DIFFERENTIAL GEOMETRY OF CURVES & SURFACES

REVISED & UPDATED SECOND EDITION

MANFREDO P. DO CARMO

DIFFERENTIAL GEOMETRY OF CURVES & SURFACES

DIFFERENTIAL GEOMETRY OF CURVES & SURFACES

REVISED & UPDATED SECOND EDITION

MANFREDO P. DO CARMO
Instituto Nacional de Matemática
Pura e Aplicada (IMPA)
Rio de Janeiro, Brazil

DOVER PUBLICATIONS, INC.
Mineola, New York

Bibliographical Note

Differential Geometry of Curves and Surfaces: Revised & Updated Second Edition is a revised, corrected, and updated second edition of the work originally published in 1976 by Prentice-Hall, Inc., Englewood Cliffs, New Jersey. The author has also provided a new Preface for this edition.

International Standard Book Number
ISBN-13: 978-0-486-80699-0
ISBN-10: 0-486-80699-5

Manufactured in the United States by LSC Communications
80699501 2016
www.doverpublications.com

To Leny,
for her indispensable assistance
in all the stages of this book

Contents

Preface to the Second Edition

In this edition, I have included many of the corrections and suggestions kindly sent to me by those who have used the book. For several reasons it is impossible to mention the names of all the people who generously donated their time doing that. Here I would like to express my deep appreciation and thank them all.

Thanks are also due to John Grafton, Senior Acquisitions Editor at Dover Publications, who believed that the book was still valuable and included in the text all of the changes I had in mind, and to the editor, James Miller, for his patience with my frequent requests.

As usual, my wife, Leny A. Cavalcante, participated in the project as if it was a work of her own; and I might say that without her this volume would not exist.

Finally, I would like to thank my son, Manfredo Jr., for helping me with several figures in this edition.

Manfredo P. do Carmo
September 20, 2016

Preface

This book is an introduction to the differential geometry of curves and surfaces, both in its local and global aspects. The presentation differs from the traditional ones by a more extensive use of elementary linear algebra and by a certain emphasis placed on basic geometrical facts, rather than on machinery or random details.

We have tried to build each chapter of the book around some simple and fundamental idea. Thus, Chapter 2 develops around the concept of a regular surface in R^3; when this concept is properly developed, it is probably the best model for differentiable manifolds. Chapter 3 is built on the Gauss normal map and contains a large amount of the local geometry of surfaces in R^3. Chapter 4 unifies the intrinsic geometry of surfaces around the concept of covariant derivative; again, our purpose was to prepare the reader for the basic notion of connection in Riemannian geometry. Finally, in Chapter 5, we use the first and second variations of arc length to derive some global properties of surfaces. Near the end of Cbapter 5 (Sec. 5-10), we show how questions on surface theory, and the experience of Chapters 2 and 4, lead naturally to the consideration of differentiable manifolds and Riemannian metrics.

To maintain the proper balance between ideas and facts, we have presented a large number of examples that are computed in detail. Furthermore, a reasonable supply of exercises is provided. Some factual material of classical differential geometry found its place in these exercises. Hints or answers are given for the exercises that are starred.

The prerequisites for reading this book are linear algebra and calculus. From linear algebra, only the most basic concepts are needed, and a standard undergraduate course on the subject should suffice. From calculus, a certain familiarity with calculus of several variables (including the statement of the implicit function theorem) is expected. For the reader's convenience, we have tried to restrict our references to R. C. Buck, *Advanced Calculus*, New York: McGraw-Hill, 1965 (quoted as Buck, *Advanced Calculus*). A certain knowledge of differential equations will be useful but it is not required.

This book is a free translation, with additional material, of a book and a set of notes, both published originally in Portuguese. Were it not for the enthusiasm and enormous help of Blaine Lawson, this book would not have come into English. A large part of the translation was done by Leny Cavalcante. I am also indebted to my colleagues and students at IMPA for their comments and support. In particular, Elon Lima read part of the Portuguese version and made valuable comments.

Robert Gardner, Jürgen Kern, Blaine Lawson, and Nolan Wallach read critically the English manuscript and helped me to avoid several mistakes, both in English and Mathematics. Roy Ogawa prepared the computer programs for some beautiful drawings that appear in the book (Figs. 1-3, 1-8, 1-9, 1-10, 1-11, 3-45 and 4-4). Jerry Kazdan devoted his time generously and literally offered hundreds of suggestions for the improvement of the manuscript. This final form of the book has benefited greatly from his advice. To all these people—and to Arthur Wester, Editor of Mathematics at Prentice-Hall, and Wilson Góes at IMPA—I extend my sincere thanks.

Rio de Janeiro Manfredo P. do Carmo

Some Remarks on Using This Book

We tried to prepare this book so it could be used in more than one type of differential geometry course. Each chapter starts with an introduction that describes the material in the chapter and explains how this material will be used later in the book. For the reader's convenience, we have used footnotes to point out the sections (or parts thereof) that can be omitted on a first reading.

Although there is enough material in the book for a full-year course (or a topics course), we tried to make the book suitable for a first course on differential geometry for students with some background in linear algebra and advanced calculus.

For a short one-quarter course (10 weeks), we suggest the use of the following material: Chapter 1: Secs. 1-2, 1-3, 1-4, 1-5 and one topic of Sec. 1-7—2 weeks. Chapter 2: Secs. 2-2 and 2-3 (omit the proofs), Secs. 2-4 and 2-5—3 weeks. Chapter 3: Secs. 3-2 and 3-3—2 weeks. Chapter 4: Secs. 4-2 (omit conformal maps and Exercises 4, 13–18, 20), 4-3 (up to Gauss theorema egregium), 4-4 (u p to Prop. 4; omit Exercises 12, 13, 16, 18–21), 4-5 (up to the local Gauss-Bonnet theorem; include applications (b) and (f))—3 weeks.

The 10-week program above is on a pretty tight schedule. A more relaxed alternative is to allow more time for the first three chapters and to present survey lectures, on the last week of the course, on geodesics, the Gauss theorema egregium, and the Gauss-Bonnet theorem (geodesics can then be defined as curves whose osculating planes contain the normals to the surface).

In a one-semester course, the first alternative could be taught more leisurely and the instructor could probably include additional material (for instance, Secs. 5-2 and 5-10 (partially), or Secs. 4-6, 5-3 and 5-4).

Please also note that an asterisk attached to an exercise does not mean the exercise is either easy or hard. It only means that a solution or hint is provided at the end of the book. Second, we have used for parametrization a bold-faced $\mathbf{x}$ and that might become clumsy when writing on the blackboard. Thus we have reserved the capital X as a suggested replacement.

Where letter symbols that would normally be italic appear in italic context, the letter symbols are set in roman. This has been done to distinguish these symbols from the surrounding text.

1 Curves

1-1. Introduction

The differential geometry of curves and surfaces has two aspects. One, which may be called classical differential geometry, started with the beginnings of calculus. Roughly speaking, classical differential geometry is the study of local properties of curves and surfaces. By local properties we mean those properties which depend only on the behavior of the curve or surface in the neighborhood of a point. The methods which have shown themselves to be adequate in the study of such properties are the methods of differential calculus. Because of this, the curves and surfaces considered in differential geometry will be defined by functions which can be differentiated a certain number of times.

The other aspect is the so-called global differential geometry. Here one studies the influence of the local properties on the behavior of the entire curve or surface. We shall come back to this aspect of differential geometry later in the book.

Perhaps the most interesting and representative part of classical differential geometry is the study of surfaces. However, some local properties of curves appear naturally while studying surfaces. We shall therefore use this first chapter for a brief treatment of curves.

The chapter has been organized in such a way that a reader interested mostly in surfaces can read only Secs. 1-2 through 1-5. Sections 1-2 through 1-4 contain essentially introductory material (parametrized curves, arc length, vector product), which will probably be known from other courses and is included here for completeness. Section 1-5 is the heart of the chapter and contains the material of curves needed for the study of surfaces. For those wishing to go a bit further on the subject of curves, we have included Secs. 1-6 and 1-7.

1-2. Parametrized Curves

We denote by R^3 the set of triples (x, y, z) of real numbers. Our goal is to characterize certain subsets of R^3 (to be called curves) that are, in a certain sense, one-dimensional and to which the methods of differential calculus can be applied. A natural way of defining such subsets is through differentiable functions. We say that a real function of a real variable is *differentiable* (or *smooth*) if it has, at all points, derivatives of all orders (which are automatically continuous). A first definition of curve, not entirely satisfactory but sufficient for the purposes of this chapter, is the following.

DEFINITION. *A* parametrized differentiable curve *is a differentiable map* $\alpha\colon \mathrm{I} \to \mathrm{R}^3$ *of an open interval* $\mathrm{I} = (\mathrm{a}, \mathrm{b})$ *of the real line* R *into* R^3.†

The word *differentiable* in this definition means that α is a correspondence which maps each $t \in I$ into a point $\alpha(t) = (x(t), y(t), z(t)) \in R^3$ in such a way that the functions $x(t)$, $y(t)$, $z(t)$ are differentiable. The variable t is called the *parameter* of the curve. The word *interval* is taken in a generalized sense, so that we do not exclude the cases $a = -\infty$, $b = +\infty$.

If we denote by $x'(t)$ the first derivative of x at the point t and use similar notations for the functions y and z, the vector $(x'(t), y'(t), z'(t)) = \alpha'(t) \in R^3$ is called the *tangent vector* (or *velocity vector*) of the curve α at t. The image set $\alpha(I) \subset R^3$ is called the *trace* of α. As illustrated by Example 5 below, one should carefully distinguish a parametrized curve, which is a map, from its trace, which is a subset of R^3.

A warning about terminology. Many people use the term “infinitely differentiable” for functions which have derivatives of all orders and reserve the word “differentiable” to mean that only the existence of the first derivative is required. We shall not follow this usage.

Example 1. The parametrized differentiable curve given by

$$\alpha(t) = (a\cos t, a\sin t, bt), \qquad t \in R,$$

has as its trace in R^3 a helix of pitch $2\pi b$ on the cylinder $x^2 + y^2 = a^2$. The parameter t here measures the angle which the x axis makes with the line joining the origin 0 to the projection of the point $\alpha(t)$ over the xy plane (see Fig. 1-1).

†In italic context, letter symbols will not be italicized so they will be clearly distinguished from the surrounding text.

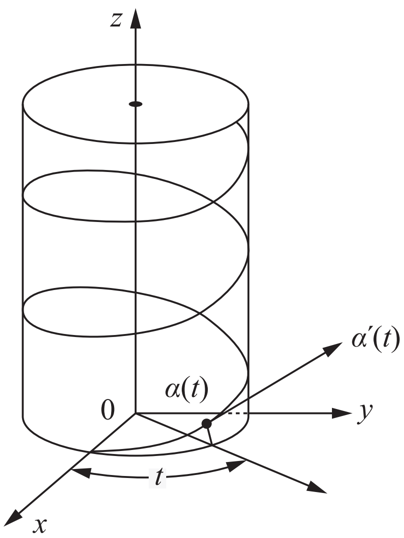


Figure 1-1

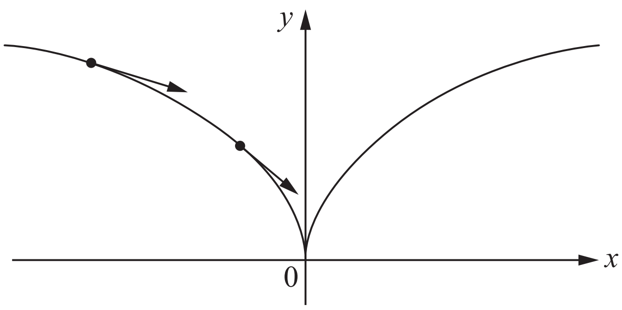


Figure 1-2

Example 2. The map $\alpha\colon R \to R^2$ given by $\alpha(t) = (t^3, t^2)$, $t \in R$, is a parametrized differentiable curve which has Fig. 1-2 as its trace. Notice that $\alpha'(0) = (0, 0)$; that is, the velocity vector is zero for $t = 0$.

Example 3. The map $\alpha\colon R \to R^2$ given by $\alpha(t) = (t^3 - 4t, t^2 - 4)$, $t \in R$, is a parametrized differentiable curve (see Fig. 1-3). Notice that $\alpha(2) = \alpha(-2) = (0, 0)$; that is, the map α is not one-to-one.

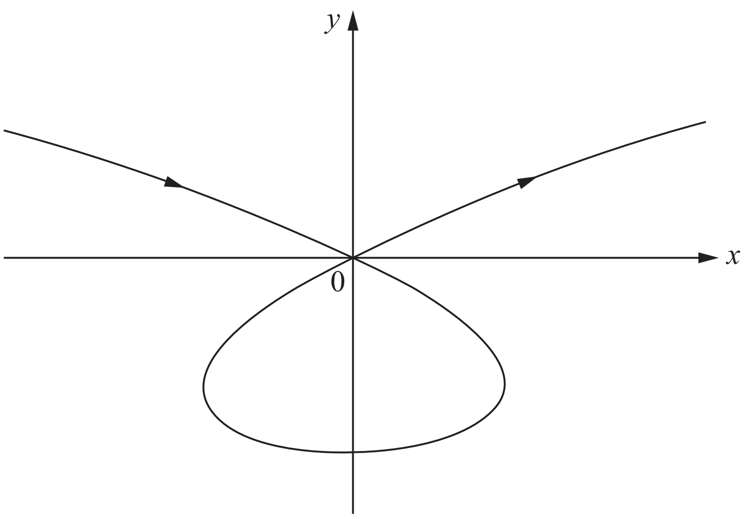


Figure 1-3

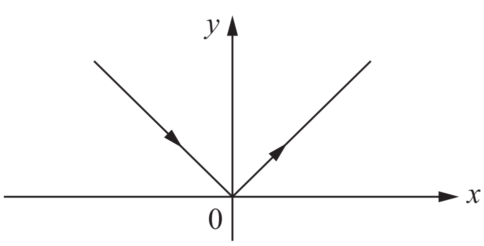


Figure 1-4

Example 4. The map $\alpha\colon R \to R^2$ given by $\alpha(t) = (t, |t|)$, $t \in R$, is *not* a parametrized differentiable curve, since $|t|$ is not differentiable at $t = 0$ (Fig. 1-4).

Example 5. The two distinct parametrized curves

$$\alpha(t) = (\cos t, \sin t),$$
$$\beta(t) = (\cos 2t, \sin 2t),$$

where $t \in (0 - \epsilon, 2\pi + \epsilon)$, $\epsilon > 0$, have the same trace, namely, the circle $x^2 + y^2 = 1$. Notice that the velocity vector of the second curve is the double of the first one (Fig. 1-5).

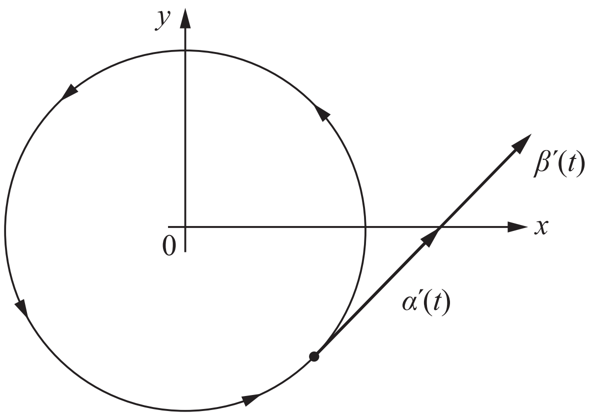


Figure 1-5

We shall now recall briefly some properties of the inner (or dot) product of vectors in R^3. Let $u = (u_1, u_2, u_3) \in R^3$ and define its *norm* (or *length*) by

$$|u| = \sqrt{u_1^2 + u_2^2 + u_3^2}.$$

Geometrically, $|u|$ is the distance from the point (u_1, u_2, u_3) to the origin $0 = (0, 0, 0)$. Now, let $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$ belong to R^3, and let θ, $0 \le \theta \le \pi$, be the angle formed by the segments $0u$ and $0v$. The *inner product* $u \cdot v$ is defined by (Fig. 1-6)

$$u \cdot v = |u||v| \cos\theta.$$

The following properties hold:

1. Assume that u and v are nonzero vectors. Then $u \cdot v = 0$ if and only if u is orthogonal to v.
2. $u \cdot v = v \cdot u$.
3. $\lambda(u \cdot v) = \lambda u \cdot v = u \cdot \lambda v$.
4. $u \cdot (v + w) = u \cdot v + u \cdot w$.

A useful expression for the inner product can be obtained as follows. Let $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$. It is easily checked

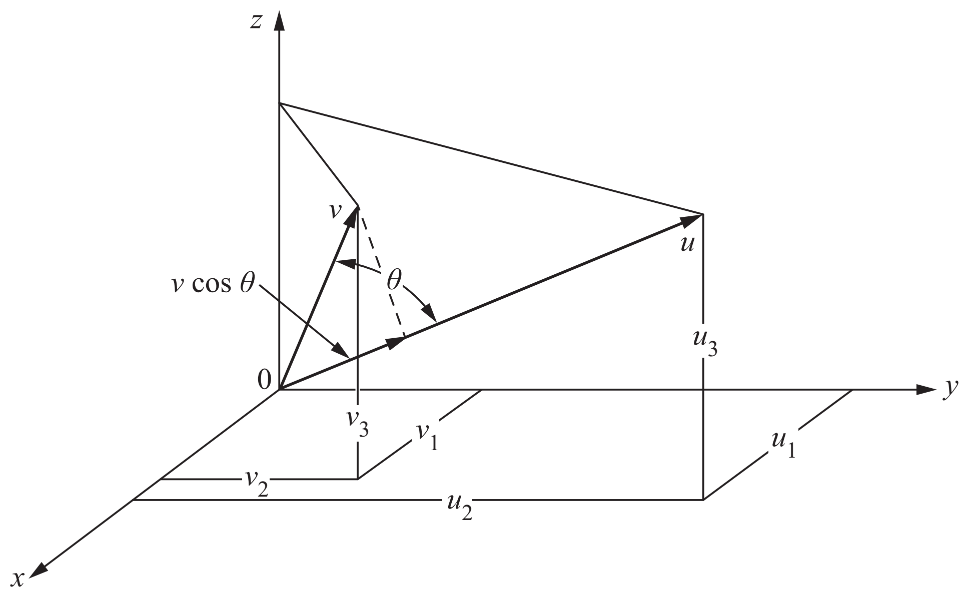


Figure 1-6

that $e_i \cdot e_j = 1$ if $i = j$ and that $e_i \cdot e_j = 0$ if $i \neq j$, where $i, j = 1, 2, 3$. Thus, by writing

$$u = u_1 e_1 + u_2 e_2 + u_3 e_3, \qquad v = v_1 e_1 + v_2 e_2 + v_3 e_3,$$

and using properties 2 to 4, we obtain

$$u \cdot v = u_1 v_1 + u_2 v_2 + u_3 v_3.$$

From the above expression it follows that if $u(t)$ and $v(t)$, $t \in I$, are differentiable curves, then $u(t) \cdot v(t)$ is a differentiable function, and

$$\frac{d}{dt}(u(t) \cdot v(t)) = u'(t) \cdot v(t) + u(t) \cdot v'(t).$$

EXERCISES

1. Find a parametrized curve $\alpha(t)$ whose trace is the circle $x^2 + y^2 = 1$ such that $\alpha(t)$ runs clockwise around the circle with $\alpha(0) = (0, 1)$.
2. Let $\alpha(t)$ be a parametrized curve which does not pass through the origin. If $\alpha(t_0)$ is a point of the trace of α closest to the origin and $\alpha'(t_0) \neq 0$, show that the position vector $\alpha(t_0)$ is orthogonal to $\alpha'(t_0)$.
3. A parametrized curve $\alpha(t)$ has the property that its second derivative $\alpha''(t)$ is identically zero. What can be said about α?
4. Let $\alpha\colon I \to R^3$ be a parametrized curve and let $v \in R^3$ be a fixed vector. Assume that $\alpha'(t)$ is orthogonal to v for all $t \in I$ and that $\alpha(0)$ is also orthogonal to v. Prove that $\alpha(t)$ is orthogonal to v for all $t \in I$.

5. Let $\alpha: I \to R^3$ be a parametrized curve, with $\alpha'(t) \neq 0$ for all $t \in I$. Show that $|\alpha(t)|$ is a nonzero constant if and only if $\alpha(t)$ is orthogonal to $\alpha'(t)$ for all $t \in I$.

1-3. Regular Curves; Arc Length

Let $\alpha: I \to R^3$ be a parametrized differentiable curve. For each $t \in I$ where $\alpha'(t) \neq 0$, there is a well-defined straight line, which contains the point $\alpha(t)$ and the vector $\alpha'(t)$. This line is called the *tangent line* to α at t. For the study of the differential geometry of a curve it is essential that there exists such a tangent line at every point. Therefore, we call any point t where $\alpha'(t) = 0$ a *singular point* of α and restrict our attention to curves without singular points. Notice that the point $t = 0$ in Example 2 of Sec. 1-2 is a singular point.

DEFINITION. *A parametrized differentiable curve* $\alpha: \mathrm{I} \to \mathrm{R}^3$ *is said to be* regular *if* $\alpha'(\mathrm{t}) \neq 0$ *for all* $\mathrm{t} \in \mathrm{I}$.

From now on we shall consider only regular parametrized differentiable curves (and, for convenience, shall usually omit the word differentiable).

Given $t_0 \in I$, the *arc length* of a regular parametrized curve $\alpha: I \to R^3$, from the point t_0, is by definition

$$s(t) = \int_{t_0}^{t} |\alpha'(t)|\, dt,$$

where

$$|\alpha'(t)| = \sqrt{(x'(t))^2 + (y'(t))^2 + (z'(t))^2}$$

is the length of the vector $\alpha'(t)$. Since $\alpha'(t) \neq 0$, the arc length s is a differentiable function of t and $ds/dt = |\alpha'(t)|$.

In Exercise 8 we shall present a geometric justification for the above definition of arc length.

It can happen that the parameter t is already the arc length measured from some point. In this case, $ds/dt = 1 = |\alpha'(t)|$; that is, the velocity vector has constant length equal to 1. Conversely, if $|\alpha'(t)| \equiv 1$, then

$$s = \int_{t_0}^{t} dt = t - t_0;$$

i.e., t is the arc length of α measured from some point.

To simplify our exposition, we shall restrict ourselves to curves parametrized by arc length; we shall see later (see Sec. 1-5) that this restriction is not essential. In general, it is not necessary to mention the origin of the arc length s, since most concepts are defined only in terms of the derivatives of $\alpha(s)$.

It is convenient to set still another convention. Given the curve α parametrized by arc length $s \in (a, b)$, we may consider the curve β defined in $(-b, -a)$ by $\beta(-s) = \alpha(s)$, which has the same trace as the first one but is described in the opposite direction. We say, then, that these two curves differ by a *change of orientation*.

EXERCISES

1. Show that the tangent lines to the regular parametrized curve $\alpha(t) = (3t, 3t^2, 2t^3)$ make a constant angle with the line $y = 0$, $z = x$.

2. A circular disk of radius 1 in the plane xy rolls without slipping along the x axis. The figure described by a point of the circumference of the disk is called a *cycloid* (Fig. 1-7).

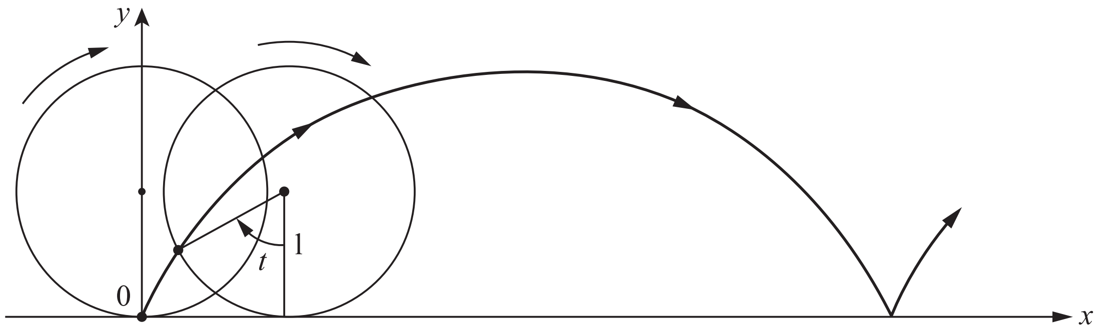


Figure 1-7. The cycloid.

***a.** Obtain a parametrized curve α: $R \to R^2$ the trace of which is the cycloid, and determine its singular points.

b. Compute the arc length of the cycloid corresponding to a complete rotation of the disk.

3. Let $0A = 2a$ be the diameter of a circle S^1 and $0y$ and AV be the tangents to S^1 at 0 and A, respectively. A half-line r is drawn from 0 which meets the circle S^1 at C and the line AV at B. On $0B$ mark off the segment $0p = CB$. If we rotate r about 0, the point p will describe a curve called the *cissoid of Diocles*. By taking $0A$ as the x axis and $0Y$ as the y axis, prove that

a. The trace of

$$\alpha(t) = \left(\frac{2at^2}{1+t^2}, \frac{2at^3}{1+t^2}\right), \qquad t \in R,$$

is the cissoid of Diocles ($t = \tan\theta$; see Fig. 1-8).

b. The origin (0, 0) is a singular point of the cissoid.

c. As $t \to \infty$, $\alpha(t)$ approaches the line $x = 2a$, and $\alpha'(t) \to 0, 2a$. Thus, as $t \to \infty$, the curve and its tangent approach the line $x = 2a$; we say that $x = 2a$ is an *asymptote* to the cissoid.

4. Let $\alpha\colon (0, \pi) \to R^2$ be given by

$$\alpha(t) = \left(\sin t, \cos t + \log \tan \frac{t}{2}\right),$$

where t is the angle that the y axis makes with the vector $\alpha'(t)$. The trace of α is called the *tractrix* (Fig. 1-9). Show that

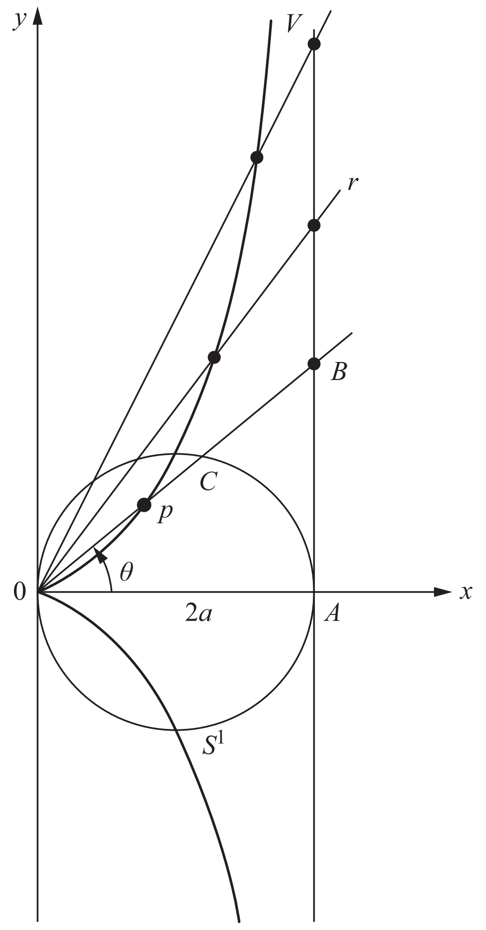


Figure 1-8. The cissoid of Diocles.

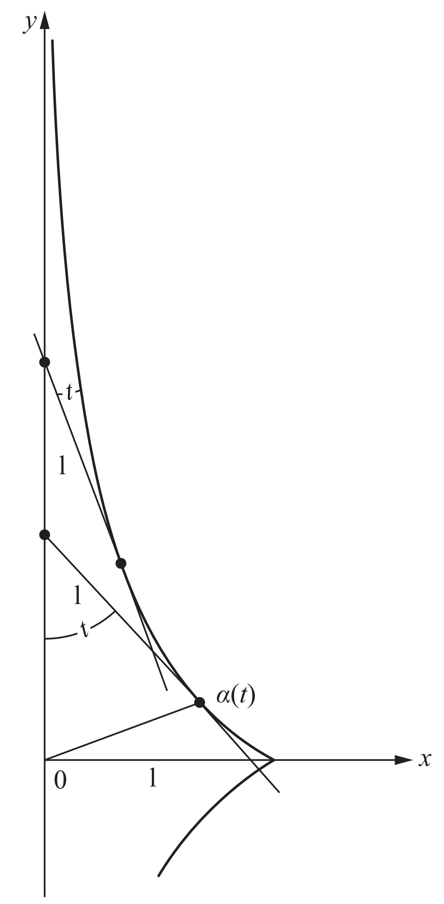


Figure 1-9. The tractrix.

a. α is a differentiable parametrized curve, regular except at $t = \pi/2$.

b. The length of the segment of the tangent of the tractrix between the point of tangency and the y axis is constantly equal to 1.

5. Let $\alpha\colon (-1, +\infty) \to R^2$ be given by

$$\alpha(t) = \left(\frac{3at}{1+t^3}, \frac{3at^2}{1+t^3}\right).$$

Prove that:

a. For $t = 0$, α is tangent to the x axis.

b. As $t \to +\infty$, $\alpha(t) \to (0, 0)$ and $\alpha'(t) \to (0, 0)$.

c. Take the curve with the opposite orientation. Now, as $t \to -1$, the curve and its tangent approach the line $x + y + a = 0$.

The figure obtained by completing the trace of α in such a way that it becomes symmetric relative to the line $y = x$ is called the *folium of Descartes* (see Fig. 1-10).

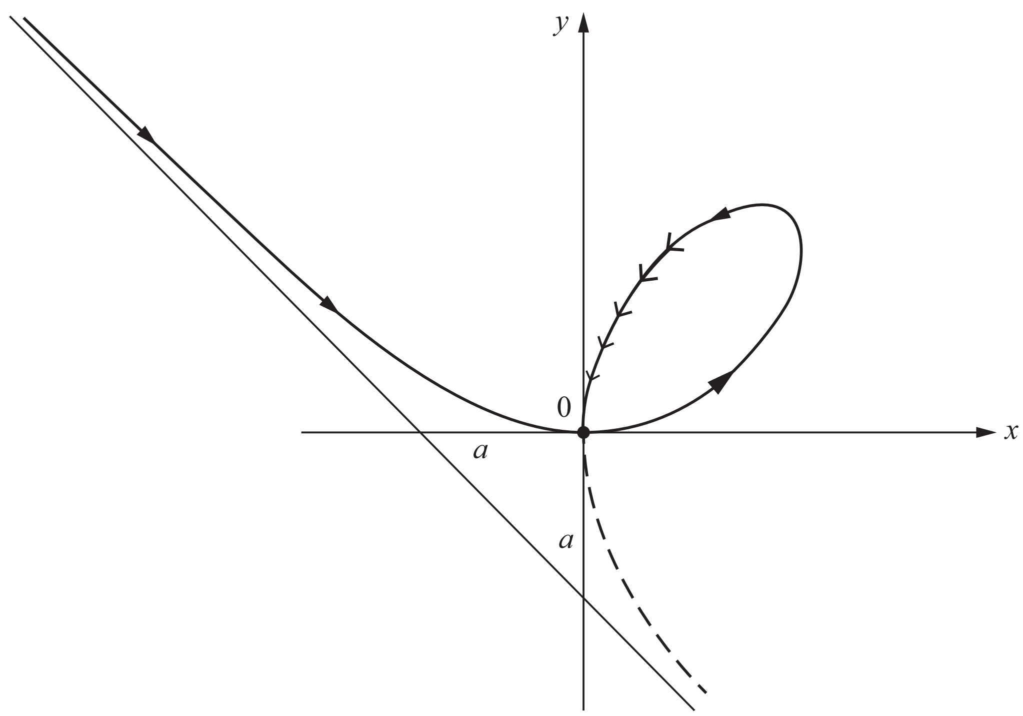


Figure 1-10. Folium of Descartes.

6. Let $\alpha(t) = (ae^{bt}\cos t, ae^{bt}\sin t)$, $t \in R$, a and b constants, $a > 0$, $b < 0$, be a parametrized curve.

a. Show that as $t \to +\infty$, $\alpha(t)$ approaches the origin 0, spiraling around it (because of this, the trace of α is called the *logarithmic spiral*; see Fig. 1-11).

b. Show that $\alpha'(t) \to (0, 0)$ as $t \to +\infty$ and that

$$\lim_{t\to+\infty} \int_{t_0}^{t} |\alpha'(t)|\, dt$$

is finite; that is, α has finite arc length in $[t_0, \infty)$.

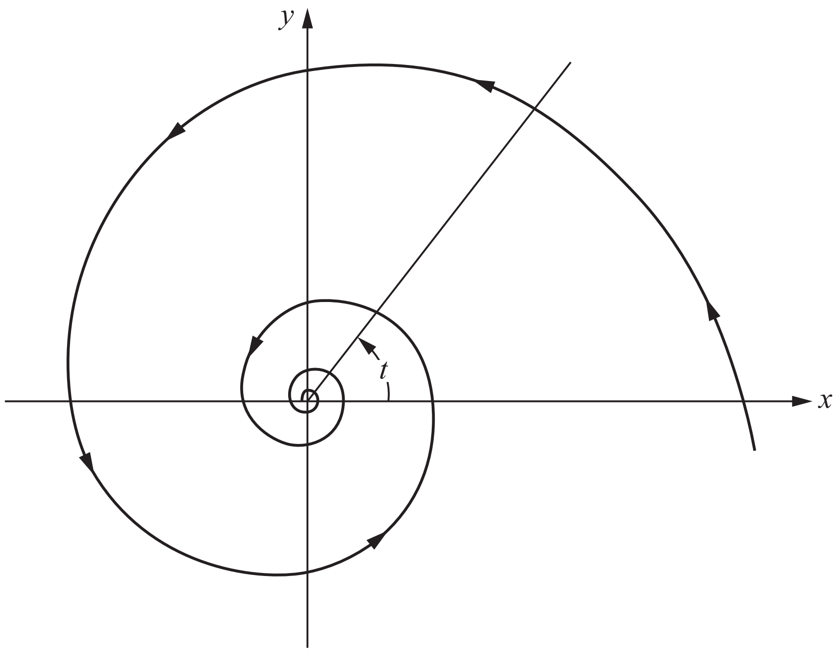


Figure 1-11. Logarithmic spiral.

7. A map α: $I \to R^3$ is *called a curve of class* C^k if each of the coordinate functions in the expression $\alpha(t) = (x(t), y(t), z(t))$ has continuous derivatives up to order k. If α is merely continuous, we say that α is of class C^0. A curve α is called *simple* if the map α is one-to-one. Thus, the curve in Example 3 of Sec. 1-2 is not simple.

Let α: $I \to R^3$ be a simple curve of class C^0. We say that α has a *weak tangent* at $t = t_0 \in I$ if the line determined by $\alpha(t_0 + h)$ and $\alpha(t_0)$ has a limit position when $h \to 0$. We say that α has a *strong tangent* at $t = t_0$ if the line determined by $\alpha(t_0 + h)$ and $\alpha(t_0 + k)$ has a limit position when $h, k \to 0$. Show that

a. $\alpha(t) = (t^3, t^2)$, $t \in R$, has a weak tangent but not a strong tangent at $t = 0$.

***b.** If α: $I \to R^3$ is of class C^1 and regular at $t = t_0$, then it has a strong tangent at $t = t_0$.

c. The curve given by

$$\alpha(t) = \begin{cases} (t^2, t^2), & t \geq 0, \\ (t^2, -t^2), & t \leq 0, \end{cases}$$

is of class C^1 but not of class C^2. Draw a sketch of the curve and its tangent vectors.

***8.** Let α: $I \to R^3$ be a differentiable curve and let $[a, b] \subset I$ be a closed interval. For every *partition*

$$a = t_0 < t_1 < \cdots < t_n = b$$

of $[a, b]$, consider the sum $\sum_{i=1}^{n} |\alpha(t_i) - \alpha(t_{i-1})| = l(\alpha, P)$, where P stands for the given partition. The norm $|P|$ of a partition P is defined as

$$|P| = \max(t_i - t_{i-1}), i = 1, \dots, n.$$

Geometrically, $l(\alpha, P)$ is the length of a polygon inscribed in $\alpha([a, b])$ with vertices in $\alpha(t_i)$ (see Fig. 1-12). The point of the exercise is to show that the arc length of $\alpha([a, b])$ is, in some sense, a limit of lengths of inscribed polygons.

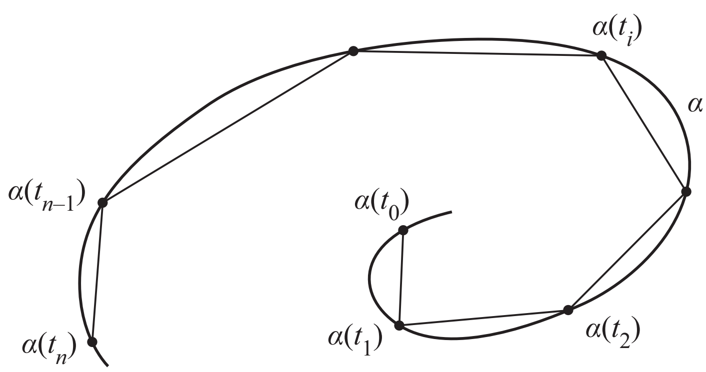


Figure 1-12

Prove that given $\epsilon > 0$ there exists $\delta > 0$ such that if $|P| < \delta$ then

$$\left| \int_a^b |\alpha'(t)|\, dt - l(\alpha, P) \right| < \epsilon.$$

9. a. Let $\alpha\colon I \to R^3$ be a curve of class C^0 (cf. Exercise 7). Use the approximation by polygons described in Exercise 8 to give a reasonable definition of arc length of α.

b. (*A Nonrectifiable Curve.*) The following example shows that, with any reasonable definition, the arc length of a C^0 curve in a closed interval may be unbounded. Let $\alpha\colon [0, 1] \to R^2$ be given as $\alpha(t) = (t, t\sin(\pi/t))$ if $t \neq 0$, and $\alpha(0) = (0, 0)$. Show, geometrically, that the arc length of the portion of the curve corresponding to $1/(n+1) \leq t \leq 1/n$ is at least $2/(n + \frac{1}{2})$. Use this to show that the length of the curve in the interval $1/N \leq t \leq 1$ is greater than $2\sum_{n=1}^{N} 1/(n+1)$, and thus it tends to infinity as $N \to \infty$.

10. (*Straight Lines as Shortest.*) Let $\alpha\colon I \to R^3$ be a parametrized curve. Let $[a, b] \subset I$ and set $\alpha(a) = p$, $\alpha(b) = q$.

a. Show that, for any constant vector v, $|v| = 1$,

$$(q - p) \cdot v = \int_a^b \alpha'(t) \cdot v\, dt \leq \int_a^b |\alpha'(t)|\, dt.$$

b. Set

$$v = \frac{q-p}{|q-p|}$$

and show that

$$|\alpha(b) - \alpha(a)| \leq \int_a^b |\alpha'(t)|\, dt;$$

that is, the curve of shortest length from $\alpha(a)$ to $\alpha(b)$ is the straight line joining these points.

1-4. The Vector Product in R^3

In this section, we shall present some properties of the vector product in R^3. They will be found useful in our later study of curves and surfaces.

It is convenient to begin by reviewing the notion of orientation of a vector space. Two ordered bases $e = \{e_i\}$ and $f = \{f_i\}$, $i = 1, \ldots, n$, of an n-dimensional vector space V have the *same orientation* if the matrix of change of basis has positive determinant. We denote this relation by $e \sim f$. From elementary properties of determinants, it follows that $e \sim f$ is an equivalence relation; i.e., it satisfies

1. $e \sim e$.
2. If $e \sim f$, then $f \sim e$.
3. If $e \sim f$, $f \sim g$, then $e \sim g$.

The set of all ordered bases of V is thus decomposed into equivalence classes (the elements of a given class are related by $\sim$) which by property 3 are disjoint. Since the determinant of a change of basis is either positive or negative, there are only two such classes.

Each of the equivalence classes determined by the above relation is called an *orientation* of V. Therefore, V has two orientations, and if we fix one of them arbitrarily, the other one is called the opposite orientation.

In the case $V = R^3$, there exists a natural ordered basis $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$, and we shall call the orientation corresponding to this basis the *positive orientation* of R^3, the other one being the *negative orientation* (of course, this applies equally well to any R^n). We also say that a given ordered basis of R^3 is *positive* (or *negative*) if it belongs to the positive (or negative) orientation of R^3. Thus, the ordered basis e_1, e_3, e_2 is a negative basis, since the matrix which changes this basis into e_1, e_2, e_3 has determinant equal to -1.

We now come to the vector product. Let $u, v \in R^3$. The *vector product* of u and v (in that order) is the unique vector $u \wedge v \in R^3$ characterized by

$$(u \wedge v) \cdot w = \det(u, v, w) \qquad \text{for all } w \in R^3.$$

Here $\det(u, v, w)$ means that if we express u, v, and w in the natural basis $\{e_i\}$,

$$u = \sum u_i e_i, \qquad v = \sum v_i e_i,$$
$$w = \sum w_i e_i, \qquad i = 1, 2, 3,$$

then

$$\det(u, v, w) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix},$$

where $|a_{ij}|$ denotes the determinant of the matrix (a_{ij}). It is immediate from the definition that

$$u \wedge v = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} e_1 - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} e_2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} e_3. \qquad \textbf{(1)}$$

Remark. It is also very frequent to write $u \wedge v$ as $u \times v$ and refer to it as the *cross product*.

The following properties can easily be checked (actually they just express the usual properties of determinants):

1. $u \wedge v = -v \wedge u$ (anticommutativity).
2. $u \wedge v$ depends linearly on u and v; i.e., for any real numbers a, b, we have
$$(au + bw) \wedge v = au \wedge v + bw \wedge v.$$
3. $u \wedge v = 0$ if and only if u and v are linearly dependent.
4. $(u \wedge v) \cdot u = 0$, $(u \wedge v) \cdot v = 0$.

It follows from property 4 that the vector product $u \wedge v \neq 0$ is normal to a plane generated by u and v. To give a geometric interpretation of its norm and its direction, we proceed as follows.

First, we observe that $(u \wedge v) \cdot (u \wedge v) = |u \wedge v|^2 > 0$. This means that the determinant of the vectors $u, v, u \wedge v$ is positive; that is, $\{u, v, u \wedge v\}$ is a positive basis.

Next, we prove the relation

$$(u \wedge v) \cdot (x \wedge y) = \begin{vmatrix} u \cdot x & v \cdot x \\ u \cdot y & v \cdot y \end{vmatrix},$$

where u, v, x, y are arbitrary vectors. This can easily be done by observing that both sides are linear in u, v, x, y. Thus, it suffices to check that

$$(e_i \wedge e_j) \cdot (e_k \wedge e_l) = \begin{vmatrix} e_i \cdot e_k & e_j \cdot e_k \\ e_i \cdot e_l & e_j \cdot e_l \end{vmatrix}$$

for all $i, j, k, l = 1, 2, 3$. This is a straightforward verification.

It follows that

$$|u \wedge v|^2 = \begin{vmatrix} u \cdot u & u \cdot v \\ u \cdot v & v \cdot v \end{vmatrix} = |u|^2|v|^2(1 - \cos^2\theta) = A^2,$$

where θ is the angle of u and v, and A is the area of the parallelogram generated by u and v.

In short, the vector product of u and v is a vector $u \wedge v$ perpendicular to a plane spanned by u and v, with a norm equal to the area of the parallelogram generated by u and v and a direction such that $\{u, v, u \wedge v\}$ is a positive basis (Fig. 1-13).

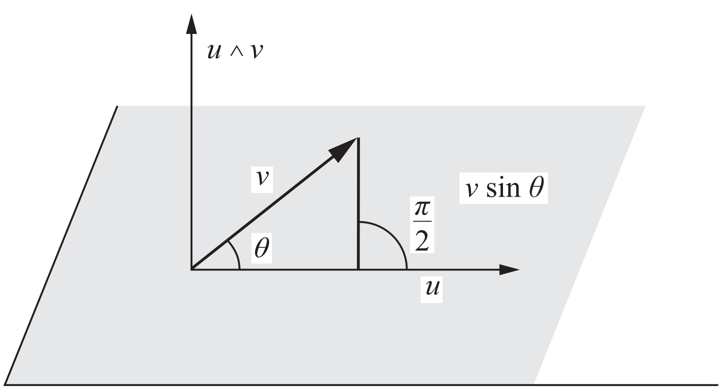


Figure 1-13

The vector product is not associative. In fact, we have the following identity:

$$(u \wedge v) \wedge w = (u \cdot w)v - (v \cdot w)u, \tag{2}$$

which can be proved as follows. First we observe that both sides are linear in u, v, w; thus, the identity will be true if it holds for all basis vectors. This last verification is, however, straightforward; for instance,

$$(e_1 \wedge e_2) \wedge e_1 = e_2 = (e_1 \cdot e_1)e_2 - (e_2 \cdot e_1)e_1.$$

Finally, let $u(t) = (u_1(t), u_2(t), u_3(t))$ and $v(t) = (v_1(t), v_2(t), v_3(t))$ be differentiable maps from the interval (a, b) to R^3, $t \in (a, b)$. It follows immediately from Eq. (1) that $u(t) \wedge v(t)$ is also differentiable and that

$$\frac{d}{dt}(u(t) \wedge v(t)) = \frac{du}{dt} \wedge v(t) + u(t) \wedge \frac{dv}{dt}.$$

Vector products appear naturally in many geometrical constructions. Actually, most of the geometry of planes and lines in R^3 can be neatly expressed in terms of vector products and determinants. We shall review some of this material in the following exercises.

EXERCISES

1. Check whether the following bases are positive:

a. The basis $\{(1, 3), (4, 2)\}$ in R^2.

b. The basis $\{(1, 3, 5), (2, 3, 7), (4, 8, 3)\}$ in R^3.

***2.** A plane P contained in R^3 is given by the equation $ax + by + cz + d = 0$. Show that the vector $v = (a, b, c)$ is perpendicular to the plane and that $|d|/\sqrt{a^2 + b^2 + c^2}$ measures the distance from the plane to the origin $(0, 0, 0)$.

***3.** Determine the angle of intersection of the two planes $5x + 3y + 2z - 4 = 0$ and $3x + 4y - 7z = 0$.

***4.** Given two planes $a_i x + b_i y + c_i z + d_i = 0$, $i = 1, 2$, prove that a necessary and sufficient condition for them to be parallel is

$$\frac{a_1}{a_2} = \frac{b_1}{b_2} = \frac{c_1}{c_2},$$

where the convention is made that if a denominator is zero, the corresponding numerator is also zero (we say that two planes are parallel if they either coincide or do not intersect).

5. Show that an equation of a plane passing through three noncolinear points $p_1 = (x_1, y_1, z_1)$, $p_2 = (x_2, y_2, z_2)$, $p_3 = (x_3, y_3, z_3)$ is given by

$$(p - p_1) \wedge (p - p_2) \cdot (p - p_3) = 0,$$

where $p = (x, y, z)$ is an arbitrary point of the plane and $p - p_1$, for instance, means the vector $(x - x_1, y - y_1, z - z_1)$.

***6.** Given two nonparallel planes $a_i x + b_i y + c_i z + d_i = 0$, $i = 1, 2$, show that their line of intersection may be parametrized as

$$x - x_0 = u_1 t, \quad y - y_0 = u_2 t, \quad z - z_0 = u_3 t,$$

where (x_0, y_0, z_0) belongs to the intersection and $u = (u_1, u_2, u_3)$ is the vector product $u = v_1 \wedge v_2$, $v_i = (a_i, b_i, c_i)$, $i = 1, 2$.

***7.** Prove that a necessary and sufficient condition for the plane

$$ax + by + cz + d = 0$$

and the line $x - x_0 = u_1 t$, $y - y_0 = u_2 t$, $z - z_0 = u_3 t$ to be parallel is

$$au_1 + bu_2 + cu_3 = 0.$$

***8.** Prove that the distance ρ between the nonparallel lines

$$\begin{aligned} x - x_0 &= u_1 t, \quad & y - y_0 &= u_2 t, \quad & z - z_0 &= u_3 t, \\ x - x_1 &= v_1 t, \quad & y - y_1 &= v_2 t, \quad & z - z_1 &= v_3 t \end{aligned}$$

is given by

$$\rho = \frac{|(u \wedge v) \cdot r|}{|u \wedge v|},$$

where $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3)$, $r = (x_0 - x_1, y_0 - y_1, z_0 - z_1)$.

9. Determine the angle of intersection of the plane $3x + 4y + 7z + 8 = 0$ and the line $x - 2 = 3t$, $y - 3 = 5t$, $z - 5 = 9t$.

10. The natural orientation of R^2 makes it possible to associate a sign to the area A of a parallelogram generated by two linearly independent vectors $u, v \in R^2$. To do this, let $\{e_i\}$, $i = 1, 2$, be the natural ordered basis of R^2, and write $u = u_1 e_1 + u_2 e_2$, $v = v_1 e_1 + v_2 e_2$. Observe the matrix relation

$$\begin{pmatrix} u \cdot u & u \cdot v \\ v \cdot u & v \cdot v \end{pmatrix} = \begin{pmatrix} u_1 & u_2 \\ v_1 & v_2 \end{pmatrix} \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix}$$

and conclude that

$$A^2 = \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}^2.$$

Since the last determinant has the same sign as the basis $\{u, v\}$, we can say that A is positive or negative according to whether the orientation of $\{u, v\}$ is positive or negative. This is called the *oriented area* in R^2.

11. a. Show that the volume V of a parallelepiped generated by three linearly independent vectors $u, v, w \in R^3$ is given by $V = |(u \wedge v) \cdot w|$, and introduce an *oriented volume* in R^3.

b. Prove that

$$V^2 = \begin{vmatrix} u \cdot u & u \cdot v & u \cdot w \\ v \cdot u & v \cdot v & v \cdot w \\ w \cdot u & w \cdot v & w \cdot w \end{vmatrix}.$$

12. Given the vectors $v \neq 0$ and w, show that there exists a vector u such that $u \wedge v = w$ if and only if v is perpendicular to w. Is this vector u uniquely determined? If not, what is the most general solution?

13. Let $u(t) = (u_1(t), u_2(t), u_3(t))$ and $v(t) = (v_1(t), v_2(t), v_3(t))$ be differentiable maps from the interval (a, b) into R^3. If the derivatives $u'(t)$ and $v'(t)$ satisfy the conditions

$$u'(t) = au(t) + bv(t), \qquad v'(t) = cu(t) - av(t),$$

where a, b, and c are constants, show that $u(t) \wedge v(t)$ is a constant vector.

14. Find all unit vectors which are perpendicular to the vector $(2, 2, 1)$ and parallel to the plane determined by the points $(0, 0, 0)$, $(1, -2, 1)$, $(-1, 1, 1)$.

1-5. The Local Theory of Curves Parametrized by Arc Length

This section contains the main results of curves which will be used in the later parts of the book.

Let $\alpha\colon I = (a, b) \to R^3$ be a curve parametrized by arc length s. Since the tangent vector $\alpha'(s)$ has unit length, the norm $|\alpha''(s)|$ of the second derivative measures the rate of change of the angle which neighboring tangents make with the tangent at s. $|\alpha''(s)|$ gives, therefore, a measure of how rapidly the curve pulls away from the tangent line at s, in a neighborhood of s (see Fig. 1-14). This suggests the following definition.

DEFINITION. *Let* α*:* $\mathrm{I} \to \mathrm{R}^3$ *be a curve parametrized by arc length* $\mathrm{s} \in \mathrm{I}$. *The number* $|\alpha''(\mathrm{s})| = \mathrm{k}(\mathrm{s})$ *is called the* curvature *of* α *at* s.

If α is a straight line, $\alpha(s) = us + v$, where u and v are constant vectors ($|u| = 1$), then $k \equiv 0$. Conversely, if $k = |\alpha''(s)| \equiv 0$, then by integration $\alpha(s) = us + v$, and the curve is a straight line.

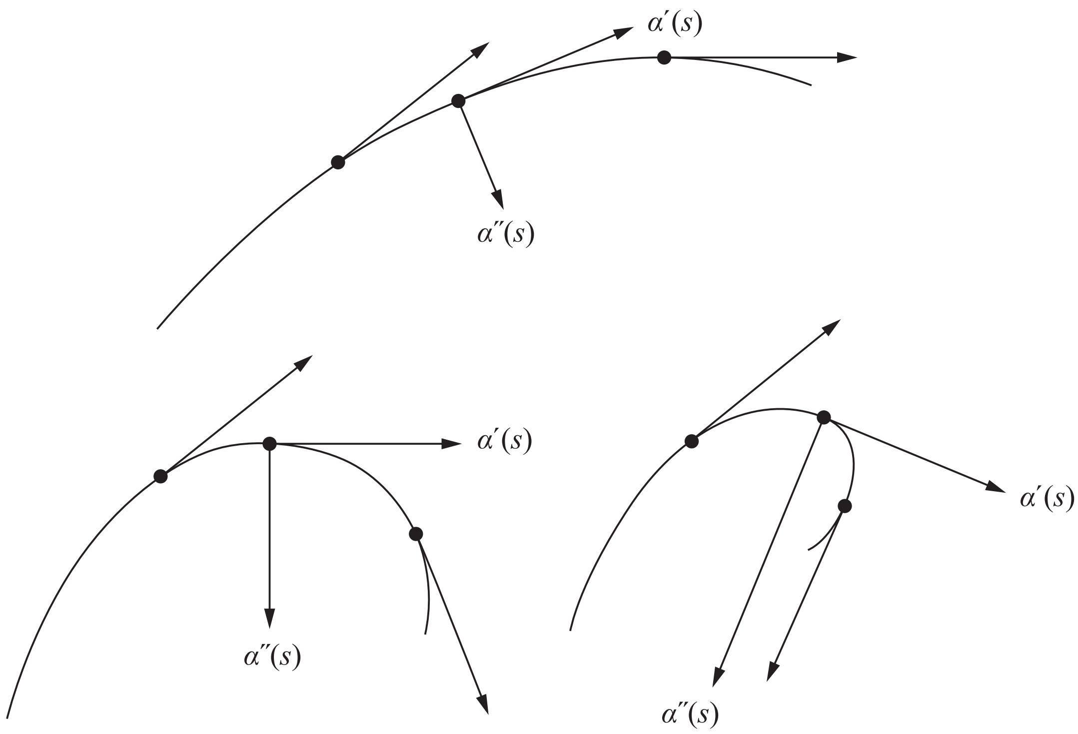


Figure 1-14

Notice that by a change of orientation, the tangent vector changes its direction; that is, if $\beta(-s) = \alpha(s)$, then

$$\frac{d\beta}{d(-s)}(-s) = -\frac{d\alpha}{ds}(s).$$

Therefore, $\alpha''(s)$ and the curvature remain invariant under a change of orientation.

At points where $k(s) \neq 0$, a unit vector $n(s)$ in the direction $\alpha''(s)$ is well defined by the equation $\alpha''(s) = k(s)n(s)$. Moreover, $\alpha''(s)$ is normal to $\alpha'(s)$, because by differentiating $\alpha'(s) \cdot \alpha'(s) = 1$ we obtain $\alpha''(s) \cdot \alpha'(s) = 0$. Thus, $n(s)$ is normal to $\alpha'(s)$ and is called the *normal vector* at s. The plane determined by the unit tangent and normal vectors, $\alpha'(s)$ and $n(s)$, is called the *osculating plane* at s. (See Fig. 1-15.)

At points where $k(s) = 0$, the normal vector (and therefore the osculating plane) is not defined (cf. Exercise 10). To proceed with the local analysis of curves, we need, in an essential way, the osculating plane. It is therefore convenient to say that $s \in I$ is a *singular point of order* 1 if $\alpha''(s) = 0$ (in this context, the points where $\alpha'(s) = 0$ are called singular points of order 0).

In what follows, we shall restrict ourselves to curves parametrized by arc length without singular points of order 1. We shall denote by $t(s) = \alpha'(s)$ the unit tangent vector of α at s. Thus, $t'(s) = k(s)n(s)$.

The unit vector $b(s) = t(s) \wedge n(s)$ is normal to the osculating plane and will be called the *binormal vector* at s. Since $b(s)$ is a unit vector, the length $|b'(s)|$ measures the rate of change of the neighboring osculating planes with

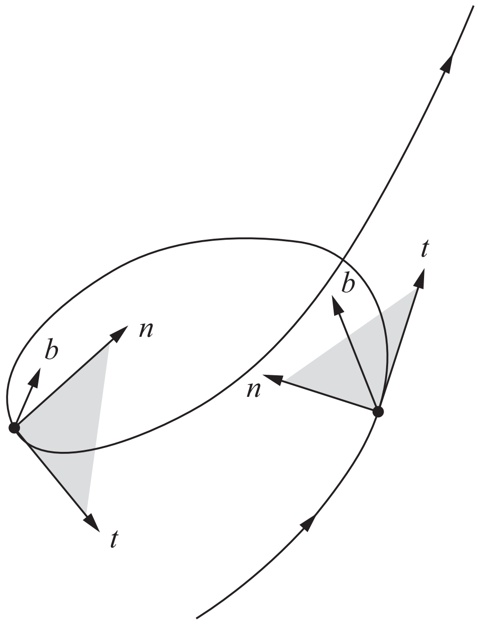


Figure 1-15

the osculating plane at s; that is, $|b'(s)|$ measures how rapidly the curve pulls away from the osculating plane at s, in a neighborhood of s (see Fig. 1-15).

To compute $b'(s)$ we observe that, on the one hand, $b'(s)$ is normal to $b(s)$ and that, on the other hand,

$$b'(s) = t'(s) \wedge n(s) + t(s) \wedge n'(s) = t(s) \wedge n'(s);$$

that is, $b'(s)$ is normal to $t(s)$. It follows that $b'(s)$ is parallel to $n(s)$, and we may write

$$b'(s) = \tau(s)n(s)$$

for some function $\tau(s)$. (*Warning*: Many authors write $-\tau(s)$ instead of our $\tau(s)$.)

DEFINITION. *Let* $\alpha\colon \mathrm{I} \to \mathrm{R}^3$ *be a curve parametrized by arc length* s *such that* $\alpha''(\mathrm{s}) \neq 0,\ \mathrm{s} \in \mathrm{I}$. *The number* $\tau(\mathrm{s})$ *defined by* $\mathrm{b}'(\mathrm{s}) = \tau(\mathrm{s})\mathrm{n}(\mathrm{s})$ *is called the* torsion *of* α *at* s.

If α is a plane curve (that is, $\alpha(I)$ is contained in a plane), then the plane of the curve agrees with the osculating plane; hence, $\tau \equiv 0$. Conversely, if $\tau \equiv 0$ (and $k \neq 0$), we have that $b(s) = b_0 =$ constant, and therefore

$$(\alpha(s) \cdot b_0)' = \alpha'(s) \cdot b_0 = 0.$$

It follows that $\alpha(s) \cdot b_0 =$ constant; hence, $\alpha(s)$ is contained in a plane normal to b_0. The condition that $k \neq 0$ everywhere is essential here. In Exercise 10 we shall give an example where τ can be defined to be identically zero and yet the curve is not a plane curve.

In contrast to the curvature, the torsion may be either positive or negative. The sign of the torsion has a geometric interpretation, to be given later (Sec. 1-6).

Notice that by changing orientation the binormal vector changes sign, since $b = t \wedge n$. It follows that $b'(s)$, and, therefore, the torsion, remain invariant under a change of orientation.

Let us summarize our position. To each value of the parameter s, we have associated three orthogonal unit vectors $t(s), n(s), b(s)$. The trihedron thus formed is referred to as the *Frenet trihedron* at s. The derivatives $t'(s) = kn$, $b'(s) = \tau n$ of the vectors $t(s)$ and $b(s)$, when expressed in the basis $\{t, n, b\}$, yield geometrical entities (curvature k and torsion τ) which give us information about the behavior of α in a neighborhood of s.

The search for other local geometrical entities would lead us to compute $n'(s)$. However, since $n = b \wedge t$, we have

$$n'(s) = b'(s) \wedge t(s) + b(s) \wedge t'(s) = -\tau b - kt,$$

and we obtain again the curvature and the torsion.

For later use, we shall call the equations

$$\begin{aligned} t' &= kn, \\ n' &= -kt - \tau b, \\ b' &= \tau n. \end{aligned}$$

the *Frenet formulas* (we have omitted the s, for convenience). In this context, the following terminology is usual. The tb plane is called the *rectifying plane*, and the nb plane the *normal plane*. The lines which contain $n(s)$ and $b(s)$ and pass through $\alpha(s)$ are called the *principal normal* and the *binormal*, respectively. The inverse $R = 1/k$ of the curvature is called the *radius of curvature* at s. Of course, a circle of radius r has radius of curvature equal to r, as one can easily verify.

Physically, we can think of a curve in R^3 as being obtained from a straight line by bending (curvature) and twisting (torsion). After reflecting on this construction, we are led to conjecture the following statement, which, roughly speaking, shows that k and τ describe completely the local behavior of the curve.

FUNDAMENTAL THEOREM OF THE LOCAL THEORY OF CURVES. *Given differentiable functions* $\mathrm{k(s)} > 0$ *and* $\tau(\mathrm{s})$, $\mathrm{s} \in \mathrm{I}$, *there exists a regular parametrized curve* $\alpha\colon \mathrm{I} \to \mathrm{R}^3$ *such that* s *is the arc length*, $\mathrm{k(s)}$ *is the curvature, and* $\tau(\mathrm{s})$ *is the torsion of* α. *Moreover, any other curve* $\bar{\alpha}$, *satisfying the same conditions, differs from* α *by a rigid motion; that is, there exists an orthogonal linear map* ρ *of* R^3, *with positive determinant, and a vector* c *such that* $\bar{\alpha} = \rho \circ \alpha + \mathrm{c}$.

The above statement is true. A complete proof involves the theorem of existence and uniqueness of solutions of ordinary differential equations and will be given in the appendix to Chap. 4. A proof of the uniqueness, up to rigid motions, of curves having the same s, $k(s)$, and $\tau(s)$ is, however, simple and can be given here.

Proof of the Uniqueness Part of the Fundamental Theorem. We first remark that arc length, curvature, and torsion are invariant under rigid motions; that means, for instance, that if $M\colon R^3 \to R^3$ is a rigid motion and $\alpha = \alpha(t)$ is a parametrized curve, then

$$\int_a^b \left|\frac{d\alpha}{dt}\right| dt = \int_a^b \left|\frac{d(M \circ \alpha)}{dt}\right| dt.$$

That is plausible, since these concepts are defined by using inner or vector products of certain derivatives (the derivatives are invariant under translations, and the inner and vector products are expressed by means of lengths and angles of vectors, and thus also invariant under rigid motions). A careful checking can be left as an exercise (see Exercise 6).

Now, assume that two curves $\alpha = \alpha(s)$ and $\bar{\alpha} = \bar{\alpha}(s)$ satisfy the conditions $k(s) = \bar{k}(s)$ and $\tau(s) = \bar{\tau}(s)$, $s \in I$. Let t_0, n_0, b_0 and $\bar{t}_0, \bar{n}_0, \bar{b}_0$ be the Frenet trihedrons at $s = s_0 \in I$ of α and $\bar{\alpha}$, respectively. Clearly, there is a rigid motion which takes $\bar{\alpha}(s_0)$ into $\alpha(s_0)$ and $\bar{t}_0, \bar{n}_0, \bar{b}_0$ into t_0, n_0, b_0. Thus, after performing this rigid motion on $\bar{\alpha}$, we have that $\bar{\alpha}(s_0) = \alpha(s_0)$ and that the Frenet trihedrons $t(s), n(s), b(s)$ and $\bar{t}(s), \bar{n}(s), \bar{b}(s)$ of α and $\bar{\alpha}$, respectively, satisfy the Frenet equations:

$$\begin{aligned}
\frac{dt}{ds} &= kn & \qquad \frac{d\bar{t}}{ds} &= k\bar{n} \\
\frac{dn}{ds} &= -kt - \tau b & \qquad \frac{d\bar{n}}{ds} &= -k\bar{t} - \tau\bar{n} \\
\frac{db}{ds} &= \tau n & \qquad \frac{d\bar{b}}{ds} &= \tau\bar{n},
\end{aligned}$$

with $t(s_0) = \bar{t}(s_0)$, $n(s_0) = \bar{n}(s_0)$, $b(s_0) = \bar{b}(s_0)$.

We now observe, by using the Frenet equations, that

$$\begin{aligned}
&\frac{1}{2}\frac{d}{ds}\{|t - \bar{t}|^2 + |n - \bar{n}|^2 + |b - \bar{b}|^2\} \\
&\quad = \langle t - \bar{t}, t' - \bar{t}'\rangle + \langle b - \bar{b}, b' - \bar{b}'\rangle + \langle n - \bar{n}, n' - \bar{n}'\rangle \\
&\quad = k\langle t - \bar{t}, n - \bar{n}\rangle + \tau\langle b - \bar{b}, n - \bar{n}\rangle - k\langle n - \bar{n}, t - \bar{t}\rangle \\
&\qquad - \tau\langle n - \bar{n}, b - \bar{b}\rangle \\
&\quad = 0
\end{aligned}$$

for all $s \in I$. Thus, the above expression is constant, and, since it is zero for $s = s_0$, it is identically zero. It follows that $t(s) = \bar{t}(s)$, $n(s) = \bar{n}(s)$, $b(s) = \bar{b}(s)$ for all $s \in I$. Since

$$\frac{d\alpha}{ds} = t = \bar{t} = \frac{d\bar{\alpha}}{ds},$$

we obtain $(d/ds)(\alpha - \bar{\alpha}) = 0$. Thus, $\alpha(s) = \bar{\alpha}(s) + a$, where a is a constant vector. Since $\alpha(s_0) = \bar{\alpha}(s_0)$, we have $a = 0$; hence, $\alpha(s) = \bar{\alpha}(s)$ for all $s \in I$.

Q.E.D.

Remark 1. In the particular case of a plane curve $\alpha\colon I \to R^2$, it is possible to give the curvature k a sign. For that, let $\{e_1, e_2\}$ be the natural basis (see Sec. 1-4) of R^2 and define the normal vector $n(s)$, $s \in I$, by requiring the basis $\{t(s), n(s)\}$ to have the same orientation as the basis $\{e_1, e_2\}$. The curvature k is then *defined* by

$$\frac{dt}{ds} = kn$$

and might be either positive or negative. It is clear that $|k|$ agrees with the previous definition and that k changes sign when we change either the orientation of α or the orientation of R^2 (Fig. 1-16).

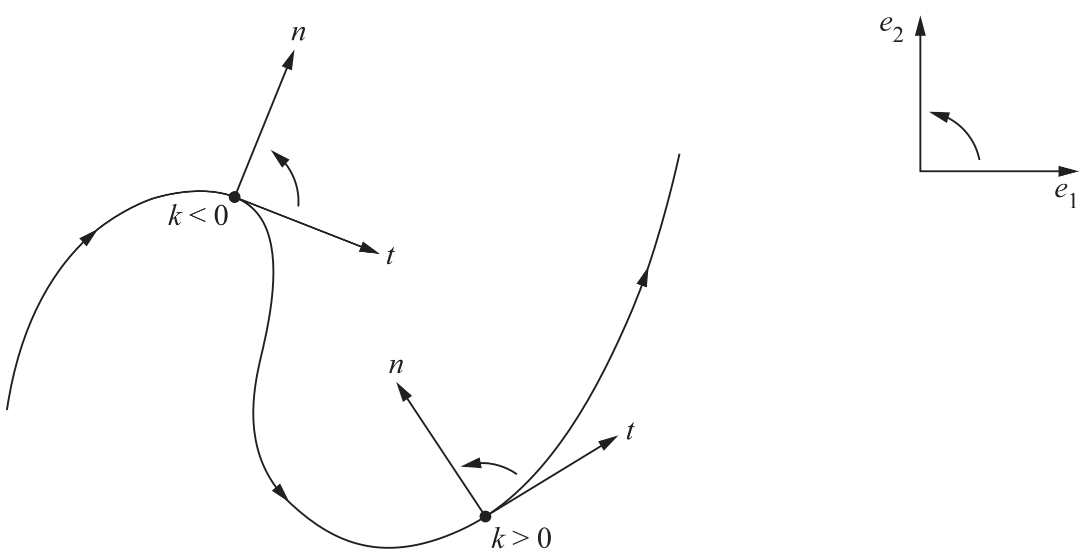


Figure 1-16

It should also be remarked that, in the case of plane curves ($\tau \equiv 0$), the proof of the fundamental theorem, refered to above, is actually very simple (see Exercise 9).

Remark 2. Given a regular parametrized curve $\alpha\colon I \to R^3$ (not necessarily parametrized by arc length), it is possible to obtain a curve $\beta\colon J \to R^3$ parametrized by arc length which has the same trace as α. In fact, let

$$s = s(t) = \int_{t_0}^{t} |\alpha'(t)|\,dt, \quad t, t_0 \in I.$$

Since $ds/dt = |\alpha'(t)| \neq 0$, the function $s = s(t)$ has a differentiable inverse $t = t(s)$, $s \in s(I) = J$, where, by an abuse of notation, t also denotes the inverse function s^{-1} of s. Now set $\beta = \alpha \circ t \colon J \to R^3$. Clearly, $\beta(J) = \alpha(I)$ and $|\beta'(s)| = |(\alpha'(t) \cdot (dt/ds)| = 1$. This shows that β has the same trace as α and is parametrized by arc length. It is usual to say that β is a *reparametrization of* $\alpha(I)$ *by arc length*.

This fact allows us to extend all local concepts previously defined to regular curves with an arbitrary parameter. Thus, we say that the curvature $k(t)$ of $\alpha \colon I \to R^3$ at $t \in I$ is the curvature of a reparametrization $\beta \colon J \to R^3$ of $\alpha(I)$ by arc length at the corresponding point $s = s(t)$. This is clearly independent of the choice of β and shows that the restriction, made at the end of Sec. 1-3, of considering only curves parametrized by arc length is not essential.

In applications, it is often convenient to have explicit formulas for the geometrical entities in terms of an arbitrary parameter; we shall present some of them in Exercise 12.

EXERCISES

Unless explicity stated, $\alpha \colon \mathrm{I} \to \mathrm{R}^3$ *is a curve parametrized by arc length* s, *with curvature* $\mathrm{k(s)} \neq 0$, *for all* $\mathrm{s} \in \mathrm{I}$.

1. Given the parametrized curve (helix)

$$\alpha(s) = \left(a \cos \frac{s}{c}, a \sin \frac{s}{c}, b\frac{s}{c}\right), \qquad s \in R,$$

where $c^2 = a^2 + b^2$,

a. Show that the parameter s is the arc length.

b. Determine the curvature and the torsion of α.

c. Determine the osculating plane of α.

d. Show that the lines containing $n(s)$ and passing through $\alpha(s)$ meet the z axis under a constant angle equal to $\pi/2$.

e. Show that the tangent lines to α make a constant angle with the z axis.

***2.** Show that the torsion τ of α is given by

$$\tau(s) = -\frac{\alpha'(s) \wedge \alpha''(s) \cdot \alpha'''(s)}{|k(s)|^2}.$$

3. Assume that $\alpha(I) \subset R^2$ (i.e., α is a plane curve) and give k a sign as in the text. Transport the vectors $t(s)$ parallel to themselves in such a way that the origins of $t(s)$ agree with the origin of R^2; the end points of $t(s)$ then describe a parametrized curve $s \to t(s)$ called the *indicatrix* *of tangents* of α. Let $\theta(s)$ be the angle from e_1 to $t(s)$ in the orientation of R^2. Prove (a) and (b) (notice that we are assuming that $k \neq 0$).

a. The indicatrix of tangents is a regular parametrized curve.

b. $dt/ds = (d\theta/ds)n$, that is, $k = d\theta/ds$.

***4.** Assume that all normals of a parametrized curve pass through a fixed point. Prove that the trace of the curve is contained in a circle.

5. A regular parametrized curve α has the property that all its tangent lines pass through a fixed point.

a. Prove that the trace of α is a (segment of a) straight line.

b. Does the conclusion in part a still hold if α is not regular?

6. A *translation* by a vector v in R^3 is the map $A\colon R^3 \to R^3$ that is given by $A(p) = p + v$, $p \in R^3$. A linear map $\rho\colon R^3 \to R^3$ is an *orthogonal transformation* when $\rho u \cdot \rho v = u \cdot v$ for all vectors $u, v \in R^3$. A *rigid motion* in R^3 is the result of composing a translation with an orthogonal transformation with positive determinant (this last condition is included because we expect rigid motions to preserve orientation).

a. Demonstrate that the norm of a vector and the angle θ between two vectors, $0 \leq \theta \leq \pi$, are invariant under orthogonal transformations with positive determinant.

b. Show that the vector product of two vectors is invariant under orthogonal transformations with positive determinant. Is the assertion still true if we drop the condition on the determinant?

c. Show that the arc length, the curvature, and the torsion of a parametrized curve are (whenever defined) invariant under rigid motions.

***7.** Let $\alpha\colon I \to R^2$ be a regular parametrized plane curve (arbitrary parameter), and define $n = n(t)$ and $k = k(t)$ as in Remark 1. Assume that $k(t) \neq 0$, $t \in I$. In this situation, the curve

$$\beta(t) = \alpha(t) + \frac{1}{k(t)}n(t), \quad t \in I,$$

is called the *evolute* of α (Fig. 1-17).

a. Show that the tangent at t of the evolute of α is the normal to α at t.

b. Consider the normal lines of α at two neighboring points t_1, t_2, $t_1 \neq t_2$. Let t_1 approach t_2 and show that the intersection points of the normals converge to a point on the trace of the evolute of α.

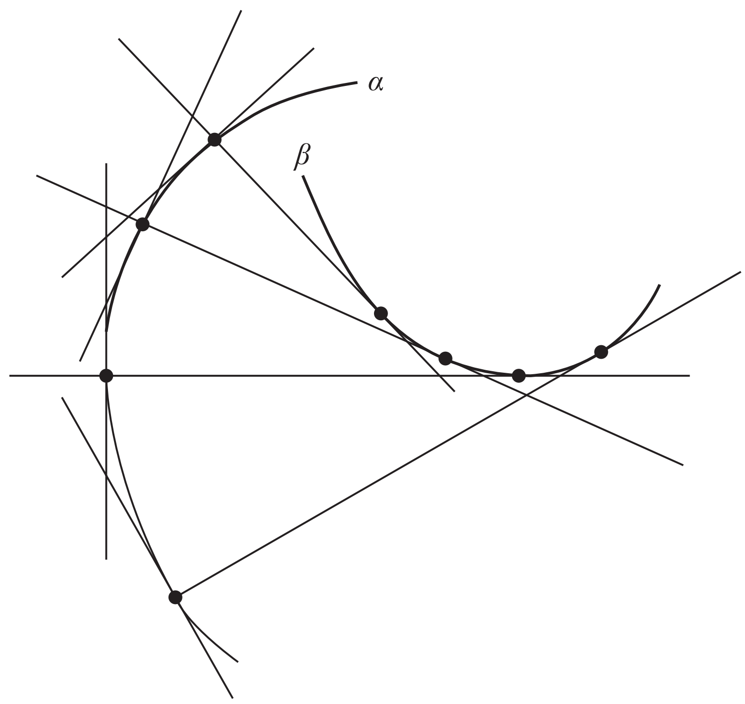


Figure 1-17

8. The trace of the parametrized curve (arbitrary parameter)

$$\alpha(t) = (t, \cosh t), \quad t \in R,$$

is called the *catenary*.

a. Show that the signed curvature (cf. Remark 1) of the catenary is

$$k(t) = \frac{1}{\cosh^2 t}.$$

b. Show that the evolute (cf. Exercise 7) of the catenary is

$$\beta(t) = (t - \sinh t \cosh t, 2\cosh t).$$

9. Given a differentiable function $k(s)$, $s \in I$, show that the parametrized plane curve having $k(s) = k$ as curvature is given by

$$\alpha(s) = \left(\int \cos\theta(s)\, ds + a, \int \sin\theta(s)\, ds + b \right),$$

where

$$\theta(s) = \int k(s)\, ds + \varphi,$$

and that the curve is determined up to a translation of the vector (a, b) and a rotation of the angle φ.

10. Consider the map

$$\alpha(t) = \begin{cases} (t, 0, e^{-1/t^2}), & t > 0 \\ (t, e^{-1/t^2}, 0), & t < 0 \\ (0, 0, 0), & t = 0 \end{cases}$$

a. Prove that α is a differentiable curve.

b. Prove that α is regular for all t and that the curvature $k(t) \neq 0$, for $t \neq 0$, $t \neq \pm\sqrt{2/3}$, and $k(0) = 0$.

c. Show that the limit of the osculating planes as $t \to 0$, $t > 0$, is the plane $y = 0$ but that the limit of the osculating planes as $t \to 0$, $t < 0$, is the plane $z = 0$ (this implies that the normal vector is discontinuous at $t = 0$ and shows why we excluded points where $k = 0$).

d. Show that τ can be defined so that $\tau \equiv 0$, even though α is not a plane curve.

11. One often gives a plane curve in polar coordinates by $\rho = \rho(\theta)$, $a \leq \theta \leq b$.

a. Show that the arc length is

$$\int_a^b \sqrt{\rho^2 + (\rho')^2}\, d\theta,$$

where the prime denotes the derivative relative to θ.

b. Show that the curvature is

$$k(\theta) = \frac{2(\rho')^2 - \rho\rho'' + \rho^2}{\{(\rho')^2 + \rho^2\}^{3/2}}.$$

12. Let $\alpha\colon I \to R^3$ be a regular parametrized curve (not necessarily by arc length) and let $\beta\colon J \to R^3$ be a reparametrization of $\alpha(I)$ by the arc length $s = s(t)$, measured from $t_0 \in I$ (see Remark 2). Let $t = t(s)$ be the inverse function of s and set $d\alpha/dt = \alpha'$, $d^2\alpha/dt^2 = \alpha''$, etc. Prove that

a. $dt/ds = 1/|\alpha'|$, $d^2t/ds^2 = -(\alpha' \cdot \alpha''/|\alpha'|^4)$.

b. The curvature of α at $t \in I$ is

$$k(t) = \frac{|\alpha' \wedge \alpha''|}{|\alpha'|^3}.$$

c. The torsion of α at $t \in I$ is

$$\tau(t) = -\frac{(\alpha' \wedge \alpha'') \cdot \alpha'''}{|\alpha' \wedge \alpha''|^2}.$$

d. If $\alpha\colon I \to R^2$ is a plane curve $\alpha(t) = (x(t), y(t))$, the signed curvature (see Remark 1) of α at t is

$$k(t) = \frac{x'y'' - x''y'}{((x')^2 + (y')^2)^{3/2}}.$$

***13.** Assume that $\tau(s) \neq 0$ and $k'(s) \neq 0$ for all $s \in I$. Show that a necessary and sufficient condition for $\alpha(I)$ to lie on a sphere is that

$$R^2 + (R')^2T^2 = \text{const.},$$

where $R = 1/k$, $T = 1/\tau$, and R' is the derivative of R relative to s.

14. Let $\alpha\colon (a, b) \to R^2$ be a regular parametrized plane curve. Assume that there exists t_0, $a < t_0 < b$, such that the distance $|\alpha(t)|$ from the origin to the trace of α will be a maximum at t_0. Prove that the curvature k of α at t_0 satisfies $|k(t_0)| \geq 1/|\alpha(t_0)|$.

***15.** Show that the knowledge of the vector function $b = b(s)$ (binormal vector) of a curve α, with nonzero torsion everywhere, determines the curvature $k(s)$ and the absolute value of the torsion $\tau(s)$ of α.

***16.** Show that the knowledge of the vector function $n = n(s)$ (normal vector) of a curve α, with nonzero torsion everywhere, determines the curvature $k(s)$ and the torsion $\tau(s)$ of α.

17. In general, a curve α is called a *helix* if the tangent lines of α make a constant angle with a fixed direction. Assume that $\tau(s) \neq 0$, $s \in I$, and prove that:

***a.** α is a helix if and only if $k/\tau = \text{const}$.

***b.** α is a helix if and only if the lines containing $n(s)$ and passing through $\alpha(s)$ are parallel to a fixed plane.

***c.** α is a helix if and only if the lines containing $b(s)$ and passing through $\alpha(s)$ make a constant angle with a fixed direction.

d. The curve

$$\alpha(s) = \left(\frac{a}{c}\int \sin\theta(s)\,ds,\ \frac{a}{c}\int \cos\theta(s)\,ds,\ \frac{b}{c}s\right),$$

where $c^2 = a^2 + b^2$, is a helix, and that $k/\tau = a/b$.

***18.** Let $\alpha\colon I \to R^3$ be a parametrized regular curve (not necessarily by arc length) with $k(t) \neq 0$, $\tau(t) \neq 0$, $t \in I$. The curve α is called a *Bertrand curve* if there exists a curve $\bar{\alpha}\colon I \to R^3$ such that the normal lines of α and $\bar{\alpha}$ at $t \in I$ are equal. In this case, $\bar{\alpha}$ is called a *Bertrand mate* of α, and we can write

$$\bar{\alpha}(t) = \alpha(t) + rn(t).$$

Prove that

a. r is constant.

b. α is a Bertrand curve if and only if there exists a linear relation

$$Ak(t) + B\tau(t) = 1, \quad t \in I,$$

where A, B are nonzero constants and k and τ are the curvature and torsion of α, respectively.

c. If α has more than one Bertrand mate, it has infinitely many Bertrand mates. This case occurs if and only if α is a circular helix.

1-6. The Local Canonical Form†

One of the most effective methods of solving problems in geometry consists of finding a coordinate system which is adapted to the problem. In the study of local properties of a curve, in the neighborhood of the point s, we have a natural coordinate system, namely the Frenet trihedron at s. It is therefore convenient to refer the curve to this trihedron.

Let $\alpha\colon I \to R^3$ be a curve parametrized by arc length without singular points of order 1. We shall write the equations of the curve, in a neighborhood of s_0, using the trihedron $t(s_0)$, $n(s_0)$, $b(s_0)$ as a basis for R^3. We may assume, without loss of generality, that $s_0 = 0$, and we shall consider the (finite) Taylor expansion

$$\alpha(s) = \alpha(0) + s\alpha'(0) + \frac{s^2}{2}\alpha''(0) + \frac{s^3}{6}\alpha'''(0) + R,$$

where $\lim_{s\to 0} R/s^3 = 0$. Since $\alpha'(0) = t$, $\alpha''(0) = kn$, and

$$\alpha'''(0) = (kn)' = k'n + kn' = k'n - k^2t - k\tau b,$$

we obtain

$$\alpha(s) - \alpha(0) = \left(s - \frac{k^2s^3}{3!}\right)t + \left(\frac{s^2k}{2} + \frac{s^3k'}{3!}\right)n - \frac{s^3}{3!}k\tau b + R,$$

where all terms are computed at $s = 0$.

Let us now take the system $Oxyz$ in such a way that the origin O agrees with $\alpha(0)$ and that $t = (1, 0, 0)$, $n = (0, 1, 0)$, $b = (0, 0, 1)$. Under these conditions, $\alpha(s) = (x(s), y(s), z(s))$ is given by

†This section may be omitted on a first reading.

$$x(s) = s - \frac{k^2 s^3}{6} + R_x,$$
$$y(s) = \frac{k}{2}s^2 + \frac{k' s^3}{6} + R_y, \quad \textbf{(1)}$$
$$z(s) = -\frac{k\tau}{6}s^3 + R_z,$$

where $R = (R_x, R_y, R_z)$. The representation (1) is called the *local canonical form* of α, in a neighborhood of $s = 0$. In Fig. 1-18 is a rough sketch of the projections of the trace of α, for s small, in the tn, tb, and nb planes.

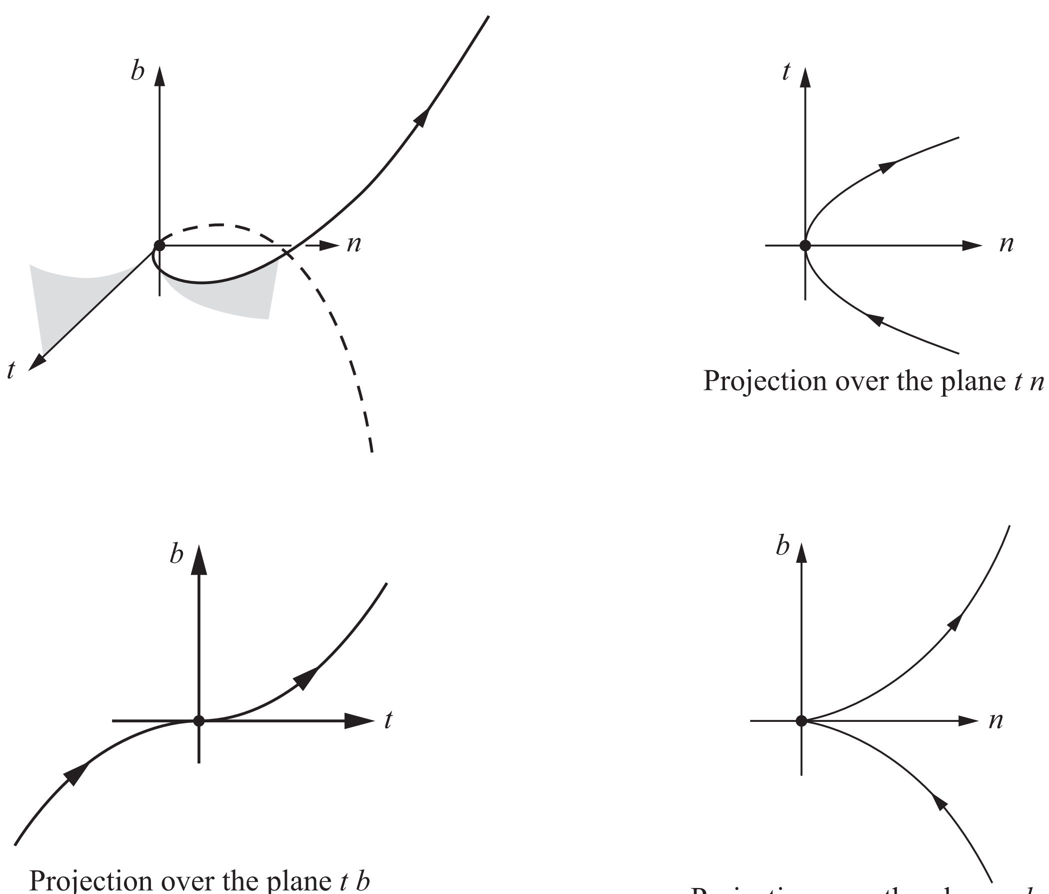


Figure 1-18

Below we shall describe some geometrical applications of the local canonical form. Further applications will be found in the Exercises.

A first application is the following interpretation of the sign of the torsion. From the third equation of (1) it follows that if $\tau < 0$ and s is sufficiently small, then $z(s)$ increases with s. Let us make the convention of calling the "positive side" of the osculating plane that side toward which b is pointing. Then, since $z(0) = 0$, when we describe the curve in the direction of increasing arc length, the curve will cross the osculating plane at $s = 0$, pointing toward the positive side (see Fig. 1-19). If, on the contrary, $\tau > 0$, the curve (described in the direction of increasing arc length) will cross the osculating plane pointing to the side opposite the positive side.

Negative torsion   Positive torsion

Figure 1-19

The helix of Exercise 1 of Sec. 1-5 has negative torsion. An example of a curve with positive torsion is the helix

$$\alpha(s) = \left(a\cos\frac{s}{c},\, a\sin\frac{s}{c},\, -b\frac{s}{c}\right)$$

obtained from the first one by a reflection in the xz plane (see Fig. 1-19).

Remark. It is also usual to define torsion by $b' = -\tau n$. With such a definition, the torsion of the helix of Exercise 1 becomes positive.

Another consequence of the canonical form is the existence of a neighborhood $J \subset I$ of $s = 0$ such that $\alpha(J)$ is entirely contained in the one side of the rectifying plane toward which the vector n is pointing (see Fig. 1-18). In fact, since $k > 0$, we obtain, for s sufficiently small, $y(s) \geq 0$, and $y(s) = 0$ if and only if $s = 0$. This proves our claim.

As a last application of the canonical form, we mention the following property of the osculating plane. The osculating plane at s is the limit position of the plane determined by the tangent line at s and the point $\alpha(s+h)$ when $h \to 0$. To prove this, let us assume that $s = 0$. Thus, every plane containing the tangent at $s = 0$ is of the form $z = cy$ or $y = 0$. The plane $y = 0$ is the rectifying plane that, as seen above, contains no points near $\alpha(0)$ (except $\alpha(0)$ itself) and that may therefore be discarded from our considerations. The condition for the plane $z = cy$ to pass through $\alpha(s+h)$ is ($s = 0$)

$$c = \frac{z(h)}{y(h)} = \frac{-\dfrac{k}{6}\tau h^3 + \cdots}{\dfrac{k}{2}h^2 + \dfrac{k^2}{6}h^3 + \cdots}.$$

Letting $h \to 0$, we see that $c \to 0$. Therefore, the limit position of the plane $z(s) = c(h)y(s)$ is the plane $z = 0$, that is, the osculating plane, as we wished.

EXERCISES

***1.** Let $\alpha\colon I \to R^3$ be a curve parametrized by arc length with curvature $k(s) \neq 0$, $s \in I$. Let P be a plane satisfying both of the following conditions:

1. P contains the tangent line at s.
2. Given any neighborhood $J \subset I$ of s, there exist points of $\alpha(J)$ in both sides of P.

Prove that P is the osculating plane of α at s.

2. Let $\alpha\colon I \to R^3$ be a curve parametrized by arc length, with curvature $k(s) \neq 0$, $s \in I$. Show that

***a.** The osculating plane at s is the limit position of the plane passing through $\alpha(s)$, $\alpha(s+h_1)$, $\alpha(s+h_2)$ when $h_1, h_2 \to 0$.

b. The limit position of the circle passing through $\alpha(s)$, $\alpha(s+h_1)$, $\alpha(s+h_2)$ when h_1, $h_2 \to 0$ is a circle in the osculating plane at s, the center of which is on the line that contains $n(s)$ and the radius of which is the radius of curvature $1/k(s)$; this circle is called the *osculating circle* at s.

3. Show that the curvature $k(t) \neq 0$ of a regular parametrized curve $\alpha\colon I \to R^3$ is the curvature at t of the plane curve $\pi \circ \alpha$, where π is the normal projection of α over the osculating plane at t.

1-7. Global Properties of Plane Curves†

In this section we want to describe some results that belong to the global differential geometry of curves. Even in the simple case of plane curves, the subject already offers examples of nontrivial theorems and interesting questions. To develop this material here, we must assume some plausible facts without proofs; we shall try to be careful by stating these facts precisely. Although we want to come back later, in a more systematic way, to global differential geometry (Chap. 5), we believe that this early presentation of the subject is both stimulating and instructive.

This section contains three topics in order of increasing difficulty: (A) the isoperimetric inequality, (B) the four-vertex theorem, and (C) the Cauchy-Crofton formula. The topics are entirely independent, and some or all of them can be omitted on a first reading.

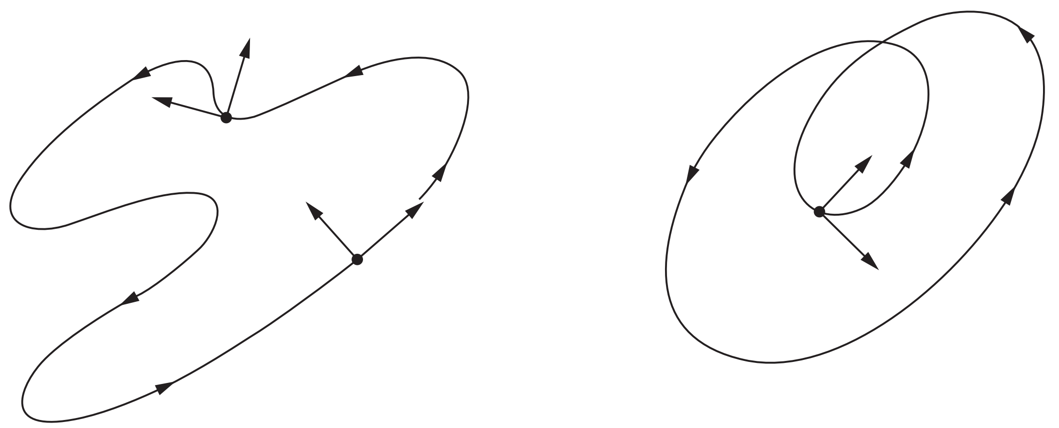
A differentiable function on a closed interval $[a, b]$ is the restriction of a differentiable function defined on an open interval containing $[a, b]$.

†This section may be omitted on a first reading.

A *closed plane curve* is a regular parametrized curve $\alpha\colon [a, b] \to R^2$ such that α and all its derivatives agree at a and b; that is,

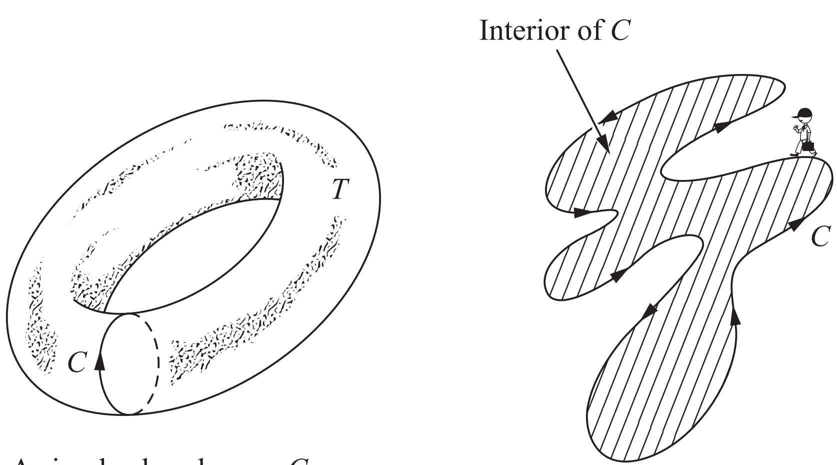
$$\alpha(a) = \alpha(b), \quad \alpha'(a) = \alpha'(b), \quad \alpha''(a) = \alpha''(b), \ldots .$$

The curve α is *simple* if it has no further self-intersections; that is, if $t_1, t_2 \in [a, b)$, $t_1 \neq t_2$, then $\alpha(t_1) \neq \alpha(t_2)$ (Fig. 1-20).



(a) A simple closed curve (b) A (nonsimple) closed curve

Figure 1-20



(a) A simple closed curve C on a torus T; C bounds no region on T. (b) C is positively oriented.

Figure 1-21

We usually consider the curve $\alpha\colon [0, l] \to R^2$ parametrized by arc length s; hence, l is the length of α. Sometimes we refer to a simple closed curve C, meaning the trace of such an object. The curvature of α will be taken with a sign, as in Remark 1 of Sec. 1-5 (see Fig. 1-20).

We assume that *a simple closed curve C in the plane bounds a region of this plane* that is called the *interior* of C. This is part of the so-called Jordan curve theorem (a proof will be given in Sec. 5-7, Theorem 1), which does not hold, for instance, for simple curves on a torus (the surface of a doughnut;

see Fig. 1-21(a)). Whenever we speak of the area bounded by a simple closed curve C, we mean the area of the interior of C. We assume further that the parameter of a simple closed curve can be so chosen that if one is going along the curve in the direction of increasing parameters, then the interior of the curve remains to the left (Fig. 1-21(b)). Such a curve will be called *positively oriented.*

A. The Isoperimetric Inequality

This is perhaps the oldest global theorem in differential geometry and is related to the following (isoperimetric) problem. *Of all simple closed curves in the plane with a given length l, which one bounds the largest area*? In this form, the problem was known to the Greeks, who also knew the solution, namely, the circle. A satisfactory proof of the fact that the circle is a solution to the isoperimetric problem took, however, a long time to appear. The main reason seems to be that the earliest proofs assumed that a solution should exist. It was only in 1870 that K. Weierstrass pointed out that many similar questions did not have solutions and gave a complete proof of the existence of a solution to the isoperimetric problem. Weierstrass' proof was somewhat hard, in the sense that it was a corollary of a theory developed by him to handle problems of maximizing (or minimizing) certain integrals (this theory is called calculus of variations and the isoperimetric problem is a typical example of the problems it deals with). Later, more direct proofs were found. The simple proof we shall present is due to E. Schmidt (1939). For another direct proof and further bibliography on the subject, one may consult Reference [10] in the Bibliography.

We shall make use of the following formula for the area A bounded by a positively oriented simple closed curve $\alpha(t) = (x(t), y(t))$, where $t \in [a, b]$ is an arbitrary parameter:

$$A = -\int_a^b y(t)x'(t)\,dt = \int_a^b x(t)y'(t)\,dt = \frac{1}{2}\int_a^b (xy' - yx')\,dt \qquad \textbf{(1)}$$

Notice that the second formula is obtained from the first one by observing that

$$\begin{aligned}\int_a^b xy'\,dt &= \int_a^b (xy)'\,dt - \int_a^b x'y\,dt = [xy(b) - xy(a)] - \int_a^b x'y\,dt \\ &= -\int_a^b x'y\,dt,\end{aligned}$$

since the curve is closed. The third formula is immediate from the first two.

To prove the first formula in Eq. (1), we consider initially the case of Fig. 1-22 where the curve is made up of two straight-line segments parallel to the y axis and two arcs that can be written in the form

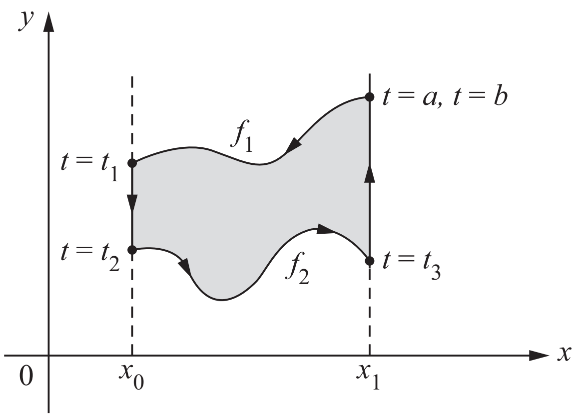


Figure 1-22

$$y = f_1(x) \quad \text{and} \quad y = f_2(x), x \in [x_0, x_1], f_1 > f_2.$$

Clearly, the area bounded by the curve is

$$A = \int_{x_0}^{x_1} f_1(x)\, dx - \int_{x_0}^{x_1} f_2(x)\, dx.$$

Since the curve is positively oriented, we obtain, with the notation of Fig. 1-22,

$$A = -\int_a^{t_1} y(t)x'(t)\, dt - \int_{t_2}^{t_3} y(t)x'(t)\, dt = -\int_a^b y(t)x'(t)\, dt,$$

since $x'(t) = 0$ along the segments parallel to the y axis. This proves Eq. (1) for this case.

To prove the general case, it must be shown that it is possible to divide the region bounded by the curve into a finite number of regions of the above type. This is clearly possible (Fig. 1-23) if *there exists a straight line E in the plane such that the distance $\rho(t)$ of $\alpha(t)$ to this line is a function with finitely many critical points* (a critical point is a point where $\rho'(t) = 0$). The last assertion is true, but we shall not go into its proof. We shall mention, however, that Eq. (1) can also be obtained by using Stokes' (Green's) theorem in the plane (see Exercise 15).

THEOREM 1 (The Isoperimetric Inequality). *Let* C *be a simple closed plane curve with length l, and let* A *be the area of the region bounded by* C. *Then*

$$l^2 - 4\pi A \geq 0, \tag{2}$$

and equality holds if and only if C *is a circle.*

Proof. Let E and E' be two parallel lines which do not meet the closed curve C, and move them together until they first meet C. We thus obtain two parallel tangent lines to C, L and L', so that the curve is entirely contained in the strip bounded by L and L'. Consider a circle S^1 which is tangent to both L and L' and does not meet C. Let O be the center of S^1 and take a coordinate system with origin at O and the x axis perpendicular to L and L' (Fig. 1-24). Parametrize C by arc length, $\alpha(s) = (x(s), y(s))$, so that it is positively oriented and the tangency points of L and L' are $s = 0$ and $s = s_1$, respectively.

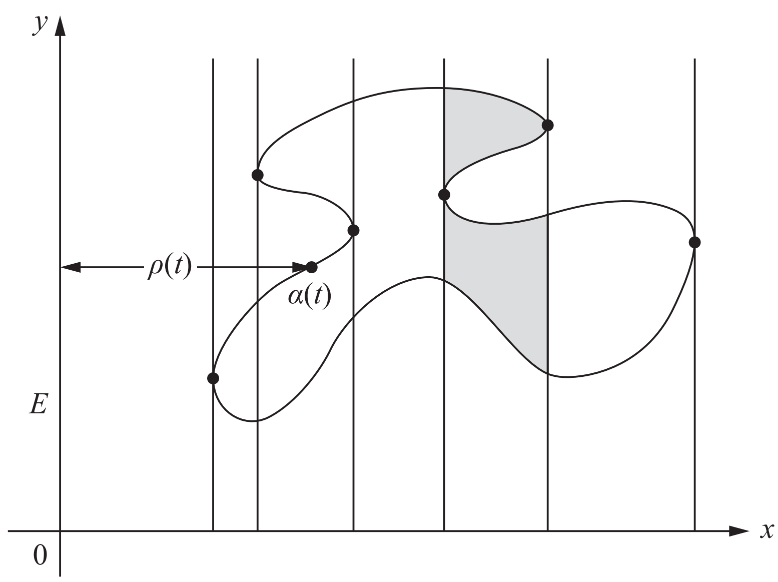


Figure 1-23

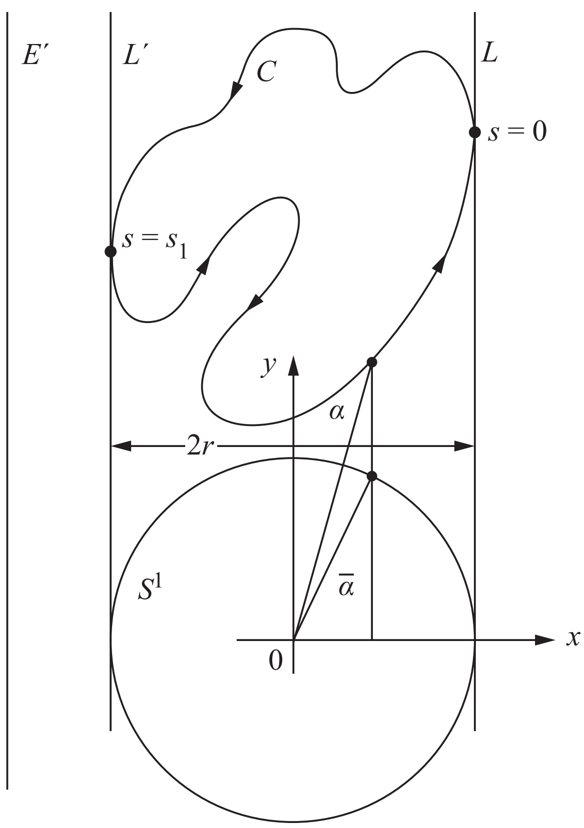


Figure 1-24

We can assume that the equation of S^1 is

$$\bar{\alpha}(s) = (\bar{x}(s), \bar{y}(s)) = (x(s), \bar{y}(s)), s \in [0, l].$$

Let $2r$ be the distance between L and L'. By using Eq. (1) and denoting by $\bar{A}$ the area bounded by S^1, we have

$$A = \int_0^l xy'\, ds, \quad \bar{A} = \pi r^2 = -\int_0^l \bar{y}x'\, ds.$$

Thus,

$$\begin{aligned} A + \pi r^2 &= \int_0^l (xy' - \bar{y}x')\, ds \le \int_0^l \sqrt{(xy' - \bar{y}x')^2}\, ds \\ &\le \int_0^l \sqrt{(x^2 + \bar{y}^2)((x')^2 + (y')^2)}\, ds = \int_0^l \sqrt{\bar{x}^2 + \bar{y}^2}\, ds \\ &= lr, \end{aligned} \tag{3}$$

where we used that the inner product of two vectors v_1 and v_2 satisfies

$$|(v_1 \cdot v_2)|^2 \le |(v_1)|^2 |(v_2)|^2. \tag{3$'$}$$

Later in the proof, we will need that equality holds in (3$'$) if and only if one vector is a multiple of the other.

We now notice the fact that the geometric mean of two positive numbers is smaller than or equal to their arithmetic mean, and equality holds if and only if they are equal. lt follows that

$$\sqrt{A}\sqrt{\pi r^2} \le \tfrac{1}{2}(A + \pi r^2) \le \tfrac{1}{2} lr. \tag{4}$$

Therefore, $4\pi A r^2 \le l^2 r^2$, and this gives Eq. (2).

Now, assume that equality holds in Eq. (2). Then equality must hold everywhere in Eqs. (3) and (4). From the equality in Eq. (4) it follows that $A = \pi r^2$. Thus, $l = 2\pi r$ and r does not depend on the choice of the direction of L. Furthermore, equality in Eq. (3) implies that

$$(x, \bar{y}) = \lambda(y', -x')$$

that is,

$$\lambda = \frac{x}{y'} = \frac{\bar{y}}{x'} = \frac{\sqrt{x^2 + \bar{y}^2}}{\sqrt{(y')^2 + (x')^2}} = \pm r.$$

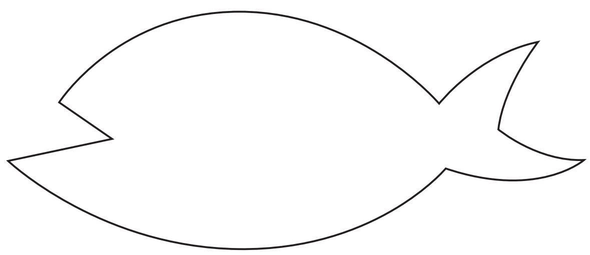
Thus, $x = \pm r y'$. Since r does not depend on the choice of the direction of L, we can interchange x and y in the last relation and obtain $y = \pm r x'$. Thus,

$$x^2 + y^2 = r^2((x')^2 + (y')^2) = r^2$$

and C is a circle, as we wished. **Q.E.D.**

Remark 1. It is easily checked that the above proof can be applied to C^1 *curves*, that is, curves $\alpha(t) = (x(t), y(t))$, $t \in [a, b]$, for which we require only that the functions $x(t)$, $y(t)$ have continuous first derivatives (which, of course, agree at a and b if the curve is closed).

Remark 2. The isoperimetric inequality holds true for a wide class of curves. Direct proofs have been found that work as long as we can define arc length and area for the curves under consideration. For the applications, it is convenient to remark that the theorem holds for *piecewise C^1 curves*, that is, continuous curves that are made up by a finite number of C^1 arcs. These curves can have a finite number of corners, where the tangent is discontinuous (Fig. 1-25).



A piecewise C^1 curve

Figure 1-25

B. The Four-Vertex Theorem

We shall need further general facts on plane closed curves.

Let $\alpha\colon [0, l] \to R^2$ be a plane closed curve given by $\alpha(s) = (x(s), y(s))$. Since s is the arc length, the tangent vector $t(s) = (x'(s), y'(s))$ has unit length. It is convenient to introduce the *tangent indicatrix* $t\colon [0, l] \to R^2$ that is given by $t(s) = (x'(s), y'(s))$; this is a differentiable curve, the trace of which is contained in a circle of radius 1 (Fig. 1-26). Observe that the velocity vector of the tangent indicatrix is

$$\begin{aligned}\frac{dt}{ds} &= (x''(s), y''(s)) \\ &= \alpha''(s) = kn,\end{aligned}$$

where n is the normal vector, oriented as in Remark 2 of Sec. 1-5, and k is the curvature of α.

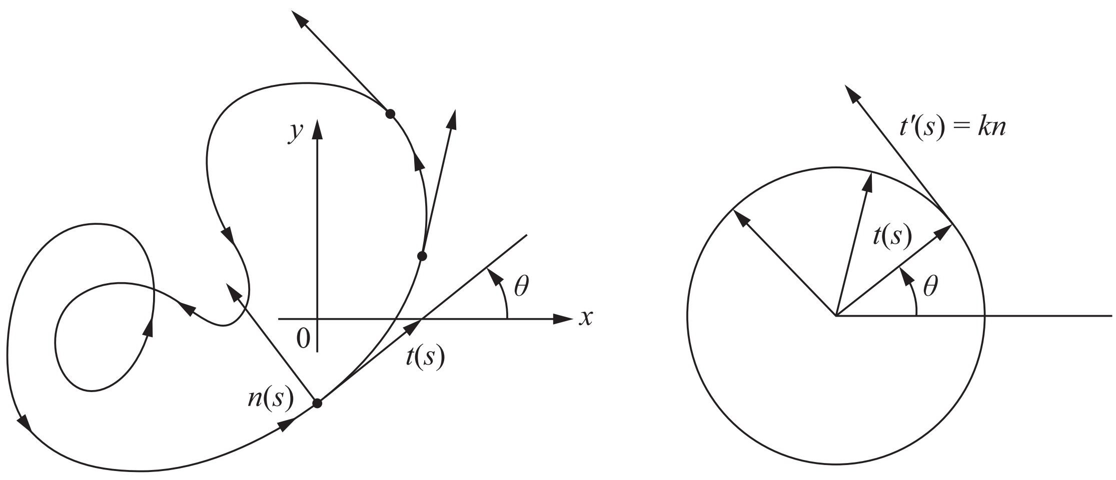


Figure 1-26

Let $\theta(s)$, $0 < \theta(s) < 2\pi$, be the angle that $t(s)$ makes with the x axis; that is, $x'(s) = \cos\theta(s)$, $y'(s) = \sin\theta(s)$. Since

$$\theta(s) = \arctan\frac{y'(s)}{x'(s)},$$

$\theta = \theta(s)$ is locally well defined (that is, it is well defined in a small interval about each s) as a differentiable function and

$$\begin{aligned}\frac{dt}{ds} &= \frac{d}{ds}(\cos\theta, \sin\theta)\\ &= \theta'(-\sin\theta, \cos\theta) = \theta' n.\end{aligned}$$

This means that $\theta'(s) = k(s)$ and suggests defining a global differentiable function $\theta\colon [0, l] \to R$ by

$$\theta(s) = \int_0^s k(s)\, ds.$$

Since

$$\theta' = k = x'y'' - x''y' = \left(\arctan\frac{y'}{x'}\right)',$$

this global function agrees, up to constants, with the previous locally defined θ. Intuitively, $\theta(s)$ measures the total rotation of the tangent vector, that is, the total angle described by the point $t(s)$ on the tangent indicatrix, as we run the curve α from 0 to s. Since α is closed, this angle is an integer multiple I of 2π; that is,

$$\int_o^l k(s)\, ds = \theta(l) - \theta(0) = 2\pi I.$$

The integer I is called the *rotation index* of the curve α.

In Fig. 1-27 are some examples of curves with their rotation indices. Observe that the rotation index changes sign when we change the orientation of the curve. Furthermore, the definition is so set that the rotation index of a positively oriented simple closed curve is positive.

An important global fact about the rotation index is given in the following theorem, which will be proved later in the book (Sec. 5-7, Theorem 2).

THE THEOREM OF TURNING TANGENTS. *The rotation index of a simple closed curve is* ± 1, *where the sign depends on the orientation of the curve.*

A regular, plane (not necessarily closed) curve $\alpha\colon [a, b] \to R^2$ is *convex* if, for all $t \in [a, b]$, the trace $\alpha([a, b])$ of α lies entirely on one side of the closed half-plane determined by the tangent line at t (Fig. 1-28).

A *vertex* of a regular plane curve $\alpha\colon [a, b] \to R^2$ is a point $t \in [a, b]$ where $k'(t) = 0$. For instance, an ellipse with unequal axes has exactly four vertices, namely the points where the axes meet the ellipse (see Exercise 3). It is an interesting global fact that this is the least number of vertices for all closed convex curves.

THEOREM 2 (The Four-Vertex Theorem). *A simple closed convex curve has at least four vertices.*

Before starting the proof, we need a lemma.

LEMMA. *Let* $\alpha\colon [0, l] \to \mathrm{R}^2$ *be a plane closed curve parametrized by arc length and let* A, B, *and* C *be arbitrary real numbers. Then*

$$\int_0^l (\mathrm{Ax} + \mathrm{By} + \mathrm{C}) \frac{\mathrm{dk}}{\mathrm{ds}}\, \mathrm{ds} = 0, \tag{5}$$

where the functions $\mathrm{x} = \mathrm{x(s)}$, $\mathrm{y} = \mathrm{y(s)}$ *are given by* $\alpha(\mathrm{s}) = (\mathrm{x(s)}, \mathrm{y(s)})$, *and* k *is the curvature of* α.

Proof of the Lemma. Recall that there exists a differentiable function $\theta\colon [0, l] \to R$ such that $x'(s) = \cos\theta$, $y'(s) = \sin\theta$. Thus, $k(s) = \theta'(s)$ and

$$x'' = -ky', \quad y'' = kx'.$$

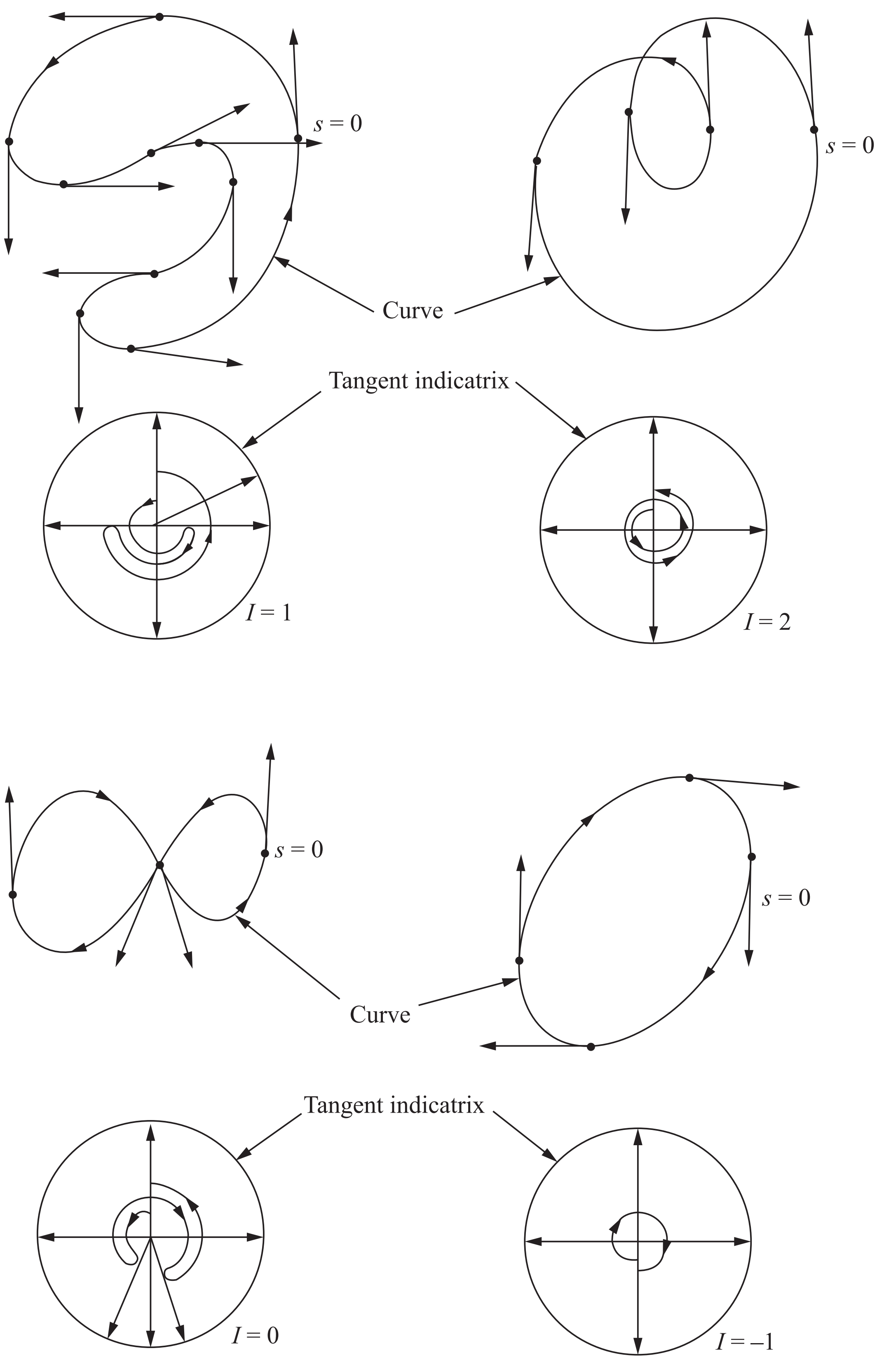
Therefore, since the functions involved agree at 0 and l,

$$\int_0^l k'\, ds = 0,$$

$$\int_0^l xk'\, ds = -\int_0^l kx'\, ds = -\int_0^l y''\, ds = 0,$$

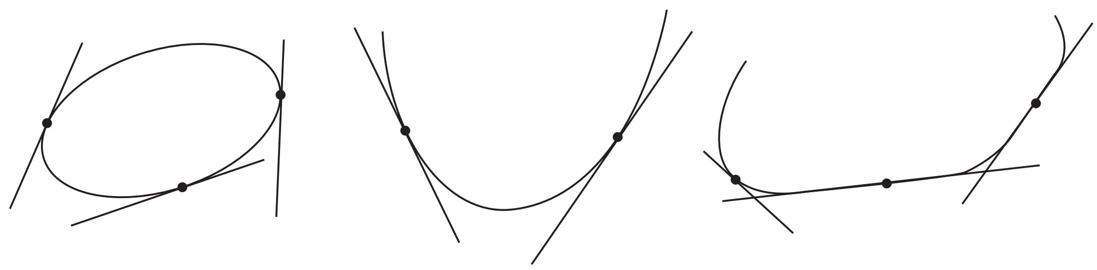
$$\int_0^l yk'\, ds = -\int_0^l ky'\, ds = \int_0^l x''\, ds = 0.$$

Q.E.D.

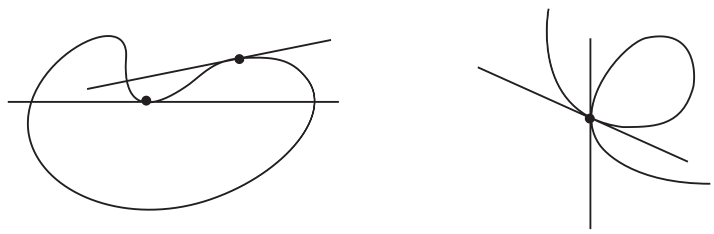


s = 0
s = 0
Curve
Tangent indicatrix
I = 1
I = 2
s = 0
s = 0
Curve
Tangent indicatrix
I = 0
I = –1

Figure 1-27



Convex curves



Nonconvex curves

Figure 1-28

Proof of the Theorem. Parametrize the curve by arc length, $\alpha\colon [0, l] \to R^2$. Since $k = k(s)$ is *a continuous function on the closed interval* $[0, l]$, *it reaches a maximum and a minimum on* $[0, l]$ (this is a basic fact in real functions; a proof can be found, for instance, in the appendix to Chap. 5, Prop. 13). Thus, α has at least two vertices, $\alpha(s_1) = p$ and $\alpha(s_2) = q$. Let L be the straight line passing through p and q, and let β and γ be the two arcs of α which are determined by the points p and q.

We claim that each of these arcs lies on a definite side of L. Otherwise, it meets L in a point r distinct from p and q (Fig. 1-29(a)). By convexity, and since p, q, r are distinct points on C, the tangent line at the intermediate point, say p, has to agree with L. Again, by convexity, this implies that L is tangent to C at the three points p, q, and r. But then the tangent to a point near p (the intermediate point) will have q and r on distinct sides, unless the whole segment rq of L belongs to C (Fig. 1-29(b)). This implies that $k = 0$ at p and q. Since these are points of maximum and minimum for k, $k \equiv 0$ on C, a contradiction.

Let $Ax + By + C = 0$ be the equation of L. If there are no further vertices, $k'(s)$ keeps a constant sign on each of the arcs β and γ. We can then arrange the signs of all the coefficients A, B, C so that the integral in Eq. (5) is positive. This contradiction shows that there is a third vertex and that $k'(s)$ changes sign on β or γ, say, on β. Since p and q are points of maximum and minimum, $k'(s)$ changes sign twice on β. Thus, there is a fourth vertex. **Q.E.D.**

The four-vertex theorem has been the subject of many investigations. The theorem also holds for simple, closed (not necessarily convex) curves, but the proof is harder. For further literature on the subject, see Reference [10].

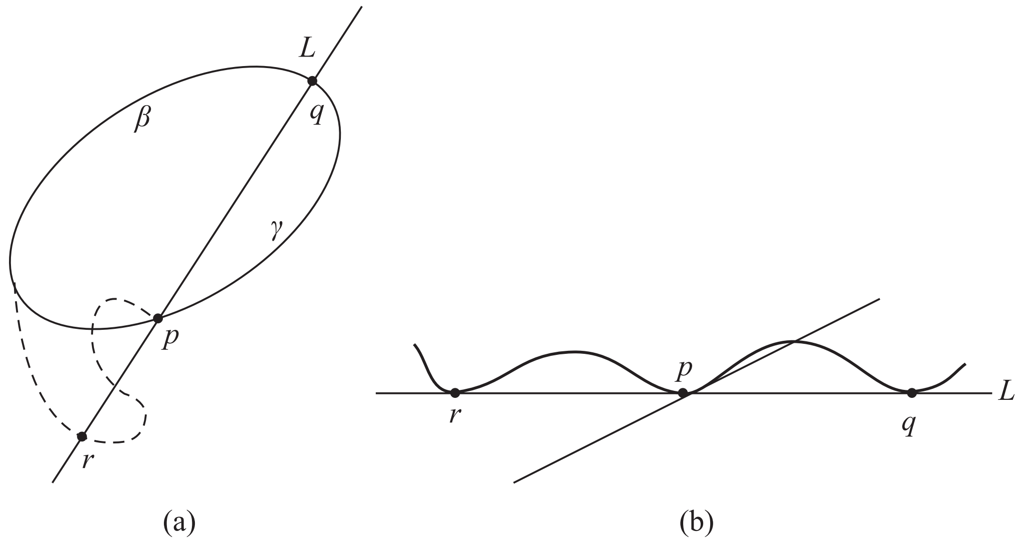


Figure 1-29

Later (Sec. 5-7, Prop. 1) we shall prove that *a plane closed curve is convex if and only if it is simple and can be oriented so that its curvature is positive or zero*. From that, and the proof given above, we see that we can reformulate the statement of the four-vertex theorem as follows. *The curvature function of a closed convex curve is* (*nonnegative and*) *either constant or else has at least two maxima and two minima*. It is then natural to ask whether such curvature functions do characterize the convex curves. More precisely, we can ask the following question. *Let* $\mathrm{k}\colon [\mathrm{a}, \mathrm{b}] \to \mathrm{R}$ *be a differentiable nonnegative function such that* k *agrees, with all its derivatives, at* a *and* b. *Assume that* k *is either constant or else has at least two maxima and two minima. Is there a simple closed curve* $\alpha\colon [\mathrm{a}, \mathrm{b}] \to \mathrm{R}^2$ *such that the curvature of* α *at* t *is* $\mathrm{k}(\mathrm{t})$*?*

For the case where $k(t)$ is strictly positive, H. Gluck answered the above question affirmatively (see H. Gluck, "The Converse to the Four Vertex Theorem," *L'Enseignement Mathématique* T. XVII, fasc. 3–4 (1971), 295–309). His methods, however, do not apply to the case $k \geq 0$.

C. The Cauchy-Crofton Formula

Our last topic in this section will be dedicated to finding a theorem which, roughly speaking, describes the following situation. Let C be a regular curve in the plane. We look at all straight lines in the plane that meet C and assign to each such line a *multiplicity* which is the number of its intersection points with C (Fig. 1-30).

We first want to find a way of assigning a measure to a given subset of straight lines in the plane. It should not be too surprising that this is possible. After all, we assign a measure (area) to point subsets of the plane. Once we realize that a straight line can be determined by two parameters (for instance,

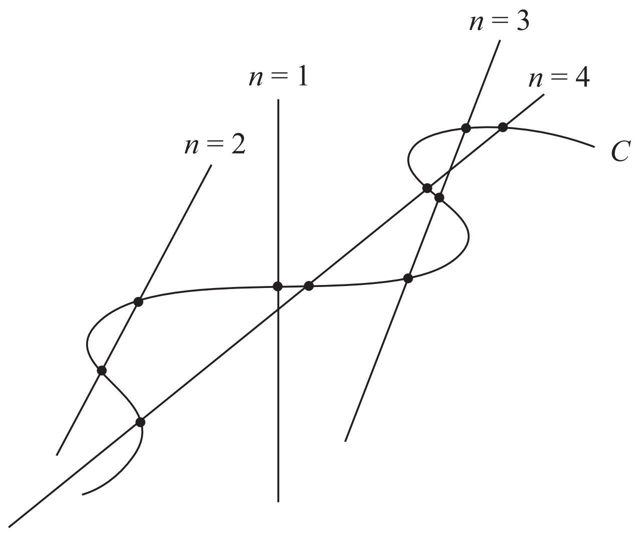


Figure 1-30. n is the multiplicity of the corresponding straight line.

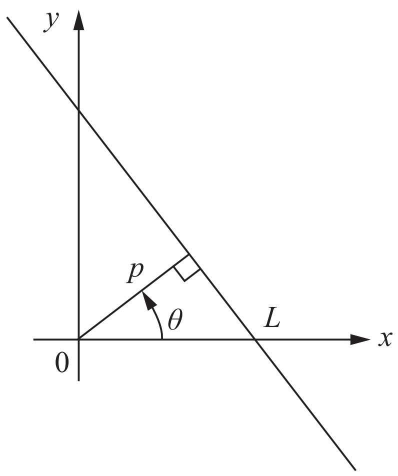


Figure 1-31. L is determined by p and θ.

p and θ in Fig. 1-31), we can think of the straight lines in the plane as points in a region of a certain plane. Thus, what we want is to find a "reasonable" way of measuring "areas" in such a plane.

Having chosen this measure, we want to apply it and find the measure of the set of straight lines (counted with multiplicities) which meet C. The result is quite interesting and can be stated as follows.

THEOREM 3 (The Cauchy-Crofton Formula). *Let* C *be a regular plane curve with length l. The measure of the set of straight lines* (*counted with multiplicities*) *which meet* C *is equal to* 2*l*.

Before going into the proof we must define what we mean by a reasonable measure in the set of straight lines in the plane. First, let us choose a convenient system of coordinates for such a set. A straight line L in the plane is determined by a unit vector $v = (\cos\theta, \sin\theta)$ normal to L and the inner product $p = v \cdot \alpha = x\cos\theta + y\sin\theta$ of v with the position vector $\alpha = (x, y)$ of L. Notice that to determine L in terms of the parameters (p, θ), we must identify $(p, \theta) \sim (p, \theta + 2k\pi)$, k an integer, and also identify $(p, \theta) \sim (-p, \theta \pm \pi)$.

Thus we can replace the set of all straight lines in the plane by the set

$$\mathfrak{L} = \{(p, \theta) \in R^2; (p, \theta) \sim (p, \theta + 2k\pi) \text{ and } (p, \theta) \sim (-p, \theta \pm \pi)\}.$$

We will show that, up to a choice of units, there is only one reasonable measure in this set.

To decide what we mean by reasonable, let us look more closely at the usual measure of areas in R^2. We need a definition.

A *rigid motion* in R^2 is a map $F: R^2 \to R^2$ given by $(\bar{x}, \bar{y}) \to (x, y)$, where (Fig. 1-32)

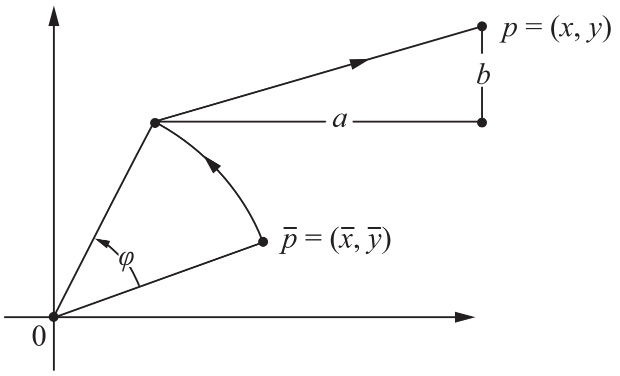


Figure 1-32

$$
\begin{aligned}
x &= a + \bar{x}\cos\varphi - \bar{y}\sin\varphi \\
y &= b + \bar{x}\sin\varphi + \bar{y}\cos\varphi.
\end{aligned} \tag{6}
$$

Now, to define the area of a set $S \subset R^2$ we consider the double integral

$$\iint_S dx\,dy;$$

that is, we integrate the "element of area" $dx\,dy$ over S. When this integral exists in some sense, we say that S is *measurable* and define the area of S as the value of the above integral. From now on, we shall assume that all the integrals involved in our discussions do exist.

Notice that we could have chosen some other element of area, say, $xy^2\,dx\,dy$. The reason for the choice of $dx\,dy$ is that, up to a factor, this is the only element of area that is invariant under rigid motions. More precisely, we have the following proposition.

PROPOSITION 1. *Let* $f(x, y)$ *be a continuous function defined in* R^2. *For any set* $S \subset R^2$, *define the area* A *of* S *by*

$$A(S) = \iint_S f(x, y)\,dx\,dy$$

(of course, we are considering only those sets for which the above integral exists). Assume that A *is invariant under rigid motions; that is, if* S *is any set and* $\bar{S} = F^{-1}(S)$, *where* F *is the rigid motion (6), we have*

$$A(\bar{S}) = \iint_{\bar{S}} f(\bar{x}, \bar{y})\,d\bar{x}\,d\bar{y} = \iint_S f(x, y)\,dx\,dy = A(S).$$

Then $f(x, y) = const.$

Proof. We recall the formula for change of variables in multiple integrals (Buck, *Advanced Calculus*, p. 301, or Exercise 15 of this section):

$$\iint_S f(x, y)\,dx\,dy = \iint_{\bar S} f(x(\bar x, \bar y), y(\bar x, \bar y))\frac{\partial(x, y)}{\partial(\bar x, \bar y)}d\bar x\,d\bar y. \tag{7}$$

Here, $x = x(\bar x, \bar y)$, $y = y(\bar x, \bar y)$ are functions with continuous partial derivatives which define the transformation of variables $T\colon R^2 \to R^2$, $\bar S = T^{-1}(S)$, and

$$\frac{\partial(x, y)}{\partial(\bar x, \bar y)} = \begin{vmatrix} \dfrac{\partial x}{\partial \bar x} & \dfrac{\partial x}{\partial \bar y} \\ \dfrac{\partial y}{\partial \bar x} & \dfrac{\partial y}{\partial \bar y} \end{vmatrix}$$

is the Jacobian of the transformation T. In our particular case, the transformation is the rigid motion (6) and the Jacobian is

$$\frac{\partial(x, y)}{\partial(\bar x, \bar y)} = \begin{vmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{vmatrix} = 1.$$

By using this fact and Eq. (7), we obtain

$$\iint_{\bar S} f(x(\bar x, \bar y), y(\bar x, \bar y))\,d\bar x\,d\bar y = \iint_{\bar S} f(\bar x, \bar y)\,d\bar x\,d\bar y.$$

Since this is true for all S, we have

$$f(x(\bar x, \bar y), y(\bar x, \bar y)) = f(\bar x, \bar y).$$

We now use the fact that for any pair of points (x, y), $(\bar x, \bar y)$ in R^2 there exists a rigid motion F such that $F(\bar x, \bar y) = (x, y)$. Thus,

$$f(x, y) = (f \circ F)(\bar x, \bar y) = f(\bar x, \bar y),$$

and $f(x, y) = \text{const.}$, as we wished. **Q.E.D.**

Remark 3. The above proof rests upon two facts: first, that the Jacobian of a rigid motion is 1, and, second, that the rigid motions are transitive on points of the plane; that is, given two points in the plane there exists a rigid motion taking one point into the other.

With these preparations, we can finally define a measure in the set $\mathfrak{L}$. We first observe that the rigid motion (6) induces a transformation on $\mathfrak{L}$. In fact, Eq. (6) maps the line $x\cos\theta + y\sin\theta = p$ into the line

$$\bar{x}\cos(\theta-\varphi)+\bar{y}\sin(\theta-\varphi)=p-a\cos\theta-b\sin\theta.$$

This means that the transformation induced by Eq. (6) on $\mathfrak{L}$ is

$$\bar{p}=p-a\cos\theta-b\sin\theta,$$
$$\bar{\theta}=\theta-\varphi.$$

It is easily checked that the Jacobian of the above transformation is 1 and that such transformations are also transitive on the set of lines in the plane. We then define the measure of a set $\mathfrak{S}\subset\mathfrak{L}$ as

$$\iint_{\mathfrak{S}} dp\, d\theta.$$

In the same way as in Prop. 1, we can then prove that this is, up to a constant factor, the only measure on $\mathfrak{L}$ that is invariant under rigid motions. This measure is, therefore, as reasonable as it can be.

We can now sketch a proof of Theorem 3.

Sketch of Proof of Theorem 3. First assume that the curve C is a segment of a straight line with length l. Since our measure is invariant under rigid motions, we can assume that the coordinate system has its origin 0 in the middle point of C and that the x axis is in the direction of C. Then the measure of the set of straight lines that meet C is (Fig. 1-33)

$$\iint dp\, d\theta=\int_0^{2\pi}\left(\int_0^{|\cos\theta|(l/2)} dp\right)d\theta=\int_0^{2\pi}\frac{l}{2}|\cos\theta|\,d\theta=2l.$$

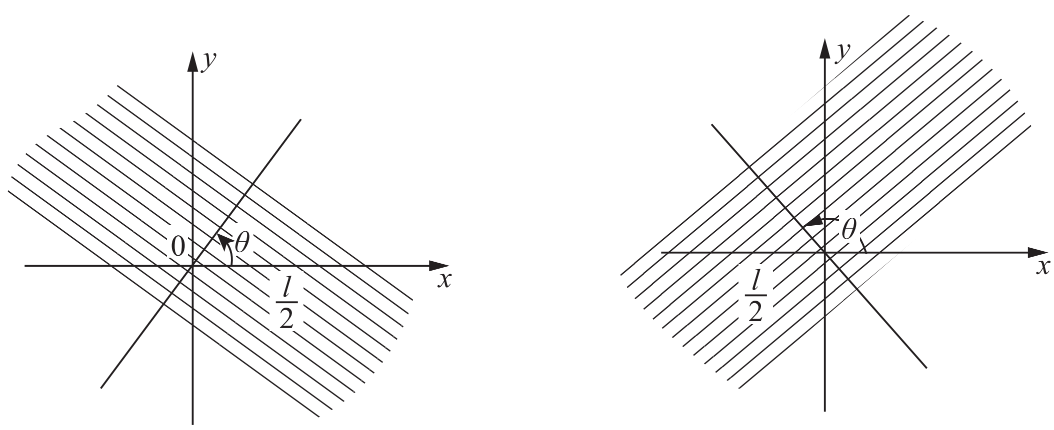


Figure 1-33

Next, let C be a polygonal line composed of a finite number of segments C_i with length $l_i(\sum l_i=l)$. Let $n=n(p,\theta)$ be the number of intersection points of the straight line (p, θ) with C. Then, by summing up the results for each segment C_i, we obtain

$$\iint n \, dp \, d\theta = 2 \sum_i l_i = 2l, \tag{8}$$

which is the Cauchy-Crofton formula for a polygonal line.

Finally, by a limiting process, it is possible to extend the above formula to any regular curve, and this will prove Theorem 3. **Q.E.D.**

It should be remarked that the general ideas of this topic belong to a branch of geometry known under the name of integral geometry. A survey of the subject can be found in L. A. Santaló, "Integral Geometry," in *Studies in Global Geometry and Analysis*, edited by S. S. Chern, The Mathematical Association of America, 1967, 147–193.

The Cauchy-Crofton formula can be used in many ways. For instance, if a curve is not rectifiable (see Exercise 9, Sec. 1-3) but the left-hand side of Eq. (8) has a meaning, this can be used to define the "length" of such a curve. Equation (8) can also be used to obtain an efficient way of estimating lengths of curves. Indeed, a good approximation for the integral in Eq. (8) is given as follows.† Consider a family of parallel straight lines such that two consecutive lines are at a distance r. Rotate this family by angles of $\pi/4$, $2\pi/4$, $3\pi/4$ in order to obtain four families of straight lines. Let n be the number of intersection points of a curve C with all these lines. Then

$$\frac{1}{2} nr \frac{\pi}{4}$$

is an approximation to the integral

$$\frac{1}{2} \iint n \, dp \, d\theta = \text{length of } C$$

and therefore gives an estimate for the length of C. To have an idea of how good this estimate can be, let us work out an example.

Example. Figure 1-34 is a drawing of an electron micrograph of a circular DNA molecule and we want to estimate its length. The four families of straight lines at a distance of 7 millimeters and angles of $\pi/4$ are drawn over the picture (a more practical way would be to have this family drawn once and for all on

†I am grateful to Robert Gardner (1939–1998) for suggesting this application and the example that follows.

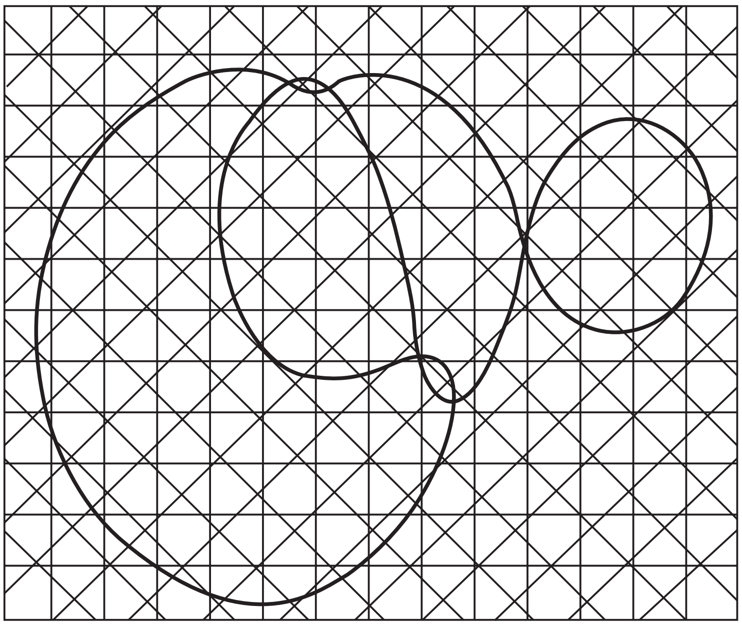


Figure 1-34. Reproduced from H. Ris and B. C. Chandler, *Cold Spring Harbor Symp. Quant. Biol.* 28, 2 (1963), with permission.

transparent paper). The number of intersection points is found to be 153. Thus,

$$\frac{1}{2}n\frac{\pi}{4} = \frac{1}{2}153\frac{3.14}{4} \sim 60.$$

Since the reference line in the picture represents 1 micrometer ($=10^{-6}$ meter) and measures, in our scale, 25 millimeters, $r = \frac{7}{25}$, and thus the length of this DNA molecule, from our values, is approximately

$$60\left(\frac{7}{25}\right) = 16.8 \text{ micrometers.}$$

The actual value is 16.3 micrometers.

EXERCISES

***1.** Is there a simple closed curve in the plane with length equal to 6 feet and bounding an area of 3 square feet?

***2.** Let $\overline{AB}$ be a segment of straight line and let $l >$ length of AB. Show that the curve C joining A and B, with length l, and such that together with $\overline{AB}$ bounds the largest possible area is an arc of a circle passing through A and B (Fig. 1-35).

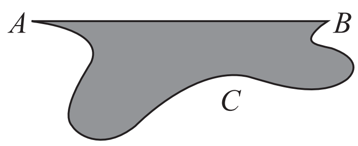


Figure 1-35

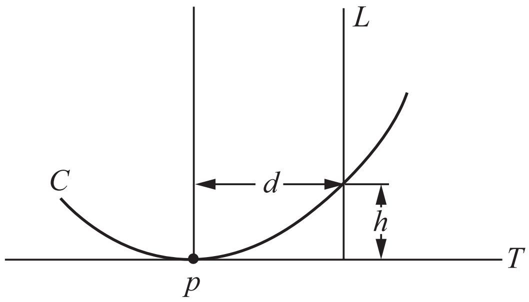


Figure 1-36

3. Compute the curvature of the ellipse

$$x = a\cos t, \quad y = b\sin t, \quad t \in [0, 2\pi], a \neq b,$$

and show that it has exactly four vertices, namely, the points $(a, 0)$, $(-a, 0)$, $(0, b)$, $(0, -b)$.

***4.** Let C be a plane curve and let T be the tangent line at a point $p \in C$. Draw a line L parallel to the normal line at p and at a distance d of p (Fig. 1-36). Let h be the length of the segment determined on L by C and T (thus, h is the "height" of C relative to T). Prove that

$$|k(p)| = \lim_{d\to 0} \frac{2h}{d^2},$$

where $k(p)$ is the curvature of C at p.

***5.** If a closed plane curve C is contained inside a disk of radius r, prove that there exists a point $p \in C$ such that the curvature k of C at p satisfies $|k| \geq 1/r$.

6. Let $\alpha(s)$, $s \in [0, l]$ be a closed convex plane curve positively oriented. The curve

$$\beta(s) = \alpha(s) - rn(s),$$

where r is a positive constant and n is the normal vector, is called a *parallel* curve to α (Fig. 1-37). Show that

a. Length of β = length of $\alpha + 2\pi r$.

b. $A(\beta) = A(\alpha) + rl + \pi r^2$.

c. $k_\beta(s) = k_\alpha(s)/(1 + rk_\alpha(s))$.

For (a)-(c), $A(\ )$ denotes the area bounded by the corresponding curve, and k_α, k_β are the curvatures of α and β, respectively.

7. Let $\alpha\colon R \to R^2$ be a plane curve defined in the entire real line R. Assume that α does not pass through the origin $0 = (0, 0)$ and that both

$$\lim_{t\to+\infty} |\alpha(t)| = \infty \quad \text{and} \quad \lim_{t\to-\infty} |\alpha(t)| = \infty.$$

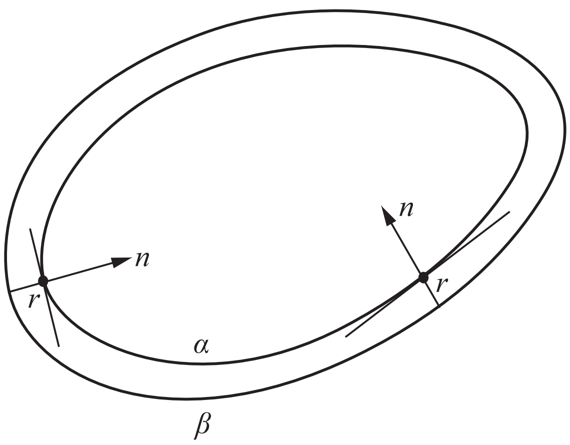


Figure 1-37

a. Prove that there exists a point $t_0 \in R$ such that $|\alpha(t_0)| \leq |\alpha(t)|$ for all $t \in R$.

b. Show, by an example, that the assertion in part a is false if one does not assume that both $\lim_{t\to+\infty} |\alpha(t)| = \infty$ and $\lim_{t\to-\infty} |\alpha(t)| = \infty$.

8.*a. Let $\alpha(s)$, $s \in [0, l]$, be a plane simple closed curve. Assume that the curvature $k(s)$ satisfies $0 < k(s) \leq c$, where c is a constant (thus, α is less curved than a circle of radius $1/c$). Prove that

$$\text{length of } \alpha \geq \frac{2\pi}{c}.$$

b. In part a replace the assumption of being simple by "α has rotation index N." Prove that

$$\text{length of } \alpha \geq \frac{2\pi N}{c}.$$

***9.** A set $K \subset R^2$ is *convex* if given any two points $p, q \in K$ the segment of straight line $\overline{pq}$ is contained in K (Fig. 1-38). Prove that a simple closed convex curve bounds a convex set.

10. Let C be a convex plane curve. Prove geometrically that C has no self-intersections.

***11.** Given a nonconvex simple closed plane curve C, we can consider its *convex hull* H (Fig. 1-39), that is, the boundary of the smallest convex set containing the interior of C. The curve H is formed by arcs of C and by the segments of the tangents to C that bridge "the nonconvex gaps" (Fig. 1-39). It can be proved that H is a C^1 closed convex curve. Use this to show that, in the isoperimetric problem, we can restrict ourselves to convex curves.

***12.** Consider a unit circle S^1 in the plane. Show that the ratio $M_1/M_2 = \frac{1}{2}$, where M_2 is the measure of the set of straight lines in the plane that meet S^1 and M_1 is the measure of all such lines that determine in S^1 a chord

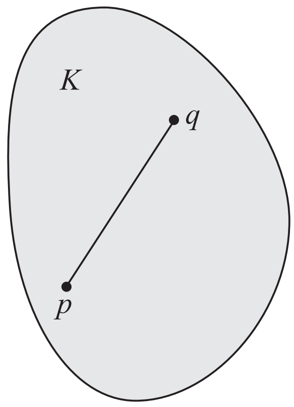


Figure 1-38

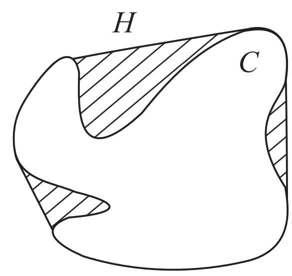


Figure 1-39

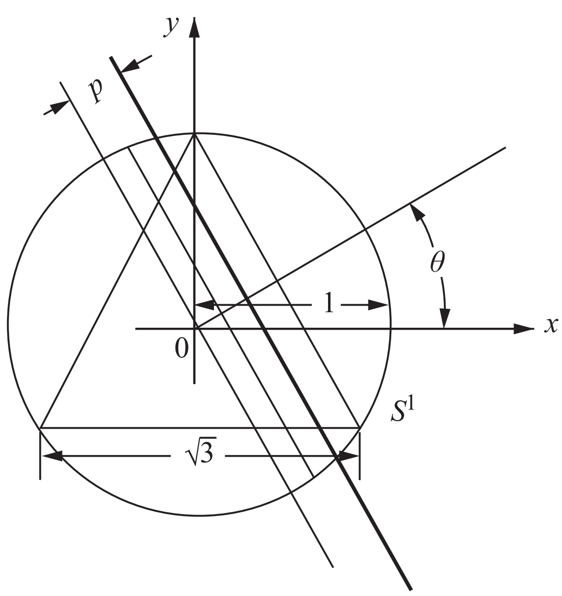


Figure 1-40

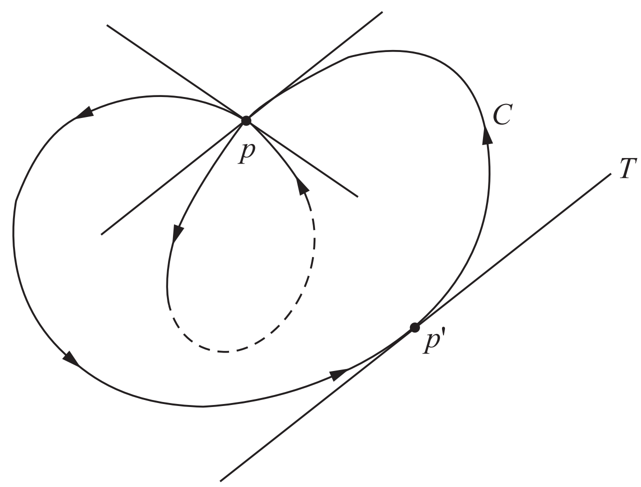


Figure 1-41

of length $> \sqrt{3}$. Intuitively, this ratio is the probability that a straight line that meets S^1 determines in S^1 a chord longer than the side of an equilateral triangle inscribed in S^1 (Fig. 1-40).

13. Let C be an oriented plane closed curve with curvature $k > 0$. Assume that C has at least one point p of self-intersection. Prove that

a. There is a point $p' \in C$ such that the tangent line T' at p' is parallel to some tangent at p.

b. The rotation angle of the tangent in the positive arc of C made up by $pp'p$ is $> \pi$ (Fig. 1-41).

c. The rotation index of C is ≥ 2.

14. a. Show that if a straight line L meets a closed convex curve C, then either L is tangent to C or L intersects C in exactly two points.

b. Use part a to show that the measure of the set of lines that meet C (without multiplicities) is equal to the length of C.

15. Green's theorem in the plane is a basic fact of calculus and can be stated as follows. Let a simple closed plane curve be given by $\alpha(t) = (x(t), y(t))$, $t \in [a, b]$. Assume that α is positively oriented, let C be its trace, and let R be the interior of C. Let $p = p(x, y)$, $q = q(x, y)$ be real functions with continuous partial derivatives p_x, p_y, q_x, q_y. Then

$$\iint_R (q_x - p_y)\, dx\, dy = \int_C \left(p\frac{dx}{dt} + q\frac{dy}{dt} \right) dt, \tag{9}$$

where in the second integral it is understood that the functions p and q are restricted to α and the integral is taken between the limits $t = a$ and $t = b$. In parts a and b below we propose to derive, from Green's theorem, a formula for the area of R and the formula for the change of variables in double integrals (cf. Eqs. (1) and (7) in the text).

a. Set $q = x$ and $p = -y$ in Eq. (9) and conclude that

$$A(R) = \iint_R dx\, dy = \frac{1}{2}\int_a^b \left(x(t)\frac{dy}{dt} - y(t)\frac{dx}{dt} \right) dt.$$

b. Let $f(x, y)$ be a real function with continuous partial derivatives and $T\colon R^2 \to R^2$ be a transformation of coordinates given by the functions $x = x(u, v)$, $y = y(u, v)$, which also admit continuous partial derivatives. Choose in Eq. (9) $p = 0$ and q so that $q_x = f$. Apply successively Green's theorem, the map T, and Green's theorem again to obtain

$$\begin{aligned} &\iint_R f(x, y)\, dx\, dy \\ &\quad = \int_C q\, dy = \int_{T^{-1}(C)} (q \circ T)(y_u u'(t) + y_v v'(t))\, dt \\ &\quad = \iint_{T^{-1}(R)} \left\{ \frac{\partial}{\partial u}((q \circ T) y_v) - \frac{\partial}{\partial v}((q \circ T) y_u) \right\} du\, dv. \end{aligned}$$

Show that

$$\begin{aligned} &\frac{\partial}{\partial u}(q(x(u, v), y(u, v)) y_v) - \frac{\partial}{\partial v}(q(x(u, v), y(u, v)) y_u) \\ &\quad = f(x(u, v), y(u, v))(x_u y_v - x_v y_u) = f\frac{\partial(x, y)}{\partial(u, v)}. \end{aligned}$$

Put that together with the above and obtain the transformation formula for double integrals:

$$\iint_R f(x, y)\, dx\, dy = \iint_{T^{-1}(R)} f(x(u, v), y(u, v)) \frac{\partial(x, y)}{\partial(u, v)}\, du\, dv.$$

2 Regular Surfaces

2-1. Introduction

In this chapter, we shall begin the study of surfaces. Whereas in the first chapter we used mainly elementary calculus of one variable, we shall now need some knowledge of calculus of several variables. Specifically, we need to know some facts about continuity and differentiability of functions and maps in R^2 and R^3. What we need can be found in any standard text of advanced calculus, for instance, Buck *Advanced Calculus*; we have included a brief review of some of this material in an appendix to Chap. 2.

In Sec. 2-2 we shall introduce the basic concept of a regular surface in R^3. In contrast to the treatment of curves in Chap. 1, regular surfaces are defined as sets rather than maps. The goal of Sec. 2-2 is to describe some criteria that are helpful in trying to decide whether a given subset of R^3 is a regular surface.

In Sec. 2-3 we shall show that it is possible to define what it means for a function on a regular surface to be differentiable, and in Sec. 2-4 we shall show that the usual notion of differential in R^2 can be extended to such functions. Thus, regular surfaces in R^3 provide a natural setting for two-dimensional calculus.

Of course, curves can also be treated from the same point of view, that is, as subsets of R^3 which provide a natural setting for one-dimensional calculus. We shall mention them briefly in Sec. 2-3.

Sections 2-2 and 2-3 are crucial to the rest of the book. A beginner may find the proofs in these sections somewhat difficult. If so, the proofs can be omitted on a first reading.

In Sec. 2-5 we shall introduce the first fundamental form, a natural instrument to treat metric questions (lengths of curves, areas of regions, etc.) on a regular surface. This will become a very important issue when we reach Chap. 4.

Sections 2-6 through 2-8 are optional on a first reading. In Sec. 2-6, we shall treat the idea of orientation on regular surfaces. This will be needed in Chaps. 3 and 4. For the benefit of those who omit this section, we shall review the notion of orientation at the beginning of Chap. 3.

2-2. Regular Surfaces; Inverse Images of Regular Values†

In this section we shall introduce the notion of a regular surface in R^3. Roughly speaking, a regular surface in R^3 is obtained by taking pieces of a plane, deforming them, and arranging them in such a way that the resulting figure has no sharp points, edges, or self-intersections and so that it makes sense to speak of a tangent plane at points of the figure. The idea is to define a set that is, in a certain sense, two-dimensional and that also is smooth enough so that the usual notions of calculus can be extended to it. By the end of Sec. 2-4, it should be completely clear that the following definition is the right one.

DEFINITION 1. *A subset* $S \subset R^3$ *is a* regular surface *if, for each* $p \in S$, *there exists a neighborhood* V *in* R^3 *and a map* $\mathbf{x}: U \to V \cap S$ *of an open set* $U \subset R^2$ *onto* $V \cap S \subset R^3$ *such that* (*Fig.* 2-1)

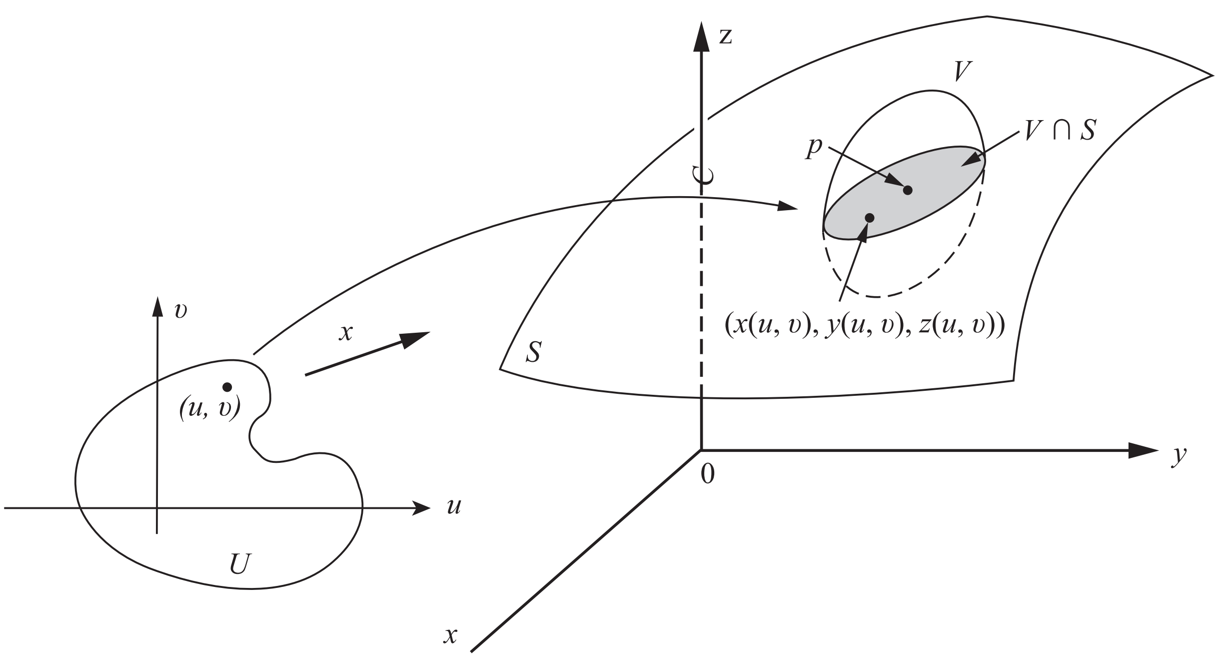


Figure 2-1

1. $\mathbf{x}$ *is differentiable. This means that if we write*

$$\mathbf{x}(u, v) = (x(u, v), y(u, v), z(u, v)), \quad (u, v) \in U,$$

†Proofs in this section may be omitted on a first reading.

the functions $\mathrm{x(u, v), y(u, v), z(u, v)}$ *have continuous partial derivatives of all orders in* U.

2. $\mathbf{x}$ *is a homeomorphism. Since* $\mathbf{x}$ *is continuous by condition* 1*, this means that* $\mathbf{x}$ *has an inverse* $\mathbf{x}^{-1}$*:* $\mathrm{V \cap S \to U}$ *which is continuous.*
3. (*The regularity condition.*) *For each* $\mathrm{q \in U}$*, the differential* $d\mathbf{x}_{\mathrm{q}}$: $\mathrm{R^2 \to R^3}$ *is one-to-one.*†

We shall explain condition 3 in a short while.

The mapping $\mathbf{x}$ is called a *parametrization* or a *system of* (local) *coordinates* in (a neighborhood of) p. The neighborhood $V \cap S$ of p in S is called a *coordinate neighborhood.*

To give condition 3 a more familiar form, let us compute the matrix of the linear map $d\mathbf{x}_q$ in the canonical bases $e_1 = (1, 0)$, $e_2 = (0, 1)$ of R^2 with coordinates (u, v) and $f_1 = (1, 0, 0)$, $f_2 = (0, 1, 0)$, $f_3 = (0, 0, 1)$ of R^3, with coordinates (x, y, z).

Let $q = (u_0, v_0)$. The vector e_1 is tangent to the curve $u \to (u, v_0)$ whose image under $\mathbf{x}$ is the curve

$$u \longrightarrow (x(u, v_0), y(u, v_0), z(u, v_0)).$$

This image curve (called the *coordinate curve* $v = v_0$) lies on S and has at $\mathbf{x}(q)$ the tangent vector (Fig. 2-2)

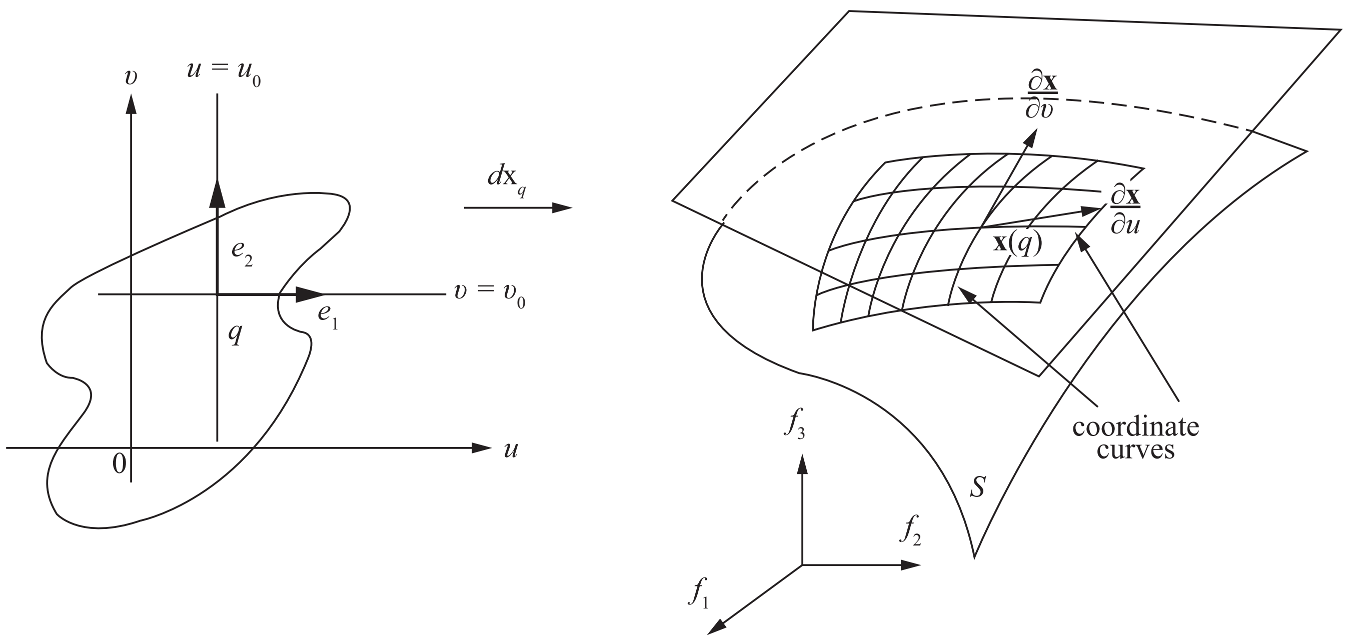


Figure 2-2

†In italic context, letter symbols are roman so they can be distinguished from the surrounding text.

$$\left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u}\right) = \frac{\partial \mathbf{x}}{\partial u},$$

where the derivatives are computed at (u_0, v_0) and a vector is indicated by its components in the basis $\{f_1, f_2, f_3\}$. By the definition of differential (appendix to Chap. 2, Def. 1),

$$d\mathbf{x}_q(e_1)\left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u}\right) = \frac{\partial \mathbf{x}}{\partial u}.$$

Similarly, using the coordinate curve $u = u_0$ (image by $\mathbf{x}$ of the curve $v \to (u_0, v)$), we obtain

$$d\mathbf{x}_q(e_2) = \left(\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v}\right) = \frac{\partial \mathbf{x}}{\partial v}.$$

Thus, the matrix of the linear map $d\mathbf{x}_q$ in the referred basis is

$$d\mathbf{x}_q = \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \\ \dfrac{\partial z}{\partial u} & \dfrac{\partial z}{\partial v} \end{pmatrix}.$$

Condition 3 of Def. 1 may now be expressed by requiring the two column vectors of this matrix to be linearly independent; or, equivalently, that the vector product $\partial\mathbf{x}/\partial u \wedge \partial\mathbf{x}/\partial v \neq 0$; or, in still another way, that one of the minors of order 2 of the matrix of $d\mathbf{x}_q$, that is, one of the Jacobian determinants

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}, \quad \frac{\partial(y, z)}{\partial(u, v)}, \quad \frac{\partial(x, z)}{\partial(u, v)},$$

be different from zero at q.

Remark 1. Definition 1 deserves a few comments. First, in contrast to our treatment of curves in Chap. 1, we have defined a surface as a subset S of R^3, and not as a map. This is achieved by covering S with the traces of parametrizations which satisfy conditions 1, 2, and 3.

Condition 1 is very natural if we expect to do some differential geometry on S. The one-to-oneness in condition 2 has the purpose of preventing self-intersections in regular surfaces. This is clearly necessary if we are to speak about, say, *the* tangent plane at a point $p \in S$ (see Fig. 2-3(a)). The continuity of the inverse in condition 2 has a more subtle purpose which can be fully understood only in the next section. For the time being, we shall mention that this condition is essential to proving that certain objects defined in terms of a parametrization do not depend on this parametrization but only on the set S itself. Finally, as we shall show in Sec. 2.4, condition 3 will guarantee the existence of a "tangent plane" at all points of S (see Fig. 2-3(b)).

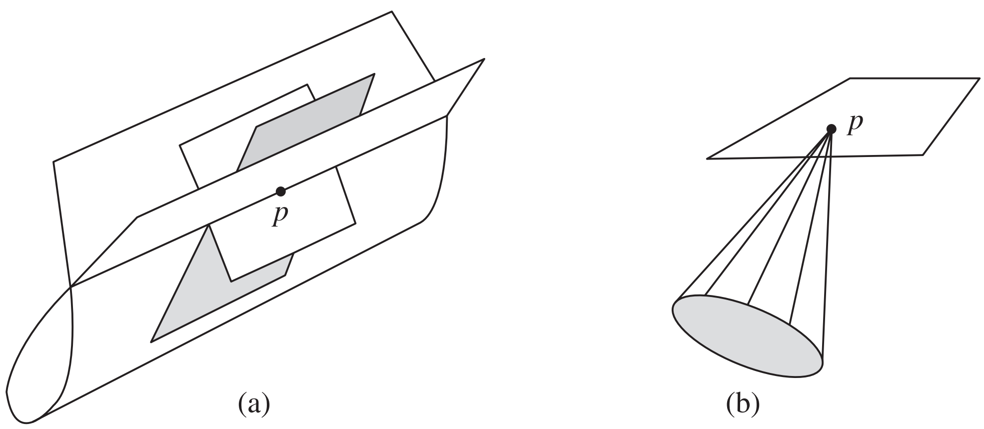


Figure 2-3. Some situations to be avoided in the definition of a regular surface.

Example 1. Let us show that the unit sphere

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

is a regular surface.

We first verify that the map $\mathbf{x}_1: U \subset R^2 \to R^3$ given by

$$\mathbf{x}_1(x, y) = (x, y, +\sqrt{1 - (x^2 + y^2)}\,), \quad (x, y) \in U,$$

where $R^2 = \{(x, y, z) \in R^3; z = 0\}$ and $U = \{(x, y) \in R^2; x^2 + y^2 < 1\}$, is a parametrization of S^2. Observe that $\mathbf{x}_1(U)$ is the (open) part of S^2 above the xy plane.

Since $x^2 + y^2 < 1$, the function $+\sqrt{1 - (x^2 + y^2)}$ has continuous partial derivatives of all orders. Thus, $\mathbf{x}_1$ is differentiable and condition 1 holds.

Condition 3 is easily verified, since

$$\frac{\partial(x, y)}{\partial(x, y)} \equiv 1.$$

To check condition 2, we observe that $\mathbf{x}_1$ is one-to-one and that $\mathbf{x}_1^{-1}$ is the restriction of the (continuous) projection $\pi(x, y, z) = (x, y)$ to the set $\mathbf{x}_1(U)$. Thus, $\mathbf{x}_1^{-1}$ is continuous in $\mathbf{x}_1(U)$.

We shall now cover the whole sphere with similar parametrizations as follows. We define $\mathbf{x}_2\colon U \subset R^2 \to R^3$ by

$$\mathbf{x}_2(x, y) = (x, y, -\sqrt{1-(x^2+y^2)}\,),$$

check that $\mathbf{x}_2$ is a parametrization, and observe that $\mathbf{x}_1(U) \cup \mathbf{x}_2(U)$ covers S^2 minus the equator

$$\{(x, y, z) \in R^3; x^2 + y^2 = 1, z = 0\}.$$

Then, using the xz and zy planes, we define the parametrizations

$$\begin{aligned}
\mathbf{x}_3(x, z) &= (x, +\sqrt{1-(x^2+z^2)}\,, z),\\
\mathbf{x}_4(x, z) &= (x, -\sqrt{1-(x^2+z^2)}\,, z),\\
\mathbf{x}_5(y, z) &= (+\sqrt{1-(y^2+z^2)}\,, y, z),\\
\mathbf{x}_6(y, z) &= (-\sqrt{1-(y^2+z^2)}\,, y, z),
\end{aligned}$$

which, together with $\mathbf{x}_1$ and $\mathbf{x}_2$, cover S^2 completely (Fig. 2-4) and show that S^2 is a regular surface.

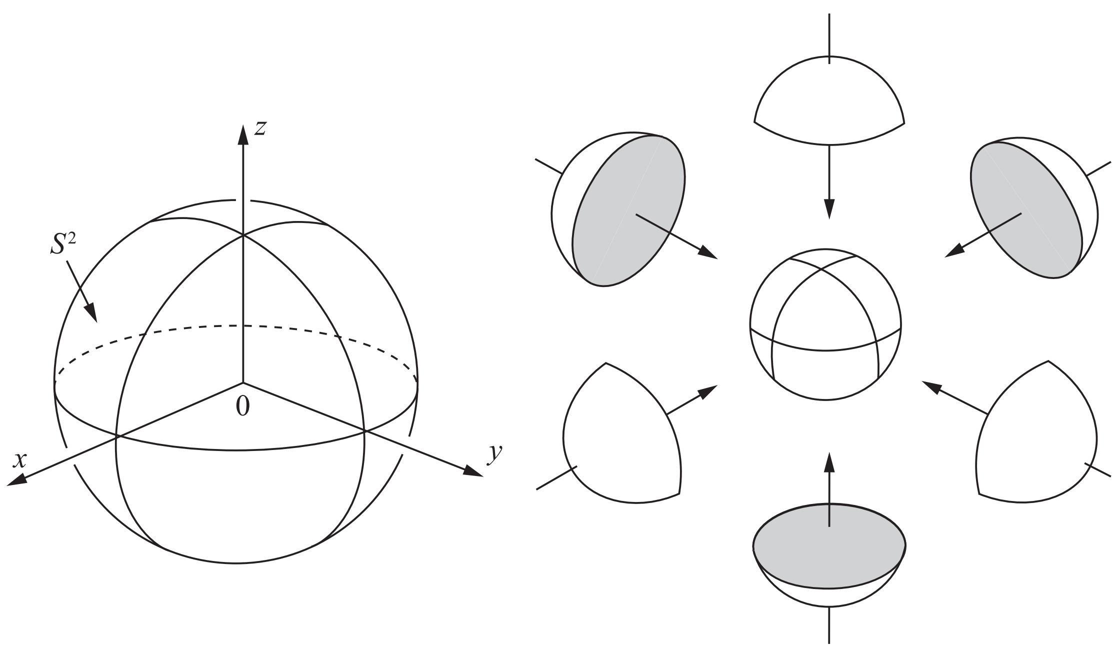


Figure 2-4

For most applications, it is convenient to relate parametrizations to the geographical coordinates on S^2. Let $V = \{(\theta, \varphi); 0 < \theta < \pi, 0 < \varphi < 2\pi\}$ and let $\mathbf{x}\colon V \to R^3$ be given by

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

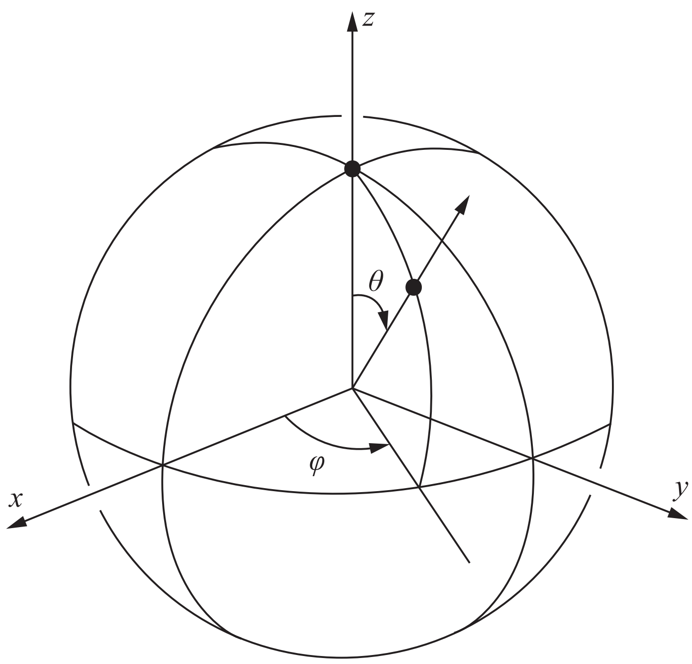


Figure 2-5

Clearly, $\mathbf{x}(V) \subset S^2$. We shall prove that $\mathbf{x}$ is a parametrization of S^2. θ is usually called the *colatitude* (the complement of the latitude) and φ the *longitude* (Fig. 2-5).

It is clear that the functions $\sin\theta\cos\varphi$, $\sin\theta\sin\varphi$, $\cos\theta$ have continuous partial derivatives of all orders; hence, $\mathbf{x}$ is differentiable. Moreover, in order that the Jacobian determinants

$$\frac{\partial(x, y)}{\partial(\theta, \varphi)} = \cos\theta\sin\theta,$$
$$\frac{\partial(y, z)}{\partial(\theta, \varphi)} = \sin^2\theta\cos\varphi,$$
$$\frac{\partial(x, z)}{\partial(\theta, \varphi)} = -\sin^2\theta\sin\varphi$$

vanish simultaneously, it is necessary that

$$\cos^2\theta\sin^2\theta + \sin^4\theta\cos^2\varphi + \sin^4\theta\sin^2\varphi = \sin^2\theta = 0.$$

This does not happen in V, and so conditions 1 and 3 of Def. 1 are satisfied.

Next, we observe that given $(x, y, z) \in S^2 - C$, where C is the semicircle

$$C = \{(x, y, z) \in S^2;\ y = 0, x \geq 0\},$$

θ is uniquely determined by $\theta = \cos^{-1} z$, since $0 < \theta < \pi$. By knowing θ, we find $\sin\varphi$ and $\cos\varphi$ from $x = \sin\theta\cos\varphi$, $y = \sin\theta\sin\varphi$, and this determines φ uniquely $(0 < \varphi < 2\pi)$. It follows that $\mathbf{x}$ has an inverse $\mathbf{x}^{-1}$. To complete the verification of condition 2, we should prove that $\mathbf{x}^{-1}$ is continuous. However, since we shall soon prove (Prop. 4) that this verification is not necessary provided we already know that the set S is a regular surface, we shall not do that here.

We remark that $\mathbf{x}(V)$ only omits a semicircle of S^2 (including the two poles) and that S^2 can be covered with the coordinate neighborhoods of two parametrizations of this type.

In Exercise 16 we shall indicate how to cover S^2 with another useful set of coordinate neighborhoods.

Example 1 shows that deciding whether a given subset of R^3 is a regular surface directly from the definition may be quite tiresome. Before going into further examples, we shall present two propositions which will simplify this task. Proposition 1 shows the relation which exists between the definition of a regular surface and the graph of a function $z = f(x, y)$. Proposition 2 uses the inverse function theorem and relates the definition of a regular surface with the subsets of the form $f(x, y, z) =$ constant.

PROPOSITION 1. *If* $f\colon U \to R$ *is a differentiable function in an open set* U *of* R^2*, then the graph of* f*, that is, the subset of* R^3 *given by* $(x, y, f(x, y))$ *for* $(x, y) \in U$*, is a regular surface.*

Proof. It suffices to show that the map $\mathbf{x}\colon U \to R^3$ given by

$$\mathbf{x}(u, v) = (u, v, f(u, v))$$

is a parametrization of the graph whose coordinate neighborhood covers every point of the graph. Condition 1 is clearly satisfied, and condition 3 also offers no difficulty since $\partial(x, y)/\partial(u, v) \equiv 1$. Finally, each point (x, y, z) of the graph is the image under $\mathbf{x}$ of the unique point $(u, v) = (x, y) \in U$. $\mathbf{x}$ is therefore one-to-one, and since $\mathbf{x}^{-1}$ is the restriction to the graph of f of the (continuous) projection of R^3 onto the xy plane, $\mathbf{x}^{-1}$ is continuous.

Q.E.D.

Before stating Prop. 2, we shall need a definition.

DEFINITION 2. *Given a differentiable map* $F\colon U \subset R^n \to R^m$ *defined in an open set* U *of* R^n *we say that* $p \in U$ *is a* critical point *of* F *if the differential* $dF_p\colon R^n \to R^m$ *is not a surjective (or onto) mapping. The image* $F(p) \in R^m$ *of a critical point is called a* critical value *of* F. *A point of* R^m *which is not a critical value is called a* regular value *of* F.

The terminology is evidently motivated by the particular case in which $f\colon U \subset R \to R$ is a real-valued function of a real variable. A point $x_0 \in U$ is critical if $f'(x_0) = 0$, that is, if the differential df_{x_0} carries all the vectors in R to the zero vector (Fig. 2-6). Notice that any point $a \notin f(U)$ is trivially a regular value of f.

If $f\colon U \subset R^3 \to R$ is a differentiable function, then df_p applied to the vector $(1, 0, 0)$ is obtained by calculating the tangent vector at $f(p)$ to the curve

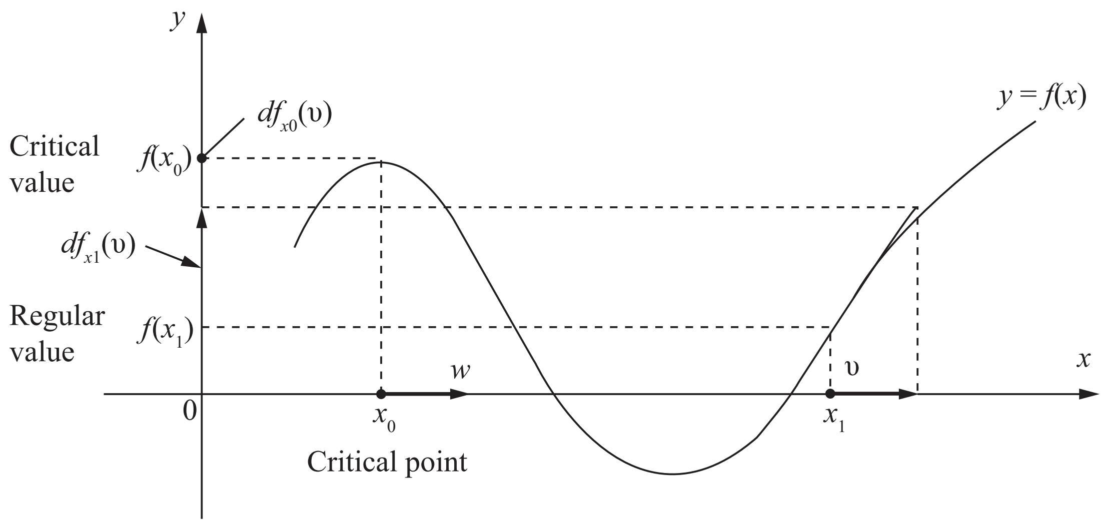


Figure 2-6

$$x \longrightarrow f(x, y_0, z_0).$$

It follows that

$$df_p(1, 0, 0) = \frac{\partial f}{\partial x}(x_0, y_0, z_0) = f_x$$

and analogously that

$$df_p(0, 1, 0) = f_y, \quad df_p(0, 0, 1) = f_z.$$

We conclude that the matrix of df_p in the basis $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ is given by

$$df_p = (f_x, f_y, f_z).$$

Note, in this case, that to say that df_p is not surjective is equivalent to saying that $f_x = f_y = f_z = 0$ at p. Hence, $a \in f(U)$ is a regular value of $f: U \subset R^3 \to R$ if and only if f_x, f_y, and f_z do not vanish simultaneously at any point in the inverse image

$$f^{-1}(a) = \{(x, y, z) \in U: f(x, y, z) = a\}.$$

PROPOSITION 2. *If* $\mathrm{f}\colon \mathrm{U} \subset \mathrm{R}^3 \to \mathrm{R}$ *is a differentiable function and* $\mathrm{a} \in \mathrm{f}(\mathrm{U})$ *is a regular value of* f, *then* $\mathrm{f}^{-1}(\mathrm{a})$ *is a regular surface in* R^3.

Proof. Let $p = (x_0, y_0, z_0)$ be a point of $f^{-1}(a)$. Since a is a regular value of f, it is possible to assume, by renaming the axes if necessary, that $f_z \neq 0$ at p. We define a mapping $F: U \subset R^3 \to R^3$ by

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

and we indicate by (u, v, t) the coordinates of a point in R^3 where F takes its values. The differential of F at p is given by

$$dF_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ f_x & f_y & f_z \end{pmatrix},$$

whence

$$\det(dF_p) = f_z \neq 0.$$

We can therefore apply the inverse function theorem (cf. the appendix to Chap. 2), which guarantees the existence of neighborhoods V of p and W of $F(p)$ such that $F: V \to W$ is invertible and the inverse $F^{-1}: W \to V$ is differentiable (Fig. 2-7). It follows that the coordinate functions of F^{-1}, i.e., the functions

$$x = u, \quad y = v, \quad z = g(u, v, t), \quad (u, v, t) \in W,$$

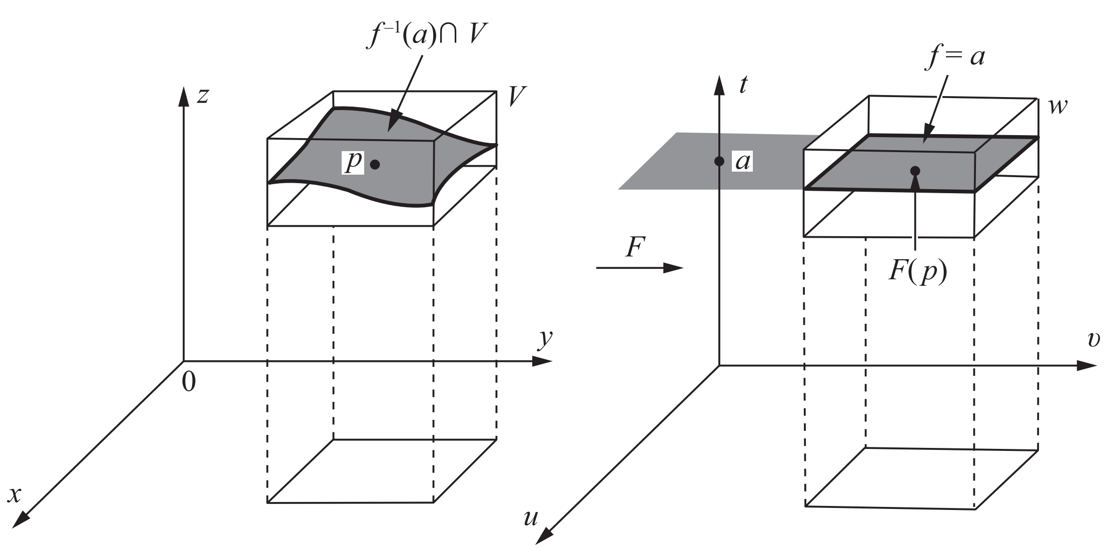


Figure 2-7

are differentiable. In particular, $z = g(u, v, a) = h(x, y)$ is a differentiable function defined in the projection of V onto the xy plane. Since

$$F(f^{-1}(a) \cap V) = W \cap \{(u, v, t); t = a\},$$

we conclude that the graph of h is $f^{-1}(a) \cap V$. By Prop. 1, $f^{-1}(a) \cap V$ is a coordinate neighborhood of p. Therefore, every $p \in f^{-1}(a)$ can be covered by a coordinate neighborhood, and so $f^{-1}(a)$ is a regular surface. **Q.E.D.**

Remark 2. The proof consists essentially of using the inverse function theorem "to solve for z" in the equation $f(x, y, z) = a$, which can be done in a neighborhood of p if $f_z(p) \neq 0$. This fact is a special case of the general implicit function theorem, which follows from the inverse function theorem and is, in fact, equivalent to it.

Example 2. The ellipsoid

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

is a regular surface. In fact, it is the set $f^{-1}(0)$ where

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

is a differentiable function and 0 is a regular value of f. This follows from the fact that the partial derivatives $f_x = 2x/a^2$, $f_y = 2y/b^2$, $f_z = 2z/c^2$ vanish simultaneously only at the point $(0, 0, 0)$, which does not belong to $f^{-1}(0)$. This example includes the sphere as a particular case ($a = b = c = 1$).

The examples of regular surfaces presented so far have been connected subsets of R^3. A surface $S \subset R^3$ is said to be *connected* if any two of its points can be joined by a continuous curve in S. In the definition of a regular surface we made no restrictions on the connectedness of the surfaces, and the following example shows that the regular surfaces given by Prop. 2 may not be connected.

Example 3. The hyperboloid of two sheets $-x^2 - y^2 + z^2 = 1$ is a regular surface, since it is given by $S = f^{-1}(0)$, where 0 is a regular value of $f(x, y, z) = -x^2 - y^2 + z^2 - 1$ (Fig. 2-8). Note that the surface S is not connected; that is, given two points in two distinct sheets ($z > 0$ and $z < 0$) it is not possible to join them by a continuous curve $\alpha(t) = (x(t), y(t), z(t))$ contained in the surface; otherwise, z changes sign and, for some t_0, we have $z(t_0) = 0$, which means that $\alpha(t_0) \notin S$.

Incidentally, the argument of Example 3 may be used to prove a property of connected surfaces that we shall use repeatedly. *If* $\mathrm{f}\colon \mathrm{S} \subset \mathrm{R}^3 \to \mathrm{R}$ *is a nonzero continuous function defined on a connected surface* S, *then* f *does not change sign on* S.

To prove this, we use the intermediate value theorem (appendix to Chap. 2, Prop. 4). Assume, by contradiction, that $f(p) > 0$ and $f(q) < 0$ for some points $p, q \in S$. Since S is connected, there exists a continuous curve $\alpha\colon [a, b] \to S$ with $\alpha(a) = p$, $\alpha(b) = q$. By applying the intermediate value theorem to the continuous function $f \circ \alpha\colon [a, b] \to R$, we find that there exists $c \in (a, b)$ with $f \circ \alpha(c) = 0$; that is, f is zero at $\alpha(c)$, a contradiction.

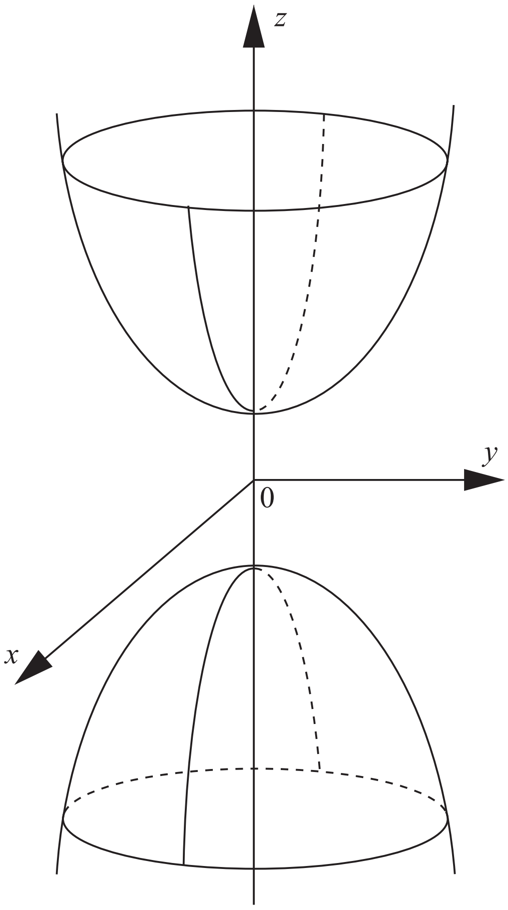


Figure 2-8. A nonconnected surface: $-y^2 - x^2 + z^2 = 1$.

Example 4. The torus T is a "surface" generated by rotating a circle S^1 of radius r about a straight line belonging to the plane of the circle and at a distance $a > r$ away from the center of the circle (Fig. 2-9).

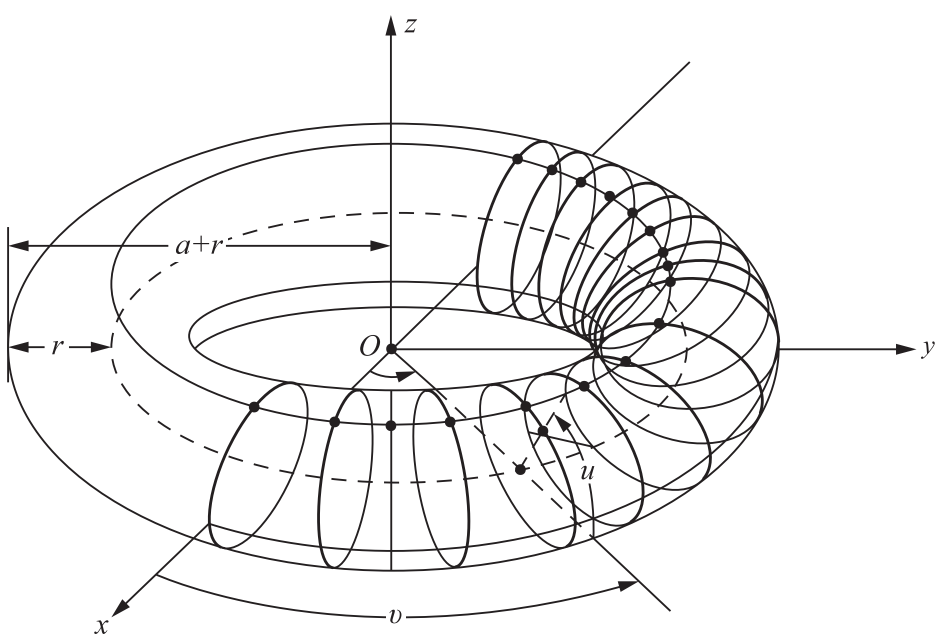


Figure 2-9

Let S^1 be the circle in the yz plane with its center in the point $(0, a, 0)$. Then S^1 is given by $(y - a)^2 + z^2 = r^2$, and the points of the figure T obtained by rotating this circle about the z axis satisfy the equation

$$z^2 = r^2 - (\sqrt{x^2 + y^2} - a)^2.$$

Therefore, T is the inverse image of r^2 by the function

$$f(x, y, z) = z^2 + (\sqrt{x^2 + y^2} - a)^2.$$

This function is differentiable for $(x, y) \neq (0, 0)$, and since

$$\frac{\partial f}{\partial z} = 2z, \quad \frac{\partial f}{\partial y} = \frac{2y(\sqrt{x^2 + y^2} - a)}{\sqrt{x^2 + y^2}},$$
$$\frac{\partial f}{\partial x} = \frac{2x(\sqrt{x^2 + y^2} - a)}{\sqrt{x^2 + y^2}},$$

r^2 is a regular value of f. It follows that the torus T is a regular surface.

Proposition 1 says that the graph of a differentiable function is a regular surface. The following proposition provides a local converse of this; that is, any regular surface is locally the graph of a differentiable function.

PROPOSITION 3. *Let* $S \subset R^3$ *be a regular surface and* $p \in S$*. Then there exists a neighborhood* V *of* p *in* S *such that* V *is the graph of a differentiable function which has one of the following three forms*: $z = f(x, y)$, $y = g(x, z)$, $x = h(y, z)$.

Proof. Let $\mathbf{x}$: $U \subset R^2 \to S$ be a parametrization of S in p, and write $\mathbf{x}(u, v) = (x(u, v), y(u, v), z(u, v))$, $(u, v) \in U$. By condition 3 of Def. 1, one of the Jacobian determinants

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

is not zero at $\mathbf{x}^{-1}(p) = q$.

Suppose first that $(\partial(x, y)/\partial(u, v))(q) \neq 0$, and consider the map $\pi \circ \mathbf{x}$: $U \to R^2$, where π is the projection $\pi(x, y, z) = (x, y)$. Then $\pi \circ \mathbf{x}(u, v) = (x(u, v), y(u, v))$, and, since $(\partial(x, y)/\partial(u, v))(q) \neq 0$, we can apply the inverse function theorem to guarantee the existence of neighborhoods V_1 of q, V_2 of $\pi \circ \mathbf{x}(q)$ such that $\pi \circ \mathbf{x}$ maps V_1 diffeomorphically onto V_2 (Fig. 2-10). It follows that π restricted to $\mathbf{x}(V_1) = V$ is one-to-one and that there is a differentiable inverse $(\pi \circ \mathbf{x})^{-1}$: $V_2 \to V_1$. Observe that, since $\mathbf{x}$ is a homeomorphism, V is a neighborhood of p in S. Now, if we compose

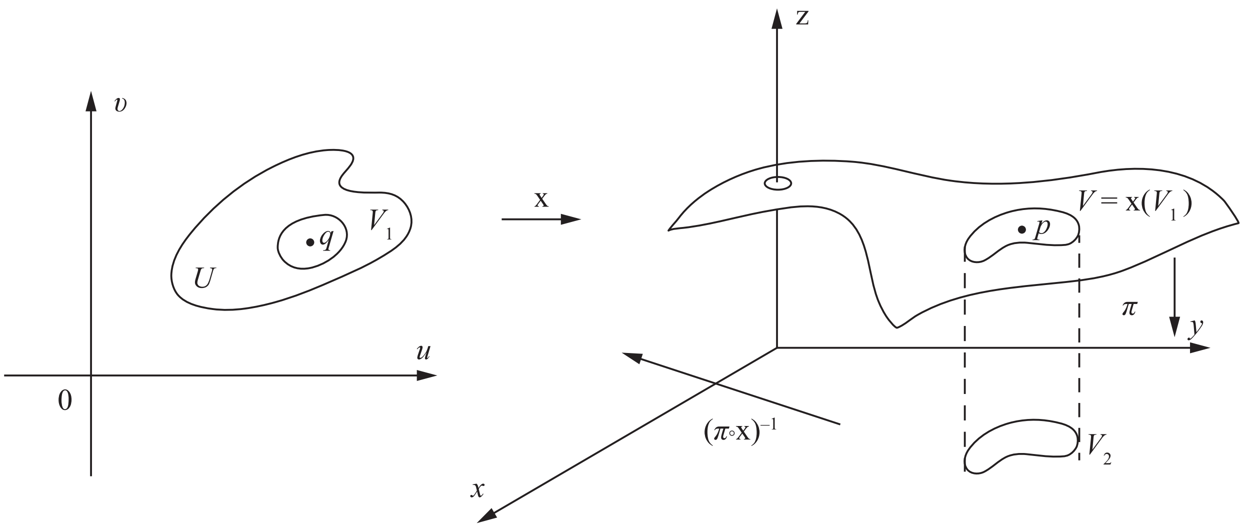


Figure 2-10

the map $(\pi \circ \mathbf{x})^{-1}$: $(x, y) \to (u(x, y), v(x, y))$ with the function $(u, v) \to z(u, v)$, we find that V is the graph of the differentiable function $z = z(u(x, y), v(x, y)) = f(x, y)$, and this settles the first case.

The remaining cases can be treated in the same way, yielding $x = h(y, z)$ and $y = g(x, z)$. **Q.E.D.**

The next proposition says that if we already know that S is a regular surface and we have a candidate $\mathbf{x}$ for a parametrization, we do not have to check that $\mathbf{x}^{-1}$ is continuous, provided that the other conditions hold. This remark was used in Example 1.

PROPOSITION 4. *Let* $p \in S$ *be a point of a regular surface* S *and let* $\mathbf{x}$: $U \subset R^2 \to R^3$ *be a map with* $p \in \mathbf{x}(U) \subset S$ *such that conditions* 1 *and* 3 *of Def.* 1 *hold. Assume that* $\mathbf{x}$ *is one-to-one. Then* $\mathbf{x}^{-1}$ *is continuous.*

Proof. We begin by proceeding in a way similar to the proof of Proposition 3. Write $\mathbf{x}(u, v) = (x(u, v), y(u, v), z(u, v))$, $(u, v) \in U$, and let $q \in U$. By conditions 1 and 3, we can assume, interchanging the coordinate axis if necessary, that $\partial(x, y)/\partial(u, v)(q) \neq 0$. Let $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$ be the projection $\pi(x, y, z) = (x, y)$. By the inverse function theorem, we obtain neighborhoods V_1 of q in U and V_2 of $\pi \circ \mathbf{x}(q)$ in $\mathbb{R}^2$ such that $\pi \circ \mathbf{x}$ applies V_1 diffeomorphically onto V_2.

Assume now that $\mathbf{x}$ is bijective. Then, restricted to $\mathbf{x}(V_1)$, $\mathbf{x}^{-1} = (\pi \circ \mathbf{x})^{-1} \circ \pi$ (Fig. 2.10). Thus $\mathbf{x}^{-1}$, as a composition of continuous maps, is continuous. **Q.E.D.**

Example 5. The one-sheeted cone C, given by

$$z = +\sqrt{x^2 + y^2}, \quad (x, y) \in R^2,$$

is not a regular surface. Observe that we cannot conclude this from the fact alone that the "natural" parametrization

$$(x, y) \to (x, y, +\sqrt{x^2 + y^2})$$

is not differentiable; there could be other parametrizations satisfying Def. 1.

To show that this is not the case, we use Prop. 3. If C were a regular surface, it would be, in a neighborhood of $(0, 0, 0) \in C$, the graph of a differentiable function having one of three forms: $y = h(x, z)$, $x = g(y, z)$, $z = f(x, y)$. The two first forms can be discarded by the simple fact that the projections of C over the xz and yz planes are not one-to-one. The last form would have to agree, in a neighborhood of $(0, 0, 0)$, with $z = +\sqrt{x^2 + y^2}$. Since $z = +\sqrt{x^2 + y^2}$ is not differentiable at $(0, 0)$, this is impossible.

Example 6. A parametrization for the torus T of Example 4 can be given by (Fig. 2-9)

$$\mathbf{x}(u, v) = ((r\cos u + a)\cos v, (r\cos u + a)\sin v, r\sin u),$$

where $0 < u < 2\pi$, $0 < v < 2\pi$.

Condition 1 of Def. 1 is easily checked, and condition 3 reduces to a straightforward computation, which is left as an exercise. Since we know that T is a regular surface, condition 2 is equivalent, by Prop. 4, to the fact that $\mathbf{x}$ is one-to-one.

To prove that $\mathbf{x}$ is one-to-one, we first observe that $\sin u = z/r$; also, if $\sqrt{x^2 + y^2} \leq a$, then $\pi/2 \leq u \leq 3\pi/2$, and if $\sqrt{x^2 + y^2} \geq a$, then either $0 < u \leq \pi/2$ or $3\pi/2 \leq u < 2\pi$. Thus, given (x, y, z), this determines u, $0 < u < 2\pi$, uniquely. By knowing u, x, and y we find $\cos v$ and $\sin v$. This determines v uniquely, $0 < v < 2\pi$. Thus, $\mathbf{x}$ is one-to-one.

It is easy to see that the torus can be covered by three such coordinate neighborhoods.

EXERCISES†

1. Show that the cylinder $\{(x, y, z) \in R^3; x^2 + y^2 = 1\}$ is a regular surface, and find parametrizations whose coordinate neighborhoods cover it.
2. Is the set $\{(x, y, z) \in R^3; z = 0 \text{ and } x^2 + y^2 \leq 1\}$ a regular surface? Is the set $\{(x, y, z) \in R^3; z = 0, \text{ and } x^2 + y^2 < 1\}$ a regular surface?
3. Show that the two-sheeted cone, with its vertex at the origin, that is, the set $\{(x, y, z) \in R^3; x^2 + y^2 - z^2 = 0\}$, is not a regular surface.

†Those who have omitted the proofs in this section should also omit Exercises 17–19.

4. Let $f(x, y, z) = z^2$. Prove that 0 is not a regular value of f and yet that $f^{-1}(0)$ is a regular surface.

***5.** Let $P = \{(x, y, z) \in R^3; x = y\}$ (a plane) and let $\mathbf{x}: U \subset R^2 \to R^3$ be given by

$$\mathbf{x}(u, v) = (u + v, u + v, uv),$$

where $U = \{(u, v) \in R^2; u > v\}$. Clearly, $\mathbf{x}(U) \subset P$. Is $\mathbf{x}$ a parametrization of P?

6. Give another proof of Prop. 1 by applying Prop. 2 to $h(x, y, z) = f(x, y) - z$.

7. Let $f(x, y, z) = (x + y + z - 1)^2$.

a. Locate the critical points and critical values of f.

b. For what values of c is the set $f(x, y, z) = c$ a regular surface?

c. Answer the questions of parts a and b for the function $f(x, y, z) = xyz^2$.

8. Let $\mathbf{x}(u, v)$ be as in Def. 1. Verify that $d\mathbf{x}_q: R^2 \to R^3$ is one-to-one if and only if

$$\frac{\partial \mathbf{x}}{\partial u} \wedge \frac{\partial \mathbf{x}}{\partial v} \neq 0.$$

9. Let V be an open set in the xy plane. Show that the set

$$\{(x, y, z) \in R^3; z = 0 \text{ and } (x, y) \in V\}$$

is a regular surface.

10. Let C be a figure "8" in the xy plane and let S be the cylindrical surface over C (Fig. 2-11); that is,

$$S = \{(x, y, z) \in R^3; (x, y) \in C\}.$$

Is the set S a regular surface?

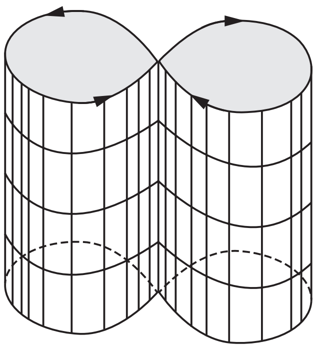


Figure 2-11

11. Show that the set $S = \{(x, y, z) \in R^3; z = x^2 - y^2\}$ is a regular surface and check that parts a and b are parametrizations for S:

a. $\mathbf{x}(u, v) = (u + v, u - v, 4uv)$, $(u, v) \in R^2$.

***b.** $\mathbf{x}(u, v) = (u \cosh v, u \sinh v, u^2)$, $(u, v) \in R^2$, $u \neq 0$.

Which parts of S do these parametrizations cover?

12. Show that $\mathbf{x}$: $U \subset R^2 \to R^3$ given by

$$\mathbf{x}(u, v) = (a \sin u \cos v, b \sin u \sin v, c \cos u), \quad a, b, c \neq 0,$$

where $0 < u < \pi$, $0 < v < 2\pi$, is a parametrization for the ellipsoid

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

Describe geometrically the curves $u = \text{const.}$ on the ellipsoid.

***13.** Find a parametrization for the hyperboloid of two sheets $\{(x, y, z) \in R^3; -x^2 - y^2 + z^2 = 1\}$.

14. A half-line $[0, \infty)$ is perpendicular to a line E and rotates about E from a given initial position while its origin 0 moves along E. The movement is such that when $[0, \infty)$ has rotated through an angle θ, the origin is at a distance $d = \sin^2(\theta/2)$ from its initial position on E. Verify that by removing the line E from the image of the rotating line we obtain a regular surface. If the movement were such that $d = \sin(\theta/2)$, what else would need to be excluded to have a regular surface?

***15.** Let two points $p(t)$ and $q(t)$ move with the same speed, p starting from $(0, 0, 0)$ and moving along the z axis and q starting at $(a, 0, 0)$, $a \neq 0$, and moving parallel to the y axis. Show that the line through $p(t)$ and $q(t)$ describes a set in R^3 given by $y(x - a) + zx = 0$. Is this a regular surface?

16. One way to define a system of coordinates for the sphere S^2, given by $x^2 + y^2 + (z - 1)^2 = 1$, is to consider the so-called *stereographic projection* π: $S^2 \sim \{N\} \to R^2$ which carries a point $p = (x, y, z)$ of the sphere S^2 minus the north pole $N = (0, 0, 2)$ onto the intersection of the xy plane with the straight line which connects N to p (Fig. 2-12). Let $(u, v) = \pi(x, y, z)$, where $(x, y, z) \in S^2 \sim \{N\}$ and $(u, v) \in xy$ plane.

a. Show that π^{-1}: $R^2 \to S^2$ is given by

$$\pi^{-1} \begin{cases} x = \dfrac{4u}{u^2 + v^2 + 4}, \\ y = \dfrac{4v}{u^2 + v^2 + 4}, \\ z = \dfrac{2(u^2 + v^2)}{u^2 + v^2 + 4}. \end{cases}$$

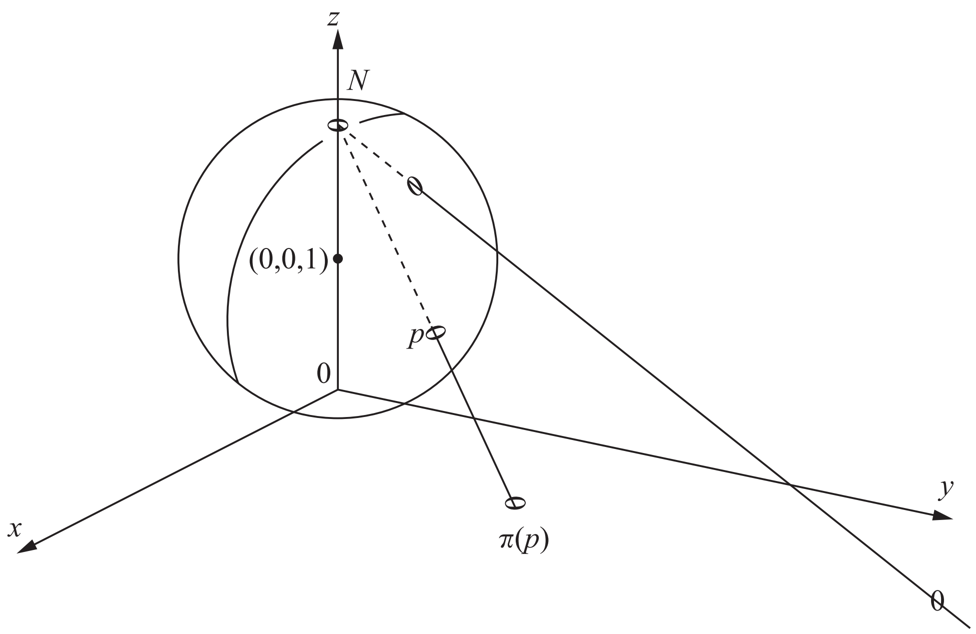


Figure 2-12. The stereographic projection.

b. Show that it is possible, using stereographic projection, to cover the sphere with two coordinate neighborhoods.

17. Define a regular curve in analogy with a regular surface. Prove that

a. The inverse image of a regular value of a differentiable function

$$f: U \subset R^2 \to R$$

is a regular plane curve. Give an example of such a curve which is not connected.

b. The inverse image of a regular value of a differentiable map

$$F: U \subset R^3 \to R^2$$

is a regular curve in R^3. Show the relationship between this proposition and the classical way of defining a curve in R^3 as the intersection of two surfaces.

***c.** The set $C = \{(x, y) \in R^2; x^2 = y^3\}$ is not a regular curve.

***18.** Suppose that $f(x, y, z) = u = \text{const.}$, $g(x, y, z) = v = \text{const.}$,

$$h(x, y, z) = w = \text{const.},$$

describe three families of regular surfaces and assume that at (x_0, y_0, z_0) the Jacobian

$$\frac{\partial(f, g, h)}{\partial(x, y, z)} \neq 0.$$

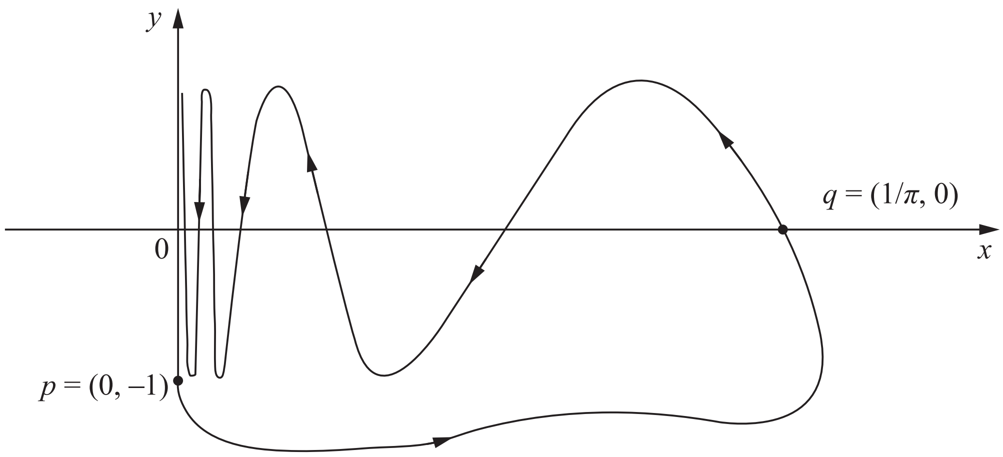


Figure 2-13

Prove that in a neighborhood of (x_0, y_0, z_0) the three families will be described by a mapping $F(u, v, w) = (x, y, z)$ of an open set of R^3 into R^3, where a local parametrization for the surface of the family $f(x, y, z) = u$, for example, is obtained by setting $u =$ const. in this mapping. Determine F for the case where the three families of surfaces are

$$f(x, y, z) = x^2 + y^2 + z^2 = u = \text{const.}; \quad \text{(spheres with center } (0, 0, 0));$$

$$g(x, y, z) = \frac{y}{x} = v = \text{const.}; \quad \text{(planes through the } z \text{ axis)};$$

$$h(x, y, z) = \frac{x^2 + y^2}{z^2} = w = \text{const.}; \quad \text{(cones with vertex at } (0, 0, 0)).$$

***19.** Let $\alpha\colon (-3, 0) \to R^2$ be defined by (Fig. 2-13)

$$\alpha(t) \begin{cases} = (0, -(t+2)), & \text{if } t \in (-3, -1), \\ = \text{regular parametrized curve joining } p = (0, -1) \text{ to } q = \left(\dfrac{1}{\pi}, 0\right), & \\ & \text{if } t \in \left(-1, -\dfrac{1}{\pi}\right), \\ = \left(-t, \sin \dfrac{1}{t}\right), & \text{if } t \in \left(-\dfrac{1}{\pi}, 0\right). \end{cases}$$

It is possible to define the curve joining p to q so that all the derivatives of α are continuous at the corresponding points and α has no self-intersections. Let C be the trace of α.

a. Is C a regular curve?

b. Let a normal line to the plane R^2 run through C so that it describes a "cylinder" S. Is S a regular surface?

2-3. Change of Parameters; Differentiable Functions on Surface†

Differential geometry is concerned with those properties of surfaces which depend on their behavior in a neighborhood of a point. The definition of a regular surface given in Sec. 2-2 is adequate for this purpose. According to this definition, each point p of a regular surface belongs to a coordinate neighborhood. The points of such a neighborhood are characterized by their coordinates, and we should be able, therefore, to define the local properties which interest us in terms of these coordinates.

For example, it is important that we be able to define what it means for a function $f: S \to R$ to be differentiable at a point p of a regular surface S. A natural way to proceed is to choose a coordinate neighborhood of p, with coordinates u, v, and say that f is differentiable at p if its expression in the coordinates u and v admits continuous partial derivatives of all orders.

The same point of S can, however, belong to various coordinate neighborhoods (in the sphere of Example 1 of Sec. 2-2 any point of the interior of the first octant belongs to three of the given coordinate neighborhoods). Moreover, other coordinate systems could be chosen in a neighborhood of p (the points referred to on the sphere could also be parametrized by geographical coordinates or by stereographic projection; cf. Exercise 16, Sec. 2-2). For the above definition to make sense, it is necessary that it does not depend on the chosen system of coordinates. In other words, it must be shown that when p belongs to two coordinate neighborhoods, with parameters (u, v) and (ξ, η), it is possible to pass from one of these pairs of coordinates to the other by means of a differentiable transformation.

The following proposition shows that this is true.

PROPOSITION 1 (Change of Parameters). *Let* p *be a point of a regular surface* S, *and let* $\mathbf{x}: \mathrm{U} \subset \mathrm{R}^2 \to \mathrm{S}$, $\mathbf{y}: \mathrm{V} \subset \mathrm{R}^2 \to \mathrm{S}$ *be two parametrizations of* S *such that* $\mathrm{p} \in \mathbf{x}(\mathrm{U}) \cap \mathrm{y}(\mathrm{V}) = \mathrm{W}$. *Then the "change of coordinates"* $\mathrm{h} = \mathbf{x}^{-1} \circ \mathbf{y}: \mathbf{y}^{-1}(\mathrm{W}) \to \mathbf{x}^{-1}(\mathrm{W})$ (Fig. 2-14) *is a diffeomorphism; that is,* h *is differentiable and has a differentiable inverse* h^{-1}.

In other words, if $\mathbf{x}$ and $\mathbf{y}$ are given by

$$\mathbf{x}(u, v) = (x(u, v), y(u, v), z(u, v)), \qquad (u, v) \in U,$$
$$\mathbf{y}(\xi, \eta) = (x(\xi, \eta), y(\xi, \eta), z(\xi, \eta)), \qquad (\xi, \eta) \in V,$$

then the change of coordinates h, given by

$$u = u(\xi, \eta), \quad v = v(\xi, \eta), \quad (\xi, \eta) \in \mathbf{y}^{-1}(W),$$

†Proofs in this section may be omitted on a first reading.

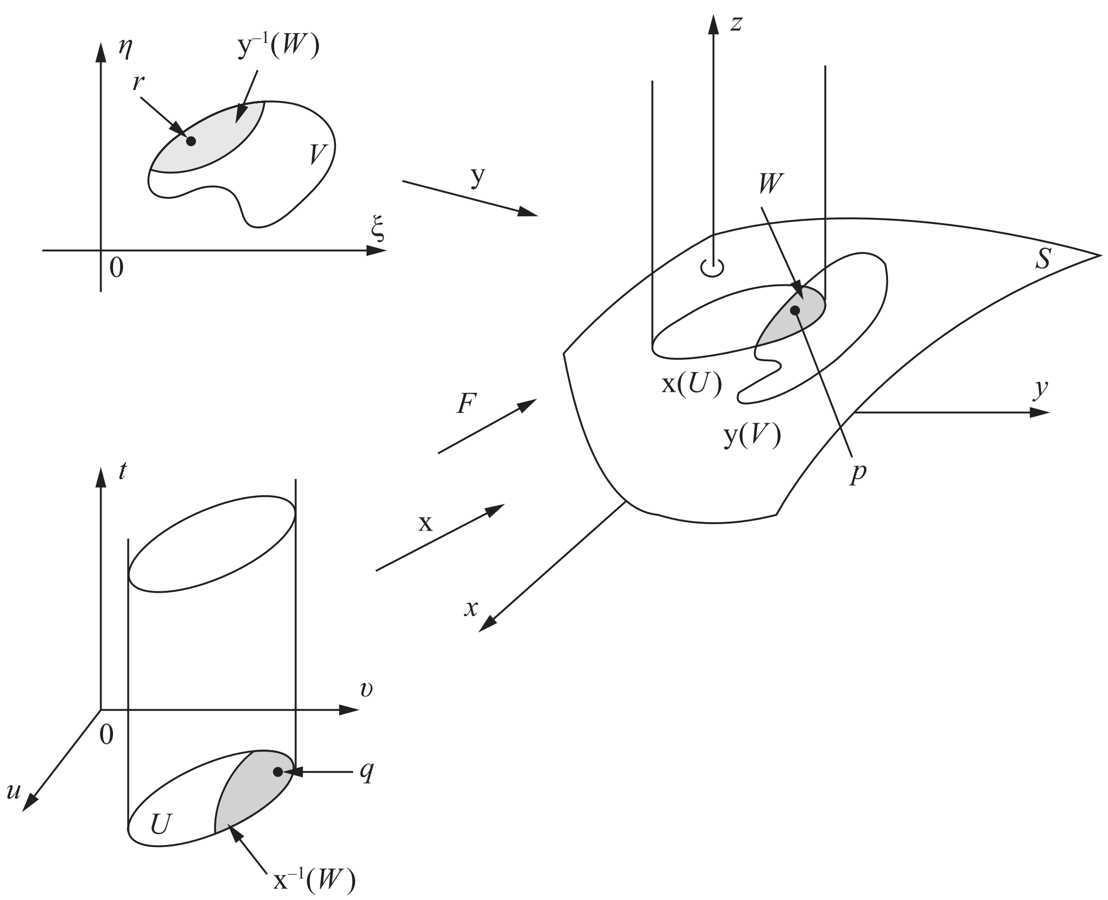


Figure 2-14

has the property that the functions u and v have continuous partial derivatives of all orders, and the map h can be inverted, yielding

$$\xi = \xi(u, v), \quad \eta = \eta(u, v), \quad (u, v) \in \mathbf{x}^{-1}(W),$$

where the functions ξ and η also have partial derivatives of all orders. Since

$$\frac{\partial(u, v)}{\partial(\xi, \eta)} \cdot \frac{\partial(\xi, \eta)}{\partial(u, v)} = 1,$$

this implies that the Jacobian determinants of both h and h^{-1} are nonzero everywhere.

Proof of Prop. 1. $h = \mathbf{x}^{-1} \circ \mathbf{y}$ is a homeomorphism, since it is composed of homeomorphisms (cf. the appendix to Chap. 2, Prop. 3). It is not possible to conclude, by an analogous argument, that h is differentiable, since $\mathbf{x}^{-1}$ is defined in an open subset of S, and we do not yet know what is meant by a differentiable function on S.

We proceed in the following way. Let $r \in \mathbf{y}^{-1}(W)$ and set $q = h(r)$. Since $\mathbf{x}(u, v) = (x(u, v), y(u, v), z(u, v))$ is a parametrization, we can assume, by renaming the axes if necessary, that

$$\frac{\partial(x, y)}{\partial(u, v)}(q) \neq 0.$$

We extend $\mathbf{x}$ to a map $F: U \times R \to R^3$ defined by

$$F(u, v, t) = (x(u, v), y(u, v), z(u, v) + t), \quad (u, v) \in U, t \in R.$$

Geometrically, F maps a vertical cylinder C over U into a "vertical cylinder" over $\mathbf{x}(U)$ by mapping each section of C with height t into the surface $\mathbf{x}(u, v) + te_3$, where e_3 is the unit vector of the z axis (Fig. 2-14).

It is clear that F is differentiable and that the restriction $F|U \times \{0\} = \mathbf{x}$. Calculating the determinant of the differential dF_q, we obtain

$$\begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} & 0 \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} & 0 \\ \dfrac{\partial z}{\partial u} & \dfrac{\partial z}{\partial v} & 1 \end{vmatrix} = \frac{\partial(x, y)}{\partial(u, v)}(q) \neq 0.$$

It is possible therefore to apply the inverse function theorem, which guarantees the existence of a neighborhood M of $\mathbf{x}(q)$ in R^3 such that F^{-1} exists and is differentiable in M.

By the continuity of $\mathbf{y}$, there exists a neighborhood N of r in V such that $\mathbf{y}(N) \subset M$ (appendix to Chap. 2, Prop. 2). Notice that, restricted to N, $h|N = F^{-1} \circ \mathbf{y}|N$ is a composition of differentiable maps. Thus, we can apply the chain rule for maps (appendix to Chap. 2, Prop. 8) and conclude that h is differentiable at r. Since r is arbitrary, h is differentiable on $\mathbf{y}^{-1}(W)$.

Exactly the same argument can be applied to show that the map h^{-1} is differentiable, and so h is a diffeomorphism. **Q.E.D.**

We shall now give an explicit definition of what is meant by a differentiable function on a regular surface.

DEFINITION 1. *Let* $\mathrm{f}: \mathrm{V} \subset \mathrm{S} \to \mathrm{R}$ *be a function defined in an open subset* V *of a regular surface* S. *Then* f *is said to be* differentiable at $\mathrm{p} \in \mathrm{V}$ *if, for some parametrization* $\mathbf{x}: \mathrm{U} \subset \mathrm{R}^2 \to \mathrm{S}$ *with* $\mathrm{p} \in \mathbf{x}(\mathrm{U}) \subset \mathrm{V}$, *the composition* $\mathrm{f} \circ \mathbf{x}: \mathrm{U} \subset \mathrm{R}^2 \to \mathrm{R}$ *is differentiable at* $\mathbf{x}^{-1}(\mathrm{p})$. f *is differentiable in* V *if it is differentiable at all points of* V.

It follows immediately from the last proposition that the definition given does not depend on the choice of the parametrization $\mathbf{x}$. In fact, if $\mathbf{y}: V \subset R^2 \to S$ is another parametrization with $p \in \mathbf{y}(V)$, and if $h = \mathbf{x}^{-1} \circ \mathbf{y}$, then $f \circ \mathbf{y} = f \circ \mathbf{x} \circ h$ is also differentiable, whence the asserted independence.

Remark 1. We shall frequently make the notational abuse of indicating f and $f \circ \mathbf{x}$ by the same symbol $f(u, v)$, and say that $f(u, v)$ is *the expression* of f in the system of coordinates $\mathbf{x}$. This is equivalent to identifying $\mathbf{x}(U)$ with U and thinking of (u, v), indifferently, as a point of U and as a point of $\mathbf{x}(U)$ with coordinates (u, v). From now on, abuses of language of this type will be used without further comment.

Example 1. Let S be a regular surface and $V \subset R^3$ be an open set such that $S \subset V$. Let $f: V \subset R^3 \to R$ be a differentiable function. Then the restriction of f to S is a differentiable function on S. In fact, for any $p \in S$ and any parametrization $\mathbf{x}: U \subset R^2 \to S$ in p, the function $f \circ \mathbf{x}: U \to R$ is differentiable. In particular, the following are differentiable functions:

1. The *height function* relative to a unit vector $v \in R^3$, $h: S \to R$, given by $h(p) = p \cdot v$, $p \in S$, where the dot denotes the usual inner product in R^3. $h(p)$ is the height of $p \in S$ relative to a plane normal to v and passing through the origin of R^3 (Fig. 2-15).

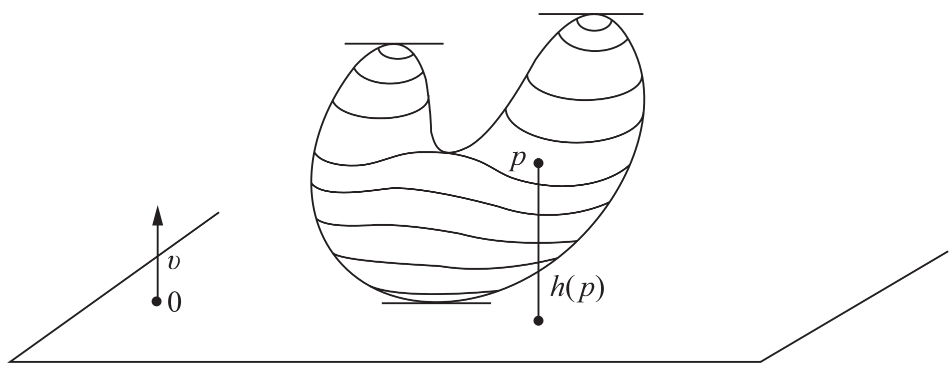


Figure 2-15

2. The square of the distance from a fixed point $p_0 \in R^3$, $f(p) = |p - p_0|^2$, $p \in S$. The need for taking the square comes from the fact that the distance $|p - p_0|$ is not differentiable at $p = p_0$.

Remark 2. The proof of Prop. 1 makes essential use of the fact that the inverse of a parametrization is continuous. Since we need Prop. 1 to be able to define differentiable functions on surfaces (a vital concept), we cannot dispose of this condition in the definition of a regular surface (cf. Remark 1 of Sec. 2-2).

The definition of differentiability can be easily extended to mappings between surfaces. A continuous map $\varphi: V_1 \subset S_1 \to S_2$ of an open set V_1 of a regular surface S_1 to a regular surface S_2 is said to be *differentiable at* $p \in V_1$ if, given parametrizations

$$\mathbf{x}_1: U_1 \subset R^2 \to S_1 \quad \mathbf{x}_2: U_2 \subset R^2 \to S_2,$$

with $p \in \mathbf{x}_1(U)$ and $\varphi(\mathbf{x}_1(U_1)) \subset \mathbf{x}_2(U_2)$, the map

$$\mathbf{x}_2^{-1} \circ \varphi \circ \mathbf{x}_1: U_1 \to U_2$$

is differentiable at $q = \mathbf{x}_1^{-1}(p)$ (Fig. 2-16).

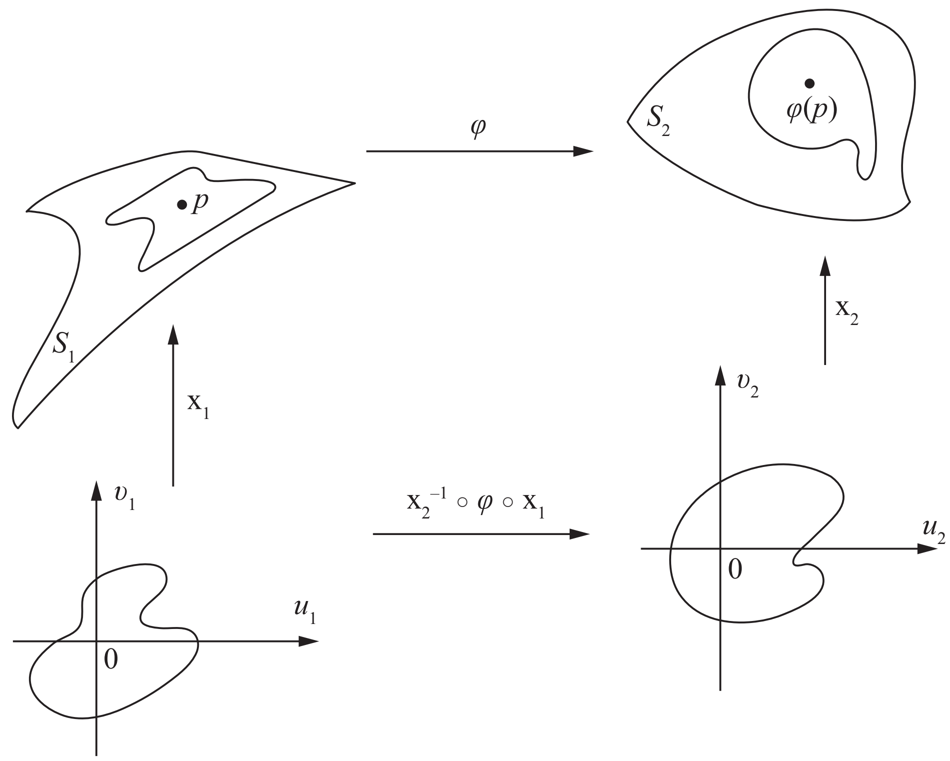


Figure 2-16

In other words, φ is differentiable if when expressed in local coordinates as $\varphi(u_1, v_1) = (\varphi_1(u_1, v_1), \varphi_2(u_1, v_1))$ the functions φ_1 and φ_2 have continuous partial derivatives of all orders.

The proof that this definition does not depend on the choice of parametrizations is left as an exercise.

We should mention that the natural notion of equivalence associated with differentiability is the notion of diffeomorphism. Two regular surfaces S_1 and S_2 are *diffeomorphic* if there exists a differentiable map φ: $S_1 \to S_2$ with a differentiable inverse φ^{-1}: $S_2 \to S_1$. Such a φ is called a *diffeomorphism* from S_1 to S_2. The notion of diffeomorphism plays the same role in the study of regular surfaces that the notion of isomorphism plays in the study of vector spaces or the notion of congruence plays in Euclidean geometry. In other words, from the point of view of differentiability, two diffeomorphic surfaces are indistinguishable.

Example 2. If $\mathbf{x}$: $U \subset R^2 \to S$ is a parametrization, $\mathbf{x}^{-1}$: $\mathbf{x}(U) \to R^2$ is differentiable. In fact, for any $p \in \mathbf{x}(U)$ and any parametrization $\mathbf{y}$: $V \subset R^2 \to S$ in p, we have that $\mathbf{x}^{-1} \circ \mathbf{y}$: $\mathbf{y}^{-1}(W) \to \mathbf{x}^{-1}(W)$, where

$$W = \mathbf{x}(U) \cap \mathbf{y}(V),$$

is differentiable. This shows that U and $\mathbf{x}(U)$ are diffeomorphic (i.e., every regular surface is locally diffeomorphic to a plane) and justifies the identification made in Remark 1.

Example 3. Let S_1 and S_2 be regular surfaces. Assume that $S_1 \subset V \subset R^3$, where V is an open set of R^3, and that $\varphi\colon V \to R^3$ is a differentiable map such that $\varphi(S_1) \subset S_2$. Then the restriction $\varphi|S_1\colon S_1 \to S_2$ is a differentiable map. In fact, given $p \in S_1$ and parametrizations $\mathbf{x}_1\colon U_1 \to S_1$, $\mathbf{x}_2\colon U_2 \to S_2$, with $p \in \mathbf{x}_1(U_1)$ and $\varphi(\mathbf{x}_1(U_1)) \subset \mathbf{x}_2(U_2)$, we have that the map

$$\mathbf{x}_2^{-1} \circ \varphi \circ \mathbf{x}_1\colon U_1 \to U_2$$

is differentiable. The following are particular cases of this general example:

1. Let S be symmetric relative to the xy plane; that is, if $(x, y, z) \in S$, then also $(x, y, -z) \in S$. Then the map $\sigma\colon S \to S$, which takes $p \in S$ into its symmetrical point, is differentiable, since it is the restriction to S of $\sigma\colon R^3 \to R^3$, $\sigma(x, y, z) = (x, y, -z)$. This, of course, generalizes to surfaces symmetric relative to any plane of R^3.
2. Let $R_{z,\theta}\colon R^3 \to R^3$ be the rotation of angle θ about the z axis, and let $S \subset R^3$ be a regular surface invariant by this rotation; i.e., if $p \in S$, $R_{z,\theta}(p) \in S$. Then the restriction $R_{z,\theta}\colon S \to S$ is a differentiable map.
3. Let $\varphi\colon R^3 \to R^3$ be given by $\varphi(x, y, z) = (xa, yb, zc)$, where a, b, and c are nonzero real numbers. φ is clearly differentiable, and the restriction $\varphi|S^2$ is a differentiable map from the sphere

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

into the ellipsoid

$$\left\{(x, y, z) \in R^3; \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1\right\}$$

(cf. Example 6 of the appendix to Chap. 2).

Remark 3. Proposition 1 implies (cf. Example 2) that a parametrization $\mathbf{x}\colon U \subset R^2 \to S$ is a diffeomorphism of U onto $\mathbf{x}(U)$. Actually, we can now characterize the regular surfaces as those subsets $S \subset R^3$ which are locally diffeomorphic to R^2; that is, for each point $p \in S$, there exists a neighborhood V of p in S, an open set $U \subset R^2$, and a map $\mathbf{x}\colon U \to V$, which is a diffeomorphism. This pretty characterization could be taken as the starting point of a treatment of surfaces (see Exercise 13).

At this stage we could return to the theory of curves and treat them from the point of view of this chapter, i.e., as subsets of R^3. We shall mention only certain fundamental points and leave the details to the reader.

The symbol I will denote an open interval of the line R. A *regular curve* in R^3 is a subset $C \subset R^3$ with the following property: For each point $p \in C$ there is a neighborhood V of p in R^3 and a differentiable homeomorphism

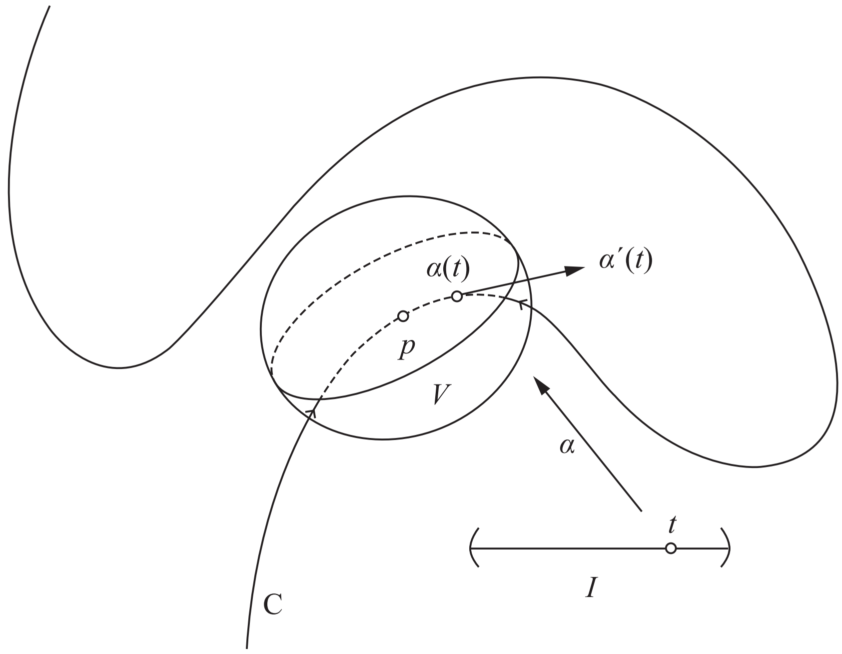


Figure 2-17. A regular curve.

$\alpha: I \subset R \to V \cap C$ such that the differential $d\alpha_t$ is one-to-one for each $t \in I$ (Fig. 2-17).

It is possible to prove (Exercise 15) that the change of parameters is given (as with surfaces) by a diffeomorphism. From this fundamental result, it is possible to decide when a given property obtained by means of a parametrization is independent of that parametrization. Such a property will then be a local property of the set C.

For example, it is shown that the arc length, defined in Chap. 1, is independent of the parametrization chosen (Exercise 15) and is, therefore, a property of the set C. Since it is always possible to locally parametrize a regular curve C by arc length, the properties (curvature, torsion, etc.) determined by means of this parametrizaion are local properties of C. This shows that the local theory of curves developed in Chap. 1 is valid for regular curves.

Sometimes a surface is defined by displacing a certain regular curve in a specified way. This occurs in the following example.

Example 4 (*Surfaces of Revolution*)**.** Let $S \subset R^3$ be the set obtained by rotating a regular connected plane curve C about an axis in the plane which does not meet the curve; we shall take the xz plane as the plane of the curve and the z axis as the rotation axis. Let

$$x = f(v), \quad z = g(v), \quad a < v < b, \quad f(v) > 0,$$

be a parametrization for C and denote by u the rotation angle about the z axis. Thus, we obtain a map

$$\mathbf{x}(u, v) = (f(v)\cos u,\ f(v)\sin u,\ g(v))$$

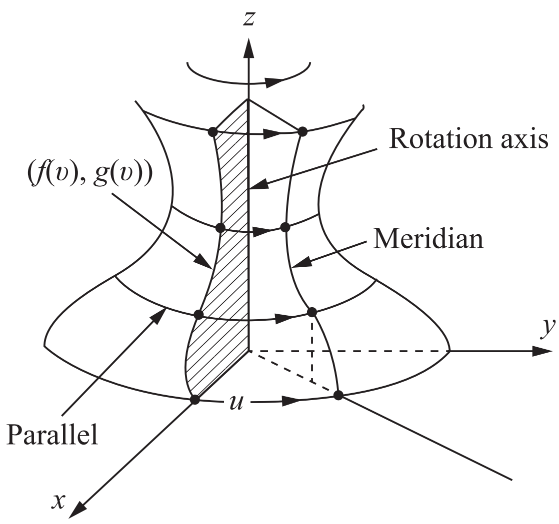


Figure 2-18. A surface of revolution.

from the open set $U = \{(u, v) \in R^2; 0 < u < 2\pi, a < v < b\}$ into S (Fig. 2-18).

We shall soon see that $\mathbf{x}$ satisfies the conditions for a parametrization in the definition of a regular surface. Since S can be entirely covered by similar parametrizations, it follows that S is a regular surface which is called a *surface of revolution*. The curve C is called the *generating curve* of S, and the z axis is the *rotation axis* of S. The circles described by the points of C are called the *parallels* of S, and the various positions of C on S are called the *meridians* of S.

To show that $\mathbf{x}$ is a parametrization of S we must check conditions 1, 2, and 3 of Def. 1, Sec. 2-2. Conditions 1 and 3 are straightforward, and we leave them to the reader. To show that $\mathbf{x}$ is a homeomorphism, we first show that $\mathbf{x}$ is one-to-one. In fact, since $(f(v), g(v))$ is a parametrization of C, given z and $x^2 + y^2 = (f(v))^2$, we can determine v uniquely. Thus, $\mathbf{x}$ is one-to-one.

We remark that, again because $(f(v), g(v))$ is a parametrization of C, v is a continuous function of z and of $\sqrt{x^2 + y^2}$ and thus a continuous function of (x, y, z).

To prove that $\mathbf{x}^{-1}$ is continuous, it remains to show that u is a continuous function of (x, y, z). To see this, we first observe that if $u \neq \pi$, we obtain, since $f(v) \neq 0$,

$$\tan\frac{u}{2} = \frac{\sin\dfrac{u}{2}}{\cos\dfrac{u}{2}} = \frac{2\sin\dfrac{u}{2}\cos\dfrac{u}{2}}{2\cos^2\dfrac{u}{2}} = \frac{\sin u}{1 + \cos u}$$

$$= \frac{\dfrac{y}{f(v)}}{1 + \dfrac{x}{f(v)}} = \frac{y}{x + \sqrt{x^2 + y^2}};$$

hence,

$$u = 2\tan^{-1}\frac{y}{x+\sqrt{x^2+y^2}}.$$

Thus, if $u \neq \pi$, u is a continuous function of (x, y, z). By the same token, if u is in a small interval about π, we obtain

$$u = 2\,\mathrm{cotan}^{-1}\frac{y}{-x+\sqrt{x^2+y^2}}.$$

Thus, u is a continuous function of (x, y, z). This shows that $\mathbf{x}^{-1}$ is continuous and completes the verification.

Remark 4. There is a slight problem with our definition of surface of revolution. If $C \subset R^2$ is a closed regular plane curve which is symmetric relative to an axis r of R^3, then, by rotating C about r, we obtain a surface which can be proved to be regular and should also be called a surface of revolution (when C is a circle and r contains a diameter of C, the surface is a sphere). To fit it in our definition, we would have to exclude two of its points, namely, the points where r meets C. For technical reasons, we want to maintain the previous terminology and shall call the latter surfaces *extended surfaces of revolution.*

A final comment should now be made on our definition of surface. We have chosen to define a (regular) surface as a subset of R^3. If we want to consider global, as well as local, properties of surfaces, this is the correct setting. The reader might have wondered, however, why we have not defined surface simply as a parametrized surface, as in the case of curves. This can be done, and in fact a certain amount of the classical literature in differential geometry was presented that way. No serious harm is done as long as only local properties are considered. However, basic global concepts, like orientation (to be treated in Secs. 2-6 and 3-1), have to be omitted, or treated inadequately, with such an approach.

In any case, the notion of parametrized surface is sometimes useful and should be included here.

DEFINITION 2. *A* parametrized surface $\mathbf{x}\colon U \subset R^2 \to R^3$ *is a differentiable map* $\mathbf{x}$ *from an open set* $U \subset R^2$ *into* R^3. *The set* $\mathbf{x}(U) \subset R^3$ *is called the* trace *of* $\mathbf{x}$. $\mathbf{x}$ *is* regular *if the differential* $d\mathbf{x}_q\colon R^2 \to R^3$ *is one-to-one for all* $q \in U$ (*i.e., the vectors* $\partial\mathbf{x}/\partial u$, $\partial\mathbf{x}/\partial v$ *are linearly independent for all* $q \in U$). *A point* $p \in U$ *where* $d\mathbf{x}_p$ *is not one-to-one is called a* singular point *of* $\mathbf{x}$.

Observe that a parametrized surface, even when regular, may have self-intersections in its trace.

Example 5. Let α: $I \to R^3$ be a nonplanar regular parametrized curve. Define

$$\mathbf{x}(t, v) = \alpha(t) + v\alpha'(t), \quad (t, v) \in I \times R.$$

$\mathbf{x}$ is a parametrized surface called the *tangent surface* of α (Fig. 2-19).

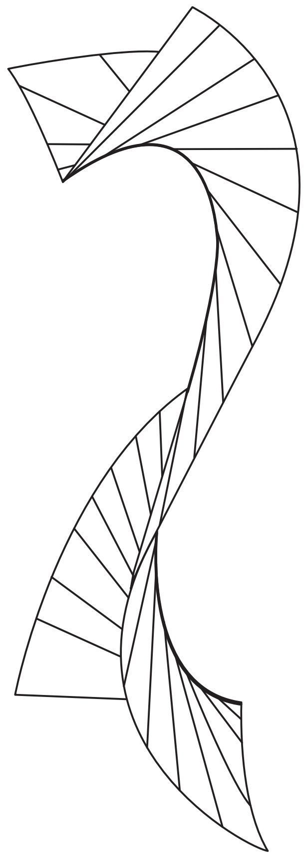


Figure 2-19. The tangent surface.

Assume now that the curvature $k(t)$, $t \in I$, of α is nonzero for all $t \in I$, and restrict the domain of $\mathbf{x}$ to $U = \{(t, v) \in I \times R; v \neq 0\}$. Then

$$\frac{\partial \mathbf{x}}{\partial t} = \alpha'(t) + v\alpha''(t), \quad \frac{\partial \mathbf{x}}{\partial v} = \alpha'(t)$$

and

$$\frac{\partial \mathbf{x}}{\partial t} \wedge \frac{\partial \mathbf{x}}{\partial v} = v\alpha''(t) \wedge \alpha'(t) \neq 0, \quad (t, v) \in U,$$

since, for all t, the curvature (cf. Exercise 12 of Sec. 1-5)

$$k(t) = \frac{|\alpha''(t) \wedge \alpha'(t)|}{|\alpha'(t)|^3}$$

is nonzero. It follows that the restriction $\mathbf{x}$: $U \to R^3$ is a regular parametrized surface, the trace of which consists of two connected pieces whose common boundary is the set $\alpha(I)$.

The following proposition shows that we can extend the local concepts and properties of differential geometry to regular parametrized surfaces.

PROPOSITION 2. *Let* $\mathbf{x}: \mathrm{U} \subset \mathrm{R}^2 \to \mathrm{R}^3$ *be a regular parametrized surface and let* $\mathrm{q} \in \mathrm{U}$. *Then there exists a neighborhood* V *of* q *in* R^2 *such that* $\mathbf{x}(\mathrm{V}) \subset \mathrm{R}^3$ *is a regular surface.*

Proof. This is again a consequence of the inverse function theorem. Write

$$\mathbf{x}(u, v) = (x(u, v), y(u, v), z(u, v)).$$

By regularity, we can assume that $(\partial(x, y)/\partial(u, v))(q) \neq 0$. Define a map $F: U \times R \to R^3$ by

$$F(u, v, t) = (x(u, v), y(u, v), z(u, v) + t), \quad (u, v) \in U, t \in R.$$

Then

$$\det(dF_q) = \frac{\partial(x, y)}{\partial(u, v)}(q) \neq 0.$$

By the inverse function theorem, there exist neighborhoods W_1 of q and W_2 of $F(q)$ such that $F: W_1 \to W_2$ is a diffeomorphism. Set $V = W_1 \cap U$ and observe that the restriction $F|V = \mathbf{x}|V$. Thus, $\mathbf{x}(V)$ is diffeomorphic to V, and hence a regular surface. **Q.E.D.**

EXERCISES†

***1.** Let $S^2 = \{(x, y, z) \in R^3; x^2 + y^2 + z^2 = 1\}$ be the unit sphere and let $A: S^2 \to S^2$ be the (*antipodal*) map $A(x, y, z) = (-x, -y, -z)$. Prove that A is a diffeomorphism.

2. Let $S \subset R^3$ be a regular surface and $\pi: S \to R^2$ be the map which takes each $p \in S$ into its orthogonal projection over $R^2 = \{(x, y, z) \in R^3; z = 0\}$. Is π differentiable?

3. Show that the paraboloid $z = x^2 + y^2$ is diffeomorphic to a plane.

4. Construct a diffeomorphism between the ellipsoid

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

and the sphere $x^2 + y^2 + z^2 = 1$.

***5.** Let $S \subset R^3$ be a regular surface, and let $d: S \to R$ be given by $d(p) = |p - p_0|$, where $p \in S$, $p_0 \neq R^3$, $p_0 \notin S$; that is, d is the distance from p to a fixed point p_0 not in S. Prove that d is differentiable.

†Those who have omitted the proofs of this section should also omit Exercises 13–16.

6. Prove that the definition of a differentiable map between surfaces does not depend on the parametrizations chosen.

7. Prove that the relation "S_1 is diffeomorphic to S_2" is an equivalence relation in the set of regular surfaces.

***8.** Let $S^2 = \{(x, y, z) \in R^3; x^2 + y^2 + z^2 = 1\}$ and $H = \{(x, y, z) \in R^3; x^2 + \mathbf{y}^2 - z^2 = 1\}$. Denote by $N = (0, 0, 1)$ and $S = (0, 0, -1)$ the north and south poles of S^2, respectively, and let $F: S^2 - \{N\} \cup \{S\} \to H$ be defined as follows: For each $p \in S^2 - \{N\} \cup \{S\}$ let the perpendicular from p to the z axis meet $0z$ at q. Consider the half-line l starting at q and containing p. Then $F(p) = l \cap H$ (Fig. 2-20). Prove that F is differentiable.

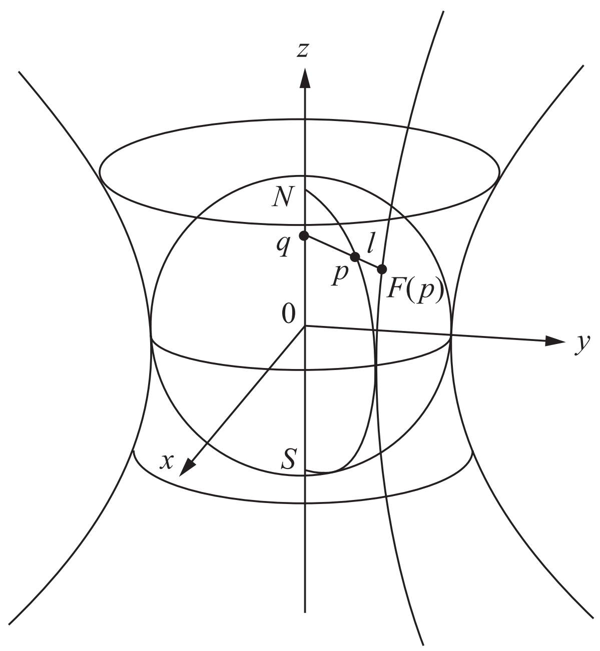


Figure 2-20

9. a. Define the notion of differentiable function on a regular curve. What does one need to prove for the definition to make sense? Do not prove it now. If you have not omitted the proofs in this section, you will be asked to do it in Exercise 15.

b. Show that the map $E: R \to S^1 = \{(x, y) \in R^2; x^2 + y^2 = 1\}$ given by

$$E(t) = (\cos t, \sin t), \qquad t \in R,$$

is differentiable (geometrically, E "wraps" R around S^1).

10. Let C be a plane regular curve which lies in one side of a straight line r of the plane and meets r at the points p, q (Fig. 2-21). What conditions should C satisfy to ensure that the rotation of C about r generates an extended (regular) surface of revolution?

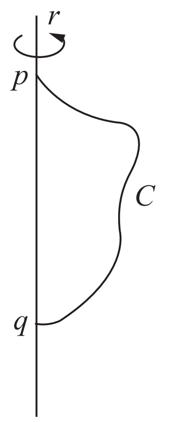


Figure 2-21

11. Prove that the rotations of a surface of revolution S about its axis are diffeomorphisms of S.

12. Parametrized surfaces are often useful to describe sets Σ which are regular surfaces except for a finite number of points and a finite number of lines. For instance, let C be the trace of a regular parametrized curve $\alpha: (a, b) \to R^3$ which does not pass through the origin $O = (0, 0, 0)$. Let Σ be the set generated by the displacement of a straight line l passing through a moving point $p \in C$ and the fixed point 0 (a cone with vertex 0; see Fig. 2-22).

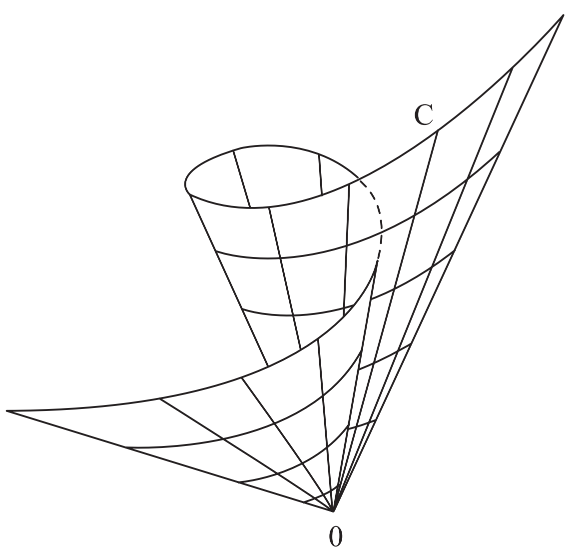


Figure 2-22

a. Find a parametrized surface **x** whose trace is Σ.

b. Find the points where **x** is not regular.

c. What should be removed from Σ so that the remaining set is a regular surface?

***13.** Show that the definition of differentiability of a function $f: V \subset S \to R$ given in the text (Def. 1) is equivalent to the following: f is differentiable in $p \in V$ if it is the restriction to V of a differentiable function defined in an open set of R^3 containing p. (Had we started with this definition of differentiability, we could have defined a surface as a set which is locally diffeomorphic to R^2; see Remark 3.)

14. Let $A \subset S$ be a subset of a regular surface S. Prove that A is itself a regular surface if and only if A is open in S; that is, $A = U \cap S$, where U is an open set in R^3.

15. Let C be a regular curve and let $\alpha: I \subset R \to C$, $\beta: J \subset R \to C$ be two parametrizations of C in a neighborhood of $p \in \alpha(I) \cap \beta(J) = W$. Let

$$h = \alpha^{-1} \circ \beta: \beta^{-l}(W) \to \alpha^{-l}(W)$$

be the change of parameters. Prove that

a. h is a diffeomorphism.

b. The absolute value of the arc length of C in W does not depend on which parametrization is chosen to define it, that is,

$$\left| \int_{t_0}^{t} |\alpha'(t)|\, dt \right| = \left| \int_{\tau_0}^{\tau} |\beta'(\tau)|\, d\tau \right|, \quad t = h(\tau), t \in I, \tau \in J.$$

***16.** Let $R^2 = \{(x, y, z) \in R^3; z = -1\}$ be identified with the complex plane $\mathbb{C}$ by setting $(x, y, -1) = x + iy = \zeta \in \mathbb{C}$. Let $P: \mathbb{C} \to \mathbb{C}$ be the complex polynomial

$$P(\zeta) = a_0\zeta^n + a_1\zeta^{n-1} + \cdots + a_n, \quad a_0 \neq 0, a_i \in \mathbb{C}, i = 0, \ldots, n.$$

Denote by π_N the stereographic projection of $S^2 = \{(x, y, z) \in R^3; x^2 + y^2 + z^2 = 1\}$ from the north pole $N = (0, 0, 1)$ onto R^2. Prove that the map $F: S^2 \to S^2$ given by

$$\begin{aligned} F(p) &= \pi_N^{-1} \circ P \circ \pi_N(p), \quad \text{if } p \in S^2 - \{N\}, \\ F(N) &= N \end{aligned}$$

is differentiable.

2-4. The Tangent Plane; The Differential of a Map

In this section we shall show that condition 3 in the definition of a regular surface S guarantees that for every $p \in S$ the set of tangent vectors to the parametrized curves of S, passing through p, constitutes a plane.

By a *tangent vector* to S, at a point $p \in S$, we mean the tangent vector $\alpha'(0)$ of a differentiable parametrized curve $\alpha: (-\epsilon, \epsilon) \to S$ with $\alpha(0) = p$.

PROPOSITION 1. *Let* $\mathbf{x}: U \subset R^2 \to S$ *be a parametrization of a regular surface* S *and let* $q \in U$. *The vector subspace of dimension* 2,

$$d\mathbf{x}_q(R^2) \subset R^3,$$

coincides with the set of tangent vectors to S *at* $\mathbf{x}(q)$.

Proof. Let w be a tangent vector at $\mathbf{x}(q)$, that is, let $w = \alpha'(0)$, where $\alpha: (-\epsilon, \epsilon) \to \mathbf{x}(U) \subset S$ is differentiable and $\alpha(0) = \mathbf{x}(q)$. By Example 2 of Sec. 2-3, the curve $\beta = \mathbf{x}^{-1} \circ \alpha: (-\epsilon, \epsilon) \to U$ is differentiable. By definition of the differential (appendix to Chap. 2, Def. 1), we have $d\mathbf{x}_q(\beta'(0)) = w$. Hence, $w \in d\mathbf{x}_q(R^2)$ (Fig. 2-23).

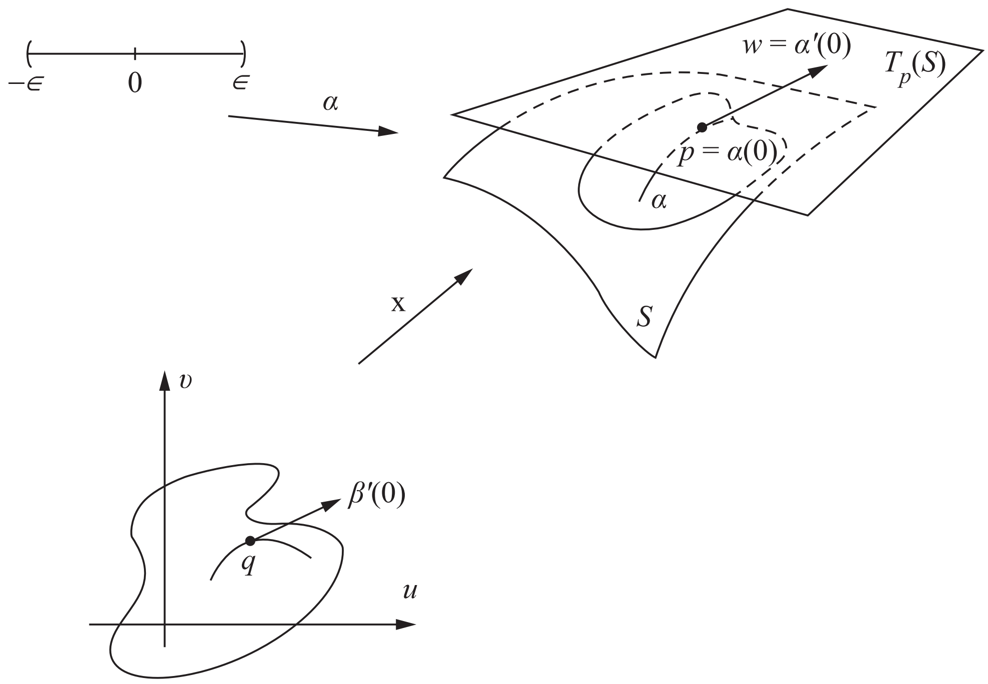


Figure 2-23

On the other hand, let $w = d\mathbf{x}_q(v)$, where $v \in R^2$. It is clear that v is the velocity vector of the curve $\gamma: (-\epsilon, \epsilon) \to U$ given by

$$\gamma(t) = tv + q, \quad t \in (-\epsilon, \epsilon).$$

By the definition of the differential, $w = \alpha'(0)$, where $\alpha = \mathbf{x} \circ \gamma$. This shows that w is a tangent vector. **Q.E.D.**

By the above proposition, the plane $d\mathbf{x}_q(R^2)$, which passes through $\mathbf{x}(q) = p$, does not depend on the parametrization $\mathbf{x}$. This plane will be called the *tangent plane* to S at p and will be denoted by $T_p(S)$. The choice of the parametrization $\mathbf{x}$ determines a basis $\{(\partial\mathbf{x}/\partial u)(q), (\partial\mathbf{x}/\partial v)(q)\}$ of $T_p(S)$, called the basis associated to $\mathbf{x}$. Sometimes it is convenient to write $\partial\mathbf{x}/\partial u = \mathbf{x}_u$ and $\partial\mathbf{x}/\partial v = \mathbf{x}_v$.

The coordinates of a vector $w \in T_p(S)$ in the basis associated to a parametrization $\mathbf{x}$ are determined as follows. w is the velocity vector $\alpha'(0)$ of a curve $\alpha = \mathbf{x} \circ \beta$, where β: $(-\epsilon, \epsilon) \to U$ is given by $\beta(t) = (u(t), v(t))$, with $\beta(0) = q = \mathbf{x}^{-1}(p)$. Thus,

$$\begin{aligned}\alpha'(0) &= \frac{d}{dt}(\mathbf{x} \circ \beta)(0) = \frac{d}{dt}\mathbf{x}(u(t), v(t))(0)\\ &= \mathbf{x}_u(q)u'(0) + \mathbf{x}_v(q)v'(0) = w.\end{aligned}$$

Thus, in the basis $\{\mathbf{x}_u(q), \mathbf{x}_v(q)\}$, w has coordinates $(u'(0), v'(0))$, where $(u(t), v(t))$ is the expression, in the parametrization $\mathbf{x}$, of a curve whose velocity vector at $t = 0$ is w.

With the notion of a tangent plane, we can talk about the differential of a (differentiable) map between surfaces. Let S_1 and S_2 be two regular surfaces and let φ: $V \subset S_1 \to S_2$ be a differentiable mapping of an open set V of S_1 into S_2. If $p \in V$, we know that every tangent vector $w \in T_p(S_1)$ is the velocity vector $\alpha'(0)$ of a differentiable parametrized curve α: $(-\epsilon, \epsilon) \to V$ with $\alpha(0) = p$. The curve $\beta = \varphi \circ \alpha$ is such that $\beta(0) = \varphi(p)$, and therefore $\beta'(0)$ is a vector of $T_{\varphi(p)}(S_2)$ (Fig. 2-24).

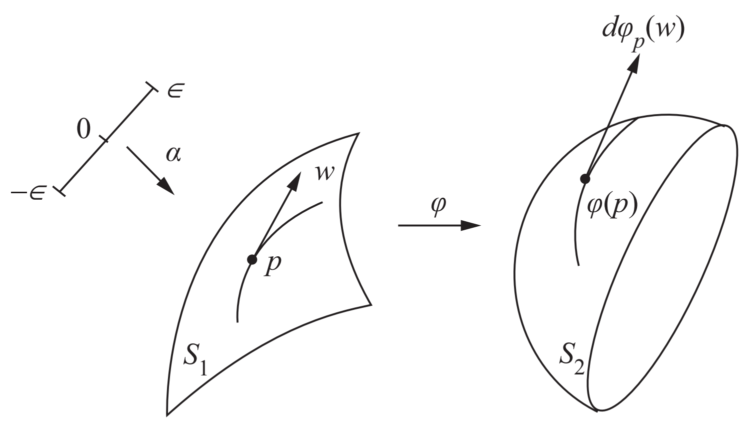


Figure 2-24

PROPOSITION 2. *In the discussion above, given* w, *the vector* $\beta'(0)$ *does not depend on the choice of* α. *The map* $\mathrm{d}\varphi_\mathrm{p}$: $\mathrm{T_p(S_1)} \to \mathrm{T_{\varphi(p)}(S_2)}$ *defined by* $\mathrm{d}\varphi_\mathrm{p}(\mathrm{w}) = \beta'(0)$ *is linear.*

Proof. The proof is similar to the one given in Euclidean spaces (cf. Prop. 7, appendix to Chap. 2). Let $\mathbf{x}(u, v)$, $\bar{\mathbf{x}}(\bar{u}, \bar{v})$ be parametrizations in neighborhoods of p and $\varphi(p)$, respectively. Suppose that φ is expressed in these coordinates by

$$\varphi(u, v) = (\varphi_1(u, v), \varphi_2(u, v))$$

and that α is expressed by

$$\alpha(t) = (u(t), v(t)), \quad t \in (-\epsilon, \epsilon).$$

Then $\beta(t) = (\varphi_1(u(t), v(t)), \varphi_2(u(t), v(t)))$, and the expression of $\beta'(0)$ in the basis $\{\bar{\mathbf{x}}_{\bar{u}}, \bar{\mathbf{x}}_{\bar{v}}\}$ is

$$\beta'(0) = \left(\frac{\partial\varphi_1}{\partial u}u'(0) + \frac{\partial\varphi_1}{\partial v}v'(0), \frac{\partial\varphi_2}{\partial u}u'(0) + \frac{\partial\varphi_2}{\partial v}v'(0)\right).$$

The relation above shows that $\beta'(0)$ depends only on the map φ and the coordinates $(u'(0), v'(0))$ of w in the basis $\{\mathbf{x}_u, x_v\}$. $\beta'(0)$ is therefore independent of α. Moreover, the same relation shows that

$$\beta'(0) = d\varphi_p(w) = \begin{pmatrix} \dfrac{\partial\varphi_1}{\partial u} & \dfrac{\partial\varphi_1}{\partial v} \\ \dfrac{\partial\varphi_2}{\partial u} & \dfrac{\partial\varphi_2}{\partial v} \end{pmatrix} \begin{pmatrix} u'(0) \\ \\ v'(0) \end{pmatrix};$$

that is, $d\varphi_p$ is a linear mapping of $T_p(S_1)$ into $T_{\varphi(p)}(S_2)$ whose matrix in the bases $\{\mathbf{x}_u, \mathbf{x}_v\}$ of $T_p(S_1)$ and $\{\bar{\mathbf{x}}_{\bar{u}}, \bar{\mathbf{x}}_{\bar{v}}\}$ of $T_{\varphi(p)}(S_2)$ is just the matrix given above. **Q.E.D.**

The linear map $d\varphi_p$ defined by Prop. 2 is called the *differential* of φ at $p \in S_1$. In a similar way we define the differential of a (differentiable) function $f: U \subset S \to R$ at $p \in U$ as a linear map $df_p: T_p(S) \to R$. We leave the details to the reader.

Example 1. Let $v \in R^3$ be a unit vector and let $h: S \to R$, $h(p) = v \cdot p$, $p \in S$, be the height function defined in Example 1 of Sec. 2-3. To compute $dh_p(w)$, $w \in T_p(S)$, choose a differentiable curve $\alpha: (-\epsilon, \epsilon) \to S$ with $\alpha(0) = p$, $\alpha'(0) = w$. Since $h(\alpha(t)) = \alpha(t) \cdot v$, we obtain

$$dh_p(w) = \frac{d}{dt}h(\alpha(t))|_{t=0} = \alpha'(0) \cdot v = w \cdot v.$$

Example 2. Let $S^2 \subset R^3$ be the unit sphere

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

and let $R_{z,\theta}: R^3 \to R^3$ be the rotation of angle θ about the z axis. Then $R_{z,\theta}$ restricted to S^2 is a differentiable map of S^2 (cf. Example 3 of Sec. 2-3). We shall compute $(dR_{z,\theta})_p(w)$, $p \in S^2$, $w \in T_p(S^2)$. Let $\alpha: (-\epsilon, \epsilon) \to S^2$ be a differentiable curve with $\alpha(0) = p$, $\alpha'(0) = w$. Then, since $R_{z,\theta}$ is linear,

$$(dR_{z,\theta})_p(w) = \frac{d}{dt}(R_{z,\theta} \circ \alpha(t))_{t=0} = R_{z,\theta}(\alpha'(0)) = R_{z,\theta}(w).$$

Observe that $R_{z,\theta}$ leaves the north pole $N = (0, 0, 1)$ fixed, and that $(dR_{z,\theta})_N: T_N(S^2) \to T_N(S^2)$ is just a rotation of angle θ in the plane $T_N(S^2)$.

In retrospect, what we have been doing up to now is extending the notions of differential calculus in R^2 to regular surfaces. Since calculus is essentially a local theory, we defined an entity (the regular surface) which locally was a plane, up to diffeomorphisms, and this extension then became natural. It might be expected therefore that the basic inverse function theorem extends to differentiable mappings between surfaces.

We shall say that a mapping φ: $U \subset S_1 \to S_2$ is a *local diffeomorphism* at $p \in U$ if there exists a neighborhood $V \subset U$ of p such that φ restricted to V is a diffeomorphism onto an open set $\varphi(V) \subset S_2$. In these terms, the version of the inverse of function theorem for surfaces is expressed as follows.

PROPOSITION 3. *If* S_1 *and* S_2 *are regular surfaces and* φ: $\mathrm{U} \subset \mathrm{S}_1 \to \mathrm{S}_2$ *is a differentiable mapping of an open set* $\mathrm{U} \subset \mathrm{S}_1$ *such that the differential* $d\varphi_{\mathrm{p}}$ *of* φ *at* $\mathrm{p} \in \mathrm{U}$ *is an isomorphism, then* φ *is a local diffeomorphism at* p.

The proof is an immediate application of the inverse function theorem in R^2 and will be left as an exercise.

Of course, all other concepts of calculus, like critical points, regular values, etc., do extend naturally to functions and maps defined on regular surfaces.

The tangent plane also allows us to speak of the angle of two intersecting surfaces at a point of intersection.

Given a point p on a regular surface S, there are two unit vectors of R^3 that are normal to the tangent plane $T_p(S)$; each of them is called *a unit normal vector* at p. The straight line that passes through p and contains a unit normal vector at p is called the *normal line* at p. The *angle* of two intersecting surfaces at an intersection point p is the angle of their tangent planes (or their normal lines) at p (Fig. 2-25).

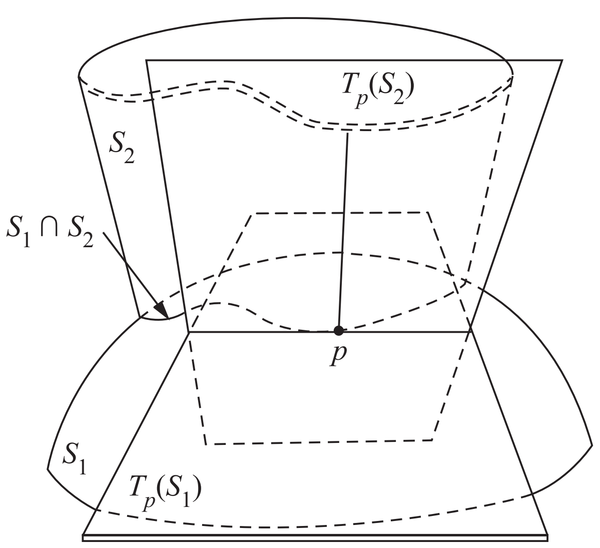


Figure 2-25

By fixing a parametrization $\mathbf{x}$: $U \subset R^2 \to S$ at $p \in S$, we can make a definite choice of a unit normal vector at each point $q \in \mathbf{x}(U)$ by the rule

$$N(q) = \frac{\mathbf{x}_u \wedge x_v}{|x_u \wedge x_v|}(q).$$

Thus, we obtain a differentiable map $N\colon \mathbf{x}(U) \to R^3$. We shall see later (Secs. 2-6 and 3-1) that it is not always possible to extend this map differentiably to the whole surface S.

Before leaving this section, we shall make some observations on questions of differentiability.

The definition given for a regular surface requires that the parametrization be of class C^∞, that is, that they possess continuous partial derivatives of *all* orders. For questions in differential geometry we need in general the existence and continuity of the partial derivatives only up to a certain order, which varies with the nature of the problem (very rarely do we need more than four derivatives).

For example, the existence and continuity of the tangent plane depends only on the existence and continuity of the first partial derivatives. It could happen, therefore, that the graph of a function $z = f(x, y)$ admits a tangent plane at every point but is not sufficiently differentiable to satisfy the definition of a regular surface. This occurs in the following example.

Example 3. Consider the graph of the function $z = \sqrt[3]{(x^2 + y^2)^2}$, generated by rotating the curve $z = x^{4/3}$ about the z axis. Since the curve is symmetric relative to the z axis and has a continuous derivative which vanishes at the origin, it is clear that the graph of $z = \sqrt[3]{(x^2 + y^2)^2}$ admits the xy plane as a tangent plane at the origin. However, the partial derivative z_{xx} does not exist at the origin, and the graph considered is not a regular surface as defined above (see Prop. 3 of Sec. 2-2).

We do not intend to get involved with this type of question. The hypothesis C^∞ in the definition was adopted precisely to avoid the study of the minimal conditions of differentiability required in each particular case. These nuances have their place, but they can eventually obscure the geometric nature of the problems treated here.

EXERCISES

***1.** Show that the equation of the tangent plane at (x_0, y_0, z_0) of a regular surface given by $f(x, y, z) = 0$, where 0 is a regular value of f, is

$$f_x(x_0, y_0, z_0)(x - x_0) + f_y(x_0, y_0, z_0)(y - y_0) + f_z(x_0, y_0, z_0)(z - z_0) = 0.$$

2. Determine the tangent planes of $x^2 + y^2 - z^2 = 1$ at the points $(x, y, 0)$ and show that they are all parallel to the z axis.

3. Show that the equation of the tangent plane of a surface which is the graph of a differentiable function $z = f(x, y)$, at the point $p_0 = (x_0, y_0)$, is given by

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Recall the definition of the differential df of a function $f\colon R^2 \to R$ and show that the tangent plane is the graph of the differential df_p.

***4.** Show that the tangent planes of a surface given by $z = xf(y/x)$, $x \neq 0$, where f is a differentiable function, all pass through the origin (0, 0, 0).

5. If a coordinate neighborhood of a regular surface can be parametrized in the form

$$\mathbf{x}(u, v) = \alpha_1(u) + \alpha_2(v),$$

where α_1 and α_2 are regular parametrized curves, show that the tangent planes along a fixed coordinate curve of this neighborhood are all parallel to a line.

6. Let $\alpha\colon I \to R^3$ be a regular parametrized curve with everywhere nonzero curvature. Consider the tangent surface of α (Example 5 of Sec. 2-3)

$$\mathbf{x}(t, v) = \alpha(t) + v\alpha'(t), \quad t \in I, v \neq 0.$$

Show that the tangent planes along the curve $\mathbf{x}$(const.,v) are all equal.

7. Let $f\colon S \to R$ be given by $f(p) = |p - p_0|^2$, where $p \in S$ and p_0 is a fixed point of R^3 (see Example 1 of Sec. 2-3). Show that $df_p(w) = 2w \cdot (p - p_0)$, $w \in T_p(S)$.

8. Prove that if $L\colon R^3 \to R^3$ is a linear map and $S \subset R^3$ is a regular surface invariant under L, i.e., $L(S) \subset S$, then the restriction $L|S$ is a differentiable map and

$$dL_p(w) = L(w), \quad p \in S, w \in T_p(S).$$

9. Show that the parametrized surface

$$x(u, v) = (v \cos u, v \sin u, au), \quad a \neq 0,$$

is regular. Compute its normal vector $N(u, v)$ and show that along the coordinate line $u = u_0$ the tangent plane of $\mathbf{x}$ rotates about this line in such a way that the tangent of its angle with the z axis is proportional to the inverse of the distance $v(= \sqrt{x^2 + y^2})$ of the point $\mathbf{x}(u_0, v)$ to the z axis.

10. (*Tubular Surfaces.*) Let $\alpha\colon I \to R^3$ be a regular parametrized curve with nonzero curvature everywhere and arc length as parameter. Let

$$\mathbf{x}(s, v) = \alpha(s) + r(n(s)\cos v + b(s)\sin v), \quad r = \text{const.} \neq 0, s \in I,$$

be a parametrized surface (the *tube* of radius r around α), where n is the normal vector and b is the binormal vector of α. Show that, when $\mathbf{x}$ is regular, its unit normal vector is

$$N(s, v) = -(n(s)\cos v + b(s)\sin v).$$

11. Show that the normals to a parametrized surface given by

$$\mathbf{x}(u, v) = (f(u)\cos v, f(u)\sin v, g(u)), \qquad f(u) \neq 0, g' \neq 0,$$

all pass through the z axis.

***12.** Show that each of the equations $(a, b, c \neq 0)$

$$x^2 + y^2 + z^2 = ax,$$
$$x^2 + y^2 + z^2 = by,$$
$$x^2 + y^2 + z^2 = cz$$

define a regular surface and that they all intersect orthogonally.

13. A *critical point* of a differentiable function $f\colon S \to R$ defined on a regular surface S is a point $p \in S$ such that $df_p = 0$.

***a.** Let $f\colon S \to R$ be given by $f(p) = |p - p_0|$, $p \in S$, $p_0 \notin S$ (cf. Exercise 5, Sec. 2-3). Show that $p \in S$ is a critical point of f if and only if the line joining p to p_0 is normal to S at p.

b. Let $h\colon S \to R$ be given by $h(p) = p \cdot v$, where $v \in R^3$ is a unit vector (cf. Example 1, Sec. 2-3). Show that $p \in S$ is a critical point of f if and only if v is a normal vector of S at p.

***14.** Let Q be the union of the three coordinate planes $x = 0$, $y = 0$, $z = 0$. Let $p = (x, y, z) \in R^3 - Q$.

a. Show that the equation in t,

$$\frac{x^2}{a-t} + \frac{y^2}{b-t} + \frac{z^2}{c-t} \equiv f(t) = 1, \qquad a > b > c > 0,$$

has three distinct real roots: t_1, t_2, t_3.

b. Show that for each $p \in R^3 - Q$, the sets given by $f(t_1) - 1 = 0$, $f(t_2) - 1 = 0$, $f(t_3) - 1 = 0$ are regular surfaces passing through p which are pairwise orthogonal.

15. Show that if all normals to a connected surface pass through a fixed point, the surface is contained in a sphere.

16. Let w be a tangent vector to a regular surface S at a point $p \in S$ and let $\mathbf{x}(u, v)$ and $\bar{\mathbf{x}}(\bar{u}, \bar{v})$ be two parametrizatioos at p. Suppose that the expressions of w in the bases associated to $\mathbf{x}(u, v)$ and $\bar{\mathbf{x}}(\bar{u}, \bar{v})$ are

$$w = \alpha_1 \mathbf{x}_u + \alpha_2 \mathbf{x}_v$$

and

$$w = \beta_1 \bar{\mathbf{x}}_{\bar{u}} + \beta_2 \mathbf{x}_{\bar{v}}.$$

Show that the coordinates of w are related by

$$\beta_1 = \alpha_1 \frac{\partial \bar{u}}{\partial u} + \alpha_2 \frac{\partial \bar{u}}{\partial v}$$
$$\beta_2 = \alpha_1 \frac{\partial \bar{v}}{\partial u} + \alpha_2 \frac{\partial \bar{v}}{\partial v},$$

where $\bar{u} = \bar{u}(u, v)$ and $\bar{v} = \bar{v}(u, v)$ are the expressions of the change of coordinates.

***17.** Two regular surfaces S_1 and S_2 intersect *transversally* if whenever $p \in S_1 \cap S_2$ then $T_p(S_1) \neq T_p(S_2)$. Prove that if S_1 intersects S_2 transversally, then $S_1 \cap S_2$ is a regular curve.

18. Prove that if a regular surface S meets a plane P in a single point p, then this plane coincides with the tangent plane of S at p.

19. Let $S \subset R^3$ be a regular surface and $P \subset R^3$ be a plane. If all points of S are on the same side of P, prove that P is tangent to S at all points of $P \cap S$.

***20.** Show that the perpendicular projections of the center $(0, 0, 0)$ of the ellipsoid

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

onto its tangent planes constitute a regular surface given by

$$\{(x, y, z) \in R^3;\ (x^2 + y^2 + z^2)^2 = a^2x^2 + b^2y^2 + c^2z^2\} - \{(0, 0, 0)\}.$$

***21.** Let $f\colon S \to R$ be a differentiable function on a connected regular surface S. Assume that $df_p = 0$ for all $p \in S$. Prove that f is constant on S.

***22.** Prove that if all normal lines to a connected regular surface S meet a fixed straight line, then S is a piece of a surface of revolution.

23. Prove that the map $F\colon S^2 \to S^2$ defined in Exercise 16 of Sec. 2-3 has only a finite number of critical points (see Exercise 13).

24. (*Chain Rule.*) Show that if $\varphi\colon S_1 \to S_2$ and $\psi\colon S_2 \to S_3$ are differentiable maps and $p \in S_1$, then

$$d(\psi \circ \varphi)_p = d\psi_{\varphi(p)} \circ d\varphi_p.$$

25. Prove that if two regular curves C_1 and C_2 of a regular surface S are tangent at a point $p \in S$, and if $\varphi\colon S \to S$ is a diffeomorphism, then $\varphi(C_1)$ and $\varphi(C_2)$ are regular curves which are tangent at $\varphi(p)$.

26. Show that if p is a point of a regular surface S, it is possible, by a convenient choice of the (x, y, z) coordinates, to represent a neighborhood of p in S in the form $z = f(x, y)$ so that $f(0, 0) = 0$, $f_x(0, 0) = 0$, $f_y(0, 0) = 0$. (This is equivalent to taking the tangent plane to S at p as the xy plane.)

27. (*Theory of Contact.*) Two regular surfaces, S and $\bar{S}$, in R^3, which have a point p in common, are said to have *contact of order* ≥ 1 at p if there exist parametrizations with the same domain $\mathbf{x}(u, v)$, $\bar{\mathbf{x}}(u, v)$ at p of S and $\bar{S}$, respectively, such that $\mathbf{x}_u = \bar{\mathbf{x}}_u$ and $\mathbf{x}_v = \bar{\mathbf{x}}_v$ at p. If, moreover, some of the second partial derivatives are different at p, the *contact* is said to be *of order exactly equal to* 1. Prove that

a. The tangent plane $T_p(S)$ of a regular surface S at the point p has contact of order ≥ 1 with the surface at p.

b. If a plane has contact of order ≥ 1 with a surface S at p, then this plane coincides with the tangent plane to S at p.

c. Two regular surfaces have contact of order ≥ 1 if and only if they have a common tangent plane at p, i.e., they are tangent at p.

d. If two regular surfaces S and $\bar{S}$ of R^3 have contact of order ≥ 1 at p and if $F\colon R^3 \to R^3$ is a diffeomorphism of R^3, then the images $F(S)$ and $F(\bar{S})$ are regular surfaces which have contact of order ≥ 1 at $f(p)$ (that is, the notion of contact of order ≥ 1 is invariant under diffeomorphisms).

e. If two surfaces have contact of order ≥ 1 at p, then $\lim_{r\to 0}(d/r) = 0$, where d is the length of the segment which is determined by the intersections with the surfaces of some parallel to the common normal, at a distance r from this normal.

28. a. Define regular value for a differentiable function $f\colon S \to R$ on a regular surface S.

b. Show that the inverse image of a regular value of a differentiable function on a regular surface S is a regular curve on S.

2-5. The First Fundamental Form; Area

So far we have looked at surfaces from the point of view of differentiability. In this section we shall begin the study of further geometric structures carried by the surface. The most important of these is perhaps the first fundamental form, which we shall now describe.

The natural inner product of $R^3 \supset S$ induces on each tangent plane $T_p(S)$ of a regular surface S an inner product, to be denoted by $\langle\ ,\ \rangle_p$: If $w_1, w_2 \in T_p(S) \subset R^3$, then $\langle w_1, w_2\rangle_p$ is equal to the inner product of w_1 and w_2 as vectors in R^3. To this inner product, which is a symmetric bilinear form (i.e., $\langle w_1, w_2\rangle = \langle w_2, w_1\rangle$ and $\langle w_1, w_2\rangle$ is linear in both w_1 and w_2), there corresponds a quadratic form $I_p\colon T_p(S) \to R$ given by

$$I_p(w) = \langle w, w\rangle_p = |w|^2 \geq 0. \tag{1}$$

DEFINITION 1. *The quadratic form* $\mathrm{I_p}$ *on* $\mathrm{T_p(S)}$, *defined by Eq. (1), is called the* first fundamental form *of the regular surface* $S \subset R^3$ *at* $\mathrm{p} \in \mathrm{S}$.

Therefore, the first fundamental form is merely the expression of how the surface S inherits the natural inner product of R^3. Geometrically, as we shall see in a while, the first fundamental form allows us to make measurements on the surface (lengths of curves, angles of tangent vectors, areas of regions) without referring back to the ambient space R^3 where the surface lies.

We shall now express the first fundamental form in the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ associated to a parametrization $\mathbf{x}(u, v)$ at p. Since a tangent vector $w \in T_p(S)$ is the tangent vector to a parametrized curve $\alpha(t) = \mathbf{x}(u(t), v(t))$, $t \in (-\epsilon, \epsilon)$, with $p = \alpha(0) = \mathbf{x}(u_0, v_0)$, we obtain

$$\begin{aligned}
I_p(\alpha'(0)) &= \langle \alpha'(0), \alpha'(0)\rangle_p \\
&= \langle \mathbf{x}_u u' + \mathbf{x}_v v', \mathbf{x}_u u' + \mathbf{x}_v v'\rangle_p \\
&= \langle \mathbf{x}_u, \mathbf{x}_u\rangle_p (u')^2 + 2\langle \mathbf{x}_u, \mathbf{x}_v\rangle_p u'v' + \langle \mathbf{x}_v, \mathbf{x}_v\rangle_p (v')^2 \\
&= E(u')^2 + 2Fu'v' + G(v')^2,
\end{aligned}$$

where the values of the functions involved are computed for $t = 0$, and

$$\begin{aligned}
E(u_0, v_0) &= \langle \mathbf{x}_u, \mathbf{x}_u\rangle_p, \\
F(u_0, v_0) &= \langle \mathbf{x}_u, \mathbf{x}_v\rangle_p, \\
G(u_0, v_0) &= \langle \mathbf{x}_v, \mathbf{x}_v\rangle_p
\end{aligned}$$

are the coefficients of the first fundamental form in the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ of $T_p(S)$. By letting p run in the coordinate neighborhood corresponding to $\mathbf{x}(u, v)$ we obtain functions $E(u, v)$, $F(u, v)$, $G(u, v)$ which are differentiable in that neighborhood.

From now on we shall drop the subscript p in the indication of the inner product $\langle\ ,\ \rangle_p$ or the quadratic form I_p when it is clear from the context which point we are referring to. It will also be convenient to denote the natural inner product of R^3 by the same symbol $\langle\ ,\ \rangle$ rather than the previous dot.

Example 1. A coordinate system for a plane $P \subset R^3$ passing through $p_0 = (x_0, y_0, z_0)$ and containing the orthonormal vectors $w_1 = (a_1, a_2, a_3)$, $w_2 = (b_1, b_2, b_3)$ is given as follows:

$$\mathbf{x}(u, v) = p_0 + uw_1 + vw_2, \quad (u, v) \in R^2.$$

To compute the first fundamental form for an arbitrary point of P we observe that $\mathbf{x}_u = w_1$, $\mathbf{x}_v = w_2$; since w_1 and w_2 are unit orthogonal vectors, the functions E, F, G are constant and given by

$$E = 1, \quad F = 0, \quad G = 1.$$

In this trivial case, the first fundamental form is essentially the Pythagorean theorem in P; i.e., the square of the length of a vector w which has coordinates a, b in the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ is equal to $a^2 + b^2$.

Example 2. The right cylinder over the circle $x^2 + y^2 = 1$ admits the parametrization $\mathbf{x}\colon U \to R^3$, where (Fig. 2-26)

$$\mathbf{x}(u, v) = (\cos u, \sin u, v),$$
$$U = \{(u, v) \in R^2; \quad 0 < u < 2\pi, \quad -\infty < v < \infty\}.$$

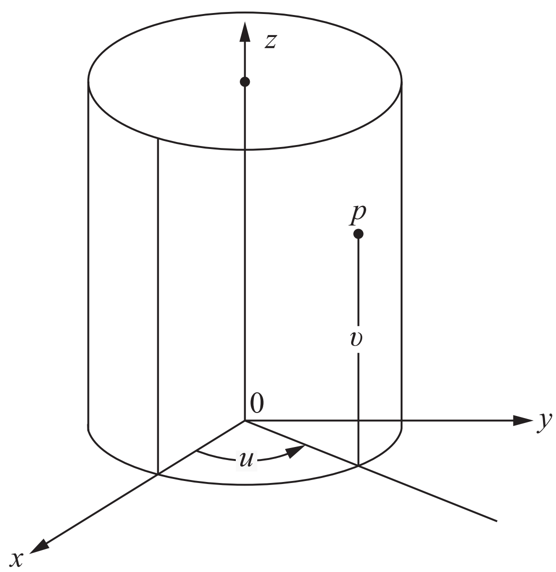


Figure 2-26

To compute the first fundamental form, we notice that

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

and therefore

$$E = \sin^2 u + \cos^2 u = 1, \quad F = 0, \quad G = 1.$$

We remark that, although the cylinder and the plane are distinct surfaces, we obtain the same result in both cases. We shall return to this subject later (Sec. 4-2).

Example 3. Consider a helix that is given by (see Example 1, Sec. 1-2) $(\cos u, \sin u, au)$. Through each point of the helix, draw a line parallel to the xy plane and intersecting the z axis. The surface generated by these lines is called a *helicoid* and admits the following parametrization:

$$\mathbf{x}(u, v) = (v \cos u, v \sin u, au), \quad 0 < u < 2\pi, \quad -\infty < v < \infty.$$

$\mathbf{x}$ applies an open strip with width 2π of the uv plane onto that part of the helicoid which corresponds to a rotation of 2π along the helix (Fig. 2-27).

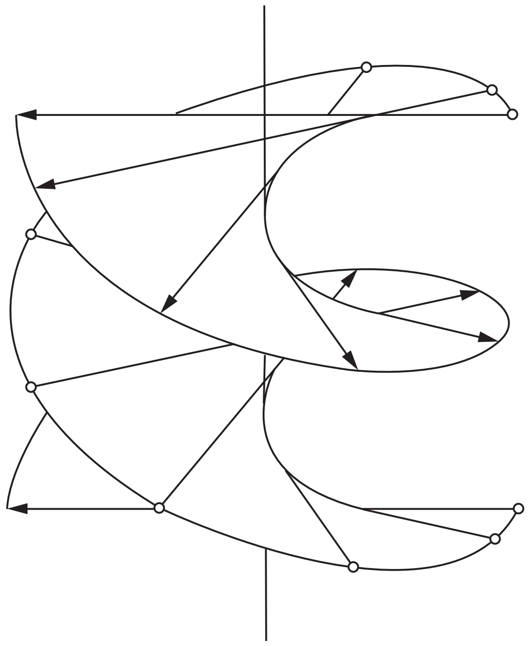


Figure 2-27. The helicoid.

The verification that the helicoid is a regular surface is straightforward and left to the reader.

The computation of the coefficients of the first fundamental form in the above parametrization gives

$$E(u, v) = v^2 + a^2, \quad F(u, v) = 0, \quad G(u, v) = 1.$$

As we mentioned before, the importance of the first fundamental form I comes from the fact that by knowing I we can treat metric questions on a regular surface without further references to the ambient space R^3. Thus, the arc length s of a parametrized curve $\alpha: I \to S$ is given by

$$s(t) = \int_o^t |\alpha'(t)|\, dt = \int_o^t \sqrt{I(\alpha'(t))}\, dt.$$

In particular, if $\alpha(t) = \mathbf{x}(u(t), v(t))$ is contained in a coordinate neighborhood corresponding to the parametrization $\mathbf{x}(u, v)$, we can compute the arc length of α between, say, 0 and t by

$$s(t) = \int_0^t \sqrt{E(u')^2 + 2Fu'v' + G(v')^2}\, dt. \qquad \textbf{(2)}$$

Also, the angle θ under which two parametrized regular curves $\alpha: I \to S$, $\beta: I \to S$ intersect at $t = t_0$ is given by

$$\cos\theta = \frac{\langle \alpha'(t_0), \beta'(t_0)\rangle}{|\alpha'(t_0)||\beta'(t_0)|}.$$

In particular, the angle φ of the coordinate curves of a parametrization $\mathbf{x}(u, v)$ is

$$\cos\varphi = \frac{\langle \mathbf{x}_u, \mathbf{x}_v \rangle}{|\mathbf{x}_u||\mathbf{x}_v|} = \frac{F}{\sqrt{EG}};$$

it follows that *the coordinate curves of a parametrization are orthogonal if and only if* F(u, v) = 0 *for all* (u, v). Such a parametrization is called an *orthogonal parametrization.*

Remark. Because of Eq. (2), many mathematicians talk about the "element" of arc length, ds of S, and write

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2,$$

meaning that if $\alpha(t) = \mathbf{x}(u(t), v(t))$ is a curve on S and $s = s(t)$ is its arc length, then

$$\left(\frac{ds}{dt}\right)^2 = E\left(\frac{du}{dt}\right)^2 + 2F\frac{du}{dt}\frac{dv}{dt} + G\left(\frac{dv}{dt}\right)^2.$$

Example 4. We shall compute the first fundamental form of a sphere at a point of the coordinate neighborhood given by the parametrization (cf. Example 1, Sec. 2-2)

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

First, observe that

$$\mathbf{x}_\theta(\theta, \varphi) = (\cos\theta\cos\varphi, \cos\theta\sin\varphi, -\sin\theta),$$
$$\mathbf{x}_\varphi(\theta, \varphi) = (-\sin\theta\sin\varphi, \sin\theta\cos\varphi, 0).$$

Hence,

$$E(\theta, \varphi) = \langle \mathbf{x}_\theta, \mathbf{x}_\theta \rangle = 1,$$
$$F(\theta, \varphi) = \langle \mathbf{x}_\theta, \mathbf{x}_\varphi \rangle = 0,$$
$$G(\theta, \varphi) = \langle \mathbf{x}_\varphi, \mathbf{x}_\varphi \rangle = \sin^2\theta.$$

Thus, if w is a tangent vector to the sphere at the point $\mathbf{x}(\theta, \varphi)$, given in the basis associated to $\mathbf{x}(\theta, \varphi)$ by

$$w = a\mathbf{x}_\theta + b\mathbf{x}_\varphi,$$

then the square of the length of w is given by

$$|w|^2 = I(w) = Ea^2 + 2Fab + Gb^2 = a^2 + b^2\sin^2\theta.$$

As an application, let us determine the curves in this coordinate neighborhood of the sphere which make a constant angle β with the meridians $\varphi = \text{const}$. These curves are called *loxodromes* (rhumb lines) of the sphere.

We may assume that the required curve $\alpha(t)$ is the image by $\mathbf{x}$ of a curve $(\theta(t), \varphi(t))$ of the $\theta\varphi$ plane. At the point $\mathbf{x}(\theta, \varphi)$ where the curve meets the meridian $\varphi = \text{const.}$, we have

$$\cos\beta = \frac{\langle \mathbf{x}_\theta, \alpha'(t)\rangle}{|\mathbf{x}_\theta||\alpha'(t)|} = \frac{\theta'}{\sqrt{(\theta')^2 + (\varphi')^2\sin^2\theta}},$$

since in the basis $\{\mathbf{x}_\theta, \mathbf{x}_\varphi\}$, the vector $\alpha'(t)$ has coordinates (θ', φ') and the vector $\mathbf{x}_\theta$ has coordinates $(1, 0)$. It follows that

$$(\theta')^2\tan^2\beta - (\varphi')^2\sin^2\theta = 0$$

or

$$\frac{\theta'}{\sin\theta} = \pm\frac{\varphi'}{\tan\beta},$$

whence, by integration, we obtain the equation of the loxodromes

$$\log\tan\left(\frac{\theta}{2}\right) = \pm(\varphi + c)\,\text{cotan}\,\beta,$$

where the constant of integration c is to be determined by giving one point $\mathbf{x}(\theta_0, \varphi_0)$ through which the curve passes.

Another metric question that can be treated by the first fundamental form is the computation (or definition) of the area of a bounded region of a regular surface S. A (regular) *domain* of S is an open and connected subset of S such that its boundary is the image in S of a circle by a differentiable homeomorphism which is regular (that is, its differential is nonzero) except at a finite number of points. A *region* of S is the union of a domain with its boundary (Fig. 2-28). A region of $S \subset R^3$ is *bounded* if it is contained in some ball of R^3.

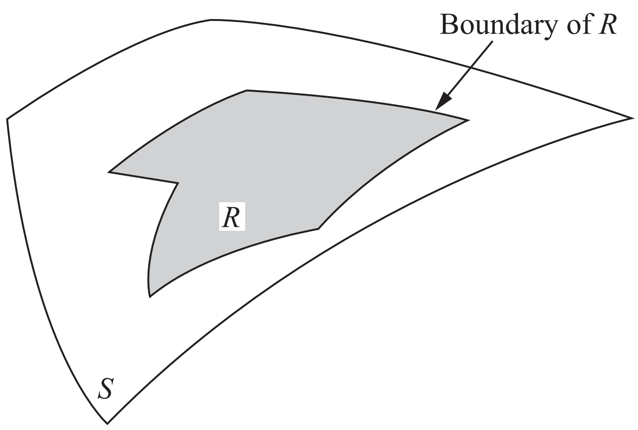


Figure 2-28

Let Q be a compact region in $\mathbb{R}^2$ that is contained in a coordinate neighborhood $\mathbf{x}\colon U \to S$. Then $\mathbf{x}(Q) = R$ is a bounded region in S.

The function $|\mathbf{x}_u \wedge \mathbf{x}_v|$, defined in U, measures the area of the parallelogram generated by the vectors $\mathbf{x}_u$ and $\mathbf{x}_v$. We first show that the integral

$$\int_Q |\mathbf{x}_u \wedge \mathbf{x}_v|\, du\, dv$$

does not depend on the parametrization $\mathbf{x}$.

In fact, let $\bar{\mathbf{x}}\colon \bar{U} \subset R^2 \to S$ be another parametrization with $R \subset \bar{\mathbf{x}}(\bar{U})$ and set $\bar{Q} = \bar{\mathbf{x}}^{-1}(R)$. Let $\partial(u, v)/\partial(\bar{u}, \bar{v})$ be the Jacobian of the change of parameters $h = \mathbf{x}^{-1} \circ \bar{\mathbf{x}}$. Then

$$\begin{aligned}\iint_{\bar{Q}} |\bar{\mathbf{x}}_{\bar{u}} \wedge \bar{\mathbf{x}}_{\bar{v}}|\, d\bar{u}\, d\bar{v} &= \iint_{\bar{Q}} |\mathbf{x}_u \wedge \mathbf{x}_v| \left|\frac{\partial(u, v)}{\partial(\bar{u}, \bar{v})}\right| d\bar{u}\, d\bar{v} \\ &= \iint_Q |\mathbf{x}_u \wedge \mathbf{x}_v|\, du\, dv,\end{aligned}$$

where the last equality comes from the theorem of change of variables in multiple integrals (cf. Buck *Advanced Calculus*, p. 304). The asserted independence is therefore proved and we can make the following definition.

DEFINITION 2. *Let* R $\subset$ S *be a bounded region of a regular surface contained in the coordinate neighborhood of the parametrization* $\mathbf{x}$: U $\subset$ R^2 $\to$ S. *The positive number*

$$\iint_Q |\mathbf{x}_u \wedge \mathbf{x}_v|\, \mathrm{du\, dv} = \mathrm{A(R)}, \quad \mathrm{Q} = \mathbf{x}^{-1}(\mathrm{R}),$$

is called the area *of* R.

There are several geometric justifications for such a definition, and one of them will be presented in Sec. 2-8.

It is convenient to observe that

$$|\mathbf{x}_u \wedge \mathbf{x}_v|^2 + \langle \mathbf{x}_u \cdot \mathbf{x}_v \rangle^2 = |\mathbf{x}_u|^2 \cdot |\mathbf{x}_v|^2,$$

which shows that the integrand of $A(R)$ can be written as

$$|\mathbf{x}_u \wedge \mathbf{x}_v| = \sqrt{EG - F^2}.$$

We should also remark that, in most examples, the restriction that the region R is contained in some coordinate neighborhood is not very serious, because there exist coordinate neighborhoods which cover the entire surface except for some curves, which do not contribute to the area.

Example 5. Let us compute the area of the torus of Example 6, Sec. 2-2. For that, we consider the coordinate neighborhood corresponding to the parametrization

$$\mathbf{x}(u, v) = ((a + r\cos u)\cos v, (a + r\cos u)\sin v, r\sin u),$$
$$0 < u < 2\pi, \quad 0 < v < 2\pi,$$

which covers the torus, except for a meridian and a parallel. The coefficients of the first fundamental form are

$$E = r^2, \quad F = 0, \quad G = (r\cos u + a)^2;$$

hence,

$$\sqrt{EG - F^2} = r(r\cos u + a).$$

Now, consider the region R_ϵ obtained as the image by $\mathbf{x}$ of the region Q_ϵ (Fig. 2-29) given by ($\epsilon > 0$ and small),

$$Q_\epsilon = \{(u, v) \in R^2; 0 + \epsilon \le u \le 2\pi - \epsilon, 0 + \epsilon \le v \le 2\pi - \epsilon\}.$$

Using Def. 2, we obtain

$$\begin{aligned} A(R_\epsilon) &= \iint_{Q_\epsilon} r(r\cos u + a)\, du\, dv \\ &= \int_{0+\epsilon}^{2\pi-\epsilon} (r^2\cos u + ra)\, du \int_{0+\epsilon}^{2\pi-\epsilon} dv \\ &= r^2(2\pi - 2\epsilon)(\sin(2\pi - \epsilon) - \sin\epsilon) + ra(2\pi - 2\epsilon)^2. \end{aligned}$$

Letting $\epsilon \to 0$ in the above expression, we obtain

$$A(T) = \lim_{\epsilon\to 0} A(R_\epsilon) = 4\pi^2 ra.$$

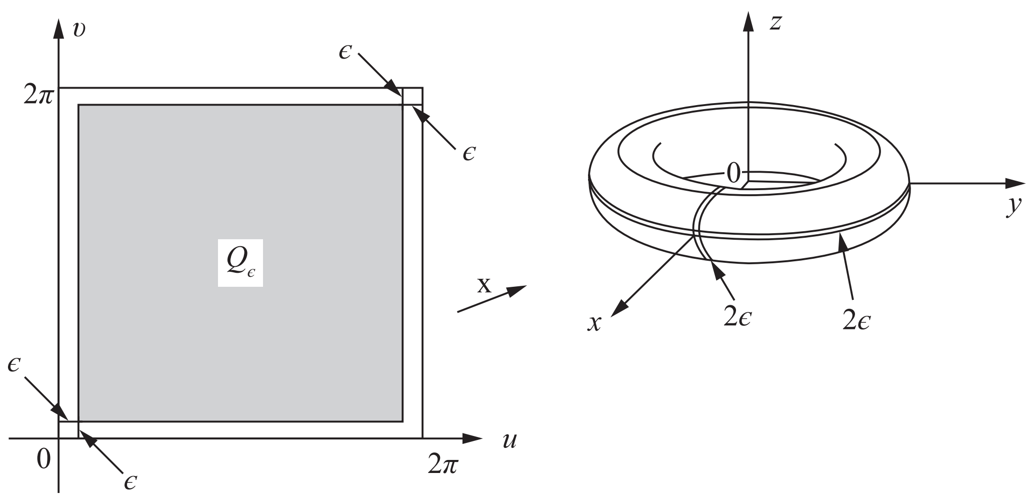


Figure 2-29

This agrees with the value found by elementary calculus, say, by using the theorem of Pappus for the area of surfaces of revolution (cf. Exercise 11).

EXERCISES

1. Compute the first fundamental forms of the following parametrized surfaces where they are regular:

a. $\mathbf{x}(u, v) = (a \sin u \cos v, b \sin u \sin v, c \cos u)$; ellipsoid.

b. $\mathbf{x}(u, v) = (au \cos v, bu \sin v, u^2)$; elliptic paraboloid.

c. $\mathbf{x}(u, v) = (au \cosh v, bu \sinh v, u^2)$; hyperbolic paraboloid.

d. $\mathbf{x}(u, v) = (a \sinh u \cos v, b \sinh u \sin v, c \cosh u)$; hyperboloid of two sheets.

2. Let $\mathbf{x}(\varphi, \theta) = (\sin\theta \cos\varphi, \sin\theta \sin\varphi, \cos\theta)$ be a parametrization of the unit sphere S^2. Let P be the plane $x = z \operatorname{cotan} \alpha$, $0 < \alpha < \pi$, and β be the acute angle which the curve $P \cap S^2$ makes with the semimeridian $\varphi = \varphi_0$. Compute $\cos\beta$.

3. Obtain the first fundamental form of the sphere in the parametrization given by stereographic projection (cf. Exercise 16, Sec. 2-2).

4. Given the parametrized surface

$$\mathbf{x}(u, v) = (u \cos v, u \sin v, \log \cos v + u), \qquad -\frac{\pi}{2} < v < \frac{\pi}{2},$$

show that the two curves $\mathbf{x}(u, v_1)$, $\mathbf{x}(u, v_2)$ determine segments of equal lengths on all curves $\mathbf{x}(u, \text{const.})$.

5. Show that the area A of a bounded region R of the surface $z = f(x, y)$ is

$$A = \iint_Q \sqrt{1 + f_x^2 + f_y^2}\, dx\, dy,$$

where Q is the normal projection of R onto the xy plane.

6. Show that

$$\mathbf{x}(u, v) = (u \sin\alpha \cos v, u \sin\alpha \sin v, u \cos\alpha)$$
$$0 < u < \infty, \quad 0 < v < 2\pi, \quad \alpha = \text{const.},$$

is a parametrization of the cone with 2α as the angle of the vertex. In the corresponding coordinate neighborhood, prove that the curve

$$\mathbf{x}(c \exp(v \sin\alpha \operatorname{cotan}\beta), v), \qquad c = \text{const.}, \beta = \text{const.},$$

intersects the generators of the cone ($v = \text{const.}$) under the constant angle β.

7. The coordinate curves of a parametrization $\mathbf{x}(u, v)$ constitute a *Tchebyshef net* if the lengths of the opposite sides of any quadrilateral formed by them are equal. Show that a necessary and sufficient condition for this is

$$\frac{\partial E}{\partial v} = \frac{\partial G}{\partial u} = 0.$$

***8.** Prove that whenever the coordinate curves constitute a Tchebyshef net (see Exercise 7) it is possible to reparametrize the coordinate neighborhood in such a way that the new coefficients of the first fundamental form are

$$E = 1, \quad F = \cos\theta, \quad G = 1,$$

where θ is the angle of the coordinate curves.

***9.** Show that a surface of revolution can always be parametrized so that

$$E = E(v), \quad F = 0, \quad G = 1.$$

10. Let $P = \{(x, y, z) \in R^3; z = 0\}$ be the xy plane and let $\mathbf{x}: U \to P$ be a parametrization of P given by

$$\mathbf{x}(\rho, \theta) = (\rho\cos\theta, \rho\sin\theta),$$

where

$$U = \{(\rho, \theta) \in R^2; \rho > 0, 0 < \theta < 2\pi\}.$$

Compute the coefficients of the first fundamental form of P in this parametrization.

11. Let S be a surface of revolution and C its generating curve (cf. Example 4, Sec. 2-3). Let s be the arc length of C and denote by $\rho = \rho(s)$ the distance to the rotation axis of the point of C corresponding to s.

a. (*Pappus' Theorem.*) Show that the area of S is

$$2\pi\int_0^l \rho(s)\, ds,$$

where l is the length of C.

b. Apply part a to compute the area of a torus of revolution.

12. Show that the area of a regular tube of radius r around a curve α (cf. Exercise 10, Sec. 2-4) is $2\pi r$ times the length of α.

13. (*Generalized Helicoids.*) A natural generalization of both surfaces of revolution and helicoids is obtained as follows. Let a regular plane curve C, which does not meet an axis E in the plane, be displaced in a rigid screw motion about E, that is, so that each point of C describes a helix (or circle) with E as axis. The set S generated by the displacement of C is called a *generalized helicoid* with *axis* E and *generator* C. If the screw motion is a pure rotation about E, S is a surface of revolution; if C is a straight line perpendicular to E, S is (a piece of) the standard helicoid (cf. Example 3).

Choose the coordinate axes so that E is the z axis and C lies in the yz plane. Prove that

a. If $(f(s), g(s))$ is a parametrization of C by arc length s, $a < s < b$, $f(s) > 0$, then $\mathbf{x}\colon U \to S$, where

$$U = \{(s, u) \in R^2;\, a < s < b, 0 < u < 2\pi\}$$

and

$$\mathbf{x}(s, u) = (f(s)\cos u,\, f(s)\sin u,\, g(s) + cu), \quad c = \text{const.},$$

is a parametrization of S. Conclude that S is a regular surface.

b. The coordinate lines of the above parametrization are orthogonal (i.e., $F = 0$) if and only if $\mathbf{x}(U)$ is either a surface of revolution or (a piece of) the standard helicoid.

14. (*Gradient on Surfaces.*) The *gradient* of a differentiable function $f\colon S \to R$ is a differentiable map grad $f\colon S \to R^3$ which assigns to each point $p \in S$ a vector grad $f(p) \in T_p(S) \subset R^3$ such that

$$\langle \operatorname{grad} f(p), v\rangle_p = df_p(v) \quad \text{for all } v \in T_p(S).$$

Show that

a. If E, F, G are the coefficients of the first fundamental form in a parametrization $\mathbf{x}\colon U \subset R^2 \to S$, then grad f on $\mathbf{x}(U)$ is given by

$$\operatorname{grad} f = \frac{f_u G - f_v F}{EG - F^2}\mathbf{x}_u + \frac{f_v E - f_u F}{EG - F^2}\mathbf{x}_v.$$

In particular, if $S = R^2$ with coordinates x, y,

$$\operatorname{grad} f = f_x e_1 + f_y e_2,$$

where $\{e_1, e_2\}$ is the canonical basis of R_2 (*thus, the definition agrees with the usual definition of gradient in the plane*).

b. If you let $p \in S$ be fixed and v vary in the unit circle $|v| = 1$ in $T_p(s)$, then $df_p(v)$ is maximum if and only if $v = \operatorname{grad} f/|\operatorname{grad} f|$ (*thus*, grad $f(p)$ *gives the direction of maximum variation of* f *at* p).

c. If grad $f \neq 0$ at all points of the *level curve* $C = \{q \in S;\, f(q) = \text{const.}\}$, then C is a regular curve on S and grad f is normal to C at all points of C.

15. (*Orthogonal Families of Curves.*)

a. Let E, F, G be the coefficients of the first fundamental form of a regular surface S in the parametrization $\mathbf{x}$: $U \subset R^2 \to S$. Let $\varphi(u, v) =$ const. and $\psi(u, v) =$ const. be two families of regular curves on $\mathbf{x}(U) \subset S$ (cf. Exercise 28, Sec. 2-4). Prove that these two families are orthogonal (i.e., whenever two curves of distinct families meet, their tangent lines are orthogonal) if and only if

$$E\varphi_v\psi_v - F(\varphi_u\psi_v + \varphi_v\psi_u) + G\varphi_u\psi_u = 0.$$

b. Apply part a to show that on the coordinate neighborhood $\mathbf{x}(U)$ of the helicoid of Example 3 the two families of regular curves

$$v \cos u = \text{const.}, \quad v \neq 0,$$
$$(v^2 + a^2)\sin^2 u = \text{const.}, \quad v \neq 0, \quad u \neq \pi,$$

are orthogonal.

2-6. Orientation of Surfaces†

In this section we shall discuss in what sense, and when, it is possible to orient a surface. Intuitively, since every point p of a regular surface S has a tangent plane $T_p(S)$, the choice of an orientation of $T_p(S)$ induces an orientation in a neighborhood of p, that is, a notion of positive movement along sufficiently small closed curves about each point of the neighborhood (Fig. 2-30). If it is possible to make this choice for each $p \in S$ so that in the intersection of any two neighborhoods the orientations coincide, then S is said to be orientable. If this is not possible, S is called nonorientable.

We shall now make these ideas precise. By fixing a parametrization $\mathbf{x}(u, v)$ of a neighborhood of a point p of a regular surface S, we determine an orientation of the tangent plane $T_p(S)$, namely, the orientation of the associated ordered basis $\{\mathbf{x}_u, \mathbf{x}_v\}$. If p belongs to the coordinate neighborhood of another parametrization $\bar{\mathbf{x}}(\bar u, \bar v)$, the new basis $\{\bar{\mathbf{x}}_{\bar u}, \bar{\mathbf{x}}_{\bar v}\}$ is expressed in terms of the first one by

$$\bar{\mathbf{x}}_{\bar u} = \mathbf{x}_u \frac{\partial u}{\partial \bar u} + \mathbf{x}_v \frac{\partial v}{\partial \bar u},$$
$$\mathbf{x}_{\bar v} = \mathbf{x}_u \frac{\partial u}{\partial \bar v} + \mathbf{x}_v \frac{\partial v}{\partial \bar v},$$

where $u = u(\bar u, \bar v)$ and $v = v(\bar u, \bar v)$ are the expressions of the change of coordinates. The bases $\{\mathbf{x}_u, \mathbf{x}_v\}$ and $\{\bar{\mathbf{x}}_{\bar u}, \bar{\mathbf{x}}_{\bar v}\}$ determine, therefore, the same orientation of $T_p(S)$ if and only if the Jacobian

†This section may be omitted on a first reading.

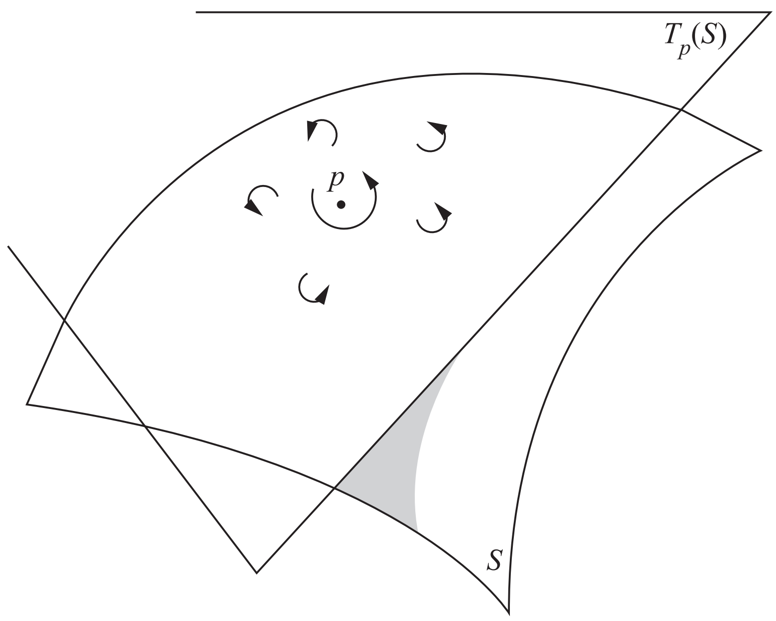


Figure 2-30

$$\frac{\partial(u, v)}{\partial(\bar{u}, \bar{v})}$$

of the coordinate change is positive.

DEFINITION 1. *A regular surface* S *is called* orientable *if it is possible to cover it with a family of coordinate neighborhoods in such a way that if a point* p ∈ S *belongs to two neighborhoods of this family, then the change of coordinates has positive Jacobian at* p*. The choice of such a family is called an* orientation *of* S*, and* S*, in this case, is called* oriented*. If such a choice is not possible, the surface is called* nonorientable*.*

Example 1. A surface which is the graph of a differentiable function (cf. Sec. 2-2, Prop. 1) is an orientable surface. In fact, all surfaces which can be covered by one coordinate neighborhood are trivially orientable.

Example 2. The sphere is an orientable surface. Instead of proceeding to a direct calculation, let us resort to a general argument. The sphere can be covered by two coordinate neighborhoods (using stereographic projection; see Exercise 16 of Sec. 2-2), with parameters (u, v) and $(\bar{u}, \bar{v})$, in such a way that the intersection W of these neighborhoods (the sphere minus two points) is a connected set. Fix a point p in W. If the Jacobian of the coordinate change at p is negative, we interchange u and v in the first system, and the Jacobian becomes positive. Since the Jacobian is different from zero in W and positive at $p \in W$, it follows from the connectedness of W that the Jacobian is everywhere positive. There exists, therefore, a family of coordinate neighborhoods satisfying Def. 1, and so the sphere is orientable.

By the argument just used, it is clear that *if a regular surface can be covered by two coordinate neighborhoods whose intersection is connected, then the surface is orientable.*

Before presenting an example of a nonorientable surface, we shall give a geometric interpretation of the idea of orientability of a regular surface in R^3.

As we have seen in Sec. 2-4, given a system of coordinates $\mathbf{x}(u, v)$ at p, we have a definite choice of a unit normal vector N at p by the rule

$$N = \frac{\mathbf{x}_u \wedge \mathbf{x}_v}{|\mathbf{x}_u \wedge \mathbf{x}_v|}(p). \tag{1}$$

Taking another system of local coordinates $\bar{\mathbf{x}}(\bar{u}, \bar{v})$ at p, we see that

$$\bar{\mathbf{x}}_{\bar{u}} \wedge \bar{\mathbf{x}}_{\bar{v}} = (\mathbf{x}_u \wedge \mathbf{x}_v)\frac{\partial(u, v)}{\partial(\bar{u}, \bar{v})}, \tag{2}$$

where $\partial(u, v)/\partial(\bar{u}, \bar{v})$ is the Jacobian of the coordinate change. Hence, N will preserve its sign or change it, depending on whether $\partial(u, v)/\partial(\bar{u}, \bar{v})$ is positive or negative, respectively.

By a differentiable *field of unit normal vectors* on an open set $U \subset S$, we shall mean a differentiable map $N: U \to R^3$ which associates to each $q \in U$ a unit normal vector $N(q) \in R^3$ to S at q.

PROPOSITION 1. *A regular surface* $\mathrm{S} \subset \mathrm{R}^3$ *is orientable if and only if there exists a differentiable field of unit normal vectors* $\mathrm{N}: \mathrm{S} \to \mathrm{R}^3$ *on* S.

Proof. If S is orientable, it is possible to cover it with a family of coordinate neighborhoods so that, in the intersection of any two of them, the change of coordinates has a positive Jacobian. At the points $p = \mathbf{x}(u, v)$ of each neighborhood, we define $N(p) = N(u, v)$ by Eq. (1). $N(p)$ is well defined, since if p belongs to two coordinate neighborhoods, with parameters (u, v) and $(\bar{u}, \bar{v})$, the normal vector $N(u, v)$ and $N(\bar{u}, \bar{v})$ coincide by Eq. (2). Moreover, by Eq. (1), the coordinates of $N(u, v)$ in R^3 are differentiable functions of (u, v), and thus the mapping $N: S \to R^3$ is differentiable, as desired.

On the other hand , let $N: S \to R^3$ be a differentiable field of unit normal vectors, and consider a family of *connected* coordinate neighborhoods covering S. For the points $p = \mathbf{x}(u, v)$ of each coordinate neighborhood $\mathbf{x}(U)$, $U \subset R^2$, it is possible, by the continuity of N and, if necessary, by interchanging u and v, to arrange that

$$N(p) = \frac{\mathbf{x}_u \wedge \mathbf{x}_v}{|\mathbf{x}_u \wedge \mathbf{x}_v|}.$$

In fact, the inner product

$$\left\langle N(p), \frac{\mathbf{x}_u \wedge \mathbf{x}_v}{|\mathbf{x}_u \wedge \mathbf{x}_v|}\right\rangle = f(p) = \pm 1$$

is a continuous function on $\mathbf{x}(U)$. Since $\mathbf{x}(U)$ is connected, the sign of f is constant. If $f = -1$, we interchange u and v in the parametrization, and the assertion follows.

Proceeding in this manner with all the coordinate neighborhoods, we have that in the intersection of any two of them, say, $\mathbf{x}(u, v)$ and $\bar{\mathbf{x}}(\bar{u}, \bar{v})$, the Jacobian

$$\frac{\partial(u, v)}{\partial(\bar{u}, \bar{v})}$$

is certainly positive; otherwise, we would have

$$\frac{\mathbf{x}_u \wedge \mathbf{x}_v}{|\mathbf{x}_u \wedge \mathbf{x}_v|} = N(p) = -\frac{\bar{\mathbf{x}}_{\bar{u}} \wedge \bar{\mathbf{x}}_{\bar{v}}}{|\bar{\mathbf{x}}_{\bar{u}} \wedge \bar{\mathbf{x}}_{\bar{v}}|} = -N(p),$$

which is a contradiction. Hence, the given family of coordinate neighborhoods after undergoing certain interchanges of u and v satisfies the conditions of Def. 1, and S is, therefore, orientable. **Q.E.D.**

Remark. As the proof shows, we need only to require the existence of a *continuous* unit vector field on S for S to be orientable. Such a vector field will be automatically differentiable.

Example 3. We shall now describe an example of a nonorientable surface, the so-called *Möbius strip*. This surface is obtained (see Fig. 2-31) by considering the circle S^1 given by $x^2 + y^2 = 4$ and the open segment AB given in the yz plane by $y = 2$, $|z| < 1$. We move the center C of AB along S^1 and turn AB about C in the Cz plane in such a manner that when C has passed through an angle u, AB has rotated by an angle $u/2$. When c completes one trip around the circle, AB returns to its initial position, with its end points inverted.

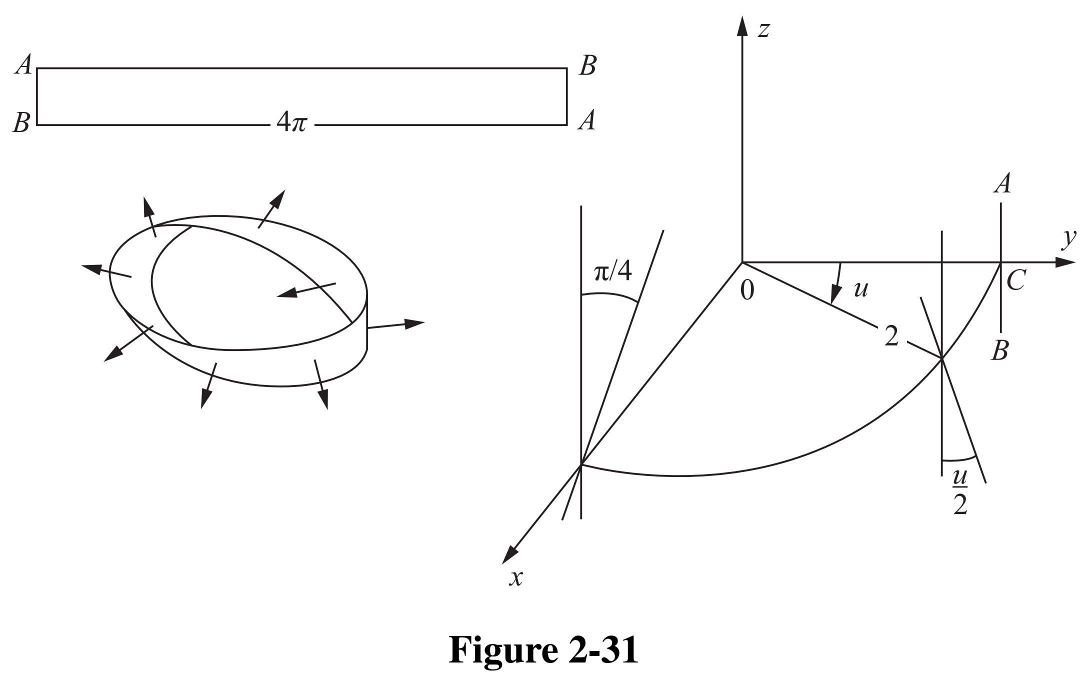


Figure 2-31

From the point of view of differentiability, it is as if we had identified the opposite (vertical) sides of a rectangle giving a twist to the rectangle so that each point of the side AB was identified with its symmetric point (Fig. 2-31).

It is geometrically evident that the Möbius strip M is a regular, nonorieotable surface. In fact, if M were orientable, there would exist a differentiable field $N: M \to R^3$ of unit normal vectors. Taking these vectors along the circle $x^2 + y^2 = 4$ we see that after making one trip the vector N returns to its original position as $-N$, which is a contradiction.

We shall now give an analytic proof of the facts mentioned above.

A system of coordinates $\mathbf{x}: U \to M$ for the Möbius strip is given by

$$\mathbf{x}(u, v) = \left(\left(2 - v \sin \frac{u}{2}\right) \sin u, \left(2 - v \sin \frac{u}{2}\right) \cos u, v \cos \frac{u}{2}\right),$$

where $0 < u < 2\pi$ and $-1 < v < 1$. The corresponding coordinate neighborhood omits the points of the open interval $u = 0$. Then by taking the origin of the u's at the x axis, we obtain another parametrization $\bar{\mathbf{x}}(\bar{u}, \bar{v})$ given by

$$\begin{aligned}
x &= \left\{2 - \bar{v} \sin\left(\frac{\pi}{4} + \frac{\bar{u}}{2}\right)\right\} \cos \bar{u},\\
y &= -\left\{2 - \bar{v} \sin\left(\frac{\pi}{4} + \frac{\bar{u}}{2}\right)\right\} \sin \bar{u},\\
z &= \bar{v} \cos\left(\frac{\pi}{4} + \frac{\bar{u}}{2}\right),
\end{aligned}$$

whose coordinate neighborhood omits the interval $u = \pi/2$. These two coordinate neighborhoods cover the Möbius strip and can be used to show that it is a regular surface.

Observe that the intersection of the two coordinate neighborhoods is not connected but consists of two connected components:

$$\begin{aligned}
W_1 &= \left\{\mathbf{x}(u, v): \frac{\pi}{2} < u < 2\pi\right\},\\
W_2 &= \left\{\mathbf{x}(u, v): 0 < u < \frac{\pi}{2}\right\}.
\end{aligned}$$

The change of coordinates is given by

$$\left.\begin{aligned}
\bar{u} &= u - \frac{\pi}{2}\\
\bar{v} &= v
\end{aligned}\right\} \quad \text{in } W_1,$$

and

$$\left.\begin{aligned}
\bar{u} &= \frac{3\pi}{2} + u\\
\bar{v} &= -v
\end{aligned}\right\} \quad \text{in } W_2.$$

It follows that

$$\frac{\partial(\bar{u}, \bar{v})}{\partial(u, v)} = 1 > 0 \quad \text{in } W_1$$

and that

$$\frac{\partial(\bar{u}, \bar{v})}{\partial(u, v)} = -1 < 0 \quad \text{in } W_2.$$

To show that the Möbius strip is nonorientable, we suppose that it is possible to define a differentiable field of unit normal vectors $N: M \to R^3$. Interchanging u and v if necessary, we can assume that

$$N(p) = \frac{\mathbf{x}_u \wedge \mathbf{x}_v}{|\mathbf{x}_u \wedge \mathbf{x}_v|}$$

for any p in the coordinate neighborhood of $\mathbf{x}(u, v)$. Analogously, we may assume that

$$N(p) = \frac{\bar{\mathbf{x}}_{\bar{u}} \wedge \bar{\mathbf{x}}_{\bar{v}}}{|\bar{\mathbf{x}}_{\bar{u}} \wedge \bar{\mathbf{x}}_{\bar{v}}|}$$

at all points of the coordinate neighborhood of $\bar{\mathbf{x}}(\bar{u}, \bar{v})$. However, the Jacobian of the change of coordinates must be -1 in either W_1 or W_2 (depending on what changes of the type $u \to v$, $\bar{u} \to \bar{v}$ has to be made). If p is a point of that component of the intersection, then $N(p) = -N(p)$, which is a contradiction.

We have already seen that a surface which is the graph of a differentiable function is orientable. We shall now show that a surface which is the inverse image of a regular value of a differentiable function is also orientable. This is one of the reasons it is relatively difficult to construct examples of nonorientable, regular surfaces in R^3.

PROPOSITION 2. *If a regular surface is given by* $S = \{(x, y, z) \in R^3; f(x, y, z) = a\}$, *where* $f: U \subset R^3 \to R$ *is differentiable and* a *is a regular value of* f, *then* S *is orientable.*

Proof. Given a point $(x_0, y_0, z_0) = p \in S$, consider the parametrized curve $(x(t), y(t), z(t))$, $t \in I$, on S passing through p for $t = t_0$. Since the curve is on S, we have

$$f(x(t), y(t), z(t)) = a$$

for all $t \in I$. By differentiating both sides of this expression with respect to t, we see that at $t = t_0$

$$f_x(p)\left(\frac{dx}{dt}\right)_{t_0} + f_y(p)\left(\frac{dy}{dt}\right)_{t_0} + f_z(p)\left(\frac{dz}{dt}\right)_{t_0} = 0.$$

This shows that the tangent vector to the curve at $t = t_0$ is perpendicular to the vector (f_x, f_y, f_z) at p. Since the curve and the point are arbitrary, we conclude that

$$N(x, y, z) = \left(\frac{f_x}{\sqrt{f_x^2 + f_y^2 + f_z^2}}, \frac{f_y}{\sqrt{f_x^2 + f_y^2 + f_z^2}}, \frac{f_z}{\sqrt{f_x^2 + f_y^2 + f_z^2}} \right)$$

is a differentiable field of unit normal vectors on S. Together with Prop. 1, this implies that S is orientable as desired. **Q.E.D.**

A final remark. Orientation is definitely not a local property of a regular surface. Locally, every regular surface is diffeomorphic to an open set in the plane, and hence orientable. Orientation is a global property, in the sense that it involves the whole surface. We shall have more to say about global properties later in this book (Chap. 5).

EXERCISES

1. Let S be a regular surface covered by coordinate neighborhoods V_1 and V_2. Assume that $V_1 \cap V_2$ has two connected components, W_1, W_2, and that the Jacobian of the change of coordinates is positive in W_1 and negative in W_2. Prove that S is nonorientable.

2. Let S_2 be an orientable regular surface and φ: $S_1 \to S_2$ be a differentiable map which is a local diffeomorphism at every $p \in S_1$. Prove that S_1 is orientable.

3. Is it possible to give a meaning to the notion of area for a Möbius strip? If so, set up an integral to compute it.

4. Let S be an orientable surface and let $\{U_\alpha\}$ and $\{V_\beta\}$ be two families of coordinate neighborhoods which cover S (that is, $\bigcup U_\alpha = S = \bigcup V_\beta$) and satisfy the conditions of Def. 1 (that is, in each of the families, the coordinate changes have positive Jacobian). We say that $\{U_\alpha\}$ and $\{V_\beta\}$ determine the *same orientation* of S if the union of the two families again satisfies the conditions of Def. 1.

Prove that a regular, connected, orientable surface can have only two distinct orientations.

5. Let φ: $S_1 \to S_2$ be a diffeomorphism.

a. Show that S_1 is orientable if and only if S_2 is orientable (*thus, orientability is preserved by diffeomorphisms*).

b. Let S_1 and S_2 be orientable and oriented. Prove that the diffeomorphism φ induces an orientation in S_2. Use the antipodal map of the sphere (Exercise 1, Sec. 2-3) to show that this orientation may be distinct (cf. Exercise 4) from the initial one (*thus, orientation itself may not be preserved by diffeomorphisms; note, however, that if* S_1 *and* S_2 *are connected, a diffeomorphism either preserves or "reverses" the orientation*).

6. Define the notion of orientation of a regular curve $C \subset R^3$, and show that if C is connected, there exist at most two distinct orientations in the sense of Exercise 4 (actually there exist exactly two, but this is harder to prove).

7. Show that if a regular surface S contains an open set diffeomorphic to a Möbius strip, then S is nonorientable.

2-7. A Characterization of Compact Orientable Surfaces†

The converse of Prop. 2 of Sec. 2-6, namely, that *an orientable surface in* R^3 *is the inverse image of a regular value of some differentiable function*, is true and nontrivial to prove. Even in the particular case of compact surfaces (defined in this section), the proof is instructive and offers an interesting example of a global theorem in differential geometry. This section will be dedicated entirely to the proof of this converse statement.

Let $S \subset R^3$ be an orientable surface. The crucial point of the proof consists of showing that one may choose, on the normal line through $p \in S$, an open interval I_p around p of length, say, $2\epsilon_p$ (ϵ_p varies with p) in such a way that if $p \neq q \in S$, then $I_p \cap I_q = \phi$. Thus, the union $\bigcup I_p$, $p \in S$, constitutes an open set V of R^3, which contains S and has the property that through each point of V there passes a unique normal line to S; V is then called a *tubular neighborhood* of S (Fig. 2-32).

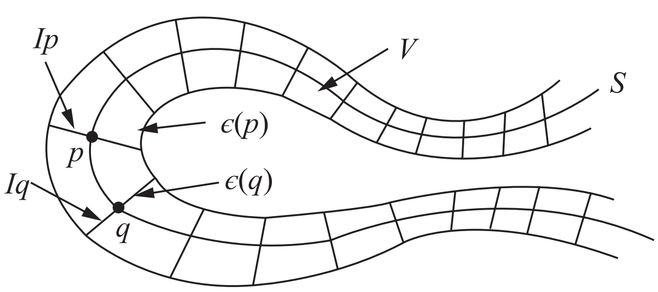


Figure 2-32. A tubular neighborhood.

Let us assume, for the moment, the existence of a tubular neighborhood V of an orientable surface S. We can then define a function $g: V \to R$ as follows: Fix an orientation for S. Observe that no two segments I_p and I_q, $p \neq q$, of the tubular neighborhood V intersect. Thus, through each point $P \in V$ there passes a unique normal line to S which meets S at a point p; by definition, $g(P)$ is the distance from p to P, with a sign given by the

†This section may be omitted on a first reading.

direction of the unit normal vector at p. If we can prove that g is a differentiable function and that 0 is a regular value of g, we shall have that $S = g^{-1}(0)$, as we wished to prove.

We shall now start the proof of the existence of a tubular neighborhood of an orientable surface. We shall first prove a local version of this fact; that is, we shall show that for each point p of a regular surface there exists a neighborhood of p which has a tubular neighborhood.

PROPOSITION 1. *Let* S *be a regular surface and* $\mathbf{x}$: U $\to$ S *be a parametrization of a neighborhood of a point* p $= \mathbf{x}(u_0, v_0) \in$ S. *Then there exists a neighborhood* W $\subset \mathbf{x}$(U) *of* p *in* S *and a number* $\epsilon > 0$ *such that the segments of the normal lines passing through points* q $\in$ W, *with center at* q *and length* 2ϵ, *are disjoint* (*that is,* W *has a tubular neighborhood*).

Proof. Consider the map F: U $\times$ R $\to$ R^3 given by

$$F(u, v; t) = \mathbf{x}(u, v) + tN(u, v), \quad (u, v) \in U, \quad t \in R,$$

where $N(u, v) = (N_x, N_y, N_z)$ is the unit normal vector at

$$\mathbf{x}(u, v) = (x(u, v), y(u, v), z(u, v)).$$

Geometrically, F maps the point $(u, v; t)$ of the "cylinder" $U \times R$ in the point of the normal line to S at a distance t from $\mathbf{x}(u, v)$. F is clearly differentiable and its Jacobian at $t = 0$ is given by

$$\begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} & \dfrac{\partial z}{\partial u} \\ \dfrac{\partial x}{\partial v} & \dfrac{\partial y}{\partial v} & \dfrac{\partial z}{\partial v} \\ N_x & N_y & N_z \end{vmatrix} = |\mathbf{x}_u \wedge \mathbf{x}_v| \neq 0.$$

By the inverse function theorem, there exists a parallelepiped in $U \times R$, say,

$$u_0 - \delta < u < u_0 + \delta, \quad v_0 - \delta < v < v_0 + \delta, \quad -\epsilon < t < \epsilon,$$

restricted to which F is one-to-one. But this means that in the image W by F of the rectangle

$$u_0 - \delta < u < u_0 + \delta, \quad v_0 - \delta < v < v_0 + \delta$$

the segments of the normal lines with centers $q \in W$ and of length $< 2\epsilon$ do not meet. **Q.E.D.**

At this point, it is convenient to observe the following. The fact that the function g: $V \to R$, defined above by assuming the existence of a tubular neighborhood V, is differentiable and has 0 as a regular value is a local fact and can be proved at once.

PROPOSITION 2. *Assume the existence of a tubular neighborhood* $\mathrm{V} \subset \mathrm{R}^3$ *of an orientable surface* $\mathrm{S} \subset \mathrm{R}^3$*, and choose an orientation for* S. *Then the function* $\mathrm{g}\colon \mathrm{V} \to \mathrm{R}$*, defined as the oriented distance from a point of* V *to the foot of the unique normal line passing through this point, is differentiable and has zero as a regular value.*

Proof. Let us look again at the map $F\colon U \times R \to R^3$ defined in Prop. 1, where we now assume that the parametrization $\mathbf{x}$ is compatible with the given orientation. Denoting by x, y, z the coordinates of $F(u, v, t) = \mathbf{x}(u, v) + tN(u, v)$ we can write

$$F(u, v, t) = (x(u, v, t), y(u, v, t), z(u, v, t)).$$

Since the Jacobian $\partial(x, y, z)/\partial(u, v, t)$ is different from zero at $t = 0$, we can invert F in some parallelepiped Q,

$$u_0 - \delta < u < u_0 + \delta, \quad v_0 - \delta < v < v_0 + \delta, \quad -\epsilon < t < \epsilon,$$

to obtain a differentiable map

$$F^{-1}(x, y, z) = (u(x, y, z), v(x, y, z), t(x, y, z)),$$

where $(x, y, z) \in F(Q) \subset V$. But the restriction to $F(Q)$ of the function $g\colon V \to R$ in the statement of Prop. 2 is precisely $t = t(x, y, z)$. Thus, g is differentiable. Furthermore, 0 is a regular value of t; otherwise

$$\frac{\partial t}{\partial x} = \frac{\partial t}{\partial y} = \frac{\partial t}{\partial z} = 0$$

for some point where $t = 0$; hence, the differential dF^{-1} would be singular for $t = 0$, which is a contradiction. **Q.E.D.**

To pass from the local to the global, that is, to prove the existence of a tubular neighborhood of an entire orientable surface, we need some topological arguments. We shall restrict ourselves to compact surfaces, which we shall now define.

Let A be a subset of R^3. We say that $p \in R^3$ is a *limit point* of A if every neighborhood of p in R^3 contains a point of A distinct from p. A is said to be *closed* if it contains all its limit points. A is *bounded* if it is contained in some ball of R^3. If A is closed and bounded, it is called a *compact set.*

The sphere and the torus are compact surfaces. The paraboloid of revolution $z = x^2 + y^2$, $(x, y) \in R^2$, is a closed surface, but, being unbounded, it is not a compact surface. The disk $x^2 + y^2 < 1$ in the plane and the Möbius strip are bounded but not closed and therefore are noncompact.

We shall need some properties of compact subsets of R^3, which we shall now state. The distance between two points $p, q \in R^3$ will be denoted by $d(p, q)$.

PROPERTY 1 (Bolzano-Weierstrass). *Let* $\mathrm{A} \subset \mathrm{R}^3$ *be a compact set. Then every infinite subset of* A *has at least one limit point in* A.

PROPERTY 2 (Heine-Borel). *Let* $\mathrm{A} \subset \mathrm{R}^3$ *be a compact set and* $\{\mathrm{U}_\alpha\}$ *be a family of open sets of* A *such that* $\bigcup_\alpha \mathrm{U}_\alpha \supset \mathrm{A}$. *Then it is possible to choose a finite number* $\mathrm{U}_{k_1}, \mathrm{U}_{k_2}, \dots, \mathrm{U}_{k_n}$ *of* U_α *such that* $\bigcup \mathrm{U}_{k_i} \supset \mathrm{A}$, $\mathrm{i} = 1, \dots, \mathrm{n}$.

PROPERTY 3 (Lebesgue). *Let* $\mathrm{A} \subset \mathrm{R}^3$ *be a compact set and* $\{\mathrm{U}_\alpha\}$ *a family of open sets of* A *such that* $\bigcup_\alpha \mathrm{U}_\alpha = \mathrm{A}$. *Then there exists a number* $\delta > 0$ (*the Lebesgue number of the family* $\{\mathrm{U}_\alpha\}$) *such that whenever two points* $\mathrm{p}, \mathrm{q} \in \mathrm{A}$ *are at a distance* $\mathrm{d}(\mathrm{p}, \mathrm{q}) < \delta$ *then* p *and* q *belong to some* U_α.

Properties 1 and 2 are usually proved in courses of advanced calculus. For completeness, we shall now prove Property 3. Later in this book (appendix to Chap. 5), we shall treat compact sets in R^n in a more systematic way and shall present proofs of Properties 1 and 2.

Proof of Property 3. Let us assume that there is no $\delta > 0$ satisfying the conditions in the statement; that is, given $1/n$ there exist points p_n and q_n such that $d(p_n, q_n) < 1/n$ but p_n and q_n do not belong to the same open set of the family $\{U_\alpha\}$. Setting $n = 1, 2, \dots,$ we obtain two infinite sets of points $\{p_n\}$ and $\{q_n\}$ which, by Property 1, have limit points p and q, respectively. Since $d(p_n, q_n) < 1/n$, we may choose these limit points in such a way that $p = q$. But $p \in U_\alpha$ for some α, because $p \in A = \bigcup_\alpha U_\alpha$, and since U_α is an open set, there is an open ball $B_\epsilon(p)$, with center in p, such that $B_\epsilon(p) \subset U_\alpha$. Since p is a limit point of $\{p_n\}$ and $\{q_n\}$, there exist, for n sufficiently large, points p_n and q_n in $B_\epsilon(p) \subset U_\alpha$; that is, p_n and q_n belong to the same U_α, which is a contradiction. **Q.E.D.**

Using Properties 2 and 3, we shall now prove the existence of a tubular neighborhood of an orientable compact surface.

PROPOSITION 3. *Let* $\mathrm{S} \subset \mathrm{R}^3$ *be a regular, compact, orientable surface. Then there exists a number* $\epsilon > 0$ *such that whenever* $\mathrm{p}, \mathrm{q} \in \mathrm{S}$ *the segments of the normal lines of length* 2ϵ, *centered in* p *and* q, *are disjoint* (*that is,* S *has a tubular neighborhood*).

Proof. By Prop. 1, for each $p \in S$ there exists a neighborhood W_p and a number $\epsilon_p > 0$ such that the proposition holds for points of W_p with $\epsilon = \epsilon_p$. Letting p run through S, we obtain a family $\{W_p\}$ with $\bigcup_{p \in S} W_p = S$. By compactness (Property 2), it is possible to choose a finite number of the W_p's, say, $W_1, \dots, W_k$ (corresponding to $\epsilon_1, \dots, \epsilon_k$) such that $\bigcup W_i = S$, $i = 1, \dots, k$. We shall show that the required ϵ is given by

$$\epsilon < \min\left(\epsilon_1, \dots, \epsilon_k, \frac{\delta}{2}\right),$$

where δ is the Lebesgue number of the family $\{W_i\}$ (Property 3).

In fact, let two points $p, q \in S$. If both belong to some W_i, $i = 1, \dots, k$, the segments of the normal lines with centers in p and q and length 2ϵ do not meet, since $\epsilon < \epsilon_i$. If p and q do not belong to the same W_i, then $d(p, q) \geq \delta$; were the segments of the normal lines, centered in p and q and of length 2ϵ, to meet at point $Q \in R^3$, we would have

$$2\epsilon \geq d(p, Q) + d(Q, q) \geq d(p, q) \geq \delta,$$

which contradicts the definition of ϵ. **Q.E.D.**

Putting together Props. 1, 2, and 3, we obtain the following theorem, which is the main goal of this section.

THEOREM. *Let* $S \subset R^3$ *be a regular compact orientable surface. Then there exists a differentiable function* $g\colon V \to R$*, defined in an open set* $V \subset R^3$*, with* $V \supset S$ *(precisely a tubular neighborhood of* S*), which has zero as a regular value and is such that* $S = g^{-1}(0)$.

Remark 1. It is possible to prove the existence of a tubular neighborhood of an orientable surface, even if the surface is not compact; the theorem is true, therefore, without the restriction of compactness. The proof is, however, more technical. In this general case, the $\epsilon(p) > 0$ is not constant as in the compact case but may vary with p.

Remark 2. It is possible to prove that a regular compact surface in R^3 is orientable; the hypothesis of orientability in the theorem (the compact case) is therefore unnecessary. A proof of this fact can be found in H. Samelson, "Orientability of Hypersurfaces in R^n," *Proc. A.M.S.* 22 (1969), 301–302.

2-8. A Geometric Definition of Area†

In this section we shall present a geometric justification for the definition of area given in Sec. 2-5. More precisely, we shall give a geometric definition of area and shall prove that in the case of a bounded region of a regular surface such a definition leads to the formula given for the area in Sec. 2-5.

To define the area of a region $R \subset S$ we shall start with a *partition* $\mathfrak{P}$ of R into a finite number of regions R_i, that is, we write $R = \bigcup_i R_i$, where the intersection of two such regions R_i is either empty or made up of boundary

†This section may be omitted on a first reading.

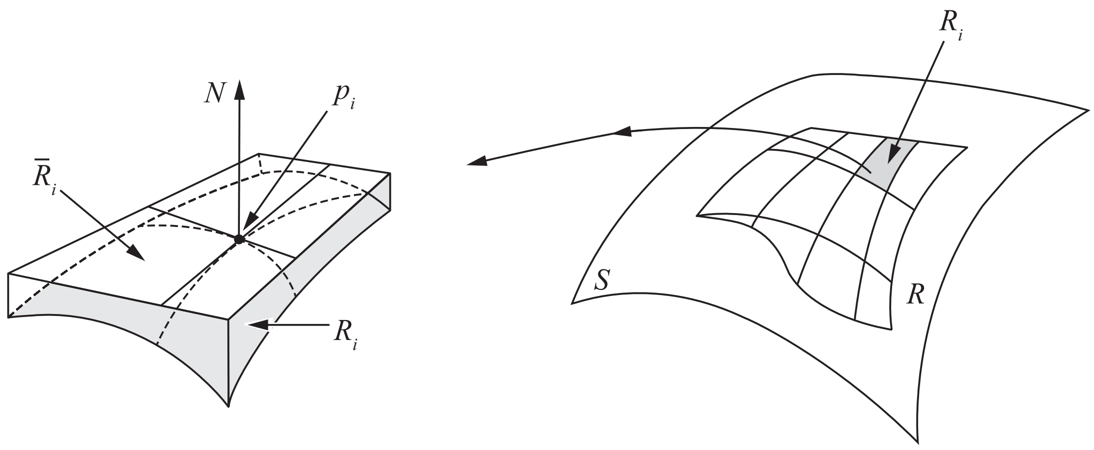


Figure 2-33

points of both regions (Fig. 2-33). The *diameter* of R_i is the supremum of the distances (in R^3) of any two points in R_i; the largest diameter of the R_i's of a given partition $\mathfrak{P}$ is called the *norm* μ of $\mathfrak{P}$. If we now take a partition of each R_i, we obtain a second partition of R, which is said to *refine* $\mathfrak{P}$.

Given a partition

$$R = \bigcup_i R_i$$

of R, we choose arbitrarily points $p_i \in R_i$ and project R_i onto the tangent plane at p_i in the direction of the normal line at p_i; this projection is denoted by $\bar{R}_i$ and its area by $A(\bar{R}_i)$. The sum $\sum_i A(\bar{R}_i)$ is an approximation of what we understand intuitively by the area of R.

If, by choosing partitions $\mathfrak{P}_1, \ldots, \mathfrak{P}_n, \ldots$ more and more refined and such that the norm μ_n of $\mathfrak{P}_n$ converges to zero, there exists a limit of $\sum_i A(\bar{R}_i)$ and this limit is independent of all choices, then we say that R has an *area* $A(R)$ defined by

$$A(R) = \lim_{\mu_n \to 0} \sum_i A(\bar{R}_i).$$

An instructive discussion of this definition can be found in R. Courant, *Differential and Integral Calculus*, Vol. II, Wiley-Interscience, New York, 1936, p. 311.

We shall show that a bounded region of a regular surface does have an area. We shall restrict ourselves to bounded regions contained in a coordinate neighborhood and shall obtain an expression for the area in terms of the coefficients of the first fundamental form in the corresponding coordinate system.

PROPOSITION. *Let* $\mathbf{x}\colon \mathrm{U} \to \mathrm{S}$ *be a coordinate system in a regular surface* S *and let* $\mathrm{R} = \mathbf{x}(\mathrm{Q})$ *be a bounded region of* S *contained in* $\mathbf{x}(\mathrm{U})$. *Then* R *has an area given by*

$$\mathrm{A}(\mathrm{R}) = \iint_{\mathrm{Q}} |\mathbf{x}_{\mathrm{u}} \wedge \mathbf{x}_{\mathrm{v}}|\, \mathrm{du}\, \mathrm{dv}.$$

Proof. Consider a partition, $R = \bigcup_i R_i$, of R. Since R is bounded and closed (hence, compact), we can assume that the partition is sufficiently refined so that any two normal lines of R_i are never orthogonal. In fact, because the normal lines vary continuously in S, there exists for each $p \in R$ a neighborhood of p in S where any two normals are never orthogonal; these neighborhoods constitute a family of open sets covering R, and considering a partition of R the norm of which is smaller than the Lebesgue number of the covering (Sec. 2-7, Property 3 of compact sets), we shall satisfy the required condition.

Fix a region R_i of the partition and choose a point $p_i \in R_i = \mathbf{x}(Q_i)$. We want to compute the area of the normal projection $\bar{R}_i$ of R_i onto the tangent plane at p_i. To do this, consider a new system of axes $p_i\bar{x}\bar{y}\bar{z}$ in R^3, obtained from $Oxyz$ by a translation Op_i, followed by a rotation which takes the z axis into the normal line at p_i in such a way that both systems have the same orientation (Fig. 2-34). In the new axes, the parametrization can be written

$$\bar{\mathbf{x}}(u, v) = (\bar{x}(u, v), \bar{y}(u, v), \bar{z}(u, v)),$$

where the explicit form of $\bar{\mathbf{x}}(u, v)$ does not interest us; it is enough to know that the vector $\bar{\mathbf{x}}(u, v)$ is obtained from the vector $\mathbf{x}(u, v)$ by a translation followed by an orthogonal linear map.

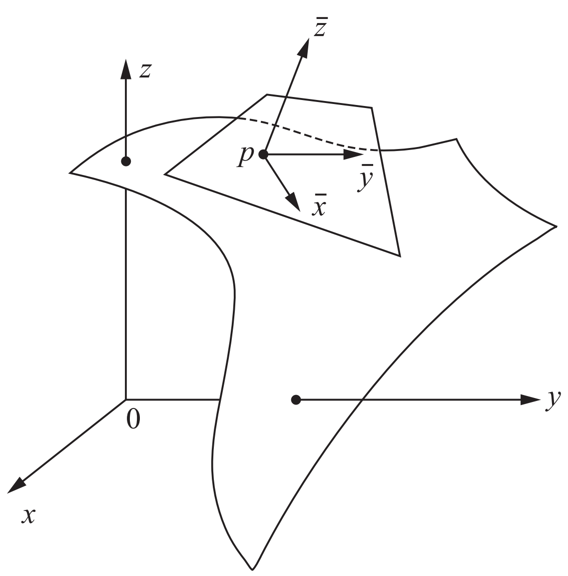


Figure 2-34

We observe that $\partial(\bar{x}, \bar{y})/\partial(u, v) \neq 0$ in Q_i; otherwise, the $\bar{z}$ component of some normal vector in R_i is zero and there are two orthogonal normal lines in R_i, a contradiction of our assumptions.

The expression of $A(\bar{R}_i)$ is given by

$$A(\bar{R}_i) = \iint_{\bar{R}_i} d\bar{x}\, d\bar{y}.$$

Since $\partial(\bar{x}, \bar{y})/\partial(u, v) \neq 0$, we can consider the change of coordinates $\bar{x} = \bar{x}(u, v)$, $\bar{y} = \bar{y}(u, v)$ and transform the above expression into

$$A(\bar{R}_i) = \iint_{Q_i} \frac{\partial(\bar{x}, \bar{y})}{\partial(u, v)}\, du\, dv.$$

We remark now that, at p_i, the vectors $\bar{\mathbf{x}}_u$ and $\bar{\mathbf{x}}_v$ belong to the $\bar{x}\bar{y}$ plane; therefore,

$$\frac{\partial \bar{z}}{\partial u} = \frac{\partial \bar{z}}{\partial v} = 0 \quad \text{at } p_i;$$

hence,

$$\left|\frac{\partial(\bar{x}, \bar{y})}{\partial(u, v)}\right| = \left|\frac{\partial \bar{\mathbf{x}}}{\partial u} \wedge \frac{\partial \bar{\mathbf{x}}}{\partial v}\right| \quad \text{at } p_i.$$

It follows that

$$\left|\frac{\partial(\bar{x}, \bar{y})}{\partial(u, v)}\right| - \left|\frac{\partial \bar{\mathbf{x}}}{\partial u} \wedge \frac{\partial \bar{\mathbf{x}}}{\partial v}\right| = \epsilon_i(u, v), \quad (u, v) \in Q_i,$$

where $\epsilon_i(u, v)$ is a continuous function in Q_i with $\epsilon_i(\mathbf{x}^{-1}(p_i)) = 0$. Since the length of a vector is preserved by translations and orthogonal linear maps, we obtain

$$\left|\frac{\partial \mathbf{x}}{\partial u} \wedge \frac{\partial \mathbf{x}}{\partial v}\right| = \left|\frac{\partial \bar{\mathbf{x}}}{\partial u} \wedge \frac{\partial \bar{\mathbf{x}}}{\partial v}\right| = \left|\frac{\partial(\bar{x}, \bar{y})}{\partial(u, v)}\right| - \epsilon_i(u, v).$$

Now let M_i and m_i be the maximum and the minimum of the continuous function $\epsilon_i(u, v)$ in the compact region Q_i; thus,

$$m_i \leq \left|\frac{\partial(\bar{x}, \bar{y})}{\partial(u, v)}\right| - \left|\frac{\partial \mathbf{x}}{\partial u} \wedge \frac{\partial \mathbf{x}}{\partial v}\right| \leq M_i;$$

hence,

$$m_i \iint_{Q_i} du\, dv \leq A(\bar{R}_i) - \iint_{Q_i} \left|\frac{\partial \mathbf{x}}{\partial u} \wedge \frac{\partial \mathbf{x}}{\partial v}\right| du\, dv \leq M_i \iint_{Q_i} du\, dv.$$

Doing the same for all R_i, we obtain

$$\sum_i m_i A(Q_i) \leq \sum_i A(\bar{R}_i) - \iint_Q |\mathbf{x}_u \wedge \mathbf{x}_v|\, du\, dv \leq \sum_i M_i A(Q_i).$$

Now, refine more and more the given partition in such a way that the norm $\mu \to 0$. Then $M_i \to m_i$. Therefore, there exists the limit of $\sum_i A(\bar{R}_i)$, given by

$$A(R) = \iint_Q \left|\frac{\partial \mathbf{x}}{\partial u} \wedge \frac{\partial \mathbf{x}}{\partial v}\right| du\, dv,$$

which is clearly independent of the choice of the partitions and of the point p_i in each partition. **Q.E.D.**

Appendix A Brief Review of Continuity and Differentiability

R^n will denote the set of n-tuples $(x_1, \ldots, x_n)$ of real numbers. Although we use only the cases $R^1 = R$, R^2, and R^3, the more general notion of R^n unifies the definitions and brings in no additional difficulties; the reader may think in R^2 or R^3, if he wishes so. In these particular cases, we shall use the following more traditional notation: x or t for R, (x, y) or (u, v) for R^2, and (x, y, z) for R^3.

A. Continuity in R^n

We start by making precise the notion of a point being ϵ-close to a given point $p_0 \in R^n$.

A *ball* (or *open ball*) in R^n with center $p_0 = (x_1^0, \ldots, x_n^0)$ and radius $\epsilon > 0$ is the set

$$B_\epsilon(p_0) = \{(x_1, \ldots, x_n) \in R^n; (x_1 - x_1^0)^2 + \cdots + (x_n - x_n^0)^2 < \epsilon^2\}.$$

Thus, in R, $B_\epsilon(p_0)$ is an open interval with center p_0 and length 2ϵ; in R^2, $B_\epsilon(p_0)$ is the interior of a disk with center p_0 and radius ϵ; in R^3, $B_\epsilon(p_0)$ is the interior of a region bounded by a sphere of center p_0 and radius ϵ (see Fig. A2-1).

A set $U \subset R^n$ is an *open set* if for each $p \in U$ there is a ball $B_\epsilon(p) \subset U$; intuitively this means that points in U are entirely surrounded by points of U, or that points sufficiently close to points of U still belong to U.

For instance, the set

$$\{(x, y) \in R^2; a < x < b, c < y < d\}$$

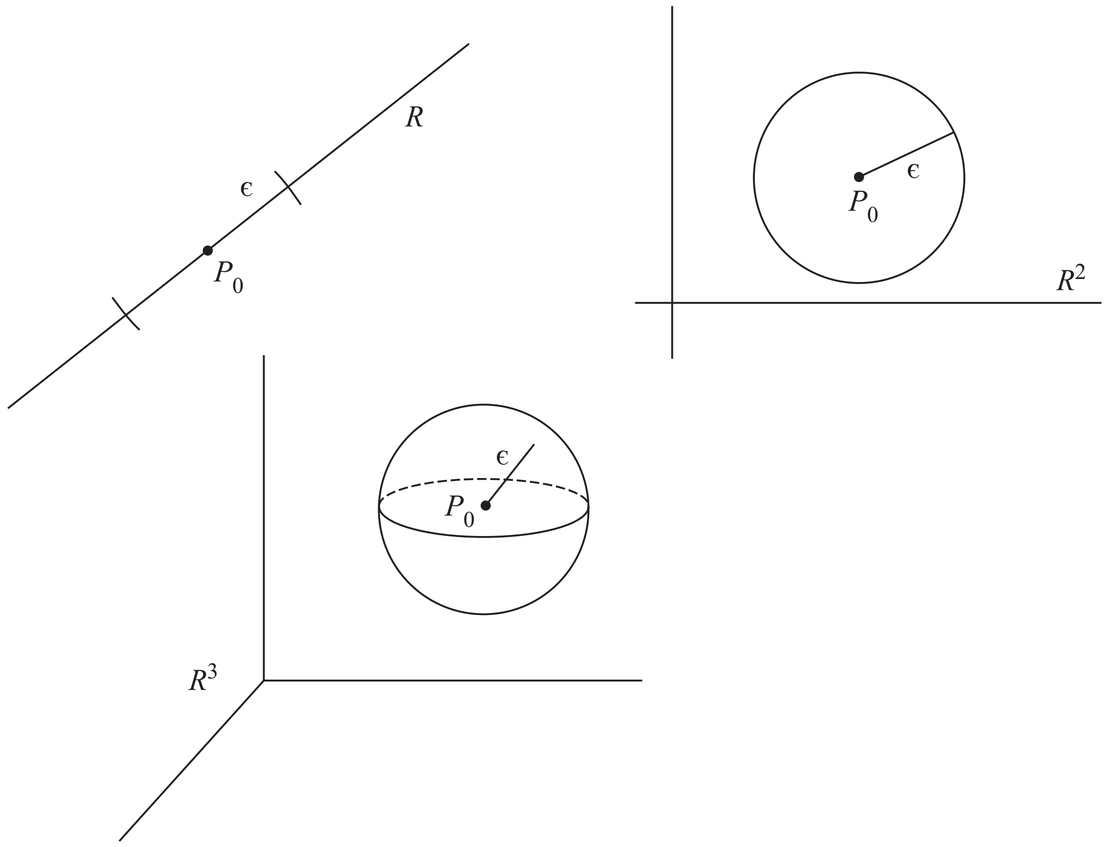


Figure A2-1

is easily seen to be open in R^2. However, if one of the strict inequalities, say $x < b$, is replaced by $x \leq b$, the set is no longer open; no ball with center at the point $(b, (d + c)/2)$, which belongs to the set, can be contained in the set (Fig. A2-2).

It is convenient to say that an open set in R^n containing a point $p \in R^n$ is a *neighborhood* of p.

From now on, $U \subset R^n$ will denote an open set in R^n.

We recall that a real function $f\colon U \subset R \to R$ of a real variable is continuous at $x_0 \in U$ if given an $\epsilon > 0$ there exists a $\delta > 0$ such that if $|x - x_0| < \delta$, then

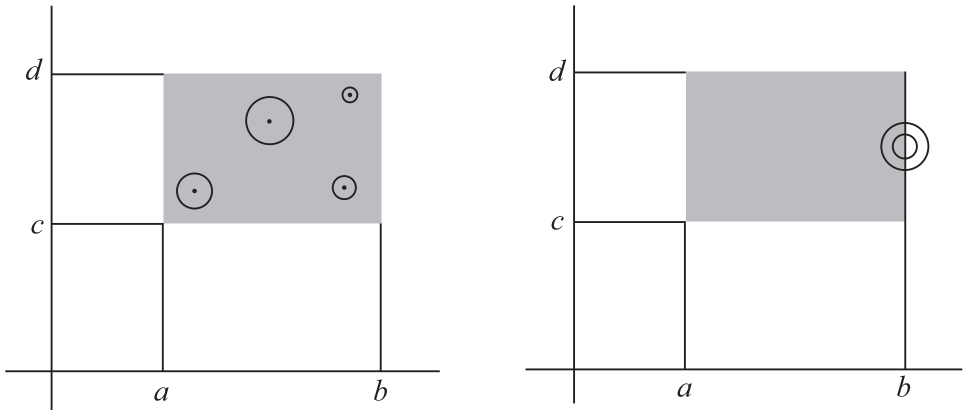


Figure A2-2

$$|f(x) - f(x_0)| < \epsilon.$$

Similarly, a real function $f: U \subset R^2 \to R$ of two real variables is continuous at $(x_0, y_0) \in U$ if given an $\epsilon > 0$ there exists a $\delta > 0$ such that if $(x - x_0)^2 + (y - y_0)^2 < \delta^2$, then

$$|f(x, y) - f(x_0, y_0)| < \epsilon.$$

The notion of ball unifies these definitions as particular cases of the following general concept:

A map $F: U \subset R^n \to R^m$ is *continuous* at $p \in U$ if given $\epsilon > 0$, there exists a $\delta > 0$ such that

$$F(B_\delta(p)) \subset B_\epsilon(F(p)).$$

In other words, F is continuous at p if points arbitrarily close to $F(p)$ are images of points sufficiently close to p. It is easily seen that in the particular cases of $n = 1, 2$ and $m = 1$, this agrees with the previous definitions. We say that F is *continuous* in U if F is *continuous* for all $p \in U$ (Fig. A2-3).

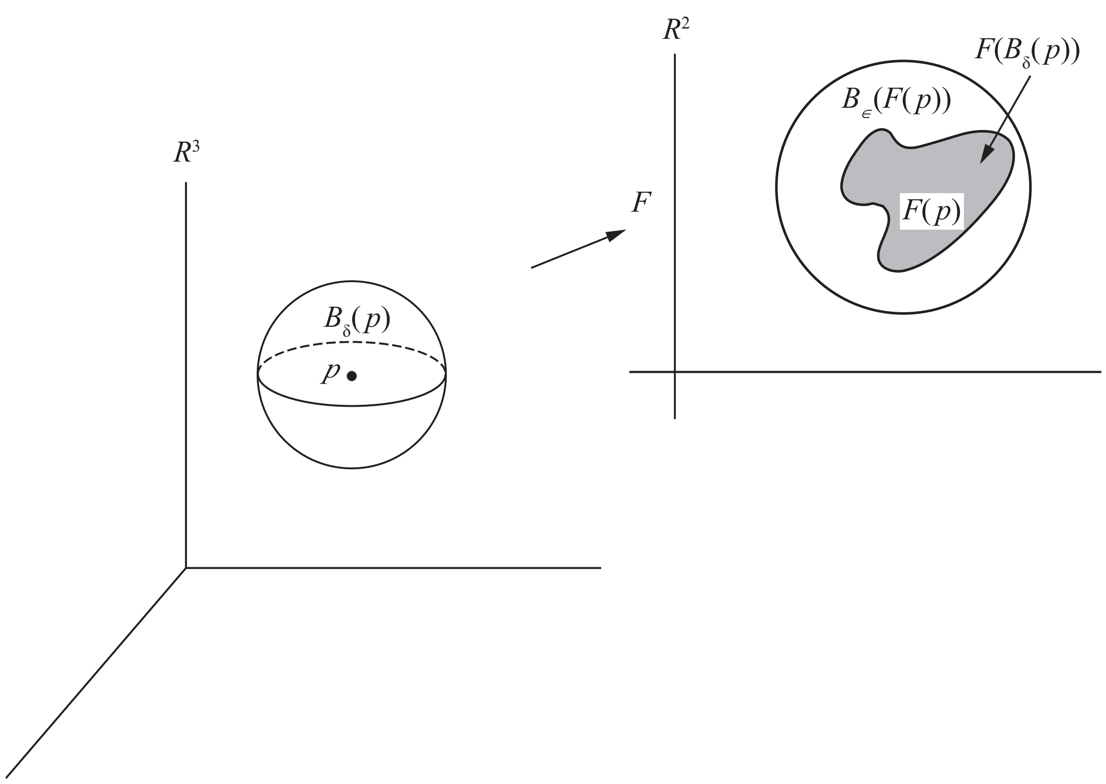


Figure A2-3

Given a map $F: U \subset R^n \to R^m$, we can determine m functions of n variables as follows. Let $p = (x_1, \ldots, x_n) \in U$ and $f(p) = (y_1, \ldots, y_m)$. Then we can write

$$y_1 = f_1(x_1, \ldots, x_n), \ldots, y_m = f_m(x_1, \ldots, x_n).$$

The functions $f_i: U \to R$, $i = 1, \ldots, m$, are the *component functions* of F.

Example 1 (*Symmetry*). Let $F\colon R^3 \to R^3$ be the map which assigns to each $p \in R^3$ the point which is symmetric to p with respect to the origin $O \in R^3$. Then $F(p) = -p$, or

$$F(x, y, z) = (-x, -y, -z),$$

and the component functions of F are

$$f_1(x, y, z) = -x, \quad f_2(x, y, z) = -y, \quad f_3(x, y, z) = -z.$$

Example 2 (*Inversion*). Let $F\colon R^2 - \{(0, 0)\} \to R^2$ be defined as follows. Denote by $|p|$ the distance to the origin $(0, 0) = O$ of a point $p \in R^2$. By definition, $F(p)$, $p \neq 0$, belongs to the half-line Op and is such that $|F(p)| \cdot |p| = 1$. Thus, $F(p) = p/|p|^2$, or

$$F(x, y) = \left(\frac{x}{x^2 + y^2}, \frac{y}{x^2 + y^2}\right), \quad (x, y) \neq (0, 0),$$

and the component functions of F are

$$f_1(x, y) = \frac{x}{x^2 + y^2}, \quad f_2(x, y) = \frac{y}{x^2 + y^2}.$$

Example 3 (*Projection*). Let $\pi\colon R^3 \to R^2$ be the projection $\pi(x, y, z) = (x, y)$. Then $f_1(x, y, z) = x$, $f_2(x, y, z) = y$.

The following proposition shows that the continuity of the map F is equivalent to the continuity of its component functions.

PROPOSITION 1. $\mathrm{F}\colon \mathrm{U} \subset \mathrm{R}^{\mathrm{n}} \to \mathrm{R}^{\mathrm{m}}$ *is continuous if and only if each component function* $\mathrm{f_i}\colon \mathrm{U} \subset \mathrm{R}^{\mathrm{n}} \to \mathrm{R}$, $\mathrm{i} = 1, \dots, \mathrm{m}$, *is continuous.*

Proof. Assume that F is continuous at $p \in U$. Then given $\epsilon > 0$, there exists $\delta > 0$ such that $F(B_\delta(p)) \subset B_\epsilon(F(p))$. Thus, if $q \in B_\delta(p)$, then

$$F(q) \in B_\epsilon(F(p)),$$

that is,

$$(f_1(q) - f_1(p))^2 + \cdots + (f_m(q) - f_m(p))^2 < \epsilon^2,$$

which implies that, for each $i = 1, \dots, m$, $|f_i(q) - f_i(p)| < \epsilon$. Therefore, given $\epsilon > 0$ there exists $\delta > 0$ such that if $q \in B_\delta(p)$, then $|f_i(q) - f_i(p)| < \epsilon$. Hence, each f_i is continuous at p.

Conversely, let f_i, $i = 1, \dots, m$, be continuous at p. Then given $\epsilon > 0$ there exists $\delta_i > 0$ such that if $q \in B_{\delta_i}(p)$, then $|f_i(q) - f_i(p)| < \epsilon/\sqrt{m}$. Set $\delta < \min \delta_i$ and let $q \in B_\delta(p)$. Then

$$(f_1(q) - f_1(p))^2 + \cdots + (f_m(q) - f_m(p))^2 < \epsilon^2,$$

and hence, the continuity of F at p. **Q.E.D.**

It follows that the maps in Examples 1, 2, and 3 are continuous.

Example 4. Let $F\colon U \subset R \to R^m$. Then

$$F(t) = (x_1(t), \dots, x_m(t)), \quad t \in U.$$

This is usually called a *vector-valued function*, and the component functions of F are the components of the vector $F(t) \in R^m$. When F is continuous, or, equivalently, the functions $x_i(t)$, $i = 1, \dots, m$, are continuous, we say that F is a *continuous curve* in R^m.

In most applications, it is convenient to express the continuity in terms of neighborhoods instead of balls.

PROPOSITION 2. *A map* $\mathrm{F}\colon \mathrm{U} \subset \mathrm{R}^n \to \mathrm{R}^m$ *is continuous at* $\mathrm{p} \in \mathrm{U}$ *if and only if, given a neighborhood* V *of* F(p) *in* R^m *there exists a neighborhood* W *of* p *in* R^n *such that* $\mathrm{F(W)} \subset \mathrm{V}$.

Proof. Assume that F is continuous at p. Since V is an open set containing $F(p)$, it contains a ball $B_\epsilon(F(p))$ for some $\epsilon > 0$. By continuity, there exists a ball $B_\delta(p) = W$ such that

$$F(W) = F(B_\delta(p)) \subset B_\epsilon(F(p)) \subset V,$$

and this proves that the condition is necessary.

Conversely, assume that the condition holds. Let $\varepsilon > 0$ be given and set $V = B_\epsilon(F(p))$. By hypothesis, there exists a neighborhood W of p in R^n such that $F(W) \subset V$. Since W is open, there exists a ball $B_\delta(p) \subset W$. Thus,

$$F(B_\delta(p)) \subset F(W) \subset V = B_\epsilon(F(p)),$$

and hence the continuity of F at p. **Q.E.D.**

The composition of continuous maps yields a continuous map. More precisely, we have the following proposition.

PROPOSITION 3. *Let* $\mathrm{F}\colon \mathrm{U} \subset \mathrm{R}^n \to \mathrm{R}^m$ *and* $\mathrm{G}\colon \mathrm{V} \subset \mathrm{R}^m \to \mathrm{R}^k$ *be continuous maps, where* U *and* V *are open set such that* $\mathrm{F(U)} \subset \mathrm{V}$. *Then* $\mathrm{G} \circ \mathrm{F}\colon \mathrm{U} \subset \mathrm{R}^n \to \mathrm{R}^k$ *is a continuous map.*

Proof. Let $p \in U$ and let W_1 be a neighborhood of $G \circ F(p)$ in R^k. By continuity of G, there is a neighborhood Q of $F(p)$ in R^m with $G(Q) \subset W_1$. By continuity of F, there is a neighborhood W_2 of (p) in R^n with $F(W_2) \subset Q$. Thus,

$$G \circ F(W_2) \subset G(Q) \subset W_1,$$

and hence the continuity of $G \circ F$. **Q.E.D.**

It is often necessary to deal with maps defined on arbitrary (not necessarily open) sets of R^n. To extend the previous ideas to this situation, we shall proceed as follows.

Let $F: A \subset R^n \to R^m$ be a map, where A is an arbitrary set in R^n. We say that F is continuous at $p \in A$ if given a neighborhood V of $F(p)$ in R^m, there exists a neighborhood W of p in R^n such that $F(W \cap A) \subset V$. For this reason, it is convenient to call $W \cap A$ a *neighborhood* of p in A. The set $B \subset A$ is *open* if, for each point $p \in B$, there exists a neighborhood of p in A entirely contained in B.

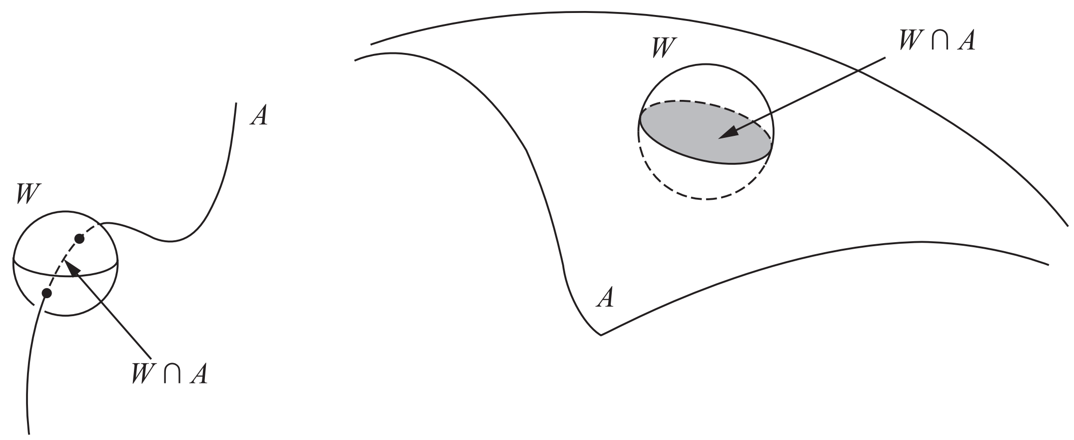


Figure A2-4

Example 5. Let

$$E = \left\{(x, y, z) \in R^3; \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1\right\}$$

be an ellipsoid, and let $\pi: R^3 \to R^2$ be the projection of Example 3. Then the restriction of π to E is a continuous map from E to R^2.

We say that a continuous map $F: A \subset R^n \to R^n$ is a *homeomorphism* onto $F(A)$ if F is one-to-one and the inverse $F^{-1}: F(A) \subset R^n \to R^n$ is continuous. In this case A and $F(A)$ are *homeomorphic sets*.

Example 6. Let $F: R^3 \to R^3$ be given by

$$F(x, y, z) = (xa, yb, zc).$$

F is clearly continuous, and the restriction of F to the sphere

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

is a continuous map $\tilde{F}: S^2 \to R^3$. Observe that $\tilde{F}(S^2) = E$, where E is the ellipsoid of Example 5. It is also clear that F is one-to-one and that

$$F^{-1}(x, y, z) = \left(\frac{x}{a}, \frac{y}{b}, \frac{z}{c}\right).$$

Thus, $\tilde{F}^{-1} = F^{-1}|E$ is continuous. Therefore, $\tilde{F}$ is a homeomorphism of the sphere S^2 onto the ellipsoid E.

Finally, we want to describe two properties of real continuous functions defined on a closed interval $[a, b]$,

$$[a, b] = \{x \in R; a \leq x \leq b\}$$

(Props. 4 and 5 below), and an important property of the closed interval $[a, b]$ itself. They will be used repeatedly in this book.

PROPOSITION 4 (The Intermediate Value Theorem). *Let* $f : [a, b] \to R$ *be a continuous function defined on the closed interval* $[a, b]$. *Assume that* $f(a)$ *and* $f(b)$ *have opposite signs; that is,* $f(a)f(b) < 0$. *Then there exists a point* $c \in (a, b)$ *such that* $f(c) = 0$.

PROPOSITION 5. *Let* $f : [a, b]$ *be a continuous function defined in the closed interval* $[a, b]$. *Then* f *reaches its maximum and its minimum in* $[a, b]$*; that is, there exist points* $x_1, x_2 \in [a, b]$ *such that* $f(x_1) < f(x) < f(x_2)$ *for all* $x \in [a, b]$.

PROPOSITION 6 (Heine-Borel). *Let* $[a, b]$ *be a closed interval and let* I_α*,* $\alpha \in A$*, be a collection of open intervals such that* $\bigcup_\alpha I_\alpha \supset [a, b]$. *Then it is possible to choose a finite number* $I_{k_1}, I_{k_2}, \ldots, I_{k_n}$ *of* I_α *such that* $\bigcup_{k_i} I_{k_i} \supset [a, b]$*,* $i = 1, \ldots, n$.

These propositions are standard theorems in courses on advanced calculus, and we shall not prove them here. However, proofs are provided in the appendix to Chap. 5 (Props. 8, 13, and 11, respectively).

B. Differentiability in R^n

Let $f : U \subset R \to R$. The *derivative* $f'(x_0)$ of f at $x_0 \in U$ is the limit (when it exists)

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}, \quad x_0 + h \in U.$$

When f has derivatives at all points of a neighborhood V of x_0, we can consider the derivative of $f' : V \to R$ at x_0, which is called the *second derivative* $f''(x_0)$ of f at x_0, and so forth. f is *differentiable* at x_0 if it has continuous derivatives of all orders at x_0. f is *differentiable* in U if it is differentiable at all points in U.

Remark. We use the word differentiable for what is sometimes called infinitely differentiable (or of class C^∞). Our usage should not be confused with the usage of elementary calculus, where a function is called differentiable if its first derivative exists.

Let F: $U \subset R^2 \to R$. The *partial derivative of* f *with respect to* x *at* $(x_0, y_0) \in U$, denoted by $(\partial f/\partial x)(x_0, y_0)$, is (when it exists) the derivative at x_0 of the function of one variable: $x \to f(x, y_0)$. Similarly, the partial derivative with respect to y at (x_0, y_0), $(\partial f/\partial y)(x_0, y_0)$, is defined as the derivative at y_0 of $y \to f(x_0, y)$. When f has partial derivatives at all points of a neighborhood V of (x_0, y_0), we can consider the *second partial derivatives at* (x_0, y_0):

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2}, \qquad \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y},$$
$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x}, \qquad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2},$$

and so forth. f is *differentiable* at (x_0, y_0) if it has continuous partial derivatives of all orders at (x_0, y_0). f is *differentiable* in U if it is differentiable at all points of U. We sometimes denote partial derivatives by

$$\frac{\partial f}{\partial x} = f_x, \quad \frac{\partial f}{\partial y} = f_y, \quad \frac{\partial^2 f}{\partial x^2} = f_{xx}, \quad \frac{\partial^2 f}{\partial x \partial y} = f_{xy}, \quad \frac{\partial^2 f}{\partial y^2} = f_{yy}.$$

It is an important fact that when f is differentiable the partial derivatives of f are independent of the order in which they are performed; that is,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}, \qquad \frac{\partial^3 f}{\partial^2 x \partial y} = \frac{\partial^3 f}{\partial x \partial y \partial x}, \qquad \text{etc.}$$

The definitions of partial derivatives and differentiability are easily extended to functions $f\colon U \subset R^n \to R$. For instance, $(\partial f/\partial x_3)(x_1^0, x_2^0, \ldots, x_n^0)$ is the derivative of the function of one variable

$$x_3 \to f(x_1^0, x_2^0, x_3, x_4^0, \ldots, x_n^0).$$

A further important fact is that partial derivatives obey the so-called *chain rule*. For instance, if $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ are real differentiable functions in $U \subset R^2$ and $f(x, y, z)$ is a real differentiable function in R^3, then the composition $f(x(u, v), y(u, v), z(u, v))$ is a differentiable function in U, and the partial derivative of f with respect to, say, u is given by

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial u}.$$

We are now interested in extending the notion of differentiability to maps $F\colon U \subset R^n \to R^m$. We say that F is *differentiable* at $p \in U$ if its component functions are differentiable at p; that is, by writing

$$F(x_1, \dots, x_n) = (f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)),$$

the functions f_i, $i = 1, \dots, m$, have continuous partial derivatives of all orders at p. F is *differentiable* in U if it is differentiable at all points in U.

For the case $m = 1$, this repeats the previous definition. For the case $n = 1$, we obtain the notion of a (parametrized) *differentiable curve* in R^m. In Chap. 1, we have already seen such an object in R^3. For our purposes, we need to extend the definition of tangent vector of Chap. 1 to the present situation. A *tangent vector* to a map $\alpha\colon U \subset R \to R^m$ at $t_0 \in U$ is the vector in R^m

$$\alpha'(t_0) = (x_1'(t_0), \dots, x_m'(t_0)).$$

Example 7. Let $F\colon U \subset R^2 \to R^3$ be given by

$$F(u, v) = (\cos u \cos v, \cos u \sin v, \cos^2 v), \quad (u, v) \in U.$$

The component functions of F, namely,

$$f_1(u, v) = \cos u \cos v, \quad f_2(u, v) = \cos u \sin v, \quad f_3(u, v) = \cos^2 v$$

have continuous partial derivatives of all orders in U. Thus, F is differentiable in U.

Example 8. Let $\alpha\colon U \subset R \to R^4$ be given by

$$\alpha(t) = (t^4, t^3, t^2, t), \quad t \in U.$$

Then α is a differentiable curve in R^4, and the tangent vector to α at t is $\alpha'(t) = (4t^3, 3t^2, 2t, 1)$.

Example 9. Given a vector $w \in R^m$ and a point $p_0 \in U \subset R^m$, we can always find a differentiable curve $\alpha\colon (-\epsilon, \epsilon) \to U$ with $\alpha(0) = p_0$ and $\alpha'(0) = w$. Simply define $\alpha(t) = p_0 + tw$, $t \in (-\epsilon, \epsilon)$. By writing $p_0 = (x_1^0, \dots, x_m^0)$ and $w = (w_1, \dots, w_m)$, the component functions of α are $x_i(t) = x_i^0 + tw_i$, $i = 1, \dots, m$. Thus, α is differentiable, $\alpha(0) = p_0$ and

$$\alpha'(0) = (x_1'(0), \dots, x_m'(0)) = (w_1, \dots, w_m) = w.$$

We shall now introduce the concept of differential of a differentiable map. It will play an important role in this book.

DEFINITION 1. *Let* $\mathrm{F}\colon \mathrm{U} \subset \mathrm{R^n} \to \mathrm{R^m}$ *be a differentiable map. To each* $\mathrm{p} \in \mathrm{U}$ *we associate a linear map* $\mathrm{dF_p}\colon \mathrm{R^n} \to \mathrm{R^m}$ *which is called the* differential *of* F *at* p *and is defined as follows. Let* $\mathrm{w} \in \mathrm{R^n}$ *and let* $\alpha\colon (-\epsilon, \epsilon) \to \mathrm{U}$

be a differentiable curve such that $\alpha(0) = \mathrm{p}$, $\alpha'(0) = \mathrm{w}$. *By the chain rule, the curve* $\beta = \mathrm{F} \circ \alpha: (-\epsilon, \epsilon) \to \mathrm{R}^{\mathrm{m}}$ *is also differentiable.* Then (Fig. A2-5)

$$\mathrm{dF_p(w)} = \beta'(0).$$

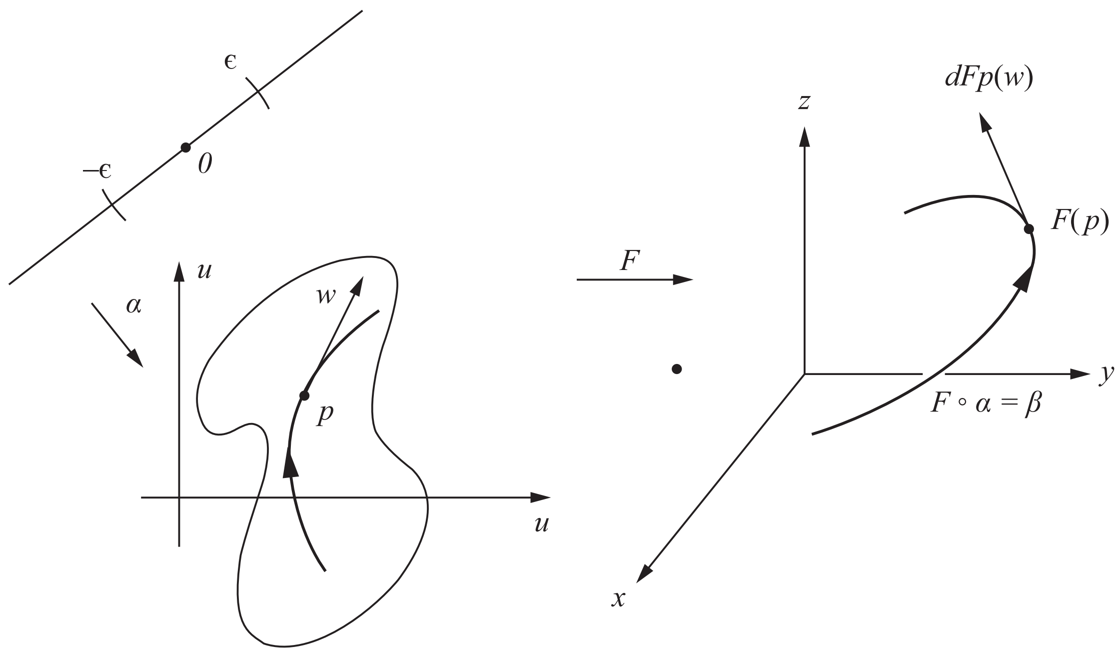


Figure A2-5

PROPOSITION 7. *The above definition of* $\mathrm{dF_p}$ *does not depend on the choice of the curve which passes through* p *with tangent vector* w, *and* $\mathrm{dF_p}$ *is, in fact, a linear map.*

Proof. To simplify notation, we work with the case $F: U \subset R^2 \to R^3$. Let (u, v) be coordinates in R^2 and (x, y, z) be coordinates in R^3. Let $e_1 = (1, 0)$, $e_2 = (0, 1)$ be the canonical basis in R^2 and $f_1 = (1, 0, 0)$, $f_2 = (0, 1, 0)$, $f_3 = (0, 0, 1)$ be the canonical basis in R^3. Then we can write $\alpha(t) = (u(t), v(t))$, $t \in (-\epsilon, \epsilon)$,

$$\alpha'(0) = w = u'(0)e_1 + v'(0)e_2,$$

$F(u, v) = (x(u, v), y(u, v), z(u, v))$, and

$$\beta(t) = F \circ \alpha(t) = (x(u(t), v(t)), y(u(t), v(t)), z(u(t), v(t))).$$

Thus, using the chain rule and taking the derivatives at $t = 0$, we obtain

$$\begin{aligned}\beta'(0) &= \left(\frac{\partial x}{\partial u}\frac{du}{dt} + \frac{\partial x}{\partial v}\frac{dv}{dt}\right) f_1 + \left(\frac{\partial y}{\partial u}\frac{du}{dt} + \frac{\partial y}{\partial v}\frac{dv}{dt}\right) f_2 \\ &\quad + \left(\frac{\partial z}{\partial u}\frac{du}{dt} + \frac{\partial z}{\partial v}\frac{dv}{dt}\right) f_3 \\ &= \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \\ \dfrac{\partial z}{\partial u} & \dfrac{\partial z}{\partial v} \end{pmatrix} \begin{pmatrix} \dfrac{du}{dt} \\ \\ \dfrac{dv}{dt} \end{pmatrix} = dF_p(w).\end{aligned}$$

This shows that dF_p is represented, in the canonical bases of R^2 and R^3, by a matrix which depends only on the partial derivatives at p of the component functions x, y, z of F. Thus, dF_p is a linear map, and clearly $dF_p(w)$ does not depend on the choice of α.

The reader will have no trouble in extending this argument to the more general situation. **Q.E.D.**

The matrix of dF_p: $R^n \to R^m$ in the canonical bases of R^n and R^m, that is, the matrix $(\partial f_i/\partial x_j)$, $i = 1, \dots, m$, $j = 1, \dots, n$, is called the *Jacobian matrix* of F at p. When $n = m$, this is a square matrix and its determinant is called the *Jacobian determinant*; it is usual to denote it by

$$\det\left(\frac{\partial f_i}{\partial x_j}\right) = \frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)}.$$

Remark. There is no agreement in the literature regarding the notation for the differential. It is also of common usage to call dF_p the derivative of F at p and to denote it by $F'(p)$.

Example 10. Let F: $R^2 \to R^2$ be given by

$$F(x, y) = (x^2 - y^2, 2xy), \quad (x, y) \in R^2.$$

F is easily seen to be differentiable, and its differential dF_p at $p = (x, y)$ is

$$dF_p = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}.$$

For instance, $dF_{(1,1)}(2, 3) = (-2, 10)$.

One of the advantages of the notion of differential of a map is that it allows us to express many facts of calculus in a geometric language. Consider, for instance, the following situation: Let $F\colon U \subset R^2 \to R^3$, $G\colon V \subset R^3 \to R^2$ be differentiable maps, where U and V are open sets such that $F(U) \subset V$. Let us agree on the following set of coordinates,

$$\underset{(u,v)}{U \subset R^2} \xrightarrow{F} \underset{(x,y,z)}{V \subset R^3} \xrightarrow{G} \underset{(\xi,\eta)}{R^2}$$

and let us write

$$F(u,v) = (x(u,v), y(u,v), z(u,v)),$$
$$G(x,y,z) = (\xi(x,y,z), \eta(x,y,z)).$$

Then

$$G \circ F(u,v) = (\xi(x(u,v), y(u,v), z(u,v)), \eta(x(u,v), y(u,v), z(u,v))),$$

and, by the chain rule, we can say that $G \circ F$ is differentiable and compute the partial derivatives of its component functions. For instance,

$$\frac{\partial \xi}{\partial u} = \frac{\partial \xi}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial \xi}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial \xi}{\partial z}\frac{\partial z}{\partial u}.$$

Now, a simple way of expressing the above situation is by using the following general fact.

PROPOSITION 8 (The Chain Rule for Maps). *Let* $\mathrm{F}\colon \mathrm{U} \subset \mathrm{R}^{\mathrm{n}} \to \mathrm{R}^{\mathrm{m}}$ *and* $\mathrm{G}\colon \mathrm{V} \subset \mathrm{R}^{\mathrm{m}} \to \mathrm{R}^{\mathrm{k}}$ *be differentiable maps, where* U *and* V *are open sets such that* $\mathrm{F(U)} \subset \mathrm{V}$. *Then* $\mathrm{G} \circ \mathrm{F}\colon \mathrm{U} \to \mathrm{R}^{\mathrm{k}}$ *is a differentiable map, and*

$$\mathrm{d(G \circ F)_p = dG_{F(p)} \circ dF_p}, \quad \mathrm{p \in U}.$$

Proof. The fact that $G \circ F$ is differentiable is a consequence of the chain rule for functions. Now, let $w_1 \in R^n$ be given and let us consider a curve $\alpha\colon (-\epsilon_2, \epsilon_2) \to U$, with $\alpha(0) = p$, $\alpha'(0) = w_1$. Set $dF_p(w_1) = w_2$ and observe that $dG_{F(p)}(w_2) = (d/dt)(G \circ F \circ \alpha)|_{t=0}$. Then

$$d(G \circ F)_p(w_1) = \frac{d}{dt}(G \circ F \circ \alpha)_{t=0} = dG_{F(p)}(w_2) = dG_{F(p)} \circ dF_p(w_1).$$

Q.E.D.

Notice that, for the particular situation we were considering before, the relation $d(G \circ F)_p = dG_{F(p)} \circ dF_p$ is equivalent to the following product of Jacobian matrices,

$$\begin{pmatrix} \dfrac{\partial \xi}{\partial u} & \dfrac{\partial \xi}{\partial v} \\ \\ \dfrac{\partial \eta}{\partial u} & \dfrac{\partial \eta}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \xi}{\partial x} & \dfrac{\partial \xi}{\partial y} & \dfrac{\partial \xi}{\partial z} \\ \\ \dfrac{\partial \eta}{\partial x} & \dfrac{\partial \eta}{\partial y} & \dfrac{\partial \eta}{\partial z} \end{pmatrix} \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \\ \dfrac{\partial z}{\partial u} & \dfrac{\partial z}{\partial v} \end{pmatrix},$$

which contains the expressions of all partial derivatives $\partial \xi/\partial u$, $\partial \xi/\partial v$, $\partial \eta/\partial u$, $\partial \eta/\partial v$. Thus, the simple expression of the chain rule for maps embodies a great deal of information on the partial derivatives of their component functions.

An important property of a differentiable function $f\colon (a, b) \subset R \to R$ defined in an open interval (a, b) is that if $f'(x) \equiv 0$ on (a, b), then f is constant on (a, b). This generalizes for differentiable functions of several variables as follows.

We say that an open set $U \subset R^n$ is *connected* if give two points $p, q \in U$ there exists a continuous map $\alpha\colon [a, b] \to U$ such that $\alpha(a) = p$ and $\alpha(b) = q$. This means that two points of U can be joined by a continuous curve in U or that U is made up of one single "piece."

PROPOSITION 9. *Let* $\mathrm{f}\colon \mathrm{U} \subset \mathrm{R}^{\mathrm{n}} \to \mathrm{R}$ *be a differentiable function defined on a connected open subset* U *of* R^{n}. *Assume that* $\mathrm{df}_{\mathrm{p}}\colon \mathrm{R}^{\mathrm{n}} \to \mathrm{R}$ *is zero at every point* $\mathrm{p} \in \mathrm{U}$. *Then* f *is constant on* U.

Proof. Let $p \in U$ and let $B_\delta(p) \subset U$ be an open ball around p and contained in U. Any point $q \in B_\epsilon(p)$ can be joined to p by the "radial" segment $\beta\colon [0, 1] \to U$, where $\beta(t) = tq + (1-t)p$, $t \in [0, 1]$ (Fig. A2-6). Since U is open, we can extend β to $(0-\epsilon, 1+\epsilon)$. Now, $f \circ \beta\colon (0-\epsilon, 1+\epsilon) \to R$ is a function defined in an open interval, and

$$d(f \circ \beta)_t = (df \circ d\beta)_t = 0,$$

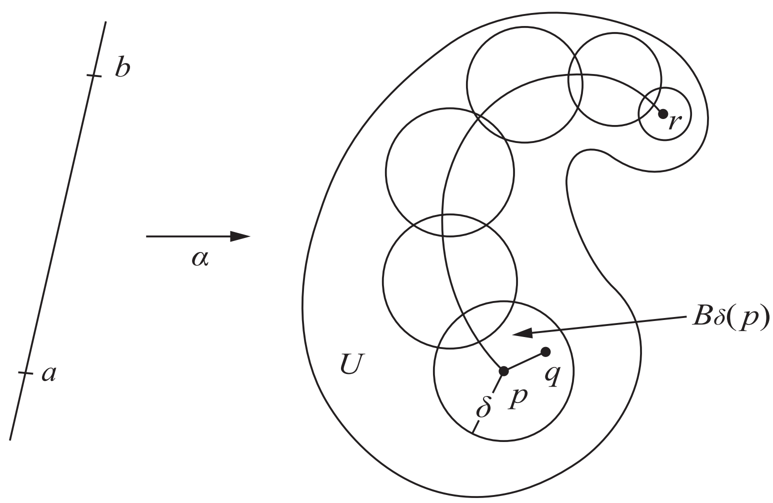


Figure A2-6

since $df \equiv 0$. Thus,

$$\frac{d}{dt}(f \circ \beta) = 0$$

for all $t \in (0-\epsilon, 1+\epsilon)$, and hence $(f \circ \beta) = \text{const}$. This means that $f(\beta(0)) = f(p) = f(\beta(1)) = f(q)$; that is, f is constant on $B_\delta(p)$.

Thus, the proposition is proved locally; that is, each point of U has a neighborhood such that f is constant on that neighborhood. Notice that so far we have not used the connectedness of U. We shall need it now to show that these constants are all the same.

Let r be an arbitrary point of U. Since U is connected, there exists a continuous curve α: $[a, b] \to U$, with $\alpha(a) = p$, $\alpha(b) = r$. The function $f \circ \alpha$: $[a, b] \to R$ is continuous in $[a, b]$. By the first part of the proof, for each $t \in [a, b]$, there exists an interval I_t, open in $[a, b]$, such that $f \circ \alpha$ is constant on I_t. Since $\bigcup_t I_t = [a, b]$, we can apply the Heine-Borel theorem (Prop. 6). Thus, we can choose a finite number $I_1, \dots, I_k$ of the intervals I_t so that $\bigcup_i I_i = [a, b]$, $i = 1, \dots, k$. We can assume, by renumbering the intervals, if necessary, that two consecutive intervals overlap. Thus, $f \circ \alpha$ is constant in the union of two consecutive intervals. It follows that f is constant on $[a, b]$; that is,

$$f(\alpha(a)) = f(p) = f(\alpha(b)) = f(r).$$

Since r is arbitrary, f is constant on U. **Q.E.D.**

One of the most important theorems of differential calculus is the so-called inverse function theorem, which, in the present notation, says the following. (Recall that a linear map A is an isomorphism if the matrix of A is invertible.)

INVERSE FUNCTION THEOREM. *Let* $F\colon U \subset R^n \to R^n$ *be a differentiable mapping and suppose that at* $p \in U$ *the differential* $dF_p\colon R^n \to R^n$ *is an isomorphism. Then there exists a neighborhood* V *of* p *in* U *and a neighborhood* W *of* $F(p)$ *in* R^n *such that* $F\colon V \to W$ *has a differentiable inverse* $F^{-1}\colon W \to V$.

A differentiable mapping F: $V \subset R^n \to W \subset R^n$, where V and W are open sets, is called a *diffeomorphism* of V with W if F has a differentiable inverse. The inverse function theorem asserts that if at a point $p \in U$ the differential dF_p is an isomorphism, then F is a diffeomorphism in a neighborhood of p. In other words, an assertion about the differential of F at a point implies a similar assertion about the behavior of F in a neighborhood of the point.

This theorem will be used repeatedly in this book. A proof can be found, for instance, in Buck, *Advanced Calculus*, p. 285.

Example 11. Let F: $R^2 \to R^2$ be given by

$$F(x, y) = (e^x \cos y, e^x \sin y), \quad (x, y) \in R^2.$$

The component functions of F, namely, $u(x, y) = e^x \cos y$, $v(x, y) = e^x \sin y$, have continuous partial derivatives of all orders. Thus, F is differentiable.

It is instructive to see, geometrically, how F transforms curves of the xy plane. For instance, the vertical line $x = x_0$ is mapped into the circle $u = e^{x_0} \cos y$, $v = e^{x_0} \sin y$ of radius e^{x_0}, and the horizontal line $y = y_0$ is mapped into the half-line $u = e^x \cos y_0$, $v = e^x \sin y_0$ with slope $\tan y_0$. It follows that (Fig. A2-7)

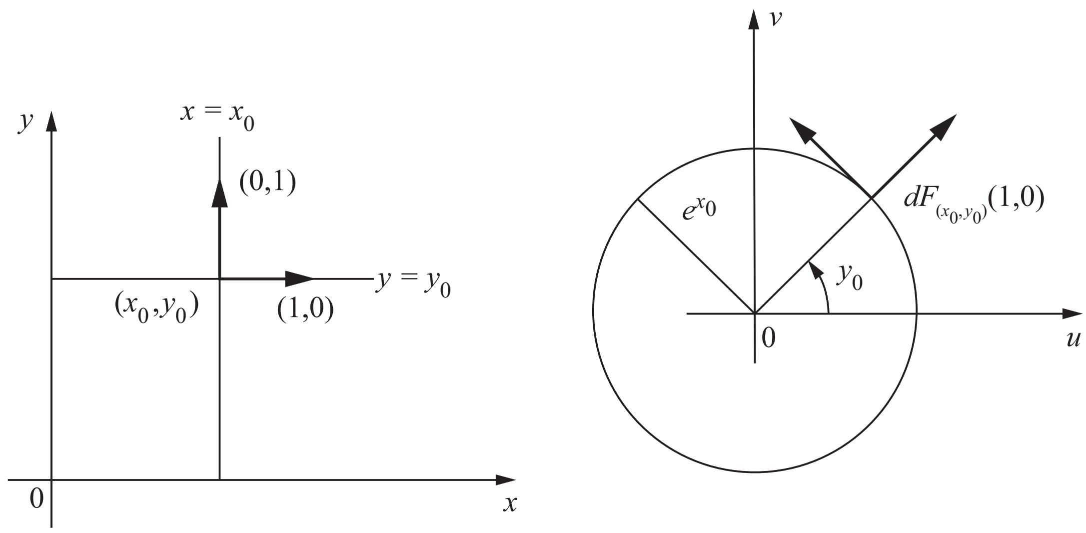


Figure A2-7

$$
\begin{aligned}
dF_{(x_0,y_0)}(1, 0) &= \frac{d}{dx}(e^x \cos y_0, e^x \sin y_0)|_{x=x_0} \\
&= (e^{x_0} \cos y_0, e^{x_0} \sin y_0), \\
dF_{(x_0,y_0)}(0, 1) &= \frac{d}{dy}(e^{x_0} \cos y, e^{x_0} \sin y)|_{y=y_0} \\
&= (-e^{x_0} \sin y_0, e^{x_0} \cos y_0).
\end{aligned}
$$

This can be most easily checked by computing the Jacobian matrix of F,

$$
dF_{(x,y)} = \begin{pmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{pmatrix} = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix},
$$

and applying it to the vectors $(1, 0)$ and $(0, 1)$ at (x_0, y_0).

We notice that the Jacobian determinant $\det(dF_{(x,y)}) = e^x \neq 0$, and thus dF_p is nonsingular for all $p = (x, y) \in R^2$ (this is also clear from the previous geometric considerations). Therefore, we can apply the inverse function theorem to conclude that F is locally a diffeomorphism.

Observe that $F(x, y) = F(x, y + 2\pi)$. Thus, F is not one-to-one and has no global inverse. For each $p \in R^2$, the inverse function theorem gives neighborhoods V of p and W of $F(p)$ so that the restriction $F\colon V \to W$ is a diffeomorphism. In our case, V may be taken as the strip $\{-\infty < x < \infty, 0 < y < 2\pi\}$ and W as $R^2 - \{(0, 0)\}$. However, as the example shows, even if the conditions of the theorem are satisfied everywhere and the domain of definition of F is very simple, a global inverse of F may fail to exist.

3 The Geometry of the Gauss Map

3-1. Introduction

As we have seen in Chap. 1, the consideration of the rate of change of the tangent line to a curve C led us to an important geometric entity, namely, the curvature of C. In this chapter we shall extend this idea to regular surfaces; that is, we shall try to measure how rapidly a surface S pulls away from the tangent plane $T_p(S)$ in a neighborhood of a point $p \in S$. This is equivalent to measuring the rate of change at p of a unit normal vector field N on a neighborhood of p. As we shall see shortly, this rate of change is given by a linear map on $T_p(S)$ which happens to be self-adjoint (see the appendix to Chap. 3). A surprisingly large number of local properties of S at p can be derived from the study of this linear map.

In Sec. 3-2, we shall introduce the relevant definitions (the Gauss map, principal curvatures and principal directions, Gaussian and mean curvatures, etc.) without using local coordinates. In this way, the geometric content of the definitions is clearly brought up. However, for computational as well as for theoretical purposes, it is important to express all concepts in local coordinates. This is taken up in Sec. 3-3.

Sections 3-2 and 3-3 contain most of the material of Chap. 3 that will be used in the remaining parts of this book. The few exceptions will be explicitly pointed out. For completeness, we have proved the main properties of self-adjoint linear maps in the appendix to Chap. 3. Furthermore, for those who have omitted Sec. 2-6, we have included a brief review of orientation for surfaces at the beginning of Sec. 3-2.

Section 3-4 contains a proof of the fact that at each point of a regular surface there exists an orthogonal parametrization, that is, a parametrization such that its coordinate curves meet orthogonally. The techniques used here are interesting in their own right and yield further results. However, for a short course it might be convenient to assume these results and omit the section.

In Sec. 3-5 we shall take up two interesting special cases of surfaces, namely, the ruled surfaces and the minimal surfaces. They are treated independently so that one (or both) of them can be omitted on a first reading.

3-2. The Definition of the Gauss Map and Its Fundamental Properties

We shall begin by briefly reviewing the notion of orientation for surfaces.

As we have seen in Sec. 2-4, given a parametrization $\mathbf{x}\colon U \subset R^2 \to S$ of a regular surface S at a point $p \in S$, we can choose a unit normal vector at each point of $\mathbf{x}(U)$ by the rule

$$N(q) = \frac{\mathbf{x}_u \wedge \mathbf{x}_v}{|\mathbf{x}_u \wedge \mathbf{x}_v|}(q), \quad q \in \mathbf{x}(U).$$

Thus, we have a differentiable map $N\colon \mathbf{x}(U) \to R^3$ that associates to each $q \in \mathbf{x}(U)$ a unit normal vector $N(q)$.

More generally, if $V \subset S$ is an open set in S and $N\colon V \to R^3$ is a differentiable map which associates to each $q \in V$ a unit normal vector at q, we say that *N is a differentiable field of unit normal vectors on V.*

It is a striking fact that not all surfaces admit a differentiable field of unit normal vectors *defined on the whole surface*. For instance, on the Möbius strip of Fig. 3-1 one cannot define such a field. This can be seen intuitively by

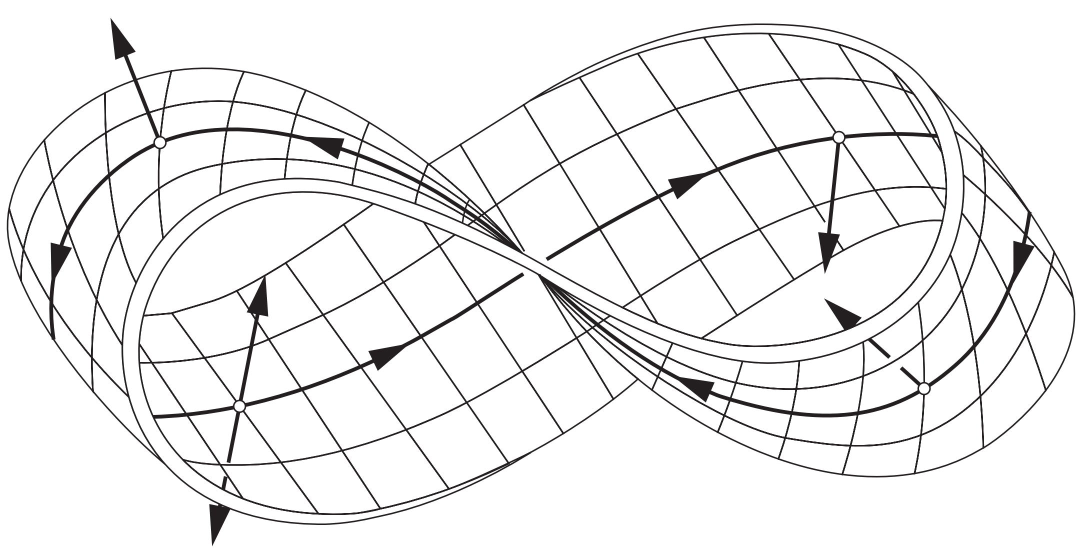


Figure 3-1. The Möbius strip.

going around once along the middle circle of the figure: After one turn, the vector field N would come back as $-N$, a contradiction to the continuity of N. Intuitively, one cannot, on the Möbius strip, make a consistent choice of a definite "side"; moving around the surface, we can go continuously to the "other side" without leaving the surface.

We shall say that a regular surface is *orientable* if it admits a differentiable field of unit normal vectors defined on the whole surface; the choice of such a field N is called an *orientation* of S.

For instance, the Möbius strip referred to above is not an orientable surface. Of course, every surface covered by a single coordinate system (for instance, surfaces represented by graphs of differentiable functions) is trivially orientable. Thus, every surface is locally orientable, and orientation is definitely a global property in the sense that it involves the whole surface.

An orientation N on S induces an orientation on each tangent space $T_p(S)$, $p \in S$, as follows. Define a basis $\{v, w\} \subset T_p(S)$ to be *positive* if $\langle v \wedge w, N\rangle$ is positive. It is easily seen that the set of all positive bases of $T_p(S)$ is an orientation for $T_p(S)$ (cf. Sec. 1-4).

Further details on the notion of orientation are given in Sec. 2-6. However, for the purpose of Chaps. 3 and 4, the present description will suffice.

Throughout this chapter, S will denote a regular orientable surface in which an orientation (i.e., a differentiable field of unit normal vectors N) has been chosen; this will be simply called a surface S with an orientation N.

DEFINITION 1. *Let* $\mathrm{S} \subset \mathrm{R}^3$ *be a surface with an orientation* N. *The map* $\mathrm{N}: \mathrm{S} \to \mathrm{R}^3$ *takes its values in the unit sphere*

$$\mathrm{S}^2 = \{(\mathrm{x}, \mathrm{y}, \mathrm{z}) \in \mathrm{R}^3; \mathrm{x}^2 + \mathrm{y}^2 + \mathrm{z}^2 = 1\}$$

The map $\mathrm{N}: \mathrm{S} \to \mathrm{S}^2$*, thus defined, is called the* Gauss map *of* S *(Fig. 3-2).*†

It is straightforward to verify that the Gauss map is differentiable. The differential dN_p of N at $p \in S$ is a linear map from $T_p(S)$ to $T_{N(p)}(S^2)$. Since $T_p(S)$ and $T_{N(p)}(S^2)$ are the same vector spaces, dN_p can be looked upon as a linear map on $T_p(S)$.

The linear map dN_p: $T_p(S) \to T_p(S)$ operates as follows. For each parametrized curve $\alpha(t)$ in S with $\alpha(0) = p$, we consider the parametrized curve $N \circ \alpha(t) = N(t)$ in the sphere S^2; this amounts to restricting the normal vector N to the curve $\alpha(t)$. The tangent vector $N'(0) = dN_p(\alpha'(0))$ is a vector in $T_p(S)$ (Fig. 3-3). It measures the rate of change of the normal vector N, restricted to the curve $\alpha(t)$, at $t = 0$. Thus, dN_p measures how N pulls

†In italic context, letter symbols set in roman rather than italics.

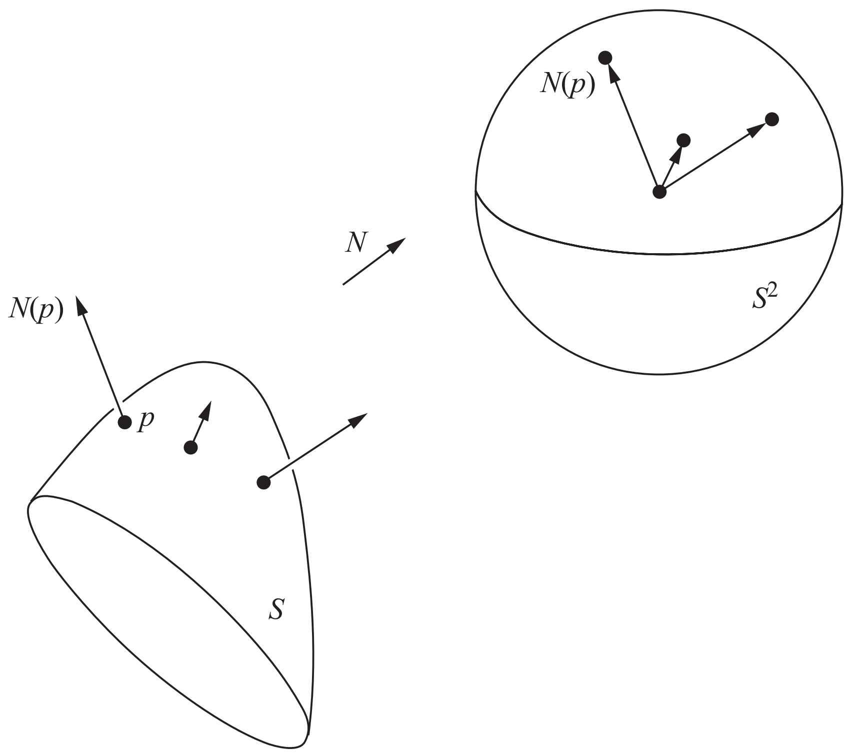


Figure 3-2. The Gauss map.

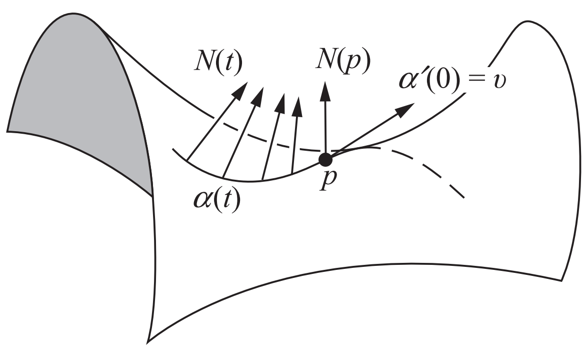


Figure 3-3

away from $N(p)$ in a neighborhood of p. In the case of curves, this measure is given by a number, the curvature. In the case of surfaces, this measure is characterized by a linear map.

Example 1. For a plane P given by $ax + by + cz + d = 0$, the unit normal vector $N = (a, b, c)/\sqrt{a^2 + b^2 + c^2}$ is constant, and therefore $dN \equiv 0$ (Fig. 3-4).

Example 2. Consider the unit sphere

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

If $\alpha(t) = (x(t), y(t), z(t))$ is a parametrized curve in S^2, then

$$2xx' + 2yy' + 2zz' = 0,$$

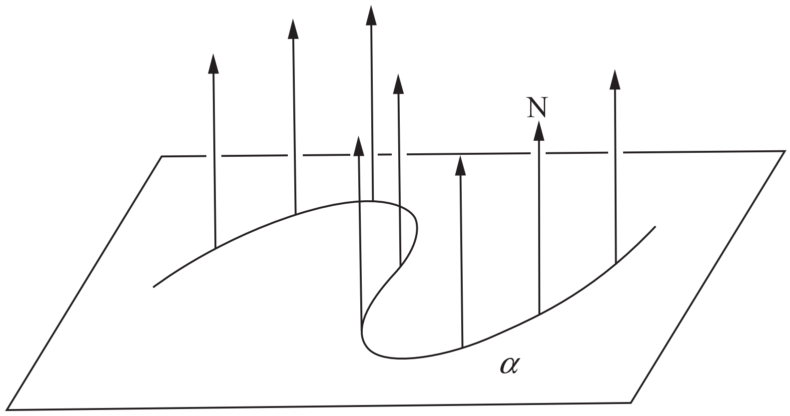


Figure 3-4. Plane: $dN_p = 0$.

which shows that the vector (x, y, z) is normal to the sphere at the point (x, y, z). Thus, $\bar{N} = (x, y, z)$ and $N = (-x, -y, -z)$ are fields of unit normal vectors in S^2. We fix an orientation in S^2 by choosing $N = (-x, -y, -z)$ as a normal field. Notice that N points toward the center of the sphere.

Restricted to the curve $\alpha(t)$, the normal vector

$$N(t) = (-x(t), -y(t), -z(t))$$

is a vector function of t, and therefore

$$dN(x'(t), y'(t), z'(t)) = N'(t) = (-x'(t), -y'(t), -z'(t));$$

that is, $dN_p(v) = -v$ for all $p \in S^2$ and all $v \in T_p(S^2)$. Notice that with the choice of $\bar{N}$ as a normal field (that is, with the opposite orientation) we would have obtained $d\bar{N}_p(v) = v$ (Fig. 3-5).

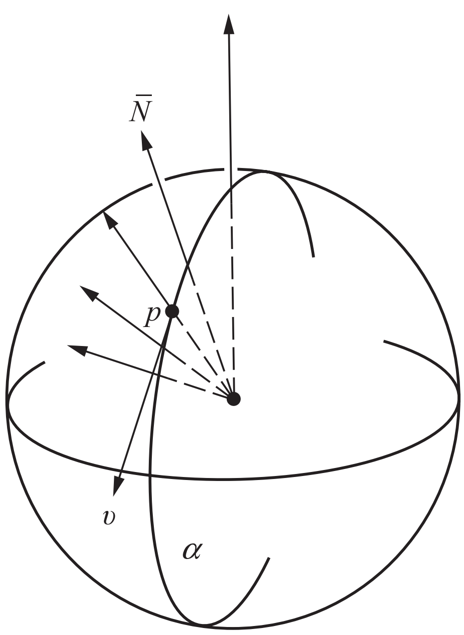


Figure 3-5. Unit sphere: $d\bar{N}_p(v) = v$.

Example 3. Consider the cylinder $\{(x, y, z) \in R^3; x^2 + y^2 = 1\}$. By an argument similar to that of the previous example, we see that $\bar{N} = (x, y, 0)$ and $N = (-x, -y, 0)$ are unit normal vectors at (x, y, z). We fix an orientation by choosing $N = (-x, -y, 0)$ as the normal vector field.

By considering a curve $(x(t), y(t), z(t))$ contained in the cylinder, that is, with $(x(t))^2 + (y(t))^2 = 1$, we are able to see that, along this curve, $N(t) = (-x(t), -y(t), 0)$ and therefore

$$dN(x'(t), y'(t), z'(t)) = N'(t) = (-x'(t), -y'(t), 0).$$

We conclude the following: If v is a vector tangent to the cylinder and parallel to the z axis, then

$$dN(v) = 0 = 0v;$$

if w is a vector tangent to the cylinder and parallel to the xy plane, then $dN(w) = -w$ (Fig. 3-6). It follows that the vectors v and w are eigenvectors of dN with eigenvalues 0 and -1, respectively (see the appendix to Chap. 3).

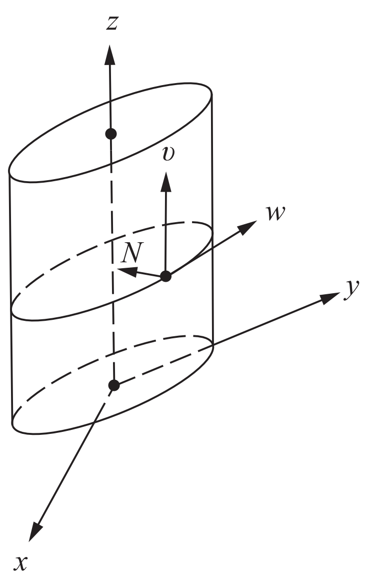


Figure 3-6

Example 4. Let us analyze the point $p = (0, 0, 0)$ of the hyperbolic paraboloid $z = y^2 - x^2$. For this, we consider a parametrization $\mathbf{x}(u, v)$ given by

$$\mathbf{x}(u, v) = (u, v, v^2 - u^2),$$

and compute the normal vector $N(u, v)$. We obtain successively

$$\begin{aligned} \mathbf{x}_u &= (1, 0, -2u), \\ \mathbf{x}_v &= (0, 1, 2v), \\ N &= \left(\frac{u}{\sqrt{u^2 + v^2 + \frac{1}{4}}}, \frac{-v}{\sqrt{u^2 + v^2 + \frac{1}{4}}}, \frac{1}{2\sqrt{u^2 + v^2 + \frac{1}{4}}} \right). \end{aligned}$$

Notice that at $p = (0, 0, 0)$ $\mathbf{x}_u$ and $\mathbf{x}_v$ agree with the unit vectors along the x and y axes, respectively. Therefore, the tangent vector at p to the curve $\alpha(t) = \mathbf{x}(u(t), v(t))$, with $\alpha(0) = p$, has, in R^3, coordinates $(u'(0), v'(0), 0)$ (Fig. 3-7). Restricting $N(u, v)$ to this curve and computing $N'(0)$, we obtain

$$N'(0) = (2u'(0), -2v'(0), 0),$$

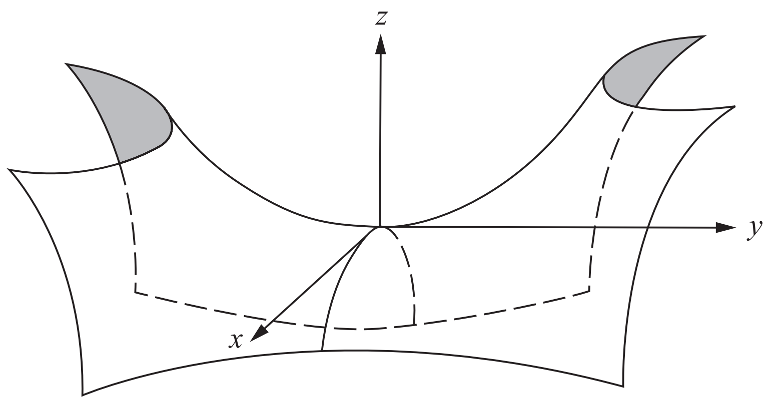


Figure 3-7

and therefore, at p,

$$dN_p(u'(0), v'(0), 0) = (2u'(0), -2v'(0), 0).$$

It follows that the vectors $(1, 0, 0)$ and $(0, 1, 0)$ are eigenvectors of dN_p with eigenvalues 2 and -2, respectively.

Example 5. The method of the previous example, applied to the point $p = (0, 0, 0)$ of the paraboloid $z = x^2 + ky^2$, $k > 0$, shows that the unit vectors of the x axis and the y axis are eigenvectors of dN_p, with eigenvalues 2 and $2k$, respectively (assuming that N is pointing outwards from the region bounded by the paraboloid).

An important fact about dN_p is contained in the following proposition.

PROPOSITION 1. *The differential* $\mathrm{dN_p}\colon \mathrm{T_p(S)} \to \mathrm{T_p(S)}$ *of the Gauss map is a self-adjoint linear map* (cf. the appendix to Chap. 3).

Proof. Since dN_p is linear, it suffices to verify that $\langle dN_p(w_1), w_2\rangle = \langle w_1, dN_p(w_2)\rangle$ for a basis $\{w_1, w_2\}$ of $T_p(S)$. Let $\mathbf{x}(u, v)$ be a parametrization of S at p and $\{\mathbf{x}_u, \mathbf{x}_v\}$ the associated basis of $T_p(S)$. If $\alpha(t) = \mathbf{x}(u(t), v(t))$ is a parametrized curve in S, with $\alpha(0) = p$, we have

$$\begin{aligned} dN_p(\alpha'(0)) &= dN_p(\mathbf{x}_u u'(0) + \mathbf{x}_v v'(0)) \\ &= \left.\frac{d}{dt} N(u(t), v(t))\right|_{t=0} \\ &= N_u u'(0) + N_v v'(0); \end{aligned}$$

in particular, $dN_p(\mathbf{x}_u) = N_u$ and $dN_p(\mathbf{x}_v) = N_v$. Therefore, to prove that dN_p is self-adjoint, it suffices to show that

$$\langle N_u, \mathbf{x}_v \rangle = \langle \mathbf{x}_u, N_v \rangle.$$

To sec this, take the derivatives of $\langle N, \mathbf{x}_u \rangle = 0$ and $\langle N, \mathbf{x}_v \rangle = 0$, relative to v and u, respectively, and obtain

$$\langle N_v, \mathbf{x}_u \rangle + \langle N, \mathbf{x}_{uv} \rangle = 0,$$
$$\langle N_u, \mathbf{x}_v \rangle + \langle N, \mathbf{x}_{vu} \rangle = 0.$$

Thus,

$$\langle N_u, \mathbf{x}_v \rangle = -\langle N, \mathbf{x}_{uv} \rangle = \langle N_v, \mathbf{x}_u \rangle. \qquad \textbf{Q.E.D.}$$

The fact that $dN_p \colon T_p(S) \to T_p(S)$ is a self-adjoint linear map allows us to associate to dN_p a quadratic form Q in $T_p(S)$, given by $Q(v) = \langle dN_p(v), v \rangle$, $v \in T_p(S)$ (cf. the appendix to Chap. 3). To obtain a geometric interpretation of this quadratic form, we need a few definitions. For reasons that will be clear shortly, we shall use the quadratic form $-Q$.

DEFINITION 2. *The quadratic form* II_p, *defined in* $\mathrm{T_p(S)}$ *by* $\mathrm{II_p(v)} = -\langle \mathrm{dN_p(v), v} \rangle$ *is called the* second fundamental form *of* S *at* p.

DEFINITION 3. *Let* C *be a regular curve in* S *passing through* $\mathrm{p} \in \mathrm{S}$, k *the curvature of* C *at* p, *and* $\cos\theta = \langle \mathrm{n, N} \rangle$, *where* n *is the normal vector to* C *and* N *is the normal vector to* S *at* p. *The number* $\mathrm{k_n = k}\cos\theta$ *is then called the* normal curvature *of* $\mathrm{C} \subset \mathrm{S}$ *at* p.

In other words, k_n is the length of the projection of the vector kn over the normal to the surface at p, with a sign given by the orientation N of S at p (Fig. 3-8).

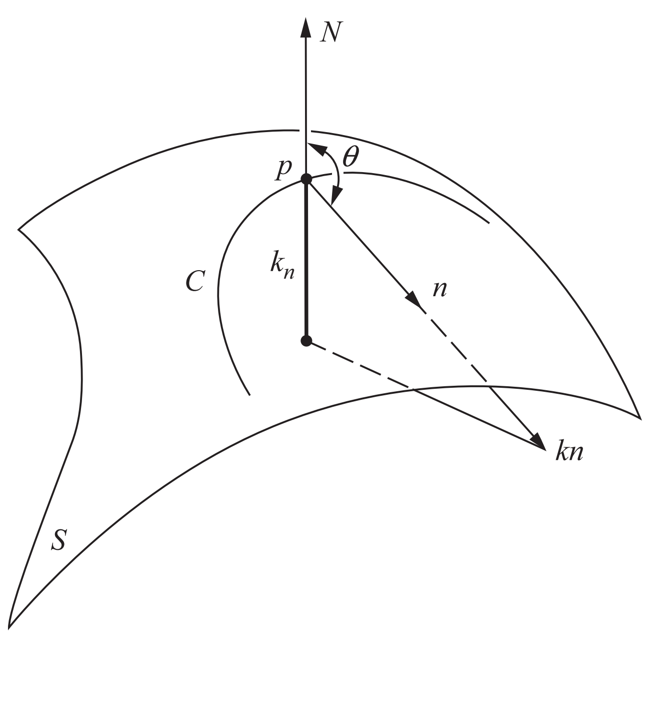


Figure 3-8

Remark. The normal curvature of C does not depend on the orientation of C but changes sign with a change of orientation for the surface.

To give an interpretation of the second fundamental form II_p, consider a regular curve $C \subset S$ parametrized by $\alpha(s)$, where s is the arc length of C, and with $\alpha(0) = p$. If we denote by $N(s)$ the restriction of the normal vector N to the curve $\alpha(s)$, we have $\langle N(s), \alpha'(s)\rangle = 0$. Hence,

$$\langle N(s), \alpha''(s)\rangle = -\langle N'(s), \alpha'(s)\rangle.$$

Therefore,

$$\begin{aligned} II_p(\alpha'(0)) &= -\langle dN_p(\alpha'(0)), \alpha'(0)\rangle \\ &= -\langle N'(0), \alpha'(0)\rangle = \langle N(0), \alpha''(0)\rangle \\ &= \langle N, kn\rangle(p) = k_n(p). \end{aligned}$$

In other words, the value of the second fundamental form II_p for a unit vector $v \in T_p(S)$ is equal to the normal curvature of a regular curve passing through p and tangent to v. In particular, we obtained the following result.

PROPOSITION 2 (Meusnier). *All curves lying on a surface* S *and having at a given point* p ∈ S *the same tangent line have at this point the same normal curvatures.*

The above proposition allows us to speak of the *normal curvature along a given direction at p.* It is convenient to use the following terminology. Given a unit vector $v \in T_p(S)$, the intersection of S with the plane containing v and $N(p)$ is called the *normal section* of S at p along v (Fig. 3-9). In a neighborhood of p, a normal section of S at p is a regular plane curve on S whose normal vector n at p is $\pm N(p)$ or zero; its curvature is therefore equal to the absolute value of the normal curvature along v at p. With this terminology, the above

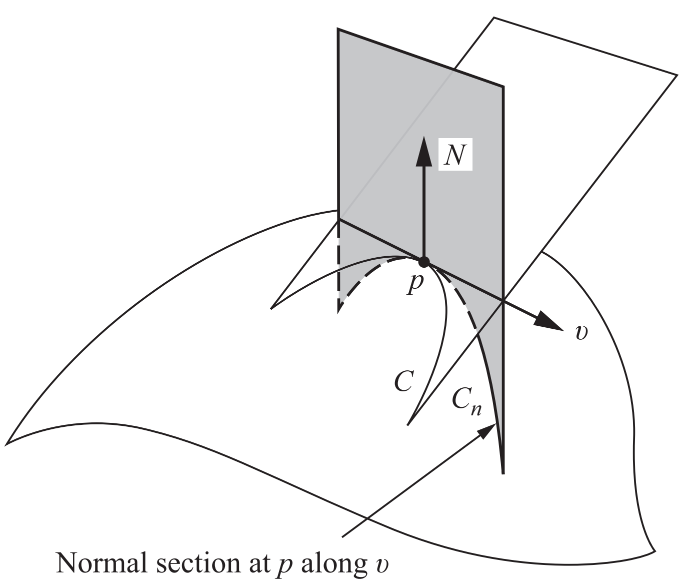


Figure 3-9. Meusnier theorem: C and C_n have the same normal curvature at p along v.

proposition says that the absolute value of the normal curvature at p of a curve $\alpha(s)$ is equal to the curvature of the normal section of S at p along $\alpha'(0)$.

Example 6. Consider the surface of revolution obtained by rotating the curve $z = y^4$ about the z axis (Fig. 3-10). We shall show that at $p = (0, 0, 0)$ the differential $dN_p = 0$. To see this, we observe that the curvature of the curve $z = y^4$ at p is equal to zero. Moreover, since the xy plane is a tangent plane to the surface at p, the normal vector $N(p)$ is parallel to the z axis. Therefore, any normal section at p is obtained from the curve $z = y^4$ by rotation; hence, it has curvature zero. It follows that all normal curvatures are zero at p, and thus $dN_p = 0$.

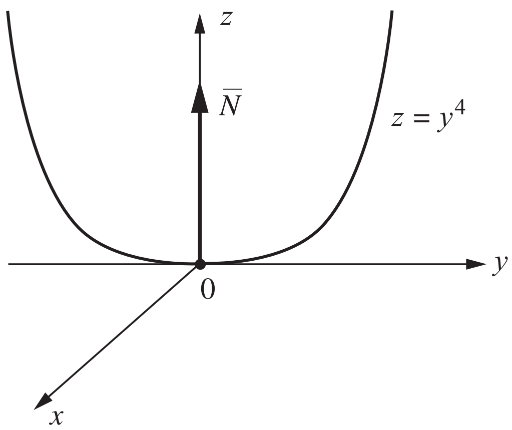


Figure 3-10

Example 7. In the plane of Example 1, all normal sections are straight lines; hence, all normal curvatures are zero. Thus, the second fundamental form is identically zero at all points. This agrees with the fact that $dN \equiv 0$.

In the sphere S^2 of Example 2, with N as orientation, the normal sections through a point $p \in S^2$ are circles with radius 1 (Fig. 3-11). Thus, all normal curvatures are equal to 1, and the second fundamental form is $II_p(v) = 1$ for all $p \in S^2$ and all $v \in T_p(S)$ with $|v| = 1$.

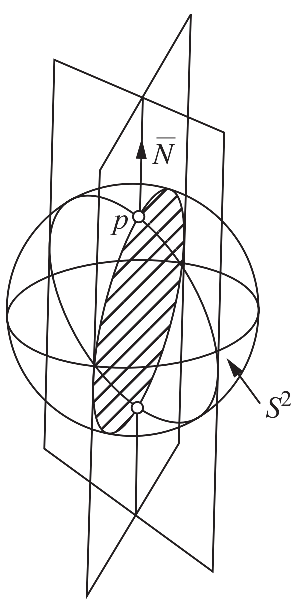


Figure 3-11. Normal sections on a sphere.

In the cylinder of Example 3, the normal sections at a point p vary from a circle perpendicular to the axis of the cylinder to a straight line parallel to the axis of the cylinder, passing through a family of ellipses (Fig. 3-12). Thus, the normal curvatures varies from 1 to 0. It is not hard to see geometrically that 1 is the maximum and 0 is the minimum of the normal curvature at p.

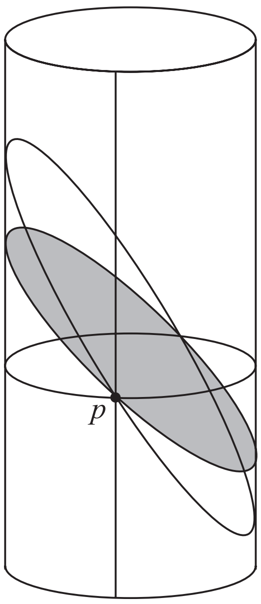


Figure 3-12. Normal sections on a cylinder.

However, an application of the theorem on quadratic forms of the appendix to Chap. 3 gives a simple proof of that. In fact, as we have seen in Example 3, the vectors w and v (corresponding to the directions of the normal curvatures 1 and 0, respectively) are eigenvectors of dN_p with eigenvalues -1 and 0, respectively. Thus, the second fundamental form takes up its extreme values in these vectors, as we claimed. Notice that this procedure allows us to check that such extreme values are l and 0.

We leave it to the reader to analyze the normal sections at the point $p = (0, 0, 0)$ of the hyperbolic paraboloid of Example 4.

Let us come back to the linear map dN_p. The theorem of the appendix to Chap. 3 shows that for each $p \in S$ there exists an orthonormal basis $\{e_1, e_2\}$ of $T_p(S)$ such that $dN_p(e_1) = -k_1 e_1$, $dN_p(e_2) = -k_2 e_2$. Moreover, k_1 and k_2 ($k_1 \geq k_2$) are the maximum and minimum of the second fundamental form II_p restricted to the unit circle of $T_p(S)$; that is, they are the extreme values of the normal curvature at p.

DEFINITION 4. *The maximum normal curvature* k_1 *and the minimum normal curvature* k_2 *are called the* principal curvatures *at* p*; the corresponding directions, that is, the directions given by the eigenvectors* e_1, e_2*, are called* principal directions *at* p.

For instance, in the plane all directions at all points are principal directions. The same happens with a sphere. In both cases, this comes from the fact that the second fundamental form at each point, restricted to the unit vectors, is constant (cf. Example 7); thus, all directions are extremals for the normal curvature.

In the cylinder of Example 3, the vectors v and w give the principal directions at p, corresponding to the principal curvatures 0 and 1, respectively. In the hyperbolic paraboloid of Example 4, the x and y axes are along the principal directions with principal curvatures -2 and 2, respectively.

DEFINITION 5. *If a regular connected curve* C *on* S *is such that for all* p $\in$ C *the tangent line of* C *is a principal direction at* p, *then* C *is said to be a* line of curvature *of* S.

PROPOSITION 3 (Olinde Rodrigues). *A necessary and sufficient condition for a connected regular curve* C *on* S *to be a line of curvature of* S *is that*

$$\mathrm{N}'(\mathrm{t}) = \lambda(\mathrm{t})\alpha'(\mathrm{t}),$$

for any parametrization $\alpha(\mathrm{t})$ *of* C, *where* $\mathrm{N}(\mathrm{t}) = \mathrm{N} \circ \alpha(\mathrm{t})$ *and* $\lambda(\mathrm{t})$ *is a differentiable function of* t. *In this case,* $-\lambda(\mathrm{t})$ *is the* (*principal*) *curvature along* $\alpha'(\mathrm{t})$.

Proof. It suffices to observe that if $\alpha'(t)$ is contained in a principal direction, then $\alpha'(t)$ is an eigenvector of dN and

$$dN(\alpha'(t)) = N'(t) = \lambda(t)\alpha'(t).$$

The converse is immediate. **Q.E.D.**

The knowledge of the principal curvatures at p allows us to compute easily the normal curvature along a given direction of $T_p(S)$. In fact, let $v \in T_p(S)$ with $|v| = 1$. Since e_1 and e_2 form an orthonormal basis of $T_p(S)$, we have

$$v = e_1 \cos\theta + e_2 \sin\theta,$$

where θ is the angle from e_1 to v in the orientation of $T_p(S)$. The normal curvature k_n along v is given by

$$\begin{aligned} k_n &= II_p(v) = -\langle dN_p(v), v\rangle \\ &= -\langle dN_p(e_1\cos\theta + e_2\sin\theta), e_1\cos\theta + e_2\sin\theta\rangle \\ &= \langle e_1k_1\cos\theta + e_2k_2\sin\theta, e_1\cos\theta + e_2\sin\theta\rangle \\ &= k_1\cos^2\theta + k_2\sin^2\theta. \end{aligned}$$

The last expression is known classically as the *Euler formula*; actually, it is just the expression of the second fundamental form in the basis $\{e_1, e_2\}$.

Given a linear map $A\colon V \to V$ of a vector space of dimension 2 and given a basis $\{v_1, v_2\}$ of V, we recall that

$$\text{determinant of } A = a_{11}a_{22} - a_{12}a_{21}, \qquad \text{trace of } A = a_{11} + a_{22},$$

where (a_{ij}) is the matrix of A in the basis $\{v_1, v_2\}$. It is known that these numbers do not depend on the choice of the basis $\{v_1, v_2\}$ and are, therefore, attached to the linear map A.

In our case, the determinant of dN is the product $(-k_1)(-k_2) = k_1k_2$ of the principal curvatures, and the trace of dN is the negative $-(k_1 + k_2)$ of the sum of principal curvatures. If we change the orientation of the surface, the determinant does not change (the fact that the dimension is even is essential here); the trace, however, changes sign.

DEFINITION 6. *Let* $\mathrm{p} \in \mathrm{S}$ *and let* $\mathrm{dN_p}\colon \mathrm{T_p(S)} \to \mathrm{T_p(S)}$ *be the differential of the Gauss map. The determinant of* $\mathrm{dN_p}$ *is the* Gaussian curvature K *of* S *at* p. *The negative of half of the trace of* $\mathrm{dN_p}$ *is called the* mean curvature H *of* S *at* p.

In terms of the principal curvatures we can write

$$K = k_1k_2, \qquad H = \frac{k_1 + k_2}{2}.$$

DEFINITION 7. *A point of a surface* S *is called*

1. Elliptic *if* $det(\mathrm{dN_p}) > 0$.
2. Hyperbolic *if* $det(\mathrm{dN_p}) < 0$.
3. Parabolic *if* $det(\mathrm{dN_p}) = 0$, *with* $\mathrm{dN_p} \neq 0$.
4. Planar *if* $\mathrm{dN_p} = 0$.

It is clear that this classification does not depend on the choice of the orientation.

At an elliptic point the Gaussian curvature is positive. Both principal curvatures have the same sign, and therefore all curves passing through this point have their normal vectors pointing toward the same side of the tangent plane. The points of a sphere are elliptic points. The point (0, 0, 0) of the paraboloid $z = x^2 + ky^2$, $k > 0$ (cf. Example 5), is also an elliptic point.

At a hyperbolic point, the Gaussian curvature is negative. The principal curvatures have opposite signs, and therefore there are curves through p whose normal vectors at p point toward any of the sides of the tangent plane at p. The point (0, 0, 0) of the hyperbolic paraboloid $z = y^2 - x^2$ (cf. Example 4) is a hyperbolic point.

At a parabolic point, the Gaussian curvature is zero, but one of the principal curvatures is not zero. The points of a cylinder (cf. Example 3) are parabolic points.

Finally, at a planar point, all principal curvatures are zero. The points of a plane trivially satisfy this condition. A nontrivial example of a planar point was given in Example 6.

DEFINITION 8. *If at* p $\in$ S, $k_1 = k_2$*, then* p *is called an* umbilical point *of* S; *in particular, the planar points* ($k_1 = k_2 = 0$) *are umbilical points.*

All the points of a sphere and a plane are umbilical points. Using the method of Example 6, we can verify that the point (0, 0, 0) of the paraboloid $z = x^2 + y^2$ is a (nonplanar) umbilical point.

We shall now prove the interesting fact that the only surfaces made up entirely of umbilical points are essentially spheres and planes.

PROPOSITION 4. *If all points of a connected surface* S *are umbilical points, then* S *is either contained in a sphere or in a plane.*

Proof. Let $p \in S$ and let $\mathbf{x}(u, v)$ be a parametrization of S at p such that the coordinate neighborhood V is connected.

Since each $q \in V$ is an umbilical point, we have, for any vector $w = a_1\mathbf{x}_u + a_2\mathbf{x}_v$ in $T_q(S)$,

$$dN(w) = \lambda(q)w,$$

where $\lambda = \lambda(q)$ is a real differentiable function in V.

We first show that $\lambda(q)$ is constant in V. For that, we write the above equation as

$$N_u a_1 + N_v a_2 = \lambda(\mathbf{x}_u a_1 + \mathbf{x}_v a_2);$$

hence, since w is arbitrary,

$$N_u = \lambda \mathbf{x}_u,$$
$$N_v = \lambda \mathbf{x}_v.$$

Differentiating the first equation in v and the second one in u and subtracting the resulting equations, we obtain

$$\lambda_u \mathbf{x}_v - \lambda_v \mathbf{x}_u = 0.$$

Since $\mathbf{x}_u$ and $\mathbf{x}_v$ are linear independent, we conclude that

$$\lambda_u = \lambda_v = 0$$

for all $q \in V$. Since V is connected, λ is constant in V, as we claimed.

If $\lambda \equiv 0$, $N_u = N_v = 0$ and therefore $N = N_0 =$ constant in V. Thus, $\langle \mathbf{x}(u, v), N_0\rangle_u = \langle \mathbf{x}(u, v), N_0\rangle_v = 0$; hence,

$$\langle \mathbf{x}(u, v), N_0\rangle = \text{const.},$$

and all points $\mathbf{x}(u, v)$ of V belong to a plane.

If $\lambda \neq 0$, then the point $\mathbf{x}(u, v) - (1/\lambda)N(u, v) = \mathbf{y}(u, v)$ is fixed, because

$$\left(\mathbf{x}(u, v) - \frac{1}{\lambda}N(u, v)\right)_u = \left(\mathbf{x}(u, v) - \frac{1}{\lambda}N(u, v)\right)_v = 0.$$

Since

$$|\mathbf{x}(u, v) - \mathbf{y}|^2 = \frac{1}{\lambda^2},$$

all points of V are contained in a sphere of center $\mathbf{y}$ and radius $1/|\lambda|$.

This proves the proposition locally, that is, for a neighborhood of a point $p \in S$. To complete the proof we observe that, since S is connected, given any other point $r \in S$, there exists a continuous curve $\alpha\colon [0, 1] \to S$ with $\alpha(0) = p$, $\alpha(1) = r$. For each point $\alpha(t) \in S$ of this curve there exists a neighborhood V_t in S contained in a sphere or in a plane and such that $\alpha^{-1}(V_t)$ is an open interval of $[0, 1]$. The union $\bigcup \alpha^{-1}(V_t)$, $t \in [0, 1]$, covers $[0, 1]$ and since $[0, 1]$ is a closed interval, it is covered by finitely many elements of the family $\{\alpha^{-1}(V_t)\}$ (cf. the Heine-Borel theorem, Prop. 6 of the appendix to Chap. 2). Thus, $\alpha([0, 1])$ is covered by a finite number of the neighborhoods V_t.

If the points of one of these neighborhoods are on a plane, all the others will be on the same plane. Since r is arbitrary, all the points of S belong to this plane.

If the points of one of these neighborhoods are on a sphere, the same argument shows that all points on S belong to a sphere, and this completes the proof. **Q.E.D.**

DEFINITION 9. *Let* p *be a point in* S. *An* asymptotic direction *of* S *at* p *is a direction of* $T_p(S)$ *for which the normal curvature is zero. An* asymptotic curve *of* S *is a regular connected curve* $C \subset S$ *such that for each* $p \in C$ *the tangent line of* C *at* p *is an asymptotic direction.*

It follows at once from the definition that at an elliptic point there are no asymptotic directions.

A useful geometric interpretation of the asymptotic directions is given by means of the Dupin indicatrix, which we shall now describe.

Let p be a point in S. The *Dupin indicatrix* at p is the set of vectors w of $T_p(S)$ such that $II_p(w) = \pm 1$.

To write the equations of the Dupin indicatrix in a more convenient form, let (ξ, η) be the Cartesian coordinates of $T_p(S)$ in the orthonormal basis $\{e_1, e_2\}$, where e_1 and e_2 are eigenvectors of dN_p. Given $w \in T_p(S)$, let ρ and θ

be "polar coordinates" defined by $w = \rho v$, with $|v| = 1$ and $v = e_1 \cos\theta + e_2 \sin\theta$, if $\rho \neq 0$. By Euler's formula,

$$\begin{aligned} \pm 1 = II_p(w) &= p^2 II_p(v) \\ &= k_1 \rho^2 \cos^2\theta + k_2 \rho^2 \sin^2\theta \\ &= k_1 \xi^2 + k_2 \eta^2, \end{aligned}$$

where $w = \xi e_1 + \eta e_2$. Thus, the coordinates (ξ, η) of a point of the Dupin indicatrix satisfy the equation

$$k_1 \xi^2 + k_2 \eta^2 = \pm 1; \qquad \textbf{(1)}$$

hence, the Dupin indicatrix is a union of conics in $T_p(S)$. We notice that the normal curvature along the direction determined by w is $k_n(v) = II_p(v) = \pm(1/\rho^2)$.

For an elliptic point, the Dupin indicatrix is an ellipse (k_1 and k_2 have the same sign); this ellipse degenerates into a circle if the point is an umbilical nonplanar point ($k_1 = k_2 \neq 0$).

For a hyperbolic point, k_1 and k_2 have opposite signs. The Dupin indicatrix is therefore made up of two hyperbolas with a common pair of asymptotic lines (Fig. 3-13). Along the directions of the asymptotes, the normal curvature is zero; they are therefore asymptotic directions. This justifies the terminology and shows that a hyperbolic point has *exactly two* asymptotic directions.

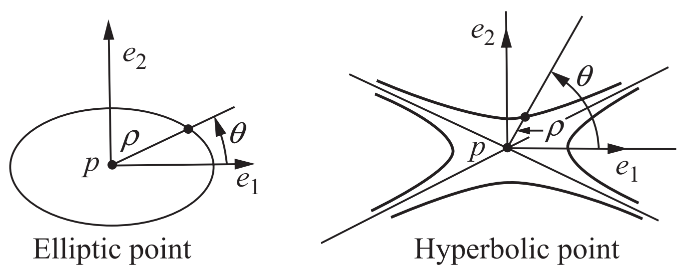


Figure 3-13. The Dupin indicatrix.

For a parabolic point, one of the principal curvatures is zero, and the Dupin indicatrix degenerates into a pair of parallel lines. The common direction of these lines is the only asymptotic direction at the given point.

In Example 5 of Sec. 3-3 we shall show an interesting property of the Dupin indicatrix.

Closely related with the concept of asymptotic direction is the concept of conjugate directions, which we shall now define.

DEFINITION 10. *Let* p *be a point on a surface* S. *Two nonzero vectors* $\mathrm{w}_1, \mathrm{w}_2 \in \mathrm{T_p(S)}$ *are* conjugate *if* $\langle \mathrm{dN_p(w_1), w_2} \rangle = \langle \mathrm{w_1, dN_p(w_2)} \rangle = 0$.

Two directions r_1, r_2 *at* p *are* conjugate *if a pair of nonzero vectors* w_1, w_2 *parallel to* r_1 *and* r_2, *respectively, are conjugate.*

It is immediate to check that the definition of conjugate directions does not depend on the choice of the vectors w_1 and w_2 on r_1 and r_2.

It follows from the definition that the principal directions are conjugate and that an asymptotic direction is conjugate to itself. Furthermore, at a nonplanar umbilic, every orthogonal pair of directions is a pair of conjugate directions, and at a planar umbilic each direction is conjugate to any other direction.

Let us assume that $p \in S$ is not an umbilical point, and let $\{e_1, e_2\}$ be the orthonormal basis of $T_p(S)$ determined by $dN_p(e_1) = -k_1 e_1$, $dN_p(e_2) = -k_2 e_2$. Let θ and φ be the angles that a pair of directions r_1 and r_2 make with e_1. We claim that r_1 and r_2 are conjugate if and only if

$$k_1 \cos\theta \cos\varphi = -k_2 \sin\theta \sin\varphi. \qquad \textbf{(2)}$$

In fact, r_1 and r_2 are conjugate if and only if the vectors

$$w_1 = e_1 \cos\theta + e_2 \sin\theta, \qquad w_2 = e_1 \cos\varphi + e_2 \sin\varphi$$

are conjugate. Thus,

$$0 = \langle dN_p(w_1), w_2 \rangle = -k_1 \cos\theta \cos\varphi - k_2 \sin\theta \sin\varphi.$$

Hence, condition (2) follows.

When both k_1 and k_2 are nonzero (i.e., p is either an elliptic or a hyperbolic point), condition (2) leads to a geometric construction of conjugate directions in terms of the Dupin indicatrix at p. We shall describe the construction at an elliptic point, the situation at a hyperbolic point being similar. Let r be a straight line through the origin of $T_p(S)$ and consider the intersection points q_1, q_2 of r with the Dupin indicatrix (Fig. 3-14). The tangent lines of the

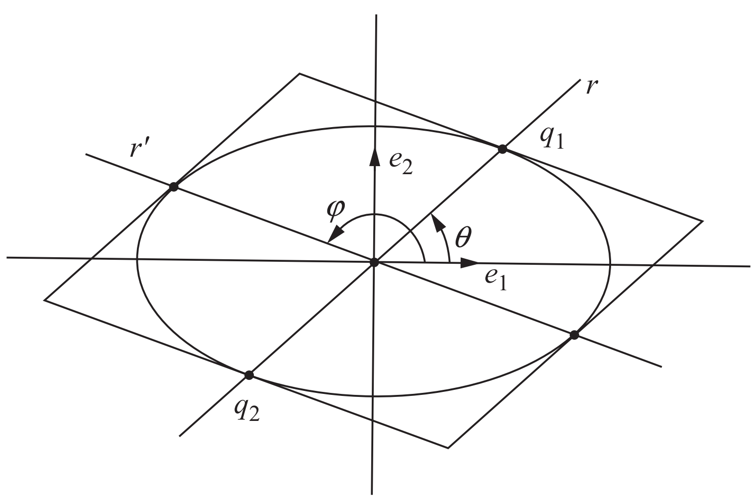


Figure 3-14. Construction of conjugate directions.

Dupin indicatrix at q_1 and q_2 are parallel, and their common direction r' is conjugate to r. We shall leave the proofs of these assertions to the Exercises (Exercise 12).

EXERCISES

1. Show that at a hyperbolic point, the principal directions bisect the asymptotic directions.

2. Show that if a surface is tangent to a plane along a curve, then the points of this curve are either parabolic or planar.

3. Let $C \subset S$ be a regular curve on a surface S with Gaussian curvature $K > 0$. Show that the curvature k of C at p satisfies

$$|k| \geq \min(|k_1|, |k_2|),$$

where k_1 and k_2 are the principal curvatures of S at p.

4. Assume that a surface S has the property that $|k_1| \leq 1$, $|k_2| \leq 1$ everywhere. Is it true that the curvature k of a curve on S also satisfies $|k| \leq 1$?

5. Show that the mean curvature H at $p \in S$ is given by

$$H = \frac{1}{\pi} \int_0^{\pi} k_n(\theta)\, d\theta,$$

where $k_n(\theta)$ is the normal curvature at p along a direction making an angle θ with a fixed direction.

6. Show that the sum of the normal curvatures for any pair of orthogonal directions, at a point $p \in S$, is constant.

7. Show that if the mean curvature is zero at a nonplanar point, then this point has two orthogonal asymptotic directions.

8. Describe the region of the unit sphere covered by the image of the Gauss map of the following surfaces:

a. Paraboloid of revolution $z = x^2 + y^2$.

b. Hyperboloid of revolution $x^2 + y^2 - z^2 = 1$.

c. Catenoid $x^2 + y^2 = \cosh^2 z$.

9. Prove that

a. The image $N \circ \alpha$ by the Gauss map $N \colon S \to S^2$ of a parametrized regular curve $\alpha \colon I \to S$ which contains no planar or parabolic points is a parametrized regular curve on the sphere S^2 (called the *spherical image* of α).

b. If $C = \alpha(I)$ is a line of curvature, and k is its curvature at p, then

$$k = |k_n k_N|,$$

where k_n is the normal curvature at p along the tangent line of C and k_N is the curvature of the spherical image $N(C) \subset S^2$ at $N(p)$.

10. Assume that the osculating plane of a line of curvature $C \subset S$, which is nowhere tangent to an asymptotic direction, makes a constant angle with the tangent plane of S along C. Prove that C is a plane curve.

11. Let p be an elliptic point of a surface S, and let r and r' be conjugate directions at p. Let r vary in $T_p(S)$ and show that the minimum of the angle of r with r' is reached at a unique pair of directions in $T_p(S)$ that are symmetric with respect to the principal directions.

12. Let p be a hyperbolic point of a surface S, and let r be a direction in $T_p(S)$. Describe and justify a geometric construction to find the conjugate direction r' of r in terms of the Dupin indicatrix (cf. the construction at the end of Sec. 3-2).

***13.** (*Theorem of Beltrami-Enneper.*) Prove that the absolute value of the torsion τ at a point of an asymptotic curve, whose curvature is nowhere zero, is given by

$$|\tau| = \sqrt{-K},$$

where K is the Gaussian curvature of the surface at the given point.

***14.** If the surface S_1 intersects the surface S_2 along the regular curve C, then the curvature k of C at $p \in C$ is given by

$$k^2 \sin^2\theta = \lambda_1^2 + \lambda_2^2 - 2\lambda_1\lambda_2 \cos\theta,$$

where λ_1 and λ_2 are the normal curvatures at p, along the tangent line to C, of S_1 and S_2, respectively, and θ is the angle made up by the normal vectors of S_1 and S_2 at p.

15. (*Theorem of Joachimstahl.*) Suppose that S_1 and S_2 intersect along a regular curve C and make an angle $\theta(p)$, $p \in C$. Assume that C is a line of curvature of S_1. Prove that $\theta(p)$ is constant if and only if C is a line of curvature of S_2.

***16.** Show that the meridians of a torus are lines of curvature.

17. Show that if $H \equiv 0$ on S and S has no planar points, then the Gauss map $N\colon S \to S^2$ has the following property:

$$\langle dN_p(w_1), dN_p(w_2)\rangle = -K(p)\langle w_1, w_2\rangle$$

for all $p \in S$ and all $w_1, w_2 \in T_p(S)$. Show that the above condition implies that the angle of two intersecting curves on S and the angle of their spherical images (cf. Exercise 9) are equal up to a sign.

***18.** Let $\lambda_1, \ldots, \lambda_m$ be the normal curvatures at $p \in S$ along directions making angles $0, 2\pi/m, \ldots, (m-1)2\pi/m$ with a principal direction, $m > 2$. Prove that

$$\lambda_1 + \cdots + \lambda_m = mH,$$

where H is the mean curvature at p.

***19.** Let $C \subset S$ be a regular curve in S. Let $p \in C$ and $\alpha(s)$ be a parametrization of C in p by arc length so that $\alpha(0) = p$. Choose in $T_p(S)$ an orthonormal positive basis $\{t, h\}$, where $t = \alpha'(0)$. The *geodesic torsion* τ_g of $C \subset S$ at p is defined by

$$\tau_g = \left\langle \frac{dN}{ds}(0), h \right\rangle.$$

Prove that

a. $\tau_g = (k_1 - k_2)\cos\varphi \sin\varphi$, where φ is the angle from e_1 to t and t is the unit tangent vector corresponding to the principal curvature k_1.

b. If τ is the torsion of C, n is the (principal) normal vector of C and $\cos\theta = \langle N, n \rangle$, then

$$\frac{d\theta}{ds} = \tau - \tau_g.$$

c. The lines of curvature of S are characterized by having geodesic torsion identically zero.

***20.** (*Dupin's Theorem.*) Three families of surfaces are said to form a *triply orthogonal system* in an open set $U \subset R^3$ if a unique surface of each family passes through each point $p \in U$ and if the three surfaces that pass through p are pairwise orthogonal. Use part c of Exercise 19 to prove Dupin's theorem: *The surfaces of a triply orthogonal system intersect each other in lines of curvature.*

3-3. The Gauss Map in Local Coordinates

In the preceding section, we introduced some concepts related to the local behavior of the Gauss map. To emphasize the geometry of the situation, the definitions were given without the use of a coordinate system. Some simple examples were then computed directly from the definitions; this procedure, however, is inefficient in handling general situations. In this section, we shall obtain the expressions of the second fundamental form and of the differential of the Gauss map in a coordinate system. This will give us a systematic method for computing specific examples. Moreover, the general expressions thus obtained are essential for a more detailed investigation of the concepts introduced above.

All parametrizations $\mathbf{x}\colon U \subset R^2 \to S$ considered in this section are assumed to be compatible with the orientation N of S; that is, in $\mathbf{x}(U)$,

$$N = \frac{\mathbf{x}_u \wedge \mathbf{x}_v}{|\mathbf{x}_u \wedge \mathbf{x}_v|}.$$

Let $\mathbf{x}(u, v)$ be a parametrization at a point $p \in S$ of a surface S, and let $\alpha(t) = \mathbf{x}(u(t), v(t))$ be a parametrized curve on S, with $\alpha(0) = p$. To simplify the notation, we shall make the convention that all functions to appear below denote their values at the point p.

The tangent vector to $\alpha(t)$ at p is $\alpha' = \mathbf{x}_u u' + \mathbf{x}_v v'$ and

$$dN(\alpha') = N'(u(t), v(t)) = N_u u' + N_v v'.$$

Since N_u and N_v belong to $T_p(S)$, we may write

$$\begin{aligned} N_u &= a_{11}\mathbf{x}_u + a_{21}\mathbf{x}_v, \\ N_v &= a_{12}\mathbf{x}_u + a_{22}\mathbf{x}_v, \end{aligned} \tag{1}$$

and therefore,

$$dN(\alpha') = (a_{11}u' + a_{12}v')\mathbf{x}_u + (a_{21}u' + a_{22}v')\mathbf{x}_v;$$

hence,

$$dN\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} u' \\ v' \end{pmatrix}.$$

This shows that in the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$, dN is given by the matrix (a_{ij}), $i, j = 1, 2$. Notice that this matrix is not necessarily symmetric, unless $\{\mathbf{x}_u, \mathbf{x}_v\}$ is an orthonormal basis.

On the other hand, the expression of the second fundamental form in the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ is given by

$$\begin{aligned} II_p(\alpha') &= -\langle dN(\alpha'), \alpha'\rangle = -\langle N_u u' + N_v v', \mathbf{x}_u u' + \mathbf{x}_v v'\rangle \\ &= e(u')^2 + 2fu'v' + g(v')^2, \end{aligned}$$

where, since $\langle N, \mathbf{x}_u\rangle = \langle N, \mathbf{x}_v\rangle = 0$,

$$\begin{aligned} e &= -\langle N_u, \mathbf{x}_u\rangle = \langle N, \mathbf{x}_{uu}\rangle, \\ f &= -\langle N_v, \mathbf{x}_u\rangle = \langle N, \mathbf{x}_{uv}\rangle = \langle N, \mathbf{x}_{vu}\rangle = -\langle N_u, \mathbf{x}_v\rangle, \\ g &= -\langle N_v, \mathbf{x}_v\rangle = \langle N, \mathbf{x}_{vv}\rangle. \end{aligned}$$

We shall now obtain the values of a_{ij} in terms of the coefficients e, f, g. From Eq. (1), we have

$$\begin{aligned} -f &= \langle N_u, \mathbf{x}_v \rangle = a_{11}F + a_{21}G, \\ -f &= \langle N_v, \mathbf{x}_u \rangle = a_{12}E + a_{22}F, \\ -e &= \langle N_u, \mathbf{x}_u \rangle = a_{11}E + a_{21}F, \\ -g &= \langle N_v, \mathbf{x}_v \rangle = a_{12}F + a_{22}G, \end{aligned} \tag{2}$$

where E, F, and G are the coefficients of the first fundamental form in the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ (cf. Sec. 2-5). Relations (2) may be expressed in matrix form by

$$-\begin{pmatrix} e & f \\ f & g \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix}; \tag{3}$$

hence,

$$\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} = -\begin{pmatrix} e & f \\ f & g \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1},$$

where $(\ \ )^{-1}$ means the inverse matrix of $(\ \ )$. It is easily checked that

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} = \frac{1}{EG - F^2} \begin{pmatrix} G & -F \\ -F & E \end{pmatrix},$$

whence the following expressions for the coefficients (a_{ij}) of the matrix of dN in the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$:

$$\begin{aligned} a_{11} &= \frac{fF - eG}{EG - F^2}, \\ a_{12} &= \frac{gF - fG}{EG - F^2}, \\ a_{21} &= \frac{eF - fE}{EG - F^2}, \\ a_{22} &= \frac{fF - gE}{EG - F^2}. \end{aligned}$$

For completeness, it should be mentioned that relations (1), with the above values, are known as the *equations of Weingarten*.

From Eq. (3) we immediately obtain

$$K = \det(a_{ij}) = \frac{eg - f^2}{EG - F^2}. \tag{4}$$

To compute the mean curvature, we recall that $-k_1, -k_2$ are the eigenvalues of dN. Therefore, k_1 and k_2 satisfy the equation

$$dN(v) = -kv = -kIv \quad \text{for some } v \in T_p(S),\ v \neq 0,$$

where I is the identity map. It follows that the linear map $dN + kI$ is not invertible; hence, it has zero determinant. Thus,

$$\det \begin{pmatrix} a_{11} + k & a_{12} \\ a_{21} & a_{22} + k \end{pmatrix} = 0$$

or

$$k^2 + k(a_{11} + a_{22}) + a_{11}a_{22} - a_{21}a_{12} = 0.$$

Since k_1 and k_2 are the roots of the above quadratic equation, we conclude that

$$H = \frac{1}{2}(k_1 + k_2) = -\frac{1}{2}(a_{11} + a_{22}) = \frac{1}{2}\frac{eG - 2fF + gE}{EG - F^2}; \tag{5}$$

hence,

$$k^2 - 2Hk + K = 0,$$

and therefore,

$$k = H \pm \sqrt{H^2 - K}. \tag{6}$$

From this relation, it follows that if we choose $k_1(q) \geq k_2(q)$, $q \in S$, then the functions k_1 and k_2 are continuous in S. Moreover, k_1 and k_2 are differentiable in S, except perhaps at the umbilical points $(H^2 = K)$ of S.

In the computations of this chapter, it will be convenient to write for short

$$\langle u \wedge v, w \rangle = (u, v, w) \quad \text{for any } u, v, w \in R^3.$$

We recall that this is merely the determinant of the 3×3 matrix whose columns (or lines) are the components of the vectors u, v, w in the canonical basis of R^3.

Example 1. We shall compute the Gaussian curvature of the points of the torus covered by the parametrization (cf. Example 6 of Sec. 2-2)

$$\mathbf{x}(u, v) = ((a + r\cos u)\cos v, (a + r\cos u)\sin v, r\sin u),$$
$$0 < u < 2\pi, \quad 0 < v < 2\pi.$$

For the computation of the coefficients e, f, g, we need to know N (and thus $\mathbf{x}_u$ and $\mathbf{x}_v$), $\mathbf{x}_{uu}$, $\mathbf{x}_{uv}$, and $\mathbf{x}_{vv}$:

$$\begin{aligned}
\mathbf{x}_u &= (-r\sin u\cos v, -r\sin u\sin v, r\cos u),\\
\mathbf{x}_v &= (-(a + r\cos u)\sin v, (a + r\cos u)\cos v, 0),\\
\mathbf{x}_{uu} &= (-r\cos u\cos v, -r\cos u\sin v, -r\sin u),\\
\mathbf{x}_{uv} &= (r\sin u\sin v, -r\sin u\cos v, 0),\\
\mathbf{x}_{vv} &= (-(a + r\cos u)\cos v, -(a + r\cos u)\sin v, 0).
\end{aligned}$$

From these, we obtain

$$\begin{aligned}
E &= \langle \mathbf{x}_u, \mathbf{x}_u\rangle = r^2, \quad F = \langle \mathbf{x}_u, \mathbf{x}_v\rangle = 0,\\
G &= \langle \mathbf{x}_v, \mathbf{x}_v\rangle = (a + r\cos u)^2.
\end{aligned}$$

Introducing the values just obtained in $e = \langle N, \mathbf{x}_{uu}\rangle$, we have, since $|\mathbf{x}_u \wedge \mathbf{x}_v| = \sqrt{EG - F^2}$,

$$e = \left\langle \frac{\mathbf{x}_u \wedge \mathbf{x}_v}{|\mathbf{x}_u \wedge \mathbf{x}_v|}, \mathbf{x}_{uu}\right\rangle = \frac{(\mathbf{x}_u, \mathbf{x}_v, \mathbf{x}_{uu})}{\sqrt{EG - F^2}} = \frac{r^2(a + r\cos u)}{r(a + r\cos u)} = r.$$

Similarly, we obtain

$$\begin{aligned}
f &= \frac{(\mathbf{x}_u, \mathbf{x}_v, \mathbf{x}_{uv})}{r(a + r\cos u)} = 0,\\
g &= \frac{(\mathbf{x}_u, \mathbf{x}_v, \mathbf{x}_{vv})}{r(a + r\cos u)} = \cos u(a + r\cos u).
\end{aligned}$$

Finally, since $K = (eg - f^2)/(EG - F^2)$, we have that

$$K = \frac{\cos u}{r(a + r\cos u)}.$$

From this expression, it follows that $K = 0$ along the parallels $u = \pi/2$ and $u = 3\pi/2$; the points of such parallels are therefore parabolic points. In the region of the torus given by $\pi/2 < u < 3\pi/2$, K is negative (notice that $r > 0$ and $a > r$); the points in this region are therefore hyperbolic points. In the region given by $0 < u < \pi/2$ or $3\pi/2 < u < 2\pi$, the curvature is positive and the points are elliptic points (Fig. 3-15).

As an application of the expression for the second fundamental form in coordinates, we shall prove a proposition which gives information about the

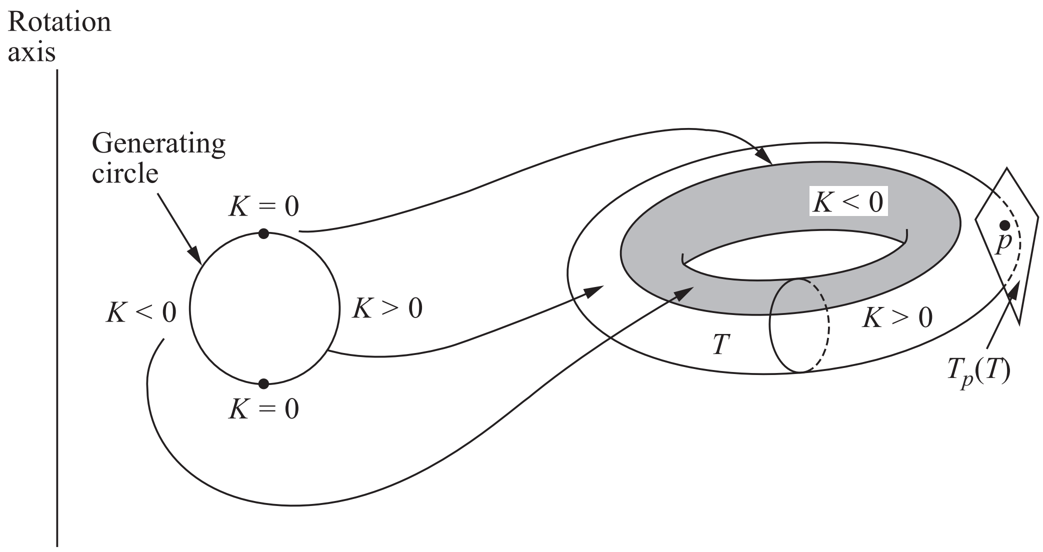


Figure 3-15

position of a surface in the neighborhood of an elliptic or a hyperbolic point, relative to the tangent plane at this point. For instance, if we look at an elliptic point of the torus of Example 1, we find that the surface lies on one side of the tangent plane at such a point (see Fig. 3-15). On the other hand, if p is a hyperbolic point of the torus T and $V \subset T$ is any neighborhood of p, we can find points of V on both sides of $T_p(S)$, however small V may be. This example reflects a general local fact that is described in the following proposition.

PROPOSITION 1. *Let* $p \in S$ *be an elliptic point of a surface* S. *Then there exists a neighborhood* V *of* p *in* S *such that all points in* V *belong to the same side of the tangent plane* $T_p(S)$. *Let* $p \in S$ *be a hyperbolic point. Then in each neighborhood of* p *there exist points of* S *in both sides of* $T_p(S)$.

Proof. Let $\mathbf{x}(u, v)$ be a parametrization in p, with $\mathbf{x}(0, 0) = p$. The distance d from a point $q = \mathbf{x}(u, v)$ to the tangent plane $T_p(S)$ is given by (Fig. 3-16)

$$d = \langle \mathbf{x}(u, v) - \mathbf{x}(0, 0), N(p) \rangle.$$

Since $\mathbf{x}(u, v)$ is differentiable, we have Taylor's formula:

$$\mathbf{x}(u, v) = \mathbf{x}(0, 0) + \mathbf{x}_u u + \mathbf{x}_v v + \tfrac{1}{2}(\mathbf{x}_{uu} u^2 + 2\mathbf{x}_{uv} uv + \mathbf{x}_{vv} v^2) + \bar{R},$$

where the derivatives are taken at $(0, 0)$ and the remainder $\bar{R}$ satisfies the condition

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

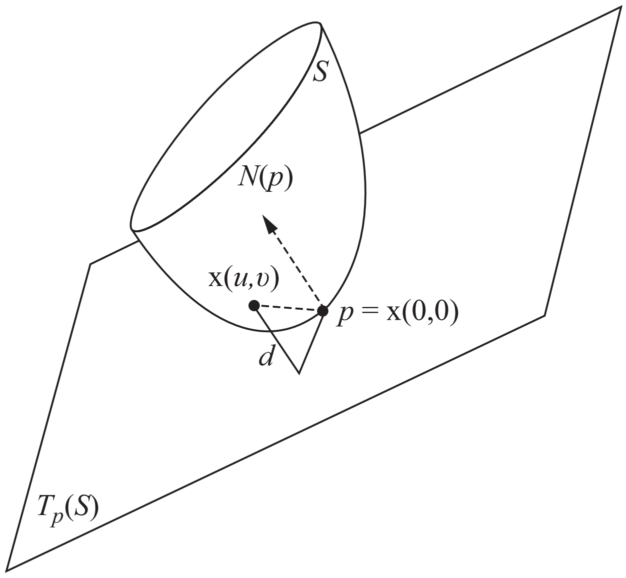


Figure 3-16

It follows that

$$\begin{aligned} d &= \langle \mathbf{x}(u, v) - \mathbf{x}(0, 0), N(p)\rangle \\ &= \tfrac{1}{2}\{\mathbf{x}_{uu}, N(p)\rangle u^2 + 2\langle \mathbf{x}_{uv}, N(p)\rangle uv + \langle \mathbf{x}_{vv}, N(p)\rangle v^2\} + R \\ &= \tfrac{1}{2}(eu^2 + 2fuv + gv^2) + R = \tfrac{1}{2}II_p(w) + R, \end{aligned}$$

where $w = \mathbf{x}_u u + \mathbf{x}_v v$, $R = \langle \bar{R}, N(p)\rangle$, and $\lim_{w\to 0}(R/|w|^2) = 0$.

For an elliptic point p, $II_p(w)$ has a fixed sign. Therefore, for all (u, v) sufficiently near p, d has the same sign as $II_p(w)$; that is, all such (u, v) belong to the same side of $T_p(S)$.

For a hyperbolic point p, in each neighborhood of p there exist points (u, v) and $(\bar{u}, \bar{v})$ such that $II_p(w/|w|)$ and $II_p(\bar{w}/|\bar{w}|)$ have opposite signs (here $\bar{w} = \mathbf{x}_u\bar{u} + \mathbf{x}_v\bar{v}$); such points belong therefore to distinct sides of $T_p(S)$.

Q.E.D.

No such statement as Prop. 1 can be made in a neighborhood of a parabolic or a planar point. In the above examples of parabolic and planar points (cf. Examples 3 and 6 of Sec. 3-2) the surface lies on one side of the tangent plane and may have a line in common with this plane. In the following examples we shall show that an entirely different situation may occur.

Example 2. The "monkey saddle" (see Fig. 3-17) is given by

$$x = u, \quad y = v, \quad z = u^3 - 3v^2u.$$

A direct computation shows that at (0, 0) the coefficients of the second fundamental form are $e = f = g = 0$; the point (0, 0) is therefore a planar point. In any neighborhood of this point, however, there are points in both sides of its tangent plane.

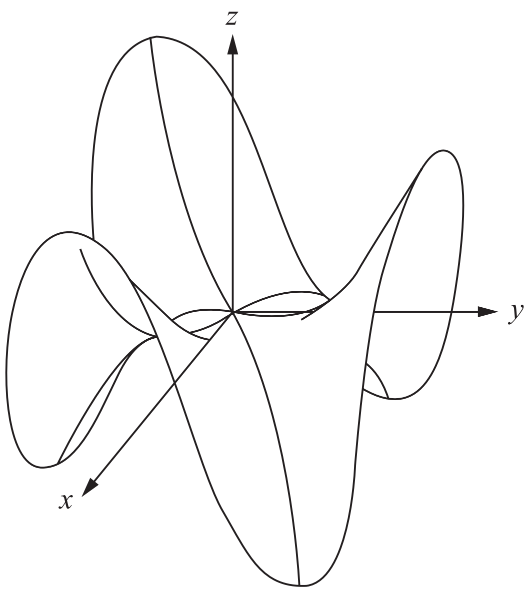


Figure 3-17

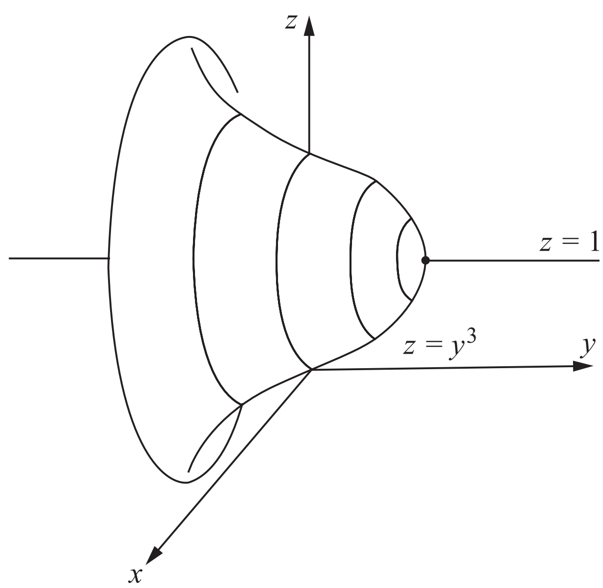


Figure 3-18

Example 3. Consider the surface obtained by rotating the curve $z = y^3$, $-1 < z < 1$, about the line $z = 1$ (see Fig. 3-18). A simple computation shows that the points generated by the rotation of the origin O are parabolic points. We shall omit this computation, because we shall prove shortly (Example 4) that the parallels and the meridians of a surface of revolution are lines of curvature; this, together with the fact that, for the points in question, the meridians (curves of the form $y = x^3$) have zero curvature and the parallel is a normal section, will imply the above statement.

Notice that in any neighborhood of such a parabolic point there exist points in both sides of the tangent plane.

The expression of the second fundamental form in local coordinates is particularly useful for the study of the asymptotic and principal directions. We first look at the asymptotic directions.

Let $\mathbf{x}(u, v)$ be a parametrization at $p \in S$, with $\mathbf{x}(0, 0) = p$, and let $e(u, v) = e$, $f(u, v) = f$, and $g(u, v) = g$ be the coefficients of the second fundamental form in this parametrization.

We recall that (see Def. 9 of Sec. 3-2) a connected regular curve C in the coordinate neighborhood of $\mathbf{x}$ is an asymptotic curve if and only if for any parametrization $\alpha(t) = \mathbf{x}(u(t), v(t))$, $t \in I$, of C we have $II(\alpha'(t)) = 0$, for all $t \in I$, that is, if and only if

$$e(u')^2 + 2fu'v' + g(v')^2 = 0, \quad t \in I. \tag{7}$$

Because of that, Eq. (7) is called the *differential equation of the asymptotic curves.* In the next section we shall give a more precise meaning to this expression. For the time being, we want to draw from Eq. (7) only the following useful conclusion: *A necessary and sufficient condition for a parametrization in a neighborhood of a hyperbolic point* ($eg - f^2 < 0$) *to be such that* *the coordinate curves of the parametrization are asymptotic curves is that* $\mathrm{e} = \mathrm{g} = 0$.

In fact, if both curves $u = \text{const.}$, $v = v(t)$ and $u = u(t)$, $v = \text{const.}$ satisfy Eq. (7), we obtain $e = g = 0$. Conversely, if this last condition holds and $f \neq 0$, Eq. (7) becomes $fu'v' = 0$, which is clearly satisfied by the coordinate lines.

We shall now consider the principal directions, maintaining the notations already established.

A connected regular curve C in the coordinate neighborhood of $\mathbf{x}$ is a line of curvature if and only if for any parametrization $\alpha(t) = \mathbf{x}(u(t), v(t))$ of C, $t \in I$, we have (cf. Prop. 3 of Sec. 3-2)

$$dN(\alpha'(t)) = \lambda(t)\alpha'(t).$$

It follows that the functions $u'(t)$, $v'(t)$ satisfy the system of equations

$$\begin{aligned} \frac{fF - eG}{EG - F^2}u' + \frac{gF - fG}{EG - F^2}v' &= \lambda u', \\ \frac{eF - fE}{EG - F^2}u' + \frac{fF - gE}{EG - F^2}v' &= \lambda v'. \end{aligned}$$

By eliminating λ in the above system, we obtain the *differential equation of the lines of curvature*,

$$(fE - eF)(u')^2 + (gE - eG)u'v' + (gF - fG)(v')^2 = 0,$$

which may be written, in a more symmetric way, as

$$\begin{vmatrix} (v')^2 & -u'v' & (u')^2 \\ E & F & G \\ e & f & g \end{vmatrix} = 0. \qquad \textbf{(8)}$$

Using the fact that the principal directions are orthogonal to each other, it follows easily from Eq. (8) that *a necessary and sufficient condition for the coordinate curves of a parametrization to be lines of curvature in a neighborhood of a nonumbilical point is that* $F = f = 0$.

Example 4 (*Surfaces of Revolution*)**.** Consider a surface of revolution parametrized by (cf. Example 4 of Sec. 2-3; we have replaced f and g by φ and ψ, respectively)

$$\mathbf{x}(u, v) = (\varphi(v)\cos u, \varphi(v)\sin u, \psi(v)),$$
$$0 < u < 2\pi, \quad a < v < b, \quad \varphi(v) \neq 0.$$

The coefficients of the first fundamental form are given by

$$E = \varphi^2, \quad F = 0, \quad G = (\varphi')^2 + (\psi')^2.$$

It is convenient to assume that the rotating curve is parametrized by arc length, that is, that

$$(\varphi')^2 + (\psi')^2 = G = 1.$$

The computation of the coefficients of the second fundamental form is straightforward and yields

$$\begin{aligned} e &= \frac{(\mathbf{x}_u, \mathbf{x}_v, \mathbf{x}_{uu})}{\sqrt{EG - F^2}} = \frac{1}{\sqrt{EG - F^2}} \begin{vmatrix} -\varphi \sin u & \varphi' \cos u & -\varphi \cos u \\ \varphi \cos u & \varphi' \sin u & -\varphi \sin u \\ 0 & \psi' & 0 \end{vmatrix} \\ &= -\varphi \psi' \\ f &= 0, \\ g &= \psi' \varphi'' - \psi'' \varphi'. \end{aligned}$$

Since $F = f = 0$, we conclude that the parallels ($v =$ const.) and the meridians ($u =$ const.) of a surface of revolution are lines of curvature of such a surface (this fact was used in Example 3).

Because

$$K = \frac{eg - f^2}{EG - F^2} = -\frac{\psi'(\psi'\varphi'' - \psi''\varphi')}{\varphi}$$

and φ is always positive, it follows that the parabolic points are given by either $\psi' = 0$ (the tangent line to the generator curve is perpendicular to the axis of rotation) or $\varphi'\psi'' - \psi'\varphi'' = 0$ (the curvature of the generator curve is zero). A point which satisfies both conditions is a planar point, since these conditions imply that $e = f = g = 0$.

It is convenient to put the Gaussian curvature in still another form. By differentiating $(\varphi')^2 + (\psi')^2 = 1$ we obtain $\varphi'\varphi'' = -\psi'\psi''$. Thus,

$$K = -\frac{\psi'(\psi'\varphi'' - \psi''\varphi')}{\varphi} = -\frac{(\psi')^2\varphi'' + (\varphi')^2\varphi''}{\varphi} = -\frac{\varphi''}{\varphi}. \tag{9}$$

Equation (9) is a convenient expression for the Gaussian curvature of a surface of revolution. It can be used, for instance, to determine the surfaces of revolution of constant Gaussian curvature (cf. Exercise 7).

To compute the principal curvatures, we first make the following general observation: *If a parametrization of a regular surface is such that $F = f = 0$,* *then the principal curvatures are given by* e/E and g/G. In fact, in this case, the Gaussian and the mean curvatures are given by (cf. Eqs. (4) and (5))

$$K = \frac{eg}{EG}, \quad H = \frac{1}{2}\frac{eG + gE}{EG}.$$

Since K is the product and $2H$ is the sum of the principal curvatures, our assertion follows at once.

Thus, the principal curvatures of a surface of revolution are given by

$$\frac{e}{E} = -\frac{\psi'\varphi}{\varphi^2} = -\frac{\psi'}{\varphi}, \quad \frac{g}{G} = \psi'\varphi'' - \psi''\varphi'; \tag{10}$$

hence, the mean curvature of such a surface is

$$H = \frac{1}{2}\frac{-\psi' + \varphi(\psi'\varphi'' - \psi''\varphi')}{\varphi}. \tag{11}$$

Example 5. Very often a surface is given as the graph of a differentiable function (cf. Prop. 1, Sec. 2-2) $z = h(x, y)$, where (x, y) belong to an open set $U \subset R^2$. It is, therefore, convenient to have at hand formulas for the relevant concepts in this case. To obtain such formulas let us parametrize the surface by

$$\mathbf{x}(u, v) = (u, v, h(u, v)), \quad (u, v) \in U,$$

where $u = x$, $v = y$. A simple computation shows that

$$\mathbf{x}_u = (1, 0, h_u), \quad \mathbf{x}_v = (0, 1, h_v), \quad \mathbf{x}_{uu} = (0, 0, h_{uu}),$$
$$\mathbf{x}_{uv} = (0, 0, h_{uu}), \quad \mathbf{x}_{vv} = (0, 0, h_{vv}).$$

Thus

$$N(x, y) = \frac{(-h_x, -h_y, 1)}{(1 + h_x^2 + h_y^2)^{1/2}}$$

is a unit normal field on the surface, and the coefficients of the second fundamental form in this orientation are given by

$$e = \frac{h_{xx}}{(1 + h_x^2 + h_y^2)^{1/2}},$$
$$f = \frac{h_{xy}}{(1 + h_x^2 + h_y^2)^{1/2}},$$
$$g = \frac{h_{yy}}{(1 + h_x^2 + h_y^2)^{1/2}}.$$

From the above expressions, any needed formula can be easily computed. For instance, from Eqs. (4) and (5) we obtain the Gaussian and mean curvatures:

$$K = \frac{h_{xx}h_{yy} - h_{xy}^2}{(1 + h_x^2 + h_y^2)^2},$$

$$2H = \frac{(1 + h_x^2)h_{yy} - 2h_x h_y h_{xy} + (1 + h_y^2)h_{xx}}{(1 + h_x^2 + h_y^2)^{3/2}}.$$

There is still another, perhaps more important, reason to study surfaces given by $z = h(x, y)$. It comes from the fact that locally any surface is the graph of a differentiable function (cf. Prop. 3, Sec. 2-2). Given a point p of a surface S, we can choose the coordinate axis of R^3 so that the origin O of the coordinates is at p and the z axis is directed along the positive normal of S at p (thus, the xy plane agrees with $T_p(S)$). It follows that a neighborhood of p in S can be represented in the form $z = h(x, y)$, $(x, y) \in U \subset R^2$, where U is an open set and h is a differentiable function (cf. Prop. 3, Sec. 2-2), with $h(0, 0) = 0$, $h_x(0, 0) = 0$, $h_y(0, 0) = 0$ (Fig. 3-19).

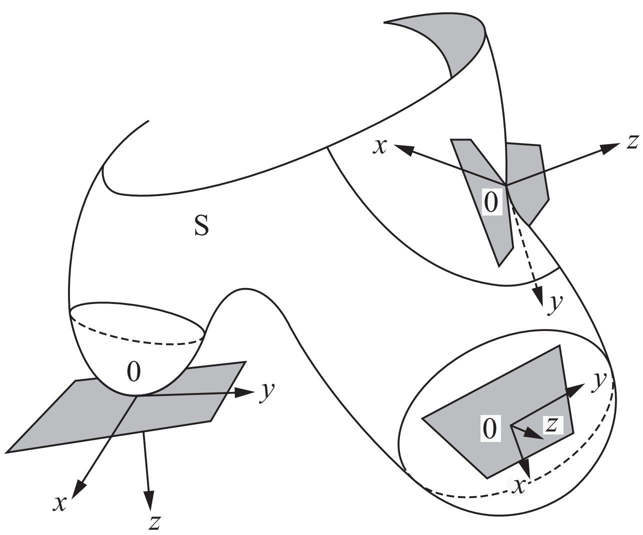


Figure 3-19. Each point of S has a neighborhood that can be written as $z = h(x, y)$.

The second fundamental form of S at p applied to the vector $(x, y) \in R^2$ becomes, in this case,

$$h_{xx}(0, 0)x^2 + 2h_{xy}(0, 0)xy + h_{yy}(0, 0)y^2.$$

In elementary calculus of two variables, the above quadratic form is known as the *Hessian* of h at $(0, 0)$. Thus, the Hessian of h at $(0, 0)$ is the second fundamental form of S at p.

Let us apply the above considerations to give a geometric interpretation of the Dupin indicatrix. With the notation as above, let $\epsilon > 0$ be a small number such that

$$C = \{(x, y) \in T_p(S); h(x, y) = \epsilon\}$$

is a regular curve (we may have to change the orientation of the surface to achieve $\epsilon > 0$). We want to show that if p is not a planar point, the curve C is "approximately" similar to the Dupin indicatrix of S at p (Fig. 3-20).

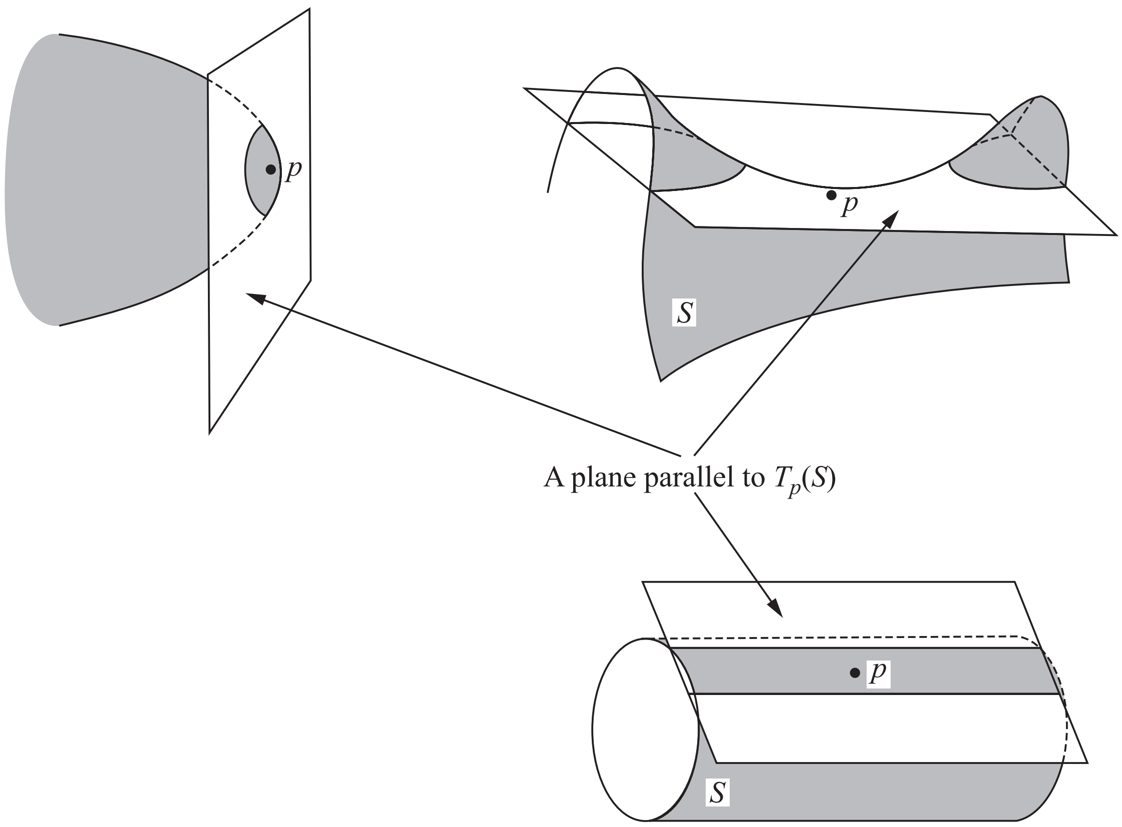


Figure 3-20

To see this, let us assume further that the x and y axes are directed along the principal directions, with the x axis along the direction of maximum principal curvature. Thus, $f = h_{xy}(0, 0) = 0$ and

$$k_1(p) = \frac{e}{E} = h_{xx}(0, 0), \quad k_2(p) = \frac{g}{G} = h_{yy}(0, 0).$$

By developing $h(x, y)$ into a Taylor's expansion about $(0, 0)$, and taking into account that $h_x(0, 0) = 0 = h_y(0, 0)$, we obtain

$$\begin{aligned} h(x, y) &= \tfrac{1}{2}(h_{xx}(0, 0)x^2 + 2h_{xy}(0, 0)xy + h_{yy}(0, 0)y^2) + R \\ &= \tfrac{1}{2}(k_1 x^2 + k_2 y^2) + R, \end{aligned}$$

where

$$\lim_{(x,y)\to(0,0)} \frac{R}{x^2 + y^2} = 0.$$

Thus, the curve C is given by

$$k_1x^2 + k_2y^2 + 2R = 2\epsilon.$$

Now, if p is not a planar point, we can consider $k_1x^2 + k_2y^2 = 2\epsilon$ as a first-order approximation of C in a neighborhood of p. By using the similarity transformation,

$$x = \bar{x}\sqrt{2\epsilon}, \quad y = \bar{y}\sqrt{2\epsilon},$$

we have that $k_1x^2 + k_2y^2 = 2\epsilon$ is transformed into the curve

$$k_1\bar{x}^2 + k_2\bar{y}^2 = 1,$$

which is the Dupin indicatrix at p. This means that *if* p *is a nonplanar point, the intersection with* S *of a plane parallel to* $\mathrm{T_p(S)}$ *and close to* p *is, in a first-order approximation, a curve similar to the Dupin indicatrix at* p.

If p is a planar point, this interpretation is no longer valid (cf. Exercise 11).

To conclude this section we shall give a geometrical interpretation of the Gaussian curvature in terms of the Gauss map $N: S \to S^2$. Actually, this was how Gauss himself introduced this curvature.

To do this, we first need a definition.

Let S and $\bar{S}$ be two oriented regular surfaces. Let $\varphi: S \to \bar{S}$ be a differentiable map and assume that for some $p \in S$, $d\varphi_p$ is nonsingular. We say that φ is *orientation-preserving* at p if given a positive basis $\{w_1, w_2\}$ in $T_p(S)$, then $\{d\varphi_p(w_1), d\varphi_p(w_2)\}$ is a positive basis in $T_{\varphi(p)}(\bar{S})$. If $\{d\varphi_p(w_1), d\varphi_p(w_2)\}$ is not a positive basis, we say that φ is *orientation-reversing* at p.

We now observe that both S and the unit sphere S^2 are embedded in R^3. Thus, an orientation N on S induces an orientation N in S^2. Let $p \in S$ be such that dN_p is nonsingular. Since for a basis $\{w_1, w_2\}$ in $T_p(S)$

$$dN_p(w_1) \wedge dN_p(w_2) = \det(dN_p)(w_1 \wedge w_2) = K w_1 \wedge w_2,$$

the Gauss map N will be orientation-preserving at $p \in S$ if $K(p) > 0$ and orientation-reversing at $p \in S$ if $K(p) < 0$. Intuitively, this means the following (Fig. 3-21): An orientation of $T_p(S)$ induces an orientation of small closed curves in S around p; the image by N of these curves will have the same or the opposite orientation to the initial one, depending on whether p is an elliptic or hyperbolic point, respectively.

To take this fact into account we shall make the convention that the area of the image by N of a region contained in a connected neighborhood $V \subset S$ where $K \neq 0$ is positive if $K > 0$ and negative if $K < 0$ (since V is connected, K does not change sign in V).

Now we can state the promised geometric interpretation of the Gaussian curvature K, for $K \neq 0$.

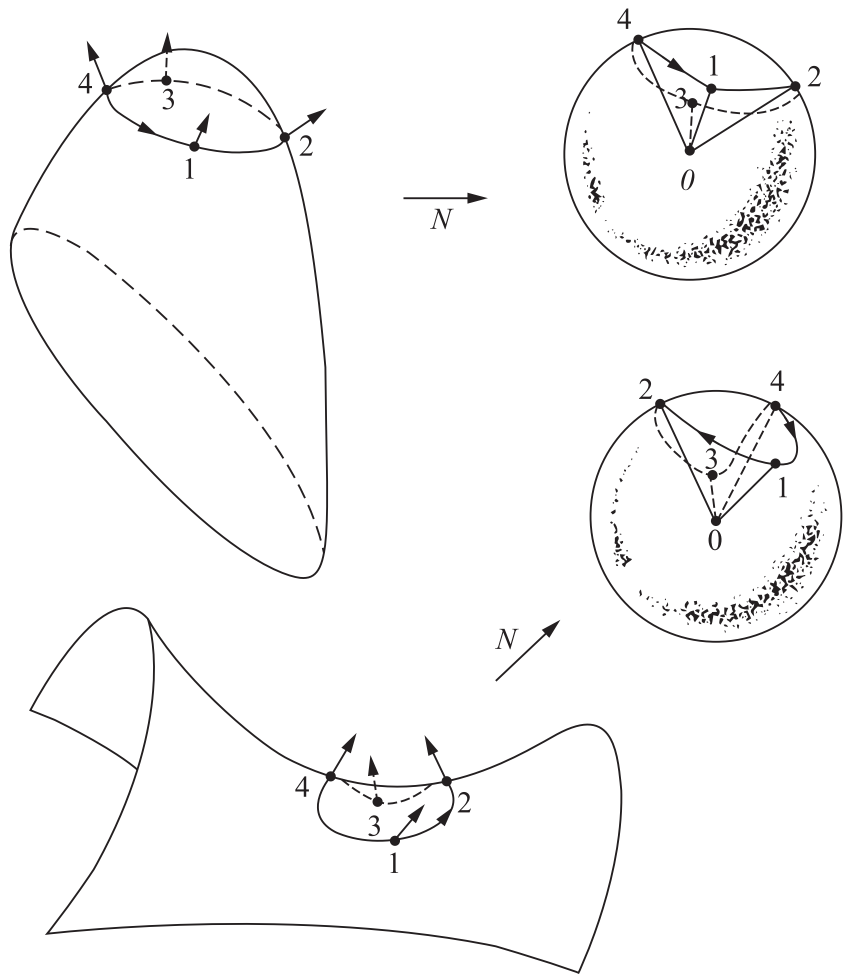


Figure 3-21. The Gauss map preserves orientation at an elliptic point and reverses it at a hyperbolic point.

PROPOSITION 2. *Let* p *be a point of a surface* S *such that the Gaussian curvature* $\mathrm{K(p)} \neq 0$, *and let* V *be a connected neighborhood of* p *where* K *does not change sign. Then*

$$\mathrm{K(p)} = \lim_{\mathrm{A}\to 0} \frac{\mathrm{A}'}{\mathrm{A}},$$

where A *is the area of a region* $\mathrm{B} \subset \mathrm{V}$ *containing* p, A′ *is the area of the image of* B *by the Gauss map* $\mathrm{N}\colon \mathrm{S} \to \mathrm{S}^2$, *and the limit is taken through a sequence of regions* $\mathrm{B_n}$ *that converges to* p, *in the sense that any sphere around* p *contains all* $\mathrm{B_n}$, *for* n *sufficiently large.*

Proof. The area A of B is given by (cf. Sec. 2-5)

$$A = \iint_R |\mathbf{x}_u \wedge \mathbf{x}_v|\, du\, dv,$$

where $\mathbf{x}(u, v)$ is a parametrization in p, whose coordinate neighborhood contains V (V can be assumed to be sufficiently small) and R is the region in the uv plane corresponding to B. The area A' of $N(B)$ is

$$A' = \iint_R |N_u \wedge N_v| \, du \, dv.$$

Using Eq. (1), the definition of K, and the above convention, we can write

$$A' = \iint_R K |\mathbf{x}_u \wedge \mathbf{x}_v| \, du \, dv. \tag{12}$$

Going to the limit and denoting also by R the area of the region R, we obtain

$$\lim_{A\to 0} \frac{A'}{A} = \lim_{R\to 0} \frac{A'/R}{A/R} = \frac{\displaystyle\lim_{R\to 0}(1/R) \iint_R K |\mathbf{x}_u \wedge \mathbf{x}_v| \, du \, dv}{\displaystyle\lim_{R\to 0}(1/R) \iint_R |\mathbf{x}_u \wedge \mathbf{x}_v| \, du \, dv}$$
$$= \frac{K |\mathbf{x}_u \wedge \mathbf{x}_v|}{|\mathbf{x}_u \wedge \mathbf{x}_v|} = K$$

(notice that we have used the mean value theorem for double integrals), and this proves the proposition. **Q.E.D.**

Remark. Comparing the proposition with the expression of the curvature

$$k = \lim_{s\to 0} \frac{\sigma}{s}$$

of a plane curve C at p (here s is the arc length of a small segment of C containing p, and σ is the arc length of its image in the indicatrix of tangents; cf. Exercise 3 of Sec. 1-5), we see that the Gaussian curvature K is the analogue, for surfaces, of the curvature k of plane curves.

EXERCISES

1. Show that at the origin $(0, 0, 0)$ of the hyperboloid $z = axy$ we have $K = -a^2$ and $H = 0$.

***2.** Determine the asymptotic curves and the lines of curvature of the helicoid $x = v \cos u$, $y = v \sin u$, $z = cu$, and show that its mean curvature is zero.

***3.** Determine the asymptotic curves of the catenoid

$$\mathbf{x}(u, v) = (\cosh v \cos u, \cosh v \sin u, v).$$

4. Determine the asymptotic curves and the lines of curvature of $z = xy$.

5. Consider the parametrized surface (Enneper's surface)

$$\mathbf{x}(u, v) = \left(u - \frac{u^3}{3} + uv^2, v - \frac{v^3}{3} + vu^2, u^2 - v^2\right)$$

and show that

a. The coefficients of the first fundamental form are

$$E = G = (1 + u^2 + v^2)^2, \quad F = 0.$$

b. The coefficients of the second fundamental form are

$$e = 2, \quad g = -2, \quad f = 0.$$

c. The principal curvatures are

$$k_1 = \frac{2}{(1 + u^2 + v^2)^2}, \quad k_2 = -\frac{2}{(1 + u^2 + v^2)^2}.$$

d. The lines of curvature are the coordinate curves.

e. The asymptotic curves are $u + v =$ const., $u - v =$ const.

6. (*A Surface with* $K \equiv -1$; *the Pseudosphere.*)

***a.** Determine an equation for the plane curve C, which is such that the segment of the tangent line between the point of tangency and some line r in the plane, which does not meet the curve, is constantly equal to 1 (this curve is called the *tractrix*; see Fig. 1-9).

b. Rotate the tractrix C about the line r; determine if the "surface" of revolution thus obtained (the *pseudosphere*; see Fig. 3-22) is regular and find out a parametrization in a neighborhood of a regular point.

c. Show that the Gaussian curvature of any regular point of the pseudosphere is -1.

7. (*Surfaces of Revolution with Constant Curvature.*) $(\varphi(v)\cos u, \varphi(v)\sin u, \psi(v))$, $\varphi \neq 0$ is given as a surface of revolution with constant Gaussian curvature K. To determine the functions φ and ψ, choose the parameter v in such a way that $(\varphi')^2 + (\psi')^2 = 1$ (geometrically, this means that v is the arc length of the generating curve $(\varphi(v), \psi(v))$). Show that

a. φ satisfies $\varphi'' + K\varphi = 0$ and ψ is given by $\psi = \int \sqrt{1 - (\varphi')^2}\, dv$; thus, $0 < u < 2\pi$, and the domain of v is such that the last integral makes sense.

b. All surfaces of revolution with constant curvature $K = 1$ which intersect perpendicularly the plane xOy are given by

$$\varphi(v) = C\cos v, \quad \psi(v) = \int_0^v \sqrt{1 - C^2\sin^2 v}\, dv,$$

where C is a constant ($C = \varphi(0)$). Determine the domain of v and draw a rough sketch of the profile of the surface in the xz plane for

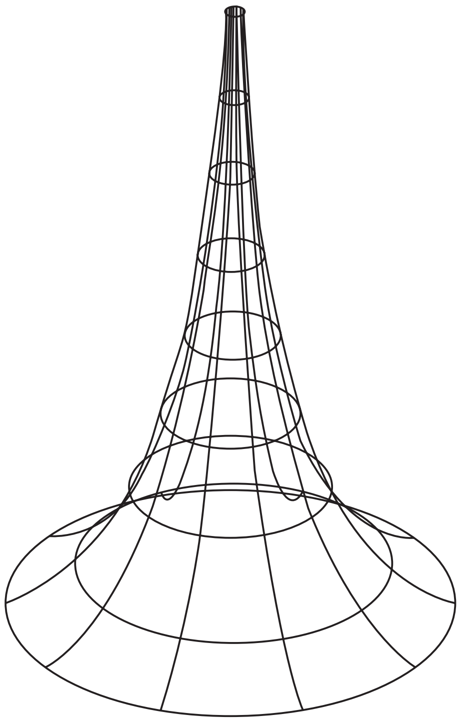


Figure 3-22. The pseudosphere.

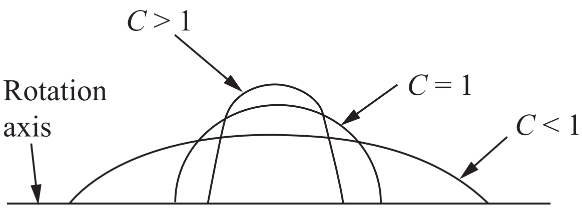


Figure 3-23

the cases $C = 1$, $C > 1$, $C < 1$. Observe that $C = 1$ gives a sphere (Fig. 3-23).

c. All surfaces of revolution with constant curvature $K = -1$ may be given by one of the following types:

1. $\varphi(v) = C\cosh v$,
   $\psi(v) = \int_0^v \sqrt{1 - C^2\sinh^2 v}\,dv$.
2. $\varphi(v) = C\sinh v$,
   $\psi(v) = \int_0^v \sqrt{1 - C^2\cosh^2 v}\,dv$.
3. $\varphi(v) = e^v$,
   $\psi(v) = \int_0^v \sqrt{1 - e^{2v}}\,dv$.

Determine the domain of v and draw a rough sketch of the profile of the surface in the xz plane.

d. The surface of type 3 in part c is the pseudosphere of Exercise 6.

e. The only surfaces of revolution with $K \equiv 0$ are the right circular cylinder, the right circular cone, and the plane.

8. (*Contact of Order* ≥ 2 *of Surfaces*.) Two surfaces S and $\bar{S}$, with a common point p, have *contact of order* ≥ 2 at p if there exist parametrizations $\mathbf{x}(u, v)$ and $\bar{\mathbf{x}}(u, v)$ in p of S and $\bar{S}$, respectively, such that

$$\mathbf{x}_u = \bar{\mathbf{x}}_u, \quad \mathbf{x}_v = \bar{\mathbf{x}}_v, \quad \mathbf{x}_{uu} = \bar{\mathbf{x}}_{uu}, \quad \mathbf{x}_{uv} = \bar{\mathbf{x}}_{uv}, \quad \mathbf{x}_{vv} = \bar{\mathbf{x}}_{vv}$$

at p. Prove the following:

***a.** Let S and $\bar{S}$ have contact of order ≥ 2 at p; $\mathbf{x}: U \to S$ and $\bar{\mathbf{x}}: U \to \bar{S}$ be arbitrary parametrizations in p of S and $\bar{S}$, respectively; and $f: V \subset R^3 \to R$ be a differentiable function in a neighborhood V of p in R^3. Then the partial derivatives of order ≤ 2 of $f \circ \bar{\mathbf{x}}: U \to R$ are zero in $\bar{\mathbf{x}}^{-1}(p)$ if and only if the partial derivatives of order ≤ 2 of $f \circ \mathbf{x}: U \to R$ are zero in $\mathbf{x}^{-1}(p)$.

***b.** Let S and $\bar{S}$ have contact of order ≥ 2 at p. Let $z = f(x, y)$, $z = \bar{f}(x, y)$ be the equations, in a neighborhood of p, of S and $\bar{S}$, respectively, where the xy plane is the common tangent plane at $p = (0, 0)$. Then the function $f(x, y) - \bar{f}(x, y)$ has all partial derivatives of order ≤ 2, at $(0, 0)$, equal to zero.

c. Let p be a point in a surface $S \subset R^3$. Let $Oxyz$ be a Cartesian coordinate system for R^3 such that $O = p$ and the xy plane is the tangent plane of S at p. Show that the paraboloid

$$z = \tfrac{1}{2}(x^2 f_{xx} + 2xyf_{xy} + y^2 f_{yy}), \tag{*}$$

obtained by neglecting third- and higher-order terms in the Taylor development around $p = (0, 0)$, has contact of order ≥ 2 at p with S (the surface $(*)$ is called the *osculating paraboloid* of S at p).

***d.** If a paraboloid (the degenerate cases of plane and parabolic cylinder are included) has contact of order ≥ 2 with a surface S at p, then it is the osculating paraboloid of S at p.

e. If two surfaces have contact of order ≥ 2 at p, then the osculating paraboloids of S and $\bar{S}$ at p coincide. Conclude that the Gaussian and mean curvatures of S and $\bar{S}$ at p are equal.

f. The notion of contact of order ≥ 2 is invariant by diffeomorphisms of R^3; that is, if S and $\bar{S}$ have contact of order ≥ 2 at p and $\varphi: R^3 \to R^3$ is a diffeomorphism, then $\varphi(S)$ and $\varphi(\bar{S})$ have contact of order ≥ 2 at $\varphi(p)$.

g. If S and $\bar{S}$ have contact of order ≥ 2 at p, then

$$\lim_{r \to 0} \frac{d}{r^2} = 0,$$

where d is the length of the segment cut by the surfaces in a straight line normal to $T_p(S) = T_p(\bar{S})$, which is at a distance r from p.

9. (*Contact of Curves.*) Define contact of order $\geq n$ (n integer ≥ 1) for regular curves in R^3 with a common point p and prove that

a. The notion of contact of order $\geq n$ is invariant by diffeomorphisms.

b. Two curves have contact of order ≥ 1 at p if and only if they are tangent at p.

10. (*Contact of Curves and Surfaces.*) A curve C and a surface S, which have a common point p, have contact of order $\geq n$ (n integer ≥ 1) at p if there exists a curve $\bar{C} \subset S$ passing through p such that C and $\bar{C}$ have contact of order $\geq n$ at p. Prove that

a. If $f(x, y, z) = 0$ is a representation of a neighborhood of p in S and $\alpha(t) = (x(t), y(t), z(t))$ is a parametrization of C in p, with $\alpha(0) = p$, then C and S have contact of order $\geq n$ if and only if

$$f(x(0), y(0), z(0)) = 0, \quad \frac{df}{dt} = 0, \ldots, \frac{d^n f}{dt^n} = 0,$$

where the derivatives are computed for $t = 0$.

b. If a plane has contact of order ≥ 2 with a curve C at p, then this is the osculating plane of C at p.

c. If a sphere has contact of order ≥ 3 with a curve C at p, and $\alpha(s)$ is a parametrization by arc length of this curve, with $\alpha(0) = p$, then the center of the sphere is given by

$$\alpha(0) + \frac{1}{k}n + \frac{k'}{k^2\tau}b.$$

Such a sphere is called the *osculating sphere* of C at p.

11. Consider the monkey saddle S of Example 2. Construct the Dupin indicatrix at $p = (0, 0, 0)$ using the definition of Sec. 3-2, and compare it with the curve obtained as the intersection of S with a plane parallel to $T_p(S)$ and close to p. Why are they not “approximately similar” (cf. Example 5 of Sec. 3-3)? Go through the argument of Example 5 of Sec. 3-3 and point out where it breaks down.

12. Consider the parametrized surface

$$\mathbf{x}(u, v) = \left(\sin u \cos v, \sin u \sin v, \cos u + \log \tan \frac{u}{2} + \varphi(v)\right),$$

where φ is a differentiable function. Prove that

a. The curves $v = $ const. are contained in planes which pass through the z axis and intersect the surface under a constant angle θ given by

$$\cos\theta = \frac{\varphi'}{\sqrt{1 + (\varphi')^2}}.$$

Conclude that the curves $v = $ const. are lines of curvature of the surface.

b. The length of the segment of a tangent line to a curve $v = \text{const.}$, determined by its point of tangency and the z axis, is constantly equal to 1. Conclude that the curves $v = \text{const.}$ are tractrices (cf. Exercise 6).

13. Let $F\colon R^3 \to R^3$ be the map (a similarity) defined by $F(p) = cp$, $p \in R^3$, c a positive constant. Let $S \subset R^3$ be a regular surface and set $F(S) = \bar{S}$. Show that $\bar{S}$ is a regular surface, and find formulas relating the Gaussian and mean curvatures, K and H, of S with the Gaussian and mean curvatures, $\bar{K}$ and $\bar{H}$, of $\bar{S}$.

14. Consider the surface obtained by rotating the curve $y = x^3$, $-1 < x < 1$, about the line $x = 1$. Show that the points obtained by rotation of the origin $(0, 0)$ of the curve are planar points of the surface.

***15.** Give an example of a surface which has an isolated parabolic point p (that is, no other parabolic point is contained in some neighborhood of p).

***16.** Show that a surface which is compact (i.e., it is bounded and closed in R^3) has an elliptic point.

17. Define Gaussian curvature for a nonorientable surface. Can you define mean curvature for a nonorientable surface?

18. Show that the Möbius strip of Fig. 3-1 can be parametrized by

$$\mathbf{x}(u, v) = \left(\left(2 - v\sin\frac{u}{2}\right)\sin u,\ \left(2 - v\sin\frac{u}{2}\right)\cos u,\ v\cos\frac{u}{2}\right)$$

and that its Gaussian curvature is

$$K = -\frac{1}{\{\frac{1}{4}v^2 + (2 - v\sin(u/2))^2\}^2}.$$

***19.** Obtain the asymptotic curves of the one-sheeted hyperboloid $x^2 + y^2 - z^2 = 1$.

20. Determine the umbilical points of the elipsoid

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

***21.** Let S be a surface with orientation N. Let $V \subset S$ be an open set in S and let $f\colon V \subset S \to R$ be any nowhere-zero differentiable function in V. Let v_1 and v_2 be two differentiable (tangent) vector fields in V such that at each point of V, v_1 and v_2 are orthonormal and $v_1 \wedge v_2 = N$.

a. Prove that the Gaussian curvature K of V is given by

$$K = \frac{\langle d(fN)(v_1) \wedge d(fN)(v_2), fN\rangle}{f^3}.$$

The virtue of this formula is that by a clever choice of f we can often simplify the computation of K, as illustrated in part b.

b. Apply the above result to show that if f is the restriction of

$$\sqrt{\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}}$$

to the ellipsoid

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

then the Gaussian curvature of the ellipsoid is

$$K = \frac{1}{a^2b^2c^2}\frac{1}{f^4}.$$

22. (*The Hessian.*) Let $h\colon S \to R$ be a differentiable function on a surface S, and let $p \in S$ be a critical point of h (i.e., $dh_p = 0$). Let $w \in T_p(S)$ and let

$$\alpha\colon (-\epsilon, \epsilon) \to S$$

be a parametrized curve with $\alpha(0) = p$, $\alpha'(0) = w$. Set

$$H_p h(w) = \left.\frac{d^2(h \circ \alpha)}{dt^2}\right|_{t=0}.$$

a. Let $\mathbf{x}\colon U \to S$ be a parametrization of S at p, and show that (the fact that p is a critical point of h is essential here)

$$H_p h(u'\mathbf{x}_u + v'\mathbf{x}_v) = h_{uu}(p)(u')^2 + 2h_{uv}(p)u'v' + h_{vv}(p)(v')^2.$$

Conclude that $H_p h\colon T_p(S) \to R$ is a well-defined (i.e., it does not depend on the choice of $\mathbf{x}$) quadratic form on $T_p(S)$. $H_p h$ is called the *Hessian* of h at p.

b. Let $h\colon S \to R$ be the height function of S relative to $T_p(S)$; that is, $h(q) = \langle q - p, N(p)\rangle$, $q \in S$. Verify that p is a critical point of h and thus that the Hessian $H_p h$ is well defined. Show that if $w \in T_p(S)$, $|w| = 1$, then

$$H_p h(w) = \text{normal curvature at } p \text{ in the direction of } w.$$

Conclude that *the Hessian at* p *of the height function relative to* $\mathrm{T_p(S)}$ *is the second fundamental form of* S *at* p.

23. (*Morse Functions on Surfaces.*) A critical point $p \in S$ of a differentiable function $h: S \to R$ is *nondegenerate* if the self-adjoint linear map $A_p h$ associated to the quadratic form $H_p h$ (cf. the appendix to Chap. 3) is nonsingular (here $H_p h$ is the Hessian of h at p; cf. Exercise 22). Otherwise, p is a *degenerate* critical point. A differentiable function on S is a *Morse function* if all its critical points are nondegenerate. Let $h_r: S \subset R^3 \to R$ be the distance function from S to r; i.e.,

$$h_r(q) = \sqrt{\langle q - r, q - r\rangle}, \quad q \in S, \quad r \in R^3, \quad r \notin S.$$

a. Show that $p \in S$ is a critical point of h_r if and only if the straight line pr is normal to S at p.

b. Let p be a critical point of $h_r: S \to R$. Let $w \in T_p(S)$, $|w| = 1$, and let $\alpha: (-\epsilon, \epsilon) \to S$ be a curve parametrized by arc length with $\alpha(0) = p$, $\alpha'(0) = w$. Prove that

$$H_p h_r(w) = \frac{1}{h_r(p)} - k_n,$$

where k_n is the normal curvature at p along the direction of w. Conclude that the orthonormal basis $\{e_1, e_2\}$, where e_1 and e_2 are along the principal directions of $T_p(S)$, diagonalizes the self-adjoint linear map $A_p h_r$. Conclude further that p is a degenerate critical point of h_r if and only if either $h_r(p) = 1/k_1$ or $h_r(p) = 1/k_2$, where k_1 and k_2 are the principal curvatures at p.

c. Show that the set

$$B = \{r \in R^3;\ h_r \text{ is a Morse function}\}$$

is a dense set in R^3; here dense in R^3 means that in each neighborhood of a given point of R^3 there exists a point of B (*this shows that on any regular surface there are "many" Morse functions*).

24. (*Local Convexity and Curvature*). A surface $S \subset R^3$ is *locally convex* at a point $p \in S$ if there exists a neighborhood $V \subset S$ of p such that V is contained in one of the closed half-spaces determined by $T_p(S)$ in R^3. If, in addition, V has only one common point with $T_p(S)$, then S is called *strictly locally convex* at p.

a. Prove that S is strictly locally convex at p if the principal curvatures of S at p are nonzero with the same sign (that is, the Gaussian curvature $K(p)$ satisfies $K(p) > 0$).

b. Prove that if S is locally convex at p, then the principal curvatures at p do not have different signs (thus, $K(p) \geq 0$).

c. To show that $K \geq 0$ does not imply local convexity, consider the surface $f(x, y) = x^3(1 + y^2)$, defined in the open set $U = \{(x, y) \in R^2; y^2 < \frac{1}{2}\}$. Show that the Gaussian curvature of this surface is nonnegative on U and yet the surface is not locally convex at $(0, 0) \in U$ (a deep theorem, due to R. Sacksteder, implies that such an example cannot be extended to the entire R^2 if we insist on keeping the curvature nonnegative; cf. Remark 3 of Sec. 5-6).

***d.** The example of part c is also very special in the following local sense. Let p be a point in a surface S, and assume that there exists a neighborhood $V \subset S$ of p such that the principal curvatures on V do not have different signs (this does not happen in the example of part c). Prove that S is locally convex at p.

3-4. Vector Fields†

In this section we shall use the fundamental theorems of ordinary differential equations (existence, uniqueness, and dependence on the initial conditions) to prove the existence of certain coordinate systems on surfaces.

If the reader is willing to assume the results of Corollaries 2, 3, and 4 at the end of this section (which can be understood without reading the section), this material may be omitted on a first reading.

We shall begin with a geometric presentation of the material on differential equations that we intend to use.

A *vector field* in an open set $U \subset R^2$ is a map which assigns to each $q \in U$ a vector $w(q) \in R^2$. The vector field w is said to be *differentiable* if writing $q = (x, y)$ and $w(q) = (a(x, y), b(x, y))$, the functions a and b are differentiable functions in U.

Geometrically, the definition corresponds to assigning to each point $(x, y) \in U$ a vector with coordinates $a(x, y)$ and $b(x, y)$ which vary differentiably with (x, y) (Fig. 3-24).

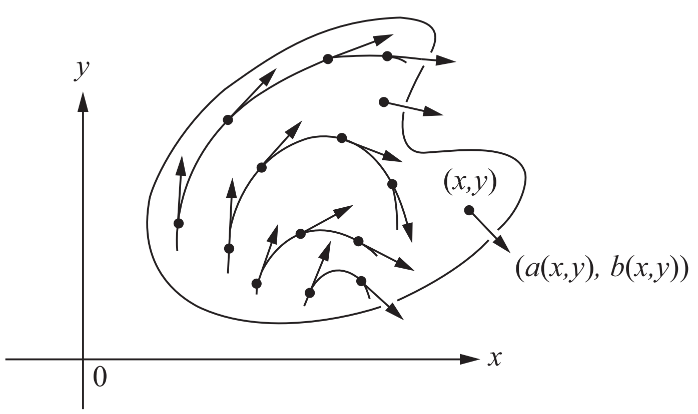


Figure 3-24

In what follows we shall consider only differentiable vector fields.

In Fig. 3-25 some examples of vector fields are shown.

†This section may be omitted on a first reading.

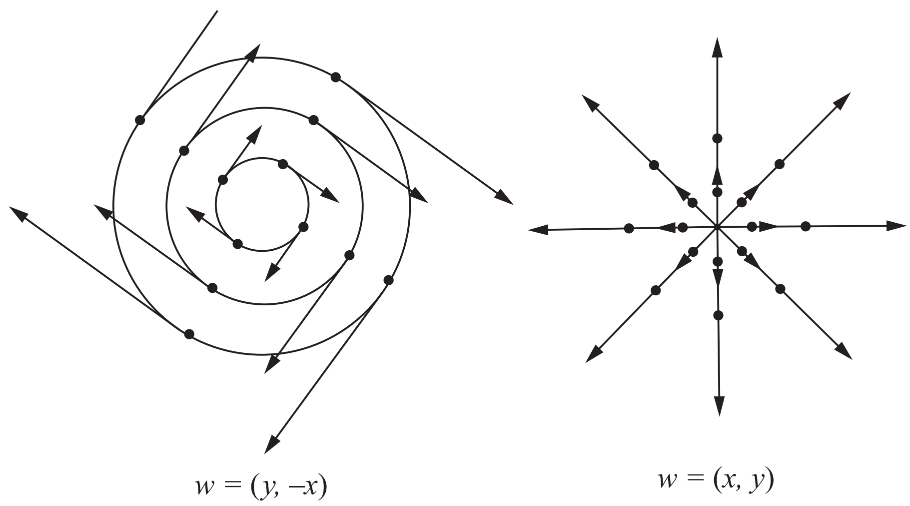


Figure 3-25

Given a vector field w, it is natural to ask whether there exists a *trajectory* of this field, that is, whether there exists a differentiable parametrized curve $\alpha(t) = (x(t), y(t))$, $t \in I$, such that $\alpha'(t) = w(\alpha(t))$.

For instance, a trajectory, passing through the point (x_0, y_0), of the vector field $w(x, y) = (x, y)$ is the straight line $\alpha(t) = (x_0 e^t, y_0 e^t)$, $t \in R$, and a trajectory of $w(x, y) = (y, -x)$, passing through (x_0, y_0), is the circle $\beta(t) = (r \sin t, r \cos t)$, $t \in R$, $r^2 = x_0^2 + y_0^2$.

In the language of ordinary differential equations, one says that the vector field w determines a system of *differential equations*,

$$\begin{aligned} \frac{dx}{dt} &= a(x, y), \\ \frac{dy}{dt} &= b(x, y), \end{aligned} \tag{1}$$

and that a trajectory of w is a *solution* to Eq. (1).

The fundamental theorem of (local) existence and uniqueness of solutions of Eq. (1) is equivalent to the following statement on trajectories (in what follows, the letters I and J will denote open intervals of the line R, containing the origin $0 \in R$).

THEOREM 1. *Let* w *be a vector field in an open set* $\mathrm{U} \subset \mathrm{R}^2$. *Given* $\mathrm{p} \in \mathrm{U}$, *there exists a trajectory* $\alpha\colon \mathrm{I} \to \mathrm{U}$ *of* w (*i.e.*, $\alpha'(\mathrm{t}) = \mathrm{w}(\alpha(\mathrm{t}))$, $\mathrm{t} \in \mathrm{I}$) *with* $\alpha(0) = \mathrm{p}$. *This trajectory is unique in the following sense: Any other trajectory* $\beta\colon \mathrm{J} \to \mathrm{U}$ *with* $\beta(0) = \mathrm{p}$ *agrees with* α *in* $\mathrm{I} \cap \mathrm{J}$.

An important complement to Theorem 1 is the fact that the trajectory passing through p "varies differentiably with p." This idea can be made precise as follows.

THEOREM 2. *Let* w *be a vector field in an open set* $\mathrm{U} \subset \mathrm{R}^2$*. For each* $\mathrm{p} \in \mathrm{U}$ *there exist a neighborhood* $\mathrm{V} \subset \mathrm{U}$ *of* p*, an interval* I*, and a mapping* $\alpha\colon \mathrm{V} \times \mathrm{I} \to \mathrm{U}$ *such that*

1. *For a fixed* $\mathrm{q} \in \mathrm{V}$*, the curve* $\alpha(\mathrm{q}, \mathrm{t})$*,* $\mathrm{t} \in \mathrm{I}$*, is the trajectory of* w *passing through* q*; that is,*

$$\alpha(\mathrm{q}, 0) = \mathrm{q}, \quad \frac{\partial\alpha}{\partial \mathrm{t}}(\mathrm{q}, \mathrm{t}) = \mathrm{w}(\alpha(\mathrm{q}, \mathrm{t})).$$

2. α *is differentiable.*

Geometrically Theorem 2 means that all trajectories which pass, for $t = 0$, in a certain neighborhood V of p may be "collected" into a single differentiable map. It is in this sense that we say that the trajectories depend differentiably on p (Fig. 3-26).

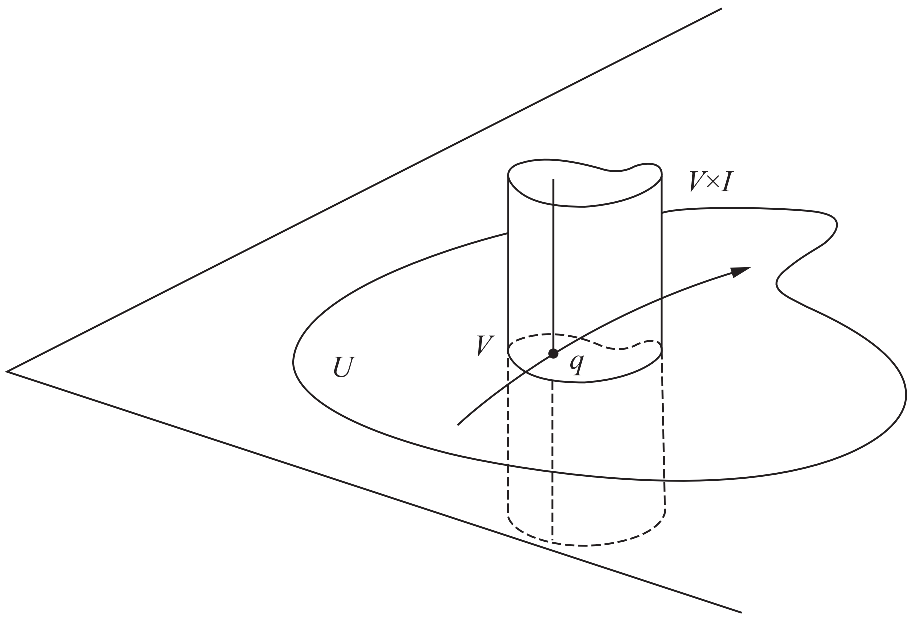


Figure 3-26

The map α is called the (*local*) *flow* of w at p.

Theorems 1 and 2 will be assumed in this book; for a proof, one can consult, for instance, W. Hurewicz, *Lectures on Ordinary Differential Equations*, M.I.T. Press, Cambridge, Mass., 1958, Chap. 2. For our purposes, we need the following consequence of these theorems.

LEMMA. *Let* w *be a vector field in an open set* $\mathrm{U} \subset \mathrm{R}^2$ *and let* $\mathrm{p} \in \mathrm{U}$ *be such that* $\mathrm{w}(\mathrm{p}) \neq 0$*. Then there exist a neighborhood* $\mathrm{W} \subset \mathrm{U}$ *of* p *and a differentiable function* $\mathrm{f}\colon \mathrm{W} \to \mathrm{R}$ *such that* f *is constant along each trajectory of* w *and* $d\mathrm{f}_{\mathrm{q}} \neq 0$ *for all* $\mathrm{q} \in \mathrm{W}$*.*

Proof. Choose a Cartesian coordinate system in R^2 such that $p = (0, 0)$ and $w(p)$ is in the direction of the x axis. Let $\alpha\colon V \times I \to U$ be the local flow at p, $V \subset U$, $t \in I$, and let $\tilde{\alpha}$ be the restriction of α to the rectangle

$$(V \times I) \cap \{(x, y, t) \in R^3; x = 0\}.$$

(See Fig. 3-27.) By the definition of local flow, $d\tilde{\alpha}_p$ maps the unit vector of the t axis into w and maps the unit vector of the y axis into itself. Therefore, $d\tilde{\alpha}_p$ is nonsingular. It follows that there exists a neighborhood $W \subset U$ of p, where $\tilde{\alpha}^{-1}$ is defined and differentiable. The projection of $\tilde{\alpha}^{-1}(x, y)$ onto the y axis is a differentiable function $\xi = f(x, y)$, which has the same value ξ for all points of the trajectory passing through $(0, \xi)$. Since $d\tilde{\alpha}_p$ is nonsingular, W may be taken sufficiently small so that $df_q \neq 0$ for all $q \in W$. f is therefore the required function. **Q.E.D.**

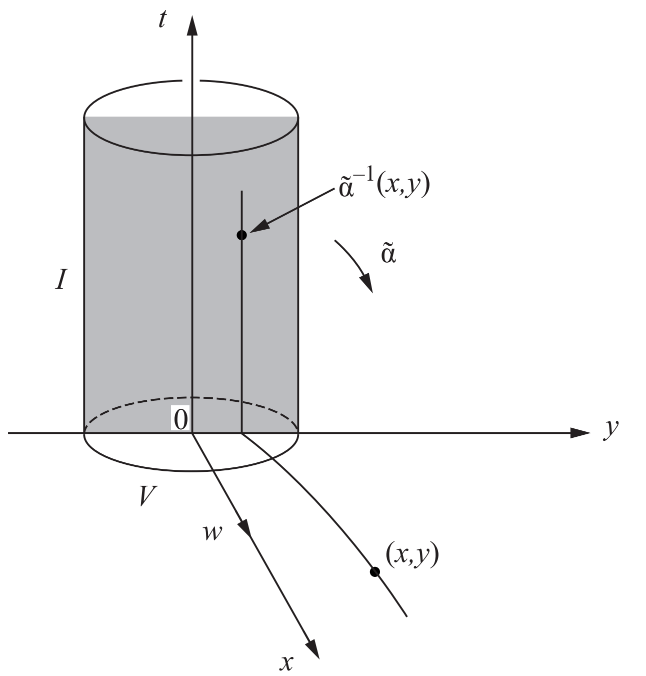


Figure 3-27

The function f of the above lemma is called a (local) *first integral* of w in a neighborhood of p. For instance, if $w(x, y) = (y, -x)$ is defined in R^2, a first integral $f: R^2 - \{(0, 0)\} \to R$ is $f(x, y) = x^2 + y^2$.

Closely associated with the concept of vector field is the concept of field of directions.

A *field of directions* r in an open set $U \subset R^2$ is a correspondence which assigns to each $p \in U$ a line $r(p)$ in R^2 passing through p. r is said to be *differentiable* at $p \in U$ if there exists a nonzero differentiable vector field w, defined in a neighborhood $V \subset U$ of p, such that for each $q \in V$, $w(q) \neq 0$ is a *basis* of $r(q)$; r is *differentiable in* U if it is differentiable for every $p \in U$.

To each nonzero differentiable vector field w in $U \subset R^2$, there corresponds a differentiable field of directions given by $r(p) =$ line generated by $w(p)$, $p \in U$.

By its very definition, each differentiable field of directions gives rise, locally, to a nonzero differentiable vector field. This, however, is not true globally, as is shown by the field of directions in $R^2 - \{(0, 0)\}$ given by the tangent lines to the curves of Fig. 3-28; any attempt to orient these curves in order to obtain a differentiable nonzero vector field leads to a contradiction.

A regular connected curve $C \subset U$ is an *integral curve* of a field of directions r defined in $U \subset R^2$ if $r(q)$ is the tangent line to C at q for all $q \in C$.

By what has been seen previously, it is clear that given a differentiable field of directions r in an open set $U \subset R^2$, there passes, for each $q \in U$, an integral curve C of r; C agrees locally with the trace of a trajectory through q of the vector field determined in U by r. In what follows, we shall consider only differentiable fields of directions and shall omit, in general, the word differentiable.

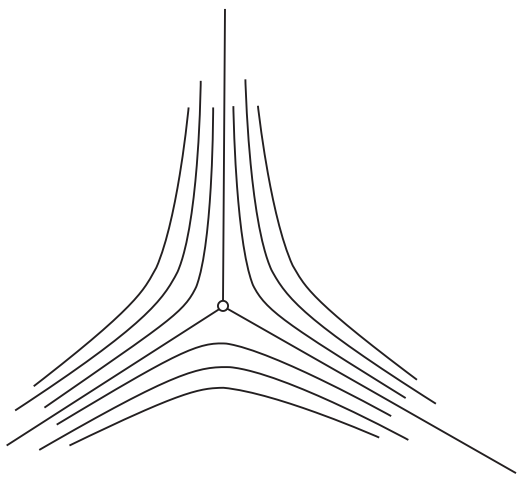


Figure 3-28. A nonorientable field of directions in $R^2 - \{(0, 0)\}$.

A natural way of describing a field of directions is as follows. We say that two nonzero vectors w_1 and w_2 at $q \in R^2$ are *equivalent* if $w_1 = \lambda w_2$ for some $\lambda \in R$, $\lambda \neq 0$. Two such vectors represent the same straight line passing through q, and, conversely, if two nonzero vectors belong to the same straight line passing through q, they are equivalent. Thus, a field of directions r on an open set $U \subset R^2$ can be given by assigning to each $q \in U$ a pair of real numbers (r_1, r_2) (the coordinates of a nonzero vector belonging to r), where we consider the pairs (r_1, r_2) and $(\lambda r_1, \lambda r_2)$, $\lambda \neq 0$, as equivalent.

In the language of differential equations, a field of directions r is usually given by

$$a(x, y)\frac{dx}{dt} + b(x, y)\frac{dy}{dt} = 0, \qquad \textbf{(2)}$$

which simply means that at a point $q = (x, y)$ we associate the line passing through q that contains the vector $(b, -a)$ or any of its nonzero multiples (Fig. 3-29). The trace of the trajectory of the vector field $(b, -a)$ is an integral curve of r. Because the parametrization plays no role in the above considerations, it is often used, instead of Eq. (2), the expression

$$a\,dx + b\,dy = 0$$

with the same meaning as before.

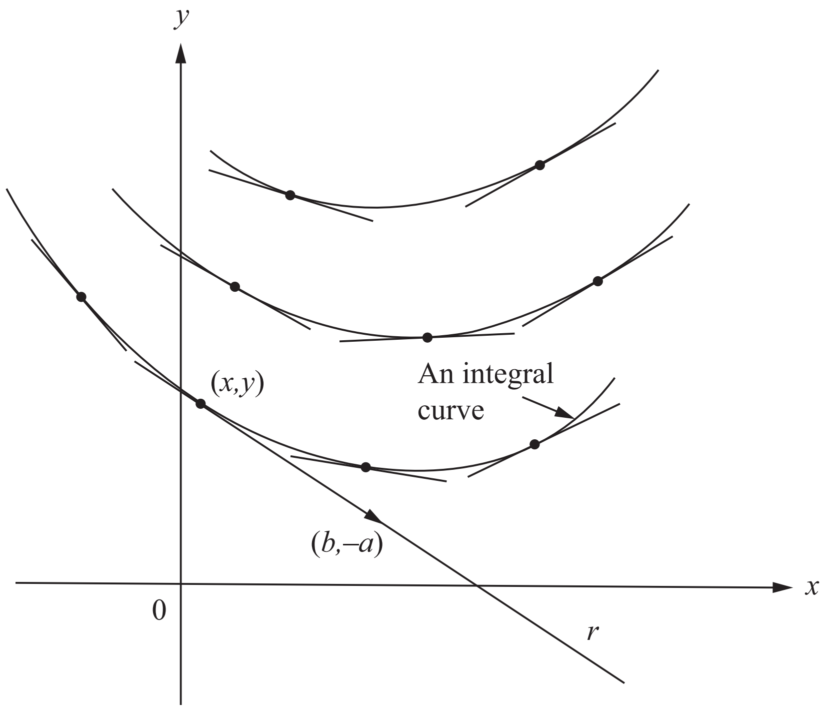


Figure 3-29. The differential equation $adx + bdy = 0$.

The ideas introduced above belong to the domain of the local facts of R^2, which depend only on the "differentiable structure" of R^2. They can, therefore, be transported to a regular surface, without further difficulties, as follows.

DEFINITION 1. *A* vector field w *in an open set* $U \subset S$ *of a regular surface* S *is a correspondence which assigns to each* $p \in U$ *a vector* $w(p) \in T_p(S)$. *The vector field* w *is differentiable at* $p \in U$ *if, for some parametrization* $\mathbf{x}(u, v)$ *at* p, *the functions* a(u, v) *and* b(u, v) *given by*

$$w(p) = a(u, v)\mathbf{x}_u + b(u, v)\mathbf{x}_v$$

are differentiable functions at p*; it is clear that this definition does not depend on the choice of* $\mathbf{x}$.

We can define, similarly, trajectories, field of directions, and integral curves. Theorems 1 and 2 and the lemma above extend easily to the present situation; up to a change of R^2 by S, the statements are exactly the same.

Example 1. A vector field in the usual torus T is obtained by parametrizing the meridians of T by arc length and defining $w(p)$ as the velocity vector of the meridian through p (Fig. 3-30). Notice that $|w(p)| = 1$ for all $p \in T$. It is left as an exercise (Exercise 2) to verify that w is differentiable.

Example 2. A similar procedure, this time on the sphere S^2 and using the semimeridians of S^2, yields a vector field w defined in the sphere minus the two poles N and S. To obtain a vector field defined in the whole sphere,

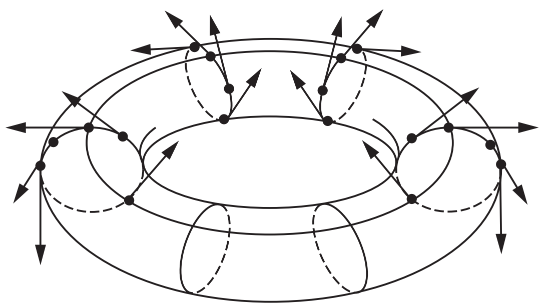


Figure 3-30

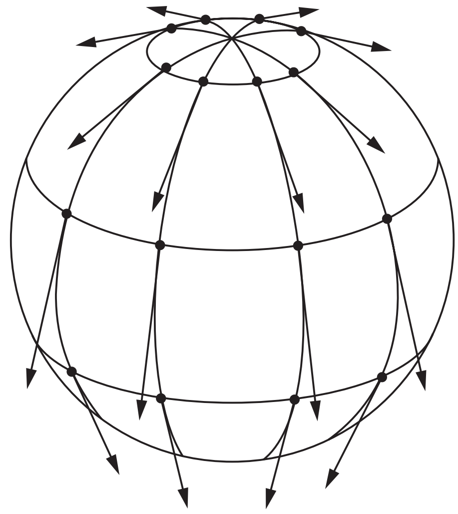


Figure 3-31

reparametrize all the semimeridians by the same parameter t, $-1 < t < 1$, and define $v(p) = (1 - t^2)w(p)$ for $p \in S^2 - \{N\} \cup \{S\}$ and $v(N) = v(S) = 0$ (Fig. 3-31).

Example 3. Let $S = \{(x, y, z) \in R^3; z = x^2 - y^2\}$ be the hyperbolic paraboloid. The intersection with S of the planes $z =$ const. $\neq 0$ determines a family of curves $\{C_\alpha\}$ such that through each point of $S - \{(0, 0, 0)\}$ there passes one curve C_α. The tangent lines to such curves give a differentiable field of directions r on $S - \{(0, 0, 0)\}$. We want to find a field of directions r' on $S - \{(0, 0, 0)\}$ that is orthogonal to r at each point and to determine the integral curves of r'. r' is called the *orthogonal field* to r, and its integral curves are called the *orthogonal family* of r (cf. Exercise 15, Sec. 2-5).

We begin by parametrizing S by

$$\mathbf{x}(u, v) = (u, v, u^2 - v^2), \quad u = x, \quad v = y.$$

The family $\{C_\alpha\}$ is given by $u^2 - v^2 =$ const. $\neq 0$ (or rather by the image under $\mathbf{x}$ of this set). If $u'\mathbf{x}_u + v'\mathbf{x}_v$ is a tangent vector of a regular parametrization of some curve C_α, we obtain, by differentiating $u^2 - v^2 =$ const.,

$$2uu' - 2vv' = 0.$$

Thus, $(u', v') = (-v, -u)$. It follows that r is given, in the parametrization $\mathbf{x}$, by the pair (v, u) or any of its nonzero multiples.

Now, let $(a(u, v), b(u, v))$ be an expression for the orthogonal field r', in the parametrization $\mathbf{x}$. Since

$$E = 1 + 4u^2, \quad F = -4uv, \quad G = 1 + 4v^2,$$

and r' is orthogonal to r at each point, we have

$$Eav + F(bv + au) + Gbu = 0$$

or

$$(1+4u^2)av - 4uv(bv+au) + (1+4v^2)bu = 0.$$

It follows that

$$va + ub = 0. \tag{3}$$

This determines the pair (a, b) at each point, up to a nonzero multiple, and hence the field r'.

To find the integral curves of r', let $u'\mathbf{x}_u + v'\mathbf{x}_v$ be a tangent vector of some regular parametrization of an integral curve of r'. Then (u', v') satisfies Eq. (3); that is,

$$vu' + uv' = 0$$

or

$$uv = \text{const.}$$

It follows that the orthogonal family of $\{C_\alpha\}$ is given by the intersections with S of the hyperbolic cylinders $xy = \text{const.} \neq 0$.

The main result of this section is the following theorem.

THEOREM. *Let* w_1 *and* w_2 *be two vector fields in an open set* $U \subset S$, *which are linearly independent at some point* $p \in U$. *Then it is possible to parametrize a neighborhood* $V \subset U$ *of* p *in such a way that for each* $q \in V$ *the coordinate curves of this parametrization passing through* q *are tangent to the lines determined by* $w_1(q)$ *and* $w_2(q)$.

Proof. Let W be a neighborhood of p where the first integrals f_1 and f_2 of w_1 and w_2, respectively, are defined. Define a map φ: $W \to R^2$ by

$$\varphi(q) = (f_1(q), f_2(q)), \quad q \in W.$$

Since f_1 is constant on the trajectories of w_1 and $(df_1) \neq 0$, we have at p

$$d\varphi_p(w_1) = ((df_1)_p(w_1), (df_2)_p(w_1)) = (0, a),$$

where $a = (df_2)_p(w_1) \neq 0$, since w_1 and w_2 are linearly independent. Similarly,

$$d\varphi_p(w_2) = (b, 0),$$

where $b = (df_1)_p(w_2) \neq 0$.

It follows that $d\varphi_p$ is nonsingular, and hence that φ is a local diffeomorphism. There exists, therefore, a neighborhood $\bar{U} \subset R^2$ of $\varphi(p)$ which is mapped diffeomorphically by $\mathbf{x} = \varphi^{-1}$ onto a neighborhood $V = \mathbf{x}(\bar{U})$ of p; that is, $\mathbf{x}$ is a parametrization of S at p, whose coordinate curves

$$f_1(q) = \text{const.}, \quad f_2(q) = \text{const.},$$

are tangent at q to the lines determined by $w_1(q)$, $w_2(q)$, respectively. **Q.E.D.**

It should be remarked that the theorem does not imply that the coordinate curves can be so parametrized that their velocity vectors are $w_1(q)$ and $w_2(q)$. The statement of the theorem applies to the coordinate curves as regular (point set) curves; more precisely, we have

COROLLARY 1. *Given two fields of directions* r *and* r′ *in an open set* $\mathrm{U} \subset \mathrm{S}$ *such that at* $\mathrm{p} \in \mathrm{U}$, $\mathrm{r(p)} \neq \mathrm{r'(p)}$, *there exists a parametrization* $\mathbf{x}$ *in a neighborhood of* p *such that the coordinate curves of* $\mathbf{x}$ *are the integral curves of* r *and* r′.

A first application of the above theorem is the proof of the existence of an orthogonal parametrization at any point of a regular surface.

COROLLARY 2. *For all* $\mathrm{p} \in \mathrm{S}$ *there exists a parametrization* $\mathbf{x}(\mathrm{u}, \mathrm{v})$ *in a neighborhood* V *of* p *such that the coordinate curves* $\mathrm{u} = \mathrm{const.}$, $\mathrm{v} = \mathrm{const.}$ *intersect orthogonally for each* $\mathrm{q} \in \mathrm{V}$ *(such an* $\mathbf{x}$ *is called an orthogonal parametrization).*

Proof. Consider an arbitrary parametrization $\bar{\mathbf{x}}: \bar{U} \to S$ at p, and define two vector fields $w_1 = \bar{\mathbf{x}}_{\bar{u}}$, $w_2 = -(\bar{F}/\bar{E})\bar{\mathbf{x}}_{\bar{u}} + \bar{\mathbf{x}}_{\bar{v}}$ in $\bar{\mathbf{x}}(\bar{U})$, where $\bar{E}$, $\bar{F}$, $\bar{G}$ are the coefficients of the first fundamental form in $\bar{\mathbf{x}}$. Since $w_1(q)$, $w_2(q)$ are orthogonal vectors, for each $q \in \bar{\mathbf{x}}(\bar{U})$, an application of the theorem yields the required parametrization. **Q.E.D.**

A second application of the theorem (more precisely, of Corollary 1) is the existence of coordinates given by the asymptotic and principal directions.

As we have seen in Sec. 3-3, the asymptotic curves are solutions of

$$e(u')^2 + 2fu'v' + g(v')^2 = 0.$$

In a neighborhood of a hyperbolic point p, we have $eg - f^2 < 0$. Rotate the plane uv so that $e(p) > 0$. Then the left-hand side of the above equation can be decomposed into two distinct linear factors, yielding

$$(Au' + Bv')(Au' + Dv') = 0, \tag{4}$$

where the coefficients are determined by

$$A^2 = e, \quad A(B + D) = 2f, \quad BD = g.$$

The above system of equations has real solutions, since $eg - f^2 < 0$. Thus, Eq. (4) gives rise to two equations:

$$Au' + Bv' = 0, \tag{4a}$$

$$Au' + Dv' = 0. \tag{4b}$$

Each of these equations determines a differentiable field of directions (for instance, Eq. (4a) determines the direction r which contains the nonzero vector $(B, -A)$), and at each point of the neighborhood in question the directions given by Eqs. (4a) and (4b) are distinct. By applying Corollary 1, we see that it is possible to parametrize a neighborhood of p in such a way that the coordinate curves are the integral curves of Eqs. (4a) and (4b). In other words,

COROLLARY 3. *Let* $\mathrm{p} \in \mathrm{S}$ *be a hyperbolic point of* S. *Then it is possible to parametrize a neighborhood of* p *in such a way that the coordinate curves of this parametrization are the asymptotic curves of* S.

Example 4. An almost trivial example, but one which illustrates the mechanism of the above method, is given by the hyperbolic paraboloid $z = x^2 - y^2$. As usual we parametrize the entire surface by

$$\mathbf{x}(u, v) = (u, v, u^2 - v^2).$$

A simple computation shows that

$$e = \frac{2}{(1 + 4u^2 + 4v^2)^{1/2}}, \quad f = 0, \quad g = -\frac{2}{(1 + 4u^2 + 4v^2)^{1/2}}.$$

Thus, the equation of the asymptotic curves can be written as

$$\frac{2}{(1 + 4u^2 + 4v^2)^{1/2}}((u')^2 - (v')^2) = 0,$$

which can be factored into two linear equations and give the two fields of directions:

$$r_1\colon\ u' + v' = 0,$$
$$r_2\colon\ u' - v' = 0.$$

The integral curves of these fields of directions are given by the two families of curves:

$$r_1\colon\ u + v = \text{const.},$$
$$r_2\colon\ u - v = \text{const.}$$

Now, the functions $f_1(u, v) = u + v$, $f_2(u, v) = u - v$ are clearly first integrals of the vector fields associated to r_1 and r_2, respectively. Thus, by setting

$$\bar{u} = u + v, \quad \bar{v} = u - v,$$

we obtain a new parametrization for the entire surface $z = x^2 - y^2$ in which the coordinate curves are the asymptotic curves of the surface.

In this particular case, the change of parameters holds for the entire surface. In general, it may fail to be globally one-to-one, even if the whole surface consists only of hyperbolic points.

Similarly, in a neighborhood of a nonumbilical point of S, it is possible to decompose the differential equation of the lines of curvature into distinct linear factors. By an analogous argument we obtain

COROLLARY 4. *Let* p $\in$ S *be a nonumbilical point of* S. *Then it is possible to parametrize a neighborhood of* p *in such a way that the coordinate curves of this parametrization are the lines of curvature of* S.

EXERCISES

1. Prove that the differentiability of a vector field does not depend on the choice of a coordinate system.
2. Prove that the vector field obtained on the torus by parametrizing all its meridians by arc length and taking their tangent vectors (Example 1) is differentiable.
3. Prove that a vector field w defined on a regular surface $S \subset R^3$ is differentiable if and only if it is differentiable as a map $w\colon S \to R^3$.
4. Let S be a surface and $\mathbf{x}\colon U \to S$ be a parametrization of S. Then

$$a(u, v)u' + b(u, v)v' = 0,$$

where a and b are differentiable functions, determines a field of directions r on $\mathbf{x}(U)$, namely, the correspondence which assigns to each $\mathbf{x}(u, v)$ the straight line containing the vector $b\mathbf{x}_u - a\mathbf{x}_v$. Show that a necessary and sufficient condition for the existence of an orthogonal field r' on $\mathbf{x}(U)$ (cf. Example 3) is that both functions

$$Eb - Fa, \qquad Fb - Ga$$

are nowhere simultaneously zero (here E, F, and G are the coefficients of the first fundamental form in $\mathbf{x}$) and that r' is then determined by

$$(Eb - Fa)u' + (Fb - Ga)v' = 0.$$

5. Let S be a surface and $\mathbf{x}\colon U \to S$ be a parametrization of S. If $ac - b^2 < 0$, show that

$$a(u, v)(u')^2 + 2b(u, v)u'v' + c(u, v)(v')^2 = 0$$

can be factored into two distinct equations, each of which determines a field of directions on $\mathbf{x}(U) \subset S$. Prove that these two fields of directions are orthogonal if and only if

$$Ec - 2Fb + Ga = 0.$$

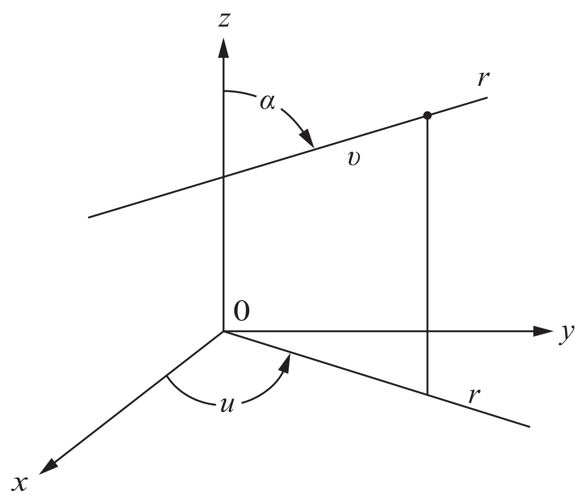


Figure 3-32

6. A straight line r meets the z axis and moves in such a way that it makes a constant angle $\alpha \neq 0$ with the z axis and each of its points describes a helix of pitch $c \neq 0$ about the z axis. The figure described by r is the trace of the parametrized surface (see Fig. 3-32)

$$\mathbf{x}(u, v) = (v \sin\alpha \cos u,\ v \sin\alpha \sin u,\ v \cos\alpha + cu).$$

$\mathbf{x}$ is easily seen to be a regular parametrized surface (cf. Exercise 13, Sec. 2-5). Restrict the parameters (u, v) to an open set U so that $\mathbf{x}(U) = S$ is a regular surface (cf. Prop. 2, Sec. 2-3).

a. Find the orthogonal family (cf. Example 3) to the family of coordinate curves $u =$ const.

b. Use the curves $u =$ const. and their orthogonal family to obtain an orthogonal parametrization for S. Show that in the new parameters $(\bar{u}, \bar{v})$ the coefficients of the first fundamental form are

$$\bar{G} = 1, \quad \bar{F} = 0, \quad \bar{E} = \{c^2 + (\bar{v} - c\bar{u}\cos\alpha)^2\}\sin^2\alpha.$$

7. Define the *derivative* w(f) *of a differentiable function* $\mathrm{f}\colon \mathrm{U} \subset \mathrm{S} \to \mathrm{R}$ *relative to a vector field* w *in* U *by*

$$\mathrm{w(f)(q)} = \frac{\mathrm{d}}{dt}(\mathrm{f} \circ \alpha)\Big|_{\mathrm{t}=0}, \quad \mathrm{q} \in \mathrm{U},$$

where $\alpha\colon I \to S$ is a curve such that $\alpha(0) = q$, $\alpha'(0) = w(q)$. Prove that

a. w is differentiable in U if and only if $w(f)$ is differentiable for all differentiable f in U.

b. Let λ and μ be real numbers and $g\colon U \subset S \to R$ be a differentiable function on U; then

$$w(\lambda f + \mu f) = \lambda w(f) + \mu w(f),$$
$$w(fg) = w(f)g + f w(g).$$

8. Show that if w is a differentiable vector field on a surface S and $w(p) \neq 0$ for some $p \in S$, then it is possible to parametrize a neighborhood of p by $\mathbf{x}(u, v)$ in such a way that $\mathbf{x}_u = w$.

9. a. Let $A\colon V \to W$ be a nonsingular linear map of vector spaces V and W of dimension 2 and endowed with inner products $\langle\ ,\ \rangle$ and $(\ ,\ )$, respectively. A is a *similitude* if there exists a real number $\lambda \neq 0$ such that $(Av_1, Av_2) = \lambda\langle v_1, v_2\rangle$ for all vectors $v_1, v_2 \in V$. Assume that A is not a similitude and show that there exists a *unique* pair of orthonormal vectors e_1 and e_2 in V such that Ae_1, Ae_2 are orthogonal in W.

b. Use part a to prove *Tissot's theorem*: Let $\varphi\colon U_1 \subset S_1 \to S_2$ be a diffeomorphism from a neighborhood U_1 of a point p of a surface S_1 into a surface S_2. Assume that the linear map $d\varphi$ is nowhere a similitude. Then it is possible to parametrize a neighborhood of p in S_1 by an orthogonal parametrization $\mathbf{x}_1\colon U \to S_1$ in such a way that $\varphi \circ \mathbf{x}_1 = \mathbf{x}_2\colon U \to S_2$ is also an orthogonal parametrization in a neighborhood of $\varphi(p) \in S_2$.

10. Let T be the torus of Example 6 of Sec. 2-2 and define a map $\varphi\colon R^2 \to T$ by

$$\varphi(u, v) = ((r\cos u + a)\cos v, (r\cos u + a)\sin v, r\sin u),$$

where u and v are the Cartesian coordinates of R^2. Let $u = at$, $v = bt$ be a straight line in R^2, passing by $(0, 0) \in R^2$, and consider the curve in T $\alpha(t) = \varphi(at, bt)$. Prove that

a. φ is a local diffeomorphism.

b. The curve $\alpha(t)$ is a regular curve; $\alpha(t)$ is a closed curve if and only if b/a is a rational number.

***c.** If b/a is irrational, the curve $\alpha(t)$ is dense in T; that is, in each neighborhood of a point $p \in T$ there exists a point of $\alpha(t)$.

***11.** Use the local uniqueness of trajectories of a vector field w in $U \subset S$ to prove the following result. Given $p \in U$, there exists a unique trajectory $\alpha\colon I \to U$ of w, with $\alpha(0) = p$, which is *maximal* in the following sense: Any other trajectory $\beta\colon J \to U$, with $\beta(0) = p$, is the restriction of α to J (i.e., $J \subset I$ and $\alpha|J = \beta$).

***12.** Prove that if w is a differentiable vector field on a compact surface S and $\alpha(t)$ is the maximal trajectory of w with $\alpha(0) = p \in S$, then $\alpha(t)$ is defined for all $t \in R$.

13. Construct a differentiable vector field on an open disk of the plane (which is not compact) such that a maximal trajectory $\alpha(t)$ is not defined for all $t \in R$ (this shows that the compactness condition of Exercise 12 is essential).

3-5. Ruled Surfaces and Minimal Surfaces†

In differential geometry one finds quite a number of special cases (surfaces of revolution, parallel surfaces, ruled surfaces, minimal surfaces, etc.) which may either become interesting in their own right (like minimal surfaces), or give a beautiful example of the power and limitations of differentiable methods in geometry. According to the spirit of this book, we have so far treated these special cases in examples and exercises.

It might be useful, however, to present some of these topics in more detail. We intend to do that now. We shall use this section to develop the theory of ruled surfaces and to give an introduction to the theory of minimal surfaces. Throughout the section it will be convenient to use the notion of parametrized surface defined in Sec. 2-3.

If the reader wishes so, the entire section or one of its topics may be omitted. Except for a reference to Sec. A in Example 6 of Sec. B, the two topics are independent and their results will not be used in any essential way in this book.

A. Ruled Surfaces

A (differentiable) *one-parameter family of* (straight) *lines* $\{\alpha(t), w(t)\}$ is a correspondence that assigns to each $t \in I$ a point $\alpha(t) \in R^3$ and a vector $w(t) \in R^3$, $w(t) \neq 0$, so that both $\alpha(t)$ and $w(t)$ depend differentiably on t. For each $t \in I$, the line L_t which passes through $\alpha(t)$ and is parallel to $w(t)$ is called *the line of the family at* t.

Given a one-parameter family of lines $\{\alpha(t), w(t)\}$, the parametrized surface

$$\mathbf{x}(t, v) = \alpha(t) + v w(t), \quad t \in I, \quad v \in R,$$

is called the *ruled surface* generated by the family $\{\alpha(t), w(t)\}$. The lines L_t are called the *rulings*, and the curve $\alpha(t)$ is called a *directrix* of the surface $\mathbf{x}$. Sometimes we use the expression ruled surface to mean the trace of $\mathbf{x}$.

†This section may be omitted on a first reading.

It should be noticed that we also allow $\mathbf{x}$ to have singular points, that is, points (t, v) where $\mathbf{x}_t \wedge x_v = 0$.

Example 1. The simplest examples of ruled surfaces are the tangent surfaces to a regular curve (cf. Example 4, Sec. 2-3), the cylinders and the cones. A *cylinder* is a ruled surface generated by a one-parameter family of lines $\{\alpha(t), w(t)\}$, $t \in I$, where $\alpha(I)$ is contained in a plane P and $w(t)$ is parallel to a fixed direction in R^3 (Fig. 3-33(a)). A *cone* is a ruled surface generated by a family $\{\alpha(t), w(t)\}$, $t \in I$, where $\alpha(I) \subset P$ and the rulings L_t all pass through a point $p \notin P$ (Fig. 3-33(b)).

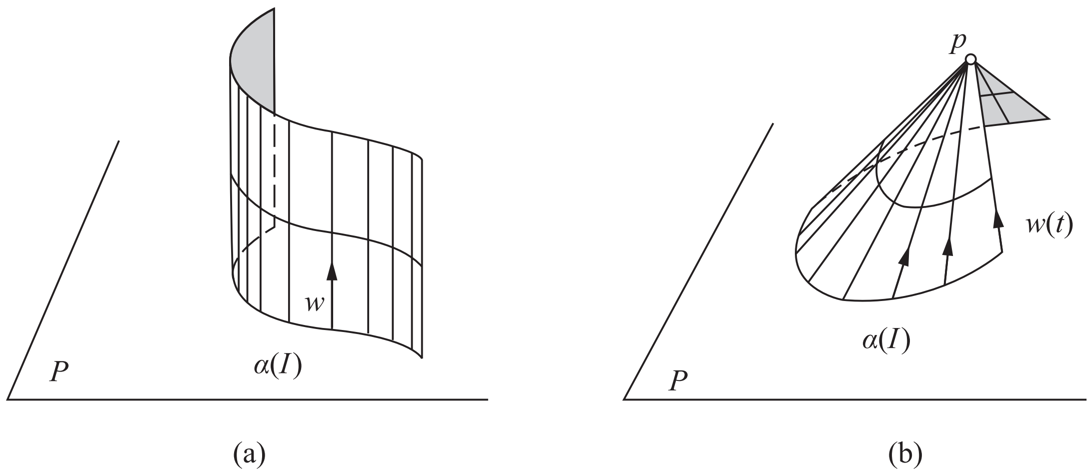


Figure 3-33

Example 2. Let S^1 be the unit circle $x^2 + y^2 = 1$ in the xy plane, and let $\alpha(s)$ be a parametrization of S^1 by arc length. For each s, let $w(s) = \alpha'(s) + e_3$, where e_3 is the unit vector of the z axis (Fig. 3-34). Then

$$\mathbf{x}(s, v) = \alpha(s) + v(\alpha'(s) + e_3)$$

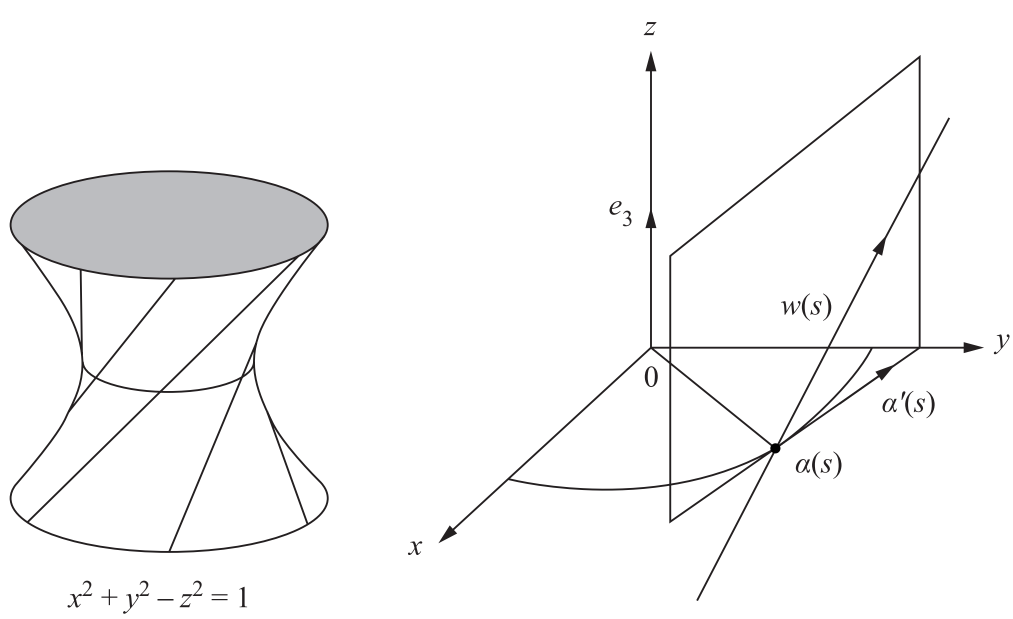


Figure 3-34. $x^2 + y^2 - z^2 = 1$ as a ruled surface.

is a ruled surface. It can be put into a more familiar form if we write

$$\mathbf{x}(s, v) = (\cos s - v \sin s, \sin s + v \cos s, v)$$

and notice that $x^2 + y^2 - z^2 = 1 + v^2 - v^2 = 1$. This shows that the trace of $\mathbf{x}$ is a hyperboloid of revolution.

It is interesting to observe that if we take $w(s) = -\alpha'(s) + e_3$, we again obtain the same surface. This shows that the hyperboloid of revolution has two sets of rulings.

We have defined ruled surfaces in such a way that allows the appearance of singularities. This is necessary if we want to include tangent surfaces and cones. We shall soon show, at least for ruled surfaces that satisfy some reasonable condition, that the singularities of such a surface (if any) will concentrate along a curve of this surface.

We shall now start the study of general ruled surfaces. We can assume, without loss of generality, that $|w(t)| = 1$, $t \in I$. To be able to develop the theory, we need the nontrivial assumption that $w'(t) \neq 0$ for all $t \in I$. If the zeros of $w'(t)$ are isolated, we can divide our surface into pieces such that the theory can be applied to each of them. However, if the zeros of $w'(t)$ have cluster points, the situation may become complicated and will not be treated here.

The assumption $w'(t) \neq 0$, $t \in I$, is usually expressed by saying that the ruled surface $\mathbf{x}$ is *noncylindrical.*

Unless otherwise stated, we shall assume that

$$\mathbf{x}(t, v) = \alpha(t) + v w(t) \qquad \textbf{(1)}$$

is a noncylindrical ruled surface with $|w(t)| = 1$, $t \in I$. Notice that the assumption $|w(t)| \equiv 1$ implies that $\langle w(t), w'(t)\rangle = 0$ for all $t \in I$.

We first want to find a parametrized curve $\beta(t)$ such that $\langle \beta'(t), w'(t)\rangle = 0$, $t \in I$, and $\beta(t)$ lies on the trace of $\mathbf{x}$; that is,

$$\beta(t) = \alpha(t) + u(t) w(t), \qquad \textbf{(2)}$$

for some real-valued function $u = u(t)$. Assuming the existence of such a curve β, one obtains

$$\beta' = \alpha' + u'w + uw';$$

hence, since $\langle w, w'\rangle = 0$,

$$0 = \langle \beta', w'\rangle = \langle \alpha', w'\rangle + u\langle w', w'\rangle.$$

It follows that $u = u(t)$ is given by

$$u = -\frac{\langle \alpha', w'\rangle}{\langle w', w'\rangle} \qquad \textbf{(3)}$$

Thus, if we define $\beta(t)$ by Eqs. (2) and (3), we obtain the required curve.

We shall now show that the curve β does not depend on the choice of the directrix α for the ruled surface. β is then called the *line of striction*, and its points are called the *central points* of the ruled surface.

To prove our claim, let $\bar{\alpha}$ be another directrix of the ruled surface; that is, let, for all (t, v),

$$\mathbf{x}(t, v) = \alpha(t) + vw(t) = \bar{\alpha}(t) + sw(t) \tag{4}$$

for some function $s = s(v)$. Then, from Eqs. (2) and (3) we obtain

$$\beta - \bar{\beta} = (\alpha - \bar{\alpha}) + \frac{\langle \bar{\alpha}' - \alpha', w' \rangle}{\langle w', w' \rangle} w,$$

where $\bar{\beta}$ is the line of striction corresponding to $\bar{\alpha}$. On the other hand, Eq. (4) implies that

$$\alpha - \bar{\alpha} = (s - v)w(t).$$

Thus,

$$\beta - \bar{\beta} = \left\{ (s - v) + \frac{\langle (v - s)w', w' \rangle}{\langle w', w' \rangle} \right\} w = 0,$$

since $\langle w, w' \rangle = 0$. This proves our claim.

We now take the line of striction as the directrix of the ruled surface and write it as follows:

$$\mathbf{x}(t, u) = \beta(t) + uw(t). \tag{5}$$

With this choice, we have

$$\mathbf{x}_t = \beta' + uw', \quad \mathbf{x}_u = w$$

and

$$\mathbf{x}_t \wedge \mathbf{x}_u = \beta' \wedge w + uw' \wedge w.$$

Since $\langle w', w \rangle = 0$ and $\langle w', \beta' \rangle = 0$, we conclude that $\beta' \wedge w = \lambda w'$ for some function $\lambda = \lambda(t)$. Thus,

$$\begin{aligned} |\mathbf{x}_t \wedge \mathbf{x}_u|^2 &= |\lambda w' + uw' \wedge w|^2 \\ &= \lambda^2 |w'|^2 + u^2 |w'|^2 = (\lambda^2 + u^2)|w'|^2. \end{aligned}$$

It follows that the only singular points of the ruled surface (5) are along the line of striction $u = 0$, and they will occur if and only if $\lambda(t) = 0$. Observe also that

$$\lambda = \frac{(\beta', w, w')}{|w'|^2},$$

where, as usual, (β', w, w') is a short for $\langle \beta' \wedge w, w' \rangle$.

Let us compute the Gaussian curvature of the surface (5) at its regular points. Since

$$\mathbf{x}_{tt} = \beta'' + uw'', \quad \mathbf{x}_{tu} = w', \quad \mathbf{x}_{uu} = 0,$$

we have, for the coefficients of the second fundamental form,

$$g = 0, \quad f = \frac{(\mathbf{x}_t, \mathbf{x}_u, \mathbf{x}_{ut})}{|\mathbf{x}_t \wedge \mathbf{x}_u|} = \frac{(\beta', w, w')}{|\mathbf{x}_t \wedge \mathbf{x}_u|};$$

hence (since $g = 0$ we do not need the value of e to compute K),

$$K = \frac{eg - f^2}{EG - F^2} = -\frac{\lambda^2 |w'|^4}{(\lambda^2 + u^2)^2 |w'|^4} = -\frac{\lambda^2}{(\lambda^2 + u^2)^2}. \qquad \textbf{(6)}$$

This shows that, at regular points, *the Gaussian curvature* K *of a ruled surface satisfies* K ≤ 0, *and* K *is zero only along those rulings which meet the line of striction at a singular point.*

Equation (6) allows us to give a geometric interpretation of the (regular) central points of a ruled surface. Indeed, the points of a ruling, except perhaps the central point, are regular points of the surface. If $\lambda \neq 0$, the function $|K(u)|$ is a continuous function on the ruling and, by Eq. (6), the central point is characterized by the fact that $|K(u)|$ has a maximum there.

For another geometrical interpretation of the line of striction see Exercise 4.

We also remark that the curvature K takes up the same values at points on a ruling that are symmetric relative to the central point (this justifies the name central).

The function $\lambda(t)$ is called the *distribution parameter* of $\mathbf{x}$. Since the line of striction is independent of the choice of the directrix, it follows that the same holds for λ. If $\mathbf{x}$ is regular, we have the following interpretation of λ. The normal vector to the surface at (t, u) is

$$N(t, u) = \frac{\mathbf{x}_t \wedge \mathbf{x}_u}{|\mathbf{x}_t \wedge \mathbf{x}_u|} = \frac{\lambda w' + uw' \wedge w}{\sqrt{\lambda^2 + u^2}|w'|}.$$

On the other hand ($\lambda \neq 0$),

$$N(t, 0) = \frac{w'}{|w'|}\frac{\lambda}{|\lambda|}$$

Therefore, if θ is the angle formed by $N(t, u)$ and $N(t, 0)$,

$$\tan\theta = \frac{u}{|\lambda|}. \qquad \textbf{(7)}$$

Thus, if θ *is the angle which the normal vector at a point of a ruling makes with the normal vector at the central point of this ruling, then* $\tan\theta$ *is proportional* *to the distance between these two points, and the coefficient of proportionality is the inverse of the distribution parameter.*

Example 3. Let S be the hyperbolic paraboloid $z = kxy$, $k \neq 0$. To show that S is a ruled surface, we observe that the lines $y = z/tk$, $x = t$, for each $t \neq 0$ belong to S. If we take the intersection of this family of lines with the plane $z = 0$, we obtain the curve $x = t$, $y = 0$, $z = 0$. Taking this curve as directrix and vectors $w(t)$ parallel to the lines $y = z/tk$, $x = t$, we obtain

$$\alpha(t) = (t, 0, 0), \quad w(t) = \frac{(0, 1, kt)}{\sqrt{1 + k^2t^2}}.$$

This gives a ruled surface (Fig. 3-35)

$$\mathbf{x}(t, v) = \alpha(t) + vw(t) = \left(t, \frac{v}{\sqrt{1 + k^2t^2}}, \frac{vkt}{\sqrt{1 + k^2t^2}}\right), \quad t \in R, v \in R,$$

the trace of which clearly agrees with S.

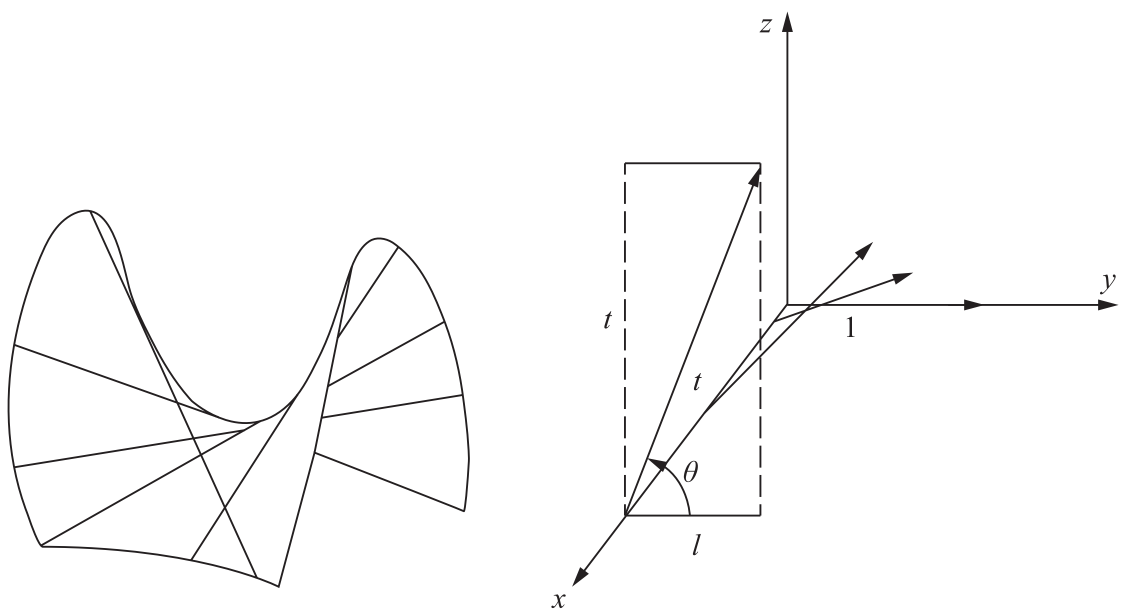


Figure 3-35. $z = xy$ as a ruled surface.

Since $\alpha'(t) = (1, 0, 0)$, we obtain that the line of striction is α itself. The distribution parameter is

$$\lambda = \frac{1}{k}.$$

We also remark that the tangent of the angle θ which $w(t)$ makes with $w(0)$ is $\tan\theta = tk$.

The last remark leads to an interesting general property of a ruled surface. If we consider the family of normal vectors along a ruling of a regular ruled surface, this family generates another ruled surface. By Eq. (7) and the last remark, the latter surface is exactly the hyperbolic paraboloid $z = kxy$, where $1/k$ is the value of the distribution parameter at the chosen ruling.

Among the ruled surfaces, the developables play a distinguished role. Let us start again with an arbitrary ruled surface (not necessarily non-cylindrical)

$$\mathbf{x}(t, v) = \alpha(t) + vw(t), \tag{8}$$

generated by the family $\{\alpha(t), w(t)\}$ with $|w(t)| \equiv 1$. The surface (8) is said to be *developable* if

$$(w, w', \alpha') \equiv 0. \tag{9}$$

To find a geometric interpretation for condition (9), we shall compute the Gaussian curvature of a developable surface at a regular point. A computation entirely similar to the one made to obtain Eq. (6) gives

$$g = 0, \qquad f = \frac{(w, w', \alpha')}{|\mathbf{x}_t \wedge \mathbf{x}_v|}.$$

By condition (9), $f \equiv 0$; hence,

$$K = \frac{eg - f^2}{EG - F^2} \equiv 0.$$

This implies that, *at regular points, the Gaussian curvature of a developable surface is identically zero*.

For another geometric interpretation of a developable surface, see Exercise 6.

We can now distinguish two nonexhaustive cases of developable surfaces:

1. $w(t) \wedge w'(t) \equiv 0$. This implies that $w'(t) \equiv 0$. Thus, $w(t)$ is constant and the ruled surface is a cylinder over a curve obtained by intersecting the cylinder with a plane normal to $w(t)$.

2. $w(t) \wedge w'(t) \neq 0$ for all $t \in I$. In this case $w'(t) \neq 0$ for all $t \in I$. Thus, the surface is noncylindrical, and we can apply our previous work. Thus, we can determine the line of striction (2) and check that the distribution parameter

$$\lambda = \frac{(\beta', w, w')}{|w'|^2} \equiv 0. \tag{10}$$

Therefore, the line of striction will be the locus of singular points of the developable surface. If $\beta'(t) \neq 0$ for all $t \in I$, it follows from Eq. (10) and the fact that $\langle \beta', w' \rangle \equiv 0$ that w is parallel to β'. Thus, the ruled surface is the tangent surface of β. If $\beta'(t) = 0$ for all $t \in I$, then the line of striction is a point, and the ruled surface is a cone with vertex at this point.

Of course, the above cases do not exhaust all possibilities. As usual, if there is a clustering of zeros of the functions involved, the analysis may become rather complicated. At any rate, away from these cluster points, a developable surface is a union of pieces of cylinders, cones, and tangent surfaces.

As we have seen, at regular points, the Gaussian curvature of a developable surface is identically zero. In Sec. 5-8 we shall prove a sort of global converse to this which implies that a regular surface $S \subset R^3$ which is closed as a subset of R^3 and has zero Gaussian curvature is a cylinder.

Example 4 (*The Envelope of the Family of Tangent Planes Along a Curve of a Surface*)**.** Let S be a regular surface and $\alpha = \alpha(s)$ a curve on S parametrized by arc length. Assume that α is nowhere tangent to an asymptotic direction. Consider the ruled surface

$$\mathbf{x}(s, v) = \alpha(s) + v\frac{N(s) \wedge N'(s)}{|N'(s)|}, \tag{11}$$

where by $N(s)$ we denote the unit normal vector of S restricted to the curve $\alpha(s)$ (since $\alpha'(s)$ is not an asymptotic direction, $N'(s) \neq 0$ for all s). We shall show that $\mathbf{x}$ is a developable surface which is regular in a neighborhood of $v = 0$ and is tangent to S along $v = 0$. Before that, however, let us give a geometric interpretation of the surface $\mathbf{x}$.

Consider the family $\{T_{\alpha(s)}(S)\}$ of tangent planes to the surface S along the curve $\alpha(s)$. If Δs is small, the two planes $T_{\alpha(s)}(S)$ and $T_{\alpha(s+\Delta s)}(S)$ of the family will intersect along a straight line parallel to the vector

$$\frac{N(s) \wedge N(s + \Delta s)}{\Delta s}.$$

If we let Δs go to zero, this straight line will approach a limiting position parallel to the vector

$$\begin{aligned} \lim_{\Delta s \to 0} \frac{N(s) \wedge N(s + \Delta s)}{\Delta s} &= \lim_{\Delta s \to 0} N(s) \wedge \frac{(N(s + \Delta s) - N(s))}{\Delta s} \\ &= N(s) \wedge N'(s). \end{aligned}$$

This means intuitively that the rulings of $\mathbf{x}$ are the limiting positions of the intersection of neighboring planes of the family $\{T_{\alpha(s)}(S)\}$. $\mathbf{x}$ is called the *envelope of the family of tangent planes of* S *along* α(s) (Fig. 3-36).

For instance, if α is a parametrization of a parallel of a sphere S^2, then the envelope of tangent planes of S^2 along α is either a cylinder, if the parallel is an equator, or a cone, if the parallel is not an equator (Fig. 3-37).

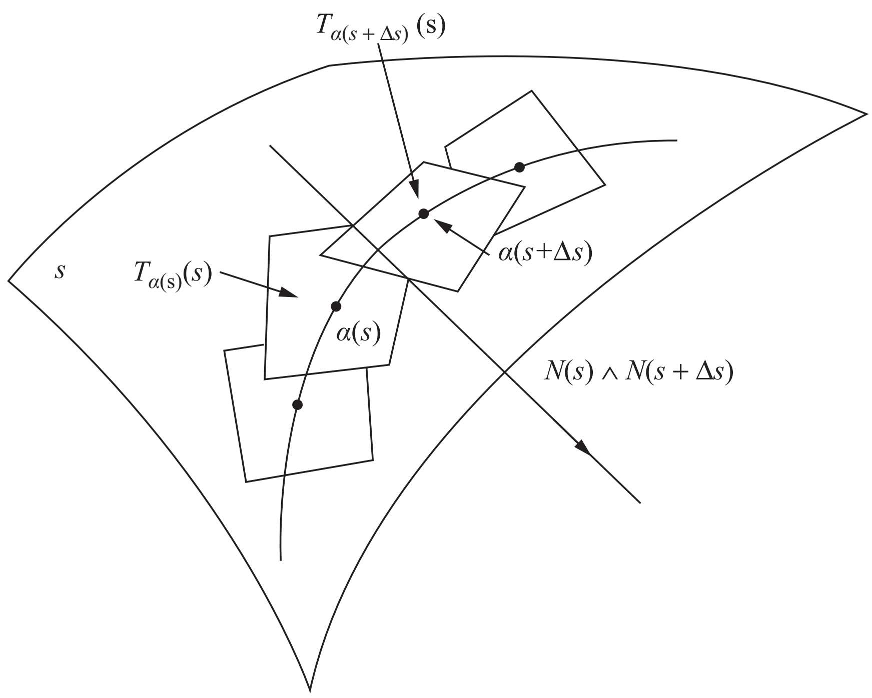


Figure 3-36

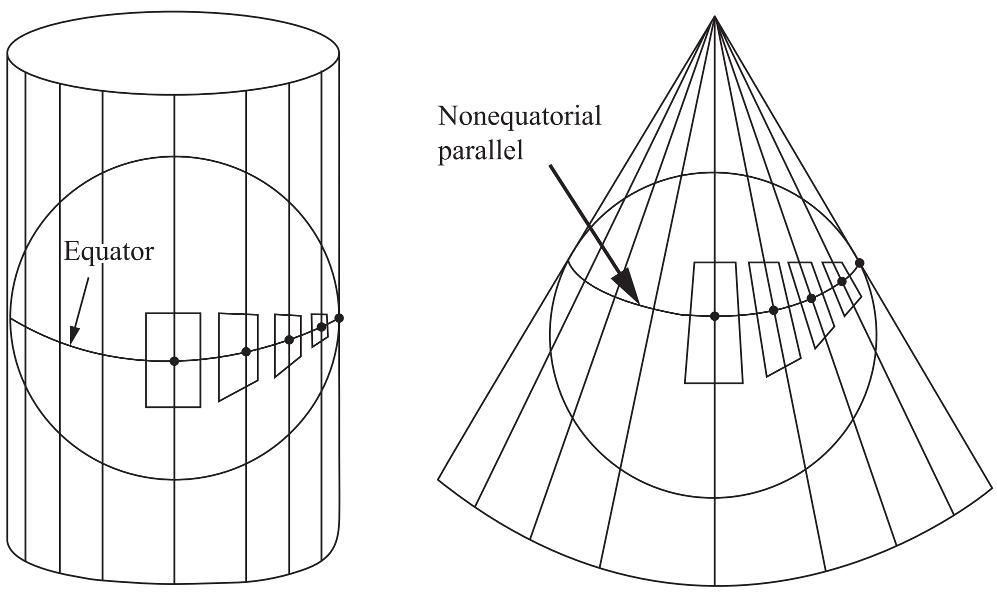


Figure 3-37. Envelopes of families of tangent planes along parallels of a sphere.

To show that **x** is a developable surface, we shall check that condition (9) holds for **x**. In fact, by a straightforward computation, we obtain

$$\left\langle \frac{N \wedge N'}{|N'|} \wedge \left(\frac{N \wedge N'}{|N'|}\right)', \alpha' \right\rangle = \left\langle \frac{N \wedge N'}{|N'|} \wedge \frac{(N \wedge N')'}{|N'|}, \alpha' \right\rangle$$
$$= \frac{1}{|N'|^2}\langle\langle N \wedge N', N''\rangle N, \alpha'\rangle = 0.$$

This proves our claim.

We shall now prove that $\mathbf{x}$ is regular in a neighborhood of $v = 0$ and that it is tangent to S along α. In fact, at $v = 0$, we have

$$\mathbf{x}_s \wedge \mathbf{x}_v = \alpha' \wedge \frac{(N \wedge N')}{|N'|} = \langle N', \alpha' \rangle \frac{N}{|N'|} = -\langle N, \alpha'' \rangle \frac{N}{|N'|}$$
$$= -\frac{(k_n N)}{|N'|},$$

where $k_n = k_n(s)$ is the normal curvature of α. Since $k_n(s)$ is nowhere zero, this shows that $\mathbf{x}$ is regular in a neighborhood of $v = 0$ and that the unit normal vector of $\mathbf{x}$ at $\mathbf{x}(s, 0)$ agrees with $N(s)$. Thus, $\mathbf{x}$ is tangent to S along $v = 0$, and this completes the proof of our assertions.

We shall summarize our conclusions as follows. Let $\alpha(s)$ *be a curve parametrized by arc length on a surface* S *and assume that* α *is nowhere tangent to an asymptotic direction. Then the envelope (11) of the family of tangent planes to* S *along* α *is a developable surface, regular in a neighborhood of* $\alpha(s)$ *and tangent to* S *along* $\alpha(s)$.

B. Minimal Surfaces

A regular parametrized surface is called *minimal* if its mean curvature vanishes everywhere. A regular surface $S \subset R^3$ is *minimal* if each of its parametrizations is minimal.

To explain why we use the word *minimal* for such surfaces, we need to introduce the notion of a variation. Let $\mathbf{x}$: $U \subset R^2 \to R^3$ be a regular parametrized surface. Choose a bounded domain $D \subset U$ (cf. Sec. 2-5) and a differentiable function h: $\bar{D} \to R$, where $\bar{D}$ is the union of the domain D with its boundary ∂D. The *normal variation* of $\mathbf{x}(\bar{D})$, determined by h, is the map (Fig. 3-38) given by,

$$\varphi\colon \bar{D} \times (-\epsilon, \epsilon) \to R^3$$
$$\varphi(u, v, t) = \mathbf{x}(u, v) + th(u, v)N(u, v), \quad (u, v) \in \bar{D}, t \in (-\epsilon, \epsilon).$$

For each fixed $t \in (-\epsilon, \epsilon)$, the map $\mathbf{x}^t$: $D \to R^3$

$$\mathbf{x}^t(u, v) = \varphi(u, v, t)$$

is a parametrized surface with

$$\frac{\partial \mathbf{x}^t}{\partial u} = \mathbf{x}_u + thN_u + th_u N,$$
$$\frac{\partial \mathbf{x}^t}{\partial v} = \mathbf{x}_v + thN_v + th_v N.$$

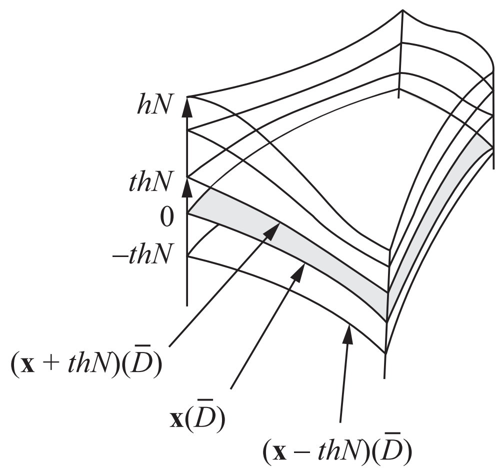


Figure 3-38. A normal variation of $\mathbf{x}(D)$.

Thus, if we denote by E^t, F^t, G^t the coefficients of the first fundamental form of $\mathbf{x}^t$, we obtain

$$\begin{aligned} E^t &= E + th(\langle \mathbf{x}_u, N_u\rangle + \langle \mathbf{x}_u, N_u\rangle) + t^2h^2\langle N_u, N_u\rangle + t^2h_uh_u, \\ F^t &= F + th(\langle \mathbf{x}_u, N_v\rangle + \langle \mathbf{x}_v, N_u\rangle) + t^2h^2\langle N_u, N_v\rangle + t^2h_uh_v, \\ G^t &= G + th(\langle \mathbf{x}_v, N_v\rangle + \langle \mathbf{x}_v, N_v\rangle) + t^2h^2\langle N_v, N_v\rangle + t^2h_vh_v. \end{aligned}$$

By using the fact that

$$\langle \mathbf{x}_u, N_u\rangle = -e, \quad \langle \mathbf{x}_u, N_v\rangle + \langle \mathbf{x}_v, N_u\rangle = -2f, \quad \langle \mathbf{x}_v, N_v\rangle = -g$$

and that the mean curvature H is (Sec. 3-3, Eq. (5))

$$H = \frac{1}{2}\frac{Eg - 2fF + Ge}{EG - F^2},$$

we obtain

$$\begin{aligned} E^tG^t - (F')^2 &= EG - F^2 - 2th(Eg - 2FJ + Ge) + R \\ &= (EG - F^2)(1 - 4thH) + R, \end{aligned}$$

where $\lim_{t\to 0}(R/t) = 0$.

It follows that if ϵ is sufficiently small, $\mathbf{x}^t$ is a regular parametrized surface. Furthermore, the area $A(t)$ of $\mathbf{x}^t(\bar{D})$ is

$$\begin{aligned} A(t) &= \int_{\bar{D}} \sqrt{E^tG^t - (F^t)^2}\, du\, dv \\ &= \int_{\bar{D}} \sqrt{1 - 4thH + \bar{R}}\sqrt{EG - F^2}\, du\, dv, \end{aligned}$$

where $\bar{R} = R/(EG - F^2)$. It follows that if ϵ is small, A is a differentiable function and its derivative at $t = 0$ is

$$A'(0) = -\int_{\bar{D}} 2hH\sqrt{EG - F^2}\, du\, dv \tag{12}$$

We are now prepared to justify the use of the word *minimal* in connection with surfaces with vanishing mean curvature.

PROPOSITION 1. *Let* $\mathbf{x}$: $\mathrm{U} \to \mathrm{R}^3$ *be a regular parametrized surface and let* $\mathrm{D} \subset \mathrm{U}$ *be a bounded domain in* U. *Then* $\mathbf{x}$ *is minimal if and only if* $\mathrm{A}'(0) = 0$ *for all such* D *and all normal variations of* $\mathbf{x}(\bar{D})$.

Proof. If $\mathbf{x}$ is minimal, $H \equiv 0$ and the condition is clearly satisfied. Conversely, assume that the condition is satisfied and that $H(q) \neq 0$ for some $q \in D$. Choose h: $\bar{D} \to R$ such that $h(q) = H(q)$, $hH > 0$, and h is identically zero outside a small neighborhood of q. Then $A'(0) < 0$ for the variation determined by this h, and that is a contradiction. **Q.E.D.**

Thus, any bounded region $\mathbf{x}(\bar{D})$ of a minimal surface $\mathbf{x}$ is a critical point for the area function of any normal variation of $\mathbf{x}(\bar{D})$. It should be noticed that this critical point may not be a minimum and that this makes the word minimal seem somewhat awkward. It is, however, a time-honored terminology which was introduced by Lagrange (who first defined a minimal surface) in 1760.

Minimal surfaces are usually associated with soap films that can be obtained by dipping a wire frame into a soap solution and withdrawing it carefully. If the experiment is well performed, a soap film is obtained that has the same frame as a boundary. It can be shown by physical considerations that the film will assume a position where at its regular points the mean curvature is zero. In this way we can "manufacture" beautiful minimal surfaces, such as the one in Fig. 3-39.

Figure 3-39

Remark 1. It should be pointed out that not all soap films are minimal surfaces according to our definition. We have assumed minimal surfaces to be regular (we could have assumed some isolated singular points, but to go beyond that would make the treatment much less elementary). However, soap

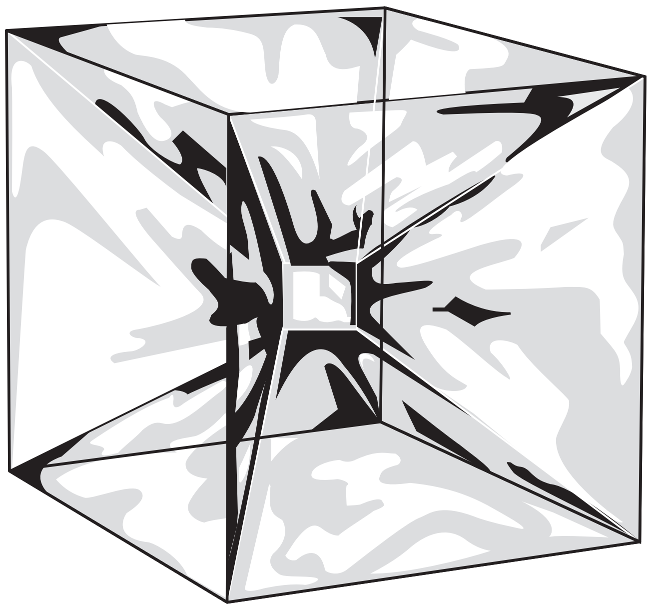


Figure 3-40

films can be formed, for instance, using a cube as a frame (Fig. 3-40), that have singularities along lines.

Remark 2. The connection between minimal surfaces and soap films motivated the celebrated Plateau's problem (Plateau was a Belgian physicist who made careful experiments with soap films around 1850). The problem can be roughly described as follows: *to prove that for each closed curve* C $\subset$ R^3 *there exists a surface* S *of minimum area with* C *as boundary*. To make the problem precise (which curves and surfaces are allowed and what is meant by C being a boundary of S) is itself a nontrivial part of the problem. A version of Plateau's problem was solved simultaneously by Douglas and Radó in 1930. Further versions (and generalizations of the problem for higher dimensions) have inspired the creation of mathematical entities which include at least as many things as soap-like films. We refer the interested reader to the Chap. 2 of Lawson [20] (references are at the end of the book) for further details and a recent bibliography of Plateau's problem.

It will be convenient to introduce, for an arbitrary parametrized regular surface, the *mean curvature vector* defined by $\mathbf{H} = HN$. The geometrical meaning of the direction of $\mathbf{H}$ can be obtained from Eq. (12). Indeed, if we choose $h = H$, we have, for this particular variation,

$$A'(0) = -2 \int_{\bar{D}} \langle \mathbf{H}, \mathbf{H} \rangle \sqrt{EG - F^2}\, du\, dv < 0.$$

This means that *if we deform* $\mathbf{x}(\bar{\mathrm{D}})$ *in the direction of the vector* $\mathbf{H}$, *the area is initially decreasing.*

The mean curvature vector has another interpretation which we shall now pursue, since it has important implications for the theory of minimal surfaces.

A regular parametrized surface $\mathbf{x} = \mathbf{x}(u, v)$ is said to be *isothermal* if $\langle \mathbf{x}_u, \mathbf{x}_u \rangle = \langle \mathbf{x}_v, \mathbf{x}_v \rangle$ and $\langle \mathbf{x}_u, \mathbf{x}_v \rangle = 0$.

PROPOSITION 2. *Let* $\mathbf{x} = \mathbf{x}(\mathrm{u}, \mathrm{v})$ *be a regular parametrized surface and assume that* $\mathbf{x}$ *is isothermal. Then*

$$\mathbf{x}_{\mathrm{uu}} + \mathbf{x}_{\mathrm{vv}} = 2\lambda^2 \mathbf{H},$$

where $\lambda^2 = \langle \mathbf{x}_{\mathrm{u}}, \mathbf{x}_{\mathrm{u}} \rangle = \langle \mathbf{x}_{\mathrm{v}}, \mathbf{x}_{\mathrm{v}} \rangle$.

Proof. Since $\mathbf{x}$ is isothermal, $\langle \mathbf{x}_u, \mathbf{x}_u \rangle = \langle \mathbf{x}_v, \mathbf{x}_v \rangle$ and $\langle \mathbf{x}_u, \mathbf{x}_v \rangle = 0$. By differentiation, we obtain

$$\langle \mathbf{x}_{uu}, \mathbf{x}_u \rangle = \langle \mathbf{x}_{vu}, \mathbf{x}_v \rangle = -\langle \mathbf{x}_u, \mathbf{x}_{vv} \rangle.$$

Thus,

$$\langle \mathbf{x}_{uu} + \mathbf{x}_{vv}, \mathbf{x}_u \rangle = 0.$$

Similarly,

$$\langle \mathbf{x}_{uu} + \mathbf{x}_{vv}, \mathbf{x}_v \rangle = 0.$$

It follows that $\mathbf{x}_{uu} + \mathbf{x}_{vv}$ is parallel to N. Since $\mathbf{x}$ is isothermal,

$$H = \frac{1}{2}\frac{g+e}{\lambda^2}.$$

Thus,

$$2\lambda^2 H = g + e = \langle N, \mathbf{x}_{uu} + \mathbf{x}_{vv} \rangle;$$

hence,

$$\mathbf{x}_{uu} + \mathbf{x}_{vv} = 2\lambda^2 \mathbf{H}.$$ **Q.E.D.**

The Laplacian Δf of a differentiable function $f : U \subset R^2 \to R$ is defined by $\Delta f = (\partial^2 f/\partial u^2) + (\partial^2 f/\partial v^2)$, $(u, v) \in U$. We say that f is *harmonic* in U if $\Delta f = 0$. From Prop. 2, we obtain

COROLLARY: *Let* $\mathbf{x}(\mathrm{u}, \mathrm{v}) = (\mathrm{x}(\mathrm{u}, \mathrm{v}), \mathrm{y}(\mathrm{u}, \mathrm{v}), \mathrm{z}(\mathrm{u}, \mathrm{v}))$ *be a parametrized surface and assume that* $\mathbf{x}$ *is isothermal. Then* $\mathbf{x}$ *is minimal if and only if its coordinate functions* x, y, z *are harmonic.*

Example 5. *The catenoid*, given by

$$\mathbf{x}(u, v) = (a \cosh v \cos u, a \cosh v \sin u, av),$$
$$0 < u < 2\pi, \quad -\infty < v < \infty.$$

This is the surface generated by rotating the catenary $y = a \cosh(z/a)$ about the z axis (Fig. 3-41). It is easily checked that $E = G = a^2 \cosh^2 v$, $F = 0$,

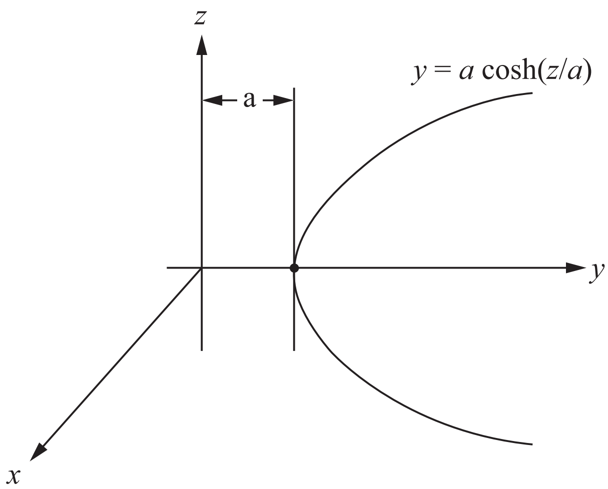


Figure 3-41

and that $\mathbf{x}_{uu} + \mathbf{x}_{vv} = 0$. Thus, the catenoid is a minimal surface. It can be characterized as the only surface of revolution which is minimal.

The last assertion can be proved as follows. We want to find a curve $y = f(x)$ such that, when rotated about the x axis, it describes a minimal surface. Since the parallels and the meridians of a surface of revolution are lines of curvature of the surface (Sec. 3-3, Example 4), we must have that the curvature of the curve $y = f(x)$ is the negative of the normal curvature of the circle generated by the point $f(x)$ (both are principal curvatures). Since the curvature of $y = f(x)$ is

$$\frac{y''}{(1 + (y')^2)^{3/2}}$$

and the normal curvature of the circle is the projection of its usual curvature $(=1/y)$ over the normal N to the surface (see Fig. 3-42), we obtain

$$\frac{y''}{(1 + (y')^2)^{3/2}} = -\frac{1}{y}\cos\varphi.$$

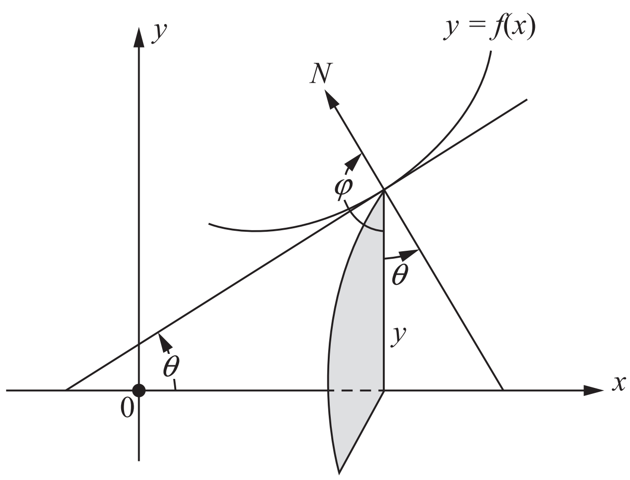


Figure 3-42

But $-\cos\varphi = \cos\theta$ (see Fig. 3-42), and since $\tan\theta = y'$, we obtain

$$\frac{y''}{(1+(y')^2)^{3/2}} = \frac{1}{y}\frac{1}{(1+(y')^2)^{1/2}}$$

as the equation to be satisfied by the curve $y = f(x)$.

Clearly, there exists a point x where $f'(x) \neq 0$. Let us work in a neighborhood of this point where $f' \neq 0$. Multiplying both members of the above equation by $2y'$, we obtain,

$$\frac{2y''y'}{1+(y')^2} = \frac{2y'}{y}.$$

Setting $1+(y')^2 = z$ (hence, $2y''y' = z'$), we have

$$\frac{z'}{z} = \frac{2y'}{y},$$

which, by integration, gives (k is a constant)

$$\log z = \log y^2 + \log k^2 = \log(yk)^2$$

or

$$1+(y')^2 = z = (yk)^2.$$

The last expression can be written

$$\frac{k\,dy}{\sqrt{(yk)^2-1}} = k\,dx,$$

which, again by integration, gives (c is a constant)

$$\cosh^{-1}(yk) = kx + c$$

or

$$y = \frac{1}{k}\cosh(kx+c).$$

Thus, in the neighborhood of a point where $f' \neq 0$, the curve $y = f(x)$ is a catenary. But then y' can only be zero at $x = 0$, and if the surface is to be connected, it is by continuity a catenoid, as we claimed.

Example 6 (*The Helicoid*)**.** (cf. Example 3, Sec. 2-5.)

$$\mathbf{x}(u, v) = (a\sinh v\cos u, a\sinh v\sin u, au).$$

It is easily checked that $E = G = a^2\cosh^2 v$, $F = 0$, and $\mathbf{x}_{uu} + \mathbf{x}_{vv} = 0$. Thus, the helicoid is a minimal surface. It has the additional property that it is the only minimal surface, other than the plane, which is also a ruled surface.

We can give a proof of the last assertion if we assume that the zeros of the Gaussian curvature of a minimal surface are isolated (for a proof, see, for instance, the survey of Osserman quoted at the end of this section, p. 76). Granted this, we shall proceed as follows.

Assume that the surface is not a plane. Then in some neighborhood W of the surface the Gaussian curvature K is strictly negative. Since the mean curvature is zero, W is covered by two families of asymptotic curves which intersect orthogonally. Since the rulings are asymptotic curves and the surface is not a plane, we can choose a point $q \in W$ such that the asymptotic curve, other than the ruling, passing through q has nonzero torsion $\tau = \sqrt{-K}$ at q. Since the osculating plane of an asymptotic curve is the tangent plane to the surface, there is a neighborhood $V \subset W$ such that the rulings of V are principal normals to the family of twisted asymptotic curves (Fig. 3-43). It is an interesting exercise in curves to prove that this can occur if and only if the twisted curves are circular helices (cf. Exercise 18, Sec. 1-5). Thus, V is a part of a helicoid. Since the torsion of a circular helix is constant, we easily see that the whole surface is part of a helicoid, as we claimed.

The helicoid and the catenoid were discovered in 1776 by Meusnier, who also proved that Lagrange's definition of minimal surfaces as critical points of a variational problem is equivalent to the vanishing of the mean curvature. For a long time, they were the only known examples of minimal surfaces. Only in 1835 did Scherk find further examples, one of which is described

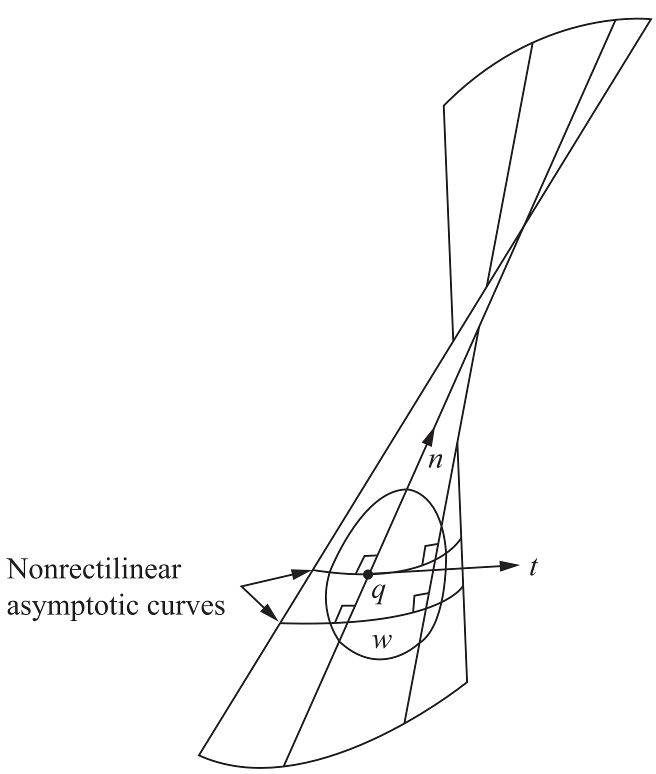


Figure 3-43

in Example 8. In Exercise 14, we shall describe an interesting connection between the helicoid and the catenoid.

Example 7 (*Enneper's Minimal Surface*)**.** Enneper's surface is the parametrized surface

$$\mathbf{x}(u, v) = \left(u - \frac{u^3}{3} + uv^2, v - \frac{v^3}{3} + vu^2, u^2 - v^2\right), \quad (u, v) \in R^2,$$

which is easily seen to be minimal (Fig. 3-44). Notice that by changing (u, v) into $(-v, u)$ we change, in the surface, (x, y, z) into $(-y, x, -z)$. Thus, if we perform a positive rotation of $\pi/2$ about the z axis and follow it by a symmetry in the xy plane, the surface remains invariant.

An interesting feature of Enneper's surface is that it has self-intersections. This can be shown by setting $u = \rho\cos\theta$, $v = \rho\sin\theta$ and writing

$$\mathbf{x}(\rho, \theta) = \left(\rho\cos\theta - \frac{\rho^3}{3}\cos 3\theta, \rho\sin\theta + \frac{\rho^3}{3}\sin 3\theta, \rho^2\cos 2\theta\right).$$

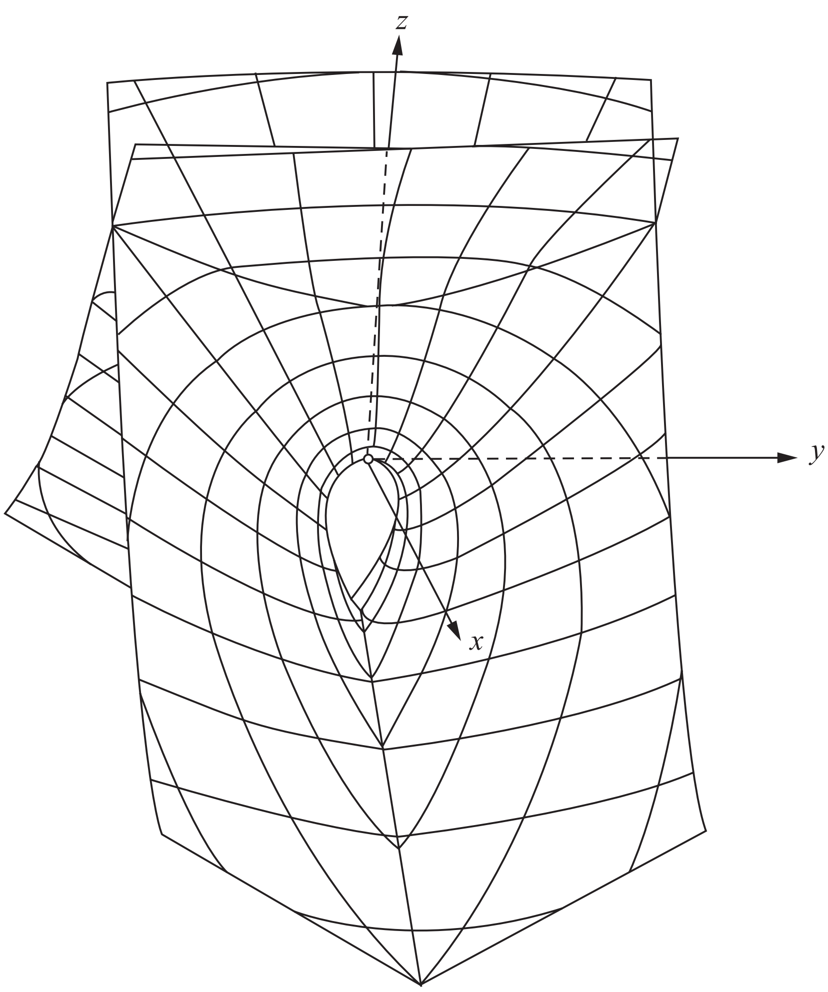


Figure 3-44. Enneper's surface. Reproduced, with modifications, from K. Leichtweiss, "Minimalflächen im Grossen," *Überblicke Math.* 2 (1969), 7–49, Fig. 4, with permission.

Thus, if $\mathbf{x}(\rho_1, \theta_1) = \mathbf{x}(\rho_2, \theta_2)$, a straightforward computation shows that

$$
\begin{aligned}
x^2 + y^2 &= \rho_1^2 + \frac{\rho_1^6}{9} - \cos 4\theta_1 \frac{2\rho_1^4}{3} \\
&= \left(\rho_1 + \frac{\rho_1^3}{3}\right)^2 - \frac{4}{3}(\rho_1^2 \cos 2\theta_1)^2 \\
&= \left(\rho_2 + \frac{\rho_2^3}{3}\right)^2 - \frac{4}{3}(\rho_2^2 \cos 2\theta_2)^2.
\end{aligned}
$$

Hence, since $\rho_1^2 \cos 2\theta_1 = \rho_2^2 \cos 2\theta_2$, we obtain

$$\rho_1 + \frac{\rho_1^3}{3} = \rho_2 + \frac{\rho_2^3}{3},$$

which implies that $\rho_1 = \rho_2$. It follows that $\cos 2\theta_1 = \cos 2\theta_2$.

If, for instance, $\rho_1 = \rho_2$ and $\theta_1 = 2\pi - \theta_2$, we obtain from

$$y(\rho_1, \theta_1) = y(\rho_2, \theta_2)$$

that $y = -y$. Hence, $y = 0$; that is, the points (ρ_1, θ_1) and (ρ_2, θ_2) belong to the curve $\sin\theta + (\rho^2/3)\sin 3\theta = 0$. Clearly, for each point (ρ, θ) belonging to this curve, the point $(\rho, 2\pi - \theta)$ also belongs to it, and

$$x(\rho, \theta) = x(\rho, 2\pi - \theta), z(\rho, \theta) = z(\rho, 2\pi - \theta).$$

Thus, the intersection of the surface with the plane $y = 0$ is a curve along which the surface intersects itself.

Similarly, it can be shown that the intersection of the surface with the plane $x = 0$ is also a curve of self-intersection (this corresponds to the case $\rho_1 = \rho_2, \theta_1 = \pi - \theta_2$). It is easily seen that they are the only self-intersections of Enneper's surface.

I want to thank Alcides Lins Neto for having worked out this example in order to draw a first sketch of Fig. 3-44.

Before going into the next example, we shall establish a useful relation between minimal surfaces and analytic functions of a complex variable. Let $\mathbb{C}$ denote the complex plane, which is, as usual, identified with R^2 by setting $\zeta = u + iv$, $\zeta \in \mathbb{C}$, $(u, v) \in R^2$. We recall that a function $f\colon U \subset \mathbb{C} \to \mathbb{C}$ is *analytic* when, by writing

$$f(\zeta) = f_1(u, v) + i f_2(u, v),$$

the real functions f_1 and f_2 have continuous partial derivatives of first order which satisfy the so-called Cauchy-Riemann equations:

$$\frac{\partial f_1}{\partial u} = \frac{\partial f_2}{\partial v}, \quad \frac{\partial f_1}{\partial v} = -\frac{\partial f_2}{\partial u}.$$

Now let $\mathbf{x}\colon U \subset R^2 \to R^3$ be a regular parametrized surface and define complex functions $\varphi_1, \varphi_2, \varphi_3$ by

$$\varphi_1(\zeta) = \frac{\partial x}{\partial u} - i\frac{\partial x}{\partial v}, \quad \varphi_2(\zeta) = \frac{\partial y}{\partial u} - i\frac{\partial y}{\partial v}, \quad \varphi_3(\zeta) = \frac{\partial z}{\partial u} - i\frac{\partial z}{\partial v},$$

where x, y, and z are the component functions of $\mathbf{x}$.

LEMMA. $\mathbf{x}$ *is isothermal if and only if* $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 \equiv 0$. *If this last condition is satisfied,* $\mathbf{x}$ *is minimal if and only if* φ_1, φ_2, *and* φ_3 *are analytic functions.*

Proof. By a simple computation, we obtain that

$$\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = E - G + 2iF,$$

whence the first part of the lemma. Furthermore, $\mathbf{x}_{uu} + \mathbf{x}_{vv} = 0$ if and only if

$$\begin{aligned} \frac{\partial}{\partial u}\left(\frac{\partial x}{\partial u}\right) &= -\frac{\partial}{\partial v}\left(\frac{\partial x}{\partial v}\right), \\ \frac{\partial}{\partial u}\left(\frac{\partial y}{\partial u}\right) &= -\frac{\partial}{\partial v}\left(\frac{\partial y}{\partial v}\right), \\ \frac{\partial}{\partial u}\left(\frac{\partial z}{\partial u}\right) &= -\frac{\partial}{\partial v}\left(\frac{\partial z}{\partial v}\right), \end{aligned}$$

which give one-half of the Cauchy-Riemann equations for φ_1, φ_2, φ_3. Since the other half is automatically satisfied, we conclude that $\mathbf{x}_{uu} + \mathbf{x}_{vv} = 0$ if and only if φ_1, φ_2, and φ_3 are analytic. **Q.E.D.**

Example 8 (*Scherk's Minimal Surface*)**.** This is given by

$$\mathbf{x}(u, v) = \left(\arg\frac{\zeta + i}{\zeta - i}, \arg\frac{\zeta + 1}{\zeta - 1}, \log\left|\frac{\zeta^2 + 1}{\zeta^2 - 1}\right|\right),$$

$$\zeta \neq \pm 1, \zeta \neq \pm i,$$

where $\zeta = u + iv$, and $\arg \zeta$ is the angle that the real axis makes with ζ.

We easily compute that

$$\begin{aligned}
\arg\frac{\zeta+i}{\zeta-i} &= \tan^{-1}\frac{2u}{u^2+v^2-1},\\
\arg\frac{\zeta+1}{\zeta-1} &= \tan^{-1}\frac{-2v}{u^2+v^2-1}\\
\log\left|\frac{\zeta^2+1}{\zeta^2-1}\right| &= \frac{1}{2}\log\frac{(u^2-v^2+1)^2+4u^2v^2}{(u^2-v^2-1)^2+4u^2v^2};
\end{aligned}$$

hence,

$$\varphi_1 = \frac{\partial x}{\partial u} - i\frac{\partial x}{\partial v} = -\frac{2}{1+\zeta^2}, \quad \varphi_2 = -\frac{2i}{1-\zeta^2}, \quad \varphi_3 = \frac{4\zeta}{1-\zeta^4}.$$

Since $\varphi_1^2+\varphi_2^2+\varphi_3^2 \equiv 0$ and φ_1, φ_2, and φ_3 are analytic, $\mathbf{x}$ is an isothermal parametrization of a minimal surface.

It is easily seen from the expressions of x, y, and z that

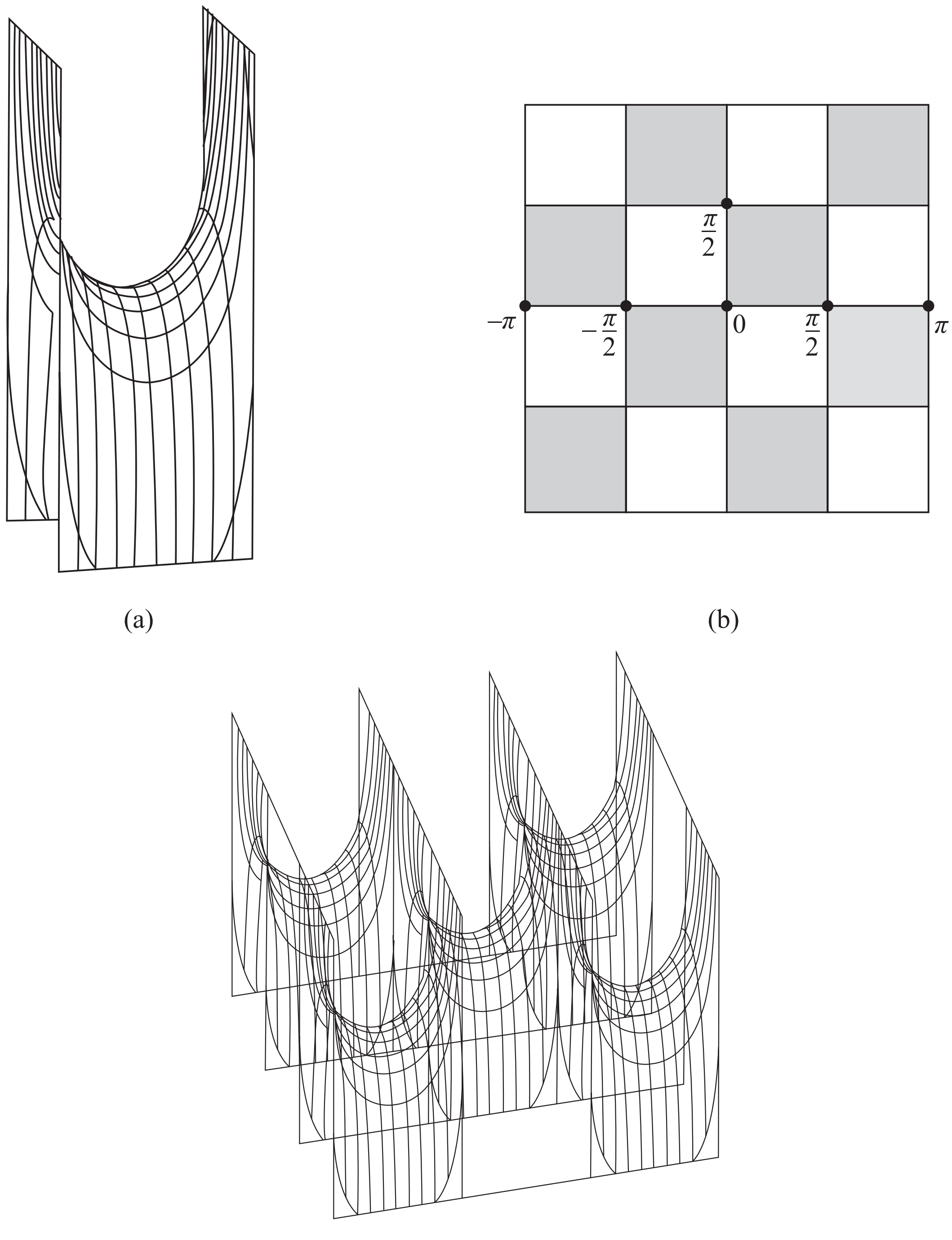
$$z = \log\frac{\cos y}{\cos x}.$$

This representation shows that Scherk's surface is defined on the chessboard pattern of Fig. 3-45 (except at the vertices of the squares, where the surface is actually a vertical line).

Minimal surfaces are perhaps the best-studied surfaces in differential geometry, and we have barely touched the subject. A very readable introduction can be found in R. Osserman, *A Survey of Minimal Surfaces*, Van Nostrand Mathematical Studies, Van Nostrand Reinhold, New York, 1969. The theory has developed into a rich branch of differential geometry in which interesting and nontrivial questions are still being investigated. It has deep connections with analytic functions of complex variables and partial differential equations. As a rule, the results of the theory have the charming quality that they are easy to visualize and very hard to prove. To convey to the reader some flavor of the subject we shall close this brief account by stating without proof one striking result.

THEOREM (Osserman). *Let* $S \subset R^3$ *be a regular, closed* (*as a subset of* R^3) *minimal surface in* R^3 *which is not a plane. Then the image of the Gauss map* $N\colon S \to S^2$ *is* dense in the sphere S^2 (*that is, arbitrarily close to any point of* S^2 *there is a point of* $N(S) \subset S^2$).

A proof of this theorem can be found in Osserman's survey, quoted above. Actually, the theorem is somewhat stronger in that it applies to complete surfaces, a concept to be defined in Sec. 5-3.



(a) (b) (c)

Figure 3-45. Scherk's surface.

EXERCISES

1. Show that the helicoid (cf. Example 3, Sec. 2-5) is a ruled surface, its line of striction is the z axis, and its distribution parameter is constant.
2. Show that on the hyperboloid of revolution $x^2 + y^2 - z^2 = 1$, the parallel of least radius is the line of striction, the rulings meet it under a constant angle, and the distribution parameter is constant.

3. Let $\alpha\colon I \to S \subset R^3$ be a curve on a regular surface S and consider the ruled surface generated by the family $\{\alpha(t), N(t)\}$, where $N(t)$ is the normal to the surface at $\alpha(t)$. Prove that $\alpha(I) \subset S$ is a line of curvature in S if and only if this ruled surface is developable.

4. Assume that a noncylindrical ruled surface

$$\mathbf{x}(t, v) = \alpha(t) + v w(t), \quad |w| = 1,$$

is regular. Let $w(t_1)$, $w(t_2)$ be the directions of two rulings of $\mathbf{x}$ and let $\mathbf{x}(t_1, v_1)$, $\mathbf{x}(t_2, v_2)$ be the feet of the common perpendicular to these two rulings. As $t_2 \to t_1$, these points tend to a point $\mathbf{x}(t_1, \bar{v})$. To determine $(t_1, \bar{v})$ prove the following:

a. The unit vector of the common perpendicular converges to a unit vector tangent to the surface at $(t_1, \bar{v})$. Conclude that, at $(t_1, \bar{v})$,

$$\langle w' \wedge w, N\rangle = 0.$$

b. $\bar{v} = -(\langle \alpha', w'\rangle / \langle w', w'\rangle)$.

Thus, $(t_1, \bar{v})$ is the central point of the ruling through t_1, and this gives another interpretation of the line of striction (assumed nonsingular).

5. A *right conoid* is a ruled surface whose rulings L_t intersect perpendicularly at fixed axis r which does not meet the directrix $\alpha\colon I \to R^3$.

a. Find a parametrization for the right conoid and determine a condition that implies it to be noncyliodrical.

b. Given a noncylindrical right conoid, find the line of striction and the distribution parameter.

6. Let

$$\mathbf{x}(t, v) = \alpha(t) + v w(t)$$

be a developable surface. Prove that at a regular point we have

$$\langle N_v, \mathbf{x}_v\rangle = \langle N_v, \mathbf{x}_t\rangle = 0.$$

Conclude that *the tangent plane of a developable surface is constant along* (the regular points of) *a fixed ruling.*

7. Let S be a regular surface and let $C \subset S$ be a regular curve on S, nowhere tangent to an asymptotic direction. Consider the envelope of the family of tangent planes of S along C. Prove that the direction of the ruling that passes through a point $p \in C$ is conjugate to the tangent direction of C at p.

8. Show that if $C \subset S^2$ is a parallel of unit sphere S^2, then the envelope of tangent planes of S^2 along C is either a cylinder, if C is an equator, or a cone, if C is not an equator.

9. (*Focal Surfaces.*) Let S be a regular surface without parabolic or umbilical points. Let $\mathbf{x}$: $U \to S$ be a parametrization of S such that the coordinate curves are lines of curvature (if U is small, this is no restriction. cf. Corollary 4, Sec. 3-4). The parametrized surfaces

$$\mathbf{y}(u, v) = \mathbf{x}(u, v) + \rho_1 N(u, v),$$
$$\mathbf{z}(u, v) = \mathbf{x}(u, v) + \rho_2 N(u, v),$$

where $\rho_1 = 1/k_1$, $\rho_2 = 1/k_2$, are called *focal surfaces* of $\mathbf{x}(U)$ (or *surfaces of centers* of $\mathbf{x}(U)$; this terminology comes from the fact that $\mathbf{y}(u, v)$, for instance, is the center of the osculating circle (cf. Sec. 1-6, Exercise 2) of the normal section at $\mathbf{x}(u, v)$ corresponding to the principal curvature k_1). Prove that

a. If $(k_1)_u$ and $(k_2)_v$ are nowhere zero, then $\mathbf{y}$ and $\mathbf{z}$ are regular parametrized surfaces.

b. At the regular points, the directions on a focal surface corresponding to the principal directions on $\mathbf{x}(U)$ are conjugate. That means, for instance, that $\mathbf{y}_u$ and $\mathbf{y}_v$ are conjugate vectors in $\mathbf{y}(U)$ for all $(u, v) \in U$.

c. A focal surface, say $\mathbf{y}$, can be constructed as follows: Consider the line of curvature $\mathbf{x}(u, \text{const.})$ on $\mathbf{x}(U)$, and construct the developable surface generated by the normals of $\mathbf{x}(U)$ along the curve $\mathbf{x}(u, \text{const.})$ (cf. Exercise 3). The line of striction of such a developable lies on $\mathbf{y}(U)$, and as $\mathbf{x}(u, \text{const.})$ describes $\mathbf{x}(U)$, this line describes $y(U)$ (Fig. 3-46).

10. Example 4 can be generalized as follows. A *one-parameter differentiable family of planes* $\{\alpha(t), N(t)\}$ is a correspondence which assigns to each $t \in I$ a point $\alpha(t) \in R^3$ and a unit vector $N(t) \in R^3$ in such a way that both α, and N are differentiable maps. A family $\{\alpha(t), N(t)\}$, $t \in I$, is said to be a *family of tangent planes* if $\alpha'(t) \neq 0$, $N'(t) \neq 0$, and $\langle \alpha'(t), N(t) \rangle = 0$ for all $t \in I$.

a. Give a proof that a differentiable one-parameter family of tangent planes $\{\alpha(t), N(t)\}$ determines a differentiable one-parameter family of lines $\{\alpha(t), (N \wedge N')/|N'|\}$ which generates a developable surface

$$\mathbf{x}(t, v) = \alpha(t) + v\frac{N \wedge N'}{|N'|}. \tag{$*$}$$

The surface $(*)$ is called the *envelope of the family* $\{\alpha(t), N(t)\}$.

b. Prove that if $\alpha'(t) \wedge (N(t) \wedge N'(t)) \neq 0$ for all $t \in I$, then the envelope $(*)$ is regular in a neighborhood of $v = 0$, and the unit normal vector of $\mathbf{x}$ at $(t, 0)$ is $N(t)$.

c. Let $\alpha = \alpha(s)$ be a curve in R^3 parametrized by arc length. Assume that the curvature $k(s)$ and the torsion $\tau(s)$ of α are nowhere zero.

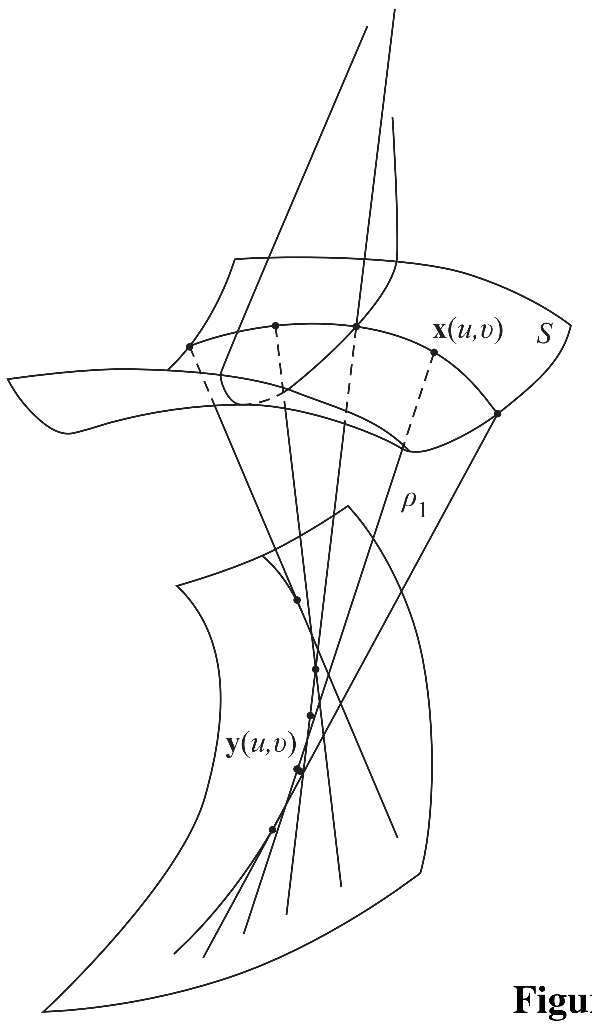


Figure 3-46. Construction of a focal surface.

Prove that the family of osculating planes $\{\alpha(s), b\{s\}\}$ is a one-parameter differentiable family of tangent planes and that the envelope of this family is the tangent surface to $\alpha(s)$ (cf. Example 5, Sec. 2-3).

11. Let $\mathbf{x} = \mathbf{x}(u, v)$ be a regular parametrized surface. A *parallel surface* to $\mathbf{x}$ is a parametrized surface

$$\mathbf{y}(u, v) = \mathbf{x}(u, v) + aN(u, v),$$

where a is a constant.

a. Prove that $y_u \wedge y_v = (1 - 2Ha + Ka^2)(\mathbf{x}_u \wedge \mathbf{x}_v)$, where K and H are the Gaussian and mean curvatures of $\mathbf{x}$, respectively.

b. Prove that at the regular points, the Gaussian curvature of $\mathbf{y}$ is

$$\frac{K}{1 - 2Ha + Ka^2}$$

and the mean curvature of $\mathbf{y}$ is

$$\frac{H - Ka}{1 - 2Ha + Ka^2}.$$

c. Let a surface $\mathbf{x}$ have constant mean curvature equal to $c \neq 0$ and consider the parallel surface to $\mathbf{x}$ at a distance $1/2c$. Prove that this parallel surface has constant Gaussian curvature equal to $4c^2$.

12. Prove that there are no compact (i.e., bounded and closed in R^3) minimal surfaces.

13. **a.** Let S be a regular surface without umbilical points. Prove that S is a minimal surface if and only if the Gauss map $N: S \to S^2$ satisfies, for all $p \in S$ and all $w_1, w_2 \in T_p(S)$,

$$\langle dN_p(w_1), dN_p(w_2)\rangle_{N(p)} = \lambda(p)\langle w_1, w_2\rangle_p,$$

where $\lambda(p) \neq 0$ is a number which depends only on p.

b. Let $\mathbf{x}: U \to S^2$ be a parametrization of the unit sphere S^2 by stereographic projection. Consider a neighborhood V of a point p of the minimal surface S in part a such that $N: S \to S^2$ restricted to V is a diffeomorphism (since $K(p) = \det(dN_p) \neq 0$, such a V exists by the inverse function theorem). Prove that the parametrization $y = N^{-1} \circ \mathbf{x}: U \to S$ is isothermal (*this gives a way of introducing isothermal parametrizations on minimal surfaces without planar points*).

14. When two differentiable functions f, $g: U \subset R^2 \to R$ satisfy the Cauchy-Riemann equations

$$\frac{\partial f}{\partial u} = \frac{\partial g}{\partial v}, \qquad \frac{\partial f}{\partial v} = -\frac{\partial g}{\partial u},$$

they are easily seen to be harmonic; in this situation, f and g are said to be *harmonic conjugate*. Let $\mathbf{x}$ and $\mathbf{y}$ be isothermal parametrizations of minimal surfaces such that their component functions are pairwise harmonic conjugate; then $\mathbf{x}$ and $\mathbf{y}$ are called *conjugate minimal surfaces*. Prove that

a. The helicoid and the catenoid are conjugate minimal surfaces.

b. Given two conjugate minimal surfaces, $\mathbf{x}$ and $\mathbf{y}$, the surface

$$\mathbf{z} = (\cos t)\mathbf{x} + (\sin t)\mathbf{y} \qquad (*)$$

is again minimal for all $t \in R$.

c. All surfaces of the one-parameter family $(*)$ have the same fundamental form: $E = \langle \mathbf{x}_u, \mathbf{x}_u\rangle = \langle \mathbf{y}_v, \mathbf{y}_v\rangle$, $F = 0$, $G = \langle \mathbf{x}_v, \mathbf{x}_v\rangle = \langle \mathbf{y}_u, \mathbf{y}_u\rangle$.

Thus, any two conjugate minimal surfaces can be joined through a one-parameter family of minimal surfaces, and the first fundamental form of this family is independent of t.

Appendix *Self-Adjoint Linear Maps and Quadratic Forms*

In this appendix, V will denote a vector space of dimension 2, endowed with an inner product $\langle\ ,\ \rangle$. All that follows can be easily extended to a finite n-dimensional vector space, but for the sake of simplicity, we shall treat only the case $n = 2$.

We say that a linear map $A: V \to V$ is *self-adjoint* if $\langle Av, w\rangle = \langle v, Aw\rangle$ for all $v, w \in V$.

Notice that if $\{e_1, e_2\}$ is an orthonormal basis for V and (α_{ij}), $i, j = 1, 2$, is the matrix of A relative to that basis, then

$$\langle Ae_j, e_i\rangle = \alpha_{ij} = \langle e_j, Ae_i\rangle = \langle Ae_i, e_j\rangle = \alpha_{ji};$$

that is, the matrix (α_{ij}) is symmetric.

To each self-adjoint linear map we associate a map $B: V \times V \to R$ defined by

$$B(v, w) = \langle Av, w\rangle.$$

B is clearly bilinear; that is, it is linear in both v and w. Moreover, the fact that A is self-adjoint implies that $B(v, w) = B(w, v)$; that is, B is a bilinear symmetric form in V.

Conversely, if B is a bilinear symmetric form in V, we can define a linear map $A: V \to V$ by $\langle Av, w\rangle = B(v, w)$ and the symmetry of B implies that A is self-adjoint.

On the other hand, to each symmetric, bilinear form B in V, there corresponds a quadratic form Q in V given by

$$Q(v) = B(v, v), \quad v \in V,$$

and the knowledge of Q determines B completely, since

$$B(u, v) = \tfrac{1}{2}[Q(u+v) - Q(u) - Q(v)].$$

Thus, a one-to-one correspondence is established between quadratic forms in V and self-adjoint linear maps of V.

The goal of this appendix is to prove that (see the theorem below) given a self-adjoint linear map $A\colon V \to V$, there exists an orthonormal basis for V such that relative to that basis the matrix of A is a diagonal matrix. Furthermore, the elements on the diagonal are the maximum and the minimum of the corresponding quadratic form restricted to the unit circle of V.

LEMMA. *If the function* $\mathrm{Q(x, y) = ax^2 + 2bxy + cy^2}$, *restricted to the unit circle* $\mathrm{x^2 + y^2 = 1}$, *has a maximum at the point* $(1, 0)$, *then* $\mathrm{b} = 0$.

Proof. Parametrize the circle $x^2 + y^2 = 1$ by $x = \cos t$, $y = \sin t$, $t \in (0 - \epsilon, 2\pi + \epsilon)$. Thus, Q, restricted to that circle, becomes a function of t:

$$Q(t) = a\cos^2 t + 2b\cos t\sin t + c\sin^2 t.$$

Since Q has a maximum at the point $(1, 0)$ we have

$$\left(\frac{dQ}{dt}\right)_{t=0} = 2b = 0.$$

Hence, $b = 0$ as we wished. **Q.E.D.**

PROPOSITION. *Given a quadratic form* Q *in* V, *there exists an orthonormal basis* $\{e_1, e_2\}$ *of* V *such that if* $\mathrm{v \in V}$ *is given by* $\mathrm{v = xe_1 + ye_2}$, *then*

$$\mathrm{Q(v)} = \lambda_1 \mathrm{x}^2 + \lambda_2 \mathrm{y}^2,$$

where λ_1 *and* λ_2 *are the maximum and minimum, respectively, of* Q *on the unit circle* $|\mathrm{v}| = 1$.

Proof. Let λ_1 be the maximum of Q on the unit circle $|v| = 1$, and let e_1 be a unit vector with $Q(e_1) = \lambda_1$. Such an e_1 exists by continuity of Q on the compact set $|v| = 1$. Let e_2 be a unit vector that is orthogonal to e_1, and set $\lambda_2 = Q(e_2)$. We shall show that the basis $\{e_1, e_2\}$ satisfies the conditions of the proposition.

Let B be the symmetric bilinear form that is associated to Q and set $v = xe_1 + ye_2$. Then

$$\begin{aligned} Q(v) = B(v, v) &= B(xe_1 + ye_2, xe_1 + ye_2) \\ &= \lambda_1 x^2 + 2bxy + \lambda_2 y^2, \end{aligned}$$

where $b = B(e_1, e_2)$. By the lemma, $b = 0$, and it only remains to prove that λ_2 is the minimum of Q in the circle $|v| = 1$. This is immediate because, for any $v = xe_1 + ye_2$ with $x^2 + y^2 = 1$, we have that

$$Q(v) = \lambda_1 x^2 + \lambda_2 y^2 \geq \lambda_2(x^2 + y^2) = \lambda_2,$$

since $\lambda_2 \leq \lambda_1$. **Q.E.D.**

We say that a vector $v \neq 0$ is an *eigenvector* of a linear map $A\colon V \to V$ if $Av = \lambda v$ for some real number λ; λ is then called an *eigenvalue* of A.

THEOREM. *Let* $\mathrm{A}\colon \mathrm{V} \to \mathrm{V}$ *be a self-adjoint linear map. Then there exists an orthonormal basis* $\{e_1, e_2\}$ *of* V *such that* $\mathrm{A}(\mathrm{e}_1) = \lambda_1 \mathrm{e}_1$, $\mathrm{A}(\mathrm{e}_2) = \lambda_2 \mathrm{e}_2$ (*that is*, e_1 *and* e_2 *are eigenvectors, and* λ_1, λ_2 *are eigenvalues of* A). *In the basis* $\{e_1, e_2\}$, *the matrix of* A *is clearly diagonal and the elements* λ_1, λ_2, $\lambda_1 \geq \lambda_2$, *on the diagonal are the maximum and the minimum, respectively, of the quadratic form* $\mathrm{Q}(\mathrm{v}) = \langle \mathrm{Av}, \mathrm{v} \rangle$ *on the unit circle of* V.

Proof. Consider the quadratic form $Q(v) = \langle Av, v \rangle$. By the proposition above, there exists an orthonormal basis $\{e_1, e_2\}$ of V, with $Q(e_1) = \lambda_1$, $Q(e_2) = \lambda_2 \leq \lambda_1$, where λ_1 and λ_2 are the maximum and minimum, respectively, of Q in the unit circle. It remains, therefore, to prove that

$$A(e_1) = \lambda_1 e_1, \quad A(e_2) = \lambda_2(e_2).$$

Since $B(e_1, e_2) = \langle Ae_1, e_2 \rangle = 0$ (by the lemma) and $e_2 \neq 0$, we have that either Ae_1 is parallel to e_1 or $Ae_1 = 0$. If Ae_1 is parallel to e_1, then $Ae_1 = \alpha e_1$, and since $\langle Ae_1, e_1 \rangle = \lambda_1 = \langle \alpha e_1, e_2 \rangle = \alpha$, we conclude that $Ae_1 = \lambda_1 e_1$; if $Ae_1 = 0$, then $\lambda_1 = \langle Ae_1, e_1 \rangle = 0$, and $Ae_1 = 0 = \lambda_1 e_1$. Thus, we have in any case that $Ae_1 = \lambda_1 e_1$.

Now using the fact that

$$B(e_1, e_2) = \langle Ae_2, e_1 \rangle = 0$$

and that

$$\langle Ae_2, e_2 \rangle = \lambda_2,$$

we can prove in the same way that $Ae_2 = \lambda_2 e_2$. **Q.E.D.**

Remark. The extension of the above results to an n-dimensional vector space, $n > 2$, requires only the following precaution. In the previous proposition, we choose the maximum $\lambda_1 = Q(e_1)$ of Q in the unit sphere, and then show that Q restricts to a quadratic form Q_1 in the subspace V_1 orthogonal to e_1. We choose for $\lambda_2 = Q_1(e_2)$ the maximum of Q_1 in the unit sphere of V_1, and so forth.