F(\mathbf{v}) - F(\mathbf{w})) \\ &= (P\mathbf{v} - P\mathbf{w})^t (P\mathbf{v} - P\mathbf{w}) \\ &= (\mathbf{v} - \mathbf{w})^t P^t P (\mathbf{v} - \mathbf{w}) \\ &= (\mathbf{v} - \mathbf{w})^t (\mathbf{v} - \mathbf{w}) \\ &= \| \mathbf{v} - \mathbf{w} \|^2 . \end{aligned}$$

Hence $F$ is an isometry.

Now define

$$G(\mathbf{v}) = P^t \mathbf{v} - P^t \mathbf{a}.$$

Then, $G$ is an isometry (because $P^t$ is orthogonal: $(P^t)^t = P = (P^{-1})^{-1} = (P^t)^{-1}$) and

$$G \circ F(\mathbf{v}) = P^t(P\mathbf{v} + \mathbf{a}) - P^t\mathbf{a} = \mathbf{v}, \quad F \circ G(\mathbf{v}) = P(P^t\mathbf{v} - P^t\mathbf{a}) + \mathbf{a} = \mathbf{v},$$

and so $G$ is the inverse of $F$ (and $F$ is bijective as it has an inverse).

For the converse, let $F$ be any isometry of $\mathbb{R}^n$. Let $\mathbf{a} = F(\mathbf{0})$ and let $\mathbf{v}_i = F(\boldsymbol{e}_i) - \mathbf{a}$. Then, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are perpendicular unit vectors. Indeed,

$$\| \mathbf{v}_i \| = \| F(\boldsymbol{e}_i) - F(\mathbf{0}) \| = \| \boldsymbol{e}_i - \mathbf{0} \| = 1,$$

showing that the $\mathbf{v}_i$ are unit vectors, and

$$\| \mathbf{v}_i - \mathbf{v}_j \| = \| \boldsymbol{e}_i - \boldsymbol{e}_j \|$$

which implies (using the fact that the $\mathbf{v}_i$ and the $\boldsymbol{e}_i$ are unit vectors) that $\mathbf{v}_i \cdot \mathbf{v}_j = \boldsymbol{e}_i \cdot \boldsymbol{e}_j = 0$ if $i \neq j$ (since $\| \mathbf{v}_i - \mathbf{v}_j \|^2 = \| \mathbf{v}_i \|^2 + \| \mathbf{v}_j \|^2 - 2\mathbf{v}_i \cdot \mathbf{v}_j$, etc.).

Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are perpendicular unit vectors, the matrix $P$ whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is orthogonal. Moreover,

$$P\boldsymbol{e}_i = \mathbf{v}_i = F(\boldsymbol{e}_i) - \mathbf{a}.$$

Hence, if $G$ denotes the isometry

$$G(\mathbf{v}) = P\mathbf{v} + \mathbf{a},$$

we have $F(\boldsymbol{e}_i) = G(\boldsymbol{e}_i)$ for $i = 1, \ldots, n$ and $F(\mathbf{0}) = G(\mathbf{0})$. Then the isometry $G^{-1} \circ F$ fixes $\mathbf{0}$ and $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, and so is the identity map by Proposition A.1.2. Hence, $F = G$. □

If $P$ is an orthogonal matrix, the determinant of $P$ must be $\pm 1$ since

$$1 = \det(P^t P) = \det(P^t)\det(P) = \det(P)^2.$$

The isometries (A.1.1) for which $\det(P) = 1$ are said to be *direct isometries*; those for which $\det(P) = -1$ are *opposite isometries*. If

$$F(\mathbf{v}) = P\mathbf{v} + \mathbf{a}, \quad G(\mathbf{v}) = Q\mathbf{v} + \mathbf{b}$$

are two isometries, the composite $G \circ F$ is the isometry

$$(G \circ F)(\mathbf{v}) = QP\mathbf{v} + Q\mathbf{a} + \mathbf{b}.$$

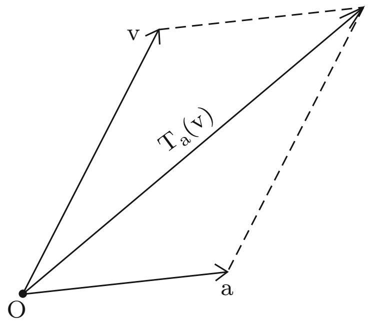
Since $\det(QP) = \det(Q)\det(P)$, the composite of two direct or two opposite isometries is direct, and the composite of a direct and an opposite isometry is opposite.

We now turn to the geometric description of isometries. The two simplest types are as follows:

*Translations* These are the direct isometries $T_{\mathbf{a}}$ given by

$$T_{\mathbf{a}}(\mathbf{v}) = \mathbf{v} + \mathbf{a}, \tag{A.1.2}$$

where $\mathbf{a}$ is a fixed vector in $\mathbb{R}^n$.



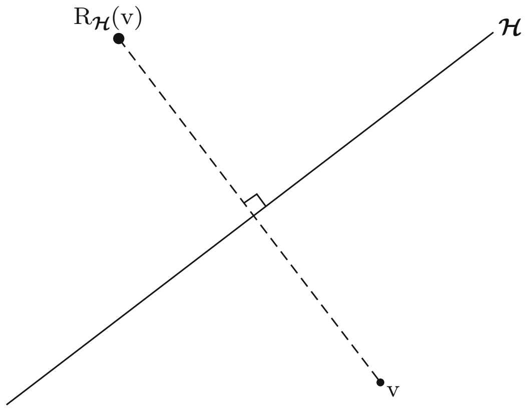

*Reflections* If $\mathbf{N} \in \mathbb{R}^n$ is a fixed non-zero vector (which we may as well assume is a unit vector) and $d \in \mathbb{R}$, the set $\mathcal{H}$ of vectors $\mathbf{v} \in \mathbb{R}^n$ such that

$$\mathbf{v} \cdot \mathbf{N} = d$$

is called a *hyperplane* (a hyperplane in $\mathbb{R}^2$ is just a straight line, and a hyperplane in $\mathbb{R}^3$ is a plane). The reflection $R_{\mathcal{H}}$ in $\mathcal{H}$ is defined by

$$R_{\mathcal{H}}(\mathbf{v}) = \mathbf{v} - 2(\mathbf{v} \cdot \mathbf{N} - d)\mathbf{N}. \tag{A.1.3}$$

We leave it to the reader to check that $R_{\mathcal{H}} \circ R_{\mathcal{H}}$ is the identity map, so that $R_{\mathcal{H}}$ is its own inverse, and that the set of points $\mathbf{v}$ fixed by $R_{\mathcal{H}}$ (i.e., such that $R_{\mathcal{H}}(\mathbf{v}) = \mathbf{v}$) is exactly the hyperplane $\mathcal{H}$. If we think of $\mathcal{H}$ as a two-way mirror, $R_{\mathcal{H}}(\mathbf{v})$ is the mirror-image of $\mathbf{v}$:




## Proposition A.1.4

Reflections are opposite isometries.

## Proof

In matrix notation, $R_{\mathcal{H}}(\mathbf{v}) = P\mathbf{v} + \mathbf{a}$ where

$$P = I - 2\mathbf{N}\mathbf{N}^t$$

and $\mathbf{a} = 2d\mathbf{N}$. Now

$$\begin{aligned} P^tP = (I - 2\mathbf{N}\mathbf{N}^t)^t(I - 2\mathbf{N}\mathbf{N}^t) &= (I - 2\mathbf{N}\mathbf{N}^t)(I - 2\mathbf{N}\mathbf{N}^t) \\ &= I - 4\mathbf{N}\mathbf{N}^t + 4\mathbf{N}\mathbf{N}^t\mathbf{N}\mathbf{N}^t = I, \end{aligned}$$

since $\mathbf{N}^t\mathbf{N} = \mathbf{N} \cdot \mathbf{N} = 1$ ($I$ denotes the identity matrix). Thus, $P$ is orthogonal and $R_{\mathcal{H}}$ is an isometry.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ be an orthonormal basis of the hyperplane $\mathbf{v} \cdot \mathbf{N} = 0$ (i.e., the hyperplane parallel to $\mathcal{H}$ passing through the origin). Let $Q$ be the $n \times n$ matrix whose columns consist of the perpendicular unit vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{N}$; then $Q$ is orthogonal. The product $PQ$ is the matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, -\mathbf{N}$, so

$$PQ = Q \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & -1 \end{pmatrix}.$$

Taking determinants of both sides and noting that $\det(Q) \neq 0$ gives $\det(P) = -1$ as required. □

It is obvious that any composite of translations is another translation. On the other hand, it turns out that *every* isometry is a composite of reflections:

## Theorem A.1.5

Every isometry of $\mathbb{R}^n$ is a composite of $\leq n + 1$ reflections.

## Proof

Let $F$ be any isometry of $\mathbb{R}^n$ and let $\boldsymbol{e}_0 = \mathbf{0}$. We construct isometries $G_i$, for $i = 0, 1, \ldots, n$, such that

(i) $G_i$ is a composite of $\leq i + 1$ reflections, and

(ii) $G_i \circ F$ fixes the points $\boldsymbol{e}_0, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_i$.

Then, $G_n \circ F$ is an isometry that fixes $\mathbf{0}$ and the points $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, and so is the identity map by Proposition A.1.2. Moreover, $G_n^{-1}$ is the composite of $\leq n+1$ reflections, for if $G_n = R_1 \circ R_2 \circ \cdots \circ R_k$, where $R_1, R_2, \ldots, R_k$ are reflections, then $G^{-1} = R_k \circ \cdots \circ R_2 \circ R_1$.

We prove the existence of the isometries $G_i$ by induction on $i$. If $\mathbf{v}_0 = \boldsymbol{e}_0$ let $G_0$ be the identity map. Otherwise, let $\mathcal{H}_0$ be the hyperplane perpendicular to the vector $\mathbf{v}_0 - \boldsymbol{e}_0$ and passing through the mid-point of the line joining $\boldsymbol{e}_0$ and $\mathbf{v}_0$ (i.e., the point $\frac{1}{2}(\boldsymbol{e}_0 + \mathbf{v}_0)$). Then, $R_{\mathcal{H}_0}$ takes $\mathbf{v}_0$ to $\boldsymbol{e}_0$ and we define $G_0 = R_{\mathcal{H}_0}$.

Suppose now that we have constructed $G_{i-1}$ satisfying the required conditions. Then, the isometry $F_i = G_{i-1} \circ F$ fixes the points $\boldsymbol{e}_0, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{i-1}$. If $F_i(\boldsymbol{e}_i) = \boldsymbol{e}_i$, define $G_i = G_{i-1}$. Otherwise, let $\mathcal{H}_i$ be the hyperplane perpendicular to $F_i(\boldsymbol{e}_i) - \boldsymbol{e}_i$ and passing through the mid-point of the line joining $\boldsymbol{e}_i$ and $F(\boldsymbol{e}_i)$. Note that the points $\boldsymbol{e}_j$ for $j = 0, 1, \ldots, i-1$ are all on $\mathcal{H}_i$ because

$$\| \boldsymbol{e}_j - F_i(\boldsymbol{e}_i) \| = \| F_i(\boldsymbol{e}_j) - F_i(\boldsymbol{e}_i) \| = \| \boldsymbol{e}_j - \boldsymbol{e}_i \| .$$

Hence, the reflection $R_{\mathcal{H}_i}$ fixes $\boldsymbol{e}_j$ for $j = 0, 1, \ldots, i-1$ and takes $F_i(\boldsymbol{e}_i)$ to $\boldsymbol{e}_i$. Define $G_i = R_{\mathcal{H}_i} \circ G_{i-1}$. This completes the inductive step. □

We now specialize to the cases of interest in this book, namely $\mathbb{R}^2$ and $\mathbb{R}^3$. We need only consider isometries that fix the origin since a general isometry can then be obtained by composing with a translation.

If $P$ is a $2 \times 2$ orthogonal matrix, its columns $\mathbf{a}_1$ and $\mathbf{a}_2$ are perpendicular unit vectors. If $\theta$ is the angle between $\mathbf{a}_1$ and the positive $x$-axis in $\mathbb{R}^2$, we then have

$$\mathbf{a}_1 = (\cos\theta, \sin\theta), \quad \mathbf{a}_2 = \pm(-\sin\theta, \cos\theta).$$

If

$$P = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

the corresponding isometry $\rho_\theta(\mathbf{v}) = P\mathbf{v}$ is the anticlockwise *rotation* through an angle $\theta$ about the origin. Since $\det(P) = 1$, $\rho_\theta$ is a direct isometry, and so by Proposition A.1.3 must be a composite of two reflections. In fact, $\rho_\theta$ is the product of the reflections in any two lines that intersect at the origin and make an angle $\theta/2$ with each other.

The other possibility is

$$P = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

The corresponding isometry is the reflection in the line passing through the origin and making an angle $\theta/2$ with the positive $x$-axis.

Thus, every direct isometry of $\mathbb{R}^2$ is the composite of a rotation and a translation, and every opposite isometry is the composite of a reflection and a translation.

For the $\mathbb{R}^3$ case, we recall from Appendix 0 that if $P$ is a square matrix, its *eigenvalues* are the roots $\lambda$ of the equation

$$\det(P - \lambda I) = 0. \tag{A.1.4}$$

When $P$ is $3 \times 3$, Eq. A.1.4 is a cubic equation for $\lambda$ which has real coefficients. It must therefore have at least one real root; we denote the corresponding eigenvector by $\mathbf{N}$, and we may as well assume that $\mathbf{N}$ is a unit vector.

Suppose now that $P$ is orthogonal. Then $\parallel P\mathbf{N} \parallel = \parallel \mathbf{N} \parallel$ and so $\lambda = \pm 1$. If $\mathbf{w}$ is any vector perpendicular to $\mathbf{N}$, $P\mathbf{w}$ is also perpendicular to $\mathbf{N}$ since (switching between vector and matrix notation)

$$(P\mathbf{w}) \cdot \mathbf{N} = \mathbf{w}^t P^t \mathbf{N} = \mathbf{w}^t P^{-1} \mathbf{N} = \pm \mathbf{w}^t \mathbf{N} = \pm \mathbf{w} \cdot \mathbf{N} = 0$$

(we used the fact that $P\mathbf{N} = \pm\mathbf{N}$ implies $P^{-1}\mathbf{N} = \pm\mathbf{N}$). So the isometry $F(\mathbf{v}) = P\mathbf{v}$ fixes each point of the line $\ell$ through the origin parallel to $\mathbf{N}$ and preserves the plane $\Pi$ through the origin perpendicular to $\ell$. Obviously the restriction $F|_\Pi$ of $F$ to $\Pi$ is an isometry of $\Pi$. By the discussion of the $\mathbb{R}^2$ case, $F|_\Pi$ must be either a rotation or a reflection.

There are now several cases to consider. If $\lambda = 1$ and $F|_\Pi$ is a rotation, then $F$ fixes each point of $\ell$ and performs a rotation through an angle $\theta$, say, in each plane perpendicular to $\ell$. We denote this isometry by $\rho_{\ell,\theta}$ and call $\ell$ the *axis of rotation*.

If $\lambda = 1$ and $F|_\Pi$ is a reflection in a line $\ell'$ in $\Pi$, then $F$ is the reflection in the plane containing $\ell$ and $\ell'$.

If $\lambda = -1$ and $F|_\Pi$ is the reflection in a line $\ell'$, then $F$ is rotation by $\pi$ about the line through the origin perpendicular to $\ell$ and $\ell'$.

Finally, if $\lambda = -1$ and $F|_\Pi$ is a rotation, then $F$ is a product of three reflections. An example of such a 'reflection-rotation' is the *antipodal map* $F(\mathbf{v}) = -\mathbf{v}$.

In particular, we have shown that every direct isometry of $\mathbb{R}^3$ is the composite of a rotation about an axis passing through the origin and a translation.

We shall need one more fact relating the isometries of $\mathbb{R}^3$ to the cross product of vectors in $\mathbb{R}^3$.

## Proposition A.1.6

Let $P$ be a $3 \times 3$ orthogonal matrix. Then, for any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$,

$$P\mathbf{v} \times P\mathbf{w} = \det(P)\, P(\mathbf{v} \times \mathbf{w}).$$

## Proof

By Theorem A.1.5 it is enough to prove this when $P$ corresponds to the reflection in a plane passing through the origin. If $\mathbf{N}$ is a unit vector, the reflection in the plane through the origin perpendicular to $\mathbf{N}$ is

$$R(\mathbf{v}) = \mathbf{v} - 2(\mathbf{v}\cdot\mathbf{N})\mathbf{N}.$$

In view of Proposition A.1.4, we have to prove that

$$\begin{aligned} &(\mathbf{v} - 2(\mathbf{v}\cdot\mathbf{N})\mathbf{N}) \times (\mathbf{w} - 2(\mathbf{w}\cdot\mathbf{N})\mathbf{N}) = -(\mathbf{v}\times\mathbf{w} - 2((\mathbf{v}\times\mathbf{w})\cdot\mathbf{N})\mathbf{N}) \\ \text{i.e.,}\quad &\mathbf{v}\times\mathbf{w} + ((\mathbf{v}\cdot\mathbf{N})\mathbf{w} - (\mathbf{w}\cdot\mathbf{N})\mathbf{v})\times\mathbf{N} = ((\mathbf{v}\times\mathbf{w})\cdot\mathbf{N})\mathbf{N}. \qquad \text{(A.1.5)} \end{aligned}$$

Applying the triple-product identity

$$(\mathbf{a}\times\mathbf{b})\times\mathbf{c} = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}, \qquad \text{(A.1.6)}$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$, to the second term on the left-hand side of Eq. A.1.5, we are reduced to proving

$$\mathbf{v}\times\mathbf{w} + ((\mathbf{v}\times\mathbf{w})\times\mathbf{N})\times\mathbf{N} = ((\mathbf{v}\times\mathbf{w})\cdot\mathbf{N})\mathbf{N},$$

and this follows from a second application of the identity (A.1.6). $\square$

# Appendix 2
## *Möbius transformations*

We collect here the main facts about Möbius transformations that we use in the exercises of Section 6.5 and in Chapter 11.

A *Möbius transformation* is a map of the form

$$M(z) = \frac{az + b}{cz + d}, \tag{A.2.1}$$

where $a, b, c$, and $d$ are complex numbers such that $ad - bc \neq 0$. Note that the map $M$ is unchanged if $a, b, c$ and $d$ are all multiplied by the same non-zero complex number.

If $c = 0$, $M(z)$ is defined for all $z \in \mathbb{C}$; if $c \neq 0$ it is defined for all $z \neq -d/c$. We can avoid this dichotomy by extending $M$ to a map on the extended complex plane $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$:

$$\begin{gathered} M(\infty) = \infty \quad \text{if } c = 0, \\ M(-d/c) = \infty, \quad M(\infty) = a/c \quad \text{if } c \neq 0. \end{gathered}$$

### Proposition A.2.1

(i) Every Möbius transformation defines a bijection $\mathbb{C}_\infty \to \mathbb{C}_\infty$ whose inverse is a Möbius transformation.

(ii) Any composite of Möbius transformations is a Möbius transformation.

## Proof

Let $M$ be as in (A.2.1) and define the Möbius transformation

$$N(z) = \frac{dz - b}{-cz + a}.$$

If $z \in \mathbb{C}$ and (if $c \neq 0$) $z \neq -d/c$ we have

$$N(M(z)) = \left( \frac{d\left(\frac{az+b}{cz+d}\right) - b}{-c\left(\frac{az+b}{cz+d}\right) + a} \right) = \frac{(ad - bc)z}{ad - bc} = z.$$

If $c \neq 0$,

$$N(M(-d/c)) = N(\infty) = -d/c.$$

Finally,

$$N(M(\infty)) = \begin{cases} N(\infty) = \infty & \text{if } c = 0 \\ N(a/c) = \infty & \text{if } c \neq 0. \end{cases}$$

Thus, $N(M(z)) = z$ for all $z \in \mathbb{C}_\infty$. Similar computations show that $M(N(z)) = z$ for all $z \in \mathbb{C}_\infty$. This shows that $N$ is the inverse of $M$; in particular, $M$ is bijective.

For (ii), let $M$ be as in (A.2.1) and let

$$M'(z) = \frac{a'z + b'}{c'z + d'}$$

be another Möbius transformation. Then,

$$M'(M(z)) = \frac{a'(az + b) + b'(cz + d)}{c'(az + b) + d'(cz + d)} = \frac{(a'a + b'c)z + a'b + b'd}{(c'a + d'c)z + c'b + d'd}.$$

Since

$$(a'a + b'c)(c'b + d'd) - (a'b + b'd)(c'a + d'c) = (a'd' - b'c')(ad - bc) \neq 0,$$

$M' \circ M$ is a Möbius transformation. □

The simplest examples of Möbius transformations are the following:

*Translations* $T_a(z) = z + a$, $a \in \mathbb{C}$

*Complex dilations* $D_a(z) = az$, $a \in \mathbb{C}$, $a \neq 0$

*The map* $K(z) = 1/z$

If we identify $\mathbb{C}$ with $\mathbb{R}^2$ in the usual way, the translations $T_a$ coincide with the translations of $\mathbb{R}^2$ considered in Appendix 1. The complex dilation $D_a$ is the composite of the real dilation $D_{|a|}$ and the rotation about the origin through an angle $\arg(a)$.

## Proposition A.2.2

Every Möbius transformation is a composite of Möbius transformations of the above three types.

## Proof

Let $M$ be as in (A.2.1). If $c = 0$ then $a \neq 0$ and

$$M(z) = az + b = a(z + b/a) = D_a T_{b/a}(z).$$

If $c \neq 0$,

$$\begin{aligned} M(z) &= \frac{a}{c} - \frac{(ad - bc)/c}{cz + d} = T_{a/c} D_{-(ad-bc)/c} K(cz + d) \\ &= T_{a/c} D_{-(ad-bc)/c} K T_d D_c(z). \end{aligned} \qquad \square$$

It is clear geometrically that translations and complex dilations take straight lines to straight lines and circles to circles, and we shall see shortly that the transformation $K$ takes lines and circles to lines and circles, but may take lines to circles and circles to lines. Because of this, it is convenient to define a Circle (capital C!) in $\mathbb{C}_\infty$ to be either a straight line or a circle in $\mathbb{C}$.

## Proposition A.2.3

Every Circle in $\mathbb{C}_\infty$ can be described by an equation of the form

$$az\bar{z} + \bar{b}z + b\bar{z} + c = 0, \tag{A.2.2}$$

where $a, c \in \mathbb{R}$, $b \in \mathbb{C}$. In fact, Eq. A.2.2 represents a straight line if $a = 0$ and $b \neq 0$, and a circle if $a \neq 0$ and $|b|^2 > ac$.

## Proof

Any straight line in the $xy$-plane has equation of the form

$$px + qy + r = 0,$$

where $p, q, r \in \mathbb{R}$ and $p$ and $q$ are not both zero. Writing $z = x + iy$, this is of the form (A.2.2) with $a = 0$, $b = p + iq$, $c = 2r$. On the other hand, every circle has equation of the form

$$(x + p)^2 + (y + q)^2 = r^2,$$

where $p, q \in \mathbb{R}$ and $r > 0$. This is of the form (A.2.2) with $a = 1$, $b = p + iq$ and $c = p^2 + q^2 - r^2$. Note that $|b|^2 - ac = r^2$. □

## Proposition A.2.4

Every Möbius transformation takes Circles to Circles.

## Proof

This is clear geometrically for translations and complex dilations, so in view of Proposition A.2.2 it is sufficient to prove it for the transformation $K$. If $w = K(z)$ then $z = 1/w$ and so if $z$ lies on the Circle (A.2.2) then

$$\begin{aligned} & \frac{a}{|w|^2} + \frac{\bar{b}}{w} + \frac{b}{\bar{w}} + c &= 0, \\ \therefore \quad & c|w|^2 + \bar{b}\bar{w} + bw + a &= 0. \end{aligned}$$

If $c = 0$, this is the equation of a line; if $c \neq 0$, it is the equation of a circle since $|\bar{b}|^2 > ca$. □

Note that this proof actually shows that $K$ takes circles passing through the origin to lines and all other circles to circles.

The other important property of Möbius transformations we shall need is

## Proposition A.2.5

Every Möbius transformation is conformal.

This means that every Möbius transformation preserves the angle between curves that intersect at a point of $\mathbb{C}$.

The reader versed in complex analysis will be able to deduce Proposition A.2.5 easily from the fact that holomorphic functions with non-vanishing derivatives are conformal. We shall give a direct proof.

## Proof A.2.5

By Proposition A.2.2, it suffices to prove that translations, complex dilations and the transformation $K$ are conformal. Now translations and rotations are isometries, hence conformal (Exercise 6.3.1), and it is easy to see that every dilation $D_a$ with $a > 0$ is conformal. Hence, we are reduced to proving that $K$ is conformal. For this, we use the method of Exercise 6.1.4.

If $u, v \in \mathbb{R}$, let

$$K(u + iv) = \tilde{u} + i\tilde{v}.$$

Then,

$$u + iv = K(\tilde{u} + i\tilde{v}) = \frac{\tilde{u} - i\tilde{v}}{\tilde{u}^2 + \tilde{v}^2},$$

$$\text{i.e.,} \quad u = \frac{\tilde{u}}{\tilde{u}^2 + \tilde{v}^2}, \quad v = \frac{-\tilde{v}}{\tilde{u}^2 + \tilde{v}^2}.$$

Hence,

$$du = \frac{\tilde{v}^2 - \tilde{u}^2}{(\tilde{u}^2 + \tilde{v}^2)^2}\, d\tilde{u} - \frac{2\tilde{u}\tilde{v}}{(\tilde{u}^2 + \tilde{v}^2)^2}\, d\tilde{v},$$

$$dv = \frac{2\tilde{u}\tilde{v}}{(\tilde{u}^2 + \tilde{v}^2)^2}\, d\tilde{u} + \frac{\tilde{v}^2 - \tilde{u}^2}{(\tilde{u}^2 + \tilde{v}^2)^2}\, d\tilde{v},$$

so the first fundamental form $du^2 + dv^2$ of $\mathbb{R}^2$ becomes

$$\frac{(\tilde{v}^2 - \tilde{u}^2)^2 + 4\tilde{u}^2\tilde{v}^2}{(\tilde{u}^2 + \tilde{v}^2)^4}(d\tilde{u}^2 + d\tilde{v}^2) = \frac{d\tilde{u}^2 + d\tilde{v}^2}{(\tilde{u}^2 + \tilde{v}^2)^2}.$$

Since this is a multiple of $d\tilde{u}^2 + d\tilde{v}^2$, $K$ is conformal. □

A *conjugate Möbius transformation* is a map of the form

$$z \mapsto \frac{a\bar{z} + b}{c\bar{z} + d},$$

where $a, b, c$, and $d$ are complex numbers such that $ad - bc \neq 0$. Of course, this map is just the composite $M \circ C$, where $M$ is the Möbius transformation in (A.2.1) and $C$ is complex-conjugation $C(z) = \bar{z}$.

There is an analogy in which Möbius and conjugate-Möbius transformations correspond to direct and opposite isometries, respectively. This is illustrated in the next result and in Proposition A.2.9 (see also the exercises in Section 6.5).

## Corollary A.2.6

The composite of a Möbius transformation and a conjugate-Möbius transformation (in either order) is conjugate-Möbius; the composite of two conjugate-Möbius transformations is Möbius.

### Proof

This follows immediately from Proposition A.2.1(ii) and the observation that, if $M$ is a Möbius transformation, so is $C \circ M \circ C$. For example, if $M$ and $N$ are Möbius transformations,

$$(M \circ C) \circ N = M \circ (C \circ N \circ C) \circ C,$$

since $C \circ C$ is the identity map. Since $M$ and $C \circ N \circ C$ are Möbius transformations, so is $M \circ (C \circ N \circ C)$, and hence $(M \circ C) \circ N$ is conjugate-Möbius. □

The properties of conjugate-Möbius transformations are easily deduced from those of Möbius transformations.

### Corollary A.2.7

Conjugate-Möbius transformations take Circles to Circles and are conformal.

### Proof

This follows from Propositions A.2.4 and A.2.5 and the fact that $C$, being the reflection in the real axis, is conformal and takes lines to lines and circles to circles. □

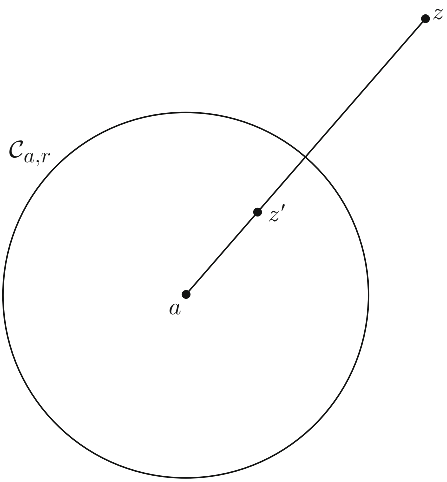
Conjugate-Möbius transformations include two important classes of geometric transformations. The first of these have already been discussed in Appendix 1, namely reflections in straight lines. We leave it to the reader to check that the reflection in the line with complex equation

$$\bar{b}z + b\bar{z} + c = 0,$$

where $b \in \mathbb{C}$, $c \in \mathbb{R}$ and $b \neq 0$, is

$$R(z) = \frac{-b\bar{z} - c}{\bar{b}}.$$

Since lines and circles are, in some sense, on the same footing in $\mathbb{C}_\infty$, it is natural to expect that there should be transformations that play the same role with respect to circles that reflections play with respect to straight lines. These are the *inversions*: the inversion $\mathcal{I}_{a,r}$ in the circle $\mathcal{C}_{a,r}$ with centre $a \in \mathbb{C}$ and radius $r > 0$ takes a point $z \in \mathbb{C}$ with $z \neq a$ to the point $z'$ on the radius of the circle passing through $z$ such that the product of the distances of $z$ and $z'$ from $a$ is $r^2$.




Thus, $z' - a = \rho(z - a)$ for some $\rho > 0$ and $|z' - a||z - a| = r^2$. These equations give $\rho = r^2/|z - a|^2$ and hence

$$\mathcal{I}_{a,r}(z) = a + \frac{r^2}{\bar{z} - \bar{a}}.$$

Since $\mathcal{I}_{a,r}$ is a conjugate-Möbius transformation, it takes Circles to Circles. In fact, since $\mathcal{I}_{a,r}(a) = \infty$, it is clear that $\mathcal{I}_{a,r}$ takes Circles passing through $a$ to lines and all other Circles to circles.

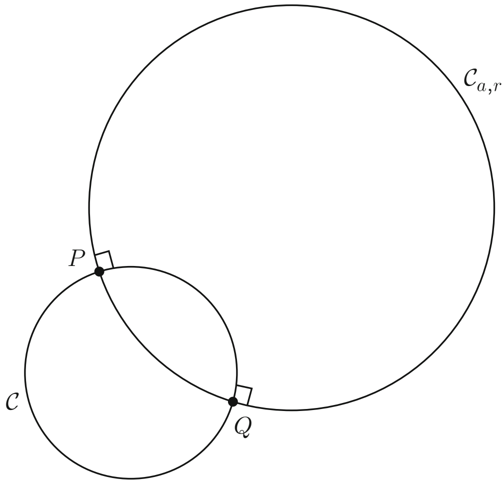
We shall also need the following property, which is not quite as obvious.

## Proposition A.2.8

The inversion $\mathcal{I}_{a,r}$ takes a circle $\mathcal{C}$ to itself if and only if $\mathcal{C}$ intersects $\mathcal{C}_{a,r}$ perpendicularly. In that case, $\mathcal{I}_{a,r}$ takes each of the two regions into which $\mathcal{C}$ divides the plane to themselves.

## Proof

Since $\mathcal{I}_{a,r}$ interchanges the interior and exterior of $\mathcal{C}_{a,r}$, it is clear that if $\mathcal{I}_{a,r}$ fixes $\mathcal{C}$ then $\mathcal{C}$ and $\mathcal{C}_{a,r}$ must intersect. Suppose for example that $\mathcal{C}$ is a circle that intersects $\mathcal{C}_{a,r}$ at points $P$ and $Q$ (the proof when $\mathcal{C}$ is a line is similar but easier). Then $\mathcal{I}_{a,r}$ interchanges the segments of $\mathcal{C}$ that are inside and outside $\mathcal{C}_{a,r}$, respectively. Since $\mathcal{I}_{a,r}$ is conformal, the angles made by the interior and exterior segments of $\mathcal{C}$ with $\mathcal{C}_{a,r}$ at $P$, say, must be equal. Since the sum of these angles is $\pi$, $\mathcal{C}_{a,r}$ and $\mathcal{C}$ must intersect perpendicularly at $P$.




Conversely, if $\mathcal{C}$ is a circle that intersects $\mathcal{C}_{a,r}$ at right angles, say at $P$ and $Q$, then $\mathcal{I}_{a,r}$ takes $\mathcal{C}$ to a circle that intersects $\mathcal{C}_{a,r}$ at right angles at $P$ and $Q$. But it is clear that there is a unique circle with these properties, namely $\mathcal{C}$ (for the centre of the circle must be the intersection of the tangent lines to $\mathcal{C}_{a,r}$ at $P$ and $Q$).

Assume now that $\mathcal{I}_{a,r}$ takes $\mathcal{C}$ to itself, and let $\mathcal{C}$ have centre $b$ and radius $s$. Since $\mathcal{C}_{a,r}$ and $\mathcal{C}$ intersect perpendicularly, $|b-a|^2 = r^2 + s^2$. Hence, $\mathcal{I}_{a,r}(b) = b - \frac{s^2}{\bar{b}-\bar{a}}$, so $|\mathcal{I}_{a,r}(b) - b| = \frac{s^2}{|b-a|} < s$ because $s < |b-a|$. Thus, $\mathcal{I}_{a,r}(b) \in \mathcal{D}$, the interior of $\mathcal{C}$. It follows that $\mathcal{I}_{a,r}$ takes every point $c \in \mathcal{D}$ to a point of $\mathcal{D}$. For suppose that $\mathcal{I}_{a,r}(c)$ is outside $\mathcal{C}$. Now, $\mathcal{I}_{a,r}$ takes the line segment joining $b$ and $c$ to a (smooth) curve $\boldsymbol{\gamma}$ joining the point $\mathcal{I}_{a,r}(b)$ inside $\mathcal{C}$ to the point $\mathcal{I}_{a,r}(c)$ outside $\mathcal{C}$, so $\boldsymbol{\gamma}$ must intersect $\mathcal{C}$ at a point $d$, say. Since $\mathcal{I}_{a,r}$ takes $\mathcal{C}$ to itself, $e = \mathcal{I}_{a,r}(d) \in \mathcal{C}$. But since $d$ is on $\boldsymbol{\gamma}$, $e$ is a point of the line segment with endpoints $b, c$. This is a contradiction. It follows that $\mathcal{I}_{a,r}$ takes $\mathcal{D}$ to itself, and since $\mathcal{I}_{a,r}$ is a bijection (indeed, it is equal to its own inverse), it must take the region exterior $\mathcal{C}$ to itself. □

The following result is analogous to Theorem A.1.5.

## Proposition A.2.9

Every Möbius transformation and every conjugate-Möbius transformation is a composite of reflections and inversions.

## Proof

By Proposition A.2.2, it is sufficient to prove that translations, complex dilations and the maps $C$ and $K$ are composites of reflections and inversions. By Theorem A.1.5, translations and rotations are composites of reflections, and $C$ is itself a reflection. We are therefore reduced to considering dilations $D_a$ with $a > 0$ and the map $K$. But $K$ is the composite of $C$ with the inversion $\mathcal{I}_{0,1}$, and $D_a$ is the composite $\mathcal{I}_{0,1} \circ \mathcal{I}_{0,a^{-1/2}}$. □

# Hints to selected exercises

1.1.6 Parametrize the ellipse by $\boldsymbol{\gamma}(t) = (p\cos t, q\sin t)$; for (i) show that the distances in question are $p(1 \pm \epsilon\cos t)$; for (ii) show that $\mathbf{n} = (q\cos t, p\sin t)$ is a vector perpendicular to the tangent line at $\boldsymbol{\gamma}(t)$ and that the distances in question are $|(\mathbf{p}-\mathbf{f}_1)\cdot\mathbf{n}|/\parallel \mathbf{n} \parallel$ and $|(\mathbf{p}-\mathbf{f}_2)\cdot\mathbf{n}|/\parallel \mathbf{n} \parallel$.

1.2.3 For the last part see the proof of Theorem 3.2.2.

1.4.5 (ii) The sequence in (i) converges to some limit $T_\infty \geq 0$ and $\boldsymbol{\gamma}$ is $T_\infty$-periodic; consider the sequence $\{T_r - T_\infty\}$. (iii) Use the mean value theorem.

1.4.6 Let $\boldsymbol{\gamma}(t) = (\gamma_1(t), \ldots, \gamma_n(t))$. Then each $\gamma_i$ is $T$-periodic and so by Exercise 1.4.5, if $\gamma_i$ is non-constant, it has a (positive) period, say $T_i$. By Exercise 1.4.4, if $\gamma_i$ is non-constant, $T = k_iT_i$ for some positive integer $k_i$. Let $k$ be the largest positive integer dividing each of the integers $k_i$. Show that $\boldsymbol{\gamma}$ is closed with period $T_0 = T/k$.

1.5.2 To guess the analogue of the condition on $f$ in Theorem 1.5.1, argue that $(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z})$ is perpendicular to the surface $f(x,y,z) = 0$, and then think about the condition that two planes intersect in a line. See Section 5.6 for a rigorous treatment.

1.5.3 This is easy.

2.2.3 It is enough to show that $\kappa_s$ changes sign when the curve is reflected in a line.

2.2.10 Think of the tangent line to $\boldsymbol{\gamma}$ at a point $\boldsymbol{\gamma}(s)$ as being rigidly fixed to $\boldsymbol{\gamma}$. When $\boldsymbol{\gamma}$ has rolled a distance $s$, the point initially at $\boldsymbol{\gamma}(s)$ has moved to the point $\mathbf{p} + s\mathbf{a}$ on $\ell$, and the tangent line to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(s)$ has become $\ell$. For the second part, note that $\rho_{-\theta(s)}(\dot{\boldsymbol{\gamma}}(s)) = \mathbf{a}$. Show that

(i) if $S$ is a skew-symmetric $n \times n$ matrix (i.e., $-S$ is equal to the transpose $S^t$ of $S$), and if $\mathbf{v}$ is any $n \times 1$ column matrix, then $\mathbf{v}^t S\mathbf{v} = 0$;

(ii) If $A$ is an orthogonal matrix (i.e., $A^t A = I$) whose entries are smooth functions of a parameter $s$, then $A^t \frac{dA}{ds}$ is skew-symmetric (the entries of $dA/ds$ are the derivatives of the entries of $A$).

2.3.2 Observe that it is enough to find *one* curve with curvature $\kappa$ and torsion $\tau$.

2.3.3 Assume that $\boldsymbol{\gamma}$ is unit-speed and show that $\mathbf{a} = \mathbf{t}\cos\theta + \mathbf{b}\sin\theta$.

2.3.4 If $\boldsymbol{\gamma}$ lies on the sphere of centre $\mathbf{a}$ and radius $r$, then $(\boldsymbol{\gamma}-\mathbf{a})\cdot(\boldsymbol{\gamma}-\mathbf{a}) = r^2$; now differentiate repeatedly. For the converse, consider $\boldsymbol{\gamma} + \rho\mathbf{n} + \dot{\rho}\sigma\mathbf{b}$.

2.3.6 Find a system of first-order differential equations satisfied by the dot products $\mathbf{v}_i \cdot \mathbf{v}_j$, and use the fact that such a system has a unique solution with given initial conditions.

3.2.1 Use the results of Appendix 1.

3.3.1 Use the inequality $2x_1x_2 \le x_1^2 + x_2^2$.

4.1.4 For the first part take $U$ to be an annulus.

5.2.3 For the first part, parametrize the line by $\boldsymbol{\gamma}(t) = \mathbf{a} + t\mathbf{b}$, and note that substituting into Eq. 5.1 gives a quadratic equation for $t$. For the second, take three points on each line and show that there is a quadric passing through all nine points.

5.6.1 Imitate the proof of Proposition 4.2.6.

5.6.4 If $\boldsymbol{\sigma}(u,v) = (f(u,v), g(u,v), h(u,v))$, show that the matrix $\begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}$ is invertible. Follow the proof of Proposition 4.2.6 to get open sets $V$ and $W$ and a smooth function $F^{-1} : V \to W$ such that $F^{-1}(f(u,v), g(u,v)) = (u,v)$ for all $(u,v) \in W$. Let $\varphi(x,y) = h(F^{-1}(x,y))$.

6.1.5 The length of the side given by $u = u_0$ is $\int_{v_0}^{v_1} \sqrt{G(u_0,v)}dv$.

6.3.8 Use Proposition A.2.5.

6.4.3 Choose a point inside the polygon and connect it to each vertex by a great circle arc.

6.5.3 It is enough to treat the case in which $\mathbf{p}$ is the north pole.

6.5.4 Call a unitary Möbius transformation as in part (i) *special* if $b \in \mathbb{R}$. Prove that every unitary Möbius transformation is a composite of special unitary Möbius transformations.

7.1.2 By computing expressions such as $(\boldsymbol{\sigma}_u \cdot \mathbf{N})_u$, prove that $\mathbf{N}_u$ and $\mathbf{N}_v$ are perpendicular to $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$, and deduce that the unit normal $\mathbf{N}$ of $\boldsymbol{\sigma}$ is a constant vector.

7.3.4 Use Exercise 6.4.7.

7.4.3 Use Example 7.4.7.

8.1.6 Any $2 \times 2$ matrix $A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ satisfies the equation $A^2 - (a+d)A + (ad - bc)I = 0$.

8.1.8 Inspect the proof of Theorem 8.1.6.

8.2.3 Consider separately the cases in which $EN = GL$ and $EN \neq GL$.

8.2.5 Use Exercise 8.2.2.

8.2.6 (i) Differentiate Eq. 8.5.

8.2.7 Use Proposition 8.1.2, the remarks following Proposition 8.2.1 and the solution of Exercise 8.1.1.

8.2.8 Use Exercises 6.1.4 and 7.1.3.

8.3.1 For (iii) use polar coordinates in the disc.

8.4.1 Use Theorem 8.2.4 and Propositions 7.3.5 and 8.4.3.

8.5.2 Use Exercise 8.1.1.

8.5.3 Use Exercise 8.2.4 and the proof of Proposition 8.1.2.

9.1.1 Use Propositions 9.1.4 and 9.1.6.

9.1.3 Use the solution of Exercise 4.2.7.

9.1.4 Take the ellipsoid to be $\frac{x^2}{p^2} + \frac{y^2}{q^2} + \frac{z^2}{r^2} = 1$ and note that $\boldsymbol{\gamma}(t) = (f(t), g(t), h(t))$ is a geodesic if and only if $(\ddot{f}, \ddot{g}, \ddot{h}) = \lambda(\frac{f}{p^2}, \frac{g}{q^2}, \frac{h}{r^2})$ for some scalar $\lambda(t)$.

9.1.5 Use Exercise 8.2.2.

9.2.2 Use Exercise 6.2.1.

9.3.3 The condition for a self-intersection is that, for some value of $w > 1$, the two values of $v$ satisfying Eq. 9.14 should differ by an integer multiple of $2\pi$.

9.4.2 Consider the intersection of $S^2$ with planes passing through $\mathbf{p}$ and $\mathbf{q}$.

9.4.3 (i) Use L'Hospital's rule.

9.5.2 (ii) If $\boldsymbol{\gamma}$ leaves $\boldsymbol{\sigma}$, it must cross the geodesic circle with centre $\mathbf{p}$ and radius $R$, say at $\mathbf{q}'$. Then its length is greater than that of the part of $\boldsymbol{\gamma}$ between $\mathbf{p}$ and $\mathbf{q}'$.

10.1.1 Compute the matrix of the Weingarten map.

10.1.3 To solve the differential equation, put $P = Lw^2$.

10.2.2 Use Exercises 6.1.4 and 9.5.1.

10.2.4 Use the geodesic equations, Exercise 9.5.1 and Corollary 10.2.3(ii).

10.3.1 As in the proof of Theorem 10.3.4, if $\kappa_1$ has a local maximum at a point $\mathbf{p}$ of the surface, then $\kappa_2 = 2H - \kappa_1$ has a local minimum there.

10.4.2 Consider the matrix $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$.

10.4.3 Use Theorem 9.2.1.

11.1.4 For the first part use Propositions 11.1.4 and A.2.3. For the second, note that $2R$ is the hyperbolic distance between $i(b+r)$ and $i(b-r)$, etc.

11.2.1 It is enough to consider the case in which $a$ and $b$ are on the imaginary axis.

11.2.2 Use Exercise 11.1.1.

11.2.3 Reduce to the case in which $l$ is the imaginary axis and use Exercise 11.1.3.

11.2.4 (i) If $P$ is any point on $\mathcal{H}$, consider the hyperbolic lines passing through $P$ perpendicular to $l$ and $m$. (iii) Let $l$ and $m$ be as in Exercise 11.2.3, and let $F$ be any isometry of $\mathcal{H}$. The proof of Proposition 11.2.3 shows that there is a composite $G$ of elementary isometries that takes $F(i)$ to $i$ and $F(l)$ to $l$. Now use (ii).

11.2.5 Imitate the proof of Proposition A.2.2.

11.3.2 Compare the solution of Exercise 11.2.4.

11.3.3 Use Exercise 11.2.5.

11.3.4 Use Exercise 11.2.5.

11.3.6 Prove this first for a right-angled triangle by applying the cosine rule in two different ways. Then deduce the general case by drawing the geodesic through $a$ that intersects the geodesic through $b$ and $c$ perpendicularly – you will have to consider separately the cases in which the point of intersection does, or does not, lie between $b$ and $c$.

11.3.8 Treat the right-angled case first using Exercise 11.3.7 and then proceed as in Exercise 11.3.6.

11.4.2 Work in $\mathcal{H}$ and assume that $l$ is the imaginary axis.

11.4.3 Work in $\mathcal{D}_P$ and assume that $a = 0$, $b \in \mathbb{R}$.

11.5.3 For the existence, note that $M(z) = (a, b; c, z)$ is a Möbius transformation that takes $(a, b, c)$ to $(\infty, 0, 1)$. For the uniqueness, note that if $M$ is a Möbius transformation that takes $(a, b, c)$ to $(\infty, 0, 1)$, then $M(z) = (\infty, 0; 1, M(z)) = (a, b; c, z)$.

12.1.1 Use Proposition 8.2.9.

12.1.3 Use Proposition 8.6.1 and Exercise 12.1.1.

12.2.4 Use Exercise 8.5.1.

12.3.1 Use Exercise 8.1.6 and Proposition 8.2.9.

12.5.1 Reparametrize by putting $\zeta = e^{\tilde{\zeta}}$.

12.5.3 (ii) Use the Weierstrass representation and inspect Eqs. 12.25 and 12.26.

13.1.2 For the last part, use Theorem 13.1.2.

13.3.2 Define $\psi_k(t) = \psi(nt - k)$, where $\psi$ is the function defined in Exercise 8.5.3(iii), and let $\varphi_k = \psi_k/(\psi_1 + \cdots + \psi_{n-1})$.

13.5.2 $E \leq \frac{1}{2}V(V - 1)$.

13.7.1 Take $\boldsymbol{\gamma}$ in Definition 13.7.2 to be the unit circle and use de Moivre's theorem.

13.7.3 If $\tilde{\boldsymbol{\xi}}$ is another reference tangent vector field, and $\theta$ is the angle between $\tilde{\boldsymbol{\xi}}$ and $\boldsymbol{\xi}$, then $d\theta/ds = -(1-\rho^2)^{-1/2}\dot{\rho}$, where $\rho = \cos\theta$. Now use Green's theorem to show that $\int_0^{\ell(\gamma)} (d\theta/ds)ds = 0$.

13.8.1 Show that the critical point is a local maximum (respectively local minimum) if and only if $\mathbf{v}^t\mathcal{H}\mathbf{v} < 0$ (resp. $> 0$) for all non-zero $2\times1$ matrices $\mathbf{v}$.

# Solutions

## Chapter 1

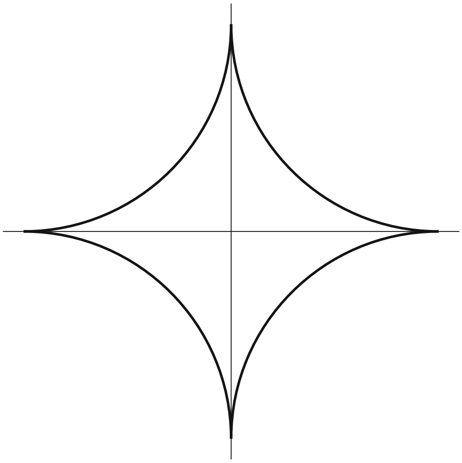
1.1.1 It is a parametrization of the part of the parabola with $x \geq 0$.

1.1.2 (i) $\boldsymbol{\gamma}(t) = (\tan t, \sec t)$ with $-\pi/2 < t < \pi/2$ and $\pi/2 < t < 3\pi/2$. Note that $\boldsymbol{\gamma}$ is defined on the union of two disjoint intervals: this corresponds to the fact that the hyperbola $y^2 - x^2 = 1$ is in two pieces, where $y \geq 1$ and where $y \leq -1$. (ii) $\boldsymbol{\gamma}(t) = (2\cos t, 3\sin t)$.

1.1.3 (i) $x + y = 1$. (ii) $y = (\ln x)^2$.

1.1.4 (i) $\dot{\boldsymbol{\gamma}}(t) = \sin 2t(-1, 1)$. (ii) $\dot{\boldsymbol{\gamma}}(t) = (e^t, 2t)$.

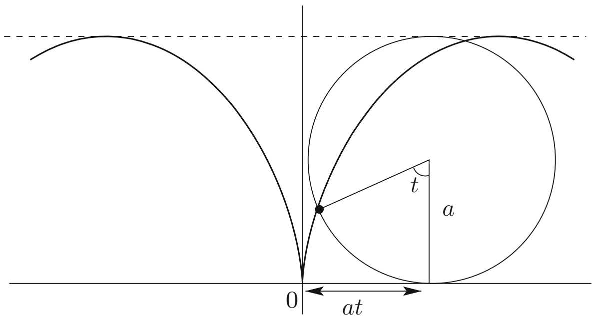
1.1.5 $\dot{\boldsymbol{\gamma}}(t) = 3\sin t \cos t(-\cos t, \sin t)$ vanishes where $\sin t = 0$ or $\cos t = 0$, i.e., $t = n\pi/2$ where $n$ is any integer. These points correspond to the four cusps of the astroid (see Exercise 1.3.3).



1.1.6 (i) The squares of the distances from $\mathbf{p}$ to the foci are $(p\cos t \pm \epsilon p)^2 + q^2\sin^2 t = (p^2 - q^2)\cos^2 t \pm 2\epsilon p^2 \cos t + p^2 = p^2(1 \pm \epsilon\cos t)^2$, so the sum of the distances is $2p$.

(ii) $\dot{\boldsymbol{\gamma}} = (-p\sin t, q\cos t)$ so if $\mathbf{n} = (q\cos t, p\sin t)$ then $\mathbf{n}\cdot\dot{\boldsymbol{\gamma}} = 0$. Hence the distances from the foci to the tangent line at $\boldsymbol{\gamma}(t)$ are $\frac{(p\cos t \mp \epsilon p, q\sin t)\cdot\mathbf{n}}{\|\mathbf{n}\|} = \frac{pq(1\mp\epsilon\cos t)}{(p^2\sin^2 t + q^2\cos^2 t)^{1/2}}$ and their product is $\frac{p^2q^2(1-\epsilon^2\cos^2 t)}{(p^2\sin^2 t+q^2\cos^2 t)} = q^2$.

(iii) It is enough to prove that $\frac{(\mathbf{p}-\mathbf{f}_1)\cdot\mathbf{n}}{\|\mathbf{p}-\mathbf{f}_1\|} = \frac{(\mathbf{p}-\mathbf{f}_2)\cdot\mathbf{n}}{\|\mathbf{p}-\mathbf{f}_2\|}$. Computation shows that both sides are equal to $q$.




1.1.7 When the circle has rotated through an angle $t$, its centre has moved to $(at, a)$, so the point on the circle initially at the origin is now at the point $(a(t - \sin t), a(1 - \cos t))$ (see the diagram above).

1.1.8 Suppose that a point $(x, y, z)$ lies on the cylinder if $x^2 + y^2 = 1/4$ and on the sphere if $(x+\frac{1}{2})^2+y^2+z^2 = 1$. From the second equation, $-1 \le z \le 1$ so let $z = \sin t$. Subtracting the two equations gives $x + \frac{1}{4} + \sin^2 t = \frac{3}{4}$, so $x = \frac{1}{2} - \sin^2 t = \cos^2 t - \frac{1}{2}$. From either equation we then get $y = \sin t\cos t$ (or $y = -\sin t\cos t$, but the two solutions are interchanged by $t \mapsto \pi - t$).

1.1.9 $\dot{\boldsymbol{\gamma}} = (-2\sin t + 2\sin 2t, 2\cos t - 2\cos 2t) = \sqrt{2}(\sqrt{2} - 1, 1)$ at $t = \pi/4$. So the tangent line is $y - (\frac{1}{\sqrt{2}} - 1) = (x - \sqrt{2})/(\sqrt{2} - 1)$ and the normal line is $y - (\frac{1}{\sqrt{2}} - 1) = -(x - \sqrt{2})(\sqrt{2} - 1)$.

1.2.1 $\dot{\boldsymbol{\gamma}}(t) = (1, \sinh t)$ so $\| \dot{\boldsymbol{\gamma}} \| = \cosh t$ and the arc-length is $s = \int_0^t \cosh u\, du = \sinh t$.

1.2.2 (i) $\| \dot{\boldsymbol{\gamma}} \|^2 = \frac{1}{4}(1 + t) + \frac{1}{4}(1 - t) + \frac{1}{2} = 1$.
(ii) $\| \dot{\boldsymbol{\gamma}} \|^2 = \frac{16}{25}\sin^2 t + \cos^2 t + \frac{9}{25}\sin^2 t = \cos^2 t + \sin^2 t = 1$.

1.2.3 Denoting $d/d\theta$ by a dot, $\dot{\boldsymbol{\gamma}} = (\dot{r}\cos\theta - r\sin\theta, \dot{r}\sin\theta + r\cos\theta)$ so $\| \dot{\boldsymbol{\gamma}} \|^2 = \dot{r}^2 + r^2$. Hence, $\boldsymbol{\gamma}$ is regular unless $r = \dot{r} = 0$ for some value of $\theta$. It is unit-speed if and only if $\dot{r}^2 = 1 - r^2$, which gives $r = \pm 1$ or $r = \pm\sin(\theta + \alpha)$ for some constant $\alpha$. To see that the latter is the equation of a circle of radius 1/2, see the diagram in the proof of Theorem 3.2.2.

1.2.4 Since $\mathbf{u}$ is a unit vector, $|\dot{\boldsymbol{\gamma}}\cdot\mathbf{u}| = \| \dot{\boldsymbol{\gamma}} \| \cos\theta$, where $\theta$ is the angle between $\dot{\boldsymbol{\gamma}}$ and $\mathbf{u}$, so $\dot{\boldsymbol{\gamma}}\cdot\mathbf{u} \leq \| \dot{\boldsymbol{\gamma}} \|$. Then, $(\mathbf{q}-\mathbf{p})\cdot\mathbf{u} = (\boldsymbol{\gamma}(b)-\boldsymbol{\gamma}(a))\cdot\mathbf{u} = \int_a^b \dot{\boldsymbol{\gamma}}\cdot\mathbf{u}\,dt \leq \int_a^b \| \dot{\boldsymbol{\gamma}} \| \; dt$. Taking $\mathbf{u} = (\mathbf{q}-\mathbf{p})/ \| \mathbf{q}-\mathbf{p} \|$ gives the result.

1.3.1 (i) $\dot{\boldsymbol{\gamma}} = \sin 2t(-1,1)$ vanishes when $t$ is an integer multiple of $\pi/2$, so $\boldsymbol{\gamma}$ is not regular. (ii) $\boldsymbol{\gamma}$ is regular since $\dot{\boldsymbol{\gamma}} \neq \mathbf{0}$ for $0 < t < \pi/2$. (iii) $\dot{\boldsymbol{\gamma}} = (1, \sinh t)$ is obviously never zero, so $\boldsymbol{\gamma}$ is regular.

1.3.2 $x = r\cos\theta = \sin^2\theta$, $y = r\sin\theta = \sin^2\theta\tan\theta$, so the parametrization in terms of $\theta$ is $\theta \mapsto (\sin^2\theta, \sin^2\theta\tan\theta)$. Since $\theta \mapsto \sin\theta$ is a bijective smooth map $(-\pi/2, \pi/2) \to (-1,1)$, with smooth inverse $t \mapsto \sin^{-1} t$, $t = \sin\theta$ is a reparametrization map. Since $\sin^2\theta = t^2$, $\sin^2\theta\tan\theta = t^3/\sqrt{1-t^2}$, so the reparametrized curve is as stated.

1.3.3 (i) $\dot{\boldsymbol{\gamma}} = \mathbf{0}$ at $t = 0 \iff m$ and $n$ are both $\geq 2$. If $m > 3$ the first components of $\ddot{\boldsymbol{\gamma}}$ and $\dddot{\boldsymbol{\gamma}}$ are both 0 at $t = 0$ so $\ddot{\boldsymbol{\gamma}}$ and $\ddot{\boldsymbol{\gamma}}$ are linearly dependent at $t = 0$; similarly if $n > 3$. So there are four cases: if $(m,n) = (2,2)$ or $(3,3)$ then either $\ddot{\boldsymbol{\gamma}}$ or $\dddot{\boldsymbol{\gamma}}$ is zero at $t = 0$, so the only possibilities for an ordinary cusp are $(m,n) = (2,3)$ and $(3,2)$ and then $\ddot{\boldsymbol{\gamma}}$ and $\dddot{\boldsymbol{\gamma}}$ are easily seen to be linearly independent at $t = 0$. (ii) Using the parametrization $\boldsymbol{\gamma}(t) = \left(t^2, \frac{t^3}{\sqrt{1-t^2}}\right)$, we get $\dot{\boldsymbol{\gamma}} = \mathbf{0}$, $\ddot{\boldsymbol{\gamma}} = (2,0)$, $\dddot{\boldsymbol{\gamma}} = (0,6)$ at $t = 0$ so the origin is an ordinary cusp. (iii) Let $\tilde{\boldsymbol{\gamma}}(\tilde{t})$ be a reparametrization of $\boldsymbol{\gamma}(t)$, and suppose $\boldsymbol{\gamma}$ has an ordinary cusp at $t = t_0$. Then, at $t = t_0$, $d\tilde{\boldsymbol{\gamma}}/d\tilde{t} = (d\boldsymbol{\gamma}/dt)(dt/d\tilde{t}) = 0$, $d^2\tilde{\boldsymbol{\gamma}}/d\tilde{t}^2 = (d^2\boldsymbol{\gamma}/dt^2)(dt/d\tilde{t})^2$, $d^3\tilde{\boldsymbol{\gamma}}/d\tilde{t}^3 = (d^3\boldsymbol{\gamma}/dt^3)(dt/d\tilde{t})^3 + 3(d^2\boldsymbol{\gamma}/dt^2)(dt/d\tilde{t})(d^2t/d\tilde{t}^2)$. Using the fact that $dt/d\tilde{t} \neq 0$, it is easy to see that $d^2\tilde{\boldsymbol{\gamma}}/d\tilde{t}^2$ and $d^3\tilde{\boldsymbol{\gamma}}/d\tilde{t}^3$ are linearly independent when $t = t_0$.

1.3.4 (i) If $\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}(\varphi(t))$, let $\psi$ be the inverse of the reparametrization map $\varphi$. Then $\tilde{\boldsymbol{\gamma}}(\psi(t)) = \boldsymbol{\gamma}(\varphi(\psi(t))) = \boldsymbol{\gamma}(t)$. (ii) If $\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}(\varphi(t))$ and $\hat{\boldsymbol{\gamma}}(t) = \tilde{\boldsymbol{\gamma}}(\psi(t))$, where $\varphi$ and $\psi$ are reparametrization maps, then $\hat{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}((\varphi \circ \psi)(t))$ and $\varphi \circ \psi$ is a reparametrization map because it is smooth and $\frac{d}{dt}(\varphi(\psi(t)) = \dot{\varphi}(\psi(t))\dot{\psi}(t) \neq 0$ as $\dot{\varphi}$ and $\dot{\psi}$ are both $\neq 0$.

1.4.1 It is closed because $\boldsymbol{\gamma}(t + 2\pi) = \boldsymbol{\gamma}(t)$ for all $t$. Suppose that $\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(u)$. Then $\cos^3 t(\cos 3t, \sin 3t) = \cos^3 u(\cos 3u, \sin 3u)$. Taking lengths gives $\cos^3 t = \pm\cos^3 u$ so $\cos t = \pm\cos u$, so $u = t, \pi - t, \pi + t$ or $2\pi - t$ (up to adding multiples of $2\pi$). The second possibility forces $t = n\pi/3$ for some integer $n$ and the third possibility is true for all $t$. Hence, the period is $\pi$ and for the self-intersections we need only consider $t = \pi/3, 2\pi/3$, giving $u = 2\pi/3, \pi/3$, respectively. Hence, there is a unique self-intersection at $\boldsymbol{\gamma}(\pi/3) = (-1/8, 0)$.

1.4.2 The curve $\tilde{\boldsymbol{\gamma}}(t) = (\cos(t^3 + t), \sin(t^3 + t))$ is a reparametrization of the circle $\boldsymbol{\gamma}(t) = (\cos t, \sin t)$ but it is not closed.

1.4.3 If $\boldsymbol{\gamma}$ is $T$-periodic then it is $kT$-periodic for all $k \neq 0$ (this can be proved by induction on $k$ if $k > 0$, or on $-k$ if $k < 0$). If $\boldsymbol{\gamma}$ is $T_1$-periodic and $T_2$-periodic then it is $k_1T_1$- and $k_2T_2$-periodic for all non-zero integers $k_1, k_2$, so $\boldsymbol{\gamma}(t + k_1T_1 + k_2T_2) = \boldsymbol{\gamma}(t + k_1T_1)$ as $\boldsymbol{\gamma}$ is $k_2T_2$-periodic, which $= \boldsymbol{\gamma}(t)$ as $\boldsymbol{\gamma}$ is $k_1T_1$-periodic.

1.4.4 If $\boldsymbol{\gamma}$ is $T$-periodic write $T = kT_0 + T_1$ where $k$ is an integer and $0 \leq T_1 < T_0$. By Exercise 1.4.3 $\boldsymbol{\gamma}$ is $T_1$-periodic; if $T_1 > 0$ this contradicts the definition of $T_0$.

1.4.5 (i) Choose $T_1 > 0$ such that $\boldsymbol{\gamma}$ is $T_1$-periodic; then $T_1$ is not the smallest positive number with this property, so there is a $T_2 > 0$ such that $\boldsymbol{\gamma}$ is $T_2$-periodic. Iterating this argument gives the desired sequence. (ii) The sequence $\{T_r\}_{r\geq 1}$ in (i) is decreasing and bounded below, so must converge to some $T_\infty \geq 0$. Then $\boldsymbol{\gamma}$ is $T_\infty$-periodic because (using continuity of $\boldsymbol{\gamma}$) $\boldsymbol{\gamma}(t + T_\infty) = \lim_{r\to\infty} \boldsymbol{\gamma}(t + T_r) = \lim_{r\to\infty} \boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(t)$. By Exercise 1.4.3, $\boldsymbol{\gamma}$ is $(T_r - T_\infty)$-periodic for all $r \geq 1$, and this sequence of positive numbers converges to 0. (iii) If $\{T_r\}$ is as in (i) and $T_r \to 0$ as $r \to \infty$, then by the mean value theorem $0 = (f(t + T_r) - f(t))/T_r = \dot{f}(t + \lambda T_r)$ for some $0 < \lambda < 1$. Letting $r \to \infty$ gives $\dot{f}(t) = 0$ for all $t$, so $f$ is constant.

1.4.6 Following the hint, since $T_0 = (k_i/k)T_i$ is an integer multiple of $T_i$, each $\gamma_i$ is $T_0$-periodic. Let $\mathcal{T}$ be the union of the finite sets of real numbers $\{T_i, 2T_i, \ldots, k_iT_i\}$ over all $i$ such that $\gamma_i$ is not constant, and let $\mathcal{P} = \{T' \in \mathcal{T} \,|\, \boldsymbol{\gamma} \text{ is } T'\text{-periodic}\}$. Then $\mathcal{P}$ is finite (because $\mathcal{T}$ is) and non-empty (because $T \in \mathcal{P}$). The smallest element of $\mathcal{P}$ is the smallest positive number $T_0'$ such that $\boldsymbol{\gamma}$ is $T_0'$-periodic (since if $\boldsymbol{\gamma}$ is $T'$-periodic either $T' > T$ or $T' \in \mathcal{P}$). By Exercise 1.4.4, $T_0 = k'T_0'$ for some integer $k'$ and then there are integers $k_i'$ such that $T_0' = k_i'T_i$ for all $i$ such that $\gamma_i$ is not constant. Then, $k_iT_i/k = k'k_iT_i$ so $kk'$ divides each $k_i$. As $k$ is the largest such divisor, $k' = 1$, so $T_0 = T_0'$.

1.5.1 $x(1 - x^2) \geq 0 \iff x \leq -1$ or $0 \leq x \leq 1$ so the curve is in (at least) two pieces. The parametrization is defined for $t \leq -1$ and $0 \leq t \leq 1$ and it covers the part of the curve with $y \geq 0$.

1.5.2 If $\boldsymbol{\gamma}(t) = (x(t), y(t), z(t))$ is a curve in the surface $f(x, y, z) = 0$, differentiating $f(x(t), y(t), z(t)) = 0$ with respect to $t$ gives $\dot{x}f_x + \dot{y}f_y + \dot{z}f_z = 0$, so $\dot{\boldsymbol{\gamma}}$ is perpendicular to $\boldsymbol{\nabla} f = (f_x, f_y, f_z)$. Since this holds for every curve in the surface, $\boldsymbol{\nabla} f$ is perpendicular to the surface. The surfaces $f = 0$ and $g = 0$ should intersect in a curve if the vectors $\boldsymbol{\nabla} f$ and $\boldsymbol{\nabla} g$ are not parallel at any point of the intersection.

1.5.3 Let $\boldsymbol{\gamma}(t) = (u(t), v(t), w(t))$ be a regular curve in $\mathbb{R}^3$. At least one of $\dot{u}, \dot{v}, \dot{w}$ is non-zero at each value of $t$. Suppose that $\dot{u}(t_0) \neq 0$ and $x_0 = u(t_0)$. As in the 'proof' of Theorem 1.5.2, there is a smooth function $h(x)$ defined for $x$ near $x_0$ such that $t = h(x)$ is the unique solution of $x = u(t)$ for each $t$ near $t_0$. Then, for $t$ near $t_0$, $\boldsymbol{\gamma}(t)$ is contained in the level curve $f(x, y, z) = g(x, y, z) = 0$, where $f(x, y, z) = y - v(h(x))$ and $g(x, y, z) = z - w(h(x))$. The functions $f$ and $g$ satisfy the conditions in the previous exercise, since $\nabla f = (-\dot{v}h', 1, 0)$, $\nabla g = (-\dot{w}h', 0, 1)$, a dash denoting $d/dx$.

# Chapter 2

2.1.1 (i) $\boldsymbol{\gamma}$ is unit-speed (Exercise 1.2.2(i)) so $\kappa = \parallel \ddot{\boldsymbol{\gamma}} \parallel = \parallel (\frac{1}{4}(1+t)^{-1/2}, \frac{1}{4}(1-t)^{-1/2}, 0) \parallel = \frac{1}{\sqrt{8(1-t^2)}}$. (ii) $\boldsymbol{\gamma}$ is unit-speed (Exercise 1.2.2(ii)) so $\kappa = \parallel \ddot{\boldsymbol{\gamma}} \parallel = \parallel (-\frac{4}{5}\cos t, \sin t, \frac{3}{5}\cos t) \parallel = 1$.
(iii) $\kappa = \frac{\|(1,\sinh t,0)\times(0,\cosh t,0)\|}{\|(1,\sinh t,0)\|^3} = \frac{\cosh t}{\cosh^3 t} = \operatorname{sech}^2 t$ using Proposition 2.1.2.
(iv) $(-3\cos^2 t\sin t, 3\sin^2 t\cos t, 0) \times (-3\cos^3 t + 6\cos t\sin^2 t, 6\sin t\cos^2 t - 3\sin^3 t, 0) = (0, 0, -9\sin^2 t\cos^2 t)$, so $\kappa = \frac{\|(0,0,-9\sin^2 t\cos^2 t)\|}{\|(-3\cos^2 t\sin t, 3\sin^2 t\cos t, 0)\|^3} = \frac{1}{3|\sin t\cos t|}$. This becomes infinite when $t$ is an integer multiple of $\pi/2$, i.e., at the four cusps $(\pm 1, 0)$ and $(0, \pm 1)$ of the astroid.

2.1.2 The proof of Proposition 1.3.5 shows that, if $\mathbf{v}(t)$ is a smooth (vector) function of $t$, then $\parallel \mathbf{v}(t) \parallel$ is a smooth (scalar) function of $t$ provided $\mathbf{v}(t)$ is non-zero for all $t$. The result now follows from the formula in Proposition 2.1.2. The curvature of the regular curve $\boldsymbol{\gamma}(t) = (t, t^3)$ is $\kappa(t) = 6|t|/(1+9t^4)^{3/2}$, which is not differentiable at $t = 0$.

2.2.1 Differentiate $\mathbf{t} \cdot \mathbf{n}_s = 0$ and use $\dot{\mathbf{t}} = \kappa_s \mathbf{n}_s$.

2.2.2 If $\boldsymbol{\gamma}$ is smooth, $\mathbf{t} = \dot{\boldsymbol{\gamma}}$ is smooth and hence so are $\dot{\mathbf{t}}$ and $\mathbf{n}_s$ (since $\mathbf{n}_s$ is obtained by applying a rotation to $\mathbf{t}$). So $\kappa_s = \dot{\mathbf{t}} \cdot \mathbf{n}_s$ is smooth.

2.2.3 For the first part, from the results in Appendix 2 it suffices to show that $\tilde{\kappa}_s = -\kappa_s$ if $M$ is the reflection in a straight line $l$. But this is clear: if we take the fixed angle $\varphi_0$ in Proposition 2.2.1 to be the angle between $l$ and the positive $x$-axis, then (in the obvious notation) $\tilde{\varphi} = -\varphi$. Conversely, if $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ have the same non-zero curvature, their signed curvatures are either the same or differ in sign. In the first case the curves differ by a direct isometry by Theorem 2.2.5; in the latter case, applying a reflection to one curve gives two curves with the same signed curvature, and these curves then differ by a direct isometry, so the original curves differ by an opposite isometry.

2.2.4 The first part is obvious as the effect of the dilation is to multiply $s$ by $a$ and leave $\varphi$ unchanged. For the second part, consider the small piece of the chain between the points with arc-length $s$ and $s+\delta s$. The net horizontal force on this piece is (in the obvious notation) $\delta(T\cos\varphi)$, and as this must vanish $T\cos\varphi$ must be a constant, say $\lambda$. The net vertical force is $\delta(T\sin\varphi)$, and this must balance the weight of the piece of chain, which is a constant multiple of $\delta s$. This shows that $T\sin\varphi = \mu s + \nu$ for some constants $\mu, \nu$, and $\nu$ must be zero because $\varphi = s = 0$ at the lowest point of $\mathcal{C}$. From $T\cos\varphi = \lambda$, $T\sin\varphi = \mu s$, we get $\tan\varphi = s/a$ where $a = \lambda/\mu$. Hence, $\sec^2\varphi\frac{d\varphi}{ds} = 1/a$, so the signed curvature is $\kappa_s = d\varphi/ds = 1/a\sec^2\varphi = 1/a(1+\tan^2\varphi) = \frac{1}{a}(1+s^2/a^2)^{-1}$. Using the first part and Example 2.2.4 gives the result.

2.2.5 We have $d\boldsymbol{\gamma}^\lambda/dt = d\boldsymbol{\gamma}/dt + \lambda d\mathbf{n}_s/dt = (1-\lambda\kappa_s)ds/dt\,\mathbf{t}$, so the arc-length $s^\lambda$ of $\boldsymbol{\gamma}^\lambda$ satisfies $ds^\lambda/dt = |1-\lambda\kappa_s|ds/dt$. The unit tangent vector of $\boldsymbol{\gamma}^\lambda$ is $\mathbf{t}^\lambda = (d\boldsymbol{\gamma}^\lambda/dt)/(ds^\lambda/dt) = \epsilon\mathbf{t}$, hence the signed unit normal of $\boldsymbol{\gamma}^\lambda$ is $\mathbf{n}_s^\lambda = \epsilon\mathbf{n}_s$. Then, the signed curvature $\kappa_s^\lambda$ of $\boldsymbol{\gamma}^\lambda$ is given by $\kappa_s^\lambda\mathbf{n}_s^\lambda = d\mathbf{t}^\lambda/ds^\lambda = (d\mathbf{t}^\lambda/dt)/|1-\lambda\kappa_s|(ds/dt) = \epsilon|1-\lambda\kappa_s|^{-1}d\mathbf{t}/ds = \kappa_s(1-\lambda\kappa_s)^{-1}\mathbf{n}_s = \epsilon\kappa_s(1-\lambda\kappa_s)^{-1}\mathbf{n}_s^\lambda = \kappa_s|1-\lambda\kappa_s|^{-1}\mathbf{n}_s^\lambda$.

2.2.6 $\boldsymbol{\epsilon}(s_0)$ lies on the perpendicular bisector of the line joining $\boldsymbol{\gamma}(s_0)$ and $\boldsymbol{\gamma}(s_0+\delta s)$. So

$$(\boldsymbol{\epsilon}(s_0) - \frac{1}{2}(\boldsymbol{\gamma}(s_0) + \boldsymbol{\gamma}(s_0+\delta s)))\cdot(\boldsymbol{\gamma}(s_0+\delta s) - \boldsymbol{\gamma}(s_0)) = 0.$$

Using Taylor's theorem, and discarding terms involving powers of $\delta s$ higher than the second, this gives (with all quantities evaluated at $s_0$) $(\boldsymbol{\epsilon}-\boldsymbol{\gamma})\cdot\dot{\boldsymbol{\gamma}}\delta s + \frac{1}{2}(\boldsymbol{\epsilon}\cdot\ddot{\boldsymbol{\gamma}} - 1 - \boldsymbol{\gamma}\cdot\ddot{\boldsymbol{\gamma}})(\delta s)^2 = 0$. This must also hold when $\delta s$ is replaced by $-\delta s$; adding and subtracting the two equations give $(\boldsymbol{\epsilon}-\boldsymbol{\gamma})\cdot\dot{\boldsymbol{\gamma}} = 0$ and $(\boldsymbol{\epsilon}-\boldsymbol{\gamma})\cdot\ddot{\boldsymbol{\gamma}} = 1$. The first equation gives $\boldsymbol{\epsilon} = \boldsymbol{\gamma} + \lambda\mathbf{n}_s$ for some scalar $\lambda$, and since $\ddot{\boldsymbol{\gamma}} = \kappa_s\mathbf{n}_s$ the second gives $\lambda = 1/\kappa_s$.

2.2.7 The tangent vector of $\boldsymbol{\epsilon}$ is $\mathbf{t} + \frac{1}{\kappa_s}(-\kappa_s\mathbf{t}) - \frac{\dot\kappa_s}{\kappa_s^2}\mathbf{n}_s = -\frac{\dot\kappa_s}{\kappa_s^2}\mathbf{n}_s$ so its arc-length is $u = \int \parallel \dot{\boldsymbol{\epsilon}} \parallel \, ds = \int \frac{\dot\kappa_s}{\kappa_s^2}ds = u_0 - \frac{1}{\kappa_s}$, where $u_0$ is a constant. Hence, the unit tangent vector of $\boldsymbol{\epsilon}$ is $-\mathbf{n}_s$ and its signed unit normal is $\mathbf{t}$. Since $-d\mathbf{n}_s/du = \kappa_s\mathbf{t}/(du/ds) = \frac{\kappa_s^3}{\dot\kappa_s}\mathbf{t}$, the signed curvature of $\boldsymbol{\epsilon}$ is $\kappa_s^3/\dot\kappa_s$.

Any point on the normal line to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(s)$ is $\boldsymbol{\gamma}(s) + \lambda\mathbf{n}_s(s)$ for some $\lambda$. Hence, the normal line intersects $\boldsymbol{\epsilon}$ at the point $\boldsymbol{\epsilon}(s)$, where $\lambda = 1/\kappa_s(s)$, and since the tangent vector of $\boldsymbol{\epsilon}$ there is parallel to $\mathbf{n}_s(s)$ by the first part, the normal line is tangent to $\boldsymbol{\epsilon}$ at $\boldsymbol{\epsilon}(s)$.

Denoting $d/dt$ by a dash, $\boldsymbol{\gamma}' = a(1-\cos t, \sin t)$ so the arc-length $s$ of $\boldsymbol{\gamma}$ is given by $ds/dt = 2a\sin(t/2)$ and $\mathbf{t} = d\boldsymbol{\gamma}/ds = (\sin(t/2), \cos(t/2))$. So $\mathbf{n}_s = (-\cos(t/2), \sin(t/2))$ and $\dot{\mathbf{t}} = (d\mathbf{t}/dt)/(ds/dt) = \frac{1}{4a\sin(t/2)}(\cos(t/2),$

$-\sin(t/2)) = -1/4a\sin(t/2)\mathbf{n}_s$, so the signed curvature of $\boldsymbol{\gamma}$ is $-1/4a\sin(t/2)$ and its evolute is

$$\begin{aligned}\boldsymbol{\epsilon}(t) &= a(t-\sin t, 1-\cos t) - 4a\sin(t/2)(-\cos(t/2), \sin(t/2))\\ &= a(t+\sin t, -1+\cos t).\end{aligned}$$

Reparametrizing $\boldsymbol{\epsilon}$ by $\tilde{t} = \pi + t$, we get $a(\tilde{t} - \sin\tilde{t}, 1-\cos\tilde{t}) + a(-\pi, -2)$, so $\boldsymbol{\epsilon}$ is obtained from a reparametrization of $\boldsymbol{\gamma}$ by translating by the vector $a(-\pi, -2)$.

2.2.8 The free part of the string is tangent to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(s)$ and has length $\ell - s$, hence the stated formula for $\boldsymbol{\iota}(s)$. The tangent vector of $\boldsymbol{\iota}$ is $\dot{\boldsymbol{\gamma}} - \dot{\boldsymbol{\gamma}} + (\ell - s)\ddot{\boldsymbol{\gamma}} = \kappa_s(\ell - s)\mathbf{n}_s$ (a dot denotes $d/ds$). The arc-length $v$ of $\boldsymbol{\iota}$ is given by $dv/ds = \kappa_s(\ell - s)$ so its unit tangent vector is $\mathbf{n}_s$ and its signed unit normal is $-\mathbf{t}$. Now $d\mathbf{n}_s/dv = \frac{1}{\kappa_s(\ell-s)}\dot{\mathbf{n}}_s = \frac{-1}{\ell-s}\mathbf{t}$, so the signed curvature of $\boldsymbol{\iota}$ is $1/(\ell - s)$.

2.2.9 The arc-length parametrization of the catenary is $\tilde{\boldsymbol{\gamma}}(s) = (\sinh^{-1} s, \sqrt{1+s^2})$. The involute $\boldsymbol{\iota}(s) = \tilde{\boldsymbol{\gamma}}(s) - s\dot{\tilde{\boldsymbol{\gamma}}}(s) = \left(\sinh^{-1} s - \frac{s}{\sqrt{1+s^2}}, \frac{1}{\sqrt{1+s^2}}\right)$ $= (u - \tanh u, \operatorname{sech} u)$ if $u = \sinh^{-1} s$. Thus, if $(x, y)$ is a point on the involute $\boldsymbol{\iota}$, $u = \cosh^{-1}(1/y)$ and $x = \cosh^{-1}(1/y) - \sqrt{1-y^2}$.

2.2.10 The rotation $\rho_{-\theta(s)}$ takes the tangent line of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(s)$ to $l$ and the line joining $q$ and $\boldsymbol{\gamma}(s)$ to a line parallel to that joining $\boldsymbol{\Gamma}(s)$ to $\mathbf{p}+s\mathbf{a}$. Hence, $\boldsymbol{\Gamma}(s) - (\mathbf{p}+s\mathbf{a}) = \rho_{-\theta(s)}(\mathbf{q} - \boldsymbol{\gamma}(s))$, which gives the stated equation. Now, $\dot{\boldsymbol{\Gamma}}(s) = \mathbf{a} + \left(\frac{d}{ds}\rho_{-\theta(s)}\right)(\mathbf{q} - \boldsymbol{\gamma}(s)) - \rho_{-\theta(s)}\dot{\boldsymbol{\gamma}}(s)$. The last term is clearly parallel to $\mathbf{a}$ and as they are both unit vectors they are equal. So we want to prove that $\left(\frac{d}{ds}\rho_{-\theta(s)}\right)(\mathbf{q} - \boldsymbol{\gamma}(s)) \cdot \rho_{-\theta(s)}(\mathbf{q} - \boldsymbol{\gamma}(s)) = 0$. If $A = \rho_{-\theta(s)}$, $\mathbf{v} = \mathbf{q} - \boldsymbol{\gamma}(s)$, we have to show (in matrix notation) $(A\mathbf{v})^t \frac{dA}{ds}\mathbf{v} = 0$, i.e., $\mathbf{v}^t A^t \frac{dA}{ds}\mathbf{v} = 0$. Since $A$ is orthogonal, this follows from parts (i) and (ii) of the hint. To prove (i), use components: $\mathbf{v}^t S\mathbf{v} = \sum_{i,j} v_i v_j S_{ij} = \sum_{i,j} v_j v_i S_{ji} = -\sum_{i,j} v_i v_j S_{ij}$. For (ii), differentiate $A^t A = I$.

2.3.1 (i) $\mathbf{t} = (\frac{1}{2}(1+t)^{1/2}, -\frac{1}{2}(1-t)^{1/2}, \frac{1}{\sqrt{2}})$ is a unit vector so $\boldsymbol{\gamma}$ is unit-speed; $\dot{\mathbf{t}} = (\frac{1}{4}(1+t)^{-1/2}, \frac{1}{4}(1-t)^{-1/2}, 0)$, so $\kappa = \parallel \dot{\mathbf{t}} \parallel = 1/\sqrt{8(1-t^2)}$; $\mathbf{n} = \frac{1}{\kappa}\dot{\mathbf{t}} = \frac{1}{\sqrt{2}}((1-t)^{1/2}, (1+t)^{1/2}, 0)$; $\mathbf{b} = \mathbf{t} \times \mathbf{n} = (-\frac{1}{2}(1+t)^{1/2}, \frac{1}{2}(1-t)^{1/2}, \frac{1}{\sqrt{2}})$; $\dot{\mathbf{b}} = (-\frac{1}{4}(1+t)^{-1/2}, -\frac{1}{4}(1-t)^{-1/2}, 0)$ so the torsion $\tau = 1/\sqrt{8(1-t^2)}$. The equation $\dot{\mathbf{n}} = -\kappa\mathbf{t} + \tau\mathbf{n}$ is easily checked.
(ii) $\mathbf{t} = (-\frac{4}{5}\sin t, -\cos t, \frac{3}{5}\sin t)$ is a unit vector so $\boldsymbol{\gamma}$ is unit-speed; $\dot{\mathbf{t}} = (-\frac{4}{5}\cos t, \sin t, \frac{3}{5}\cos t)$, so $\kappa = \parallel \dot{\mathbf{t}} \parallel = 1$; $\mathbf{n} = \frac{1}{\kappa}\dot{\mathbf{t}} = (-\frac{4}{5}\cos t, \sin t, \frac{3}{5}\cos t)$; $\mathbf{b} = \mathbf{t} \times \mathbf{n} = (-\frac{3}{5}, 0, -\frac{4}{5})$, so $\dot{\mathbf{b}} = \mathbf{0}$ and $\tau = 0$. By the proof of Proposition 2.3.5, $\boldsymbol{\gamma}$ is a circle of radius $1/\kappa = 1$ with centre $\boldsymbol{\gamma} + \frac{1}{\kappa}\mathbf{n} = (0, 1, 0)$ in the plane passing through $(0, 1, 0)$ perpendicular to $\mathbf{b} = (-\frac{3}{5}, 0, -\frac{4}{5})$, i.e., the plane $3x + 4z = 0$.

2.3.2 Let $a = \kappa/(\kappa^2+\tau^2)$, $b = \tau/(\kappa^2+\tau^2)$. By Examples 2.1.3 and 2.3.2, the circular helix with parameters $a$ and $b$ has curvature $a/(a^2+b^2) = \kappa$ and torsion $b/(a^2+b^2) = \tau$. By Theorem 2.3.6, every curve with curvature $\kappa$ and torsion $\tau$ is obtained by applying a direct isometry to this helix.

2.3.3 Differentiating $\mathbf{t}\cdot\mathbf{a}$ (= constant) gives $\mathbf{n}\cdot\mathbf{a} = 0$; since $\mathbf{t}, \mathbf{n}, \mathbf{b}$ are an orthonormal basis of $\mathbb{R}^3$, $\mathbf{a} = \mathbf{t}\cos\theta + \mu\mathbf{b}$ for some scalar $\mu$; since $\mathbf{a}$ is a unit vector, $\mu = \pm\sin\theta$; differentiating $\mathbf{a} = \mathbf{t}\cos\theta \pm \mathbf{b}\sin\theta$ gives $\tau = \kappa\cot\theta$. Conversely, if $\tau = \lambda\kappa$, there exists $\theta$ with $\lambda = \cot\theta$; differentiating shows that $\mathbf{a} = \mathbf{t}\cos\theta + \mathbf{b}\sin\theta$ is a constant vector and $\mathbf{t}\cdot\mathbf{a} = \cos\theta$ so $\theta$ is the angle between $\mathbf{t}$ and $\mathbf{a}$. For the circular helix in Example 2.1.3, the angle between the tangent vector $d\boldsymbol{\gamma}/d\theta = (-a\sin\theta, a\cos\theta, b)$ and the $z$-axis is the constant $\cos^{-1}(b/\sqrt{a^2+b^2})$.

2.3.4 Differentiating $(\boldsymbol{\gamma}-\mathbf{a})\cdot(\boldsymbol{\gamma}-\mathbf{a}) = r^2$ repeatedly gives $\mathbf{t}\cdot(\boldsymbol{\gamma}-\mathbf{a}) = 0$; $\mathbf{t}\cdot\mathbf{t} + \kappa\mathbf{n}\cdot(\boldsymbol{\gamma}-\mathbf{a}) = 0$, so $\mathbf{n}\cdot(\boldsymbol{\gamma}-\mathbf{a}) = -1/\kappa$; $\mathbf{n}\cdot\mathbf{t} + (-\kappa\mathbf{t}+\tau\mathbf{b})\cdot(\boldsymbol{\gamma}-\mathbf{a}) = \dot\kappa/\kappa^2$, and so $\mathbf{b}\cdot(\boldsymbol{\gamma}-\mathbf{a}) = \dot\kappa/\tau\kappa^2$; and finally $\mathbf{b}\cdot\mathbf{t} - \tau\mathbf{n}\cdot(\boldsymbol{\gamma}-\mathbf{a}) = (\dot\kappa/\tau\kappa^2)^{\cdot}$, and so $\tau/\kappa = (\dot\kappa/\tau\kappa^2)^{\cdot}$. Conversely, if Eq. 2.22 holds, then $\rho = -\sigma(\dot\rho\sigma)^{\cdot}$, so $(\rho^2 + (\dot\rho\sigma)^2)^{\cdot} = 2\rho\dot\rho + 2(\dot\rho\sigma)(\dot\rho\sigma)^{\cdot} = 0$, hence $\rho^2 + (\dot\rho\sigma)^2$ is a constant, say $r^2$ (where $r > 0$). Let $\mathbf{a} = \boldsymbol{\gamma} + \rho\mathbf{n} + \dot\rho\sigma\mathbf{b}$; then $\dot{\mathbf{a}} = \mathbf{t} + \dot\rho\mathbf{n} + \rho(-\kappa\mathbf{t}+\tau\mathbf{b}) + (\dot\rho\sigma)^{\cdot}\mathbf{b} + (\dot\rho\sigma)(-\tau\mathbf{n}) = \mathbf{0}$ using Eq. 2.22; so $\mathbf{a}$ is a constant vector and $\parallel \boldsymbol{\gamma}-\mathbf{a} \parallel^2 = \rho^2 + (\dot\rho\sigma)^2 = r^2$, hence $\boldsymbol{\gamma}$ is contained in the sphere with centre $\mathbf{a}$ and radius $r$.

2.3.5 $\dot{\boldsymbol{\Gamma}} = P\dot{\boldsymbol{\gamma}}$ so $\mathbf{T} = P\mathbf{t}$ and $\parallel \dot{\boldsymbol{\Gamma}} \parallel^2 = (P\dot{\boldsymbol{\gamma}})\cdot(P\dot{\boldsymbol{\gamma}}) = \dot{\boldsymbol{\gamma}}\cdot\dot{\boldsymbol{\gamma}}$ since $P$ is orthogonal. Then, $\ddot{\boldsymbol{\Gamma}} = P\ddot{\boldsymbol{\gamma}}$, taking lengths shows that $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$ have the same curvature $\kappa$, and then dividing by $\kappa$ gives $\mathbf{N} = P\mathbf{n}$. Then $\mathbf{B} = P\mathbf{t}\times P\mathbf{n}$. If $P$ corresponds to a direct isometry (i.e., a rotation), this is equal to $P(\mathbf{t}\times\mathbf{n}) = P\mathbf{b}$, but if $P$ corresponds to an opposite isometry, $P\mathbf{t}\times P\mathbf{n} = -P\mathbf{b}$ (Proposition A.1.6).

2.3.6 Let $\lambda_{ij} = \mathbf{v}_i\cdot\mathbf{v}_j$. The vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are orthonormal if and only if $\lambda_{ij} = \delta_{ij}$ (= 1 if $i = j$ and = 0 if $i \neq j$). So it is enough to prove that $\lambda_{ij} = \delta_{ij}$ for all values of $s$ given that it holds for $s = s_0$. Differentiating $\mathbf{v}_i\cdot\mathbf{v}_j$ gives $\dot\lambda_{ij} = \sum_{k=1}^{3}(a_{ik}\lambda_{kj} + a_{jk}\lambda_{ik})$. Now $\lambda_{ij} = \delta_{ij}$ is a solution of this system of differential equations because $a_{ij} + a_{ji} = 0$. But the theory of ordinary differential equations tells us that there is a unique solution with given values when $s = s_0$.

# Chapter 3

3.1.1 $\dot{\boldsymbol{\gamma}} = (-\sin t - a\sin 2t, \cos t + a\cos 2t)$ so $\| \dot{\boldsymbol{\gamma}} \|^2 = 1 + a^2 + 2a\cos t$. This is $\geq 1 + a^2 - 2|a| = (1-|a|)^2$ so $\boldsymbol{\gamma}$ is regular if $|a| \neq 1$. If $|a| = 1$ then $\| \dot{\boldsymbol{\gamma}} \| = 2(1 + a\cos t)$ so the origin is a singular point of $\boldsymbol{\gamma}$. If $a = 0$ then $\boldsymbol{\gamma}$ is a circle. If $0 < |a| < 1$, then $\boldsymbol{\gamma}(t_1) = \boldsymbol{\gamma}(t_2) \Longrightarrow 1 + a\cos t_1 = 1 + a\cos t_2 \Longrightarrow \cos t_1 = \cos t_2 \Longrightarrow t_2 = t_1$ or $2\pi - t_1$. In the latter case, $\boldsymbol{\gamma}(t_2) = ((1 + a\cos t_1)\cos t_1, -(1 + a\cos t_1)\sin t_1)$ so $\boldsymbol{\gamma}(t_1) = \boldsymbol{\gamma}(t_2) \Longrightarrow \sin t_1 = 0 \Longrightarrow t_1 = 0$ or $\pi$. In all cases, $t_2 - t_1$ is a multiple of $2\pi$, so $\boldsymbol{\gamma}$ is a closed curve with period $2\pi$ without self-intersections. If $|a| > 1$, $\boldsymbol{\gamma}$ passes through the origin when $\cos t = -1/a$, which has two roots with $0 \leq t < 2\pi$, say $t_1 < t_2$, so the origin is a self-intersection. The picture is qualitatively similar to that in Example 1.1.7 (which is the case $a = 2$), so the complement of the image of $\boldsymbol{\gamma}$ is the union of two bounded regions enclosed by the part of the curve with $t_1 \leq t \leq t_2$, and an unbounded region.

3.2.1 By Appendix 1, any isometry $M$ of $\mathbb{R}^2$ is of the form $M(\mathbf{v}) = P\mathbf{v} + \mathbf{b}$, where $P$ is a $3 \times 3$ orthogonal matrix and $\mathbf{b}$ is a constant vector. If $\tilde{\boldsymbol{\gamma}} = M(\boldsymbol{\gamma})$, then $\dot{\tilde{\boldsymbol{\gamma}}} = P\dot{\boldsymbol{\gamma}}$, so $\| \dot{\tilde{\boldsymbol{\gamma}}} \| = \| \dot{\boldsymbol{\gamma}} \|$, which implies that $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ have the same length. If we think of $\boldsymbol{\gamma}$ as a curve in the $xy$-plane in $\mathbb{R}^3$, Eq. 3.2 can be written $\mathcal{A}(\boldsymbol{\gamma}) = \int_0^T (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \mathbf{k}\, dt$, where $\mathbf{k} = (0, 0, 1)$. It now follows from Proposition A.1.6 that $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ have the same area (note that if $M$ is opposite, the area appears to change sign, but it does not because in that case $\tilde{\boldsymbol{\gamma}}$ is negatively oriented when $\boldsymbol{\gamma}$ is positively-oriented).

3.2.2 Parametrizing the ellipse by $\boldsymbol{\gamma}(t) = (p\cos t, q\sin t)$, with $0 \leq t \leq 2\pi$, its area is $\frac{1}{2}\int_0^{2\pi}(pq\sin^2 t + pq\cos^2 t)dt = \pi pq$. By the isoperimetric inequality, the length $\ell$ of the ellipse satisfies $\ell \geq \sqrt{4\pi \times \pi pq} = 2\pi\sqrt{pq}$, with equality if and only if the ellipse is a circle, i.e., $p = q$. But its length is $\int_0^{2\pi} \| \dot{\boldsymbol{\gamma}} \| dt = \int_0^{2\pi} \sqrt{p^2\sin^2 t + q^2\cos^2 t}\, dt$.

3.3.1 Let $(x_1, y_1)$ and $(x_2, y_2)$ be points in the interior of the ellipse, so that $\frac{x_i^2}{p^2} + \frac{y_i^2}{q^2} < 1$ for $i = 1, 2$. A point of the line segment joining the two points is $(tx_1 + (1-t)x_2, ty_1 + (1-t)y_2)$ for some $0 \leq t \leq 1$. This is in the interior of the ellipse because

$$\frac{(tx_1 + (1-t)x_2)^2}{p^2} + \frac{(ty_1 + (1-t)y_2)^2}{q^2}$$
$$= t^2\left(\frac{x_1^2}{p^2} + \frac{y_1^2}{q^2}\right) + (1-t)^2\left(\frac{x_2^2}{p^2} + \frac{y_2^2}{q^2}\right) + 2t(1-t)\left(\frac{x_1x_2}{p^2} + \frac{y_1y_2}{q^2}\right)$$

$$< t^2 + (1-t)^2 + t(1-t)\left(\frac{x_1^2}{p^2} + \frac{y_1^2}{q^2} + \frac{x_2^2}{p^2} + \frac{y_2^2}{q^2}\right)$$
$$\leq t^2 + (1-t)^2 + 2t(1-t) = 1.$$

3.3.2 $\dot{\boldsymbol{\gamma}} = (-\sin t - 2\sin 2t, \cos t + 2\cos 2t)$ and $\parallel \dot{\boldsymbol{\gamma}} \parallel = \sqrt{5 + 4\cos t}$, so the angle $\varphi$ between $\dot{\boldsymbol{\gamma}}$ and the $x$-axis is given by $\cos\varphi = \frac{-\sin t - 2\sin 2t}{\sqrt{5+4\cos t}}$, $\sin\varphi = \frac{\cos t + 2\cos 2t}{\sqrt{5+4\cos t}}$. Differentiating the second equation gives $\dot{\varphi}\cos\varphi = \frac{-\sin t(24\cos^2 t + 42\cos t + 9)}{(5+4\cos t)^{3/2}}$, so $\dot{\varphi} = \frac{\sin t(24\cos^2 t + 42\cos t + 9)}{(5+4\cos t)(\sin t + 2\sin 2t)} = \frac{9+6\cos t}{5+4\cos t}$. Hence, if $s$ is the arc-length of $\boldsymbol{\gamma}$, $\kappa_s = d\varphi/ds = (d\varphi/dt)/(ds/dt) = (9 + 6\cos t)/(5 + 4\cos t)^{3/2}$, so $\dot{\kappa}_s = 12\sin t(2 + \cos t)/(5 + 4\cos t)^{5/2}$. This vanishes at only two points of the curve, where $t = 0$ and $t = \pi$.

3.3.3 From $\boldsymbol{\epsilon}(s) = \boldsymbol{\gamma}(s) + \frac{1}{\kappa_s}\mathbf{n}_s$ we get $\dot{\boldsymbol{\epsilon}} = -\dot{\kappa}_s\mathbf{n}_s/\kappa_s^2$, so $\boldsymbol{\epsilon}$ has a singular point where $\dot{\kappa}_s = 0$, i.e., where $\boldsymbol{\gamma}$ has a vertex.

# Chapter 4

4.1.1 Let $U$ be an open disc in $\mathbb{R}^2$ and $\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y) \in U,\ z = 0\}$. If $W = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y) \in U\}$, then $W$ is an open subset of $\mathbb{R}^3$, and $\mathcal{S} \cap W$ is homeomorphic to $U$ by $(x, y, 0) \mapsto (x, y)$. So $\mathcal{S}$ is a surface.

4.1.2 The image of $\boldsymbol{\sigma}_\pm^x$ is the intersection of the sphere with the open set $\pm x > 0$ in $\mathbb{R}^3$, and its inverse is the projection $(x, y, z) \mapsto (y, z)$. Similarly for $\boldsymbol{\sigma}_\pm^y$ and $\boldsymbol{\sigma}_\pm^z$. A point of the sphere not in the image of any of the six patches would have to have $x$, $y$ and $z$ all zero, which is impossible.

4.1.3 Multiplying the two equations gives $(x^2 - z^2)\sin\theta\cos\theta = (1 - y^2)\sin\theta\cos\theta$, so $x^2 + y^2 - z^2 = 1$ unless $\cos\theta = 0$ or $\sin\theta = 0$; if $\cos\theta = 0$, then $x = -z$ and $y = 1$ and if $\sin\theta = 0$ then $x = z$ and $y = -1$, and both of these lines are also contained in the surface. The given line $L_\theta$ passes through $(\sin 2\theta, -\cos 2\theta, 0)$ and is parallel to the vector $(\cos 2\theta, \sin 2\theta, 1)$; it follows that we get all of the lines by taking $0 \leq \theta < \pi$. Let $(x, y, z)$ be a point of the surface; if $x \neq z$, let $\theta$ be such that $\cot\theta = (1 - y)/(x - z)$; then $(x, y, z)$ is on $L_\theta$; similarly if $x \neq -z$. The only remaining cases are the points $(0, 0, \pm 1)$, which lie on the lines $L_{\pi/2}$ and $L_0$. To get a surface patch covering $\mathcal{S}$, define $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ by $\boldsymbol{\sigma}(u, v) = (\sin 2\theta, -\cos 2\theta, 0) + t(\cos 2\theta, \sin 2\theta, 1)$, By the preceding paragraph, this patch covers the whole surface.

Let $M_\varphi$ be the line $(x - z)\cos\varphi = (1 + y)\sin\varphi$, $(x + z)\sin\varphi = (1 - y)$ $\cos\varphi$. By the same argument as above, $M_\varphi$ is contained in the surface and every point of the surface lies on some $M_\varphi$ with $0 \le \varphi < \pi$. If $\theta + \varphi$ is not a multiple of $\pi$, the lines $L_\theta$ and $M_\varphi$ intersect in the point $\left(\frac{\cos(\theta-\varphi)}{\sin(\theta+\varphi)}, \frac{\sin(\theta-\varphi)}{\sin(\theta+\varphi)}, \frac{\cos(\theta+\varphi)}{\sin(\theta+\varphi)}\right)$; for each $\theta$ with $0 \le \theta < \pi$, there is exactly one $\varphi$ with $0 \le \varphi < \pi$ such that $\theta + \varphi$ is a multiple of $\pi$, and the lines $L_\theta$ and $M_\varphi$ do not intersect. If $(x, y, z)$ lies on both $L_\theta$ and $L_\varphi$, with $\theta \ne \varphi$, then $(1-y)\tan\theta = (1-y)\tan\varphi$ and $(1+y)\cot\theta = (1-y)\cot\varphi$, which gives both $y = 1$ and $y = -1$ (the case in which $\theta = 0$ and $\varphi = \pi/2$, or vice versa, has to be treated separately, but the conclusion is the same). This shows that $L_\theta$ and $L_\varphi$ do not intersect; similarly, $M_\theta$ and $M_\varphi$ do not intersect.

4.1.4 For the first part, let $U = \{(u, v) \in \mathbb{R}^2 \mid 0 < u^2 + v^2 < \pi^2\}$, let $r = \sqrt{u^2 + v^2}$, and define $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ by $\boldsymbol{\sigma}(u, v) = (\frac{u}{r}, \frac{v}{r}, \tan(r - \frac{\pi}{2}))$. If $S^2$ could be covered by a single surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$, then $S^2$ would be homeomorphic to the open subset $U$ of $\mathbb{R}^2$. As $S^2$ is a closed and bounded subset of $\mathbb{R}^3$, it is compact. Hence, $U$ would be compact, and hence closed. But, since $\mathbb{R}^2$ is connected, the only non-empty subset of $\mathbb{R}^2$ that is both open and closed is $\mathbb{R}^2$ itself, and this is not compact as it is not bounded.

4.1.5 If $\{\boldsymbol{\sigma}_\alpha : U_\alpha \to \mathbb{R}^3\}$ is an atlas for a surface $\mathcal{S}$, and if $W$ is an open subset of $\mathbb{R}^3$, the restrictions $\{\boldsymbol{\sigma}|_{U_\alpha \cap \boldsymbol{\sigma}_\alpha^{-1}(W)}\}$ form an atlas of $\mathcal{S} \cap W$ (one should discard the restrictions for which $U_\alpha \cap \boldsymbol{\sigma}_\alpha^{-1}(W)$ is empty).

4.2.1 $\boldsymbol{\sigma}$ is obviously smooth and $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (-f_u, -f_v, 1)$ is nowhere zero, so $\boldsymbol{\sigma}$ is regular.

4.2.2 $\boldsymbol{\sigma}^z_\pm$ is a special case of Exercise 4.2.1, with $f = \pm\sqrt{1 - u^2 - v^2}$ ($\sqrt{1 - u^2 - v^2}$ is smooth because $1 - u^2 - v^2 > 0$ if $(u, v) \in U$); similarly for the other patches. The transition map from $\boldsymbol{\sigma}^x_+$ to $\boldsymbol{\sigma}^y_+$, for example, is $\Phi(\tilde{u}, \tilde{v}) = (u, v)$, where $\boldsymbol{\sigma}^y_+(\tilde{u}, \tilde{v}) = \boldsymbol{\sigma}^x_+(u, v)$; so $u = \sqrt{1 - \tilde{u}^2 - \tilde{v}^2}$, $v = \tilde{v}$, and this is smooth since $1 - \tilde{u}^2 - \tilde{v}^2 > 0$ if $(\tilde{u}, \tilde{v}) \in U$.

4.2.3 (i) is clearly injective and is regular because $\boldsymbol{\sigma}$ is smooth and $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (-v, -u, 1)$ is never zero. (ii) is injective but is not regular since $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (0, -3v^2, 2v)$ vanishes when $v = 0$. (iii) is not injective because $\boldsymbol{\sigma}(u, v) = \boldsymbol{\sigma}(-u-1, v)$ and is also not regular since $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = 0, 2v(1 + 2u, 1 + 2u)$ vanishes when $u = -1/2$.

4.2.4 This is similar to Example 4.1.4, but using the 'latitude-longitude' patch $\boldsymbol{\sigma}(\theta, \varphi) = (p\cos\theta\cos\varphi, q\cos\theta\sin\varphi, r\sin\theta)$.

4.2.5 A typical point on the circle $\mathcal{C}$ has coordinates $(a + b\cos\theta, 0, b\sin\theta)$; rotating this about the $z$-axis through an angle $\varphi$ gives the point $\boldsymbol{\sigma}(\theta, \varphi)$; the torus is covered by the four patches obtained by taking $(\theta, \varphi)$ to lie

in one of the following open sets:
(i) $0 < \theta < 2\pi, 0 < \varphi < 2\pi$, (ii) $0 < \theta < 2\pi, -\pi < \varphi < \pi$,
(iii) $-\pi < \theta < \pi, 0 < \varphi < 2\pi$, (iv) $-\pi < \theta < \pi, -\pi < \varphi < \pi$. Each patch is regular because

$$\boldsymbol{\sigma}_\theta \times \boldsymbol{\sigma}_\varphi = -b(a + b\cos\theta)(\cos\theta\cos\varphi, \cos\theta\sin\varphi, \sin\theta)$$

is never zero (since $a + b\cos\theta \geq a - b > 0$).

4.2.6 Suppose the centre of the propeller is initially at the origin. At time $t$, the centre is at $(0, 0, \alpha t)$ where $\alpha$ is the speed of the aeroplane. If the propeller is initially along the $x$-axis, the point initially at $(v, 0, 0)$ is therefore at the point $(v\cos\omega t, v\sin\omega t, \alpha t)$ at time $t$, where $\omega$ is the angular velocity of the propeller. Let $u = \omega t$, $\lambda = \alpha/\omega$. Next, $\boldsymbol{\sigma}_u = (-v\sin u, v\cos u, \lambda)$, $\boldsymbol{\sigma}_v = (\cos u, \sin u, 0)$, so the standard unit normal is $\mathbf{N} = (\lambda^2 + v^2)^{-1/2}(-\lambda\sin u, \lambda\cos u, -v)$. If $\theta$ is the angle between $\mathbf{N}$ and the $z$-axis, $\cos\theta = -v/(\lambda^2 + v^2)^{1/2}$ and hence $\cot\theta = \pm v/\lambda$, while the distance from the $z$-axis is $v$.

4.2.7 $\boldsymbol{\sigma}$ is the tube swept out by a circle of radius $a$ in a plane perpendicular to $\boldsymbol{\gamma}$ as its centre moves along $\boldsymbol{\gamma}$. $\boldsymbol{\sigma}_s = (1 - \kappa a\cos\theta)\mathbf{t} - \tau a\sin\theta\mathbf{n} + \tau a\cos\theta\mathbf{b}$, $\boldsymbol{\sigma}_\theta = -a\sin\theta\mathbf{n} + a\cos\theta\mathbf{b}$, giving $\boldsymbol{\sigma}_s \times \boldsymbol{\sigma}_\theta = -a(1 - \kappa a\cos\theta)(\cos\theta\mathbf{n} + \sin\theta\mathbf{b})$; this is never zero since $\kappa a < 1$ implies that $1 - \kappa a\cos\theta > 0$ for all $\theta$.

4.2.8 If $\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} + \mathbf{a}$, where $\mathbf{a}$ is a constant vector, then $\tilde{\boldsymbol{\sigma}}$ is smooth if $\boldsymbol{\sigma}$ is smooth, and $\tilde{\boldsymbol{\sigma}}_u = \boldsymbol{\sigma}_u$, $\tilde{\boldsymbol{\sigma}}_v = \boldsymbol{\sigma}_v$, so $\tilde{\boldsymbol{\sigma}}$ is regular if $\boldsymbol{\sigma}$ is regular. If $A$ is an invertible $3 \times 3$ matrix and $\tilde{\boldsymbol{\sigma}} = A\boldsymbol{\sigma}$, then $\tilde{\boldsymbol{\sigma}}$ is smooth if $\boldsymbol{\sigma}$ is smooth and $\tilde{\boldsymbol{\sigma}}_u = A\boldsymbol{\sigma}_u$, $\tilde{\boldsymbol{\sigma}}_v = A\boldsymbol{\sigma}_v$, so since $A$ is invertible $\tilde{\boldsymbol{\sigma}}_u$ and $\tilde{\boldsymbol{\sigma}}_v$ are linearly independent if $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent.

4.2.9 See Exercise 4.1.5. The restriction of a smooth map $U \to \mathbb{R}^2$, where $U$ is an open subset of $\mathbb{R}^2$, to an open subset of $U$ is smooth.

4.3.1 If $\mathcal{S}$ is covered by a single surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$, then $f : \mathcal{S} \to \mathbb{R}$ is smooth if and only if $f \circ \boldsymbol{\sigma} : U \to \mathbb{R}$ is smooth. We must check that, if $\tilde{\boldsymbol{\sigma}} : \tilde{U} \to \mathbb{R}^3$ is another patch covering $\mathcal{S}$, then $f \circ \tilde{\boldsymbol{\sigma}}$ is smooth if and only if $f \circ \boldsymbol{\sigma}$ is smooth. This is true because $f \circ \tilde{\boldsymbol{\sigma}} = (f \circ \boldsymbol{\sigma}) \circ \Phi$, where $\Phi$ is the transition map from $\boldsymbol{\sigma}$ to $\tilde{\boldsymbol{\sigma}}$, and both $\Phi$ and $\Phi^{-1}$ are smooth. The last part is true because if $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a smooth map, where $U$ is an open subset of $\mathbb{R}^2$, then each component of $\boldsymbol{\sigma}$ (which is a map $U \to \mathbb{R}$) is smooth (this is because a vector function such as $\boldsymbol{\sigma}$ is differentiated 'componentwise').

4.3.2 $f$ is not a diffeomorphism as it is not injective because: $f(0, y, z) = f(0, y, z + 2\pi)$. Take an atlas for the cone consisting of the patches $\boldsymbol{\sigma}(u, v) = (u\cos v, u\sin v, u)$, defined on the open sets $U_1 = \{(u, v) | u > 0,$

$0 < v < 2\pi\}$ and $U_2 = \{(u, v) | u > 0, -\pi < v < \pi\}$ (call these $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$), and parametrize the half-plane by $\boldsymbol{\pi}(u, v) = (0, u, v)$ with $u > 0$. If $(0, a, b)$ is any point in the plane, assume first that $b$ is not an even multiple of $\pi$, say $2n\pi < b < 2(n+1)\pi$ for some integer $n$. Then, $f(\boldsymbol{\pi}(u, v)) = \boldsymbol{\sigma}_1(u, v - 2n\pi)$ if $2n\pi < v < 2(n+1)\pi$. So $f$ is a diffeomorphism from the open subset $\{(0, y, z) | 2n\pi < z < 2(n+1)\pi\}$ of the half-plane to the cone with the half-line $y = 0, x = z > 0$ removed. Similarly if $b$ is not an odd multiple of $\pi$. This proves that $f$ is a local diffeomorphism.

4.4.1 (i) At $(1, 1, 0)$, $\boldsymbol{\sigma}_u = (1, 0, 2)$, $\boldsymbol{\sigma}_v = (0, 1, -2)$ so $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (-2, 2, 1)$ and the tangent plane is $-2x + 2y + z = 0$. (ii) At $(1, 0, 1)$, where $r = 1, \theta = 0$, $\boldsymbol{\sigma}_r = (1, 0, 2), \boldsymbol{\sigma}_\theta = (0, 1, 2)$ so $\boldsymbol{\sigma}_r \times \boldsymbol{\sigma}_\theta = (-2, -2, 1)$ and the equation of the tangent plane is $-2x - 2y + z = 0$.

4.4.2 Let $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ be a reparametrization of $\boldsymbol{\sigma}$. Then, $\boldsymbol{\sigma}_u = \frac{\partial \tilde{u}}{\partial u} \tilde{\boldsymbol{\sigma}}_{\tilde{u}} + \frac{\partial \tilde{v}}{\partial u} \tilde{\boldsymbol{\sigma}}_{\tilde{v}}$, $\boldsymbol{\sigma}_v = \frac{\partial \tilde{u}}{\partial v} \tilde{\boldsymbol{\sigma}}_{\tilde{u}} + \frac{\partial \tilde{v}}{\partial v} \tilde{\boldsymbol{\sigma}}_{\tilde{v}}$, so $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linear combinations of $\tilde{\boldsymbol{\sigma}}_{\tilde{u}}$ and $\tilde{\boldsymbol{\sigma}}_{\tilde{v}}$. Hence, any linear combination of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ is a linear combination of $\tilde{\boldsymbol{\sigma}}_{\tilde{u}}$ and $\tilde{\boldsymbol{\sigma}}_{\tilde{v}}$. The converse is also true since $\boldsymbol{\sigma}$ is a reparametrization of $\tilde{\boldsymbol{\sigma}}$.

4.4.3 If $\boldsymbol{\gamma}(t) = (x(t), y(t), z(t))$ then $\frac{d}{dt} F(\boldsymbol{\gamma}(t)) = F_x \dot{x} + F_y \dot{y} + F_z \dot{z} = \boldsymbol{\nabla} F \cdot \dot{\boldsymbol{\gamma}}$. Since $\boldsymbol{\nabla}_{\mathcal{S}} F - \boldsymbol{\nabla} F$ is perpendicular to $T_{\mathbf{p}} \mathcal{S}$, it is perpendicular to $\dot{\boldsymbol{\gamma}}(t_0)$ for every curve $\boldsymbol{\gamma}$ on $\mathcal{S}$ passing through $\mathbf{p}$ when $t = t_0$. It follows that $\boldsymbol{\nabla}_{\mathcal{S}} F \cdot \dot{\boldsymbol{\gamma}} = \boldsymbol{\nabla} F \cdot \dot{\boldsymbol{\gamma}}$ at $\mathbf{p}$. If the restriction of $F$ to $\mathcal{S}$ has a local maximum or a local minimum at $\mathbf{p}$, so does $F(\boldsymbol{\gamma}(t))$ for all curves $\boldsymbol{\gamma}$ on $\mathcal{S}$ passing through $\mathbf{p}$, hence $\frac{d}{dt} F(\boldsymbol{\gamma}(t)) = 0$ at $\mathbf{p}$, which implies that $\boldsymbol{\nabla} F$ is perpendicular to $\dot{\boldsymbol{\gamma}}$, and hence perpendicular to the tangent plane of $\mathcal{S}$ at $\mathbf{p}$. This means that $\boldsymbol{\nabla}_{\mathcal{S}} F = \mathbf{0}$.

4.4.4 $d(f \circ \boldsymbol{\gamma}))/dt = D_{\boldsymbol{\gamma}(t)} f(\dot{\boldsymbol{\gamma}}(t))$ is non-zero because $\dot{\boldsymbol{\gamma}}$ is non-zero ($\boldsymbol{\gamma}$ is regular) and $D_{\boldsymbol{\gamma}(t)} f$ is invertible (Proposition 4.4.6).

4.5.1 The transition map $\Phi(t, \theta) = (\tilde{t}, \tilde{\theta})$ is defined on the union of the rectangles given by $0 < \theta < \pi$ and $\pi < \theta < 2\pi$ (and $-1/2 < t < 1/2$). Obviously $\Phi(t, \theta) = (t, \theta)$ if $0 < \theta < \pi$. If $\pi < \theta < 2\pi$, we must have $\tilde{\theta} = \theta - 2\pi$. Since $\sin \frac{\tilde{\theta}}{2} = -\sin \frac{\theta}{2}$, $\cos \frac{\tilde{\theta}}{2} = -\cos \frac{\theta}{2}$, $\boldsymbol{\sigma}(t, \theta) = \tilde{\boldsymbol{\sigma}}(\tilde{t}, \tilde{\theta})$ forces $\tilde{t} = -t$. So $\Phi(t, \theta) = (t, \theta)$ if $0 < \theta < \pi$, and $= (-t, \theta - 2\pi)$ if $\pi < \theta < 2\pi$. The Jacobian determinant is $+1$ on the first rectangle, $-1$ on the second.

4.5.2 Let $\{\boldsymbol{\sigma}_\alpha : U_\alpha \to \mathbb{R}^3\}$ be an atlas for $\mathcal{S}$ such that the transition map $\Phi_{\alpha\beta}$ between $\boldsymbol{\sigma}_\alpha$ and $\boldsymbol{\sigma}_\beta$ satisfies $\det(J(\Phi_{\alpha\beta})) > 0$ for all $\alpha, \beta$ (Definition 4.5.1). By Proposition 4.3.1, $\{f \circ \boldsymbol{\sigma}_\alpha\}$ is an atlas for $\tilde{\mathcal{S}}$, and the transition maps for this atlas are the same as those for the atlas of $\mathcal{S}$, because $(f \circ \boldsymbol{\sigma}_\beta)^{-1} \circ (f \circ \boldsymbol{\sigma}_\alpha) = \boldsymbol{\sigma}_\beta^{-1} \circ \boldsymbol{\sigma}_\alpha$ (where this composite is defined). So the atlas $\{f \circ \boldsymbol{\sigma}_\alpha\}$ gives $\tilde{\mathcal{S}}$ the structure of an oriented surface.

# Chapter 5

5.1.1 $f_x = 2x$, $f_y = 2y$ and $f_z = 4z^3$ vanish simultaneously only when $x = y = z = 0$, but this does not satisfy $x^2 + y^2 + z^4 = 1$. So by Theorem 5.1.1 this is a smooth surface. (ii) Let $f(x, y, z)$ be the left-hand side minus the right-hand side; then, $f_x = 4x(x^2 + y^2 + z^2 - a^2 - b^2)$, $f_y = 4y(x^2 + y^2 + z^2 - a^2 - b^2)$, $f_z = 4z(x^2 + y^2 + z^2 + a^2 - b^2)$; if $f_z = 0$ then $z = 0$ since $x^2 + y^2 + z^2 + a^2 - b^2 > 0$ everywhere on the torus; if $f_x = f_y = 0$ too, then since the origin is not on the torus, we must have $x^2 + y^2 = a^2 + b^2$, but then substituting into the equation of the torus gives $(2a^2)^2 = 4a^2(a^2 + b^2)$, a contradiction. For the last part, let $\boldsymbol{\sigma}(\theta, \varphi) = (x, y, z)$ be the parametrization in Exercise 4.2.5. Then, $x^2 + y^2 + z^2 + a^2 - b^2 = 2a(a + b\cos\theta)$, so $(x^2 + y^2 + z^2 + a^2 - b^2)^2 = 4a^2(a + b\cos\theta)^2 = 4a^2(x^2 + y^2)$. Conversely, if $(x, y, z)$ satisfies the given equation, let $r = \sqrt{x^2 + y^2}$. A little algebra gives $(r^2 + z^2 - a^2 - b^2)^2 = 4a^2(b^2 - z^2)$. Hence, $|z| \leq b$ so $z = b\sin\theta$ for some $\theta \in \mathbb{R}$. Then we find $r^2 = a^2 + b^2\cos^2\theta \pm 2ab\cos\theta$, so (since $r \geq 0$) $r = a \pm b\cos\theta$. With the plus sign $(x, y, z) = \boldsymbol{\sigma}(\theta, \varphi)$ for some $\varphi \in \mathbb{R}$; with the minus sign, $(x, y, z) = \boldsymbol{\sigma}(\pi - \theta, \varphi)$ for some $\varphi$. Thus, the image of $\boldsymbol{\sigma} : \mathbb{R}^2 \to \mathbb{R}^3$ coincides with the set of solutions to the given equation.

5.1.2 See the solution of Exercise 4.4.3 for the first part. Since $\mathcal{S}$ has a (smooth) choice of unit normal $\boldsymbol{\nabla} f \parallel \boldsymbol{\nabla} f \parallel$ at each point, it is orientable. The solution of Exercise 4.4.3 also shows that if the restriction of $F$ to $\mathcal{S}$ has a local maximum or a local minimum at $\mathbf{p}$, then $\boldsymbol{\nabla} F$ is perpendicular to the tangent plane of $\mathcal{S}$ at $\mathbf{p}$. But $\boldsymbol{\nabla} f$ is also perpendicular to the tangent plane. Hence, if the restriction of $F$ to $\mathcal{S}$ has a local maximum or a local minimum at $\mathbf{p}$, then $\boldsymbol{\nabla} F$ is parallel to $\boldsymbol{\nabla} f$ at $\mathbf{p}$, i.e., $\boldsymbol{\nabla} F = \lambda \boldsymbol{\nabla} f$ for some scalar $\lambda$.

5.1.3 Let $f(x, y, z) = xyz - 1$, $F(x, y, z) = x^2 + y^2 + z^2$. Then, $f = 0$ is a smooth surface $\mathcal{S}$ by Theorem 5.1.1 and $F$ defines a smooth function on $\mathcal{S}$. To see that $F$ has a smallest value on $\mathcal{S}$, let $\mathcal{B}$ be the closed ball given by $x^2 + y^2 + z^2 \leq 3$. Then, $\mathcal{B} \cap \mathcal{S}$ is compact as it is closed and bounded and it is non-empty because it contains the point $(1, 1, 1)$. Hence, the continuous positive function $F$ must attain its lower bound, say $\ell$, on $\mathcal{B} \cap \mathcal{S}$, and $\ell \leq 3$ since $F(1, 1, 1) = 3$. Obviously $F(x, y, z) > 3$ if $(x, y, z) \notin \mathcal{B}$, so $\ell$ is the smallest value of $F$ on $\mathcal{S}$.

By Exercise 5.1.3, the local maxima or minima of $F$ on $\mathcal{S}$ occur where $(2x, 2y, 2z) = \lambda(yz, xz, xy)$ for some $\lambda$. Since $xyz = 1$ on $\mathcal{S}$ this gives $x^2 = y^2 = z^2 = \lambda/2$, so $x, y, z$ are equal up to sign. Since their product is

1, there are four possibilities: $(x, y, z) = (1, 1, 1)$, $(1, -1, -1)$, $(-1, 1, -1)$ or $(-1, -1, 1)$. The value of $F$ is 3 at each of these points, which is the smallest value of $F$ on $\mathcal{S}$ from above. The distance between any two of these points of $\mathbb{R}^3$ is the same $(2\sqrt{2})$, so they form the vertices of a regular tetrahedron.

5.2.1 (i) $(p\cos u \cos v, q\cos u \sin v, r\sin u)$ (cf. Exercise 4.2.4); (ii) see Exercise 4.1.3; (iii) $(u, v, \pm\sqrt{1 + \frac{u^2}{p^2} + \frac{v^2}{q^2}})$; (iv), (v), (vi) see Exercise 4.2.1; (vii) $(p\cos u, q\cos u, v)$; (viii) $(\pm p\cosh u, q\sinh u, v)$; (ix) $(u, u^2/p^2, v)$; (x) $(0, u, v)$; (xi) $(\pm p, u, v)$.

5.2.2 In the notation of Theorem 5.2.2, $A = \begin{pmatrix} 1 & -1/3 & 0 \\ -1/3 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$. The eigenvalues are $2/3, 4/3, -2$ and the corresponding unit eigenvectors are the columns of $P = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$. If $\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = P\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, then $z' = z$ and the quadric becomes $\frac{2}{3}x'^2 + \frac{4}{3}y'^2 - 2z'^2 + 4z' = c$, i.e., $\frac{2}{3}x'^2 + \frac{4}{3}y'^2 - 2(z'-1)^2 = c - 2$. Comparing with the standard forms in Theorem 5.2.2 gives the stated results when $c > 2$ and $c < 2$. If $c = 2$ we have a cone with axis the $z$-axis (which is the same as the $z'$-axis), vertex at $x' = y' = 0, z' = 1$, i.e., $x = y = 0, z = 1$, and cross-section perpendicular to the $z$-axis an ellipse $\frac{2}{3}x'^2 + \frac{4}{3}y'^2 =$ constant.

5.2.3 Substituting the components $(x, y, z)$ of $\boldsymbol{\gamma}(t) = \mathbf{a} + t\mathbf{b}$ into the equation of the quadric gives a quadratic equation for $t$; if the quadric contains three points on the line, this quadratic equation has three roots, hence is identically zero, so the quadric contains the whole line.

For the second part, take three points on each of the given lines; substituting the coordinates of these nine points into the equation of the quadric gives a system of nine homogeneous linear equations for the ten coefficients $a_1, \ldots, c$ of the quadric; such a system always has a non-trivial solution. By the first part, the resulting quadric contains all three lines.

5.2.4 Let $L_1, L_2, L_3$ be three lines from the first family; by the preceding exercise, there is a quadric $\mathcal{Q}$ containing all three lines; all but finitely many lines of the second family intersect each of the three lines; if $L'$ is such a line, $\mathcal{Q}$ contains three points of $L'$, and hence the whole of $L'$ by the preceding exercise; so $\mathcal{Q}$ contains all but finitely many lines of the second family; since any quadric is a closed subset of $\mathbb{R}^3$, $\mathcal{Q}$ must contain all the lines of the second family, and hence must contain $\mathcal{S}$.

5.3.1 From Example 5.3.2, the surface can be parametrized by $\boldsymbol{\sigma}(u, v) = (\cosh u \cos v, \cosh u \sin v, u)$, with $u \in \mathbb{R}$ and $-\pi < v < \pi$ or $0 < v < 2\pi$.

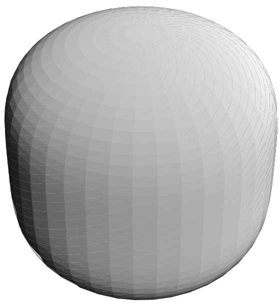
5.3.2 $\| \boldsymbol{\sigma}(u, v) \|^2 = \operatorname{sech}^2 u(\cos^2 v + \sin^2 v) + \tanh^2 u = \operatorname{sech}^2 u + \tanh^2 u = 1$, so $\boldsymbol{\sigma}$ parametrizes an open subset of $S^2$; $\boldsymbol{\sigma}$ is clearly smooth; and $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = -\operatorname{sech}^2 u\, \boldsymbol{\sigma}(u, v)$ is never zero, so $\boldsymbol{\sigma}$ is regular. Meridians correspond to the parameter curves $v = \text{constant}$, and parallels to the curves $u = \text{constant}$.

5.3.3 (i) $\tilde{\boldsymbol{\gamma}} \cdot \mathbf{a} = 0$ so $\tilde{\boldsymbol{\gamma}}$ is contained in the plane perpendicular to $\mathbf{a}$ and passing through the origin; (ii) simple algebra; (iii) $\tilde{v}$ is clearly a smooth function of $(u, v)$ and the Jacobian matrix of the map $(u, v) \mapsto (u, \tilde{v})$ is $\begin{pmatrix} 1 & 0 \\ \dot{\boldsymbol{\gamma}} \cdot \mathbf{a} & 1 \end{pmatrix}$, where a dot denotes $d/du$; this matrix is invertible so $\tilde{\boldsymbol{\gamma}}$ is a reparametrization of $\boldsymbol{\gamma}$.

5.3.4 $\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}, \boldsymbol{\sigma}_v = \boldsymbol{\delta}$ (a dot denotes $d/du$) so $\dot{\boldsymbol{\delta}}(u)$ is perpendicular to the surface at $\boldsymbol{\sigma}(u, v) \iff \dot{\boldsymbol{\delta}} \cdot (\dot{\boldsymbol{\gamma}} + v\dot{\boldsymbol{\delta}}) = 0, \dot{\boldsymbol{\delta}} \cdot \boldsymbol{\delta} = 0$. The second equation follows from $\| \boldsymbol{\delta} \| = 1$ so the two conditions are satisfied $\iff v = -(\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\delta}})/ \| \dot{\boldsymbol{\delta}} \|^2$. Hence, $\boldsymbol{\Gamma}(u) = \boldsymbol{\gamma} - (\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\delta}})\boldsymbol{\delta}/ \| \dot{\boldsymbol{\delta}} \|^2$. Using $\dot{\boldsymbol{\delta}} \cdot \boldsymbol{\delta} = 0$ again, $\dot{\boldsymbol{\Gamma}} \cdot \dot{\boldsymbol{\delta}} = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\delta}} - (\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\delta}})\dot{\boldsymbol{\delta}} \cdot \dot{\boldsymbol{\delta}}/ \| \dot{\boldsymbol{\delta}} \|^2 = 0$.

5.4.1 Both surfaces are closed subsets of $\mathbb{R}^3$, as they are of the form $f(x, y, z) = 0$, where $f : \mathbb{R}^3 \to \mathbb{R}$ is a continuous function (equal to $x^2 - y^2 + z^4 - 1$ and $x^2 + y^2 + z^4 - 1$ in the two cases). The surface in (i) is not bounded, and hence not compact, since it contains the point $(1, a^2, a)$ for all real numbers $a$; that in (ii) is bounded, and hence compact, since $x^2 + y^2 + z^4 = 1 \implies -1 \leq x, y, z \leq 1$.

The surface in (ii) is obtained by rotating the curve $x^2 + z^4 = 1$ in the $xz$-plane around the $z$-axis:



5.4.2 A closed curve $\gamma$ with period $T$ can be identified with the unit circle by $\gamma(t) \mapsto (\cos(2\pi t/T), \sin(2\pi t/T))$. This gives rise to a diffeomorphism from the tube around $\gamma$ to a tube around the circle, i.e., a torus. We have to make the tube have a sufficiently small radius to avoid self-intersections.

5.5.1 Both parts are geometrically obvious.

5.5.2 Let $(a, b, c) \in \mathbb{R}^3$ with $a$ and $b$ non-zero. Then $F_t(a, b, c) \to \infty$ as $t \to \infty$ and as $t$ approaches $p^2$ and $q^2$ from the left; and $F_t(a, b, c) \to -\infty$ as $t \to -\infty$ and as $t$ approaches $p^2$ and $q^2$ from the right. From this and the fact that $F_t(a, b, c) = 0$ is equivalent to a cubic equation for $t$, it follows that there exist unique numbers $u, v, w$ with $u < p^2$, $p^2 < v < q^2$ and $q^2 < w$ such that $F_t(a, b, c) = 0$ when $t = u, v$ or $w$. The surfaces $F_u(x, y, z) = 0$ and $F_w(x, y, z) = 0$ are elliptic paraboloids and $F_v(x, y, z) = 0$ is a hyperbolic paraboloid, and we have shown that there is one surface of each type passing through each point $(a, b, c)$. To parametrize these surfaces, write $F_t(x, y, z) = 0$ as the cubic equation $x^2(q^2-t)+y^2(p^2-t)-2z(p^2-t)(q^2-t)+t(p^2-t)(q^2-t) = 0$, and note that the left-hand side must be equal to $(t-u)(t-v)(t-w)$; putting $t = p^2, q^2$ and then equating coefficients of $t^2$ (say) gives $x = \pm\sqrt{\frac{(p^2-u)(p^2-v)(p^2-w)}{q^2-p^2}}$, $y = \pm\sqrt{\frac{(q^2-u)(q^2-v)(q^2-w)}{p^2-q^2}}$, $z = \frac{1}{2}(u + v + w - p^2 - q^2)$.

5.6.1 Let $F : W \to V$ be the smooth bijective map constructed in the proof of Proposition 4.2.6. Then, $(u(t), v(t)) = F^{-1}(\gamma(t))$ is smooth.

5.6.2 Suppose, for example, that $f_y \neq 0$ at $(x_0, y_0)$. Let $F(x, y) = (x, f(x, y))$; then $F$ is smooth and its Jacobian matrix $\begin{pmatrix} 1 & f_x \\ 0 & f_y \end{pmatrix}$ is invertible at $(x_0, y_0)$. By the inverse function theorem, $F$ has a smooth inverse $G$ defined on an open subset of $\mathbb{R}^2$ containing $F(x_0, y_0) = (x_0, 0)$, and $G$ must be of the form $G(x, z) = (x, g(x, z))$ for some smooth function $g$. Then $\gamma(t) = (t, g(t, 0))$ is a parametrization of the level curve $f(x, y) = 0$ containing $(x_0, y_0)$.

The matrix $\begin{pmatrix} f_x & f_y & f_z \\ g_x & g_y & g_z \end{pmatrix}$ has rank 2 everywhere; suppose that, at some point $(x_0, y_0, z_0)$ on the level curve, the $2 \times 2$ submatrix $\begin{pmatrix} f_y & f_z \\ g_y & g_z \end{pmatrix}$ is invertible. Then, the function $F(x, y, z) = (x, f(x, y, z), g(x, y, z))$

is smooth and its Jacobian matrix $\begin{pmatrix} 1 & 0 & 0 \\ f_x & f_y & f_z \\ g_x & g_y & g_z \end{pmatrix}$ is invertible at $(x_0, y_0, z_0)$. Let $G(x, u, v) = (x, \varphi(x, u, v), \psi(x, u, v))$ be the smooth inverse of $F$ defined near $(x_0, 0, 0)$. Then $\boldsymbol{\gamma}(t) = (t, \varphi(t, 0, 0), \psi(t, 0, 0))$ is a parametrization of the level curve $f(x, y, z) = g(x, y, z) = 0$ containing $(x_0, y_0, z_0)$.

5.6.3 Let $\boldsymbol{\sigma}(u, v) = (f(u, v), g(u, v), h(u, v))$. The condition $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \neq \mathbf{0}$ at $(u_0, v_0)$ means that the matrix $\begin{pmatrix} f_u & f_v \\ g_u & g_v \\ h_u & h_v \end{pmatrix}$ has rank 2 at $(u_0, v_0)$, so at least one $2 \times 2$ submatrix is invertible, say $\begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}$. If $F(u, v) = (f(u, v), g(u, v))$, then as in the proof of Proposition 4.2.6 there is an open subset $V$ of $\mathbb{R}^2$ containing $F(u_0, v_0)$ and an open subset $W$ of $U$ containing $(u_0, v_0)$ such that $F : W \to V$ is bijective, in particular injective. Then the restriction of $\boldsymbol{\sigma}$ to $W$ is injective.

5.6.4 Let $\boldsymbol{\sigma}(u, v) = (f(u, v), g(u, v), h(u, v))$. The condition that $\mathbf{N}(u_0, v_0)$ is not parallel to the $xy$-plane means that the matrix $\begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}$ is invertible at $(u_0, v_0)$. If $F(u, v) = (f(u, v), g(u, v))$, then as in the proof of Proposition 4.2.6 there is an open subset $V$ of $\mathbb{R}^2$ containing $F(u_0, v_0)$ and an open subset $W$ of $U$ containing $(u_0, v_0)$ such that $F : W \to V$ is bijective with smooth inverse. If $F^{-1}(u, v) = (\alpha(u, v), \beta(u, v))$, then near $(x_0, y_0, z_0)$ the surface coincides with the graph $z = h(\alpha(x, y), \beta(x, y))$. If $\mathbf{N}(u_0, v_0)$ is parallel to the $xy$-plane, then at least one of the other two $2 \times 2$ submatrices of the Jacobian matrix of $\boldsymbol{\sigma}(u, v)$ is invertible, and then the surface coincides near $(x_0, y_0, z_0)$ with a graph of the form $x = \varphi(y, z)$ or $y = \varphi(x, z)$.

# Chapter 6

6.1.1 (i) Quadric cone $x^2 + z^2 = y^2$; $\boldsymbol{\sigma}_u = (\cosh u \sinh v, \cosh u \cosh v, \cosh u)$, $\boldsymbol{\sigma}_v = (\sinh u \cosh v, \sinh u \sinh v, 0)$, $\| \boldsymbol{\sigma}_u \|^2 = 2\cosh^2 u \cosh^2 v$, $\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 2 \sinh u \cosh u \sinh v \cosh v$, $\| \boldsymbol{\sigma}_v \|^2 = \sinh^2 u \cosh 2v$, and the first fundamental form is $2\cosh^2 u \cosh^2 v \, du^2 + \sinh 2u \sinh 2v \, dudv + \sinh^2 u \cosh 2v \, dv^2$.

(ii) Paraboloid of revolution; $(2 + 4u^2)\, du^2 + 8uv \, dudv + (2 + 4v^2)\, dv^2$.

(iii) Hyperbolic cylinder; $(\cosh^2 u + \sinh^2 u)\, du^2 + dv^2$.
(iv) Paraboloid of revolution; $(1+4u^2)\, du^2 + 8uv\, dudv + (1+4v^2)\, dv^2$.

6.1.2 Applying a translation to a surface patch $\boldsymbol{\sigma}$ does not change $\boldsymbol{\sigma}_u$ or $\boldsymbol{\sigma}_v$. If $P$ is a $3 \times 3$ orthogonal matrix, $(P\boldsymbol{\sigma})_u = P(\boldsymbol{\sigma}_u), (P\boldsymbol{\sigma})_v = P(\boldsymbol{\sigma}_v)$, and $P$ preserves dot products ($P(\mathbf{p}) \cdot P(\mathbf{q}) = \mathbf{p} \cdot \mathbf{q}$ for all vectors $\mathbf{p}, \mathbf{q} \in \mathbb{R}^3$). Applying the dilation $(x, y, z) \mapsto a(x, y, z)$, where $a$ is a non-zero constant, multiplies $\boldsymbol{\sigma}$ by $a$ and hence the first fundamental form by $a^2$.

6.1.3 Since both sides define bilinear forms on the tangent plane, it suffices to prove that the two sides agree when $\mathbf{v}, \mathbf{w}$ belong to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$. This is easily checked using $du(\boldsymbol{\sigma}_u) = dv(\boldsymbol{\sigma}_v) = 1$, $du(\boldsymbol{\sigma}_v) = dv(\boldsymbol{\sigma}_u) = 0$.

6.1.4 By the chain rule, $\tilde{\boldsymbol{\sigma}}_{\tilde{u}} = \boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{u}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{u}}$, $\tilde{\boldsymbol{\sigma}}_{\tilde{v}} = \boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{v}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{v}}$, which gives $\tilde{E} = \tilde{\boldsymbol{\sigma}}_{\tilde{u}}.\tilde{\boldsymbol{\sigma}}_{\tilde{u}} = E\left(\frac{\partial u}{\partial \tilde{u}}\right)^2 + 2F \frac{\partial u}{\partial \tilde{u}}\frac{\partial v}{\partial \tilde{u}} + G\left(\frac{\partial v}{\partial \tilde{u}}\right)^2$. Similar expressions for $\tilde{F}$ and $\tilde{G}$ can be found; multiplying out the matrices shows that these formulas are equivalent to the matrix equation in the question. Following the procedure given, $Edu^2 + 2Fdudv + Gdv^2 = E\left(\frac{\partial u}{\partial \tilde{u}}d\tilde{u} + \frac{\partial u}{\partial \tilde{v}}d\tilde{v}\right)^2 + 2F\left(\frac{\partial u}{\partial \tilde{u}}d\tilde{u} + \frac{\partial u}{\partial \tilde{v}}d\tilde{v}\right)\left(\frac{\partial v}{\partial \tilde{u}}d\tilde{u} + \frac{\partial v}{\partial \tilde{v}}d\tilde{v}\right) + G\left(\frac{\partial v}{\partial \tilde{u}}d\tilde{u} + \frac{\partial v}{\partial \tilde{v}}d\tilde{v}\right)^2$. The coefficient of $d\tilde{u}^2$ is $E\left(\frac{\partial u}{\partial \tilde{u}}\right)^2 + 2F\frac{\partial u}{\partial \tilde{u}}\frac{\partial v}{\partial \tilde{u}} + G\left(\frac{\partial v}{\partial \tilde{u}}\right)^2$, which agrees with the expression for $\tilde{E}$ found above. Similarly for $\tilde{F}$ and $\tilde{G}$.

6.1.5 (i) $\Longleftrightarrow$ (ii): $E_v = G_u = 0 \Longleftrightarrow \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uv} = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uv} = 0 \Longleftrightarrow \boldsymbol{\sigma}_{uv}$ is parallel to $\mathbf{N}$. Consider the quadrilateral bounded by the parameter curves $u = u_0, u = u_1, v = v_0, v = v_1$. The length of the side given by $u = u_0$ is $\int_{v_0}^{v_1} \parallel (\boldsymbol{\sigma}_v(u_0, v) \parallel dv = \int_{v_0}^{v_1} \sqrt{G(u_0, v)}dv$. (i) $\Longrightarrow$ (iii): If $G_u = 0$, $G$ depends only on $v$ so this integral is unchanged when $u_0$ is replaced by $u_1$. So the two sides $u = u_0$ and $u = u_1$ have the same length; and similarly for the other two sides. (iii) $\Longrightarrow$ (i): If the lengths are equal then $\int_{v_0}^{v_1} \sqrt{G(u, v)}dv$ is independent of $u$; differentiating with respect to $u$ gives $\int_{v_0}^{v_1} \frac{G_u}{2\sqrt{G}}dv = 0$ for all $v_0, v_1$, so $G_u = 0$; and similarly $E_v = 0$.

Assuming conditions (i)–(iii) are satisfied, define $\tilde{u} = \int \sqrt{E(u)}\, du$, $\tilde{v} = \int \sqrt{G(v)}\, dv$. Then, $(u, v) \mapsto (\tilde{u}, \tilde{v})$ is a reparametrization map because its Jacobian matrix $\begin{pmatrix} \sqrt{E} & 0 \\ 0 & \sqrt{G} \end{pmatrix}$ has non-zero determinant $\sqrt{EG}$. The first fundamental form of the reparametrization $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ of $\boldsymbol{\sigma}(u, v)$ is $d\tilde{u}^2 + \frac{2F}{\sqrt{EG}}d\tilde{u}d\tilde{v} + d\tilde{v}^2$. Since $EG - F^2 > 0$, we have $-1 < \frac{2F}{\sqrt{EG}} < 1$ so there is a smooth function $\theta(\tilde{u}, \tilde{v})$ with $0 < \theta < \pi$ such that $\cos\theta = \frac{2F}{\sqrt{EG}}$. This gives the first fundamental form as $d\tilde{u}^2 + 2\cos\theta d\tilde{u}d\tilde{v} + d\tilde{v}^2$.

Since $\tilde{u} = \frac{1}{2}(\hat{u}+\hat{v})$, $\tilde{v} = \frac{1}{2}(\hat{u}-\hat{v})$, the first fundamental form becomes

$$\frac{1}{4}(d\hat{u}+d\hat{v})^2 + \frac{1}{2}\cos\theta(d\hat{u}^2 - d\hat{v}^2) + \frac{1}{4}(d\hat{u}-d\hat{v})^2$$
$$= \frac{1}{2}(1+\cos\theta)d\hat{u}^2 + \frac{1}{2}(1-\cos\theta)d\hat{v}^2 = \cos^2\frac{\theta}{2}d\hat{u}^2 + \sin^2\frac{\theta}{2}d\hat{v}^2.$$

6.2.1 The map is $\boldsymbol{\sigma}(u,v) \mapsto \left(u\sqrt{2}\cos\frac{v}{\sqrt{2}}, u\sqrt{2}\sin\frac{v}{\sqrt{2}}, 0\right) = \tilde{\boldsymbol{\sigma}}(u,v)$, say. The image of this map is the sector of the $xy$-plane whose polar coordinates $(r,\theta)$ satisfy $0 < \theta < \pi\sqrt{2}$. The first fundamental form of $\boldsymbol{\sigma}$ is $2\,du^2 + u^2\,dv^2$; $\tilde{\boldsymbol{\sigma}}_u = \left(\sqrt{2}\cos\frac{v}{\sqrt{2}}, \sqrt{2}\sin\frac{v}{\sqrt{2}}, 0\right)$,$\tilde{\boldsymbol{\sigma}}_v = \left(-u\sin\frac{v}{\sqrt{2}}, u\cos\frac{v}{\sqrt{2}}, 0\right)$, so $\parallel \tilde{\boldsymbol{\sigma}}_u \parallel^2 = 2$, $\tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v = 0$, $\parallel \tilde{\boldsymbol{\sigma}}_v \parallel^2 = u^2$ and the first fundamental form of $\tilde{\boldsymbol{\sigma}}$ is also $2\,du^2 + u^2\,dv^2$.

6.2.2 No: the part of the ruling $(t,0,t)$ with $1 \leq t \leq 2$ (say) has length $\sqrt{2}$ and is mapped to the straight line segment $(t,0,0)$ with $1 \leq t \leq 2$, which has length 1.

6.2.3 A straightforward calculation shows that the first fundamental form of $\boldsymbol{\sigma}^t$ is $\cosh^2 u(du^2+dv^2)$; in particular, it is independent of $t$. Hence, $\boldsymbol{\sigma}(u,v) \mapsto \boldsymbol{\sigma}^t(u,v)$ is an isometry for all $t$. Taking $t = \pi/2$ gives the isometry from the catenoid to the helicoid; under this map, the parallels $u =$ constant on the catenoid go to circular helices on the helicoid, and the meridians $v =$ constant go to the rulings of the helicoid.

6.2.4 The line of striction is given by $v = -(\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\delta}})/\parallel \dot{\boldsymbol{\delta}} \parallel^2$ (Exercise 5.3.4), where in this case $\boldsymbol{\delta} = \dot{\boldsymbol{\gamma}}$. Since $\boldsymbol{\gamma}$ is unit-speed, $\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 0$ so $v = 0$ and we get the curve $\boldsymbol{\gamma}$ itself. For the second part, we can assume that $u_0 = 0$ and by applying an isometry of $\mathbb{R}^3$ that $\boldsymbol{\gamma}(0) = \mathbf{0}, \mathbf{t}(0) = \mathbf{i}, \mathbf{n}(0) = \mathbf{j}, \mathbf{b}(0) = \mathbf{k}$ (in the usual notation). Then, using Frenet-Serret, $\ddot{\boldsymbol{\gamma}}(0) = \kappa(0)\mathbf{j}, \dddot{\boldsymbol{\gamma}}(0) = (-\kappa(0)^2, \dot{\kappa}(0), \kappa(0)\tau(0))$ so, neglecting higher powers of $u$ in each component, $\boldsymbol{\gamma}(u) = \boldsymbol{\gamma}(0) + u \cdot \boldsymbol{\gamma}(0) + \frac{1}{2}\ddot{\boldsymbol{\gamma}}(0)u^2 + \frac{1}{6}\dddot{\boldsymbol{\gamma}}(0)u^3 + \cdots = (u, \frac{1}{2}\kappa(0)u^2, \frac{1}{6}\kappa(0)\tau(0)u^3)$. The intersection of the surface with the plane perpendicular to $\mathbf{t}(0) = \mathbf{i}$ is given by setting the $x$-component of $\boldsymbol{\sigma}(u,v)$ equal to zero. This gives $v = -u+$ higher terms, so neglecting such terms, $u = -v$. Then the intersection is $\boldsymbol{\Gamma}(v) = \boldsymbol{\sigma}(-v,v) = \boldsymbol{\gamma}(-v) + v\dot{\boldsymbol{\gamma}}(-v) = (0, -\frac{1}{2}\kappa(0)v^2, \frac{1}{3}\kappa(0)\tau(0)v^3)$.

6.3.1 If the first fundamental forms of two surfaces are equal, they are certainly proportional, so any isometry is a conformal map. Stereographic projection is a conformal map from $S^2$ to the plane, but it is not an isometry since $\lambda \neq 1$ (see Example 6.3.5).

6.3.2 The first fundamental form of the given surface patch is $(1+u^2+v^2)^2$ $(du^2+dv^2)$; this is a multiple of $du^2+dv^2$ so the patch is conformal.

6.3.3 The first fundamental form of $\tilde{\boldsymbol{\sigma}}(u,v)$ is $\left(\frac{d\psi}{du}\right)^2 du^2 + \cos^2\psi(u)dv^2$. So $\tilde{\boldsymbol{\sigma}}$ is conformal $\iff d\psi/du = \pm\cos\psi$. Taking the plus sign, $u = \int \sec\psi d\psi = \ln(\sec\psi + \tan\psi)$, so $\frac{1+\sin\psi}{\cos\psi} = e^u$. Then $2\cosh u = e^u + e^{-u} = \frac{1+\sin\psi}{\cos\psi} + \frac{\cos\psi}{1+\sin\psi} = 2\sec\psi$. Hence, $\cos\psi = \mathrm{sech}u, \sin\psi = \tanh u$ and $\tilde{\boldsymbol{\sigma}}(u,v)$ is the patch in Exercise 5.3.2.

6.3.4 $\Phi$ is conformal if and only if $f_u^2 + g_u^2 = f_v^2 + g_v^2$ and $f_u f_v + g_u g_v = 0$. Let $z = f_u + ig_u$, $w = f_v + ig_v$; then $\Phi$ is conformal if and only if $z\bar{z} = w\bar{w}$ and $z\bar{w} + \bar{z}w = 0$, where the bar denotes complex conjugate; if $z = 0$, then $w = 0$ and all four equations are certainly satisfied; if $z \neq 0$, the equations give $z^2 = -w^2$, so $z = \pm iw$; these are easily seen to be equivalent to the first pair of equations in the statement of the exercise if the sign is $+$, and to the second pair if the sign is $-$. We have $\det(J(\Phi)) = \begin{vmatrix} f_u & g_u \\ f_v & g_v \end{vmatrix} = \pm(f_u^2 + f_v^2)$, with a plus sign if the first pair of equations hold and a minus sign if the second pair of equations hold.

6.3.5 Let $\mathcal{S}$ be an orientable surface. Fix a smooth choice of unit normal at each point of $\mathcal{S}$, and let $\mathcal{A}$ be the atlas for $\mathcal{S}$ consisting of all the surface patches for $\mathcal{S}$ whose standard unit normal agrees with the chosen normal. On the other hand, by Theorem 6.3.6 $\mathcal{S}$ has an atlas consisting of conformal parametrizations; let $\tilde{\mathcal{A}}$ be the maximal such atlas (i.e., the set of all conformal parametrizations of $\mathcal{S}$). Then, $\mathcal{A} \cap \tilde{\mathcal{A}}$ is an atlas for $\mathcal{S}$. Indeed, if $\mathbf{p} \in \mathcal{S}$, let $\boldsymbol{\sigma}$ be any conformal parametrization of $\mathcal{S}$ containing $\mathbf{p}$. If $\boldsymbol{\sigma}$ has the wrong orientation (so that $\boldsymbol{\sigma} \notin \tilde{\mathcal{A}}$), then $\tilde{\boldsymbol{\sigma}}(u,v) = \boldsymbol{\sigma}(-u,v)$ is a conformal parametrization containing $\mathbf{p}$ that has the correct orientation. Thus, in any case there is a surface patch of $\mathcal{S}$ containing $\mathbf{p}$ that is both conformal and correctly oriented. Let $\Phi$ be the transition map between two of the patches in the atlas $\mathcal{A} \cap \tilde{\mathcal{A}}$. Then, $\Phi$ is a conformal diffeomorphism between open subsets of $\mathbb{R}^2$. By Exercise 6.3.4, $\Phi$ is either holomorphic or anti-holomorphic, and in the latter case $\det(J(\Phi)) < 0$, contradicting the fact that $\Phi$ is the transition map between two correctly oriented surface patches. Hence, $\Phi$ must be holomorphic.

6.3.6 Following Example 6.3.5, we find $\tilde{\Pi}(x,y,z) = (\frac{x}{z+1}, \frac{y}{z+1}, 0)$. Identifying $(u,v) \in \mathbb{R}^2$ with $w = u + iv \in \mathbb{C}$, we find $\tilde{\boldsymbol{\sigma}}_1(w) = (\frac{2w}{|w|^2+1}, \frac{1-|w|^2}{1+|w|^2})$. Then $\boldsymbol{\sigma}_1(w) = \tilde{\boldsymbol{\sigma}}_1(1/\bar{w})$, so the transition map is $w \mapsto 1/\bar{w}$. This is not holomorphic, so the atlas $\{\boldsymbol{\sigma}_1, \tilde{\boldsymbol{\sigma}}_1\}$ does not give $S^2$ the structure of a Riemann surface. If $\hat{\boldsymbol{\sigma}}_1(w) = \tilde{\boldsymbol{\sigma}}_1(\bar{w})$, the transition map between $\boldsymbol{\sigma}_1$ and $\hat{\boldsymbol{\sigma}}_1$ is $w \mapsto 1/w$. This is holomorphic (when $w \neq 0$, which holds on the overlap of the two patches), so the atlas $\{\boldsymbol{\sigma}_1, \hat{\boldsymbol{\sigma}}_1\}$ gives $S^2$ the structure of a Riemann surface.

6.3.7 Any circle on $S^2$ is the intersection of $S^2$ with a plane, and so (see Appendix 2) has equation of the form $aw + \bar{a}\bar{w} + bz = c$, where $a \in \mathbb{C}$, $b, c \in \mathbb{R}$ are constants (and $a$ and $b$ are not both zero). Substituting $w = \frac{2\xi}{|\xi|^2+1}, z = \frac{|\xi|^2-1}{|\xi|^2+1}$ gives $(b-c)|\xi|^2 + 2a\xi + 2\bar{a}\bar{\xi} = b + c$, which is the equation of a Circle in $\mathbb{C}_\infty$ (a line if $b = c$, a circle otherwise). The converse is proved similarly.

6.3.8 The expression of the map $\Pi^{-1} \circ M \circ \Pi$ in terms of the atlas $\{\boldsymbol{\sigma}_1, \tilde{\boldsymbol{\sigma}}_1\}$ of $S^2$ in Exercise 6.3.6, which consists of conformal patches, is of the form $w \mapsto M(w), w \mapsto M(1/\bar{w}), w \mapsto \overline{M(w)}^{-1}$, or $w \mapsto \overline{M(1/\bar{w})}^{-1}$, i.e., a Möbius or conjugate-Möbius transformation. Since such transformations are conformal (Appendix 2), the result follows.

6.4.1 Parametrize the paraboloid by $\boldsymbol{\sigma}(u, v) = (u, v, u^2+v^2)$; its first fundamental form is $(1 + 4u^2)du^2 + 8uv\,dudv + (1 + 4v^2)dv^2$. Hence, the required area is $\int \sqrt{1 + 4(u^2 + v^2)}\,dudv$, taken over the disc $u^2 + v^2 < 1$. Let $u = r \sin\theta, v = r\cos\theta$; then the area is $2\pi \int_0^1 \sqrt{1 + 4r^2}r\,dr = \frac{\pi}{6}(5^{3/2} - 1)$. This is less than the area $2\pi$ of the hemisphere.

6.4.2 If $\mathcal{S}$ is a sphere with centre the origin and radius $R$, the map $S^2 \to \mathcal{S}$ given by $\mathbf{p} \mapsto R\mathbf{p}$ multiplies the first fundamental form by $R^2$, and so is conformal but multiplies areas by $R^2$. It follows from Theorem 6.4.7 that the sum of the angles of a spherical triangle of area $A$ on $\mathcal{S}$ is $\pi + A/R^2$. In this case, $R$ is the radius of the earth and $A$ is $\geq$ the area of Australia, so the sum of the angles is $\geq \pi + (7{,}500{,}000)/(6{,}500)^2 = \pi + \frac{30}{169}$ radians. Hence, at least one angle of the triangle must be at least one third of this, i.e., $\pi + \frac{10}{169}$ radians.

6.4.3 Take a point $\mathbf{p}$ inside the polygon and join it to each vertex of the polygon by an arc of a great circle. This gives $n$ triangles whose sides are arcs of great circles. The sum of their angles is the sum of their areas (i.e., the area of the polygon) minus $n\pi$ by Theorem 6.4.7, and is also the sum of the angles of the polygon plus $2\pi$ (the angle around $\mathbf{p}$).

6.4.4 The sum of the angles around any vertex is $2\pi$, so the sum of the angles of all the polygons is $2\pi V$. By the preceding exercise, the sum of the angles of a polygon with $n$ sides is $(n-2)\pi$ plus its area. Summing over all polygons gives $2\pi V = 4\pi + \sum_{\text{polygons}}(n-2)\pi$, since the sum of the areas of all the polygons is the area $4\pi$ of the sphere. Since two polygons meet along each edge, $\sum_{\text{polygons}} n = 2E$, and since there are $F$ polygons altogether, we get $2\pi V = 2\pi E - 2\pi F + 4\pi$, which is equivalent to $V - E + F = 2$.

6.4.5 (i) is obvious as a local isometry preserves $E, F, G$ and hence $\sqrt{EG - F^2}$. If $\tilde{E} = \lambda E$, $\tilde{F} = \lambda F$ and $\tilde{G} = \lambda G$, and if $\tilde{E}\tilde{G} - \tilde{F}^2 = EG - F^2$, then

$\lambda^2 = 1$ and so $\lambda = 1$ (as $E, \tilde{E}$ are $> 0$). This proves (ii). The map from $S^2$ to the unit cylinder in the proof of Theorem 6.4.6 is an equiareal map that is not a local isometry.

6.4.6 Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$; $f$ is equiareal $\iff \int_R (E_1G_1 - F_1^2)^{1/2}\,dudv = \int_R (E_2G_2 - F_2^2)^{1/2}\,dudv$ for all regions $R \subseteq U$. This holds $\iff$ the two integrands are equal everywhere, i.e., $\iff E_1G_1 - F_1^2 = E_2G_2 - F_2^2$.

6.4.7 Since $\mathbf{N}$ is perpendicular to the tangent plane, $\mathbf{N} \times \boldsymbol{\sigma}_u$ is parallel to the tangent plane, and so $= \alpha\boldsymbol{\sigma}_u + \beta\boldsymbol{\sigma}_v$ for some $\alpha, \beta$. Now $(\mathbf{N} \times \boldsymbol{\sigma}_u) \cdot \boldsymbol{\sigma}_u = 0$, $(\mathbf{N} \times \boldsymbol{\sigma}_u) \cdot \boldsymbol{\sigma}_v = (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) \cdot \mathbf{N} = \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel \mathbf{N} \cdot \mathbf{N} = \sqrt{EG - F^2}$ by Proposition 6.4.2. This gives the two equations $\alpha E + \beta F = 0$, $\alpha F + \beta G = \sqrt{EG - F^2}$, which imply $\alpha = -F/\sqrt{EG - F^2}$, $\beta = E/\sqrt{EG - F^2}$. The formula for $\mathbf{N} \times \boldsymbol{\sigma}_v$ is proved similarly.

6.5.1 If the internal angles are equal to $\alpha$, Theorem 6.4.7 gives $3\alpha - \pi = 4\pi/4$, so $\alpha = 2\pi/3$. Corollary 6.5.6 then gives the length of a side as $A = \cos^{-1}(-1/3)$.

6.5.2 Using the notation following Proposition 6.5.8, there is a rotation $R_1$ of $S^2$ that takes $\mathbf{a}'$ to $\mathbf{a}$; then a further rotation $R_2$ around the diameter through $\mathbf{a}$ that makes the side through $\mathbf{a}$ and $R_1(\mathbf{b}')$ coincide with the side through $\mathbf{a}$ and $\mathbf{b}$. By Corollary 6.5.6 the two triangles have sides of the same length, so we must have $\mathbf{b} = R_2R_1(\mathbf{b}')$. If $\mathbf{c}$ and $\mathbf{c}'$ are on the same side of the plane containing the side through $\mathbf{a}$ and $\mathbf{b}$, we shall then have $\mathbf{c} = R_2R_1(\mathbf{c}')$ and the isometry $R_2R_1$ takes the triangle with vertices $\mathbf{a}', \mathbf{b}', \mathbf{c}'$ to the triangle with vertices $\mathbf{a}, \mathbf{b}, \mathbf{c}$; if they are on opposite sides the isometry $R_3R_2R_1$ does this, where $R_3$ is reflection in the plane containing the side through $\mathbf{a}$ and $\mathbf{b}$.

6.5.3 By applying an isometry of $\mathbb{R}^3$, which leaves lengths and areas unchanged, we can assume that $\mathbf{p}$ is the north pole $(0, 0, 1)$. The spherical circle of radius $R$ and centre $\mathbf{p}$ is then the circle of latitude $\varphi = \pi/2 - R$ (Example 4.1.4), which is a circle of radius $\sin R$. The area inside it is, by Example 6.1.3 and Proposition 6.4.2, $\int_0^{2\pi}\int_{\pi/2-R}^{\pi/2} \cos\theta d\theta d\varphi = 2\pi(1 - \cos R)$. The maximum value of $R$ is $\pi$; if $\pi/2 \le R \le \pi$, one replaces $R$ by $\pi - R$ in (i) and (ii).

6.5.4 (i) If $M'(w) = \frac{a'w+b'}{c'w+d'}$ is another unitary Möbius transformation, then $(M' \circ M)(w) = \frac{Aw+B}{Cw+D}$, where $A = a'a + b'c, B = a'b + b'd, C = c'a + d'c, D = c'b + d'd$. Thus, $\bar{A} = \bar{a}'\bar{a} + \bar{b}'\bar{c} = d'd + (-c')(-b) = D$ and similarly $C = -\bar{B}$. Inverses are dealt with similarly. (ii) Denoting $(x, y, z) \in \mathbb{R}^3$ by $(\xi, z)$ with $\xi = x + iy \in \mathbb{C}$, the plane through the origin perpendicular to $(a, b)$ is $\bar{w}a + w\bar{a} + 2bz = 0$, and reflection in it is

$F(\xi,z) = (\xi,z) - 2\frac{\bar{w}a+w\bar{a}+2bz}{|a|^2+b^2}(a,b)$. Taking $\xi = \frac{2w}{|w|^2+1}$, $z = \frac{|w^2-1}{|w|^2+1}$, we find that $F(\xi,z) = (\xi',z')$, where $\xi' = \frac{2(|a|^2+b^2)w-2a(\bar{w}a+w\bar{a}+b(|w|^2-1))}{(|w|^2+1)(|a|^2+b^2)}$, $z' = \frac{(|a|^2+b^2)(|w|^2-1)-2b(\bar{w}a+w\bar{a}+b(|w|^2-1))}{(|w|^2+1)(|a|^2+b^2)}$, which gives $w' = \frac{\xi'}{1-z'} = \frac{-ab|w|^2+b^2w-a^2\bar{w}+ab}{b^2|w|^2+b\bar{a}w+ba\bar{w}+|a|^2} = \frac{(-a\bar{w}+b)(bw+a)}{(b\bar{w}+\bar{a})(bw+a)} = \frac{-a\bar{w}+b}{b\bar{w}+\bar{a}}$. (iii) By Proposition 6.5.7, if $F$ is any isometry of $S^2$, $F_\infty = (M_1 \circ J)\circ(M_2 \circ J)\circ\cdots\circ(M_k \circ J)$ for some $k$. Since $J\circ M\circ J$ is easily seen to be a unitary Möbius transformation if $M$ is one, part (i) implies that $F_\infty$ is a unitary Möbius transformation if $k$ is even, and of the form $M \circ J$ with $M$ unitary Möbius if $k$ is odd.

(iv) If $a \in \mathbb{C}, b \in \mathbb{R}$, call the unitary Möbius transformation $M(w) = \frac{aw+b}{-bw+\bar{a}}$ *special unitary.* Then $M = F_\infty \circ J$ where $F$ is as in (ii). Since $J = R_\infty$ where $R$ is reflection in the $yz$-plane, $M = (F\circ R)_\infty$ corresponds to the isometry $F \circ R$ of $S^2$. It therefore suffices to prove that every unitary Möbius transformation is a composite of finitely many special unitary Möbius transformations. If $M'(w) = \frac{Aw+B}{-\bar{B}w+\bar{A}}$ is any unitary Möbius transformation, where $A, B \in \mathbb{C}$, let $B = be^{i\theta}$ with $b,\theta \in \mathbb{R}$. Then $M' = \rho\circ M\circ\rho^{-1}$, where $\rho(w) = e^{i\theta}$ and $M(w) = \frac{Aw+b}{-bw+\bar{A}}$ are both special unitary Möbius transformations ($\rho(w) = \frac{aw+b}{-bw+\bar{a}}$ with $a = e^{i\theta/2}, b = 0$).

# Chapter 7

7.1.1 $\boldsymbol{\sigma}_u = (1,0,2u)$, $\boldsymbol{\sigma}_v = (0,1,2v)$, so $\mathbf{N} = \lambda(-2u,-2v,1)$, where $\lambda = (1+4u^2+4v^2)^{-1/2}$; $\boldsymbol{\sigma}_{uu} = (0,0,2)$, $\boldsymbol{\sigma}_{uv} = \mathbf{0}$, $\boldsymbol{\sigma}_{vv} = (0,0,2)$, so $L = 2\lambda$, $M = 0$, $N = 2\lambda$, and the second fundamental form is $2\lambda(du^2+dv^2)$.

7.1.2 $\boldsymbol{\sigma}_u \cdot \mathbf{N}_u = -\boldsymbol{\sigma}_{uu}\cdot\mathbf{N}$ (since $\boldsymbol{\sigma}_u\cdot\mathbf{N} = 0$), so $\mathbf{N}_u\cdot\boldsymbol{\sigma}_u = 0$; similarly, $\mathbf{N}_u\cdot\boldsymbol{\sigma}_v = \mathbf{N}_v\cdot\boldsymbol{\sigma}_u = \mathbf{N}_v\cdot\boldsymbol{\sigma}_v = 0$; hence, $\mathbf{N}_u$ and $\mathbf{N}_v$ are perpendicular to both $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$, and so are parallel to $\mathbf{N}$. On the other hand, $\mathbf{N}_u$ and $\mathbf{N}_v$ are perpendicular to $\mathbf{N}$ since $\mathbf{N}$ is a unit vector. Thus, $\mathbf{N}_u = \mathbf{N}_v = \mathbf{0}$, and hence $\mathbf{N}$ is constant. Then, $(\boldsymbol{\sigma}\cdot\mathbf{N})_u = \boldsymbol{\sigma}_u\cdot\mathbf{N} = 0$, and similarly $(\boldsymbol{\sigma}\cdot\mathbf{N})_v = 0$, so $\boldsymbol{\sigma}\cdot\mathbf{N}$ is constant, say equal to $d$, and then $\boldsymbol{\sigma}$ is an open subset of the plane $\mathbf{v}\cdot\mathbf{N} = d$.

7.1.3 From Section 4.5, $\tilde{\mathbf{N}} = \pm\mathbf{N}$, the sign being that of $\det(J)$. From $\tilde{\boldsymbol{\sigma}}_{\tilde{u}} = \boldsymbol{\sigma}_u\frac{\partial u}{\partial\tilde{u}} + \boldsymbol{\sigma}_v\frac{\partial v}{\partial\tilde{u}}$, $\tilde{\boldsymbol{\sigma}}_{\tilde{v}} = \boldsymbol{\sigma}_u\frac{\partial u}{\partial\tilde{v}} + \boldsymbol{\sigma}_v\frac{\partial v}{\partial\tilde{v}}$, we get

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}\tilde{u}} = \boldsymbol{\sigma}_u\frac{\partial^2 u}{\partial\tilde{u}^2} + \boldsymbol{\sigma}_v\frac{\partial^2 v}{\partial\tilde{u}^2} + \boldsymbol{\sigma}_{uu}\left(\frac{\partial u}{\partial\tilde{u}}\right)^2 + 2\boldsymbol{\sigma}_{uv}\frac{\partial u}{\partial\tilde{u}}\frac{\partial v}{\partial\tilde{u}} + \boldsymbol{\sigma}_{vv}\left(\frac{\partial v}{\partial\tilde{u}}\right)^2.$$

So $\tilde{L} = \pm\left(L\left(\frac{\partial u}{\partial\tilde{u}}\right)^2 + 2M\frac{\partial u}{\partial\tilde{u}}\frac{\partial v}{\partial\tilde{u}} + N\left(\frac{\partial v}{\partial\tilde{u}}\right)^2\right)$, since $\boldsymbol{\sigma}_u\cdot\mathbf{N} = \boldsymbol{\sigma}_v\cdot\mathbf{N} = 0$. This, together with similar formulas for $\tilde{M}$ and $\tilde{N}$, are equivalent to the matrix equation in the question.

7.1.4 Let $\boldsymbol{\sigma}$ be a surface patch, $P$ a $3\times 3$ orthogonal matrix, $\mathbf{a}\in\mathbb{R}^3$ a constant vector, and $\tilde{\boldsymbol{\sigma}} = P\boldsymbol{\sigma}+\mathbf{a}$. Then, $\tilde{\boldsymbol{\sigma}}_u = P\boldsymbol{\sigma}_u$, $\tilde{\boldsymbol{\sigma}}_v = P\boldsymbol{\sigma}_v$, so $\tilde{\boldsymbol{\sigma}}_u\times\tilde{\boldsymbol{\sigma}}_v = \pm\boldsymbol{\sigma}_u\times\boldsymbol{\sigma}_v$ (Proposition A.1.6), the sign being $+$ if the isometry $\mathbf{v}\mapsto P\mathbf{v}+\mathbf{a}$ is direct and $-$ if it is opposite. It follows that (in the obvious notation), $\tilde{L}=\pm L, \tilde{M}=\pm M, \tilde{N}=\pm N$. The dilation $\mathbf{v}\mapsto a\mathbf{v}$, where $a$ is a non-zero constant, multiplies $\boldsymbol{\sigma}$ by $a$ and hence multiplies each of $L, M, N$ by $a$.

7.2.1 The paraboloid is the level surface $f=0$ where $f(x,y,z)=z-x^2-y^2$ and $\mathbf{N} = \frac{(f_x,f_y,f_z)}{\|(f_x,f_y,f_z)\|}$ is the corresponding unit normal. So $\mathcal{G}(x,y,z) = \frac{(-2x,-2y,1)}{(4x^2+4y^2+1)^{1/2}}$.

7.2.2 This is obvious since changing the orientation changes the Gauss map $\mathcal{G}$ to $-\mathcal{G}$.

7.3.1 Let $t$ be the parameter for $\boldsymbol{\gamma}$, let $s$ be arc-length along $\boldsymbol{\gamma}$, and denote $d/dt$ by a dot and $d/ds$ by a dash. Then, $\dot{\boldsymbol{\gamma}} = \frac{ds}{dt}\boldsymbol{\gamma}'$, $\ddot{\boldsymbol{\gamma}} = \left(\frac{ds}{dt}\right)^2\boldsymbol{\gamma}' + \frac{d^2s}{dt^2}\boldsymbol{\gamma}'$. By Proposition 7.3.5, $\kappa_n = \langle\langle\boldsymbol{\gamma}',\boldsymbol{\gamma}'\rangle\rangle = \langle\langle\frac{1}{(ds/dt)}\dot{\boldsymbol{\gamma}}, \frac{1}{(ds/dt)}\dot{\boldsymbol{\gamma}}\rangle\rangle = \langle\langle\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}}\rangle\rangle/(ds/dt)^2 = \langle\langle\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}}\rangle\rangle/\langle\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}}\rangle$. For the second part, since $\boldsymbol{\gamma}'\cdot(\mathbf{N}\times\boldsymbol{\gamma}')=\mathbf{0}$, we have $\ddot{\boldsymbol{\gamma}}\cdot(\mathbf{N}\times\dot{\boldsymbol{\gamma}}) = \left(\frac{ds}{dt}\right)^3\boldsymbol{\gamma}''\cdot(\mathbf{N}\times\boldsymbol{\gamma}') = \langle\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}}\rangle^{3/2}\kappa_g$.

7.3.2 Let $\boldsymbol{\gamma}$ be a unit-speed curve on the sphere of centre $\mathbf{a}$ and radius $r$. Then, $(\boldsymbol{\gamma}-\mathbf{a})\cdot(\boldsymbol{\gamma}-\mathbf{a}) = r^2$; differentiating gives $\dot{\boldsymbol{\gamma}}\cdot(\boldsymbol{\gamma}-\mathbf{a})=0$, so $\ddot{\boldsymbol{\gamma}}\cdot(\boldsymbol{\gamma}-\mathbf{a}) = -\dot{\boldsymbol{\gamma}}\cdot\dot{\boldsymbol{\gamma}} = -1$. At the point $\boldsymbol{\gamma}(t)$, the unit normal of the sphere is $\mathbf{N} = \pm\frac{1}{r}(\boldsymbol{\gamma}(t)-\mathbf{a})$, so $\kappa_n = \ddot{\boldsymbol{\gamma}}\cdot\mathbf{N} = \pm\frac{1}{r}\ddot{\boldsymbol{\gamma}}\cdot(\boldsymbol{\gamma}-\mathbf{a}) = \mp\frac{1}{r}$.

7.3.3 If the sphere has radius $R$, the parallel with latitude $\theta$ has radius $r=R\cos\theta$; if $\mathbf{p}$ is a point of this circle, its principal normal at $\mathbf{p}$ is parallel to the line through $\mathbf{p}$ perpendicular to the $z$-axis, while the unit normal to the sphere is parallel to the line through $\mathbf{p}$ and the centre of the sphere. The angle $\psi$ in Eq. 7.10 is therefore equal to $\theta$ or $\pi-\theta$ so $\kappa_g = \pm\frac{1}{r}\sin\theta = \pm\frac{1}{R}\tan\theta$. Note that this is zero if and only if the parallel is a great circle.

7.3.4 We have $\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$, so by Exercise 6.4.7, $\mathbf{N}\times\dot{\boldsymbol{\gamma}} = \frac{\dot{u}(E\boldsymbol{\sigma}_v - F\boldsymbol{\sigma}_u)+\dot{v}(F\boldsymbol{\sigma}_v-G\boldsymbol{\sigma}_u)}{\sqrt{EG-F^2}}$, $\ddot{\boldsymbol{\gamma}} = \ddot{u}\boldsymbol{\sigma}_u + \ddot{v}\boldsymbol{\sigma}_v + \dot{u}^2\boldsymbol{\sigma}_{uu} + 2\dot{u}\dot{v}\boldsymbol{\sigma}_{uv} + \dot{v}^2\boldsymbol{\sigma}_{vv}$. Hence, $\kappa_g = \ddot{\boldsymbol{\gamma}}\cdot(\mathbf{N}\times\dot{\boldsymbol{\gamma}}) = (\dot{u}\ddot{v}-\dot{v}\ddot{u})\sqrt{EG-F^2}+A\dot{u}^3+B\dot{u}^2\dot{v}+C\dot{u}\dot{v}^2+D\dot{v}^3$, where $A = \boldsymbol{\sigma}_{uu}\cdot(E\boldsymbol{\sigma}_v - F\boldsymbol{\sigma}_u) = E((\boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_v)_u - \boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_{uv}) - \frac{1}{2}F(\boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_u)_u = E(F_u-\frac{1}{2}E_v)-\frac{1}{2}FE_u$, with similar expressions for $B, C, D$.

If $F=0$, we find by this method that $A = -\frac{1}{2}E_v\sqrt{E/G}$, $B = G_u\sqrt{E/G}-\frac{1}{2}E_u\sqrt{G/E}$, $C = \frac{1}{2}G_v\sqrt{E/G}-E_v\sqrt{G/E}$, $D = \frac{1}{2}G_u\sqrt{G/E}$.

7.3.5 $\kappa_1 = \kappa\mathbf{N}_1\cdot\mathbf{n}$, $\kappa_2 = \kappa\mathbf{N}_2\cdot\mathbf{n}$, so $\kappa_1\mathbf{N}_2-\kappa_2\mathbf{N}_1 = \kappa((\mathbf{N}_1\cdot\mathbf{n})\mathbf{N}_2-(\mathbf{N}_2\cdot\mathbf{n})\mathbf{N}_1) = \kappa(\mathbf{N}_1\times\mathbf{N}_2)\times\mathbf{n}$. Taking the squared length of each side, we get $\kappa_1^2+\kappa_2^2-2\kappa_1\kappa_2\mathbf{N}_1\cdot\mathbf{N}_2 = \kappa^2\parallel(\mathbf{N}_1\times\mathbf{N}_2)\times\mathbf{n}\parallel^2$. Now, $\mathbf{N}_1\cdot\mathbf{N}_2 = \cos\alpha$;

$\dot{\boldsymbol{\gamma}}$ is perpendicular to $\mathbf{N}_1$ and $\mathbf{N}_2$, so $\mathbf{N}_1 \times \mathbf{N}_2$ is parallel to $\dot{\boldsymbol{\gamma}}$, hence perpendicular to $\mathbf{n}$; hence,$\| (\mathbf{N}_1 \times \mathbf{N}_2) \times \mathbf{n} \| = \| \mathbf{N}_1 \times \mathbf{N}_2 \| \| \mathbf{n} \| = \sin \alpha$.

7.3.6 A straight line has a unit-speed parametrization $\boldsymbol{\gamma}(t) = \mathbf{p} + t\,\mathbf{q}$ (with $\mathbf{q}$ a unit vector), so $\ddot{\boldsymbol{\gamma}} = \mathbf{0}$ and hence $\kappa_n = \ddot{\boldsymbol{\gamma}} \cdot \mathbf{N} = 0$. In general, $\kappa_n = 0 \iff \ddot{\boldsymbol{\gamma}}$ is perpendicular to $\mathbf{N} \iff \mathbf{N}$ is perpendicular to $\mathbf{n} \iff \mathbf{N}$ is parallel to $\mathbf{b}$ (since $\mathbf{N}$ is perpendicular to $\mathbf{t}$).

7.3.7 The second fundamental form is $(-du^2 + u^2\,dv^2)/u\sqrt{1+u^2}$, so a curve $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ is asymptotic if and only if $-\dot{u}^2 + u^2\dot{v}^2 = 0$, i.e., $dv/du = \dot{v}/\dot{u} = \pm 1/u$, so $\ln u = \pm(v + c)$, where $c$ is a constant.

7.4.1 $\tilde{\mathbf{v}}$ is a smooth function of $t$ and lies in $T_{\boldsymbol{\gamma}(\varphi(t))}\mathcal{S} = T_{\tilde{\boldsymbol{\gamma}}(t)}\mathcal{S}$, so $\tilde{\mathbf{v}}$ is a tangent vector field along $\tilde{\boldsymbol{\gamma}}$. The formula follows from Eq. 7.11 and the fact that $\frac{d\mathbf{v}}{dt} = \frac{d\tilde{\mathbf{v}}}{d\tilde{t}}\frac{d\varphi}{dt}$, where $\tilde{t} = \varphi(t)$. The last part follows since $\dot{\varphi} \neq 0$ so $\nabla_{\tilde{\boldsymbol{\gamma}}}\tilde{\mathbf{v}} = \mathbf{0} \iff \nabla_{\boldsymbol{\gamma}}\mathbf{v} = \mathbf{0}$.

7.4.2 If $\mathbf{p}$ and $\mathbf{q}$ correspond to the parameter values $t = a$ and $t = b$, respectively, let $\boldsymbol{\Gamma}(t) = \boldsymbol{\gamma}(a + b - t)$ (thus, $\boldsymbol{\Gamma}$ is $\boldsymbol{\gamma}$ 'traversed backwards'). We show that $\Pi_{\boldsymbol{\Gamma}}^{\mathbf{qp}}$ is the inverse of $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}$. Let $\mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ and let $\mathbf{v}$ be the tangent vector field parallel along $\boldsymbol{\gamma}$ such that $\mathbf{v}(a) = \mathbf{w}$. Then, $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}(\mathbf{w}) = \mathbf{v}(b)$. By Exercise 7.4.1, $\mathbf{V}(t) = \mathbf{v}(a + b - t)$ is parallel along $\boldsymbol{\Gamma}$ so $\Pi_{\boldsymbol{\Gamma}}^{\mathbf{qp}}(\mathbf{v}(b)) = \Pi_{\boldsymbol{\Gamma}}^{\mathbf{qp}}(\mathbf{V}(a)) = \mathbf{V}(b) = \mathbf{v}(a) = \mathbf{w}$. This proves that $\Pi_{\boldsymbol{\Gamma}}^{\mathbf{qp}} \circ \Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}}$ is the identity map on $T_{\mathbf{p}}\mathcal{S}$. One proves similarly (or by interchanging the roles of $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$) that $\Pi_{\boldsymbol{\gamma}}^{\mathbf{pq}} \circ \Pi_{\boldsymbol{\Gamma}}^{\mathbf{qp}}$ is the identity map on $T_{\mathbf{q}}\mathcal{S}$.

7.4.3 Let $\alpha, \beta, \gamma$ be the internal angles of the triangle at $\mathbf{p}, \mathbf{q}, \mathbf{r}$, respectively. Since the arc through $\mathbf{p}$ and $\mathbf{q}$ is part of a great circle, the tangent vector of the arc is parallel along the arc (Exercise 7.4.7). So the result of parallel transporting $\mathbf{v}_0$ to $\mathbf{q}$ along the arc $\overline{\mathbf{pq}}$ through $\mathbf{p}$ and $\mathbf{q}$ is a vector $\mathbf{v}_1$ tangent to $\overline{\mathbf{pq}}$ at $\mathbf{q}$. Now $\mathbf{v}_1$ makes an angle $\pi - \beta$ with the arc $\overline{\mathbf{qr}}$ at $\mathbf{q}$, so parallel transporting $\mathbf{v}_1$ along $\overline{\mathbf{qr}}$ to $\mathbf{r}$ gives a vector $\mathbf{v}_2$ which makes an angle $(\pi - \beta) + (\pi - \gamma)$ with the arc $\overline{\mathbf{rp}}$ at $\mathbf{r}$. Parallel transporting $\mathbf{v}_2$ along $\overline{\mathbf{rp}}$ to $\mathbf{p}$ then gives a vector $\mathbf{v}_3$ which makes an angle $(\pi - \beta) + (\pi - \gamma) + (\pi - \alpha)$ with $\mathbf{v}_0$. Since $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ all have the same length (Proposition 7.4.9(ii)), the result follows from Theorem 6.4.7.

# Chapter 8

8.1.1 Parametrize the surface by $\boldsymbol{\sigma}(x, y) = (x, y, f(x, y))$. Then, $\boldsymbol{\sigma}_x = (1, 0, f_x)$, $\boldsymbol{\sigma}_y = (0, 1, f_y)$, $\mathbf{N} = (1 + f_x^2 + f_y^2)^{-1/2}(-f_x, -f_y, 1)$, $\boldsymbol{\sigma}_{xx} = (0, 0, f_{xx})$, $\boldsymbol{\sigma}_{xy} = (0, 0, f_{xy})$, $\boldsymbol{\sigma}_{yy} = (0, 0, f_{yy})$. This gives $E = 1 + f_x^2$, $F = f_x f_y$,

$G = 1 + f_y^2$ and $L = (1 + f_x^2 + f_y^2)^{-1/2} f_{xx}$, $M = (1 + f_x^2 + f_y^2)^{-1/2} f_{xy}$, $N = (1 + f_x^2 + f_y^2)^{-1/2} f_{yy}$. By Corollary 8.1.3, $K = \frac{f_{xx}f_{yy}-f_{xy}^2}{(1+f_x^2+f_y^2)^2}$, $H = \frac{(1+f_y^2)f_{xx}-2f_xf_yf_{xy}+(1+f_x^2)f_{yy}}{2(1+f_x^2+f_y^2)^{3/2}}$.

8.1.2 For the helicoid $\boldsymbol{\sigma}(u, v) = (v \cos u, v \sin u, \lambda u)$, $\boldsymbol{\sigma}_u = (-v \sin u, v \cos u, \lambda)$, $\boldsymbol{\sigma}_v = (\cos u, \sin u, 0)$, $\mathbf{N} = (\lambda^2 + v^2)^{-1/2}(-\lambda \sin u, \lambda \cos u, -v)$, $\boldsymbol{\sigma}_{uu} = (-v \cos u, -v \sin u, 0)$, $\boldsymbol{\sigma}_{uv} = (-\sin u, \cos u, 0)$, $\boldsymbol{\sigma}_{vv} = \mathbf{0}$. This gives $E = \lambda^2 + v^2, F = 0, G = 1$ and $L = N = 0, M = \lambda/\sqrt{\lambda^2 + v^2}$. Hence, $K = (LN - M^2)/(EG - F^2) = -\lambda^2/(\lambda^2 + v^2)^2$.

For the catenoid $\boldsymbol{\sigma}(u, v) = (\cosh u \cos v, \cosh u \sin v, u)$, $\boldsymbol{\sigma}_u = (\sinh u \cos v, \sinh u \sin v, 1)$, $\boldsymbol{\sigma}_v = (-\cosh u \sin v, \cosh u \cos v, 0)$, $\mathbf{N} = \operatorname{sech} u(-\cos v, -\sin v, \sinh u)$, $\boldsymbol{\sigma}_{uu} = (\cosh u \cos v, \cosh u \sin v, 0)$, $\boldsymbol{\sigma}_{uv} = (-\sinh u \sin v, \sinh u \cos v, 0)$, $\boldsymbol{\sigma}_{vv} = (-\cosh u \cos v, -\cosh u \sin v, 0)$. This gives $E = G = \cosh^2 u, F = 0$ and $L = -1, M = 0, N = 1$. Hence, $K = (LN - M^2)/(EG - F^2) = -\operatorname{sech}^4 u$.

8.1.3 Since $\boldsymbol{\sigma}$ is smooth and $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$ is never zero, $\mathbf{N} = \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v / \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel$ is smooth. Hence, $E, F, G, L, M$ and $N$ are smooth. Since $EG - F^2 > 0$ (by the remark following Proposition 6.4.2), the formulas in Corollary 8.1.3 show that $H$ and $K$ are smooth.

8.1.4 From Example 8.1.5, $K = 0 \iff \dot{\boldsymbol{\delta}} \cdot \mathbf{N} = 0 \iff \dot{\boldsymbol{\delta}} \cdot ((\mathbf{t} + v\dot{\boldsymbol{\delta}}) \times \boldsymbol{\delta}) = 0 \iff \dot{\boldsymbol{\delta}} \cdot (\mathbf{t} \times \boldsymbol{\delta}) = 0$. If $\boldsymbol{\delta} = \mathbf{n}$, $\dot{\boldsymbol{\delta}} = -\kappa\mathbf{t} + \tau\mathbf{b}$, $\mathbf{t} \times \boldsymbol{\delta} = \mathbf{b}$, so $K = 0 \iff \tau = 0 \iff \boldsymbol{\gamma}$ is planar (by Proposition 2.3.3). If $\boldsymbol{\delta} = \mathbf{b}$, $\dot{\boldsymbol{\delta}} = -\tau\mathbf{n}$, $\mathbf{t} \times \boldsymbol{\delta} = -\mathbf{n}$, so again $K = 0 \iff \tau = 0$.

8.1.5 The dilation $(x, y, z) \mapsto (ax, ay, az)$, where $a$ is a non-zero constant, multiplies $E, F, G$ by $a^2$ and $L, M, N$ by $a$, hence $H$ by $a^{-1}$ and $K$ by $a^{-2}$ (using Corollary 8.1.3).

8.1.6 This follows immediately from Definition 8.1.1 and the hint.

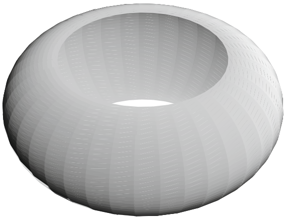
8.1.7 Suppose that the cone is the union of the straight lines joining points of a curve $\mathcal{C}$ to a vertex $\mathbf{v}$. It is clear that the Gauss map $\mathcal{G}$ is constant along the rulings of the cone, so the image of the cone under $\mathcal{G}$ is the same as the image of $\mathcal{C}$ under $\mathcal{G}$, which is a curve.

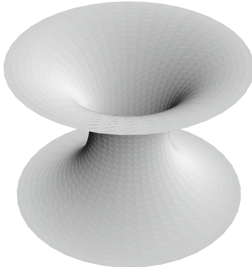
8.1.8 By Eq. 8.2, the area of $\boldsymbol{\sigma}(R)$ is

$$\int_R \parallel \mathbf{N}_u \times \mathbf{N}_v \parallel dudv = \int_R |K| \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel dudv = \int_R |K| d\mathcal{A}_{\boldsymbol{\sigma}}.$$

8.1.9 Using the parametrization $\boldsymbol{\sigma}$ in Exercise 4.2.5, we find that $E = b^2, F = 0$, $G = (a + b \cos\theta)^2$ and $L = b, M = 0, N = (a + b \cos\theta) \cos\theta$. This gives

$K = \cos\theta/b(a+b\cos\theta)$. It follows that $\mathcal{S}^+$ and $\mathcal{S}^-$ are the annular regions on the torus given by $-\pi/2 \leq u \leq \pi/2$ and $\pi/2 \leq u \leq 3\pi/2$, respectively.





$\mathcal{S}^+$ $\mathcal{S}^-$

It is clear that as a point $\mathbf{p}$ moves over $\mathcal{S}^+$ (resp. $\mathcal{S}^-$), the unit normal at $\mathbf{p}$ covers the whole of the unit sphere. Hence, $\int_{\mathcal{S}^+} |K| d\mathcal{A} = \int_{\mathcal{S}^-} |K| d\mathcal{A} = 4\pi$ by the preceding exercise; since $|K| = \pm K$ on $\mathcal{S}^\pm$, this gives the result.

8.1.10 $\nabla_u \mathbf{w} = \mathbf{w}_u - (\mathbf{w}_u \cdot \mathbf{N})\mathbf{N}$ so $\nabla_v(\nabla_u \mathbf{w}) = \mathbf{w}_{uv} - (\mathbf{w}_{uv} \cdot \mathbf{N})\mathbf{N} - (\mathbf{w}_u \cdot \mathbf{N}_v)\mathbf{N} - (\mathbf{w}_u \cdot \mathbf{N})\mathbf{N}_v - (\mathbf{w}_{uv} \cdot \mathbf{N})\mathbf{N} + (\mathbf{w}_{uv} \cdot \mathbf{N})\mathbf{N} + (\mathbf{w}_u \cdot \mathbf{N}_v)\mathbf{N} + (\mathbf{w}_u \cdot \mathbf{N})(\mathbf{N}_v \cdot \mathbf{N})\mathbf{N} = \mathbf{w}_{uv} - (\mathbf{w}_{uv} \cdot \mathbf{N})\mathbf{N} - (\mathbf{w}_u \cdot \mathbf{N})\mathbf{N}_v$. Interchanging $u$ and $v$ and subtracting gives the first formula. Replacing $\mathbf{w}$ by $\lambda\mathbf{w}$ in this formula gives $\lambda\{(\mathbf{w}_v \cdot \mathbf{N})\mathbf{N}_u - (\mathbf{w}_u \cdot \mathbf{N})\mathbf{N}_v\} + \lambda_v(\mathbf{w} \cdot \mathbf{N})\mathbf{N}_u - \lambda_u(\mathbf{w} \cdot \mathbf{N})\mathbf{N}_v = \lambda\{(\mathbf{w}_v \cdot \mathbf{N})\mathbf{N}_u - (\mathbf{w}_u \cdot \mathbf{N})\mathbf{N}_v\}$ since $\mathbf{w} \cdot \mathbf{N} = 0$. It is also obvious that $\nabla_v(\nabla_u(\mathbf{w}_1 + \mathbf{w}_2)) - \nabla_u(\nabla_v(\mathbf{w}_1 + \mathbf{w}_2)) = (\nabla_v(\nabla_u \mathbf{w}_1) - \nabla_u(\nabla_v \mathbf{w}_1)) + (\nabla_v(\nabla_u \mathbf{w}_2) - \nabla_u(\nabla_v \mathbf{w}_2))$ for any two tangent vector fields $\mathbf{w}_1, \mathbf{w}_2$.

Now $\nabla_v(\nabla_u \boldsymbol{\sigma}_u) - \nabla_u(\nabla_v \boldsymbol{\sigma}_u) = (\boldsymbol{\sigma}_{uv} \cdot \mathbf{N})\mathbf{N}_u - (\boldsymbol{\sigma}_{uu} \cdot \mathbf{N})\mathbf{N}_v = M\mathbf{N}_u - L\mathbf{N}_v$ (in the usual notation). Using Proposition 8.1.2, this is equal to $M(a\boldsymbol{\sigma}_u + b\boldsymbol{\sigma}_v) - L(c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v)$ and using the explicit expressions for $a, b, c, d$ in Proposition 8.1.2 this becomes $K(E\boldsymbol{\sigma}_v - F\boldsymbol{\sigma}_u)$. Similarly, $\nabla_v(\nabla_u \boldsymbol{\sigma}_v) - \nabla_u(\nabla_v \boldsymbol{\sigma}_v) = K(-G\boldsymbol{\sigma}_u + F\boldsymbol{\sigma}_v)$.

If $\nabla_v(\nabla_u \mathbf{w}) - \nabla_u(\nabla_v \mathbf{w}) = \mathbf{0}$ (*) for all $\mathbf{w}$, then taking $\mathbf{w} = \boldsymbol{\sigma}_u$ gives $K = 0$ since $E \neq 0$, $\boldsymbol{\sigma}_v \neq \mathbf{0}$. Conversely, if $K = 0$ then (*) holds for $\mathbf{w} = \boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$, and hence by the first part of the exercise it holds for $\alpha\boldsymbol{\sigma}_u + \beta\boldsymbol{\sigma}_v$ for all smooth functions $\alpha, \beta$ of $(u, v)$. But every tangent vector field $\mathbf{w}$ is of this form.

8.2.1 For the helicoid $\boldsymbol{\sigma}(u, v) = (v\cos u, v\sin u, \lambda u)$, the first and second fundamental forms are $(\lambda^2 + v^2)du^2 + dv^2$ and $2\lambda du dv/\sqrt{\lambda^2 + v^2}$, respectively. Hence, the principal curvatures are the roots of $\begin{vmatrix} -\kappa(\lambda^2 + v^2) & \frac{\lambda}{\sqrt{\lambda^2+v^2}} \\ \frac{\lambda}{\sqrt{\lambda^2+v^2}} & -\kappa \end{vmatrix} = 0$, i.e., $\kappa = \pm\lambda/(\lambda^2 + v^2)$. For the catenoid $\boldsymbol{\sigma}(u, v) = (\cosh u \cos v, \cosh u \sin v, u)$, the first and second fundamental forms are $\cosh^2 u(du^2+dv^2)$ and $-du^2+dv^2$, so the principal curvatures are the roots of $\begin{vmatrix} -1-\kappa\cosh^2 u & 0 \\ 0 & 1-\kappa\cosh^2 u \end{vmatrix} = 0$, i.e., $\kappa = \pm\mathrm{sech}^2 u$.

8.2.2 This is obvious, since $\mathcal{W}(\dot{\boldsymbol{\gamma}}) = -\dot{\mathbf{N}}$.

8.2.3 $\boldsymbol{\gamma}$ is a line of curvature $\Longleftrightarrow \dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_n$ is a principal vector for all $t \Longleftrightarrow \begin{pmatrix} L & M \\ L & N \end{pmatrix} = \kappa \begin{pmatrix} E & F \\ F & G \end{pmatrix}$ for some scalar $\kappa$. Writing this matrix equation as two scalar equations and then eliminating $\kappa$ gives the stated equation. For the second part, if the second fundamental form is a multiple of the first, the Weingarten map is a scalar multiple of the identity map, so every tangent vector is principal and every curve on the surface is a line of curvature. If $F = M = 0$ the matrices $\mathcal{F}_I$ and $\mathcal{F}_{II}$ are diagonal, hence so is the matrix $\mathcal{F}_I^{-1}\mathcal{F}_{II}$ of the Weingarten map with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$. This means that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are principal vectors, i.e., that the parameter curves $v =$ constant and $u =$ constant are lines of curvature. Conversely, if every parameter curve is a line of curvature, the stated equation must hold if $\dot{u} = 0$ and if $\dot{v} = 0$. This gives $EM = FL$ and $FN = GM$, which imply that $(EN - GL)F = EGM - EGM = 0$ and so $F = 0$ and then $GM = 0$ so $M = 0$. If $EN = GL$ the equation in the exercise implies that every curve is a line of curvature, so every tangent vector is principal, so (i) holds. Condition (i) implies that the two principal curvatures are equal everywhere, i.e., every point is an umbilic, so $\boldsymbol{\sigma}$ is an open subset of a plane or a sphere by Proposition 8.2.9. From Examples 6.1.3 and 7.1.2, the first and second fundamental forms of a surface of revolution are $du^2 + f(u)^2 dv^2$ and $(\dot{f}\ddot{g} - \ddot{f}\dot{g})du^2 + f\dot{g}dv^2$, respectively. Since the terms $dudv$ are absent, the vectors $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are principal; but these are tangent to the meridians and parallels, respectively.

8.2.4 Let $\mathbf{N}_1$ be a unit normal of $\mathcal{S}$. Then, $K = 0 \Longleftrightarrow \dot{\mathbf{N}}_1.(\mathbf{t} \times \mathbf{N}_1) = 0$. Since $\dot{\mathbf{N}}_1$ is perpendicular to $\mathbf{N}_1$ and $\mathbf{N}_1$ is perpendicular to $\mathbf{t}$, this condition holds $\Longleftrightarrow \dot{\mathbf{N}}_1$ is parallel to $\mathbf{t}$, i.e., $\Longleftrightarrow \dot{\mathbf{N}}_1 = -\lambda\dot{\boldsymbol{\gamma}}$ for some scalar $\lambda$. Now use Exercise 8.2.2.

8.2.5 Let $\mathbf{N}_1$ and $\mathbf{N}_2$ be unit normals of the two surfaces; if $\boldsymbol{\gamma}$ is a unit-speed parametrization of $\mathcal{C}$, then $\dot{\mathbf{N}}_1 = -\lambda_1\dot{\boldsymbol{\gamma}}$ for some scalar $\lambda_1$ by Exercise 8.2.2. If $\mathcal{C}$ is a line of curvature of $\mathcal{S}_2$, then $\dot{\mathbf{N}}_2 = -\lambda_2\dot{\boldsymbol{\gamma}}$ for some scalar $\lambda_2$, and then $(\mathbf{N}_1\cdot\mathbf{N}_2)^{\cdot} = -\lambda_1\dot{\boldsymbol{\gamma}}\cdot\mathbf{N}_2 - \lambda_2\dot{\boldsymbol{\gamma}}\cdot\mathbf{N}_1 = 0$, so $\mathbf{N}_1\cdot\mathbf{N}_2$ is constant along $\boldsymbol{\gamma}$, showing that the angle between $\mathcal{S}_1$ and $\mathcal{S}_2$ is constant. Conversely, if $\mathbf{N}_1 \cdot \mathbf{N}_2$ is constant, then $\mathbf{N}_1.\dot{\mathbf{N}}_2 = 0$ since $\dot{\mathbf{N}}_1 \cdot \mathbf{N}_2 = -\lambda_1\dot{\boldsymbol{\gamma}} \cdot \mathbf{N}_2 = 0$; thus, $\dot{\mathbf{N}}_2$ is perpendicular to $\mathbf{N}_1$, and is also perpendicular to $\mathbf{N}_2$ as $\mathbf{N}_2$ is a unit vector; but $\dot{\boldsymbol{\gamma}}$ is also perpendicular to $\mathbf{N}_1$ and $\mathbf{N}_2$; hence, $\dot{\mathbf{N}}_2$ must be parallel to $\dot{\boldsymbol{\gamma}}$, so there is a scalar $\lambda_2$ (say) such that $\dot{\mathbf{N}}_2 = -\lambda_2\dot{\boldsymbol{\gamma}}$.

8.2.6 (i) Differentiate the three equations in (8.5) with respect to $w, u$ and $v$, respectively; this gives $\boldsymbol{\sigma}_{uw} \cdot \boldsymbol{\sigma}_v + \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{vw} = 0$, $\boldsymbol{\sigma}_{uv} \cdot \boldsymbol{\sigma}_w + \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uw} = 0$, $\boldsymbol{\sigma}_{vw} \cdot \boldsymbol{\sigma}_u + \boldsymbol{\sigma}_w \cdot \boldsymbol{\sigma}_{uv} = 0$. Subtracting the second equation from the sum of the other two gives $\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{vw} = 0$, and similarly $\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_{uw} = \boldsymbol{\sigma}_w \cdot \boldsymbol{\sigma}_{uv} = 0$. (ii) Since $\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_w = 0$, it follows that the matrix $\mathcal{F}_I$ for the $u = u_0$ surface is diagonal (and similarly for the others). Let $\mathbf{N}$ be the unit normal of the $u = u_0$ surface; $\mathbf{N}$ is parallel to $\boldsymbol{\sigma}_v \times \boldsymbol{\sigma}_w$ by definition, and hence to $\boldsymbol{\sigma}_u$ since $\boldsymbol{\sigma}_u$, $\boldsymbol{\sigma}_v$ and $\boldsymbol{\sigma}_w$ are perpendicular; by (i), $\boldsymbol{\sigma}_{vw} \cdot \boldsymbol{\sigma}_u = 0$, hence $\boldsymbol{\sigma}_{vw} \cdot \mathbf{N} = 0$, proving that the matrix $\mathcal{F}_{II}$ for the $u = u_0$ surface is diagonal. (iii) By part (ii), the parameter curves of each surface $u = u_0$ are lines of curvature. But the parameter curve $v = v_0$, say, on this surface is the curve of intersection of the $u = u_0$ surface with the $v = v_0$ surface.

8.2.7 On the open subset of the ellipsoid with $z \neq 0$, we can use the parametrization $\boldsymbol{\sigma}(x, y) = (x, y, z)$, where $z = \pm r\sqrt{1 - \frac{x^2}{p^2} - \frac{y^2}{q^2}}$. By Proposition 8.1.2 and the remarks following Proposition 8.2.1, the condition for an umbilic is that $\mathcal{F}_{II} = \kappa \mathcal{F}_I$ for some scalar $\kappa$. The formulas in the solution of Exercise 8.1.1 lead to the equations $z_{xx} = \lambda(1+z_x^2)$, $z_{xy} = \lambda z_x z_y$, $z_{yy} = \lambda(1 + z_y^2)$, where $\lambda = \kappa\sqrt{1 + z_x^2 + z_y^2}$. If $x$ and $y$ are both non-zero, the middle equation gives $\lambda = -1/z$, and substituting into the first equation gives the contradiction $p^2 = r^2$. Hence, either $x = 0$ or $y = 0$. If $x = 0$, the equations have the four solutions

$$x = 0, \quad y = \pm q\sqrt{\frac{q^2 - p^2}{q^2 - r^2}}, \quad z = \pm r\sqrt{\frac{r^2 - p^2}{r^2 - q^2}}.$$

Similarly, one finds the following eight other candidates for umbilics:

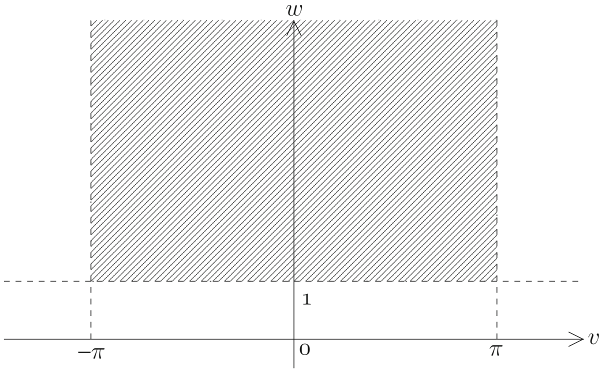
$$x = \pm p\sqrt{\frac{p^2 - q^2}{p^2 - r^2}}, \quad y = 0, \quad z = \pm r\sqrt{\frac{r^2 - q^2}{r^2 - p^2}},$$

$$x = \pm p\sqrt{\frac{p^2 - r^2}{p^2 - q^2}}, \quad y = \pm q\sqrt{\frac{q^2 - r^2}{q^2 - p^2}}, \quad z = 0.$$

Of these 12 points, exactly 4 are real, depending on the relative sizes of $p^2, q^2$ and $r^2$.

If $p = q \neq r$, the only umbilics are the two points $(0, 0, \pm r)$. If $p = q = r$ every point of the ellipsoid (now a sphere) is an umbilic.

8.2.8 By Proposition 8.2.3, the principal curvatures are the roots of the quadratic equation $\kappa^2 - 2H\kappa + K = 0$, i.e., $H \pm \sqrt{H^2 - K}$. If there are no umbilics, we must have $H^2 > K$, and then the principal curvatures are smooth because $H$ and $K$ are (Exercise 8.1.3). The second part follows from Exercises 6.1.4 and 7.1.3 and Proposition 8.2.6.

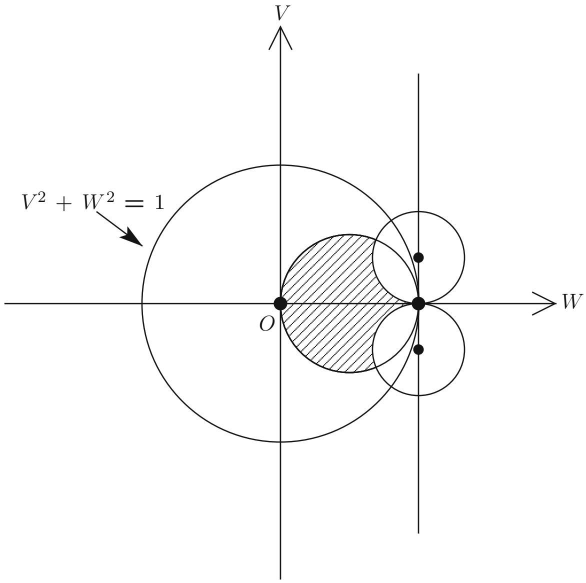



8.3.1 (i) Setting $\tilde{u} = v, \tilde{v} = w = e^{-u}$, we have $u = -\ln\tilde{v}, v = \tilde{u}$ so, in the notation of Exercise 6.1.4, $J = \begin{pmatrix} 0 & -\frac{1}{\tilde{v}} \\ 1 & 0 \end{pmatrix}$. Since $J$ is invertible, $(u, v) \mapsto (v, w)$ is a reparametrization map. The first fundamental form in terms of $v, w$ is given by $\begin{pmatrix} \tilde{E} & \tilde{F} \\ \tilde{F} & \tilde{G} \end{pmatrix} = J^t \begin{pmatrix} E & F \\ F & G \end{pmatrix} J = \begin{pmatrix} 0 & 1 \\ -\frac{1}{\tilde{v}} & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & f(u)^2 \end{pmatrix}\begin{pmatrix} 0 & -\frac{1}{\tilde{v}} \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{w^2} & 0 \\ 0 & \frac{1}{w^2} \end{pmatrix}$, so the first fundamental form is $(dv^2 + dw^2)/w^2$.

(ii) We find that the matrix $\tilde{J} = \begin{pmatrix} \frac{\partial v}{\partial V} & \frac{\partial v}{\partial W} \\ \frac{\partial w}{\partial V} & \frac{\partial w}{\partial W} \end{pmatrix} = \begin{pmatrix} v(w+1) & \frac{1}{2}(v^2 - (w+1)^2) \\ -\frac{1}{2}(v^2 - (w+1)^2) & v(w+1) \end{pmatrix}$, so the first fundamental form matrix in terms of $V$ and $W$ is $\tilde{J}^t \begin{pmatrix} \frac{1}{w^2} & 0 \\ 0 & \frac{1}{w^2} \end{pmatrix} \tilde{J} = \frac{(v^2+(w+1)^2)^2}{4w^2} I = \frac{4}{(1-V^2-W^2)^2} I$, after some tedious algebra.

In (i), $u < 0$ and $-\pi < v < \pi$ corresponds to $-\pi < v < \pi$ and $w > 1$, a semi-infinite rectangle in the upper half of the $vw$-plane.

To find the corresponding region in (ii), it is convenient to introduce the complex numbers $z = v + iw$, $Z = V + iW$. Then, the equations in (ii) are equivalent to $Z = \frac{z-i}{z+i}$, $z = \frac{Z+1}{i(Z-1)}$. The line $v = \pi$ in the $vw$-plane corresponds to $z + \bar{z} = 2\pi$ (the bar denoting complex conjugate), i.e., $\frac{Z+1}{i(Z-1)} - \frac{\bar{Z}+1}{i(\bar{Z}-1)} = 2\pi$, which simplifies to $|Z - (1 - \frac{i}{\pi})|^2 = \frac{1}{\pi^2}$; so $v = \pi$ corresponds to the circle in the $VW$-plane with centre $1 - \frac{i}{\pi}$ and radius $\frac{1}{\pi}$. Similarly, $v = -\pi$ corresponds to the circle with centre $1 + \frac{i}{\pi}$ and radius $\frac{1}{\pi}$. Finally, $w = 1$ corresponds to $z - \bar{z} = 2i$, i.e., $\frac{Z+1}{i(Z-1)} + \frac{\bar{Z}+1}{i(\bar{Z}-1)} = 2i$. This simplifies to $|Z - \frac{1}{2}|^2 = \frac{1}{4}$; so $w = 1$ corresponds to the circle with centre $1/2$ and radius $1/2$ in the $VW$-plane. The required region in the $VW$-plane is that bounded by these three circles:




For (iii) we follow the hint and make use of polar coordinates on the disc, $V = r\cos\theta$, $W = r\sin\theta$, $\bar{V} = \bar{r}\cos\bar{\theta}$, $\bar{W} = \bar{r}\sin\bar{\theta}$. We find that $\bar{r} = \frac{2r}{r^2+1}$, $\bar{\theta} = \theta$. Suppose that the first fundamental form in terms of these parameters is $Ed\bar{r}^2 + 2Fd\bar{r}d\bar{\theta} + Gd\bar{\theta}^2$. Since $\frac{d\bar{r}}{dr} = \frac{2(1-r^2)}{(1+r^2)^2}$, the first fundamental form is $\frac{4(1-r^2)^2}{(1+r^2)^4}Edr^2 + \frac{4(1-r^2)}{(1+r^2)^2}Fdrd\theta + Gd\theta^2$. Equating this to $\frac{4(dV^2+dW^2)}{(1-V^2-W^2)^2} = \frac{4(dr^2+r^2d\theta^2)}{(1-r^2)^2}$, we get $E = \frac{(1+r^2)^4}{(1-r^2)^4} = \frac{1}{(1-\bar{r}^2)^2}$, $F = 0$, $G = \frac{4r^2}{(1-r^2)^2} = \frac{\bar{r}^2}{1-\bar{r}^2}$. Converting back to the parameters $(\bar{V}, \bar{W})$, we have $\bar{r}d\bar{r} = \bar{V}d\bar{V} + \bar{W}d\bar{W}, \bar{r}^2d\bar{\theta} = \bar{V}d\bar{W} - \bar{W}d\bar{V}$, so the first fundamental form becomes $\frac{(\bar{V}d\bar{V}+\bar{W}d\bar{W})^2+(1-\bar{r}^2)(\bar{V}d\bar{W}-\bar{W}d\bar{V})^2}{\bar{r}^2(1-\bar{r}^2)^2} =$
$\frac{(\bar{V}^2+(1-\bar{r}^2)\bar{W}^2)d\bar{V}^2+2\bar{r}^2\bar{V}\bar{W}d\bar{V}d\bar{W}+(\bar{W}^2+(1-\bar{r}^2)\bar{V}^2)d\bar{W}^2}{\bar{r}^2(1-\bar{r}^2)^2}$
$= \frac{(1-\bar{W}^2)d\bar{V}^2+2\bar{V}\bar{W}d\bar{V}d\bar{W}+(1-\bar{V}^2)d\bar{W}^2}{(1-\bar{V}^2-\bar{W}^2)^2}$.

8.4.1 Let $\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$ be a patch of $\mathcal{S}$ containing $\mathbf{p} = \tilde{\boldsymbol{\sigma}}(\tilde{u}_0, \tilde{v}_0)$. The Gaussian curvature $K$ of $\mathcal{S}$ is $< 0$ at $\mathbf{p}$; since $K$ is a smooth function of $(\tilde{u}, \tilde{v})$ (Exercise 8.1.3), $K(\tilde{u}, \tilde{v}) < 0$ for $(\tilde{u}, \tilde{v})$ in some open set $\tilde{U}$ containing $(\tilde{u}_0, \tilde{v}_0)$; then every point of $\tilde{\boldsymbol{\sigma}}(\tilde{U})$ is hyperbolic. Let $\kappa_1, \kappa_2$ be the principal curvatures of $\tilde{\boldsymbol{\sigma}}$, let $0 < \theta < \pi/2$ be such that $\tan\theta = \sqrt{-\kappa_1/\kappa_2}$, and let $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ be the unit tangent vectors of $\tilde{\boldsymbol{\sigma}}$ making angles $\theta$ and $-\theta$, respectively, with the principal vector corresponding to $\kappa_1$ (see Theorem 8.2.4). Applying Proposition 8.4.3 gives the result. For the last part, put $\dot{v} = 0$ in the formula for $\kappa_n$ in Proposition 7.3.5: this shows that $L = 0$ if the parameter curves $v =$ constant are asymptotic. Similarly $N = 0$ if the parameter curves $u =$ constant are asymptotic.

8.5.1 By Corollary 8.1.3 and the fact that $\boldsymbol{\sigma}$ is conformal, the mean curvature of $\boldsymbol{\sigma}$ is $H = \frac{L+N}{2E}$, so $H = 0 \iff L+N = 0$, i.e., $\iff (\boldsymbol{\sigma}_{uu}+\boldsymbol{\sigma}_{vv})\cdot\mathbf{N} = 0$ (*). Obviously, then, $H = 0$ if $\Delta\boldsymbol{\sigma} = \boldsymbol{\sigma}_{uu}+\boldsymbol{\sigma}_{vv} = \mathbf{0}$. For the converse, we have to show that $\Delta\boldsymbol{\sigma} = \mathbf{0}$ if (*) holds. It is enough to prove that $\Delta\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}_u = \Delta\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}_v = 0$, since $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v, \mathbf{N}\}$ is a basis of $\mathbb{R}^3$. We compute $\Delta\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}_u = \boldsymbol{\sigma}_{uu}\cdot\boldsymbol{\sigma}_u + \boldsymbol{\sigma}_{vv}\cdot\boldsymbol{\sigma}_u = \frac{1}{2}(\boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_u)_u + (\boldsymbol{\sigma}_v\cdot\boldsymbol{\sigma}_u)_v - (\boldsymbol{\sigma}_v\cdot\boldsymbol{\sigma}_{uv}) = \frac{1}{2}(\boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_u - \boldsymbol{\sigma}_v\cdot\boldsymbol{\sigma}_v)_u + (\boldsymbol{\sigma}_v\cdot\boldsymbol{\sigma}_u)_v$. Since $\boldsymbol{\sigma}$ is conformal, $\boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_u = \boldsymbol{\sigma}_v\cdot\boldsymbol{\sigma}_v$ and $\boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_v = 0$. Hence, $\Delta\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}_u = 0$. Similarly, $\Delta\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}_v = 0$. The first fundamental form of the given surface patch is $(1+u^2+v^2)^2(du^2+dv^2)$, so it is conformal, and $\boldsymbol{\sigma}_{uu}+\boldsymbol{\sigma}_{vv} = (-2u, 2v, 2) + (2u, -2v, -2) = \mathbf{0}$.

8.5.2 Using the formula in Exercise 8.1.1 with $f(x, y) = \ln\cos y - \ln\cos x$ gives $H = \frac{\sec^2 x(1+\tan^2 y) - \sec^2 y(1+\tan^2 x)}{2(1+\tan^2 x+\tan^2 y)^{3/2}} = 0$.

8.5.3 $\boldsymbol{\Sigma}_u = \boldsymbol{\sigma}_u + w\mathbf{N}_u$, $\boldsymbol{\Sigma}_v = \boldsymbol{\sigma}_v + w\mathbf{N}_v$, $\boldsymbol{\Sigma}_w = \mathbf{N}$. $\boldsymbol{\Sigma}_u\cdot\boldsymbol{\Sigma}_w = 0$ since $\boldsymbol{\sigma}_u\cdot\mathbf{N} = \mathbf{N}_u\cdot\mathbf{N} = 0$, and similarly $\boldsymbol{\Sigma}_v\cdot\boldsymbol{\Sigma}_w = 0$. Finally, $\boldsymbol{\Sigma}_u\cdot\boldsymbol{\Sigma}_v = \boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_v + w(\boldsymbol{\sigma}_u\cdot\mathbf{N}_v + \boldsymbol{\sigma}_v\cdot\mathbf{N}_u) + w^2\mathbf{N}_u\cdot\mathbf{N}_v = F - 2wM + w^2\mathbf{N}_u\cdot\mathbf{N}_v = w^2\mathbf{N}_u\cdot\mathbf{N}_v$; by the proof of Proposition 8.1.2, $\mathbf{N}_u = -\frac{L}{E}\boldsymbol{\sigma}_u$, $\mathbf{N}_v = -\frac{N}{G}\boldsymbol{\sigma}_v$, so $\mathbf{N}_u\cdot\mathbf{N}_v = \frac{LN}{EG}F = 0$. Every surface $u = u_0$ (a constant) is ruled as it is the union of the straight lines given by $v =$ constant; by Exercise 8.2.4, this surface is flat provided the curve $\boldsymbol{\gamma}(v) = \boldsymbol{\sigma}(u_0, v)$ is a line of curvature of $\mathcal{S}$, i.e., if $\boldsymbol{\sigma}_v$ is a principal vector; but this is true since the matrices $\mathcal{F}_I$ and $\mathcal{F}_{II}$ are diagonal. Similarly for the surfaces $v =$ constant.

# Chapter 9

9.1.1 By Exercise 4.1.3, there are two straight lines on the hyperboloid passing through $(1, 0, 0)$; by Proposition 9.1.4, they are geodesics. The circle given by $z = 0, x^2+y^2 = 1$ and the hyperbola given by $y = 0, x^2 - z^2 = 1$ are both normal sections, hence geodesics by Proposition 9.1.6.

9.1.2 Let $\kappa(\boldsymbol{\gamma}) = \ddot{\boldsymbol{\gamma}}\cdot(\mathbf{N}\times\dot{\boldsymbol{\gamma}})$. Note that if $\boldsymbol{\Gamma}(t) = \boldsymbol{\gamma}(\varphi(t))$ is a reparametrization of $\boldsymbol{\gamma}$, then $\kappa(\boldsymbol{\Gamma}) = \left(\frac{d\varphi}{dt}\right)^3\kappa(\boldsymbol{\gamma})$. In particular, $\kappa(\boldsymbol{\gamma}) = 0 \iff \kappa(\boldsymbol{\Gamma}) = 0$. For (i), let $\boldsymbol{\gamma}$ be a pre-geodesic and let $\boldsymbol{\Gamma}$ be a geodesic reparametrization of $\boldsymbol{\gamma}$. By Proposition 9.1.2 $\boldsymbol{\Gamma}$ has constant speed, say $v$, and then $\tilde{\boldsymbol{\Gamma}}(t) = \boldsymbol{\Gamma}(t/v)$ is a unit-speed geodesic. By Proposition 9.1.3, $\kappa(\tilde{\boldsymbol{\Gamma}}) = 0$, hence $\kappa(\boldsymbol{\Gamma}) = 0$, hence $\kappa(\boldsymbol{\gamma}) = 0$. Conversely, if $\kappa(\boldsymbol{\gamma}) = 0$ and if $\boldsymbol{\Gamma}$ is a unit-speed reparametrization of $\boldsymbol{\gamma}$, then $\kappa(\boldsymbol{\Gamma}) = 0$ so $\boldsymbol{\Gamma}$ is a geodesic by Proposition 9.1.3. Part (ii) is obvious. For (iii), let $\boldsymbol{\gamma}$ be a constant speed pre-geodesic, say with speed $v$. Then $\boldsymbol{\Gamma}(t) = \boldsymbol{\gamma}(t/v)$ is a unit-speed pre-geodesic, hence a geodesic by (i) and Proposition 9.1.3.

Since $\ddot{\boldsymbol\gamma} = v^2\ddot{\boldsymbol\Gamma}$, $\ddot{\boldsymbol\gamma}$ is perpendicular to the surface, so $\boldsymbol\gamma$ is a geodesic. Finally, (iv) follows from (iii) and Proposition 9.1.2.

9.1.3 Let $\Pi_s$ be the plane through $\boldsymbol\gamma(s)$ perpendicular to $\mathbf{t}(s)$; the parameter curve $s$ = constant is the intersection of the surface with $\Pi_s$. From the solution to Exercise 4.2.7, the standard unit normal of $\boldsymbol\sigma$ is $\mathbf{N} = -(\cos\theta\,\mathbf{n} + \sin\theta\,\mathbf{b})$. Since this is perpendicular to $\mathbf{t}$, the circles in question are normal sections.

9.1.4 Take the ellipsoid to be $\frac{x^2}{p^2} + \frac{y^2}{q^2} + \frac{z^2}{r^2} = 1$; the vector $(\frac{x}{p^2}, \frac{y}{q^2}, \frac{z}{r^2})$ is normal to the ellipsoid by Exercise 5.1.2. If $\boldsymbol\gamma(t) = (f(t), g(t), h(t))$ is a curve on the ellipsoid, $R = \left(\frac{\dot f^2}{p^2} + \frac{\dot g^2}{q^2} + \frac{\dot h^2}{r^2}\right)^{-1/2}$, $S = \left(\frac{f^2}{p^4} + \frac{g^2}{q^4} + \frac{h^2}{r^4}\right)^{-1/2}$. Now, $\boldsymbol\gamma$ is a geodesic $\Longleftrightarrow$ $\ddot{\boldsymbol\gamma}$ is parallel to the normal $\Longleftrightarrow$ $(\ddot f, \ddot g, \ddot h) = \lambda\left(\frac{f}{p^2}, \frac{g}{q^2}, \frac{h}{r^2}\right)$ for some scalar $\lambda(t)$. From $\frac{f^2}{p^2} + \frac{g^2}{q^2} + \frac{h^2}{r^2} = 1$ we get $\frac{f\dot f}{p^2} + \frac{g\dot g}{q^2} + \frac{h\dot h}{r^2} = 0$, hence $\frac{\dot f^2}{p^2} + \frac{\dot g^2}{q^2} + \frac{\dot h^2}{r^2} + \frac{f\ddot f}{p^2} + \frac{g\ddot g}{q^2} + \frac{h\ddot h}{r^2} = 0$, i.e., $\frac{\dot f^2}{p^2} + \frac{\dot g^2}{q^2} + \frac{\dot h^2}{r^2} + \lambda\left(\frac{f^2}{p^4} + \frac{g^2}{q^4} + \frac{h^2}{r^4}\right) = 0$, which gives $\lambda = -S^2/R^2$. The curvature $\parallel \ddot{\boldsymbol\gamma} \parallel = (\ddot f^2 + \ddot g^2 + \ddot h^2)^{1/2} = |\lambda| \left(\frac{f^2}{p^4} + \frac{g^2}{q^4} + \frac{h^2}{r^4}\right)^{1/2} = \frac{|\lambda|}{S} = \frac{S}{R^2}$. Finally, $\frac{1}{2}\frac{d}{dt}\left(\frac{1}{R^2S^2}\right) = \left(\frac{f\dot f}{p^4} + \frac{g\dot g}{q^4} + \frac{h\dot h}{r^4}\right) \left(\frac{\dot f^2}{p^2} + \frac{\dot g^2}{q^2} + \frac{\dot h^2}{r^2}\right) + \left(\frac{f^2}{p^4} + \frac{g^2}{q^4} + \frac{h^2}{r^4}\right)\left(\frac{\dot f\ddot f}{p^2} + \frac{\dot g\ddot g}{q^2} + \frac{\dot h\ddot h}{r^2}\right) = \frac{1}{R^2}\left(\frac{f\dot f}{p^4} + \frac{g\dot g}{q^4} + \frac{h\dot h}{r^4}\right) + \frac{\lambda}{S^2}\left(\frac{f\dot f}{p^4} + \frac{g\dot g}{q^4} + \frac{h\dot h}{r^4}\right) = 0$, since $\lambda = -S^2/R^2$. Hence, $RS$ is constant.

9.1.5 Suppose that a geodesic $\boldsymbol\gamma$ lies in the plane $\mathbf{v}\cdot\mathbf{a} = b$, where $\mathbf{a}$ and $b$ are constants. Then $\dot{\boldsymbol\gamma}\cdot\mathbf{a} = \ddot{\boldsymbol\gamma}\cdot\mathbf{a} = 0$. Since $\ddot{\boldsymbol\gamma}$ is parallel to $\mathbf{N}$ (the unit normal of the surface), $\mathbf{N}\cdot\mathbf{a} = 0$, so $\dot{\mathbf{N}}\cdot\mathbf{a} = 0$. Since $\mathbf{N}$, $\dot{\boldsymbol\gamma}$ and $\dot{\mathbf{N}}$ are all parallel to the plane and the last two vectors are perpendicular to the first, they are parallel. Hence $\boldsymbol\gamma$ is a line of curvature by Exercise 8.2.2. Conversely, if $\boldsymbol\gamma$ is both a geodesic and a line of curvature, we may assume $\boldsymbol\gamma$ has unit-speed (for the unit-speed reparametrization of $\boldsymbol\gamma$ would still be a geodesic and a line of curvature). Let $\mathbf{a} = \mathbf{N}\times\dot{\boldsymbol\gamma}$. Then $\dot{\mathbf{a}} = \dot{\mathbf{N}}\times\dot{\boldsymbol\gamma} + \mathbf{N}\times\ddot{\boldsymbol\gamma} = 0$ since the first term vanishes because $\boldsymbol\gamma$ is a line of curvature and the second because $\boldsymbol\gamma$ is a geodesic. So $\mathbf{a}$ is constant. And $\dot{\boldsymbol\gamma}\cdot\mathbf{a} = 0$ so $\boldsymbol\gamma\cdot\mathbf{a}$ is a constant, say $b$. Hence $\boldsymbol\gamma$ lies in the plane $\mathbf{v}\cdot\mathbf{a} = b$.

9.1.6 For (i) note that $\ddot{\boldsymbol\gamma}$ is a non-zero vector parallel to both $\mathbf{N}_1$ and $\mathbf{N}_2$, so $\mathbf{N}_1$ and $\mathbf{N}_2$ must be parallel. For an example, take $\mathcal{S}_1$ and $\mathcal{S}_2$ to be the sphere and cylinder in Theorem 6.4.6. Now suppose that $\mathcal{S}_1$ and $\mathcal{S}_2$ intersect perpendicularly. Then, $\mathbf{N}_1, \mathbf{N}_2$ and $\dot{\boldsymbol\gamma}$ are perpendicular unit vectors. From $\dot{\boldsymbol\gamma}\cdot\mathbf{N}_2 = 0$ we get $\ddot{\boldsymbol\gamma}\cdot\mathbf{N}_2 + \dot{\boldsymbol\gamma}\cdot\dot{\mathbf{N}}_2 = 0$. If $\boldsymbol\gamma$ is a geodesic on $\mathcal{S}_1$, $\ddot{\boldsymbol\gamma}\cdot\mathbf{N}_2 = 0$ since $\ddot{\boldsymbol\gamma}$ is parallel to $\mathbf{N}_1$, so $\dot{\mathbf{N}}_2$ is perpendicular to $\dot{\boldsymbol\gamma}$. Since $\dot{\mathbf{N}}_2$ is also perpendicular to $\mathbf{N}_2$, it must be parallel to $\mathbf{N}_1$. Conversely, if $\dot{\mathbf{N}}_2$ is parallel to $\mathbf{N}_1$, then $\dot{\boldsymbol\gamma}\cdot\dot{\mathbf{N}}_2 = 0$ so $\ddot{\boldsymbol\gamma}$ is perpendicular to $\mathbf{N}_2$. Since $\ddot{\boldsymbol\gamma}$ is also perpendicular to $\dot{\boldsymbol\gamma}$, it must be parallel to $\mathbf{N}_1$. Finally, if $\boldsymbol\gamma$ is a geodesic on both $\mathcal{S}_1$ and $\mathcal{S}_2$, then $\dot{\mathbf{N}}_1$ is parallel to $\mathbf{N}_2$ and $\dot{\mathbf{N}}_2$ is parallel to $\mathbf{N}_1$. It follows that $(\mathbf{N}_1 \times \mathbf{N}_2)^{\cdot} = \dot{\mathbf{N}}_1 \times \mathbf{N}_2 + \mathbf{N}_1 \times \dot{\mathbf{N}}_2 = \mathbf{0}$ so $\mathbf{N}_1 \times \mathbf{N}_2$ is a constant vector. Since $\dot{\boldsymbol{\gamma}}$ is a unit vector parallel to $\mathbf{N}_1 \times \mathbf{N}_2$, $\dot{\boldsymbol{\gamma}}$ is constant, so $\mathcal{C}$ is part of a straight line.

9.2.1 If $\mathbf{p}$ and $\mathbf{q}$ lie on the same parallel of the cylinder, there are exactly two geodesics joining them, namely the two circular arcs of the parallel of which $\mathbf{p}$ and $\mathbf{q}$ are the endpoints. If $\mathbf{p}$ and $\mathbf{q}$ are not on the same parallel, there are infinitely many circular helices joining $\mathbf{p}$ and $\mathbf{q}$ (see Example 9.2.8).

9.2.2 Take the cone to be $\boldsymbol{\sigma}(u,v) = (u\cos v, u\sin v, u)$. By Exercise 6.2.1, $\boldsymbol{\sigma}$ is locally isometric to an open subset of the $xy$-plane by $\boldsymbol{\sigma}(u,v) \mapsto (u\sqrt{2}\cos\frac{v}{\sqrt{2}}, u\sqrt{2}\sin\frac{v}{\sqrt{2}}, 0)$. By Corollary 9.2.7, the geodesics on the cone correspond to the straight lines in the $xy$-plane. Any such line, other than the axes $x = 0$ and $y = 0$, has equation $ax + by = 1$, where $a, b$ are constants; this line corresponds to the curve $v \mapsto \left(\frac{\cos v}{\sqrt{2}(a\cos\frac{v}{\sqrt{2}}+b\sin\frac{v}{\sqrt{2}})}, \frac{\sin v}{\sqrt{2}(a\cos\frac{v}{\sqrt{2}}+b\sin\frac{v}{\sqrt{2}})}, \frac{1}{\sqrt{2}(a\cos\frac{v}{\sqrt{2}}+b\sin\frac{v}{\sqrt{2}})}\right)$; the $x$- and $y$-axes correspond to straight lines on the cone.

9.2.3 Parametrize the cylinder by $\boldsymbol{\sigma}(u,v) = (\cos u, \sin u, v)$. Then, $E = G = 1$, $F = 0$, so the geodesic equations are $\ddot{u} = \ddot{v} = 0$. Hence, $u = a + bt, v = c + dt$, where $a, b, c$, and $d$ are constants. If $b = 0$ this is a straight line on the cylinder; otherwise, it is a circular helix.

9.2.4 For the first part,

$$\begin{aligned}
(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{\cdot} &= (E_u\dot{u} + E_v\dot{v})\,\dot{u}^2 + 2(F_u\dot{u} + F_v\dot{v})\dot{u}\dot{v} \\
&\quad +(G_u\dot{u} + G_v\dot{v})\dot{v}^2 + 2E\dot{u}\ddot{u} + 2F(\dot{u}\ddot{v} + \ddot{u}\dot{v}) + 2G\dot{v}\ddot{v} \\
&= \dot{u}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2) + \dot{v}(E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2) \\
&\quad +2E\dot{u}\ddot{u} + 2F(\dot{u}\ddot{v} + \ddot{u}\dot{v}) + 2G\dot{v}\ddot{v} \\
&= 2(E_u\dot{u} + F\dot{v})^{\cdot}\,\dot{u} + 2(F\dot{u} + G\dot{v})^{\cdot}\,\dot{v} + 2(E\dot{u} + F\dot{v})\ddot{u} + 2(F\dot{u} + G\dot{v})\ddot{v} \\
&\quad \text{(by the geodesic equations)} \\
&= 2[(E\dot{u} + F\dot{v})\dot{u}]^{\cdot} + 2[(F\dot{u} + G\dot{v})\dot{v}]^{\cdot} = 2(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{\cdot}.
\end{aligned}$$

Hence, $(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{\cdot} = 0$ and so $\parallel \dot{\boldsymbol{\gamma}} \parallel^2 = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2$ is constant.

Suppose now that (i) and (ii) hold. Differentiating $E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2 =$ constant gives $(E_u\dot{u} + E_v\dot{v})\dot{u}^2 + 2(F_u\dot{u} + F_v\dot{v})\dot{u}\dot{v} + (G_u\dot{u} + G_v\dot{v})\dot{v}^2 = -2(E\dot{u} + F\dot{v})\ddot{u} - 2(F\dot{u} + G\dot{v})\ddot{v}$, i.e., $(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2)\dot{u} + (E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2)\dot{v} = -2(E\dot{u} + F\dot{v})\ddot{u} - 2(F\dot{u} + G\dot{v})\ddot{v}$. Using (ii) we get $2\dot{u}\frac{d}{dt}(E\dot{u} + F\dot{v}) + 2(E\dot{u} + F\dot{v})\ddot{v} = -2(F\dot{u} + G\dot{v})\ddot{v} - (E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2)\dot{v}$. The left-hand side of this equation equals $2\frac{d}{dt}(\dot{u}(E\dot{u} + F\dot{v})) =$

$-2\frac{d}{dt}(F\dot{u}\dot{v}+G\dot{v}^2) = -2(F\dot{u}+G\dot{v})\ddot{v} - 2\dot{v}\frac{d}{dt}(F\dot{u}+G\dot{v})$. Combining the last two equations gives (iii) provided $\dot{v} \neq 0$.

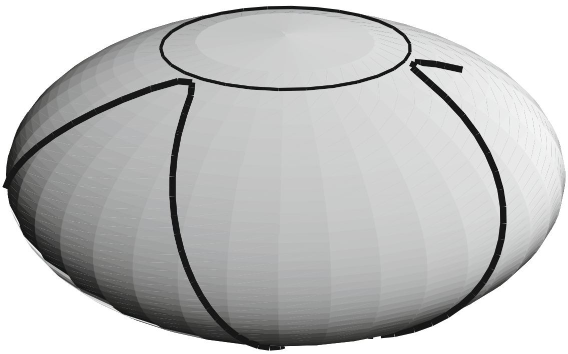
9.2.5 $E = 1, F = 0, G = 1+u^2$, so $\boldsymbol{\gamma}$ is unit-speed $\iff \dot{u}^2+(1+u^2)\dot{v}^2 = 1$. The second equation in (9.2) gives $\frac{d}{dt}((1+u^2)\dot{v}) = 0$, i.e., $\dot{v} = \frac{a}{1+u^2}$, where $a$ is a constant. So $\dot{u}^2 = 1 - \frac{a^2}{1+u^2}$ and, along the geodesic, $\frac{dv}{du} = \frac{\dot{v}}{\dot{u}} = \pm\frac{a}{\sqrt{(1-a^2+u^2)(1+u^2)}}$. If $a = 0$, then $v =$ constant and we have a ruling. If $a = 1$, then $dv/du = \pm 1/u\sqrt{1+u^2}$, which can be integrated to give $v = v_0 \mp \sinh^{-1}\frac{1}{u}$, where $v_0$ is a constant.

For the last part, note that $\left(\frac{du}{dv}\right)^2 = \frac{(u^2+1)(u^2+1-a^2)}{a^2}$ so (i) if $a^2 > 1$ then $du/dv = 0$ for $u = \pm\sqrt{a^2-1}$ and this is the minimum distance of the geodesic from the $z$-axis; (ii) if $a^2 < 1$ then $|du/dv| > a^{-2} - 1$ so $u$ will decrease to zero and the geodesic will cross the $z$-axis; (iii) if $a^2 = 1$ then $du/dv = \pm(u^2+1)$ so $u = \pm\tan(v+c)$ where $c$ is a constant. The information given implies that, when $u = D$, $\cos\alpha = \dot{\boldsymbol{\gamma}}\cdot\boldsymbol{\sigma}_u = \dot{u}$ (since $E = 1, F = 0$) so $a^2 = (1+D^2)\sin^2\alpha$. Then, $a^2$ is $> 1$, $< 1$ or $= 1$ according as $D$ is $>$, $<$ or $= \cot\alpha$.

9.2.6 This is straightforward algebra.

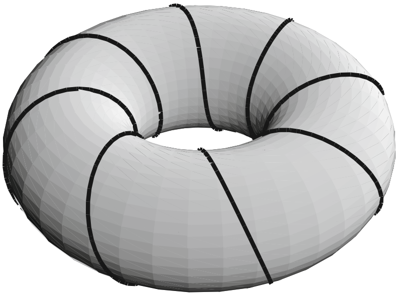
9.3.1 They are normal sections.

9.3.2 (i) Let the spheroid be obtained by rotating the ellipse $\frac{x^2}{p^2}+\frac{z^2}{q^2} = 1$ around the $z$-axis, where $p, q > 0$. Then, $p$ is the maximum distance of a point of the spheroid from the $z$-axis, so the angular momentum $\Omega$ of a geodesic must be $\leq p$ (we can assume that $\Omega \geq 0$). If $\Omega = 0$, the geodesic is a meridian. If $0 < \Omega < a$, the geodesic is confined to the annular region on the spheroid contained between the circles $z = \pm q\sqrt{1-\frac{\Omega^2}{p^2}}$, and the discussion in Example 9.3.3 shows that the geodesic 'bounces' between these two circles (see diagram below).

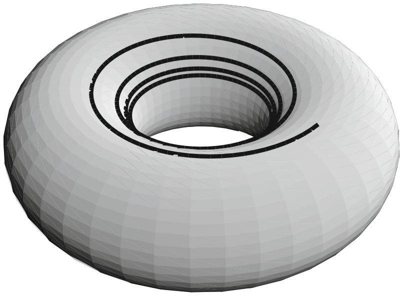


If $\Omega = p$, Eq. 9.10 shows that the geodesic must be the parallel $z = 0$.

(ii) Let the torus be as in Exercise 4.2.5. If $\Omega = 0$, the geodesic is a meridian (a circle). If $0 < \Omega < a - b$, the geodesic spirals around the torus. If $\Omega = a - b$, the geodesic is either the parallel of radius $a - b$ or spirals around the torus approaching this parallel asymptotically (but never crossing it):

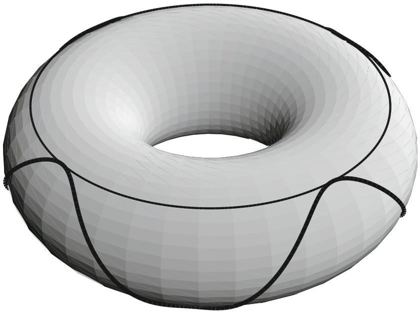


$0 < \Omega < a - b$



$\Omega = a - b$

If $a - b < \Omega < a + b$, the geodesic is confined to the annular region consisting of the part of the torus a distance $\geq \Omega$ from the axis, and bounces between the two parallels which bound this region:



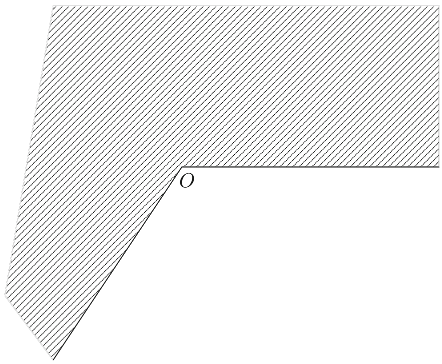
If $\Omega = a + b$, the geodesic must be the parallel of radius $a + b$.

9.3.3 The two solutions of Eq. 9.14 are $v = v_0 \pm \sqrt{\frac{1}{\Omega^2} - w^2}$, so the condition for a self-intersection is that, for some $w > 1$, $2\sqrt{\frac{1}{\Omega^2} - w^2} = 2k\pi$ for some integer $k > 0$. This holds $\Longleftrightarrow 2\sqrt{\frac{1}{\Omega^2} - 1} > 2\pi$, i.e., $\Omega < (1 + \pi^2)^{-1/2}$. In this case, there are $k$ self-intersections, where $k$ is the largest integer such that $2k\pi < 2\sqrt{\frac{1}{\Omega^2} - 1}$.

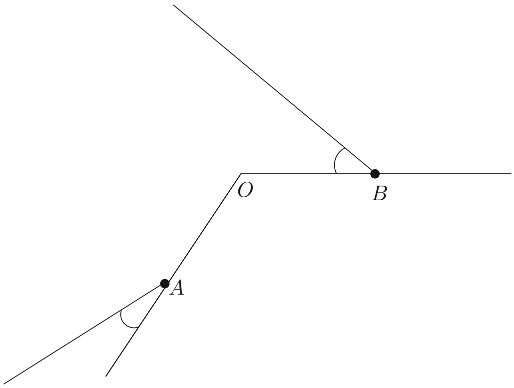
9.3.4 From the solution to Exercise 8.3.1, $Z = V + iW = \frac{z-i}{z+i}$, where $z = v + iw$. This is a Möbius transformation, so it takes lines and circles to lines and circles and preserves angles (Appendix 2). Since the geodesics on the pseudosphere correspond to straight lines and circles in the $vw$-plane perpendicular to the $v$-axis, they correspond in the $VW$-plane to straight lines and circles perpendicular to the image of the $V$-axis under the transformation $z \mapsto \frac{z-i}{z+i}$, i.e., the unit circle $V^2 + W^2 = 1$.

A straight line $a\bar{V} + b\bar{W} = c$ in the $\bar{V}\bar{W}$-plane (where $a, b, c$ are constants) corresponds to the curve $2aV + 2bW = c(V^2 + W^2 + 1)$ in the $VW$-plane. If $c = 0$, this is a straight line through the origin, which corresponds to a geodesic on the pseudosphere by the first part. If $c \neq 0$ it is the equation of a circle with centre $(a/c, b/c)$ and squared radius $(a^2 + b^2 - c^2)/c^2$. This circle intersects the boundary circle $V^2 + W^2 = 1$ orthogonally because the square of the distance between the centres of the two circles is equal to the sum of the squares of their radii. Hence this circle also corresponds to a geodesic on the pseudosphere. This proves that every straight line in the $\bar{V}\bar{W}$-plane corresponds to a geodesic on the pseudosphere. That every geodesic on the pseudosphere arises from a straight line in the $\bar{V}\bar{W}$-plane in this way can be proved by similar arguments, or by noting that there is a straight line in the $\bar{V}\bar{W}$-plane passing through any point of the disc $\bar{V}^2 + \bar{W}^2 < 1$ in any direction and using Proposition 9.2.4.

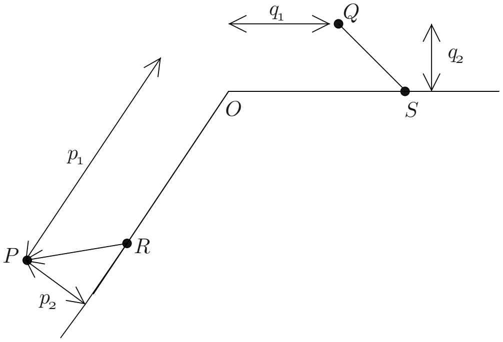
9.4.1 From Exercise 6.2.1, the cone is isometric to the 'sector' $\mathcal{S}$ of the plane with vertex at the origin and angle $\pi\sqrt{2}$:




Geodesics on the cone correspond to possibly broken line segments in $\mathcal{S}$: if a line segment meets the boundary of $\mathcal{S}$ at a point $A$, say, it may continue from the point $B$ on the other boundary line at the same distance as $A$ from the origin and with the indicated angles being equal.

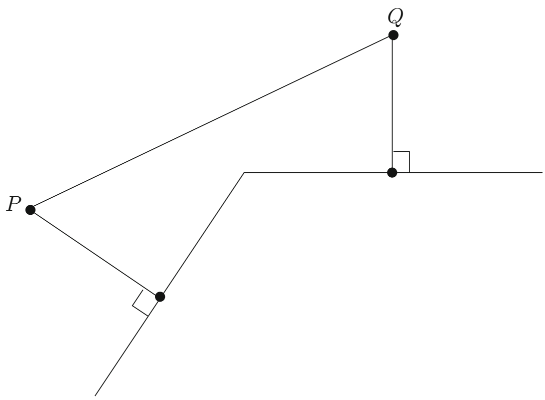



(i) TRUE: if two points $P$ and $Q$ can be joined by a line segment in $\mathcal{S}$ there is no problem; otherwise, $P$ and $Q$ can be joined by a broken line segment satisfying the conditions above:




To see that this is always possible, let $p_1$, $p_2$, $q_1$ and $q_2$ be the indicated distances, and let $R$ and $S$ be the points on the boundary of the sector at a distance $(p_2q_1 + p_1q_2)/(p_2 + q_2)$ from the origin. Then, the broken line segment joining $P$ and $R$ followed by that joining $S$ and $Q$ is the desired geodesic.

(ii) FALSE:



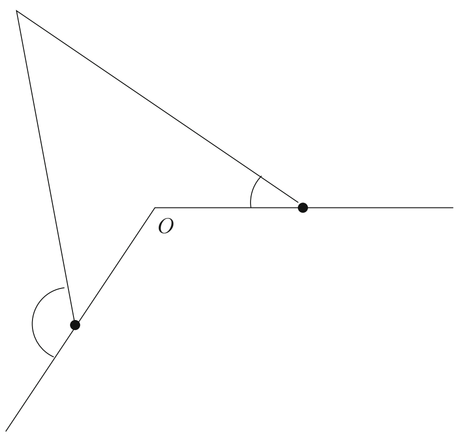

(iii) FALSE: many meet in two points, such as the two geodesics joining $P$ and $Q$ in the diagram in (ii).

(iv) TRUE: the meridians do not intersect (remember that the vertex of the cone has been removed), and parallel straight lines that are entirely contained in $\mathcal{S}$ do not intersect.

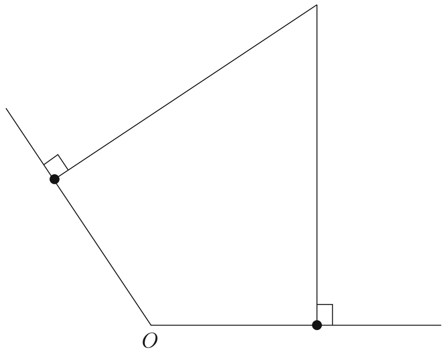
(v) TRUE: since (broken) line segments in $\mathcal{S}$ can clearly be continued indefinitely in both directions.

(vi) FALSE: if $A$ and $B$ are points on the boundary of the sector at the same distance from $O$ (see the diagram at the top of the previous page) and if $C$ is a point in the sector such that the straight line segments $AC$ and $BC$ are in the sector, then $AC$ and $BC$ are geodesics joining the same two pints of the cone but they have different length unless $C$ lies on the bisector of the angle of the sector.

(vii ) TRUE: a situation of the form



in which the indicated angles are equal is clearly impossible. But the answer to this part of the question depends on the angle of the cone: if the angle is $\alpha$, instead of $\pi/4$, lines can self-intersect if $\alpha < \pi/6$, for then the corresponding sector in the plane has angle $< \pi$.




9.4.2 We consider the intersection of $S^2$ with the plane passing through $\mathbf{p}$ and $\mathbf{q}$ and making an angle $\theta$ with the $xy$-plane, where $-\pi/2 < \theta < \pi/2$. This intersection is a circle $\mathcal{C}_\theta$ but it is not a great circle unless $\theta = 0$. Hence, if $\theta \neq 0$, the short segment of $\mathcal{C}_\theta$ joining $\mathbf{p}$ and $\mathbf{q}$ is not a geodesic and so has length $> \pi/2$ (the length of the shortest geodesic joining $\mathbf{p}$ and $\mathbf{q}$). Since the length of $\mathcal{C}_\theta$ is $\leq 2\pi$, the length of the long segment of $\mathcal{C}_\theta$ joining $\mathbf{p}$ and $\mathbf{q}$ has length $< 3\pi/2$ if $\theta \neq 0$, i.e., strictly less than the length of the long segment of the geodesic $\mathcal{C}_0$ joining $\mathbf{p}$ and $\mathbf{q}$. So the long geodesic segment is not a local minimum of the length of curves joining $\mathbf{p}$ and $\mathbf{q}$.

9.4.3 (i) This is obvious if $n \geq 0$ since $e^{-1/t^2} \to 0$ as $t \to 0$. We prove that $t^{-n}e^{-1/t^2} \to 0$ as $t \to 0$ by induction on $n \geq 0$. We know the result if $n = 0$, and if $n > 0$ we can apply L'Hopital's rule: $\lim_{t\to 0} \frac{t^{-n}}{e^{1/t^2}} = \lim_{t\to 0} \frac{nt^{-n-1}}{\frac{2}{t^3}e^{1/t^2}} = \lim_{t\to 0} \frac{n}{2}\frac{t^{-(n-2)}}{e^{1/t^2}}$, which vanishes by the induction hypothesis.

(ii) We prove by induction on $n$ that $\theta$ is $n$-times differentiable with

$$\frac{d^n\theta}{dt^n} = \begin{cases} \frac{P_n(t)}{t^{3n}}e^{-1/t^2} & \text{if } t \neq 0, \\ 0 & \text{if } t = 0, \end{cases}$$

where $P_n$ is a polynomial in $t$. For $n = 0$, the assertion holds with $P_0 = 1$. Assuming the result for some $n \geq 0$, $\frac{d^{n+1}\theta}{dt^{n+1}} = \left(\frac{-3nP_n}{t^{3n+1}} + \frac{P_n'}{t^{3n}} + \frac{2P_n}{t^{3n+3}}\right)e^{-1/t^2}$ if $t \neq 0$, so we take $P_{n+1} = (2 - 3nt^2)P_n + t^3P_n'$. If $t = 0$, $\frac{d^{n+1}\theta}{dt^{n+1}} = \lim_{t\to 0} \frac{P_n(t)}{t^{3n+1}}e^{-1/t^2} = P_n(0)\lim_{t\to 0} \frac{e^{-1/t^2}}{t^{3n+1}} = 0$ by part (i). Parts (iii) and (iv) are obvious.

9.5.1 Since $\boldsymbol{\gamma}^\theta$ is unit-speed, $\boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_r = 1$, so $\int_0^R \boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_r \, dr = R$. Differentiating with respect to $\theta$ gives $\int_0^R \boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_{r\theta} \, dr = 0$, and then integrating by parts gives

$$\boldsymbol{\sigma}_\theta \cdot \boldsymbol{\sigma}_r \Big|_{r=0}^{r=R} - \int_0^R \boldsymbol{\sigma}_\theta \cdot \boldsymbol{\sigma}_{rr} \, dr = 0.$$

Now $\boldsymbol{\sigma}(0, \theta) = \mathbf{p}$ for all $\theta$, so $\boldsymbol{\sigma}_\theta = \mathbf{0}$ when $r = 0$. So we must show that the integral in the last equation vanishes. But, $\boldsymbol{\sigma}_{rr} = \ddot{\gamma}^\theta$, the dot denoting the derivative with respect to the parameter $r$ of the geodesic $\boldsymbol{\gamma}^\theta$, so $\boldsymbol{\sigma}_{rr}$ is parallel to the unit normal $\mathbf{N}$ of $\boldsymbol{\sigma}$; since $\boldsymbol{\sigma}_\theta \cdot \mathbf{N} = 0$, it follows that $\boldsymbol{\sigma}_\theta \cdot \boldsymbol{\sigma}_{rr} = 0$. The first fundamental form is as indicated since $\boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_r = 1$ and $\boldsymbol{\sigma}_r \cdot \boldsymbol{\sigma}_\theta = 0$.

9.5.2 (i) The length of the part of $\boldsymbol{\gamma}$ between $\mathbf{p}$ and $\mathbf{q}$ is $\int_0^1 \sqrt{\dot{f}^2 + G\dot{g}^2} dt \geq \int_0^1 \sqrt{\dot{f}^2} dt = f(1) - f(0) = R$. (ii) Use the hint, noting that the length of the part of $\boldsymbol{\gamma}$ between $\mathbf{p}$ and $\mathbf{q}'$ is $\geq R$. (iii) If the part of $\boldsymbol{\gamma}$ between $\mathbf{p}$ and $\mathbf{q}$ has length $R$, then it must stay inside the geodesic circle with centre $\mathbf{p}$ and radius $R$ by (ii), and then we must have $\int_0^1 \sqrt{\dot{f}^2 + G\dot{g}^2} dt = \int_0^1 \sqrt{\dot{f}^2} dt$. Then $G\dot{g} = 0$ for all $t \in (0, 1)$, so $\dot{g} = 0$ (as $G > 0$) and so $g$ is a constant which must be $\alpha$ as $\boldsymbol{\gamma}$ passes through $\mathbf{q}$. This means that $\boldsymbol{\gamma}$ is a parametrization of the radial line $\theta = \alpha$.

# Chapter 10

10.1.1 The matrix of the Weingarten map with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is $\mathcal{F}_I^{-1}\mathcal{F}_{II} = = \begin{pmatrix} \cos^2 v & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -\cos^2 v & 0 \\ 0 & -1 \end{pmatrix} = -I$, so $\mathbf{N}_u = \boldsymbol{\sigma}_u$, $\mathbf{N}_v = \boldsymbol{\sigma}_v$. Thus, $\mathbf{N} = \boldsymbol{\sigma} - \mathbf{a}$, where $\mathbf{a}$ is a constant vector. Hence, $\parallel \boldsymbol{\sigma} - \mathbf{a} \parallel = 1$, showing that the surface is an open subset of the sphere $\mathcal{S}$ of radius 1 and centre $\mathbf{a}$. The standard latitude-longitude parametrization $\boldsymbol{\sigma}(u, v)$ of $S^2$ has first and second fundamental forms both given by $du^2 + \cos^2 u dv^2$, so the parametrization $\boldsymbol{\sigma}(v, u) + \mathbf{a}$ of $\mathcal{S}$ has the given first and second fundamental forms (the second fundamental form changes sign because $\boldsymbol{\sigma}_v \times \boldsymbol{\sigma}_u = -\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$).

10.1.2 $\Gamma_{22}^1 = \sin u \cos u$ and the other Christoffel symbols are zero; the second Codazzi–Mainardi equation is not satisfied.

10.1.3 The Christoffel symbols are $\Gamma_{11}^1 = 0$, $\Gamma_{11}^2 = 1/w$, $\Gamma_{12}^1 = -1/w$, $\Gamma_{12}^2 = 0$, $\Gamma_{22}^1 = 0$, $\Gamma_{22}^2 = -1/w$. Using the first equation in Proposition 10.1.2 we get $K = -1$. The Codazzi-Mainardi equations are $L_w = -(L + N)/w$,

$N_v = 0$. Hence, $N$ depends only on $w$, and since $-1 = K = LN/EG$, we have $LN = -1/w^4$ so $L$ also depends only on $w$; the first Codazzi-Mainardi equation gives $dL/dw = -L/w + 1/Lw^5$, which is the stated differential equation. Putting $P = Lw^2$ we get $dP/dw = (1+P^2)/wP$ which integrates to give $1 + P^2 = Cw^2$, where $C > 0$ is a constant, i.e., $L = \pm\sqrt{Cw^2 - 1}/w^2$. Hence, the second fundamental form is only defined for $w \geq C^{-1/2}$ or $w \leq -C^{-1/2}$.

The first fundamental form in this exercise is the same as that of a suitable parametrization of the pseudosphere (see Exercise 8.3.1(i)). We saw that the pseudosphere corresponds to (part of) the region $w > 1$.

10.1.4 The Christoffel symbols are $\Gamma^1_{11} = E_u/2E$, $\Gamma^2_{11} = -E_v/2G$, $\Gamma^1_{12} = E_v/2E$, $\Gamma^2_{12} = G_u/2G$, $\Gamma^1_{22} = -G_u/2E$, $\Gamma^2_{22} = G_v/2G$. The first Codazzi–Mainardi equation is

$$L_v = \frac{LE_v}{2E} - N\left(\frac{-E_v}{2G}\right) = \frac{1}{2}E_v\left(\frac{L}{E} + \frac{N}{G}\right),$$

and similarly for the other equation. Finally,

$$(\kappa_1)_v = \frac{E_v}{2E}\left(\frac{L}{E} + \frac{N}{G}\right) - \frac{LE_v}{E^2} = \frac{E_v}{2E}\left(\frac{N}{G} - \frac{L}{E}\right) = \frac{E_v}{2E}(\kappa_2 - \kappa_1),$$

and similarly for $(\kappa_2)_u$.

10.2.1 By Corollary 10.2.3(i), $K = -\frac{1}{2e^\lambda}\left(\frac{\partial}{\partial u}\left(\frac{(e^\lambda)_u}{e^\lambda}\right) + \frac{\partial}{\partial v}\left(\frac{(e^\lambda)_v}{e^\lambda}\right)\right) = -\frac{1}{2e^\lambda}(\lambda_{uu} + \lambda_{vv})$.

10.2.2 By Exercise 6.1.4, $\begin{pmatrix} \tilde{E} & \tilde{F} \\ \tilde{F} & \tilde{G} \end{pmatrix} = J^t \begin{pmatrix} E & F \\ F & G \end{pmatrix} J$, where $J = \begin{pmatrix} \frac{\partial r}{\partial u} & \frac{\partial r}{\partial v} \\ \frac{\partial \theta}{\partial u} & \frac{\partial \theta}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{u}{r} & \frac{v}{r} \\ \frac{-v}{r^2} & \frac{u}{r^2} \end{pmatrix}$. By Exercise 9.5.1, $E = 1, F = 0$, and we get the stated formulas for $\tilde{E}, \tilde{F}, \tilde{G}$. From $\tilde{E} - 1 = \frac{v^2}{r^2}\left(\frac{G}{r^2} - 1\right)$, $\tilde{G} - 1 = \frac{u^2}{r^2}\left(\frac{G}{r^2} - 1\right)$, we get $u^2(\tilde{E} - 1) = v^2(\tilde{G} - 1)$. Since $\tilde{E}$ and $\tilde{G}$ are smooth functions of $(u, v)$, they have Taylor expansions $\tilde{E} = \sum_{i+j\leq 2} e_{ij}u^iv^j + o(r^2)$, $\tilde{G} = \sum_{i+j\leq 2} g_{ij}u^iv^j + o(r^2)$, where $o(r^k)$ denotes terms such that $o(r^k)/r^k \to 0$ as $r \to 0$. Equating coefficients on both sides of $u^2(\tilde{E}-1) = v^2(\tilde{G}-1)$ shows that all the $e$'s and $g$'s are zero except $e_{02} = g_{20} = k$, say. Then, $\tilde{E} = 1 + kv^2 + o(r^2)$, which implies that $G = r^2 + kr^4 + o(r^4)$. By Corollary 10.2.3(ii), $K = -\frac{1}{\sqrt{G}}\frac{\partial^2\sqrt{G}}{\partial r^2}$. From the first part, $\sqrt{G} = r + \frac{1}{2}kr^3 + o(r^3)$, hence $K = -3k + o(1)$. Taking $r = 0$ gives $K(P) = -3k$.

10.2.3 (i) $C_R = \int_0^{2\pi} \parallel \boldsymbol{\sigma}_\theta \parallel\, d\theta = \int_0^{2\pi} \sqrt{G}\, d\theta = \int_0^{2\pi} \left(R - \frac{1}{6}K(P)R^3 + o(R^3)\right) d\theta = 2\pi\left(R - \frac{1}{6}K(P)R^3 + o(R^3)\right)$. (ii) Since $d\mathcal{A}_{\boldsymbol{\sigma}} = \sqrt{G}\, dr d\theta$, the area

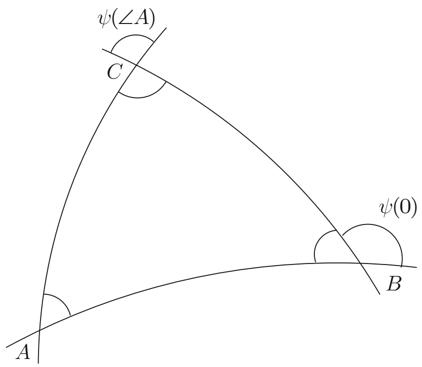
$A_R = \int_0^R \int_0^{2\pi} \sqrt{G}\, drd\theta$ is equal to $2\pi \int_0^R \left(r - \frac{1}{6}K(P)r^3 + o(r^3)\right) dr = \pi R^2 \left(1 - \frac{K(P)}{12}R^2 + o(R^2)\right)$. If $\mathcal{S} = S^2$, Exercise 6.5.3 gives $C_R = 2\pi \sin R = 2\pi(R = \frac{1}{6}R^3 + o(R^3))$, $A_R = 2\pi(1 - \cos R) = 2\pi(\frac{1}{2}R^2 - \frac{1}{24}R^4 + o(R^4))$. Since $K = 1$ these formulas agree with those in (i) and (ii).

10.2.4 (i) Let $s$ be the arc-length of $\boldsymbol{\gamma}$, so that $ds/d\theta = \lambda$, and denote $d/ds$ by a dot. The first of the geodesic equations (9.2) applied to $\boldsymbol{\gamma}$ gives $\ddot{r} = \frac{1}{2}G_r\dot{\theta}^2$. Since $r = f(\theta)$, this gives $\frac{1}{\lambda}\left(\frac{1}{\lambda}f'\right)' = \frac{1}{2\lambda^2}G_r$. This simplifies to give the stated equation. (ii) Since $\boldsymbol{\sigma}_r$ and $\dot{\boldsymbol{\gamma}}$ are unit vectors, $\cos\psi = \boldsymbol{\sigma}_r \cdot \boldsymbol{\gamma} = \frac{1}{\lambda}\boldsymbol{\sigma}_r \cdot (f'\boldsymbol{\sigma}_r + \boldsymbol{\sigma}_\theta) = f'/\lambda$. Also, $\boldsymbol{\sigma}_r \times \dot{\boldsymbol{\gamma}} = \frac{1}{\lambda}(\boldsymbol{\sigma}_r \times \boldsymbol{\sigma}_\theta) = \frac{\sqrt{G}}{\lambda}\mathbf{N}$, so $\sin\psi = \sqrt{G}/\lambda$. Hence, $\left(\frac{f'}{\lambda}\right)' = -\psi'\sin\psi = -\frac{\sqrt{G}}{\lambda}\psi'$, and so $\psi' = -\frac{1}{\sqrt{G}}\left(f'' - \frac{f'\lambda'}{\lambda}\right) = -\frac{1}{2\sqrt{G}}\frac{\partial G}{\partial r} = -\frac{\partial\sqrt{G}}{\partial r}$.

(iii) Using the formula for $K$ in Corollary 10.2.3(ii) and the expression for the first fundamental form of $\boldsymbol{\sigma}$ in Exercise 9.5.1, we get

$$\begin{aligned}\int_{\mathcal{T}} K d\mathcal{A}_{\boldsymbol{\sigma}} &= \int_0^\alpha \int_0^{f(\theta)} -\frac{1}{\sqrt{G}}\frac{\partial^2\sqrt{G}}{\partial r^2}\sqrt{G}\, drd\theta \\ &= -\int_0^\alpha \left.\frac{\partial\sqrt{G}}{\partial r}\right|_{r=0}^{r=f(\theta)} d\theta = \int_0^\alpha \left(\psi' + \left.\frac{\partial\sqrt{G}}{\partial r}\right|_{r=0}\right) d\theta.\end{aligned}$$

By Exercise 10.2.2, $\sqrt{G} = r + o(r)$ so $\partial\sqrt{G}/\partial r = 1$ at $r = 0$. Hence, $\int_{ABC} K d\mathcal{A}_{\boldsymbol{\sigma}} = \psi(\alpha) - \psi(0) + \alpha = \gamma - (\pi - \beta) + \alpha = \alpha + \beta + \gamma - \pi$.




10.2.5 With the notation of Example 4.5.3 we have, on the median circle $t = 0$, $\boldsymbol{\sigma}_t = \left(-\sin\frac{\theta}{2}\cos\theta, -\sin\frac{\theta}{2}\sin\theta, \cos\frac{\theta}{2}\right)$, $\boldsymbol{\sigma}_\theta = (-\sin\theta, \cos\theta, 0)$, hence $E = 1, F = 0, G = 1$ and $\mathbf{N} = \left(-\cos\frac{\theta}{2}\cos\theta, \sin\frac{\theta}{2}\sin\theta, -\sin\frac{\theta}{2}\right)$; $\boldsymbol{\sigma}_{tt} = \mathbf{0}$, $\boldsymbol{\sigma}_{t\theta} = \Big(-\frac{1}{2}\cos\frac{\theta}{2}\cos\theta + \sin\frac{\theta}{2}\sin\theta, -\frac{1}{2}\cos\frac{\theta}{2}\sin\theta - \sin\frac{\theta}{2}\cos\theta, -\frac{1}{2}\sin\frac{\theta}{2}\Big)$, giving $L = 0, M = \frac{1}{2}$. Hence, $K = (LN - M^2)/(EG - F^2) = -1/4$. Since $K \neq 0$, the Theorema Egregium implies that the Möbius band is not locally isometric to a plane.

10.2.6 The catenoid has first fundamental form $\cosh^2 u(du^2 + dv^2)$ and its Gaussian curvature is $K = -\text{sech}^4 u$ (Exercise 8.1.2). If $f$ is an isometry of the catenoid, let $f(\boldsymbol{\sigma}(u,v)) = \boldsymbol{\sigma}(\tilde{u},\tilde{v})$. By the Theorema Egregium, $\text{sech}^4 u = \text{sech}^4 \tilde{u}$, so $\tilde{u} = \pm u$; reflecting in the plane $z = 0$ changes $u$ to $-u$, so assume that the sign is $+$. Let $\tilde{v} = f(u,v)$; the first fundamental form of $\boldsymbol{\sigma}(u, f(u,v))$ is $(\cosh^2 u + f_u^2)du^2 + 2f_u f_v du dv + f_v^2 \cosh^2 u dv^2$; hence, $\cosh^2 u = \cosh^2 u + f_u^2$, $f_u f_v = 0$ and $f_v^2 \cosh^2 u = \cosh^2 u$. So $f_u = 0$, $f_v = \pm 1$ and $f = \pm v + \alpha$, where $\alpha$ is a constant. If the sign is $+$ we have a rotation by $\alpha$ about the $z$-axis; if the sign is $-$ we have a reflection in the plane containing the $z$-axis and making an angle $\alpha/2$ with the $xz$-plane.

10.3.1 Arguing as in the proof of Theorem 10.3.4, we suppose that $J$ attains its maximum value $> 0$ at some point $\mathbf{p} \in \mathcal{S}$ contained in a patch $\boldsymbol{\sigma}$ of $\mathcal{S}$. We can assume that the principal curvatures $\kappa_1$ and $\kappa_2$ of $\boldsymbol{\sigma}$ satisfy $\kappa_1 > \kappa_2 > 0$ everywhere. Since $H = \frac{1}{2}(\kappa_1 + \kappa_2)$, $\kappa_1 > H$ and $J = 4(\kappa_1 - H)^2$. Thus, $J$ increases with $\kappa_1$ when $\kappa_1 > H$, so $\kappa_1$ must have a maximum at $\mathbf{p}$, and then $\kappa_2 = 2H - \kappa_1$ has a minimum there. By Lemma 10.3.5, $K \leq 0$ at $\mathbf{p}$, contradicting the assumption that $K > 0$ everywhere.

10.3.2 We start with the parametrization $\boldsymbol{\sigma}(U,V) = (f(U)\cos V, f(U)\sin V, g(U))$, where $f(U) = e^U$, $g(U) = \int \sqrt{1-e^{2U}}dU$. The first and second fundamental forms are $dU^2 + e^{2U}dV^2$ and $\frac{-e^U}{\sqrt{1-e^{2U}}}dU^2 + e^U\sqrt{1-e^{2U}}dV^2$, respectively. In the notation of the proof of Proposition 10.3.2, $\kappa_1 = -1/e^U\sqrt{1-e^{2U}}$, $\kappa_2 = e^{-U}\sqrt{1-e^{2U}}$. So we are in case (ii) of the proof and $\tan\omega = \sqrt{e^{-2U}-1}$. We find that $e(U) = E/\sin^2\omega = \frac{1}{1-e^{2U}}$, $g(V) = G\sec^2\omega = 1$. So $\tilde{V} = V$ and $\tilde{U} = \int \frac{dU}{\sqrt{1-e^{2U}}} = -\cosh^{-1}(e^{-U}) - c$ for some constant $c$. So $U = -\ln(\cosh(\tilde{U} + c))$. Hence, $\theta = 2\omega = 2\tan^{-1}\sqrt{e^{-2U}-1} = 2\tan^{-1}\sqrt{\cosh^2(\tilde{U}+c)-1} = 2\tan^{-1}\sinh(\tilde{U}+c)$. Finally, $u = \frac{1}{2}(\tilde{U}+\tilde{V}), v = \frac{1}{2}(\tilde{V}-\tilde{U})$, so $\tilde{U} = u - v$.

10.4.1 Let $f : \mathcal{S} \to \tilde{\mathcal{S}}$ be a local diffeomorphism that takes unit-speed geodesics to unit-speed geodesics. Let $\mathbf{p} \in \mathcal{S}$ and let $\mathbf{0} \neq \mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$. There is a unique geodesic $\boldsymbol{\gamma}(t)$ on $\mathcal{S}$ such that $\boldsymbol{\gamma}(0) = \mathbf{p}$ and $\dot{\boldsymbol{\gamma}}(0) = \mathbf{v}/\parallel \mathbf{v} \parallel$. Then $\boldsymbol{\gamma}$ is unit-speed so $\tilde{\boldsymbol{\gamma}} = f \circ \boldsymbol{\gamma}$ is a unit-speed geodesic. In particular, $\dot{\tilde{\boldsymbol{\gamma}}} = D_{\mathbf{p}}f(\mathbf{v}/\parallel \mathbf{v} \parallel)$ is a unit vector, i.e., $\parallel D_{\mathbf{p}}f(\mathbf{v}) \parallel = \parallel \mathbf{v} \parallel$. This means that $D_{\mathbf{p}}f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\tilde{\mathcal{S}}$ is an isometry, so $f$ is a local isometry.

10.4.2 Local isometries take geodesics to geodesics by Corollary 9.2.7. If we apply a dilation $\mathbf{v} \mapsto a\mathbf{v}$, where $a \neq 0$ is a constant, to a surface, the first fundamental form gets multiplied by $a^2$ and so the Christoffel symbols are unchanged (see Proposition 7.4.4). By Proposition 9.2.3, the geodesic equations are unchanged. It follows that dilations take geodesics to geodesics. Hence, any composite of local isometries and dilations also takes geodesics to geodesics. The converse is false: the map from the $xy$-plane to itself given by $(x, y) \mapsto (x, 2y)$ takes geodesics to geodesics (as it takes straight lines to straight lines) but is not the composite of a dilation and a local isometry.

10.4.3 (i) This is true because $F$ is conformal. (ii) The parameter curve $u \mapsto \boldsymbol{\sigma}(u, v_0)$ is a geodesic on $\boldsymbol{\sigma}$ for any fixed $v_0$ by construction of the geodesic patch $\boldsymbol{\sigma}$. Since $F$ is a geodesic local diffeomorphism, $u \mapsto F(\boldsymbol{\sigma}(u, v_0))$ is a pre-geodesic on $\tilde{\boldsymbol{\sigma}}$. Hence, for some smooth function $u(t)$, $t \mapsto F(\boldsymbol{\sigma}(u(t), v_0))$ is a geodesic on $\tilde{\boldsymbol{\sigma}}$. The second geodesic equation in Theorem 9.2.1 gives $\lambda_v \dot{u}^2 = 0$, so $\lambda_v = 0$ and $\lambda$ is independent of $v$. (iii) Let $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$; we can assume that $\boldsymbol{\gamma}$ is unit-speed. Since the first fundamental form of $\boldsymbol{\sigma}$ is $du^2 + Gdv^2$, the parameter curves $v =$ constant and $u =$ constant intersect orthogonally and unit vectors parallel to them are $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v/\sqrt{G}$, respectively. If the oriented angle between $\dot{\boldsymbol{\gamma}}$ and the curve $v =$ constant is $\theta$, we have $\dot{\boldsymbol{\gamma}} = \cos\theta \boldsymbol{\sigma}_u + \sin\theta \boldsymbol{\sigma}_v/\sqrt{G} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v$. Hence, $\dot{u} = \cos\theta$ and $\dot{v} = \frac{\sin\theta}{\sqrt{G}}$. The first geodesic equation in Theorem 9.2.1 gives $\ddot{u} = \frac{1}{2}G_u\dot{v}^2$, i.e., $\dot{\theta}\sin\theta = \frac{1}{2}G_u\dot{v}^2$. Substituting for $\dot{v}$ gives $\frac{d\theta}{dv} = \frac{\dot{\theta}}{\dot{v}} = \frac{G_u\dot{v}}{2\sin\theta} = G_u/2\sqrt{G}$. (iv) Apply (iii) to $F \circ \boldsymbol{\gamma}$ and use the fact that $F$ is conformal. (v) Parts (iii) and (iv) imply $(\lambda G)_u = \lambda G_u$, hence $\lambda_u G = 0$, hence $\lambda_u = 0$, i.e., $\lambda$ is independent of $u$. By (ii), $\lambda$ is constant. (vi) If $D_{\lambda^{-1/2}}$ is the dilation by a factor $\lambda^{-1/2}$, the composite $D_{\lambda^{-1/2}} \circ F$ preserves the first fundamental form and so is a local isometry, say $\mathcal{G}$. Then, $F = D_{\lambda^{1/2}} \circ \mathcal{G}$.

# Chapter 11

11.1.1 Let $l$ meet the real axis at $b$ and suppose that $\mathfrak{Re}(a) > b$ (the case $\mathfrak{Re}(a) < b$ is similar). The semicircle with centre $d$ on the real axis and radius $|a - d|$ passes through $a$ and does not meet $l$ provided that $|a - d| \leq |d - b|$, i.e., provided $d \geq (|a|^2 - b^2)/(\mathfrak{Re}(a) - b)$.

11.1.2 Suppose that $a$ and $b$ lie on a half-line geodesic, say $a = r + is, b = r + it$, where $r, s, t \in \mathbb{R}$ and $t > s$. Then, $d_{\mathcal{H}}(a, b) = \int_s^t \frac{dw}{w} = \ln(t/s) = d$, say, so $t/s = e^d$. On the other hand, the formula in Proposition 11.1.4 gives $2\tanh^{-1}\left(\frac{t-s}{t+s}\right) = 2\tanh^{-1}\left(\frac{e^d-1}{e^d+1}\right) = 2\tanh^{-1}(\tanh\frac{d}{2}) = d$.

11.1.3 The required hyperbolic line cannot be a half-line, so must be a semicircle with centre the origin, and it must have radius $|a|$.

11.1.4 By Proposition 11.1.4, $z \in \mathcal{C}_{a,R} \iff 2\tanh^{-1}\left|\frac{z-a}{z-\bar{a}}\right| = R$. If $\lambda = \tanh(R/2)$, this is equivalent to $(1-\lambda^2)|z|^2 - (\bar{a}-\lambda^2 a)z - (a-\lambda^2\bar{a})\bar{z} + (1-\lambda^2)|a|^2 = 0$. According to Proposition A.2.3, this is the equation of a circle provided that $\lambda^2 < 1$, which is obvious, and $|\bar{a}-\lambda^2 a|^2 > (1-\lambda^2)^2|a|^2$. This condition reduces to $2|a|^2 > a^2+\bar{a}^2$. Writing $a = |a|e^{i\theta}$ this becomes $\cos 2\theta < 1$, which is true because $a \in \mathcal{H}$ implies $0 < \theta < \pi$.

$\mathcal{C}_{ic,R}$ will be a circle with centre on the imaginary axis, say at $ib$. Then $\mathcal{C}_{a,R}$ intersects the imaginary axis at the points $i(b \pm r)$, so these two points must be a hyperbolic distance $2R$ apart, i.e., $2R = 2\tanh^{-1}\left|\frac{2ir}{2ib}\right|$. This gives $r = b\tanh R$, which is equivalent to $R = \frac{1}{2}\ln\frac{b+r}{b-r}$. Next, the points $i(b\pm r)$ are the same hyperbolic distance from $ic$, so $\left|\frac{i(b+r)-ic}{i(b+r)+ic}\right| = \left|\frac{i(b-r)-ic}{i(b-r)+ic}\right|$. This gives $(b+r-c)(b-r+c) = (c+r-b)(c+r+b)$, which simplifies to $c^2 = b^2 - r^2$. Parametrizing $\mathcal{C}_{ic,R}$ by $v = r\cos\theta, w = b + r\sin\theta$, and denoting $d/d\theta$ by a dot, the circumference of $\mathcal{C}_{ic,R}$ is $\int_0^{2\pi}\frac{\sqrt{\dot{v}^2+\dot{w}^2}}{w^2}d\theta = \int_0^{2\pi}\frac{r}{b+r\sin\theta}d\theta = \frac{2\pi r}{\sqrt{b^2-r^2}} = 2\pi\sinh R$. The area inside $\mathcal{C}_{ic,R}$ is $\int_{\mathrm{int}(\mathcal{C}_{ic,R})}\frac{dvdw}{w} = \int_{\mathcal{C}_{ic,R}}\frac{dv}{w}$ by Green's theorem, which $= \int_0^{2\pi}\frac{-r\sin\theta}{b+r\sin\theta}d\theta = \frac{2\pi b}{\sqrt{b^2-r^2}} - 2\pi = 2\pi(\cosh R - 1)$. The circumference is $2\pi\sinh R = 2\pi(R + \frac{1}{6}R^3 + o(R^3))$ and the area is $2\pi(\cosh R - 1) = 2\pi(\frac{1}{2}R^2 + \frac{1}{24}R^4 + o(R^4)))$. Since $K = -1$ these formulas are consistent with those in Exercise 10.2.7. (See Exercise 10.2.3 for the $o(\ )$ notation.)

11.2.1 By Proposition 11.2.3, there is an isometry that takes $a$ to $i$ and $b$ to $ir$, say, where $r > 0$. Since isometries leave distances unchanged, we need only prove the result for $a = i, b = ir$. Assume that $r > 1$ (the case $r < 1$ is similar). Then, $d_{\mathcal{H}}(a,b) = \ln r$. On the other hand, if $\boldsymbol{\gamma}(t) = v(t) + iw(t)$ is any curve in $\mathcal{H}$ with $\boldsymbol{\gamma}(t_0) = i, \boldsymbol{\gamma}(t_1) = ir$, say, the length of the part of $\boldsymbol{\gamma}$ between $a$ and $b$ is $\int_{t_0}^{t_1}\frac{\sqrt{\dot{v}^2+\dot{w}^2}}{w}dt \geq \int_{t_0}^{t_1}\frac{\dot{w}}{w}dt = \int_1^r\frac{dw}{w} = \ln r$.

11.2.2 By applying an isometry we can assume that $l$ is a half-line, and then the result was proved in Exercise 11.1.1.

11.2.3 By applying an isometry, we can assume that $l$ is the imaginary axis. Then, $m$ must be the semicircle with centre the origin and radius $|a|$. Let $a = \rho e^{i\theta}$, where $\rho > 0, -\pi < \theta < \pi$, and let $c = it$, $t > 0$. Since $\tanh^{-1}x$ is a strictly increasing function of $x$, we have to show that $\left|\frac{a-it}{a+it}\right| > \left|\frac{a-i\rho}{a+i\rho}\right|$ if $t \neq \rho$. The second expression equals $\frac{1-\sin\theta}{1+\sin\theta}$, and the difference is $\frac{2(\rho-t)^2}{|a+it|^2}\frac{\sin\theta}{1+\sin\theta}$, which is $> 0$ if $t \neq \rho$.

11.2.4 (i) If $a \in \mathcal{H}$, let $l'$ and $m'$ be the unique hyperbolic lines passing through $a$ and perpendicular to $l$ and $m$, respectively (Exercise 11.2.3). Let $b$ and $c$ be the intersections of $l'$ and $m'$ with $l$ and $m$, respectively. We are given that $F(b) = b$ and $F(c) = c$, so $F(l') = l'$ and $F(m') = m'$ as $l'$ and $m'$ are the unique hyperbolic lines passing through $b$ and $c$ and perpendicular to $l$ and $m$. Since $a$ is the unique point of intersection of $l'$ and $m'$, we must have $F(a) = a$. (ii) Either $F$ or $\mathcal{I}_{0,1} \circ F$ fixes $l$, $m$ and the interior of the semicircle $m$. Next, either $F$, $\mathcal{I}_{0,1} \circ F$, $\mathcal{R}_0 \circ F$ or $\mathcal{R}_0 \circ \mathcal{I}_{0,1} \circ F$ fixes $l$, $m$, the interior of the semicircle $m$ and the region $\mathcal{H}_{>0} = \{z \in \mathcal{H} \,|\, \mathfrak{Re}(z) > 0\}$ to the right of $l$. Let $G$ be this isometry. Then, $G$ fixes each point of $m$ because there is a unique point of $m$ at any given distance $d > 0$ from $i$ in the region $\mathcal{H}_{>0}$. Similarly, $G$ fixes each point of $l$. Hence, $G$ is the identity by (i). (iii) Let $F$ be any isometry of $\mathcal{H}$. By the proof of Proposition 11.2.3, there is an isometry $G$ that is a composite of elementary isometries and which takes $F(i)$ to $i$ and $F(l)$ to $l$. Then, $G \circ F$ is an isometry that fixes $l$ and $i$. As $m$ is the unique hyperbolic line intersecting $l$ perpendicularly at $i$, $G \circ F$ fixes $m$. By (iii), $G \circ F$ is one of four composites of elementary isometries. It follows that $F$ is a composite of elementary isometries. (iv) By (iii) it suffices to prove that every elementary isometry is a composite of reflections and inversions in lines and circles perpendicular to the real axis. For reflections and inversions there is nothing to prove, so we need only consider translations and dilations. But, if $a \in \mathbb{R}$, $\mathcal{T}_a = \mathcal{R}_0 \circ \mathcal{R}_1$, where $\mathcal{R}_1$ is the reflection in the line $\mathfrak{Re}(z) = a/2$; and if $a > 0$, $\mathcal{D}_a = \mathcal{I}_0 \circ \mathcal{I}$, where $\mathcal{I}$ is inversion in the circle with centre the origin and radius $\sqrt{a}$.

11.2.5 (i) This is obvious from the proof of Proposition A.1.2(ii). (ii) If $a, b, c, d \in \mathbb{R}$, a calculation shows that $\mathfrak{Im}(M(z)) = \frac{ad-bc}{|cz+d|^2}\mathfrak{Im}(z)$. If $ad - bc > 0$ this is $> 0$ whenever $\mathfrak{Im}(z) > 0$. Conversely, suppose that $M$ takes $\mathcal{H}$ to itself. Assume that $c \neq 0$ and $d \neq 0$ (the cases in which $c = 0$ or $d = 0$ are similar but easier). Then, $M$ must take the real axis to itself (as it must take the lower half-plane $-\mathcal{H}$ to itself), i.e., $\frac{az+b}{cz+d} \in \mathbb{R}$ if $z \in \mathbb{R}$. Taking $z = 0$ gives $b/d = \lambda \in \mathbb{R}$, say. Letting $z \to \infty$ gives $a/c = \mu \in \mathbb{R}$, say. Then, $\frac{az+b}{cz+d} = \mu + \frac{\lambda-\mu}{\frac{c}{d}z+1}$. This is real whenever $z$ is real, so $c/d = \nu \in \mathbb{R}$, say. Hence, $a, b, c, d$ are, up to an overall multiple, equal to the real numbers $\mu\nu, \lambda, \nu, 1$, respectively. The condition $ad - bc > 0$ now follows from the previous calculation. (iii) Following the proof (and the notation) of Proposition A.2.2, we have $M = \mathcal{T}_{b/d} \circ \mathcal{D}_{a/d}$ if $c = 0$ (and then $d \neq 0$), while if $c \neq 0$, $M = \mathcal{T}_{a/c} \circ \mathcal{D}_{(ad-bc)/c^2} \circ (-K) \circ \mathcal{T}_{d/c}$. Hence, it suffices to show that $-K$ is a composite of elementary isometries. But $-K = \mathcal{R}_0 \circ \mathcal{I}_{0,1}$ in the notation of Exercise 11.2.4. (iv) $J$ is reflection in the imaginary axis and hence an isometry of $\mathcal{H}$, so the result follows from (i). (v) This follows from the fact that, if $M$ is a real Möbius transformation, so is $J \circ M \circ J$. For example, if $M_1, M_2$ are real Möbius transformations, then $(M_1 \circ J) \circ (M_2 \circ J) = M_1 \circ (J \circ M_2 \circ J)$ is a composite of real Möbius transformations, hence a real Möbius transformation by (i). (vi) By (v) and Exercise 11.2.4(iii), it suffices to prove that every elementary isometry of $\mathcal{H}$ is a Möbius isometry. If $a \in \mathbb{R}$, $\mathcal{T}_a$ is real Möbius and $\mathcal{R}_a = \mathcal{T}_{2a} \circ J$. If $a > 0$ then $\mathcal{D}_a$ is real Möbius. Finally, if $a \in \mathbb{R}, r > 0$ then $\mathcal{I}_{a,r} = \mathcal{T}_a \circ \mathcal{D}_{r^2} \circ \mathcal{I}_{0,1} \circ \mathcal{T}_{-a}$ (see the proof of Proposition 11.2.1), so it suffices to prove that $\mathcal{I}_{0,1}$ is a Möbius isometry; but $\mathcal{I}_{0,1} = (-K) \circ J$, where $-K(z) = -1/z$ is real Möbius.

11.3.1 The distance we want is $2\tanh^{-1}\left|\frac{\mathcal{P}^{-1}(b)-\mathcal{P}^{-1}(a)}{\mathcal{P}^{-1}(b)-\mathcal{P}^{-1}(a)}\right|$. Now use $\mathcal{P}^{-1}(z) = \frac{z+1}{i(z-1)}$. The algebra is straightforward.

11.3.2 By Proposition 11.2.3, there is an isometry $F$ that takes $l$ to the real axis and the point of intersection of $l$ and $m$ to the origin. Then $F$ must take $m$ to the imaginary axis as this is the unique hyperbolic line through the origin perpendicular to the real axis. The number of such isometries is the number of isometries that take the real axis to itself and the imaginary axis to itself. If $G$ is such an isometry, then either $G$, $\mathcal{R}_0 \circ G$, $\mathcal{R}_1 \circ G$ or $\mathcal{R}_1 \circ \mathcal{R}_0 \circ G$ fix the real and imaginary axes and also each quadrant into which the disc $\mathcal{D}_P$ is divided by the axes. If $H$ is this isometry, the argument used in the solution of Exercise 11.2.4(ii) shows that $H$ must be the identity map. Hence, $G = \mathcal{R}_0$, $\mathcal{R}_1$, $\mathcal{R}_0 \circ \mathcal{R}_1$ or the identity map.

11.3.3 Let us call a Möbius transformation of the type in the statement of the exercise a *hyperbolic* Möbius transformation. Since $\mathcal{P}$ is a Möbius transformation, the Möbius transformations that take $\mathcal{D}_P$ to itself are those of the form $\mathcal{P}M\mathcal{P}^{-1}$, where $M$ is a Möbius transformation that takes $\mathcal{H}$ to itself, i.e., a real Möbius transformation (Exercise 11.2.5). If $M(z) = \frac{az+b}{cz+d}$, where $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$, we find that $\mathcal{P}M\mathcal{P}^{-1}(z) = \frac{(a+d+i(b-c))z+a-d-i(b+c)}{(a-d+i(b+c))z+a+d-i(b-c)}$. Since $|a + d + i(b - c)|^2 - |a - d - i(b + c)|^2 = 4(ad - bc) > 0$, $\mathcal{P}M\mathcal{P}^{-1}$ is hyperbolic. Conversely, we have to show that if $M$ is a hyperbolic Möbius transformation, then $\mathcal{P}^{-1}M\mathcal{P}$ is a real Möbius transformation. The calculation is similar to that already given.

11.3.4 By Exercise 11.2.5(iii), the isometries of $\mathcal{H}$ are of the form $M$ or $M \circ J$ where $M$ is real Möbius and $J(z) = -\bar{z}$. Hence, the isometries of $\mathcal{D}_P$ are $\mathcal{P}M\mathcal{P}^{-1}$ and $\mathcal{P}(M \circ J)\mathcal{P}^{-1} = \mathcal{P}M\mathcal{P}^{-1} \circ \mathcal{P}J\mathcal{P}^{-1}$. But $\mathcal{P}M\mathcal{P}^{-1}$ is hyperbolic and $\mathcal{P}J\mathcal{P}^{-1}(z) = \mathcal{P}J\left(\frac{z+1}{i(z-1)}\right) = \mathcal{P}\left(\frac{\bar{z}+1}{i(\bar{z}-1)}\right) = \mathcal{P}(\mathcal{P}^{-1}(\bar{z})) = \bar{z}$.

11.3.5 We know that every isometry of $\mathcal{H}$ is the composite of reflections $\mathcal{R}_a$ and inversions $\mathcal{I}_{a,r}$ with $a \in \mathbb{R}$, $r > 0$ (Exercise 11.2.4(iv)). Now, $\mathcal{I}_{a,r} = \mathcal{T}_a \circ \mathcal{D}_{r^2} \circ \mathcal{I}_{0,1} \circ \mathcal{T}_{-a}$ and any translation $\mathcal{T}_a$ ($a \in \mathbb{R}$) is the composite of reflections $\mathcal{R}_0 \circ \mathcal{R}_{a/2}$. It therefore suffices to show that, if $F$ is any isometry of $\mathcal{H}$ of the form $\mathcal{R}_a$ ($a \in \mathbb{R}$), $\mathcal{D}_a$ ($a > 0$) or $\mathcal{I}_{0,1}$, then $\mathcal{P} \circ F \circ \mathcal{P}^{-1}$ is a composite of isometries of $\mathcal{D}_P$ of the types in Proposition 11.3.3. We find that (a) if $a \neq 0$, $\mathcal{P} \circ \mathcal{R}_a \circ \mathcal{P}^{-1} = \mathcal{I}_{b,r}$, where $b = \frac{1+ia}{ia}$, $r = 1/|a|$; (b) $\mathcal{P} \circ \mathcal{R}_0 \circ \mathcal{P}^{-1}$ is reflection in the real axis; (c) $\mathcal{P} \circ \mathcal{I}_{0,1} \circ \mathcal{P}^{-1}$ is reflection in the imaginary axis; (d) if $a > 0$, $\mathcal{P} \circ \mathcal{D}_a \circ \mathcal{P}^{-1}$ is the composite of two inversions of the type in Proposition 11.3.3(i), namely $\mathcal{I}_{c,\sqrt{c^2-1}} \circ \mathcal{I}_{b,\sqrt{b^2-1}}$, where $b, c$ are real numbers such that $b^2 > 1$, $c^2 > 1$ and $a = f(c)/f(b)$ where $f(x) = \frac{x+1}{x-1}$. (We can take $b$ to be any real number $> 1$ and distinct from $1/f(a)$, then choose $c = f(af(b))$; then, $f(c) = f(f(af(b))) = af(b)$ using the property $f(f(x)) = x$ for all $x \neq 1$.)

11.3.6 Suppose first that $\gamma = \pi/2$. The cosine rule gives $\cosh C = \cosh A \cosh B$ and $\cosh A = \cosh B \cosh C - \sinh B \sinh C \cos\alpha$. Eliminating $\cosh C$ gives $\cos\alpha = \cosh A \sinh B / \sinh C$. Hence, $\sin^2\alpha \sinh^2 C = \sinh^2 C - \cosh^2 A \sinh^2 B = \cosh^2 A \cosh^2 B - 1 - \cosh^2 A \sinh^2 B = \cosh^2 A - 1 = \sinh^2 A$, so $\sin\alpha / \sinh A = 1/\sinh C$. Interchanging the roles of $A$ and $B$ gives $\sin\beta / \sinh B = 1/\sinh C$.

In the general case suppose that the hyperbolic line through the vertex of the triangle with angle $\alpha$ intersects the opposite side at a point which divides that side into segments of lengths $A'$ and $A''$, so that $\beta$ is the angle between the sides of lengths $C$ and $A'$. Suppose also that this hyperbolic line segment has length $D$. Then the original triangle is divided into two triangles, one with angles $\pi/2, \beta, \alpha'$ and sides of lengths $C, D, A'$ and the other with angles $\pi/2, \gamma, \alpha''$ and sides of lengths $B, D, A''$. Applying the first part to each of these triangles gives $\sin\beta / \sinh D = 1/\sinh C$ and $\sin\gamma / \sinh D = 1/\sinh B$. Hence, $\sin\beta / \sinh B = \sin\gamma / \sinh C$. The other equation is proved by interchanging the roles of $A$ and $B$ (for example). The case in which the hyperbolic line through a vertex meets the hyperbolic line through the other two vertices in a point outside the triangle is similar.

11.3.7 (i) Using the sine rule, $\cos^2\alpha \sinh^2 C = \sinh^2 C - \sinh^2 A = \cosh^2 C - \cosh^2 A = \cosh^2 A(\cosh^2 B - 1)$ (using the cosine rule) $= \cosh^2 A \sinh^2 B$. (ii) Using the sine and cosine rules, $\sin\beta = \sinh B / \sinh C$ and $\cos\alpha = (\cosh B \cosh C - \cosh A)/\sinh B \sinh C$. Hence, $\cos\alpha / \sin\beta = (\cosh B \cosh C - \cosh A)/\sinh^2 B = (\cosh^2 B \cosh A - \cosh A)/\sinh^2 B = \cosh A$. (iii) Using $\sin\beta = \sinh B / \sinh C$ and the cosine rule for $\cos\beta$, we get $\cot\beta = \frac{\cosh A \cosh C - \cosh B}{\sinh A \sinh B} = \frac{\sinh^2 A \cosh B}{\sinh A \sinh B} = \frac{\sinh A}{\tanh B}$.

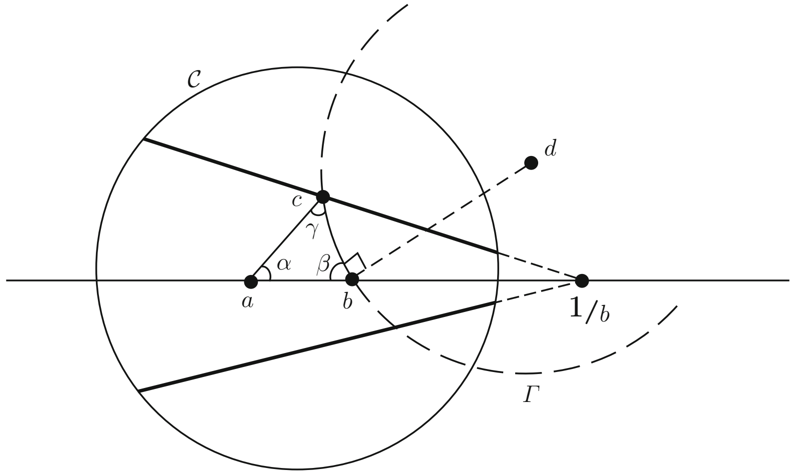
11.3.8 If $\gamma = \pi/2$ the formula we want is that in Exercise 11.3.7(ii). In the general case, we use the method (and notation) of the solution of Exercise 11.3.6. In the case where the hyperbolic line through a vertex perpendicular to the opposite side meets that side at a point inside the triangle, applying Exercise 11.3.7(ii) to the two right-angled triangles gives $\cosh A' = \cos\alpha'/\sin\beta$, $\cosh A'' = \cos\alpha''/\sin\gamma$, so $\cosh A = \cosh(A' + A'') = \cosh A' \cosh A'' + \sinh A' \sinh A'' = \frac{\cos\alpha' \cos\alpha''}{\sin\beta\sin\gamma} + \frac{\tanh^2 D}{\tan\beta\tan\gamma}$ using Exercise 11.3.7(iii). By 11.3.7(ii), $\cosh D = \frac{\cos\beta}{\sin\alpha'} = \frac{\cos\gamma}{\sin\alpha''}$ so $\cosh A \sin\beta\sin\gamma = \cos\alpha + \sin\alpha' \sin\alpha'' + \tanh^2 D \cos\beta\cos\gamma = \cos\alpha + \operatorname{sech}^2 D \cos\beta\cos\gamma + \tanh^2 D \cos\beta\cos\gamma = \cos\alpha + \cos\beta\cos\gamma$. The case in which the perpendicular meets the opposite side at a point outside the triangle is similar.

11.3.9 Let $\boldsymbol{\gamma}(t) = (x(t), y(t), z(t))$ be a curve on $S^2$. Then, $\Pi(\boldsymbol{\gamma}(t)) = (u(t), v(t))$, where $u = \frac{x}{1-z}, v = \frac{y}{1-z}$. Denoting $d/dt$ by a dot, $\dot{u} = \frac{(1-z)\dot{x}+x\dot{z}}{(1-z)^2}$ with a similar formula for $\dot{v}$, which give $4\frac{\dot{u}^2+\dot{v}^2}{1+u^2+v^2} = \frac{1}{(1-z)^2}\big((1-z)^2(\dot{x}^2+\dot{y}^2) + (x^2+y^2)\dot{z}^2 + 2(x\dot{x}+y\dot{y})\dot{z}(1-z)\big)$. Using $x^2+y^2 = 1-z^2$, which implies $x\dot{x}+y\dot{y} = -z\dot{z}$, this expression simplifies to $\dot{x}^2+\dot{y}^2+\dot{z}^2$. Hence, the length of $\Pi\circ\boldsymbol{\gamma}$ calculated using the given first fundamental form on $\mathbb{R}^2$ is $\int \frac{2\sqrt{\dot{u}^2+\dot{v}^2}}{1+u^2+v^2}dt = \int \sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}dt$, which is the length of $\boldsymbol{\gamma}$. Hence, $\Pi$ is an isometry.

11.4.1 Let $l$ and $m$ be two distinct hyperbolic lines in $\mathcal{H}$ that do not intersect at any point of $\mathcal{H}$. If $l$ and $m$ are both half-lines they are parallel as they do not have a common perpendicular. If at least one of $l$ and $m$ is a semicircle, then $l$ and $m$ are parallel if they intersect at a point of the real axis, and ultra-parallel otherwise.

11.4.2 We work in $\mathcal{H}$ and assume that $l$ is the imaginary axis (by applying a suitable isometry). If $a = v + iw$, the semicircle geodesic through $a$ intersects $l$ at $i\sqrt{v^2+w^2} = ir$, say. The distance of $a$ from $l$ is $2\tanh^{-1}\left|\frac{ir-v-iw}{ir-v+iw}\right|$. Setting this equal to a constant, say $D$, gives (after some algebra) $v^2/w^2 = 2\sinh^2(D/2)$. This is the equation of a pair of lines passing through the origin. As they are not perpendicular to the real axis (unless $D = 0$), they are not hyperbolic lines.

11.4.3 We work in $\mathcal{D}_P$. By applying an isometry we can assume that $a$ is the origin and $b > 0$. Suppose that the hyperbolic triangle with vertices $a, b$, and $c$ has internal angles $\alpha, \beta$, and $\gamma$, and assume that $\mathfrak{Re}(c) > 0$.




The hyperbolic line through $b$ and $c$ is part of a circle $\Gamma$ with centre $d$ and radius $r$, say. The line through $b$ and $d$ makes an angle $\pi/2 - \beta$ with the real axis, so $d = b + r\sin\beta + ir\cos\beta$. Similarly, $d = c + r\sin(\alpha+\gamma) - ir\cos(\alpha+\gamma) = c + r\sin(A+\beta) + ir\cos(A+\beta)$ using $A = \pi - \alpha - \beta - \gamma$. Writing $c = v + iw$ we get $v = b - r\sin A\cos\beta + r(1-\cos A)\sin\beta$, $w = r(1-\cos A)\cos\beta + r\sin A\sin\beta$. Then, $v(1-\cos A) + w\sin A = (b + 2r\sin\beta)(1-\cos A)$. But, since $\gamma$ intersects the boundary $\mathcal{C}$ of $\mathcal{D}_P$ perpendicularly, $r^2 + 1 = |d|^2 = (b + r\sin\beta)^2 + r^2\cos^2\beta = r^2 + b^2 + 2br\sin\beta$, so $2br\sin\beta = 1 - b^2$. Hence, $v(1-\cos A) + w\sin A = \frac{1-\cos A}{b}$. This is the equation of a straight line that intersects the real axis at $1/b$ and makes an angle $A/2$ with the (negative) real axis. The set of points $c$ for which the triangle with vertices $a, b, c$ has area $A$ is the union of this line together with its reflection in the real axis. These lines are not hyperbolic lines as they do not pass through the origin.

11.5.1 From Example 6.3.5, $\Pi^{-1}(v,w) = \left(\frac{2v}{v^2+w^2+1}, \frac{2w}{v^2+w^2+1}, \frac{v^2+w^2-1}{v^2+w^2+1}\right)$, so $\mathcal{K}(v,w) = \frac{2(v,w)}{v^2+w^2+1}$.

11.5.2 From Appendix 2, every Möbius transformation is a composite of transformations of the form $z \mapsto 1/z$, $z \mapsto z + \lambda$, $z \mapsto \lambda z$ (where $\lambda \neq 0$ in the last case). Hence, it suffices to establish Eq. 11.11 when $M$ is of this form. For the last two types this is obvious; for the first, $(a^{-1}, b^{-1}; c^{-1}, d^{-1}) = \frac{(a^{-1}-c^{-1})(b^{-1}-d^{-1})}{(a^{-1}-d^{-1})(b^{-1}-d^{-1})} = (a, b; c, d)$ on multiplying numerator and denominator by $abcd$. This argument is only valid provided none of $a, b, c$, and $d$ is 0 or $\infty$, but a similar argument works in the other cases. For example, if $a = \infty$ but $b, c, d \neq 0$, we have to show that

$(0, b^{-1}; c^{-1}, d^{-1}) = (\infty, b; c, d)$. This is proved by multiplying numerator and denominator of $(0, b^{-1}; c^{-1}, d^{-1}) = \frac{-c^{-1}(b^{-1}-d^{-1})}{-d^{-1}(b^{-1}-c^{-1})}$ by $bcd$.

If $M : \mathbb{C}_\infty \to \mathbb{C}_\infty$ is a bijection satisfying Eq. 11.11, let $b, c, d \in \mathbb{C}_\infty$ be such that $M(b) = 1, M(c) = 0, M(d) = \infty$. Then, $M(z) = (M(z), 1; 0, \infty) = (M(z), M(b); M(c), M(d)) = (z, b; c, d)$, so $M$ is a Möbius transformation.

11.5.3 It is enough to prove the existence when $a' = \infty, b' = 0, c' = 1$. For if $M$ and $M'$ are Möbius transformations taking $(a, b, c)$ and $(a', b', c')$ to $(\infty, 0, 1)$, then $M'^{-1} \circ M$ is a Möbius transformation taking $(a, b, c)$ to $(a', b', c')$. But $M(z) = (a, b; c, z)$ is a Möbius transformation that takes $(a, b, c)$ to $(\infty, 0, 1)$. For the uniqueness, note that if $M$ is a Möbius transformation that takes $(a, b, c)$ to $(\infty, 0, 1)$, then $M(z) = (\infty, 0; 1, M(z)) = (a, b; c, z)$.

11.5.4 $(a, -1/\bar{a}; b, -1/\bar{b}) = \frac{(a-b)(-\frac{1}{\bar{a}}+\frac{1}{\bar{b}})}{(a+\frac{1}{\bar{b}})(-\frac{1}{\bar{a}}-b)} = \frac{(a-b)(\bar{b}-\bar{a})}{(1+\bar{a}b)(1+a\bar{b})} = -\left|\frac{a-b}{1+\bar{a}b}\right|^2 = -\tan^2 \frac{1}{2}d$ by Proposition 6.5.2.

11.5.5 The reflection in the line through the origin making an angle $\theta$ with the real axis is $\mathcal{R}(z) = e^{2i\theta}\bar{z}$. Then, $\mathcal{K}(\mathcal{R}(z)) = \frac{2e^{2i\theta}\bar{z}}{|z|^2+1} = e^{2i\theta}\overline{\mathcal{K}(z)} = \mathcal{R}(\mathcal{K}(z))$.

11.5.6 This follows from Exercises 11.3.5 and 11.5.5 and Proposition 11.5.4.

# Chapter 12

12.1.1 $\kappa_1 + \kappa_2 = 0 \Longrightarrow \kappa_2 = -\kappa_1 \Longrightarrow K = \kappa_1\kappa_2 = -\kappa_1^2 \leq 0$. $K = 0 \Longleftrightarrow \kappa_1^2 = 0 \Longleftrightarrow \kappa_1 = \kappa_2 = 0 \Longleftrightarrow$ the surface is an open subset of a plane (by Proposition 8.2.9).

12.1.2 From Eq. 8.15, $\boldsymbol{\sigma}_u^\lambda \times \boldsymbol{\sigma}_v^\lambda = (1-\lambda\kappa_1)(1-\lambda\kappa_2)\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$, where $\kappa_1, \kappa_2$ are the principal curvatures of $\boldsymbol{\sigma}$. Since $\boldsymbol{\sigma}$ is minimal, $\kappa_2 = -\kappa_1$ so $\mathcal{A}_{\boldsymbol{\sigma}^\lambda}(U) = \int_U (1-\lambda^2\kappa_1^2) \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel dudv = \mathcal{A}_{\boldsymbol{\sigma}}(U) - \lambda^2 \int_U \kappa_1^2 \parallel \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \parallel dudv$. Since the integrand in the last integral is $\geq 0$ everywhere, the stated inequality follows. Equality holds $\Longleftrightarrow$ the last integral vanishes, which happens $\Longleftrightarrow$ the integrand vanishes everywhere, i.e., $\Longleftrightarrow \kappa_1 = 0$ everywhere. In that case $\kappa_2 = -\kappa_1 = 0$ also, and $\boldsymbol{\sigma}$ is an open subset of the plane by Proposition 8.2.9.

12.1.3 By Proposition 8.6.1, a compact minimal surface would have $K > 0$ at some point, contradicting Exercise 12.1.1.

12.1.4 The first part follows from Exercises 6.1.2 and 7.1.4. The map which wraps the plane onto the unit cylinder (Example 6.2.4) is a local isometry, but the plane is a minimal surface and the cylinder is not.

12.2.1 By the solution of Exercise 8.1.2, the helicoid $\boldsymbol{\sigma}(u,v) = (v\cos u, v\sin u, \lambda u)$ has $E = \lambda^2 + v^2, F = 0, G = 1, L = 0, M = \lambda/(\lambda^2+v^2)^{1/2}, N = 0$, so $H = \frac{LG-2MF+NG}{2(EG-F^2)} = 0$.

12.2.2 Calculation shows that the first and second fundamental forms of $\boldsymbol{\sigma}^t$ are $\cosh^2 u(du^2+dv^2)$ and $-\cos t\,du^2 - 2\sin t\,dudv + \cos t\,dv^2$, respectively, so $H = \frac{-\cos t\cosh^2 u+\cos t\cosh^2 u}{2\cosh^4 u} = 0$.

12.2.3 From Example 5.3.1, the cylinder can be parametrized by $\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\mathbf{a}$, where $\boldsymbol{\gamma}$ is unit-speed, $\parallel \mathbf{a} \parallel = 1$ and $\boldsymbol{\gamma}$ is contained in a plane $\Pi$ perpendicular to $\mathbf{a}$. We have $\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}} = \mathbf{t}$ (a dot denoting $d/du$), $\boldsymbol{\sigma}_v = \mathbf{a}$, so $E = 1, F = 0, G = 1$; $\mathbf{N} = \mathbf{t}\times\mathbf{a}$, $\boldsymbol{\sigma}_{uu} = \dot{\mathbf{t}} = \kappa\mathbf{n}$, $\boldsymbol{\sigma}_{uv} = \boldsymbol{\sigma}_{vv} = \mathbf{0}$, so $L = \kappa\mathbf{n}\cdot(\mathbf{t}\times\mathbf{a})$, $M = N = 0$. Now $\mathbf{t}\times\mathbf{a}$ is a unit vector parallel to $\Pi$ and perpendicular to $\mathbf{t}$, hence parallel to $\mathbf{n}$; so $L = \pm\kappa$ and $H = \pm\kappa/2$. So $H = 0 \iff \kappa = 0 \iff \boldsymbol{\gamma}$ is part of a straight line $\iff$ the cylinder is an open subset of a plane.

12.2.4 The first fundamental form is $(\cosh v + 1)(\cosh v - \cos u)(du^2 + dv^2)$, so $\boldsymbol{\sigma}$ is conformal. By Exercise 8.5.1, to show that $\boldsymbol{\sigma}$ is minimal we must show that $\boldsymbol{\sigma}_{uu} + \boldsymbol{\sigma}_{vv} = \mathbf{0}$; but this is so, since $\boldsymbol{\sigma}_{uu} = (\sin u\cosh v, \cos u\cosh v, \sin\frac{u}{2}\sinh\frac{v}{2}) = -\boldsymbol{\sigma}_{vv}$.

(i) $\boldsymbol{\sigma}(0,v) = (0, 1-\cosh v, 0)$, which is the $y$-axis. Any straight line is a geodesic. (ii) $\boldsymbol{\sigma}(\pi,v) = (\pi, 1+\cosh v, -4\sinh\frac{v}{2})$, which is a curve in the plane $x = \pi$ such that $z^2 = 16\sinh^2\frac{v}{2} = 8(\cosh v - 1) = 8(y-2)$, i.e., a parabola. The geodesic equations are $\frac{d}{dt}(E\dot{u}) = \frac{1}{2}E_u(\dot{u}^2+\dot{v}^2)$, $\frac{d}{dt}(E\dot{v}) = \frac{1}{2}E_v(\dot{u}^2+\dot{v}^2)$, where a dot denotes the derivative with respect to the parameter $t$ of the geodesic and $E = (\cosh v+1)(\cosh v - \cos u)$. When $u = \pi$, the unit-speed condition is $E\dot{v}^2 = 1$, so $\dot{v} = 1/(\cosh v+1)$. Hence, the first geodesic equation is $0 = \frac{1}{2}E_u\dot{v}^2$, which holds because $E_u = \sin u(\cosh v+1) = 0$ when $u = \pi$; the second geodesic equation is $\frac{d}{dt}(\cosh v+1) = (\cosh v+1)\sinh v\,\dot{v}^2 = \sinh v\,\dot{v}$, which obviously holds. (iii) $\boldsymbol{\sigma}(u,0) = (u-\sin u, 1-\cos u, 0)$, which is the cycloid of Exercise 1.1.7 (in the $xy$-plane, with $a = 1$ and with $t$ replaced by $u$). The second geodesic equation is satisfied because $E_v = \sinh v(2\cosh v+1-\cos u) = 0$ when $v = 0$. The unit-speed condition is $2(1-\cos u)\dot{u}^2 = 1$, so $\dot{u} = 1/2\sin\frac{u}{2}$. The first geodesic equation is $\frac{d}{dt}(4\sin^2\frac{u}{2}\dot{u}) = \sin u\dot{u}^2$, i.e., $\frac{d}{dt}(2\sin\frac{u}{2}) = \cos\frac{u}{2}\dot{u}$, which obviously holds.

12.3.1 (i) From the proof of Theorem 12.3.2, the Gauss map is conformal $\iff$ $\mathcal{W}^2 = \lambda\cdot\mathrm{id}$, where $\lambda$ is a smooth function on $\mathcal{S}$. If $H \neq 0$ at $\mathbf{p}$, then $H \neq 0$

on an open subset $\mathcal{O}$ of $\mathcal{S}$ containing $\mathbf{p}$. By Exercise 8.1.6, $\mathcal{W} = \frac{\lambda^2+K}{2H}\cdot\mathrm{id}$, so every point of $\mathcal{O}$ is an umbilic and $\mathcal{O}$ is an open subset of a plane or a sphere by Proposition 8.2.9. Since $H \neq 0$ the planar case is impossible. Part (ii) is now obvious. For (iii), assume that $\mathcal{S}$ is not minimal. Then, there is a point $\mathbf{p} \in \mathcal{S}$ at which $H \neq 0$, say $H = \mu$. The argument in (i) and (ii) shows that the set $\mathcal{S}_\mu$ of points of $\mathcal{S}$ at which $H = \mu$ is a (non-empty) open subset of $\mathcal{S}$; it is also a closed subset because $H$ is a continuous function on $\mathcal{S}$. Since $\mathcal{S}$ is connected, $\mathcal{S}_\mu = \mathcal{S}$. Hence, every point of $\mathcal{S}$ is an umbilic, and so $\mathcal{S}$ is an open subset of a sphere (the planar case is impossible as $\mu \neq 0$).

12.3.2 (i) From Example 12.1.4, $\mathbf{N} = (-\operatorname{sech} u \cos v, -\operatorname{sech} u \sin v, \tanh u)$. Hence, if $\mathbf{N}(u,v) = \mathbf{N}(u',v')$, then $u = u'$ since $u \mapsto \tanh u$ is injective, so $\cos v = \cos v'$ and $\sin v = \sin v'$, hence $v = v'$; thus, $\mathbf{N}$ is injective. If $\mathbf{N} = (x,y,z)$, then $x^2 + y^2 = \operatorname{sech}^2 u \neq 0$, so the image of $\mathbf{N}$ does not contain the poles. Given a point $(x,y,z) \in S^2$ other than the poles, let $u = \pm\operatorname{sech}^{-1}\sqrt{x^2+y^2}$, the sign being that of $z$, and let $v$ be such that $\cos v = -x/\sqrt{x^2+y^2}$, $\sin v = -y/\sqrt{x^2+y^2}$; then, $\mathbf{N}(u,v) = (x,y,z)$.

(ii) By the solution of Exercise 8.1.2, $\mathbf{N} = (\lambda^2+v^2)^{-1/2}(-\lambda \sin u, \lambda \cos u, -v)$. Since $\mathbf{N}(u,v) = \mathbf{N}(u+2k\pi, v)$ for all integers $k$, the infinitely many points $\boldsymbol{\sigma}(u+2k\pi, v) = \boldsymbol{\sigma}(u,v) + (0,0,2k\pi)$ of the helicoid all have the same image under the Gauss map. (Of course, this is geometrically obvious because the helicoid itself is left unchanged by the translation by $2\pi$ parallel to the $z$-axis.) If $\mathbf{N} = (x,y,z)$, then $x^2+y^2 = \lambda^2/(\lambda^2+v^2) \neq 0$, so the image of $\mathbf{N}$ does not contain the poles. If $(x,y,z) \in S^2$ and $x^2+y^2 \neq 0$, let $v = -\lambda z/\sqrt{x^2+y^2}$ and let $u$ be such that $\sin u = -x/\sqrt{x^2+y^2}$, $\cos u = -y/\sqrt{x^2+y^2}$; then $\mathbf{N}(u,v) = (x,y,z)$.

12.4.1 By Proposition 12.3.2, if $K \neq 0$ the Gauss map $\mathcal{G} : \mathcal{S} \to S^2$ is a conformal local diffeomorphism. Let $R$ be a rotation of $\mathbb{R}^3$ about the origin that takes $\mathcal{G}(\mathbf{p})$ to the south pole of $S^2$ (or any point other than the north pole). There is an open subset $\mathcal{O}$ of $\mathcal{S}$ containing $\mathbf{p}$ such that $\mathcal{G}(\mathcal{O})$ does not contain the north pole. By Example 6.3.5, $\Pi \circ R \circ \mathcal{G}$ is a conformal diffeomorphism from $\mathcal{O}$ to an open subset $U$ of $\mathbb{R}^2$. The inverse of this diffeomorphism is the desired surface patch $\boldsymbol{\sigma}$.

12.5.1 $\boldsymbol{\varphi} = \boldsymbol{\sigma}_u - i\boldsymbol{\sigma}_v = (1-u^2+v^2-2iuv, 2uv - i(1-v^2+u^2), 2u+2iv) = (1-\zeta^2, -i(1+\zeta^2), 2\zeta)$. So the conjugate surface is, up to a translation, $\tilde{\boldsymbol{\sigma}}(u,v) = \mathfrak{Re}\int (i(1-\zeta^2), 1+\zeta^2, 2i\zeta)\, d\zeta = \mathfrak{Re}\left(i\left(\zeta - \frac{\zeta^3}{3}\right), \zeta + \frac{\zeta^3}{3}, i\zeta^2\right) = \left(-v + u^2 v - \frac{v^3}{3}, u + \frac{u^3}{3} - uv^2, -2uv\right)$. Let $U = (u-v)/\sqrt{2}$, $V = (u+v)/\sqrt{2}$, $\tilde{\boldsymbol{\sigma}}(U,V) = \boldsymbol{\sigma}(u,v)$; then,

$$\tilde{\boldsymbol{\sigma}}(U,V) = \left( \frac{1}{\sqrt{2}} \left( U - V + UV^2 - U^2V + \frac{1}{3}V^3 - \frac{1}{3}U^3 \right) \right),$$
$$\frac{1}{\sqrt{2}} \left( U + V + UV^2 + U^2V - \frac{1}{3}V^3 - \frac{1}{3}U^3 \right), U^2 - V^2 \;).$$

Applying the $\pi/4$ rotation $(x,y,z) \mapsto \left(\frac{1}{\sqrt{2}}(x+y), \frac{1}{\sqrt{2}}(y-x), z\right)$ to $\tilde{\boldsymbol{\sigma}}(U,V)$ then gives $\left(U - \frac{1}{3}U^3 + UV^2, V - \frac{1}{3}V^3 + U^2V, U^2 - V^2\right)$, which is Enneper's surface again.

12.5.2 $\boldsymbol{\varphi} = \left(\frac{1}{2}(1-\zeta^{-4})(1-\zeta^2), \frac{i}{2}(1-\zeta^{-4})(1+\zeta^2), \zeta(1-\zeta^{-4})\right)$, so

$$\begin{aligned} \boldsymbol{\sigma} &= \mathfrak{Re}\left( \frac{1}{2}\left(\zeta - \frac{\zeta^3}{3} - \zeta^{-1} + \frac{\zeta^{-3}}{3}\right), \right. \\ &\qquad \left. \frac{i}{2}\left(\zeta + \frac{\zeta^3}{3} + \zeta^{-1} + \frac{\zeta^{-3}}{3}\right), \frac{\zeta^2}{2} + \frac{\zeta^{-2}}{2} \right) \\ &= \mathfrak{Re}\left( -\frac{1}{6}(\zeta - \zeta^{-1})^3, \frac{i}{6}(\zeta + \zeta^{-1})^3, \frac{1}{2}(\zeta + \zeta^{-1})^2 \right), \end{aligned}$$

up to a translation. Put $\zeta = e^{\tilde{\zeta}}$, $\tilde{\zeta} = \tilde{u} + i\tilde{v}$. Then, $\boldsymbol{\sigma}(u,v) = \tilde{\boldsymbol{\sigma}}(\tilde{u},\tilde{v})$, where $\tilde{\boldsymbol{\sigma}}(\tilde{u},\tilde{v}) =$

$$\begin{aligned} &\mathfrak{Re}\left( -\frac{4}{3}\sinh^3\tilde{\zeta}, \frac{4i}{3}\cosh^3\tilde{\zeta}, 2\cosh^2\tilde{\zeta} \right) \\ &= \left( 4\sinh\tilde{u}\cos\tilde{v}(\cosh^2\tilde{u}\sin^2\tilde{v} - \frac{1}{3}\sinh^2\tilde{u}\cos^2\tilde{v}), \right. \\ &\qquad 4\sinh\tilde{u}\sin\tilde{v}(\frac{1}{3}\sinh^2\tilde{u}\sin^2\tilde{v} - \cosh^2\tilde{u}\cos^2\tilde{v}), \\ &\qquad \left. 2(\cosh^2\tilde{u}\cos^2\tilde{v} - \sinh^2\tilde{u}\sin^2\tilde{v}) \right). \end{aligned}$$

12.5.3 The first part is obvious. (i) If $a \in \mathbb{R}$, the identity $\boldsymbol{\sigma}^a_u - i\boldsymbol{\sigma}^q_v = a(\boldsymbol{\sigma}_u - i\boldsymbol{\sigma}_v)$ implies that $\boldsymbol{\sigma}^a = a\boldsymbol{\sigma} + \mathbf{a}$, where $\mathbf{a}$ is a constant vector. Hence, $\boldsymbol{\sigma}^a$ is obtained from $\boldsymbol{\sigma}$ by applying the dilation $D_a$ followed by the translation $T_{\mathbf{a}}$ (Appendix 1). (ii) If $f$ and $g$ are the functions in the Weierstrass representation of $\boldsymbol{\sigma}$ (Proposition 12.5.4), those in the Weierstrass representation of $\boldsymbol{\sigma}^a$ are $af$ and $g$ (see Eq. 12.22). By Eq. 12.25, replacing $f$ by $af$ leaves the first fundamental form unchanged, so the map $\boldsymbol{\sigma}(u,v) \mapsto \boldsymbol{\sigma}^a(u,v)$ is an isometry, and by Eq. 12.26 $\mathbf{N}$ does not depend on $f$, so the tangent planes of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^a$ at corresponding points are parallel.

12.5.4 We have $\boldsymbol{\sigma}_u^{e^{it}} - i\boldsymbol{\sigma}_v^{e^{it}} = e^{it}(\boldsymbol{\sigma}_u - i\boldsymbol{\sigma}_v)$. Since $\boldsymbol{\sigma}(u,v) = (\cosh u \cos v, \cosh u \sin v, u)$, we get $\boldsymbol{\sigma}_u^{e^{it}} = (\cos t \sinh u \cos v - \sin t \cosh u \sin v, \sin t \cosh u \cos v + \cos t \sinh u \sin v, \cos t)$, $\boldsymbol{\sigma}_v^{e^{it}} = (-\cos t \cosh u \sin v - \sin t \sinh u \cos v, \cos t \cosh u \cos v - \sin t \sinh u \sin v, -\sin t)$. Integrating gives $\boldsymbol{\sigma}^{e^{it}}(u,v) = \cos t(\cosh u \cos v, \cosh u \sin v, u) + \sin t(-\sinh u \sin v, \sinh u \cos v, -v) = \cos t\boldsymbol{\sigma}(u,v) + \sin t\hat{\boldsymbol{\sigma}}(u,v)$, say (up to a translation). In the notation of Exercise 6.2.3, $\tilde{\boldsymbol{\sigma}}(\sinh u, \frac{\pi}{2} + v) = (-\sinh u \sin v, \sinh u \cos v, \frac{\pi}{2} + v)$. Reflecting in the $xy$-plane and then translating by $\pi/2$ along the $z$-axis takes $\tilde{\boldsymbol{\sigma}}(\sinh u, \frac{\pi}{2} + v)$ to $\hat{\boldsymbol{\sigma}}(u,v)$.

12.5.5 (i) $\boldsymbol{\varphi}$ is never zero since we have arranged that $F'$ and $G'$ are never both zero. Condition (ii) in Theorem 12.5.2 is obvious. When $v = 0$, $F'(z) = \frac{\partial}{\partial u}F(u,0) = F_u$, etc, so $\frac{d}{du}\boldsymbol{\sigma}(u,0) = (\dot f, \dot g, 0) = \dot{\boldsymbol{\gamma}}$ (a dot denoting $d/du$). This proves (ii). When $v = 0$, $\boldsymbol{\varphi} = (\dot f, \dot g, i\sqrt{\dot f^2 + \dot g^2})$. Using Eq. 12.26, $\mathbf{N} = (-\dot g, \dot f, 0)/\sqrt{\dot f^2 + \dot g^2})$. Then, $\mathbf{N} \times \dot{\boldsymbol{\gamma}} = (0, 0, -\sqrt{\dot f^2 + \dot g^2})$ and finally $\ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) = 0$. It follows that $\boldsymbol{\gamma}$ is a pre-geodesic on $\boldsymbol{\sigma}$ (Exercise 9.1.2). If $\boldsymbol{\gamma}$ is the cycloid, $F(z) = z - \sin z, G(z) = 1 - \cos z$, so $\boldsymbol{\sigma}_u - i\boldsymbol{\sigma}_v = \boldsymbol{\varphi} = (1 - \cos z, \sin z, 2i \sin z)$. This gives

$$\begin{aligned}
\boldsymbol{\sigma}_u &= (1 - \cos u \cosh v, \sin u \cosh v, -2\cos \tfrac{u}{2} \sinh \tfrac{v}{2}) \\
\boldsymbol{\sigma}_v &= (-\sin u \sinh v, -\cos u \sinh v, -2\sin \tfrac{u}{2} \cosh \tfrac{v}{2}).
\end{aligned}$$

Integrating gives

$$\boldsymbol{\sigma}(u,v) = (u - \sin u \cosh v, -\cos u \cosh v, -4\sin \frac{u}{2} \sinh \frac{v}{2}),$$

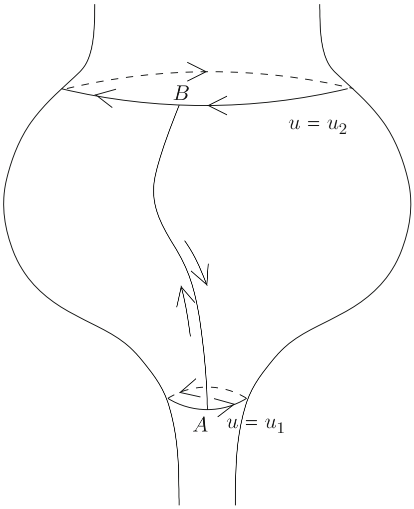
up to a translation. Translating by $(0,1,0)$ gives Catalan's surface.

# Chapter 13

13.1.1 If $\boldsymbol{\gamma}$ is a *simple* closed geodesic, Theorem 13.1.2 gives $\int_{\mathrm{int}(\boldsymbol{\gamma})} K d\mathcal{A} = 2\pi$; since $K \leq 0$, this is impossible. The parallels of a cylinder are not the images under a surface patch $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ of a *simple* closed curve $\boldsymbol{\pi}$ in the plane such that $\mathrm{int}(\boldsymbol{\pi})$ is contained in $U$. Note that the whole cylinder can actually be covered by a single patch (see Exercise 4.1.4) in which $U$ is an annulus, but that the parallels correspond to circles going 'around the hole' in the annulus.

13.1.2 Since the unit normal $\mathbf{N}$ of $S^2$ is equal to $\pm\mathbf{n}$, the geodesic curvature $\kappa_g$ of $\mathbf{n}$ is, up to a sign, $\mathbf{n}'' \cdot (\mathbf{n} \times \mathbf{n}')$. Let $t$ be the arc-length of $\boldsymbol{\gamma}$ and denote

$d/dt$ by a dot. Then, $ds/dt = \parallel \dot{\mathbf{n}} \parallel = \parallel -\kappa\mathbf{t} + \tau\mathbf{b} \parallel = \sqrt{\kappa^2 + \tau^2} = R$, say, where $\mathbf{t} = \dot{\boldsymbol{\gamma}}$. Then, $\mathbf{n}' = (-\kappa\mathbf{t} + \tau\mathbf{b})/R$, $\mathbf{n} \times \mathbf{n}' = (\kappa\mathbf{b} + \tau\mathbf{t})/R$, and $\mathbf{n}'' = \frac{1}{R}\frac{d}{dt}\left(\frac{-\kappa\mathbf{t}+\tau\mathbf{b}}{R}\right) = -R^{-1}(\kappa/R)^{\cdot}\mathbf{t} + R^{-1}(\tau/R)^{\cdot}\mathbf{b} - R^{-2}(\kappa^2 + \tau^2)\mathbf{n}$. These formulas give $\mathbf{n}'' \cdot (\mathbf{n} \times \mathbf{n}') = -R^{-2}\tau(\kappa/R)^{\cdot} + R^{-2}\kappa(\tau/R)^{\cdot} = (\kappa\dot{\tau} - \tau\dot{\kappa})/R^3$. Since $\dot{\kappa} = R\kappa'$, etc., $\kappa_g = \pm\frac{\kappa\tau' - \tau\kappa'}{\kappa^2+\tau^2} = \pm\frac{d}{ds}\tan^{-1}\frac{\tau}{\kappa}$. Applying Theorem 13.1.2 to the curve $\mathbf{n}$ on $S^2$, and noting that $K = 1$ for $S^2$ and that $\int_0^{\ell(\mathbf{n})} \kappa_g \, dt = 0$ because $\kappa_g$ is the derivative of an $\ell(\mathbf{n})$-periodic function (where $\ell(\mathbf{n})$ is the length of the closed curve $\mathbf{n}$), we get that the area inside $\mathbf{n}$ is $2\pi$.



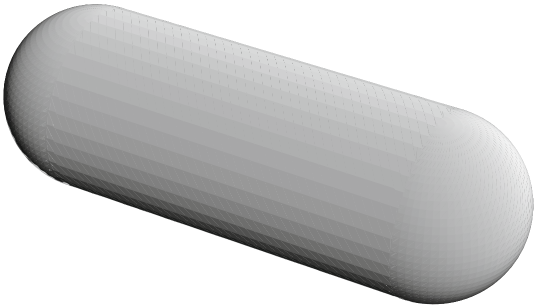

13.2.1 The parallel $u = u_1$ is the circle $\boldsymbol{\gamma}_1(v) = (f(u_1)\cos v, f(u_1)\sin v, g(u_1))$; if $s$ is the arc-length of $\boldsymbol{\gamma}_1$, $ds/dv = f(u_1)$. Denote $d/ds$ by a dot and $d/du$ by a dash. Then, $\dot{\boldsymbol{\gamma}} = (-\sin v, \cos v, 0)$, $\ddot{\boldsymbol{\gamma}} = -\frac{1}{f(u_1)}(\cos v, \sin v, 0)$, and the unit normal of the surface is $\mathbf{N} = (-g'\cos v, -g'\sin v, f')$. This gives the geodesic curvature of $\boldsymbol{\gamma}$ as $\kappa_g = \ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) = \frac{f'(u_1)}{f(u_1)}$. Since $\ell(\boldsymbol{\gamma}_1) = 2\pi f(u_1)$, $\int_0^{\ell(\boldsymbol{\gamma}_1)} \kappa_g \, ds = 2\pi f'(u_1)$. Similarly for $\boldsymbol{\gamma}_2$. By Example 8.1.4, $K = -f''/f$, so $\int_R K d\mathcal{A}_{\boldsymbol{\sigma}} = \int_0^{2\pi}\int_{u_1}^{u_2} -\frac{f''}{f} f du dv = 2\pi(f'(u_1) - f'(u_2))$. Hence, $\int_0^{\ell(\boldsymbol{\gamma}_1)} \kappa_g \, ds - \int_0^{\ell(\boldsymbol{\gamma}_2)} \kappa_g \, ds = \int_R K d\mathcal{A}_{\boldsymbol{\sigma}}$. This equation is the result of applying Theorem 13.2.2 to the curvilinear polygon shown above.

13.3.1 This can be proved by expressing $\mathbf{a}, \mathbf{b}, \mathbf{c}$ in component form and computing both sides. Alternatively, one may observe that both sides of the equation are linear in each of $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ (separately), and change sign when any two of $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ are interchanged. This means that it is enough to prove the formula when $\mathbf{a} = \mathbf{i}, \mathbf{b} = \mathbf{j}, \mathbf{c} = \mathbf{k}$, when both sides are obviously equal to 1.

13.3.2 Define $\varphi_k$ as in the hint. The stated properties are easily checked.

13.4.1 By Corollary 13.4.8, $\int_{\mathcal{S}} K d\mathcal{A} = 4\pi(1-g)$, and $g = 1$ since $\mathcal{S}$ is diffeomorphic to $T_1$. By Proposition 8.6.1, $K > 0$ at some point of $\mathcal{S}$.

13.4.2 $K > 0 \Longrightarrow \int_{\mathcal{S}} K d\mathcal{A} > 0 \Longrightarrow g < 1$ by Corollary 13.4.8; since $g$ is a non-negative integer, $g = 0$ so $\mathcal{S}$ is diffeomorphic to a sphere. The converse is false: for example, a 'cigar tube' is diffeomorphic to a sphere but $K = 0$ on the cylindrical part.



13.5.1 Assuming that every country has $\geq$ six neighbours, the argument in the proof of Theorem 13.5.1 gives $E \geq 3F$ and $2E \geq 3V$, so $V - E + F \leq \frac{2E}{3} - E + \frac{E}{3} = 0$, contradicting $V - E + F = 2$.

13.5.2 $3F = 2E$ because each face has three edges and each edge is an edge of two faces. From $\chi = V - E + F$, we get $\chi = V - E + \frac{2}{3}E$, so $E = 3(V - \chi)$. Since each edge has two vertices and two edges cannot intersect in more than one vertex, $E \leq \frac{1}{2}V(V-1)$; hence, $3(V-\chi) \leq \frac{1}{2}V(V-1)$, which is equivalent to $V^2 - 7V + 6\chi \geq 0$. The roots of the quadratic are $\frac{1}{2}\left(7 \pm \sqrt{49 - 7\chi}\right)$, so $V \leq \frac{1}{2}\left(7 - \sqrt{49-7\chi}\right)$ or $V \geq \frac{1}{2}\left(7 + \sqrt{49 - 7\chi}\right)$. Since $\chi = 2, 0, -2, \ldots$, the first condition gives $V \leq 3$, which would allow only one triangle; hence, the second condition must hold.

13.6.1 The circle $\theta = \theta_0$ is a circle in the plane $z = b\sin\theta_0$ with centre on the $z$-axis, so its principal normal is a unit vector perpendicular to the circle and in this plane, hence equal (up to a sign) to $(\cos\varphi, \sin\varphi, 0)$. The unit normal of $\boldsymbol{\sigma}$ is $\mathbf{N} = (-\cos\theta\cos\varphi, -\cos\theta\sin\varphi, -\sin\theta)$, so the angle between $\mathbf{N}$ and $\mathbf{n}$ at a point of the circle $\theta = \theta_0$ is $\theta_0$. The radius of the circle is $a + b\cos\theta_0$, so its geodesic curvature is $\frac{\sin\theta_0}{a + b\cos\theta_0}$. Hence, $\int \kappa_g ds = 2\pi\sin\theta_0$ and the holonomy is $2\pi - 2\pi\sin\theta_0$. The circles $\varphi =$ constant are geodesics (as they are meridians on a surface of revolution) so $\kappa_g = 0$ and the holonomy is $2\pi - 0 = 2\pi$.

13.6.2 For the circle $v = 1$ on the cone, the radius is 1 and the angle between the principal normal $\mathbf{n}$ of the circle and the unit normal $\mathbf{N}$ of the cone is $\pi/4$, so the holonomy is $2\pi - 2\pi/\sqrt{2} = (2-\sqrt{2})\pi$. The cone is flat so if the converse of Proposition 13.6.5 were true the holonomy around any closed curve on the cone would be zero.

13.7.1 Take the reference tangent vector field to be $\boldsymbol{\xi} = (1, 0)$, and take the simple closed curve $\boldsymbol{\gamma}(s) = (\cos s, \sin s)$. At $\boldsymbol{\gamma}(s)$, we have $\mathbf{V} = (\alpha, \beta)$, where

$$\alpha + i\beta = \begin{cases} (\cos s + i \sin s)^k & \text{if } k > 0, \\ (\cos s - i \sin s)^{-k} & \text{if } k < 0. \end{cases}$$

By de Moivre's theorem, $\alpha = \cos ks$, $\beta = \sin ks$ in both cases. Hence, the angle $\psi$ between $\mathbf{V}$ and $\boldsymbol{\xi}$ is equal to $ks$, and Definition 13.7.2 shows that the multiplicity is $k$.

13.7.2 If $\boldsymbol{\sigma}(u, v) = \tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v})$, where $(\tilde{u}, \tilde{v}) \mapsto (u, v)$ is a reparametrization map, then $\mathbf{V} = \alpha\boldsymbol{\sigma}_u + \beta\boldsymbol{\sigma}_v = \tilde{\alpha}\tilde{\boldsymbol{\sigma}}_{\tilde{u}} + \tilde{\beta}\tilde{\boldsymbol{\sigma}}_{\tilde{v}} \implies \tilde{\alpha} = \alpha\frac{\partial\tilde{u}}{\partial u} + \beta\frac{\partial\tilde{v}}{\partial u}$, $\tilde{\beta} = \alpha\frac{\partial\tilde{u}}{\partial v} + \beta\frac{\partial\tilde{v}}{\partial v}$. Hence, $\tilde{\alpha}$ and $\tilde{\beta}$ are smooth if $\alpha$ and $\beta$ are smooth. Since the components of the vectors $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are smooth, if $\mathbf{V}$ is smooth so are its components. If the components of $\mathbf{V} = \alpha\boldsymbol{\sigma}_v + \beta\boldsymbol{\sigma}_v$ are smooth, then $\mathbf{V}\cdot\boldsymbol{\sigma}_u$ and $\mathbf{V}\cdot\boldsymbol{\sigma}_v$ are smooth functions, hence $\alpha = \frac{G(\mathbf{V}\cdot\boldsymbol{\sigma}_u) - F(\mathbf{V}\cdot\boldsymbol{\sigma}_v)}{EG-F^2}$, $\beta = \frac{E(\mathbf{V}\cdot\boldsymbol{\sigma}_v) - F(\mathbf{V}\cdot\boldsymbol{\sigma}_u)}{EG-F^2}$ are smooth functions, so $\mathbf{V}$ is smooth.

13.7.3 If $\tilde{\psi}$ is the angle between $\mathbf{V}$ and $\tilde{\boldsymbol{\xi}}$, we have $\tilde{\psi} - \psi = \theta$ (up to multiples of $2\pi$); so we must show that $\int_0^{\ell(\boldsymbol{\gamma})} \dot{\theta}\, ds = 0$ (a dot denotes $d/ds$). This is not obvious since $\theta$ is not a well-defined smooth function of $s$ (although $d\theta/ds$ is well defined). However, $\rho = \cos\theta$ is well defined and smooth, since $\rho = \boldsymbol{\xi}.\tilde{\boldsymbol{\xi}}/ \parallel \boldsymbol{\xi} \parallel\parallel \tilde{\boldsymbol{\xi}} \parallel$. Now, $\dot{\rho} = -\dot{\theta}\sin\theta$, so we must prove that $\int_0^{\ell(\boldsymbol{\gamma})} \frac{\dot{\rho}}{\sqrt{1-\rho^2}} ds = 0$. Using Green's theorem, this integral is equal to

$$\int_{\boldsymbol{\pi}} \frac{\rho_u du + \rho_v dv}{\sqrt{1-\rho^2}} = \int_{\mathrm{int}(\boldsymbol{\pi})} \left( \frac{\partial}{\partial u}\left(\frac{\rho_v}{\sqrt{1-\rho^2}}\right) - \frac{\partial}{\partial v}\left(\frac{\rho_u}{\sqrt{1-\rho^2}}\right)\right) dudv,$$

where $\boldsymbol{\pi}$ is the curve in $U$ such that $\boldsymbol{\gamma}(s) = \boldsymbol{\sigma}(\boldsymbol{\pi}(s))$; and this line integral vanishes because

$$\frac{\partial}{\partial u}\left(\frac{\rho_v}{\sqrt{1-\rho^2}}\right) = \frac{\partial}{\partial v}\left(\frac{\rho_u}{\sqrt{1-\rho^2}}\right) \quad \left(= \frac{\rho_{uv}(1-\rho^2) + \rho\rho_u\rho_v}{(1-\rho^2)^{3/2}}\right).$$

13.8.1 Let $F : \mathcal{S} \to \mathbb{R}$ be a smooth function on a surface $\mathcal{S}$, let $\mathbf{p} \in \mathcal{S}$, let $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ be patches of $\mathcal{S}$ containing $\mathbf{p}$, say $\boldsymbol{\sigma}(u_0, v_0) = \tilde{\boldsymbol{\sigma}}(\tilde{u}_0, \tilde{v}_0) = \mathbf{p}$, and let $f = F \circ \boldsymbol{\sigma}$ and $\tilde{f} = F \circ \tilde{\boldsymbol{\sigma}}$. Then, $\tilde{f}_{\tilde{u}} = f_u\frac{\partial u}{\partial\tilde{u}} + f_v\frac{\partial v}{\partial\tilde{u}}$, $\tilde{f}_{\tilde{v}} = f_u\frac{\partial u}{\partial\tilde{v}} + f_v\frac{\partial v}{\partial\tilde{v}}$, so if $f_u = f_v = 0$ at $(u_0, v_0)$, then $\tilde{f}_{\tilde{u}} = \tilde{f}_{\tilde{v}} = 0$ at $(\tilde{u}_0, \tilde{v}_0)$.

Since $f_u = f_v = 0$ at $\mathbf{p}$, we have $\tilde{f}_{\tilde{u}\tilde{u}} = f_{uu}\left(\frac{\partial u}{\partial \tilde{u}}\right)^2 + 2f_{uv}\frac{\partial u}{\partial \tilde{u}}\frac{\partial v}{\partial \tilde{u}} + f_{vv}\left(\frac{\partial v}{\partial \tilde{u}}\right)^2$, with similar expressions for $\tilde{f}_{\tilde{u}\tilde{v}}$ and $\tilde{f}_{\tilde{v}\tilde{v}}$. This gives, in an obvious notation, $\tilde{\mathcal{H}} = J^t\mathcal{H}J$, where $J = \begin{pmatrix} \frac{\partial u}{\partial \tilde{u}} & \frac{\partial u}{\partial \tilde{v}} \\ \frac{\partial v}{\partial \tilde{u}} & \frac{\partial v}{\partial \tilde{v}} \end{pmatrix}$ is the Jacobian matrix of the reparametrization map $(\tilde{u}, \tilde{v}) \mapsto (u, v)$.

Since $J$ is invertible, $\tilde{\mathcal{H}}$ is invertible if $\mathcal{H}$ is invertible. Since the matrix $\mathcal{H}$ is real and symmetric, it has eigenvectors $\mathbf{v}_1, \mathbf{v}_2$, with eigenvalues $\lambda_1, \lambda_2$, say, such that $\mathbf{v}_i^t\mathbf{v}_j = 1$ if $i = j$ and $= 0$ if $i \neq j$. Then, if $\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2$ is any vector, where $\alpha_1, \alpha_2$ are scalars, $\mathbf{v}^t\mathcal{H}\mathbf{v} = \lambda_1\alpha_1^2 + \lambda_2\alpha_2^2$; hence, $\mathbf{v}^t\mathcal{H}\mathbf{v} > 0$ (resp. $< 0$) for all $\mathbf{v} \neq \mathbf{0} \iff \lambda_1$ and $\lambda_2$ are both $> 0$ (resp. both $< 0$) $\iff \mathbf{p}$ is a local minimum (resp. local maximum); and hence $\mathbf{p}$ is a saddle point $\iff \mathbf{v}^t\mathcal{H}\mathbf{v}$ can be both $> 0$ and $< 0$, depending on the choice of $\mathbf{v}$. Since $J$ is invertible, a vector $\tilde{\mathbf{v}} \neq \mathbf{0} \iff \mathbf{v} = J\tilde{\mathbf{v}} \neq \mathbf{0}$; and $\tilde{\mathbf{v}}^t\tilde{\mathcal{H}}\tilde{\mathbf{v}} = \tilde{\mathbf{v}}^t J^t\mathcal{H}J\tilde{\mathbf{v}} = \mathbf{v}^t\mathcal{H}\mathbf{v}$. The assertions in the last sentence of the exercise follow from this.

13.8.2 (i) $f_x = 2x - 2y$, $f_y = -2x + 8y$, so $f_x = f_y = 0$ at the origin. $f_{xx} = 2$, $f_{xy} = -2, f_{yy} = 8$, so $\mathcal{H} = \begin{pmatrix} 2 & -2 \\ -2 & 8 \end{pmatrix}$. $\mathcal{H}$ is invertible so the origin is non-degenerate; and the eigenvalues $5 \pm \sqrt{13}$ of $\mathcal{H}$ are both $> 0$, so it is a local minimum. (ii) $f_x = f_y = 0$ and $\mathcal{H} = \begin{pmatrix} 2 & 4 \\ 4 & 0 \end{pmatrix}$ at the origin; $\det\mathcal{H} = -16 < 0$, so the eigenvalues of $\mathcal{H}$ are of opposite sign and the origin is a saddle point. (iii) $f_x = f_y = 0$ and $\mathcal{H} = 0$ at the origin, which is therefore a degenerate critical point.

13.8.3 Using the parametrization $\boldsymbol{\sigma}$ in Exercise 4.2.5 (with $a = 2, b = 1$) gives $f(\theta, \varphi) = F(\boldsymbol{\sigma}(\theta, \varphi)) = (2 + \cos\theta)\cos\varphi + 3$. Then, $f_\theta = -\sin\theta\cos\varphi$, $f_\varphi = -(2 + \cos\theta)\sin\varphi$; since $2 + \cos\theta > 0$, $f_\varphi = 0 \implies \varphi = 0$ or $\pi$, and then $f_\theta = 0 \implies \theta = 0$ or $\pi$; so there are four critical points, $\mathbf{p} = (3, 0, 0)$, $\mathbf{q} = (1, 0, 0)$, $\mathbf{r} = (-1, 0, 0)$ and $\mathbf{s} = (-3, 0, 0)$. Next, $\mathcal{H} = \begin{pmatrix} -\cos\theta\cos\varphi & \sin\theta\sin\varphi \\ \sin\theta\sin\varphi & -(2 + \cos\theta)\cos\varphi \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -3 \end{pmatrix}$ at $\mathbf{p}$, $= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ at $\mathbf{q}$, $= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ at $R$, and $= \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$ at $\mathbf{s}$; hence, $\mathbf{p}$ is a local maximum, $\mathbf{q}$ and $\mathbf{r}$ are saddle points, and $\mathbf{s}$ is a local minimum (all of which is geometrically obvious).

# Index