Chapter 4
Calculus on a Surface

This chapter begins with the definition of a surface in $\mathbf{R}^3$ and with some standard ways to construct surfaces. Although this concept is a more-or-less familiar one, it is not as widely known as it should be that each surface has a differential and integral calculus strictly comparable with the usual calculus on the Euclidean plane $\mathbf{R}^2$. The elements of this calculus—functions, vector fields, differential forms, mappings—belong strictly to the surface and not to the Euclidean space $\mathbf{R}^3$ in which the surface is located. Indeed, we shall see in the final section that this calculus survives undamaged when $\mathbf{R}^3$ is removed, leaving just the surface and nothing more.

4.1 Surfaces in $\mathbf{R}^3$

A surface in $\mathbf{R}^3$ is, to begin with, a *subset* of $\mathbf{R}^3$, that is, a certain collection of points of $\mathbf{R}^3$. Of course, not all subsets are surfaces: We must certainly require that a surface be smooth and two-dimensional. These requirements will be expressed in mathematical terms by the next two definitions.

1.1 Definition A *coordinate patch* $\mathbf{x}\colon D \to \mathbf{R}^3$ is a one-to-one regular mapping of an open set D of $\mathbf{R}^2$ into $\mathbf{R}^3$.

The image $\mathbf{x}(D)$ of a coordinate patch $\mathbf{x}$—that is, the set of all values of $\mathbf{x}$—is a smooth two-dimensional subset of $\mathbf{R}^3$ (Fig. 4.1). Regularity (Definition 7.9 of Chapter 1), for a patch as for a curve, is a basic smoothness condition; the one-to-one requirement is included to prevent $\mathbf{x}(D)$ from cutting across itself. Initially, in order to avoid certain technical difficulties (Example

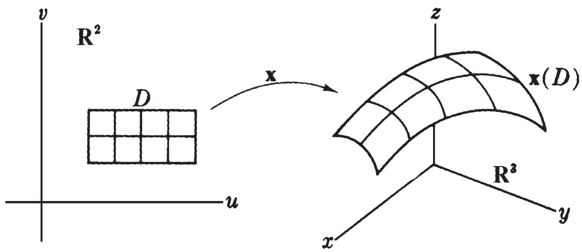


FIG. 4.1

1.6), we must use *proper* patches, those for which the inverse function $\mathbf{x}^{-1}$: $\mathbf{x}(D) \to D$ is continuous (that is, has continuous coordinate functions). If we think of D as a thin sheet of rubber, then $\mathbf{x}(D)$ is gotten by bending and stretching D in a not too violent fashion.

To construct a suitable definition of surface we start from the rough idea that *any small enough region in a surface M resembles a region in the plane* $\mathbf{R}^2$. The discussion above shows that this can be stated somewhat more precisely as, *near each of its points, M can be expressed as the image of a proper patch.* (When the image of a patch $\mathbf{x}$ is contained in M, we say that $\mathbf{x}$ is a patch *in* M.) To get the final form of the definition, it remains only to define a *neighborhood* $\mathscr{N}$ *of* $\mathbf{p}$ *in M* to consist of all points of M whose Euclidean distance from $\mathbf{p}$ is less than some number $\varepsilon > 0$.

1.2 Definition A *surface in* $\mathbf{R}^3$ is a subset M of $\mathbf{R}^3$ such that for each point $\mathbf{p}$ of M there exists a proper patch in M whose image contains a neighborhood of $\mathbf{p}$ in M (Fig. 4.2).

The familiar surfaces used in elementary calculus satisfy this definition; for example, let us verify that the unit sphere Σ in $\mathbf{R}^3$ is a surface. By definition,

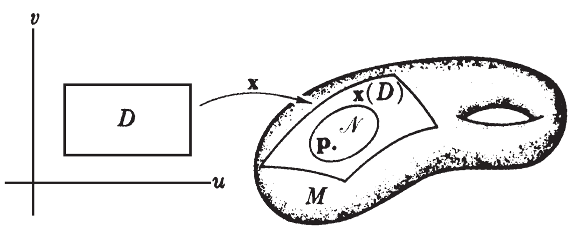


FIG. 4.2

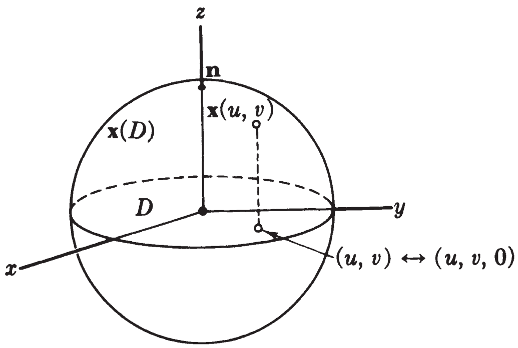


FIG. 4.3

Σ consists of all points at unit distance from the origin—that is, all points **p** such that

$$\| \mathbf{p} \| = (p_1^2 + p_2^2 + p_3^2)^{1/2} = 1.$$

To check the definition above, we start by finding a proper patch in Σ covering a neighborhood of the north pole (0, 0, 1). Note that by dropping each point (q_1, q_2, q_3) of the northern hemisphere of Σ onto the xy plane at $(q_1, q_2, 0)$ we get a one-to-one correspondence of this hemisphere with a disk D of radius 1 in the xy plane (see Fig. 4.3). If this plane is identified with $\mathbf{R}^2$ by means of the natural association $(q_1, q_2, 0) \leftrightarrow (q_1, q_2)$, then D becomes the disk in $\mathbf{R}^2$ consisting of all points (u, v) such that $u^2 + v^2 < 1$. Expressing this correspondence as a function on D yields the formula

$$\mathbf{x}(u, v) = \left(u, v, \sqrt{1 - u^2 - v^2}\right).$$

Thus **x** is a one-to-one function from D onto the northern hemisphere of Σ. We claim that **x** is a proper patch. The coordinate functions of **x** are differentiable on D, so **x** is a mapping. To show that **x** is regular, we compute its Jacobian matrix (or transpose)

$$\begin{pmatrix} \dfrac{\partial u}{\partial u} & \dfrac{\partial v}{\partial u} & \dfrac{\partial f}{\partial u} \\[2ex] \dfrac{\partial u}{\partial v} & \dfrac{\partial v}{\partial v} & \dfrac{\partial f}{\partial v} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dfrac{\partial f}{\partial u} \\[2ex] 0 & 1 & \dfrac{\partial f}{\partial v} \end{pmatrix}.$$

where $f = \sqrt{1 - u^2 - v^2}$. Evidently the rows of this matrix are always linearly independent, so its rank at each point is 2. Thus, by the criterion following Definition 7.9 of Chapter 1, **x** is regular and hence is a patch. Furthermore, **x** is proper, since its inverse function $\mathbf{x}^{-1}\colon \mathbf{x}(D) \to D$ is given by the formula

$$\mathbf{x}^{-1}(p_1, p_2, p_3) = (p_1, p_2)$$

and hence is certainly continuous. Finally, we observe that the patch $\mathbf{x}$ covers a neighborhood of (0, 0, 1) in Σ. Indeed, it covers a neighborhood of every point $\mathbf{q}$ in the northern hemisphere of Σ.

In a strictly analogous way, we can find a proper patch covering each of the other five coordinate hemispheres of Σ, and thus verify, by Definition 1.2, that Σ is a surface. Our real purpose here has been to illustrate Definition 1.2—we soon find a much quicker way to prove (in particular) that spheres are surfaces.

The argument above shows that if f is *any* differentiable real-valued function on an open set D in $\mathbf{R}^2$, then the function $\mathbf{x}$: $D \to \mathbf{R}^3$ such that

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

is a proper patch. We shall call patches of this type *Monge patches.*

We turn now to some standard methods of constructing surfaces. Note that the image $M = \mathbf{x}(D)$ of just one proper patch automatically satisfies 1.2; M is then called a *simple* surface. (Thus Definition 1.2 says that any surface in $\mathbf{R}^3$ can be constructed by gluing together simple surfaces.)

1.3 Example The surface M: $z = f(x, y)$. Every differentiable real-valued function f on $\mathbf{R}^2$ determines a surface M in $\mathbf{R}^3$: the *graph* of f, that is, the set of all points of $\mathbf{R}^3$ whose coordinates satisfy the equation $z = f(x, y)$. Evidently M is the image of the Monge patch

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

hence by the remarks above, M is a simple surface.

If g is a real-valued function on $\mathbf{R}^3$ and c is a number, denote by M: $g = c$ the set of all points $\mathbf{p}$ such that $g(\mathbf{p}) = c$. For example, if g is a temperature distribution in space, then M: $g = c$ consists of all points of temperature c. There is a simple condition that tells when such a subset of $\mathbf{R}^3$ is a surface.

1.4 Theorem Let g be a differentiable real-valued function on $\mathbf{R}^3$, and c a number. The subset M: $g(x, y, z) = c$ of $\mathbf{R}^3$ is a surface if the differential dg is not zero at any point of M.

(In Definition 1.2 and in this theorem we are tacitly assuming that M has some points in it; thus the equation $x^2 + y^2 + z^2 = -1$, for example, does not define a surface.)

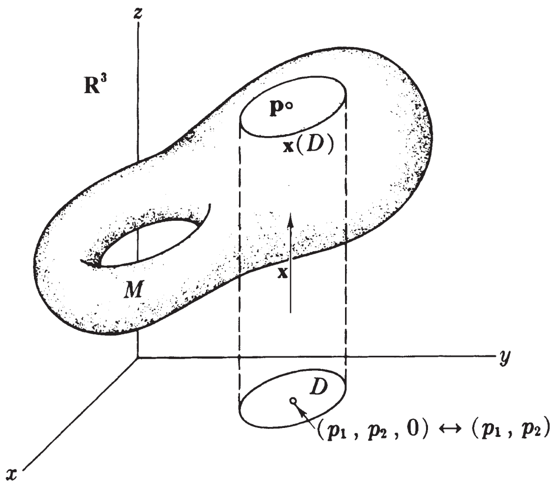


FIG. 4.4

Proof. All we do is give geometric content to a famous result of advanced calculus—the implicit function theorem. If **p** is a point of M, we must find a proper patch covering a neighborhood of **p** in M (Fig. 4.4). Since

$$dg = \frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy + \frac{\partial g}{\partial z}dz,$$

the hypothesis on dg is equivalent to assuming that at least one of these partial derivatives is not zero at **p**, say $(\partial g/\partial z)(\mathbf{p}) \neq 0$. In this case, the implicit function theorem says that near **p** the equation $g(x, y, z) = c$ can be solved for z. More precisely, it asserts that there is a differentiable real-valued function h defined on a neighborhood D of (p_1, p_2) such that

(1) For each point (u, v) in D, the point $(u, v, h(u, v))$ lies in M; that is, $g(u, v, h(u, v)) = c$.

(2) Points of the form $(u, v, h(u, v))$, with (u, v) in D, fill a neighborhood of **p** in M.

It follows immediately that the Monge patch $\mathbf{x}\colon D \to \mathbf{R}^3$ such that

$$x(u, v) = (u, v, h(u, v))$$

satisfies the requirements in Definition 1.2. Since **p** was an arbitrary point of M, we conclude that M is a surface. ◆

From now on we use the notation $M\colon g = c$ only when $dg \neq 0$ on M. Then M is a surface said to be defined *implicitly* by the equation $g = c$. It is now easy to prove that spheres are surfaces. The sphere Σ in $\mathbf{R}^3$ of radius $r > 0$ and center $\mathbf{c} = (c_1, c_2, c_3)$ is the set of all points at distance r from c. If $g = \sum (x_i - c_i)^2$ then Σ is defined implicitly by the equation $g = r^2$. Now,

$$dg = 2\sum (x_i - c_i) dx_i.$$

Hence dg is zero only at the point $\mathbf{c}$, which is not in Σ. Thus Σ is a surface.

An important class of surfaces is gotten by rotating curves.

1.5 Example Surfaces of revolution. Let C be a curve in a plane $P \subset \mathbf{R}^3$, and let A be a line in P that does not meet C. When this *profile curve* C is revolved around the *axis* A, it sweeps out a *surface of revolution* M in $\mathbf{R}^3$.

Let us check that M really is a surface. For simplicity, suppose that P is a coordinate plane and A is a coordinate axis—say, the xy plane and x axis, respectively. Since C must not meet A, we put it in the upper half, $y > 0$, of the xy plane. As C is revolved, each of its points $(q_1, q_2, 0)$ gives rise to a whole circle of points

$$(q_1, q_2 \cos v, q_2 \sin v) \quad \text{for } 0 \leqq v \leqq 2\pi.$$

Thus *a point* $\mathbf{p} = (p_1, p_2, p_3)$ *is in* M *if and only if the point*

$$\bar{\mathbf{p}} = \left(p_1, \sqrt{p_2^2 + p_3^2}, 0\right)$$

is in C (Fig. 4.5).

If the profile curve is C: $f(x, y) = c$, we define a function g on $\mathbf{R}^3$ by

$$g(x, y, z) = f\left(x, \sqrt{y^2 + z^2}\right).$$

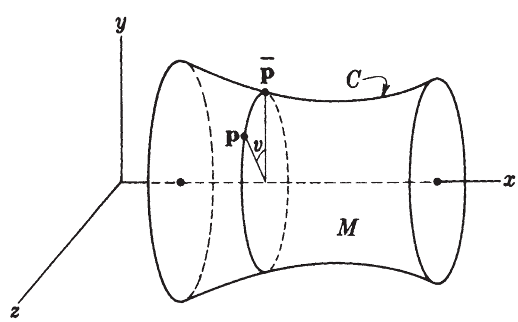


FIG. 4.5

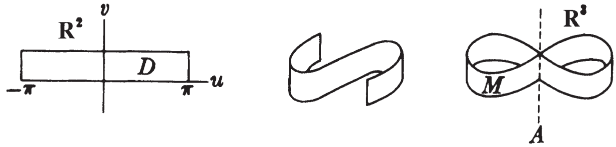


FIG. 4.6

Then the argument above shows that the resulting surface of revolution is exactly M: $g(x, y, z) = c$. Using the chain rule, it is not hard to show that dg is never zero on M, so M is a surface.

The circles in M generated under revolution by each point of C are called the *parallels* of M; the different positions of C as it is rotated are called the *meridians* of M. This terminology derives from the geography of the sphere; however, a sphere is *not* a surface of revolution as defined above. Its profile curve must twice meet the axis of revolution, so two "parallels" reduce to single points. To simplify the statements of later theorems, we use a slightly different terminology in this case; see Exercise 12.

The necessity of the properness condition on the patches in Definition 1.2 is shown by the following example.

1.6 Example Suppose that a rectangular strip of tin is bent into a figure 8, as in Fig. 4.6. The configuration M that results does not satisfy our intuitive picture of what a surface should be, for along the axis A, M is not like the plane $\mathbf{R}^2$ but is instead like *two* intersecting planes. To express this construction in mathematical terms, let D be the rectangle $-\pi < u < \pi$, $0 < v < 1$ in $\mathbf{R}^2$ and define $\mathbf{x}\colon D \to \mathbf{R}^3$ by $\mathbf{x}(u, v) = (\sin u, \sin 2u, v)$. It is easy to check that $\mathbf{x}$ is a patch, but its image $M = \mathbf{x}(D)$ is not a surface: $\mathbf{x}$ is not a *proper* patch. Continuity fails for $\mathbf{x}^{-1}\colon M \to D$ since, roughly speaking, to restore M to D, $\mathbf{x}^{-1}$ must tear M along the axis A (the z axis of $\mathbf{R}^3$).

By Example 1.5, the familiar *torus of revolution* T is a surface (Fig. 4.16). With somewhat more work, one could construct *double toruses* of various shapes, as in Fig. 4.7. By adding "handles" and "tubes" to existing surfaces one can—in principle, at least—construct surfaces of any desired degree of complexity (Fig. 4.8).

FIG. 4.7

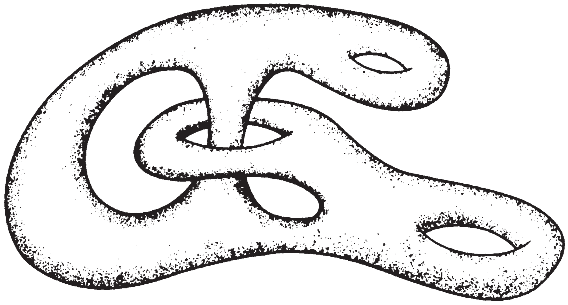


FIG. 4.8

Exercises

1. None of the following subsets M of $\mathbf{R}^3$ are surfaces. At which points $\mathbf{p}$ is it impossible to find a proper patch in M that will cover a neighborhood of $\mathbf{p}$ in M? (Sketch M—formal proofs not required.)

(a) Cone M: $z^2 = x^2 + y^2$

(b) Closed disk M: $x^2 + y^2 \leqq 1$, $z = 0$.

(c) Folded plane M: $xy = 0$, $x \geqq 0$, $y \geqq 0$.

2. A *plane* in $\mathbf{R}^3$ is a surface M: $ax + by + cz = d$, where the numbers a, b, c are necessarily not all zero. Prove that every plane in $\mathbf{R}^3$ may be described by a vector equation as on page 62.

3. Sketch the general shape of the surface M: $z = ax^2 + by^2$ in each of the following cases:

(a) $a > b > 0$. (b) $a > 0 > b$.

(c) $a > b = 0$. (d) $a = b = 0$.

4. In which of the following cases is the mapping $\mathbf{x}$: $\mathbf{R}^2 \to \mathbf{R}^3$ a patch?

(a) $\mathbf{x}(u, v) = (u, uv, v)$. (b) $\mathbf{x}(u, v) = (u^2, u^3, v)$.

(c) $\mathbf{x}(u, v) = (u, u^2, v + v^3)$. (d) $\mathbf{x}(u, v) = (\cos 2\pi u, \sin 2\pi u, v)$.

(Recall that $\mathbf{x}$ is one-to-one if and only if $\mathbf{x}(u, v) = \mathbf{x}(u_1, v_1)$ implies $(u, v) = (u_1, v_1)$.)

5. (a) Prove that M: $(x^2 + y^2)^2 + 3z^2 = 1$ is a surface.

(b) For which values of c is M: $z(z - 2) + xy = c$ a surface?

6. Determine the intersection $z = 0$ of the *monkey saddle*

$$M: z = f(x, y), \quad f(x, y) = y^3 - 3yx^2,$$

with the xy plane. On which regions of the plane is $f > 0$? $f < 0$? How does this surface get its name? (*Hint:* see Fig. 5.19.)

7. Let $\mathbf{x}: D \to \mathbf{R}^3$ be a mapping, with

$$\mathbf{x}(u, v) = (x_1(u, v), x_2(u, v), x_3(u, v)).$$

(a) Prove that a point $\mathbf{p} = (p_1, p_2, p_3)$ of $\mathbf{R}^3$ is in the image $\mathbf{x}(D)$ if and only if the equations

$$p_1 = x_1(u, v), \quad p_2 = x_2(u, v), \quad p_3 = x_3(u, v)$$

can be solved for u and v, with (u, v) in D.

(b) If for every point $\mathbf{p}$ in $\mathbf{x}(D)$ these equations have the *unique* solution $u = f_1(p_1, p_2, p_3)$, $v = f_2(p_1, p_2, p_3)$, with (u, v) in D, prove that $\mathbf{x}$ is one-to-one and that $\mathbf{x}^{-1}: \mathbf{x}(D) \to D$ is given by the formula

$$\mathbf{x}^{-1}(\mathbf{p}) = (f_1(\mathbf{p}), f_2(\mathbf{p})).$$

8. Let $\mathbf{x}: D \to \mathbf{R}^3$ be the function given by

$$\mathbf{x}(u, v) = (u^2, uv, v^2)$$

on the first quadrant D: $u > 0$, $v > 0$. Show that $\mathbf{x}$ is one-to-one and find a formula for its inverse function $\mathbf{x}^{-1}: \mathbf{x}(D) \to D$. Then prove that $\mathbf{x}$ is a proper patch.

9. Let $\mathbf{x}: \mathbf{R}^2 \to \mathbf{R}^3$ be the mapping

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

Show that $\mathbf{x}$ is a proper patch and that the image of $\mathbf{x}$ is the entire surface M: $z = (x^2 - y^2)/4$.

10. If $F: \mathbf{R}^3 \to \mathbf{R}^3$ is a diffeomorphism and M is a surface in $\mathbf{R}^3$, prove that the image $F(M)$ is also a surface in $\mathbf{R}^3$. (*Hint:* If $\mathbf{x}$ is a patch in M, then the composite function $F(\mathbf{x})$ is regular, since $F(\mathbf{x})_* = F_*\mathbf{x}_*$ by Ex. 9 of Sec. 1.7.)

11. Prove this special case of Exercise 10: If F is a diffeomorphism of $\mathbf{R}^3$, then the image of the surface M: $g = c$ is $\bar{M}$: $\bar{g} = c$, where $\bar{g} = g(F^{-1})$ and $\bar{M}$ is a surface. (*Hint:* If $dg(\mathbf{v}) \neq 0$ at $\mathbf{p}$ in M, show by using Ex. 7 of Sec. 1.7 that $d\bar{g}(F_*\mathbf{v}) \neq 0$ at $F(\mathbf{p})$.)

12. Let C be a Curve in the xy plane that is symmetric about the x axis. Assume C crosses the x axis and always does so orthogonally. Explain why there can be only one or two crossings. Thus C is either an arc or is closed (Fig. 4.9). Revolving C about the x axis gives a surface M, called an *augmented surface of revolution*. Explain how to define patches in M at the crossing points.

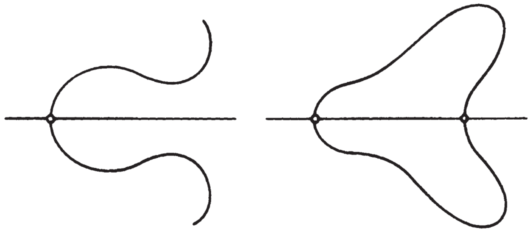


FIG. 4.9

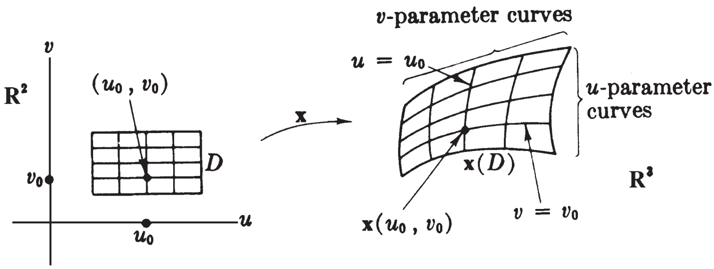


FIG. 4.10

4.2 Patch Computations

In Section 1, coordinate patches were used to define a surface; now we consider some properties of patches that will be useful in studying surfaces.

Let $\mathbf{x}: D \to \mathbf{R}^3$ be a coordinate patch. Holding u or v constant in the function $(u, v) \to \mathbf{x}(u, v)$ produces curves. Explicitly, for each point (u_0, v_0) in D the curve

$$u \to \mathbf{x}(u, v_0)$$

is called the *u-parameter curve*, $v = v_0$, of $\mathbf{x}$; and the curve

$$v \to \mathbf{x}(u_0, v)$$

is the *v-parameter curve*, $u = u_0$ (Fig. 4.10).

Thus, the image $\mathbf{x}(D)$ is covered by these two families of curves, which are the images under $\mathbf{x}$ of the horizontal and vertical lines in D, and one curve from each family goes through each point of $\mathbf{x}(D)$.

2.1 Definition If $\mathbf{x}: D \to \mathbf{R}^3$ is a patch, for each point (u_0, v_0) in D:

(1) The velocity vector at u_0 of the u-parameter curve, $v = v_0$, is denoted by $\mathbf{x}_u(u_0, v_0)$.

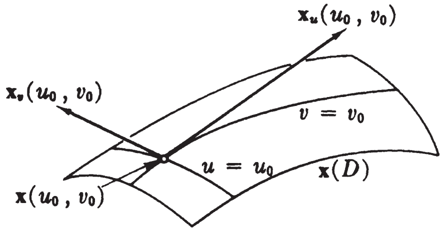


FIG. 4.11

(2) The velocity vector at v_0 of the v-parameter curve, $u = u_0$, is denoted by $\mathbf{x}_v(u_0, v_0)$.

The vectors $\mathbf{x}_u(u_0, v_0)$ and $\mathbf{x}_v(u_0, v_0)$ are called the *partial velocities* of $\mathbf{x}$ at (u_0, v_0) (Fig. 4.11).

Thus $\mathbf{x}_u$ and $\mathbf{x}_v$ are functions on D whose values at each point (u_0, v_0) are tangent vectors to $\mathbf{R}^3$ at $\mathbf{x}(u_0, v_0)$. The subscripts u and v are intended to suggest partial differentiation. Indeed if the patch is given in terms of its Euclidean coordinate functions by a formula

$$\mathbf{x}(u, v) = (x_1(u, v), x_2(u, v), x_3(u, v)),$$

then it follows from the definition above that the partial velocity functions are given by

$$\mathbf{x}_u = \left(\frac{\partial x_1}{\partial u}, \frac{\partial x_2}{\partial u}, \frac{\partial x_3}{\partial u}\right)_x,$$

$$\mathbf{x}_v = \left(\frac{\partial x_1}{\partial v}, \frac{\partial x_2}{\partial v}, \frac{\partial x_3}{\partial v}\right)_x.$$

The subscript $\mathbf{x}$ (frequently omitted) is a reminder that $\mathbf{x}_u(u, v)$ and $\mathbf{x}_v(u, v)$ have point of application $\mathbf{x}(u, v)$.

2.2 Example The *geographical patch* in the sphere. Let Σ be the sphere of radius $r > 0$ centered at the origin of $\mathbf{R}^3$. Longitude and latitude on the earth suggest a patch in Σ quite different from the Monge patch used on Σ in Section 1. The point $\mathbf{x}(u, v)$ of Σ with longitude u $(-\pi < u < \pi)$ and latitude v $(-\pi/2 < v < \pi/2)$ has Euclidean coordinates (Fig. 4.12).

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

With the domain D of $\mathbf{x}$ defined by these inequalities, the image $\mathbf{x}(D)$ of $\mathbf{x}$ is all of Σ except one semicircle from north pole to south pole. The u-

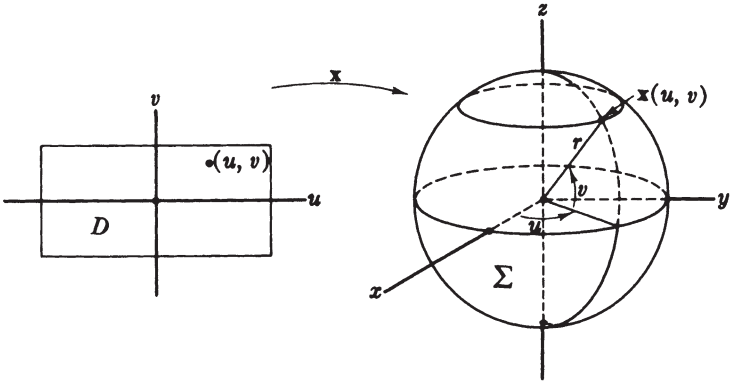


FIG. 4.12

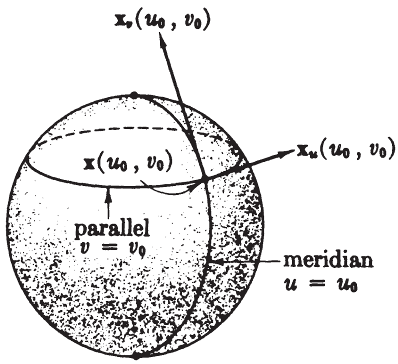


FIG. 4.13

parameter curve, $v = v_0$, is a circle—the parallel of latitude v_0. The v-parameter curve, $u = u_0$, is a semicircle—the meridian of longitude u_0.

We compute the partial velocities of $\mathbf{x}$ to be

$$\mathbf{x}_u(u, v) = r\,(-\cos v \sin u, \cos v \cos u, 0),$$

$$\mathbf{x}_v(u, v) = r\,(-\sin v \cos u, -\sin v \sin u, \cos v).$$

where r denotes a scalar multiplication. Evidently $\mathbf{x}_u$ always points due east, and $\mathbf{x}_v$ due north (Fig. 4.13). In a moment we shall give a formal proof that $\mathbf{x}$ is a patch in Σ.

To test whether a given subset M of $\mathbf{R}^3$ is a surface, Definition 1.2 demands proper patches (and Example 1.6 shows why). But once we know that M is a surface, the properness condition need no longer concern us (Exercise 14 of Section 3). Furthermore, in many situations the one-to-one restriction on patches can also be dropped.

2.3 Definition A regular mapping $\mathbf{x}\colon D \to \mathbf{R}^3$ whose image lies in a surface M is called a *parametrization* of the region $\mathbf{x}(D)$ in M.

(Thus a patch is merely a one-to-one parametrization.) In favorable cases this image $\mathbf{x}(D)$ may be the whole surface M, and we then have the analogue of the more familiar notion of parametrization of a Curve (see end of Section 1.4). Parametrizations will be of first importance in practical computations with surfaces, so we consider some ways of determining whether a mapping $\mathbf{x}\colon D \to \mathbf{R}^3$ is a parametrization of (part of) a given surface M.

The image of $\mathbf{x}$ must, of course, lie in M. Note that if the surface is given in the implicit form $M\colon g = c$, this means that the composite function $g(\mathbf{x})$ must have constant value c.

To test whether $\mathbf{x}$ is regular, note first that parameter curves and partial velocities $\mathbf{x}_u$ and $\mathbf{x}_v$ are well-defined for an arbitrary differentiable mapping $\mathbf{x}\colon D \to \mathbf{R}^3$. Also, the last two rows of the cross product

$$\mathbf{x}_u \times \mathbf{x}_v = \begin{vmatrix} U_1 & U_2 & U_3 \\ \dfrac{\partial x_1}{\partial u} & \dfrac{\partial x_2}{\partial u} & \dfrac{\partial x_3}{\partial u} \\ \dfrac{\partial x_1}{\partial v} & \dfrac{\partial x_2}{\partial v} & \dfrac{\partial x_3}{\partial v} \end{vmatrix}$$

give the (transposed) Jacobian matrix of $\mathbf{x}$ at each point. Thus the regularity of $\mathbf{x}$ is equivalent to the condition that $\mathbf{x}_u \times \mathbf{x}_v$ *is never zero*, or, by properties of the cross product, that at each point (u,v) of D *the partial velocity vectors of* $\mathbf{x}$ *are linearly independent.*

Let us try out these methods on the mapping $\mathbf{x}$ given in Example 2.2. Since the sphere is defined implicitly by $g = x^2 + y^2 + z^2 = r^2$, we must show that $g(\mathbf{x}) = r^2$. Substituting the coordinate functions of $\mathbf{x}$ for x, y, and z gives

$$\begin{aligned} r^{-2}g(\mathbf{x}) &= (\cos v \cos u)^2 + (\cos v \sin u)^2 + \sin^2 v \\ &= \cos^2 v + \sin^2 v = 1. \end{aligned}$$

A short computation using the formulas for $\mathbf{x}_u$ and $\mathbf{x}_v$, given in Example 2.2, yields

$$r^{-2}\mathbf{x}_u \times \mathbf{x}_v = \cos u \cos^2 v U_1 + \sin u \cos^2 v U_2 + \cos v \sin v U_3.$$

Since $-\pi/2 < v < \pi/2$ in the domain D of $\mathbf{x}$, $\cos v$ is never zero there; but $\sin u$ and $\cos u$ are never zero simultaneously, so $\mathbf{x}_u \times \mathbf{x}_v$, is never zero on D. Thus $\mathbf{x}$ is regular—and hence is a parametrization. In fact, it remains a parametrization if the condition $-\pi < u < \pi$ is dropped, thus replacing D by the infinite strip $-\pi/2 < v < \pi/2$. In this case the u-parameter curves are periodic parametrizations of the meridians, and $\mathbf{x}$ covers the entire sphere except for the poles $(0, 0, \pm 1)$.

To show that $\mathbf{x}$ on the original domain D is a patch, it remains only to show that it is one-to-one on D, that is,

$$\mathbf{x}(u, v) = \mathbf{x}(u_1, v_1) \Rightarrow (u, v) = (u_1, v_1).$$

In view of the definition of $\mathbf{x}$, the vector equation here gives the three scalar equations

$$r \cos v \cos u = r \cos v_1 \cos u_1,$$

$$r \cos v \sin u = r \cos v_1 \sin u_1,$$

$$r \sin v = r \sin v_1.$$

Since $-\pi/2 < v < \pi/2$ in D, the last equation implies $v = v_1$. Thus $r \cos v = r \cos v_1 > 0$ can be canceled from the first two equations, and we conclude that $u = u_1$ as well.

The geographical definition of $\mathbf{x}$ in Example 2.2 makes the preceding results seem almost obvious, but the methods used will serve in more difficult cases.

2.4 Example Parametrization of a surface of revolution. Suppose that M is obtained, as in Example 1.5, by revolving a curve C in the upper half of the xy plane about the x axis. Now let

$$\alpha(u) = (g(u), h(u), 0)$$

be a parametrization of C (note that $h > 0$). As we observed in Example 1.5, when the point $(g(u), h(u), 0)$ on the profile curve C has been rotated through an angle v, it reaches a point $\mathbf{x}(u, v)$ with the same x coordinate $g(u)$, but new y and z coordinates $h(u) \cos v$ and $h(u) \sin v$, respectively (Fig. 4.14). Thus

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

Evidently this formula defines a mapping into M whose image is all of M. A short computation shows that $\mathbf{x}_u$ and $\mathbf{x}_v$ are always linearly independent, so $\mathbf{x}$ is a parametrization of M. The domain D of $\mathbf{x}$ consists of all points (u, v) for which u is in the domain of α. The u-parameter curves of $\mathbf{x}$ parametrize

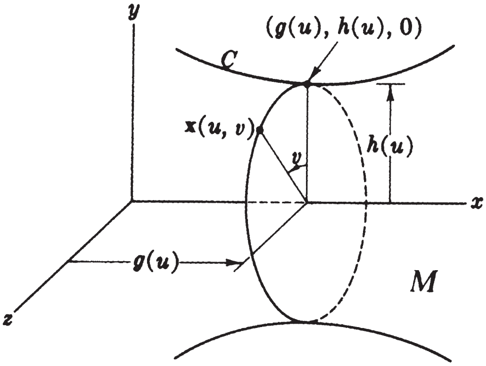


FIG. 4.14

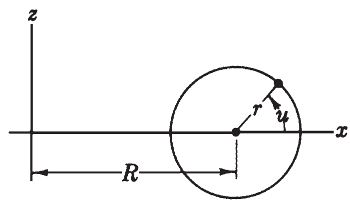


FIG. 4.15

the meridians of M, the v-parameter curves the parallels. (Thus the parametrization $\mathbf{x}$: $D \to M$ is never one-to-one.)

Obviously we are not limited to rotating curves in the xy plane about the x axis. But with other choices of coordinates, we maintain the same geometric meaning for the functions g and h: g measures distance *along* the axis of revolution, while h measures distance *from* the axis of revolution.

Actually, the geographical patch in the sphere is one instance of Example 2.4 (with u and v reversed); here is another.

2.5 Example *Torus of revolution T.* This is the surface of revolution obtained when the profile curve C is a circle. Suppose that C is the circle in the xz plane with radius $r > 0$ and center $(R, 0, 0)$. We shall rotate about the z axis; hence we must require $R > r$ to keep C from meeting the axis of revolution. A natural parametrization (Fig. 4.15) for C is

$$\alpha(u) = (R + r \cos u, r \sin u).$$

Thus by the remarks above we must have $g(u) = r \sin u$ (distance *along* the z axis) and $h(u) = R + r \cos u$ (distance *from* the z axis). The general argument in Example 2.4—with coordinate axes permuted—then yields the parametrization

$$\begin{aligned} \mathbf{x}(u, v) &= (h(u) \cos v, h(u) \sin v, g(u)) \\ &= ((R + r \cos u) \cos v, (R + r \cos u) \sin v, r \sin u). \end{aligned}$$

We call $\mathbf{x}$ the *usual parametrization* of the torus (Fig. 4.16). Its domain is the whole plane $\mathbf{R}^2$, and it is periodic in both u and v:

$$\mathbf{x}(u + 2\pi, v) = \mathbf{x}(u, v + 2\pi) = \mathbf{x}(u, v) \quad \text{for all } (u, v).$$

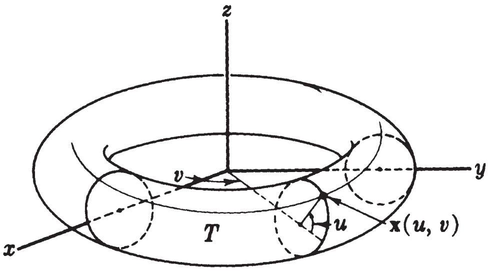


FIG. 4.16

2.6 Definition A *ruled surface* is a surface swept out by a straight line L moving along a curve β. The various positions of the generating line L are called the *rulings* of the surface. Such a surface always has a *ruled parametrization*,

$$\mathbf{x}(u, v) = \beta(u) + v\delta(u).$$

We call β the *base curve* and δ the *director curve*, although δ is usually pictured as a vector field on β pointing along the line L.

Several examples of ruled surfaces are given in the following exercises. It is usually necessary to put restrictions on β and δ to ensure that $\mathbf{x}$ is a parametrization.

There are infinitely many different parametrizations and patches in any surface. Those we have discussed occur frequently and are fitted in a natural way to their surfaces.

Exercises

1. Find a parametrization of the entire surface obtained by revolving:
(a) C: $y = \cosh x$ around the x axis (catenoid).
(b) C: $(x - 2)^2 + y^2 = 1$ around the y axis (torus).
(c) C: $z = x^2$ around the z axis (paraboloid).

2. Partial velocities $\mathbf{x}_u$ and $\mathbf{x}_v$ are defined for an arbitrary mapping $\mathbf{x}$: $D \to \mathbf{R}^3$, so we can consider the real-valued functions

$$E = \mathbf{x}_u \bullet \mathbf{x}_u, \quad F = \mathbf{x}_u \bullet \mathbf{x}_v, \quad G = \mathbf{x}_v \bullet \mathbf{x}_v$$

on D. Prove

$$\|\mathbf{x}_u \times \mathbf{x}_v\|^2 = EG - F^2.$$

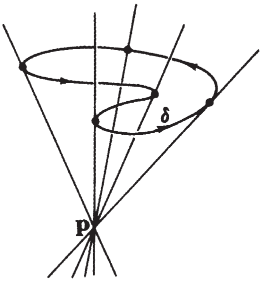


FIG. 4.17

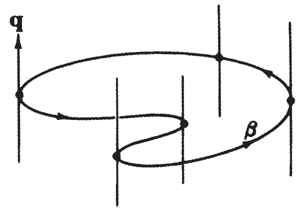


FIG. 4.18

Deduce that **x** is a regular mapping if and only if $EG - F^2$ is never zero. (This is often the easiest way to check regularity. We will see, beginning in the next chapter, that the functions E, F, G are fundamental to the geometry of surfaces.)

3. A *generalized cone* is a ruled surface with a parametrization of the form

$$\mathbf{x}(u, v) = \mathbf{p} + v\delta(u).$$

Thus all rulings pass through the vertex **p** (Fig. 4.17). Show that **x** is regular if and only if v and $\delta \times \delta'$ are never zero. (Thus the vertex is never part of the cone. Unless the term *generalized* is used, we assume that δ is a closed curve and require either $v > 0$ or $v < 0$.)

4. A *generalized cylinder* is a ruled surface for which the rulings are all Euclidean parallel (Fig. 4.18). Thus there is always a parametrization of the form

$$\mathbf{x}(u, v) = \beta(u) + v\mathbf{q} \quad (\mathbf{q} \in \mathbf{R}^3).$$

Prove: (a) Regularity of **x** is equivalent to $\beta' \times \mathbf{q}$ never zero.

(b) If $C: f(x, y) = a$ is a Curve in the plane, show that in $\mathbf{R}^3$ the same equation defines a surface $\tilde{C}$. If $t \to (x(t), y(t))$ is a parametrization of C, find a parametrization of $\tilde{C}$ that shows it is a generalized cylinder.

Generalized cylinders are a rather broad category—including Euclidean planes when β is a straight line—so unless the term *generalized* is used, we assume that cylinders are over *closed* curves β.

5. A line L is attached orthogonally to an axis A (Fig. 4.19). If L moves steadily along A, rotating at constant speed, then L sweeps out a *helicoid* H.

When A is the z axis, H is the image of the mapping $\mathbf{x}: \mathbf{R}^2 \to \mathbf{R}^3$ such that

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

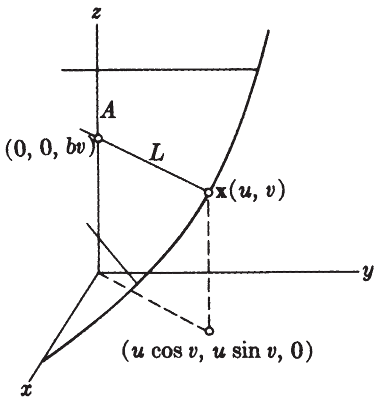


FIG. 4.19

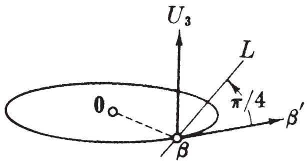


FIG. 4.20

(a) Prove that **x** is a patch.
(b) Describe its parameter curves.
(c) Express the helicoid in the implicit form $g = c$.
(d) (*Computer graphics.*) Plot one full turn ($0 \leqq v \leqq 2\pi$) of a helicoid with $b = 1/2$. Restrict the rulings to $-1 \leqq u \leqq 1$.

6. (a) Show that the *saddle surface* M: $z = xy$ is doubly ruled: Find two ruled parametrizations with different rulings.
(b) (*Computer graphics.*) Plot a representative portion of M, using a patch for which the parameter curves are rulings.

7. Let β be a unit-speed parametrization of the unit circle in the xy plane. Construct a ruled surface as follows: Move a line L along β in such a way that L is always orthogonal to the radius of the circle and makes constant angle $\pi/4$ with β' (Fig. 4.20).
(a) Derive this parametrization of the resulting ruled surface M:

$$\mathbf{x}(u, v) = \beta(u) + v(\beta'(u) + U_3).$$

(b) Express **x** explicitly in terms of v and coordinate functions for β.
(c) Deduce that M is given implicitly by the equation

$$x^2 + y^2 - z^2 = 1.$$

(d) Show that if the angle $\pi/4$ above is changed to $-\pi/4$, the same surface M results. Thus M is doubly ruled.
(e) Sketch this surface M showing the two rulings through each of the points (1, 0, 0) and (2, 1, 2).

8. Let M be the surface of revolution gotten by revolving the curve $t \to (g(t), h(t), 0)$ about the x axis ($h > 0$). Show that:

(a) If g' is never zero, then M has a parametrization of the form

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

(b) If h' is never zero, then M has a parametrization of the form

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

A *quadric surface* is a surface M: $g = 0$ in $\mathbf{R}^3$ such that g contains at most quadratic terms in x_1, x_2, x_3, that is,

$$g = \sum_{i,j} a_{ij} x_i x_j + \sum_i b_i x_i + c.$$

Trivial cases excepted, every quadric surface is congruent to one of the five types described in the next two exercises. (Use of computers is optional in these exercises.)

9. In each case, (i) show that M is a surface, and sketch its general shape when $a = 3$, $b = 2$, $c = 1$; (ii) show that $\mathbf{x}$ is a parametrization in M and describe what part of M it covers.

(a) *Ellipsoid.* M: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$,

$\mathbf{x}(u, v) = (a\cos u \cos v, b\cos u \sin v, c \sin u)$ on D: $-\pi/2 < u < \pi/2$.

(b) *Hyperboloid of one sheet* (Fig. 4.21).

M: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$, $\mathbf{x}(u, v) = (a\cosh u \cos v, b\cosh u \sin v, c\sinh u)$ on $\mathbf{R}^2$.

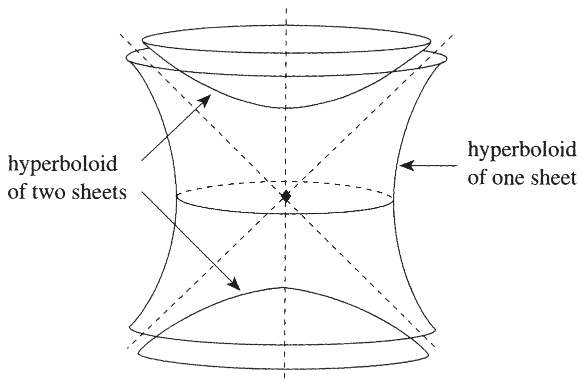


FIG. 4.21

(c) *Hyperboloid of two sheets* (Fig. 4.21).

M: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = -1$, $\mathbf{x}(u, v) = (a \sinh u \cos v,\ b \sinh u \sin v,\ c \cosh u)$ on D: $u \neq 0$.

10. Sketch the following surfaces (graphs of functions) for $a = 2$, $b = 1$:

(a) *Elliptic paraboloid.* M: $z = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$. Show that

$$\mathbf{x}(u, v) = (au \cos v, bu \sin v, u^2), \quad u > 0,$$

is a parametrization that omits only one point of M.

(b) *Hyperbolic paraboloid.* M: $z = \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2}$.

Show that M is covered by the single patch

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

11. *Doubly ruled quadrics.*

(a) Show that the hyperbolic paraboloid M in the preceding exercise is doubly ruled.

(b) (*Computer graphics.*) For $a = 2$, $b = 1$ use the patch in (b) of Exercise 10 to plot a portion of M. (Keep the same scale on all axes; the parameter curves will be the rulings.)

(c) Find two different ruled parametrizations of the hyperboloid of one sheet by using the scheme in the special case, Exercise 7.

(d) (*Computer graphics.*) Plot a portion of each of these parametrizations, taking $a = 1.5$, $b = 1$, $c = 2$.

4.3 Differentiable Functions and Tangent Vectors

We now begin an exposition of the calculus on a surface M in $\mathbf{R}^3$. The space $\mathbf{R}^3$ will gradually fade out of the picture, since our ultimate goal is a calculus for M alone. Generally speaking, we shall follow the order of topics in Chapter 1, making such changes as are necessary to adapt the calculus of the plane $\mathbf{R}^2$ to a surface M.

Suppose that f a is real-valued function defined on a surface M. If $\mathbf{x}$: $D \to M$ is a coordinate patch in M, then the composite function $f(\mathbf{x})$ is called a *coordinate expression* for f; it is just an ordinary real-valued function $(u, v) \to f(\mathbf{x}(u, v))$. We define f to be *differentiable* provided all its coordinate expressions are differentiable in the usual Euclidean sense (Definition 1.3 of Chapter 1).

For a function F: $\mathbf{R}^n \to M$, each patch $\mathbf{x}$ in M gives a *coordinate expression* $\mathbf{x}^{-1}(F)$ for F. Evidently this composite function is defined only on the set $\mathcal{O}$ of all points $\mathbf{p}$ of $\mathbf{R}^n$ such that $F(\mathbf{p})$ is in $\mathbf{x}(D)$. Again we define F to be *differentiable* provided all its coordinate expressions are differentiable in the usual Euclidean sense. We must understand that this includes the requirement that $\mathcal{O}$ be an open set of R^n, so that the differentiability of $\mathbf{x}^{-1}(F)$:$\mathcal{O} \to \mathbf{R}^2$ is well-defined, as in Section 7 of Chapter 1

In particular, a *curve* α: $I \to M$ in a surface M is, as before, a differentiable function from an open interval I into M.

To see how this definition works out in practice, we examine an important special case.

3.1 Lemma If α is a curve α: $I \to M$ whose route lies in the image $\mathbf{x}(D)$ of a single patch $\mathbf{x}$, then there exist unique differentiable functions a_1, a_2 on I such that

$$\alpha(t) = \mathbf{x}(a_1(t), a_2(t)) \quad \text{for all } t,$$

or in functional notation, $\alpha = \mathbf{x}(a_1, a_2)$. (See Fig. 4.22.)

Proof. By definition, the coordinate expression $\mathbf{x}^{-1}\,\alpha$: $I \to D$ is differentiable—it is just a curve in $\mathbf{R}^2$ whose route lies in the domain D of $\mathbf{x}$. If a_1, a_2 are the Euclidean coordinate functions of $\mathbf{x}^{-1}\ \alpha$, then

$$\alpha = \mathbf{x}\mathbf{x}^{-1}\alpha = \mathbf{x}(a_1, a_2).$$

These are the only such functions, for if $\alpha = \mathbf{x}(b_1, b_2)$, then

$$(a_1, a_2) = \mathbf{x}^{-1}\alpha = \mathbf{x}^{-1}\mathbf{x}(b_1, b_2) = (b_1, b_2). \qquad \blacklozenge$$

These functions a_1, a_2 are called the *coordinate functions* of the curve α with respect to the patch $\mathbf{x}$.

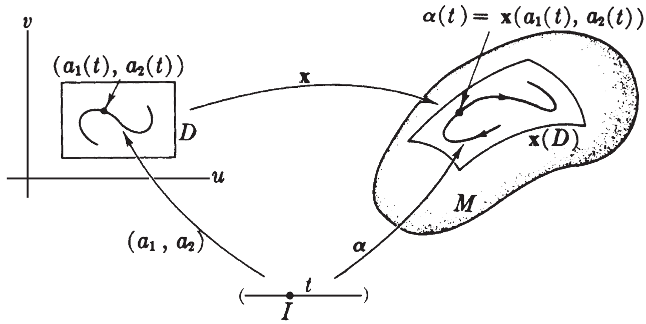


FIG. 4.22

For an arbitrary patch $\mathbf{x}: D \to M$, it is natural to think of the domain D as a *map* of the region $\mathbf{x}(D)$ in M. The functions $\mathbf{x}$ and $\mathbf{x}^{-1}$ establish a one-to-one correspondence between objects in $\mathbf{x}(D)$ and objects in D. If a curve α in $\mathbf{x}(D)$ represents the voyage of a ship, the coordinate curve (a_1, a_2) plots its position on the map D.

A rigorous proof of the following technical fact requires the methods of advanced calculus, and we shall not attempt to give a proof here.

3.2 Theorem Let M be a surface in $\mathbf{R}^3$. If $F: \mathbf{R}^n \to \mathbf{R}^3$ is a (differentiable) mapping whose image lies in M, then considered as a function F. $\mathbf{R}^n \to M$ into M, F is differentiable (as defined above).

This theorem links the calculus of M tightly to the calculus of $\mathbf{R}^3$. For example, it implies the "obvious" result that a curve in $\mathbf{R}^3$ that lies in M is a curve of M.

Since a patch is a differentiable function from an open set of $\mathbf{R}^2$ into $\mathbf{R}^3$, it follows that a patch is a differentiable function into M. Hence its coordinate expressions are all differentiable, so *patches overlap smoothly.*

3.3 Corollary If $\mathbf{x}$ and $\mathbf{y}$ are patches in a surface M in $\mathbf{R}^3$ whose images overlap, then the composite functions $\mathbf{x}^{-1}\mathbf{y}$ and $\mathbf{y}^{-1}\mathbf{x}$ are (differentiable) mappings defined on open sets of $\mathbf{R}^2$.

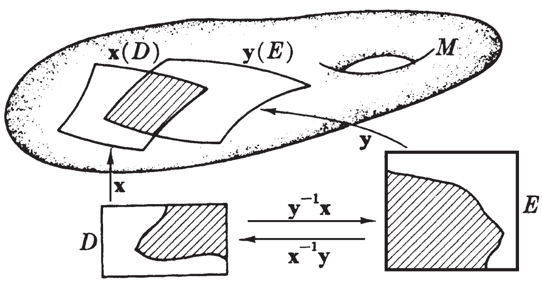


FIG. 4.23

The function $\mathbf{y}^{-1}\mathbf{x}$, for example, is defined only for those points (u, v) in D such that $\mathbf{x}(u, v)$ lies in the image $\mathbf{y}(E)$ of $\mathbf{y}$ (Fig. 4.23).

By an argument like that for Lemma 3.1, Corollary 3.3 can be rewritten:

3.4 Corollary If $\mathbf{x}$ and $\mathbf{y}$ are overlapping patches in M, then there exist unique differentiable functions $\bar{u}$ and $\bar{v}$ such that

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

for all (u, v) in the domain of $\mathbf{x}^{-1}\mathbf{y}$. In functional notation: $\mathbf{y} = \mathbf{x}(\bar{u}, \bar{v})$.

There are, of course, symmetrical equations expressing **x** in terms of **y**.

Corollary 3.3 makes it much easier to prove differentiability. For example, if f is a real-valued function on M, instead of verifying that *all* coordinate expressions $f(\mathbf{x})$ are Euclidean differentiable, we need only do so for enough patches **x** to cover all of M (so a single patch will often be enough). The proof is an exercise in checking domains of composite functions: For an *arbitrary* patch **y**, $f\mathbf{x}$ and $\mathbf{x}^{-1}\mathbf{y}$ differentiable imply $f\mathbf{x}\mathbf{x}^{-1}\mathbf{y}$ differentiable. This function is in general not $f\mathbf{y}$, because its domain is too small. But since there are enough **x**'s to cover M, such functions constitute all of $f\mathbf{y}$, and thus prove that it is differentiable.

It is intuitively clear what it means for a vector to be tangent to a surface M in $\mathbf{R}^3$. A formal definition can be based on the idea that a curve in M must have all its velocity vectors tangent to M.

3.5 Definition Let **p** be a point of a surface M in $\mathbf{R}^3$. A tangent vector **v** to $\mathbf{R}^3$ at **p** is *tangent to M at* **p** provided **v** is a velocity of some curve in M (Fig. 4.24).

The set of all tangent vectors to M at **p** is called the *tangent plane of M at* **p** and is denoted by $T_p(M)$. The following result shows, in particular, that at each point **p** of M the tangent plane $T_p(M)$ is actually a 2-dimensional vector subspace of the tangent space $T_p(\mathbf{R}^3)$.

3.6 Lemma Let **p** be a point of a surface M in $\mathbf{R}^3$, and let **x** be a patch in M such that $\mathbf{x}(u_0, v_0) = \mathbf{p}$. A tangent vector **v** to $\mathbf{R}^3$ at **p** is tangent to M if and only if **v** can be written as a linear combination of $\mathbf{x}_u(u_0, v_0)$ and $\mathbf{x}_v(u_0, v_0)$.

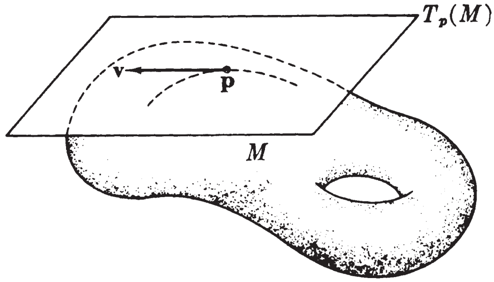


FIG. 4.24

Since partial velocities are always linearly independent, we deduce that they provided a *basis* for the tangent plane of M at each point of $\mathbf{x}(D)$.

Proof. Note that the parameter curves of $\mathbf{x}$ are curves in M, so $\mathbf{x}_u$ and $\mathbf{x}_v$ are always tangent to M at $\mathbf{p}$.

First suppose that $\mathbf{v}$ is tangent to M at $\mathbf{p}$; thus there is a curve α in M such that $\alpha(0) = \mathbf{p}$ and $\alpha'(0) = \mathbf{v}$. Now by Lemma 3.1, α may be written

$$\alpha = \mathbf{x}(a_1, a_2);$$

hence by the chain rule,

$$\alpha' = \mathbf{x}_u(a_1, a_2)\frac{da_1}{dt} + \mathbf{x}_v(a_1, a_2)\frac{da_2}{dt}.$$

But since $\alpha(0) = \mathbf{p} = \mathbf{x}(u_0, v_0)$, we have $a_1(0) = u_0$, $a_2(0) = v_0$. Hence evaluation at $t = 0$ yields

$$\mathbf{v} = \alpha'(0) = \frac{da_1}{dt}(0)\mathbf{x}_u(u_0, v_0) + \frac{da_2}{dt}(0)\mathbf{x}_v(u_0, v_0).$$

Conversely, suppose that a tangent vector $\mathbf{v}$ to $\mathbf{R}^3$ can be written

$$\mathbf{v} = c_1\mathbf{x}_u(u_0, v_0) + c_2\mathbf{x}_v(u_0, v_0).$$

By computations as above, $\mathbf{v}$ is the velocity vector at $t = 0$ of the curve

$$t \to \mathbf{x}(u_0 + tc_1, v_0 + tc_2).$$

Thus $\mathbf{v}$ is tangent to M at $\mathbf{p}$. ◆

A reasonable deduction, based on the general properties of derivatives, is that the tangent plane $T_p(M)$ is the linear approximation of the surface M near $\mathbf{p}$.

3.7 Definition A Euclidean vector field Z on a surface M in $\mathbf{R}^3$ is a function that assigns to each point $\mathbf{p}$ of M a tangent vector $Z(\mathbf{p})$ to $\mathbf{R}^3$ at $\mathbf{p}$.

A Euclidean vector field V for which each vector $V(\mathbf{p})$ is tangent to M at $\mathbf{p}$ is called a *tangent vector field* on M (Fig. 4.25). Frequently these vector fields are defined, not on all of M, but only on some region in M. As usual, we always assume differentiability.

A Euclidean vector $\mathbf{z}$ at a point $\mathbf{p}$ of M is *normal* to M if it is orthogonal to the tangent plane $T_p(M)$—that is, to every tangent vector to M at $\mathbf{p}$. And a Euclidean vector field Z on M is a *normal vector field* on M provided each vector $Z(\mathbf{p})$ is normal to M.

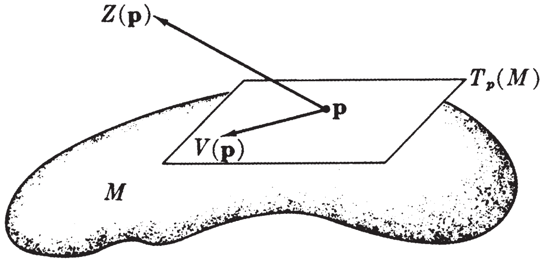


FIG. 4.25

Because $T_p(M)$ is a two-dimensional subspace of $T_p(\mathbf{R}^3)$, there is only one direction normal to M at $\mathbf{p}$: All normal vectors $\mathbf{z}$ at $\mathbf{p}$ are collinear.

Thus if $\mathbf{z}$ is not zero, it follows that $T_p(M)$ *consists of precisely those vectors in* $T_p(\mathbf{R}^3)$ *that are orthogonal* to $\mathbf{z}$.

It is particularly easy to deal with tangent and normal vector fields on a surface given in implicit form.

3.8 Lemma If M: $g = c$ is a surface in $\mathbf{R}^3$, then the *gradient* vector field $\nabla g = \sum \partial g/\partial x_i U_i$ (considered only at points of M) is a nonvanishing normal vector field on the entire surface M.

Proof. The gradient is nonvanishing (that is, never zero) on M since in the implicit case we require that the partial derivatives $\partial g/\partial x_i$ cannot simultaneously be zero at any point of M.

We must show that $(\nabla g)(\mathbf{p}) \bullet \mathbf{v} = 0$ for every tangent vector $\mathbf{v}$ to M at $\mathbf{p}$. First note that if α is a curve in M, then $g(\alpha) = g(\alpha_1, \alpha_2, \alpha_3)$ has constant value c. Thus by the chain rule,

$$\sum \frac{\partial g}{\partial x_i}(\alpha)\frac{d\alpha_i}{dt} = 0.$$

Now choose α to have initial velocity

$$\alpha'(0) = \mathbf{v} = (v_1, v_2, v_3)$$

at $\alpha(0) = \mathbf{p}$. Then

$$0 = \sum \frac{\partial g}{\partial x_i}(\alpha(0))\frac{d\alpha_i}{dt}(0) = \sum \frac{\partial g}{\partial x_i}(\mathbf{p})v_i = (\nabla g)(\mathbf{p}) \bullet \mathbf{v}.$$

◆

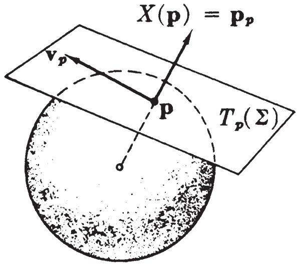


FIG. 4.26

3.9 Example Vector fields on the sphere Σ: $g = \sum x_i^2 = r^2$. The lemma shows that

$$X = \frac{1}{2}\nabla g = \sum x_i U_i$$

is a normal vector field on Σ (Fig. 4.26). This is geometrically evident, since $X(\mathbf{p}) = \sum p_i U_i(\mathbf{p})$ is the vector $\mathbf{p}$ with point of application $\mathbf{p}$! It follows by a remark above that $\mathbf{v}_p$ is tangent to Σ if and only if the dot product $\mathbf{v}_p \bullet \mathbf{p}_p = \mathbf{v} \bullet \mathbf{p}$ is zero. Similarly, a vector field V on Σ is a *tangent* vector field if and only if $V \bullet X = 0$. For example, $V(\mathbf{p}) = (-p_2, p_1, 0)$ defines a tangent vector field on Σ that points "due east" and vanishes at the north and south poles $(0, 0, \pm r)$.

We must emphsasize that only the *tangent* vector fields on M belong to the calculus of M itself, since they derive ultimately from curves in M (Definition 3.5). This is certainly not the case with *normal* vector fields. However, as we shall see in the next chapter, normal vector fields are quite useful in studying M from the viewpoint of an observer in $\mathbf{R}^3$.

Finally, we shall adapt the notion of directional derivatives to a surface. Definition 3.1 of Chapter 1 uses straight lines in $\mathbf{R}^3$; thus we must use the less restrictive formulation based on Lemma 4.6 of Chapter 1.

3.10 Definition Let $\mathbf{v}$ be a tangent vector to M at $\mathbf{p}$, and let f be a differentiable real-valued function on M. The *derivative* $\mathbf{v}[f]$ *of f with respect to* $\mathbf{v}$ is the common value of $(d/dt)(f\alpha)(0)$ for all curves α in M with initial velocity $\mathbf{v}$.

Directional derivatives on a surface have exactly the same linear and Leibnizian properties as in the Euclidean case (Theorem 3.3 of Chapter 1).

Exercises

1. Let $\mathbf{x}$ be the geographical patch in the sphere Σ (Ex. 2.2). Find the coordinate expression $f(\mathbf{x})$ for the following functions on Σ:
(a) $f(\mathbf{p}) = p_1^2 + p_2^2$. (b) $f(\mathbf{p}) = (p_1 - p_2)^2 + p_3^2$.

2. Let $\mathbf{x}$ be the usual parametrization of the torus (Ex. 2.5).
(a) Find the Euclidean coordinates $\alpha_1, \alpha_2, \alpha_3$ of the curve $\alpha(t) = \mathbf{x}(t, t)$.
(b) Show that α is periodic, and find its period $p > 0$, the smallest number such that $\alpha(t + p) = \alpha(t)$ for all t.

3. (a) Prove Corollary 3.4.
(b) Derive the chain rule

$$\mathbf{y}_u = \frac{\partial \bar{u}}{\partial u}\mathbf{x}_u + \frac{\partial \bar{v}}{\partial u}\mathbf{x}_v, \quad \mathbf{y}_v = \frac{\partial \bar{u}}{\partial v}\mathbf{x}_u + \frac{\partial \bar{v}}{\partial v}\mathbf{x}_v,$$

where $\mathbf{x}_u$ and $\mathbf{x}_v$ are evaluated on $(\bar{u}, \bar{v})$.
(c) Deduce that $\mathbf{y}_u \times \mathbf{y}_v = J\mathbf{x}_u \times \mathbf{x}_v$, where J is the Jacobian of the mapping $\mathbf{x}^{-1}\mathbf{y} = (\bar{u}, \bar{v})\colon D \to \mathbf{R}^2$.

4. Let $\mathbf{x}$ be a patch in M.
(a) If $\mathbf{x}_*$ is the tangent map of $\mathbf{x}$ (Sec. 7 of Ch. 1), show that

$$\mathbf{x}_*(U_1) = \mathbf{x}_u, \quad \mathbf{x}_*(U_2) = \mathbf{x}_v,$$

where U_1, U_2 is the natural frame field on $\mathbf{R}^2$.
(b) If f is a differentiable function on M, prove

$$\mathbf{x}_u[f] = \frac{\partial}{\partial u}(f(\mathbf{x})), \quad \mathbf{x}_v[f] = \frac{\partial}{\partial v}(f(\mathbf{x})).$$

5. Prove that:
(a) $\mathbf{v} = (v_1, v_2, v_3)$ is tangent to M: $z = f(x, y)$ at a point $\mathbf{p}$ of M if and only if

$$v_3 = \frac{\partial f}{\partial x}(p_1, p_2)v_1 + \frac{\partial f}{\partial y}(p_1, p_2)v_2.$$

(b) if $\mathbf{x}$ is a patch in an arbitrary surface M, then $\mathbf{v}$ is tangent to M at $\mathbf{x}(u, v)$ if and only if

$$\mathbf{v} \bullet \mathbf{x}_u(u, v) \times \mathbf{x}_v(u, v) = 0.$$

6. Let $\mathbf{x}$ and $\mathbf{y}$ be the patches in the unit sphere Σ that are defined on the unit disk D: $u^2 + v^2 < 1$ by

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

where $f = \sqrt{1 - u^2 - v^2}$.

(a) On a sketch of Σ indicate the images $\mathbf{x}(D)$ and $\mathbf{y}(D)$, and the region on which they overlap.

(b) At which points of D is $\mathbf{y}^{-1}\mathbf{x}$ defined? Find a formula for this function.

(c) At which points of D is $\mathbf{x}^{-1}\mathbf{y}$ defined? Find a formula for this function.

7. Find a nonvanishing normal vector field on M: $z = xy$ and two tangent vector fields that are linearly independent at each point.

8. Let C be the circular cone parametrized by

$$\mathbf{x}(u, v) = v(\cos u, \sin u, 1).$$

If α is the curve $\mathbf{x}\left(\sqrt{2}t, e^t\right)$:

(a) Express α' in terms of $\mathbf{x}_u$ and $\mathbf{x}_v$.

(b) Show that at each point of α, the velocity α' bisects the angle between $\mathbf{x}_u$ and $\mathbf{x}_v$. (*Hint:* Verify that

$$\alpha' \bullet \left(\frac{\mathbf{x}_u}{\|\mathbf{x}_u\|}\right) = \alpha' \bullet \left(\frac{\mathbf{x}_v}{\|\mathbf{x}_v\|}\right),$$

where $\mathbf{x}_u$ and $\mathbf{x}_v$ are evaluated on $\left(\sqrt{2}t, e^t\right)$.)

(c) Make a sketch of the cone C showing the curve α.

9. If $\mathbf{z}$ is a nonzero vector normal to M at $\mathbf{p}$, let $\overline{T}_p(M)$ be the Euclidean plane through $\mathbf{p}$ orthogonal to $\mathbf{z}$. Prove:

(a) If each tangent vector $\mathbf{v}_p$ to M at $\mathbf{p}$ is replaced by its *tip* $\mathbf{p} + \mathbf{v}$, then $T_p(M)$ becomes $\overline{T}_p(M)$. Thus $\overline{T}_p(M)$ gives a concrete representation of $T_p(M)$ in $\mathbf{R}^3$. It is called the *Euclidean tangent plane to M* at *p*.

(b) If $\mathbf{x}$ is a patch in M, then $\overline{T}_{x(u,v)}(M)$ consists of all points $\mathbf{r}$ in $\mathbf{R}^3$ such that $(\mathbf{r} - \mathbf{x}(u, v)) \bullet \mathbf{x}_u(u, v) \times \mathbf{x}_v(u, v) = 0$.

(c) If M is given implicitly by $g = c$, then $\overline{T}_P(M)$ consists of all points $\mathbf{r}$ in $\mathbf{R}^3$ such that $(\mathbf{r} - \mathbf{p}) \bullet (\nabla g)(\mathbf{p}) = 0$.

10. In each case below find an equation of the form $ax + by + cz = d$ for the plane $\overline{T}_p(M)$.

(a) $\mathbf{p} = (0, 0, 0)$ and M is the sphere

$$x^2 + y^2 + (z - 1)^2 = 1.$$

(b) $\mathbf{p} = (1, -2, 3)$ and M is the ellipsoid

$$\frac{x^2}{4} + \frac{y^2}{16} + \frac{z^2}{18} = 1.$$

(c) $\mathbf{p} = \mathbf{x}(2, \pi/4)$, where M is the helicoid parametrized by

$$\mathbf{x}(u, v) = (u \cos v, u \sin v, 2v).$$

11. (*Continuation of Ex.* 2.) With $\mathbf{x}$ the usual parametrization of the torus of revolution T, consider the curve α: $\mathbf{R} \to T$ such that $\alpha(t) = \mathbf{x}(at, bt)$.

(a) If a/b is a rational number, show that α is a simple closed curve in T, that is, periodic with no self-crossings.

(b) If a/b is irrational, show α is one-to-one. Such a curve is called a *winding line* on the torus. It is *dense* in T in the sense that given any $\varepsilon > 0$, α comes within distance ε of every point of T.

(c) (*Computer graphics.*) For reference, plot the torus T with $R = 3$, $r = 1$ (see Ex. 2.5). Then plot the following curves in T:

(i) $\alpha(t) = \mathbf{x}(3t, 5t)$ on intervals $0 \leqq t \leqq b$, for $b = \pi, 2\pi$, and larger values. Estimate the period of α, in this case the smallest number $T > 0$ such that $\alpha(T) = \alpha(0)$.

(ii) $\alpha(t) = \mathbf{x}(\pi t, 5t)$ on intervals $0 \leqq t \leqq b$, for increasing values of b. (Keep the curve reasonably smooth.)

12. A Euclidean vector field $Z = \sum z_i U_i$ on M is *differentiable* provided its coordinate functions z_1, z_2, z_3 (on M) are differentiable. If V is a tangent vector field on M, show that

(a) For every patch $\mathbf{x}$: $D \to M$, V can be written as

$$V(\mathbf{x}(u, v)) = f(u, v)\mathbf{x}_u(u, v) + g(u, v)\mathbf{x}_v(u, v).$$

(b) V is differentiable if and only if the functions f and g (on D) are differentiable.

The following exercises deal with *open sets* in a surface M in $\mathbf{R}^3$, that is, sets $\mathscr{U}$ in M that contain a neighborhood in M of each of their points.

13. Prove that if $\mathbf{y}$: $E \to M$ is a proper patch, then $\mathbf{y}$ carries open sets in E to open sets in M. Deduce that if $\mathbf{x}$: $D \to M$ is an arbitrary patch, then the image $\mathbf{x}(D)$ is an open set in M. (*Hint:* To prove the latter assertion, use Cor. 3.3.)

14. Prove that *every* patch $\mathbf{x}$: $D \to M$ in a surface M in $\mathbf{R}^3$ is proper. (*Hint:* Use Ex. 13. Note that $(\mathbf{x}^{-1}\mathbf{y})\mathbf{y}^{-1}$ is continuous and agrees with $\mathbf{x}^{-1}$ on an open set in $\mathbf{x}(D)$.)

15. If $\mathscr{U}$ is a subset of a surface M in $\mathbf{R}^3$, prove that $\mathscr{U}$ is itself a surface in $\mathbf{R}^3$ if and only if $\mathscr{U}$ is an open set of M.

4.4 Differential Forms on a Surface

In Chapter 1 we discussed differential forms on $\mathbf{R}^3$ only in sufficient detail to take care of the Cartan structural equations (Theorem 8.3 of Chapter 2). In the next three sections we shall give a rather complete treatment of forms *on a surface.*

Forms are just what we will need to describe the geometry of a surface, but this is only one example of their usefulness. Surfaces and Euclidean spaces are merely special cases of the general notion of *manifold* (Section 8). Every manifold has a differential and integral calculus—expressed in terms of functions, vector fields, and forms—that generalizes the usual elementary calculus on the real line. Thus forms are fundamental to all the many branches of mathematics and its applications that are based on calculus. In the special case of a surface, the calculus of forms is rather easy, but it still gives a remarkably accurate picture of the most general case.

Just as for $\mathbf{R}^3$, a 0-*form* f on a surface M is simply a (differentiable) real-valued function on M, and a 1-*form* ϕ on M is a real-valued function on tangent vectors to M that is linear at each point (Definition 5.1 of Chapter 1). We did not give a precise definition of 2-forms in Chapter 1, but we shall do so now. A 2-form will be a two-dimensional analogue of a 1-form: a real-valued function, not on single tangent vectors, but on *pairs* of tangent vectors. (In this context the term "pair" will always imply that the tangent vectors have the same point of application.)

4.1 Definition A 2-*form* η on a surface M is a real-valued function on all ordered pairs of tangent vectors $\mathbf{v}$, $\mathbf{w}$ to M such that

(1) $\eta(\mathbf{v}, \mathbf{w})$ is linear in $\mathbf{v}$ and in $\mathbf{w}$;

(2) $\eta(\mathbf{v}, \mathbf{w}) = -\eta(\mathbf{w}, \mathbf{v})$.

Since a surface is two-dimensional, *all p-forms with $p > 2$ are zero*, by definition. This fact considerably simplifies the theory of differential forms on a surface.

At the end of this section we will show that our new definitions are consistent with the informal exposition given in Chapter 1, Section 6.

Forms are added in the usual pointwise fashion; we add only forms of the same *degree* $p = 0, 1, 2$. Just as a 1-form ϕ is evaluated on a vector field V, now a 2-form η is evaluated on a pair of vector fields V, W to give a real-valued function $\eta(V, W)$ on the surface M. Of course, we shall always assume that the forms we deal with are differentiable—that is, convert differentiable vector fields into differentiable functions.

Note that the alternation rule (2) in Definition 4.1 implies that

$$\eta(\mathbf{v}, \mathbf{v}) = 0$$

for any tangent vector $\mathbf{v}$. This rule also shows that 2-forms are related to determinants.

4.2 Lemma Let η be a 2-form on a surface M, and let $\mathbf{v}$ and $\mathbf{w}$ be (linearly independent) tangent vectors at some point of M. Then

$$\eta(a\mathbf{v} + b\mathbf{w}, c\mathbf{v} + d\mathbf{w}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \eta(\mathbf{v}, \mathbf{w}).$$

Proof. Since η is linear in its first variable, its value on the pair of tangent vectors $a\mathbf{v} + b\mathbf{w}$, $c\mathbf{v} + d\mathbf{w}$ is $a\eta(\mathbf{v}, c\mathbf{v} + d\mathbf{w}) + b\eta(\mathbf{w}, c\mathbf{v} + d\mathbf{w})$. Using the linearity of η in its second variable, we get

$$ac\,\eta(\mathbf{v}, \mathbf{v}) + ad\,\eta(\mathbf{v}, \mathbf{w}) + bc\,\eta(\mathbf{w}, \mathbf{v}) + bd\,\eta(\mathbf{w}, \mathbf{w}).$$

Then the alternation rule (2) gives

$$\eta(a\mathbf{v} + b\mathbf{w}, c\mathbf{v} + d\mathbf{w}) = (ad - bc)\,\eta(\mathbf{v}, \mathbf{w}).$$ ◆

Thus the values of a 2-form on *all* pairs of tangent vectors at a point are completely determined by its value on any *one* linearly independent pair. This remark is used frequently in later work.

Wherever they appear, differential forms satisfy certain general properties, established (at least partially) in Chapter 1 for forms on $\mathbf{R}^3$. To begin with, *the wedge product of a p-form and a q-form is always a* $(p + q)$*-form.* If p or q is zero, the wedge product is just the usual multiplication by a function. On a surface, the wedge product is always zero if $p + q > 2$. So we need a definition only for the case $p = q = 1$.

4.3 Definition If ϕ and ψ are 1-forms on a surface M, the *wedge product* $\phi \wedge \psi$ is the 2-form on M such that

$$(\phi \wedge \psi)(\mathbf{v}, \mathbf{w}) = \phi(\mathbf{v})\psi(\mathbf{w}) - \phi(\mathbf{w})\psi(\mathbf{v})$$

for all pairs $\mathbf{v}$, $\mathbf{w}$ of tangent vectors to M.

Note that $\phi \wedge \psi$ really is a 2-form on M, since it is a real-valued function on all pairs of tangent vectors and satisfies the conditions in Definition 4.1. The wedge product has all the usual algebraic properties except commutativity; in general, *if ξ is a p-form and η is a q-form, then*

$$\xi \wedge \eta = (-1)^{pq}\eta \wedge \xi.$$

On a surface the only minus sign occurs in the multiplication of 1-forms, where just as in Chapter 1, we have $\phi \wedge \psi = -\psi \wedge \phi$, and hence $\phi \wedge \phi = 0$.

The differential calculus of forms is based on the exterior derivative d. For a 0-form (function) f on a surface, the exterior derivative is, as before, the

1-form df such that $df(\mathbf{v}) = \mathbf{v}[f]$. Wherever forms appear, the exterior derivative of a p-form is a $(p + 1)$-form. Thus, for surfaces the only new definition we need is that of the exterior derivative $d\phi$ of a 1-form ϕ.

4.4 Definition Let ϕ be a 1-form on a surface M. Then the *exterior derivative* $d\phi$ of ϕ is the 2-form such that for any patch $\mathbf{x}$ in M,

$$d\phi(\mathbf{x}_u, \mathbf{x}_v) = \frac{\partial}{\partial u}(\phi(\mathbf{x}_v)) - \frac{\partial}{\partial v}(\phi(\mathbf{x}_u)).$$

As it stands, this is not yet a valid definition; there is a problem of consistency. What we have actually defined is a form $d_{\mathbf{x}}\phi$ on the image of each patch $\mathbf{x}$ in M. So what we must prove is that on the region where two patches overlap, the forms $d_{\mathbf{x}}\phi$ and $d_{\mathbf{y}}\phi$ are equal. Only then will we have obtained from ϕ a single form $d\phi$ on M.

4.5 Lemma Let ϕ be a 1-form on M. If $\mathbf{x}$ and $\mathbf{y}$ are patches in M, then $d_{\mathbf{x}}\phi = d_{\mathbf{y}}\phi$ on the overlap of $\mathbf{x}(D)$ and $\mathbf{y}(E)$.

Proof. Because $\mathbf{y}_u$ and $\mathbf{y}_v$ are linearly independent at each point, it suffices by Lemma 4.2 to show that

$$(d_{\mathbf{y}}\phi)(\mathbf{y}_u, \mathbf{y}_v) = (d_{\mathbf{x}}\phi)(\mathbf{y}_u, \mathbf{y}_v).$$

Now, as in Corollary 3.4, we write $\mathbf{y} = \mathbf{x}(\bar{u}, \bar{v})$ and deduce by the chain rule that

$$\mathbf{y}_u = \frac{\partial \bar{u}}{\partial u}\mathbf{x}_u + \frac{\partial \bar{v}}{\partial u}\mathbf{x}_v,$$

$$\mathbf{y}_v = \frac{\partial \bar{u}}{\partial v}\mathbf{x}_u + \frac{\partial \bar{v}}{\partial v}\mathbf{x}_v, \tag{1}$$

where $\mathbf{x}_u$ and $\mathbf{x}_v$ are henceforth *evaluated on* $(\bar{u}, \bar{v})$. Then by Lemma 4.2,

$$(d_{\mathbf{x}}\phi)(\mathbf{y}_u, \mathbf{y}_v) = J(d_{\mathbf{x}}\phi)(\mathbf{x}_u, \mathbf{x}_v), \tag{2}$$

where J is the Jacobian $(\partial\bar{u}/\partial u)\,(\partial\bar{v}/\partial v) - (\partial\bar{u}/\partial v)\,(\partial\bar{v}/\partial u)$. Thus it is clear from Definition 4.4 that to prove $(d_y\phi)\,(\mathbf{y}_u, \mathbf{y}_v) = (d_{\mathbf{x}}\phi)\,(\mathbf{y}_u, \mathbf{y}_v)$, all we need is the equation.

$$\frac{\partial}{\partial u}(\phi(\mathbf{y}_v)) - \frac{\partial}{\partial v}(\phi(\mathbf{y}_v)) = J\left\{\frac{\partial}{\partial \bar{u}}(\phi(\mathbf{x}_v)) - \frac{\partial}{\partial \bar{v}}(\phi(\mathbf{x}_v))\right\}. \tag{3}$$

It suffices to operate on $(\partial/\partial u)(\phi(\mathbf{y}_v))$, for merely reversing u and v will then yield $(\partial/\partial v)(\phi(\mathbf{y}_u))$. Since (3) requires us to *subtract* these two derivatives, we can *discard any terms that will cancel* when u and v are everywhere reversed.

Applying ϕ to the second equation in (1) yields

$$\phi(\mathbf{y}_v) = \phi(\mathbf{x}_u)\frac{\partial \bar{u}}{\partial v} + \phi(\mathbf{x}_v)\frac{\partial \bar{v}}{\partial v}.$$

Hence by the chain rule,

$$\frac{\partial}{\partial u}(\phi(\mathbf{y}_v)) = \frac{\partial}{\partial u}(\phi(\mathbf{x}_u))\frac{\partial \bar{u}}{\partial v} + \frac{\partial}{\partial u}(\phi(\mathbf{x}_v))\frac{\partial \bar{v}}{\partial v} + \cdots, \tag{4}$$

where in accordance with the remark above we have discarded two symmetric terms. Next we use the chain rule—and the same remark—to get

$$\frac{\partial}{\partial u}(\phi(\mathbf{y}_v)) = \left(\frac{\partial}{\partial \bar{v}}(\phi(\mathbf{x}_u))\frac{\partial \bar{v}}{\partial u} + \cdots\right)\frac{\partial \bar{u}}{\partial v} + \left(\frac{\partial}{\partial \bar{u}}(\phi(\mathbf{x}_v))\frac{\partial \bar{u}}{\partial u} + \cdots\right)\frac{\partial \bar{v}}{\partial v}. \tag{5}$$

Now reverse u and v in (5) (and also $\bar{u}$ and $\bar{v}$) and subtract from (5) itself. The result is precisely equation (3). ◆

It is difficult to exaggerate the importance of the exterior derivative. We have already seen in Chapter 1 that it generalizes the familiar notion of differential of a function, and that it contains the three fundamental derivative operations of classical vector analysis (Exercise 8 in Section 1.6). In Chapter 2 it is essential to the Cartan structural equations (Theorem 8.3). Perhaps the clearest statement of its meaning will come in Stokes's theorem (6.5), which could actually be used to define the exterior derivative of a 1-form.

On a surface, the exterior derivative of a wedge product displays the same linear and Leibnizian properties (Theorem 6.4 of Chapter 2) as in $\mathbf{R}^3$; see Exercise 3. For practical computations these properties are apt to be more efficient than a direct appeal to the definition. Examples of this technique appear in subsequent exercises.

The most striking property of this notion of derivative is that there are no *second* exterior derivatives: Wherever forms appear, *the exterior derivative applied twice always gives zero.* For a surface, we need only prove this for 0-forms, since even for a 1-form ϕ, the second derivative $d(d\phi)$ is a 3-form, and hence is automatically zero.

4.6 Theorem If f is a real-valued function on M, then $d(df) = 0$.

Proof. Let $\psi = df$, so we must show $d\psi = 0$. It suffices by Lemma 4.2 to prove that for any patch $\mathbf{x}$ in M we have $(d\psi)(\mathbf{x}_u, \mathbf{x}_v) = 0$. Now using Exercise 4 of Section 3, we get

$$\psi(\mathbf{x}_u) = df(\mathbf{x}_u) = \mathbf{x}_u[f] = \frac{\partial}{\partial u}(f(\mathbf{x}))$$

and similarly

$$\psi(\mathbf{x}_v) = \frac{\partial}{\partial v}(f(\mathbf{x})).$$

Hence

$$d\psi(\mathbf{x}_u, \mathbf{x}_v) = \frac{\partial}{\partial u}(\psi(\mathbf{x}_v)) - \frac{\partial}{\partial v}(\psi(\mathbf{x}_u)) = \frac{\partial^2(f(\mathbf{x}))}{\partial u \partial v} - \frac{\partial^2(f(\mathbf{x}))}{\partial v \partial u} = 0. \quad ◆$$

Many computations and proofs reduce to the problem of showing that two forms are equal. As we have seen, to do so it is not necessary to check that the forms have the same value on *all* tangent vectors. In particular, if $\mathbf{x}$ is a coordinate patch, then

(1) for 1-forms on $\mathbf{x}(D)$: $\phi = \psi$ if and only if $\phi(\mathbf{x}_u) = \psi(\mathbf{x}_u)$ and $\phi(\mathbf{x}_v) = \psi(\mathbf{x}_v)$;
(2) for 2-forms on $\mathbf{x}(D)$: $\mu = \nu$ if and only if $\mu(\mathbf{x}_u, \mathbf{x}_v) = \nu(\mathbf{x}_u, \mathbf{x}_v)$.

(To prove these criteria, we express arbitrary tangent vectors as linear combinations of $\mathbf{x}_u$ and $\mathbf{x}_v$.) More generally, $\mathbf{x}_u$ and $\mathbf{x}_v$ may be replaced by any two vector fields that are linearly independent at each point.

Let us now check that the rigorous results proved in this section are consistent with the rules of operation stated in Chapter 1, Section 6.

4.7 Example Differential forms on the plane $\mathbf{R}^2$. Let $u_1 = u$ and $u_2 = v$ be the natural coordinate functions, and U_1, U_2 the natural frame field on $\mathbf{R}^2$. The differential calculus of forms on $\mathbf{R}^2$ is expressed in terms of u_1 and u_2 as follows:

If f is a function, ϕ a 1-form, and η a 2-form, then

(1) $\phi = f_1 du_1 + f_2 du_2$, where $f_i = \phi(U_i)$.
(2) $\eta = g\, du_1 du_2$, where $g = \eta(U_1, U_2)$.
(3) for $\psi = g_1 du_1 + g_2 du_2$ and ϕ as above,

$$\phi \wedge \psi = (f_1 g_2 - f_2 g_1) du_1\, du_2.$$

(4) $df = \dfrac{\partial f}{\partial u_1} du_1 + \dfrac{\partial f}{\partial u_2} du_2.$

(5) $d\phi = \left(\dfrac{\partial f_2}{\partial u_1} - \dfrac{\partial f_1}{\partial u_2}\right) du_1\, du_2$ (ϕ as above).

For the proof of these formulas, see Exercise 4.

Similar definitions and coordinate expressions may be established on any Euclidean space. In the case of the real line $\mathbf{R}^1$, the natural frame field reduces to the single vector field U_1 for which $U_1 = [f] = df/dt$. All p-forms with $p > 0$ are zero, and for a 1-form, $\phi = \phi(U_1)\, dt$.

Wherever differential forms are used, the following conditions are fundamental.

4.8 Definition A differential form ϕ is *closed* if its exterior derivative is zero, $d\phi = 0$; and ϕ is *exact* if it is the exterior derivative of some form, $\phi = d\xi$.

Since d applied twice is always 0, *every exact form is closed*. In the case of a surface, since d increases degrees by 1, every 2-form is closed and no 0-form (i.e., function) is exact. Thus 1-forms are the important case, and for a 1-form ϕ exactness always means that there is a *function* f such that $df = \phi$. The analytical and topological consequences of these definitions run deep.

Exercises

1. Prove the Leibnizian formulas

$$d(fg) = g\, df + f\, dg, \quad d(f\phi) = df \wedge \phi + f\, d\phi,$$

where f and g are functions on M and ϕ is a 1-form.
(*Hint:* By definition, $(f\phi)(\mathbf{v}_p) = f(\mathbf{p})\phi(\mathbf{v}_p)$; hence $f\phi$ evaluated on $\mathbf{x}_u$ is $f(\mathbf{x})\phi(\mathbf{x}_u)$.)

2. (a) Prove formulas (1) and (2) in Example 4.7 using the remark preceding Example 4.7. (*Hint:* Show $(du_1 du_2)\,(U_1, U_2) = 1$.)
(b) Derive the remaining formulas using the properties of d and the wedge product.

3. If f is a real-valued function on a surface, and g is a function on the real line, show that

$$\mathbf{v}_p[g(f)] = g'(f)\mathbf{v}_p[f].$$

Deduce that

$$d(g(f)) = g'(f)df.$$

4. If f, g, and h are functions on a surface M, and ϕ is a 1-form, prove:
(a) $d(fgh) = ghdf + fhdg + fgdh$,
(b) $d(\phi f) = fd\phi - \phi \wedge df, \quad (\phi f = f\phi)$,
(c) $(df \wedge dg)\,(\mathbf{v}, \mathbf{w}) = \mathbf{v}[f]\mathbf{w}[g] - \mathbf{v}[g]\mathbf{w}[f]$.

5. Suppose that M is covered by open sets $\mathscr{U}_1, \ldots, \mathscr{U}_k$, and on each $\mathscr{U}_i$ there is defined a function f_i such that $f_i - f_j$ is constant on the overlap of $\mathscr{U}_i$ and $\mathscr{U}_j$. Show that there is a 1-form ϕ on M such that $\phi = df_i$ on each $\mathscr{U}_i$. Generalize to the case of 1-forms ϕ_i such that $\phi_i - \phi_j$ is closed.

6. Let $\mathbf{y}$: $E \to M$ be an arbitrary mapping of an open set of $\mathbf{R}^2$ into a surface M. If ϕ is a 1-form on M, show that the formula

$$d\phi(\mathbf{y}_u, \mathbf{y}_v) = \frac{\partial}{\partial u}(\phi(\mathbf{y}_v)) - \frac{\partial}{\partial v}(\phi(\mathbf{y}_u))$$

is still valid even when $\mathbf{y}$ is not regular or one-to-one.
(*Hint:* In the proof of Lem. 4.5, check that equation (3) is still valid in this case.)

A patch $\mathbf{x}$ in M establishes a one-to-one correspondence between an open set D of $\mathbf{R}^2$ and an open set $\mathbf{x}(D)$ of M. Although we have emphasized the function $\mathbf{x}$: $D \to \mathbf{x}(D)$, there are some advantages to emphasizing instead the inverse function $\mathbf{x}^{-1}$: $\mathbf{x}(D) \to D$.

7. If $\mathbf{x}$: $D \to M$ is a patch in M, let $\tilde{u}$ and $\tilde{v}$ be the coordinate functions of $\mathbf{x}^{-1}$, so $\mathbf{x}^{-1}(\mathbf{p}) = (\tilde{u}(\mathbf{p}), \tilde{v}(\mathbf{p}))$ for all $\mathbf{p}$ in $\mathbf{x}(D)$. Show that
(a) $\tilde{u}$ and $\tilde{v}$ are differentiable functions on $\mathbf{x}(D)$ such that:

$$\tilde{u}(\mathbf{x}(u, v)) = u, \quad \tilde{v}(\mathbf{x}(u, v)) = v.$$

These functions constitute the *coordinate system* associated with $\mathbf{x}$.

(b) $d\tilde{u}(\mathbf{x}_u) = 1, \quad d\tilde{u}(\mathbf{x}_v) = 0,$

$d\tilde{v}(\mathbf{x}_u) = 0, \quad d\tilde{v}(\mathbf{x}_v) = 1.$

(c) If ϕ is a 1-form and η is a 2-form, then

$$\phi = f\, d\tilde{u} + g\, d\tilde{v}, \quad \text{where } f(\mathbf{x}) = \phi(\mathbf{x}_u),\ g(\mathbf{x}) = \phi(\mathbf{x}_v);$$
$$\eta = h\, d\tilde{u}\, d\tilde{v}, \quad \text{where } h(\mathbf{x}) = \eta(\mathbf{x}_u, \mathbf{x}_v).$$

(*Hint:* for (b) use Ex. 4(b) of Sec. 3.)

8. Identify (or describe) the associated coordinate system $\tilde{u}$, $\tilde{v}$ of
(a) The polar coordinate patch $\mathbf{x}(u, v) = (u\cos v, u\sin v)$ defined on the domain D: $u > 0$, $0 < v < 2\pi$.
(b) The identity patch $\mathbf{x}(u, v) = (u, v)$ in $\mathbf{R}^2$.
(c) The geographical patch $\mathbf{x}$ in the sphere.

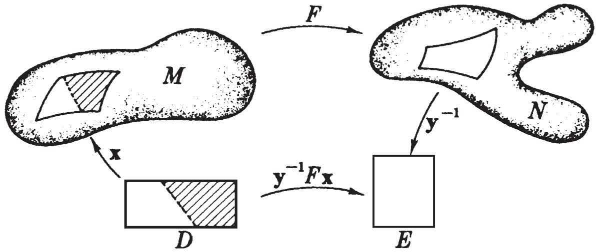


FIG. 4.27

4.5 Mappings of Surfaces

To define differentiability of a function *from* a surface *to* a surface, we follow the same general scheme used in Section 3 and require that all its coordinate expressions be differentiable.

5.1 Definition A function F: $M \to N$ from one surface to another is *differentiable* provided that for each patch $\mathbf{x}$ in M and $\mathbf{y}$ in N the composite function $\mathbf{y}^{-1}F\mathbf{x}$ is Euclidean differentiable (and defined on an open set of $\mathbf{R}^2$). F is then called a *mapping of surfaces*.

Evidently the function $\mathbf{y}^{-1}F\mathbf{x}$ is defined at all points (u, v) of D such that $F(\mathbf{x}(u, v))$ lies in the image of $\mathbf{y}$ (Fig. 4.27). As in Section 3 we deduce from Corollary 3.3 that in applying this definition, it suffices to check enough patches to cover both M and N.

5.2 Example (1) Let Σ be the unit sphere in $\mathbf{R}^3$ (center at $\mathbf{0}$) but with *north and south poles removed*, and let C be the cylinder based on the unit circle in the xy plane. So C is in contact with the sphere along the equator. We define a mapping F: $\Sigma \to C$ as follows: If $\mathbf{p}$ is a point of Σ, draw the line orthogonally out from the z azis through $\mathbf{p}$, and let $F(\mathbf{p})$ be the point at which this line first meets C, as in Fig. 4.28. To prove that F is a mapping, we use the geographical patch $\mathbf{x}$ in Σ (Example 2.2), and for C the patch $\mathbf{y}(u, v) = (\cos u, \sin u, v)$. Now $\mathbf{x}(u, v) = (\cos v \cos u, \cos v \sin u, \sin v)$, so from the definition of F we get

$$F(\mathbf{x}(u, v)) = (\cos u, \sin u, \sin v).$$

But this point of C is $\mathbf{y}(u, \sin v)$; hence

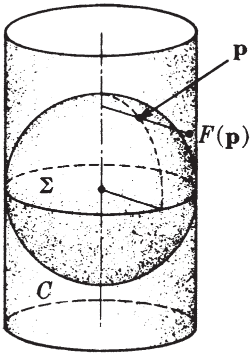


FIG. 4.28

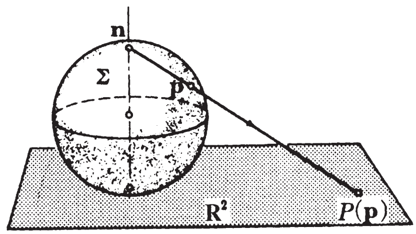


FIG. 4.29

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

Applying $\mathbf{y}^{-1}$ to both sides of this equation gives

$$(\mathbf{y}^{-1}F\mathbf{x})(u, v) = (u, \sin v),$$

so $\mathbf{y}^{-1}F\mathbf{x}$ is certainly differentiable. (Actually, **x** does not entirely cover Σ, but the missing semicircle can be covered by a patch like **x**.) We conclude that F is a mapping.

(2) Stereographic projection of the punctured sphere Σ onto the plane. Let Σ be a unit sphere resting on the xy plane at the origin, so the center of Σ is at (0, 0, 1). *Delete the north pole* $\mathbf{n} = (0, 0, 2)$ from Σ. Now imagine that there is a light source at the north pole, and for each point **p** of Σ, let $P(\mathbf{p})$ be the shadow of **p** in the xy plane (Fig. 4.29). As usual, we identify the xy plane with $\mathbf{R}^2$ by $(p_1, p_2, 0) \leftrightarrow (p_1, p_2)$. Thus we have defined a function P from Σ onto $\mathbf{R}^2$. Evidently P has the form

$$P(p_1, p_2, p_3) = \left(\frac{Rp_1}{r}, \frac{Rp_2}{r}\right),$$

where r and R are the distances from **p** and $P(\mathbf{p})$, respectively, to the z axis. But from the similar triangles in Fig. 4.30, we see that $R/2 = r/(2 - p_3)$; hence

$$P(p_1, p_2, p_3) = \left(\frac{2p_1}{2 - p_3}, \frac{2p_2}{2 - p_3}\right).$$

Now if **x** is any patch in Σ, the composite function $P\mathbf{x}$ is Euclidean differentiable, so $P\colon \Sigma \to \mathbf{R}^2$ is a mapping.

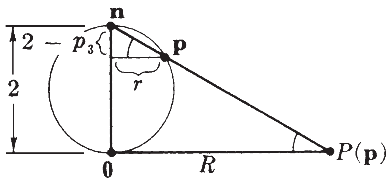


FIG. 4.30

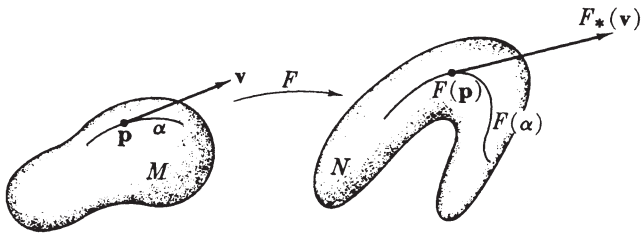


FIG. 4.31

Just as for mappings of Euclidean space, each mapping of surfaces has a tangent map.

5.3 Definition Let F: $M \to N$ be a mapping of surfaces. The *tangent map* $F*$ of F assigns to each tangent vector $\mathbf{v}$ to M the tangent vector $F*(\mathbf{v})$ to N such that if $\mathbf{v}$ is the initial velocity of a curve α in M, then $F*(\mathbf{v})$ is the initial velocity of the image curve $F(\alpha)$ in N (Fig. 4.31).

Furthermore, at each point $\mathbf{p}$, the tangent map $F*$ is a linear transformation from the tangent plane $T_p(M)$ to the tangent plane $T_{F(p)}(N)$ (see Exercise 9). It follows immediately from the definition that $F*$ preserves velocities of curves: If $\bar{\alpha} = F(\alpha)$ is the image in N of a curve α in M, then $F*(\alpha') = \bar{\alpha}'$. As in the Euclidean case, we deduce the convenient property that the tangent map of a composition is the composition of the tangent maps (Exercise 10).

The tangent map of a mapping F: $M \to N$ may be computed in terms of partial velocities as follows. If $\mathbf{x}$: $D \to M$ is a parametrization in M, let $\mathbf{y}$ be the composite mapping $F(\mathbf{x})$: $D \to N$ (which need not be a parametrization). Obviously, F carries the parameter curves of $\mathbf{x}$ to the corresponding parameter curves of $\mathbf{y}$. Since $F*$ preserves velocities of curves, it follows at once that

$$F*(\mathbf{x}_u) = \mathbf{y}_u, \quad F*(\mathbf{x}_v) = \mathbf{y}_v.$$

Since $\mathbf{x}_u$ and $\mathbf{x}_v$ give a basis for the tangent space of M at each point of $\mathbf{x}(D)$, these readily computable formulas completely determine $F*$.

The discussion of regular mappings in Section 7 of Chapter 1 translates easily to the case of a mapping of surfaces $F: M \to N$. F is *regular* provided all of its derivative maps $F*_p: T_p(M) \to T_{F(p)}(N)$ are one-to-one. Since these tangent planes have the same dimension, we know from linear algebra that the one-to-one requirement is equivalent to $F*$ being a linear isomorphism. A mapping $F: M \to N$ that has an inverse mapping $F^{-1}: N \to M$ is called a *diffeomorphism*. We may think of a diffeomorphism F as smoothly distorting M to produce N. By applying the Euclidean formulation of the inverse function theorem to a coordinate expression $\mathbf{y}^{-1}F\mathbf{x}$ for F, we can deduce this extension of the inverse function theorem (7.10 of Chapter 1).

5.4 Theorem Let $F: M \to N$ be a mapping of surfaces, and suppose that $F*_p: T_p(M) \to T_{F(p)}(N)$ is a linear isomorphism at some one point $\mathbf{p}$ of M. Then there exists a neighborhood $\mathscr{U}$ of $\mathbf{p}$ in M such that the restriction of F to $\mathscr{U}$ is a diffeomorphism onto a neighborhood $\mathscr{V}$ of $F(\mathbf{p})$ in N.

An immediate consequence is this useful result: *A one-to-one regular mapping F of M onto N is a diffeomorphism*. For since F is one-to-one and onto, it has a unique inverse function F^{-1}, and F^{-1} is a differentiable mapping since on each neighborhood $\mathscr{V}$ as above, it coincides with the inverse of the diffeomorphism $\mathscr{U} \to \mathscr{V}$. Surfaces M and N are said to be *diffeomorphic* if there exists a diffeomorphism from M to N.

Diffeomorphisms have little respect for size or shape; here are some examples.

5.5 Example (1) Any open rectangle in the plane $\mathbf{R}^2$ is diffeomorphic to the entire plane. Take $R: -\pi/2 < u, v < \pi/2$ for simplicity. Then $F(u,v) = (\tan u, \tan v)$ is a mapping of R onto $\mathbf{R}^2$. Using a branch of the inverse tangent function, the mapping $F^{-1}(u_1, v_1) = (\tan^{-1}(u_1), \tan^{-1}(v_1))$ is a differentiable inverse of F, so F is a diffeomorphism.

(2) The sphere Σ minus one point is also diffeomorphic to the entire plane. Stereographic projection P, as in Example 5.2(2), is a one-to-one mapping of the punctured sphere Σ_0 onto $\mathbf{R}^2$. A variant

$$\mathbf{x}(u, v) = (\cos v \cos u, \cos v \sin u, 1 + \sin v)$$

of the usual geographical parametrization is a parametrization of $\Sigma_0 - \mathbf{0}$. The formula for P in Example 5.2 gives

$$\mathbf{y}(u, v) = P(\mathbf{x}(u, v)) = \frac{2 \cos v}{1 - \sin v}(\cos u, \sin u).$$

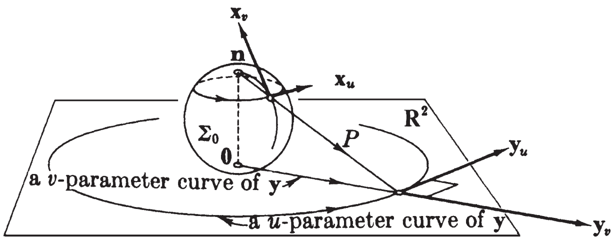


FIG. 4.32

Since $P*(\mathbf{x}_u) = \mathbf{y}_u$ and $P*(\mathbf{x}_v) = \mathbf{y}_v$, the regularity of F can be checked by computing $\mathbf{y}_u$ and $\mathbf{y}_v$. These turn out to be orthogonal and nonzero, hence linearly independent, as suggested by Fig. 4.32.

(3) A cylinder C over a closed curve is diffeomorphic to the plane minus one point. For simplicity, take C: $x^2 + y^2 = 1$, and define a mapping F: $C \to \mathbf{R}^2$ by $F(x, y, z) = e^z(x, y)$. Since e^z takes on all values $r > 0$, F maps C onto $\mathbf{R}^2 - \mathbf{0}$.

For the inverse of F, experimentation suggests

$$G(u, v) = \left(\frac{u}{\sqrt{u^2 + v^2}}, \frac{v}{\sqrt{u^2 + v^2}}, \log \sqrt{u^2 + v^2} \right).$$

To prove $G = F^{-1}$, compute $G(F(x, y, z)) = (x, y, z)$ and $F(G(u, v)) = (u, v)$.

Differential forms have the remarkable property that they can be moved from one surface to another by means of an arbitrary mapping.† Let us experiment with a 0-form, that is, a real-valued function. If F: $M \to N$ is a mapping of surfaces and f is a function on M, there is simply no reasonable general way to move f over to a function on N. But if instead, f is a function on N, the problem is easy; we pull f back to the composite function $f(F)$ on M. The corresponding pull-back for 1-forms and 2-forms is accomplished as follows.

5.6 Definition Let F: $M \to N$ be a mapping of surfaces.

(1) If ϕ is a 1-form on N, let $F^*\phi$ be the 1-form on M such that

$$(F^*\phi)(\mathbf{v}) = \phi(F_* \mathbf{v})$$

for all tangent vectors $\mathbf{v}$ to M.

(2) If η is a 2-form on N, let $F^*\eta$ be the 2-form on M such that

† This is not true for vector fields.

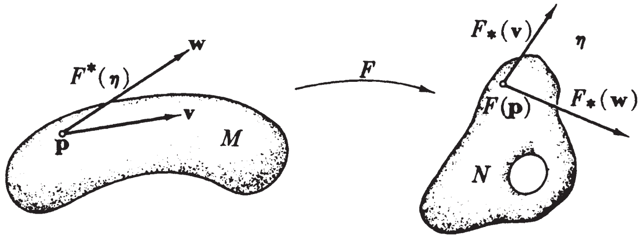


FIG. 4.33

$$(F^*\eta)(\mathbf{v},\mathbf{w}) = \eta(F_*\mathbf{v},\ F_*\mathbf{w})$$

for all pairs of tangent vectors $\mathbf{v}$, $\mathbf{w}$ on M (Fig. 4.33).

When we are dealing with a function f in its role as a 0-form, we shall sometimes write F^*f instead of $f(F)$, in accordance with the notation for the pullback of 1-forms and 2-forms.

The essential operations on forms are sum, wedge product, and exterior derivative; all are preserved by mappings.

5.7 Theorem Let F: $M \to N$ be a mapping of surfaces, and let ξ and η be forms on N. Then

(1) $F^*(\xi + \eta) = F^*\xi + F^*\eta$,
(2) $F^*(\xi \wedge \eta) = F^*\xi \wedge F^*\eta$,
(3) $F^*(d\xi) = d(F^*\xi)$.

Proof. In (1), ξ and η are both assumed to be p-forms (degree $p = 0, 1, 2$) and the proof is a routine computation. In (2), we must allow ξ and η to have different degrees. When, say, ξ is a function f, the given formula means simply $F^*(f\eta) = f(F)F^*(\eta)$. In any case, the proof of (2) is also a straightforward computation. But (3) is more interesting. The easier case when ξ is a function is left as an exercise, and we address ourselves to the case where ξ is a 1-form.

It suffices to show that for every patch $\mathbf{x}$: $D \to M$,

$$(d(F^*\xi))(\mathbf{x}_u, \mathbf{x}_v) = (F^*(d\xi))(\mathbf{x}_u, \mathbf{x}_v).$$

Let $\mathbf{y} = F(\mathbf{x})$, and recall that $F_*(\mathbf{x}_u) = \mathbf{y}_u$ and $F_*(\mathbf{x}_v) = \mathbf{y}_v$. Thus, using the definitions of d and F^*, we get

$$\begin{aligned} d(F^*\xi)(\mathbf{x}_u, \mathbf{x}_v) &= \frac{\partial}{\partial u}\{(F^*\xi)(\mathbf{x}_v)\} - \frac{\partial}{\partial v}\{(F^*\xi)(\mathbf{x}_u)\} \\ &= \frac{\partial}{\partial u}\{\xi(F_*\mathbf{x}_v)\} - \frac{\partial}{\partial v}\{\xi(F_*\mathbf{x}_u)\} \\ &= \frac{\partial}{\partial u}\{\xi(\mathbf{y}_v)\} - \frac{\partial}{\partial v}\{\xi(\mathbf{y}_u)\}. \end{aligned}$$

Even if $\mathbf{y}$ is not a patch, Exercise 6 of Section 4 shows that this last expression is still equal to $d\xi(\mathbf{y}_u, \mathbf{y}_v)$. But

$$d\xi(\mathbf{y}_u, \mathbf{y}_v) = d\xi(F_*\mathbf{x}_u, F_*\mathbf{x}_v) = (F^*(d\xi))(\mathbf{x}_u, \mathbf{x}_v).$$

Thus we conclude that $d(F^*\xi)$ and $F^*(d\xi)$ have the same value on $(\mathbf{x}_u, \mathbf{x}_v)$. ◆

The elegant formulas in Theorem 5.7 are the key to the deeper study of mappings. In Chapter 6 we shall apply them to the connection forms of frame fields to get fundamental information about the geometry of mappings of surfaces.

Exercises

1. Let M and N be surfaces in $\mathbf{R}^3$. If $F\colon \mathbf{R}^3 \to \mathbf{R}^3$ is a mapping such that the image $F(M)$ of M is contained in N, then the restriction of F to M is a function $F|M\colon M \to N$. Prove that $F|M$ is a mapping of surfaces. (*Hint:* Use Thm. 3.2.)

2. Let Σ be the sphere of radius r with center at the origin of $\mathbf{R}^3$. Describe the effect of the following mappings $F\colon \Sigma \to \Sigma$ on the meridians and parallels of Σ.

(a) $F(\mathbf{p}) = -\mathbf{p}$.

(b) $F(p_1, p_2, p_3) = (p_3, p_1, p_2)$.

(c) $F(p_1, p_2, p_3) = \left(\dfrac{p_1 + p_2}{\sqrt{2}}, \dfrac{p_1 - p_2}{\sqrt{2}}, -p_3\right).$

3. Let M be a *simple surface*, that is, one that is the image of a single proper patch $\mathbf{x}\colon D \to \mathbf{R}^3$. If $\mathbf{y}\colon D \to N$ is any mapping into a surface N, show that the function $F\colon M \to N$ such that

$$F(\mathbf{x}(u, v)) = \mathbf{y}(u, v) \quad \text{for all } (u, v) \text{ in } D$$

is a mapping of surfaces. (*Hint*: Write $F = \mathbf{y}\mathbf{x}^{-1}$, and use Cor. 3.3.)

4. Use the preceding exercise to construct a mapping of the helicoid H (Ex. 2.5) onto the torus T (Ex. 2.5) such that the rulings of H are carried to the meridians of T.

5. If Σ is the sphere $\| \mathbf{p} \| = r$, the mapping A: $\Sigma \to \Sigma$ such that $A(\mathbf{p}) = -\mathbf{p}$ is called the *antipodal map* of Σ. Prove that A is a diffeomorphism and that $A_*(\mathbf{v}_p) = (-\mathbf{v})_{-p}$.

6. A regular mapping F: $M \to N$ of surfaces is often called a *local diffeomorphism*. For such a mapping F, prove that, in fact, every point $\mathbf{p}$ of M has a neighborhood $\mathscr{U}$ such that $F|\mathscr{U}$ is a diffeomorphism of $\mathscr{U}$ onto a neighborhood of $F(\mathbf{p})$ in N.

7. If $\mathbf{x}$: $D \to M$ is a parametrization, prove that the restriction of $\mathbf{x}$ to a sufficiently small neighborhood of a point (u_0, v_0) in D is a patch in M. (Thus any parametrization can be cut into patches.)

8. Let F: $M \to N$ be a mapping. If $\mathbf{x}$ is a patch in M, then as in the text, let $\mathbf{y} = F(\mathbf{x})$. (Although $\mathbf{y}$ maps into N, it is not necessarily a patch.) For a curve

$$\alpha(t) = \mathbf{x}(a_1(t), a_2(t))$$

in M, show that the image curve $\bar{\alpha} = F(\alpha)$ in N, has velocity

$$\bar{\alpha}' = \frac{da_1}{dt}\mathbf{y}_u(a_1, a_2) + \frac{da_2}{dt}\mathbf{y}_v(a_1, a_2).$$

9. Prove: (a) The invariance property needed to justify the definition (5.3) of the tangent map.

(b) Tangent maps F_*: $T_p(M) \to T_{F(p)}(N)$ are linear transformations.

10. Given mappings $M \xrightarrow{F} N \xrightarrow{G} P$, let GF: $M \to P$ be the composite mapping. Show that

(a) GF is differentiable, (b) $(GF)_* = G_*F_*$,

(c) $(GF)^* = F^*G^*$,

that is, for any form ξ on P, $(GF)^*(\xi) = F^*(G^*(\xi))$. (Note the reversal of factors, caused by the fact that forms travel in the opposite direction from points and tangent vectors.)

11. Prove that every surface of revolution is diffeomorphic to either a torus or a cylinder. (*Hint:* Parametrize profile curves on the same interval.) (As Fig. 4.9 suggests, every *augmented* surface of revolution is diffeomorphic to either a plane or a sphere.)

12. (a) Show that the inverse mapping P^{-1} of the stereographic projection P: $\Sigma_0 \to \mathbf{R}^2$ is given by

$$P^{-1}(u, v) = \frac{(4u, 4v, 2f)}{f + 4}, \quad \text{where } f = u^2 + v^2.$$

(Check that both PP^{-1} and $P^{-1}P$ are identity maps.)

(b) Deduce that the entire sphere Σ can be covered by only two patches. (The scheme in Section 1 requires six.)

13. (*Consistent formulas.*) If $G: \tilde{M} \to M$ is a regular mapping onto M, and $\tilde{F}: \tilde{M} \to N$ is an arbitrary mapping, we say that the formula $F(G(\mathbf{q})) = \tilde{F}(\mathbf{q})$ is *consistent* provided

$$G(\mathbf{q}_1) = G(\mathbf{q}_2) \Rightarrow \tilde{F}(\mathbf{q}_1) = \tilde{F}(\mathbf{q}_2)$$

for $\mathbf{q}_1$, $\mathbf{q}_2$ in $\tilde{M}$. Prove:

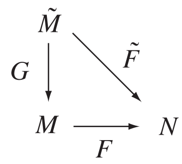
(a) In this case, F is a well-defined differentiable mapping from M to N.

(b) Furthermore, if the reverse implication

$$\tilde{F}(\mathbf{q}_1) = \tilde{F}(\mathbf{q}_2) \Rightarrow G(\mathbf{q}_1) = G(\mathbf{q}_2)$$

also holds, then F is one-to-one.

This result is helpful in constructing maps $F: M \to N$ with specified properties. Often G will be a parametrization of M.



4.6 Integration of Forms

Differential forms are no less important in integral calculus than in differential calculus. Indeed, they are just what is needed to establish integration theory on an arbitrary surface. In a sense, integration takes place only on Euclidean space, so a form on a surface is integrated by first pulling it back to Euclidean space.

Consider the one-dimensional case. Let $\alpha: [a, b] \to M$ be a curve segment on a surface M. The pullback $\alpha^*\phi$ of a 1-form ϕ on M to the interval $[a, b]$ has the expression $f(t)dt$, where by the remarks following Example 4.7,

$$f(t) = (\alpha^*\phi)(U_1(t)) = \phi(\alpha_*(U_1(t))) = \phi(\alpha'(t)).$$

Thus the scheme mentioned above yields the following result:

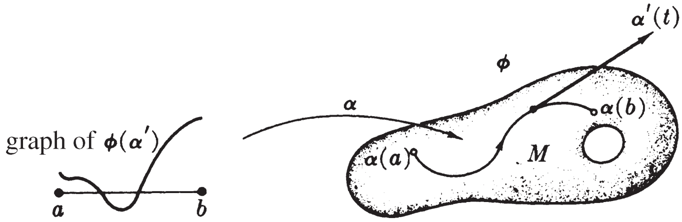


FIG. 4.34

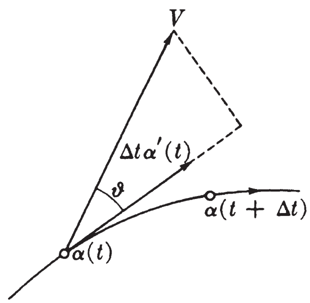


FIG. 4.35

6.1 Definition Let ϕ be a 1-form on M, and let α: $[a, b] \to M$ be a curve segment in M (Fig. 4.34). Then the *integral of ϕ over α* is

$$\int_a \phi = \int_{[a,b]} \alpha^* \phi = \int_a^b \phi(\alpha'(t))dt.$$

The integral $\int_\alpha \phi$, often called a *line integral*, has a wide variety of uses in science and engineering. For example, let us consider a vector field V on a surface M as a *force field*, and a curve α: $[a, b] \to M$ as a description of a moving particle, with $\alpha(t)$ its position at time t. What is the total amount of *work* W done by the force on the particle as it moves from $\mathbf{p} = \alpha(a)$ to $\mathbf{q} = \alpha(b)$? The discussion of velocity in Chapter 1, Section 4, shows that for Δt small, the subsegment of α from $\alpha(t)$ to $\alpha(t + \Delta t)$ is approximated by the straight line segment $\Delta t\ \alpha(t)$. Work is done on the particle only by the component of force *tangent* to α, that is,

$$V(\alpha) \bullet \frac{\alpha'}{\|\alpha'\|} = \|V(\alpha)\| \cos \vartheta,$$

where ϑ is the angle between $V(\alpha(t))$ and $\alpha'(t)$ (Fig. 4.35). Thus the work done by the force during time Δt is approximately the *force* (as above) times the *distance* $\|\alpha'(t)\|\ \Delta t$. Adding these contributions over the whole time interval $[a, b]$ and taking the usual limit, we get

$$W = \int_a^b V(\alpha(t)) \bullet \alpha'(t)dt.$$

To express this more simply, we introduce the 1-form ϕ *dual* to the vector field V; its value on a tangent vector $\mathbf{w}$ at $\mathbf{p}$ is $\mathbf{w} \bullet V(\mathbf{p})$. Then by Definition 6.1, the total work is just

$$W = \int_a \phi.$$

We emphasize that this notion of line integral—like everything we do with forms—applies without essential change if the surface M is replaced by a Euclidean space or, indeed, by any *manifold* (Section 8).

When the 1-form being integrated is the differential of a function, we get the following generalization of the fundamental theorem of calculus.

6.2 Theorem Let f be a function on M, and let $\alpha\colon [a, b] \to M$ be a curve segment in M from $\mathbf{p} = \alpha(a)$ to $\mathbf{q} = \alpha(b)$. Then

$$\int_\alpha df = f(\mathbf{q}) - f(\mathbf{p}).$$

Proof. By definition,

$$\int_\alpha df = \int_a^b df(\alpha')dt.$$

But

$$df(\alpha') = \alpha'[f] = \frac{d}{dt}(f(\alpha)).$$

Hence, by the fundamental theorem of calculus,

$$\int_a df = \int_a^b \frac{d}{dt}(f(\alpha))\,dt = f(\alpha(b)) - f(\alpha(a)) = f(\mathbf{q}) - f(\mathbf{p}).$$ ◆

Thus the integral $\int_a df$ is *path independent:* its value is the same for all curves from $\mathbf{p}$ to $\mathbf{q}$. Hence it is zero for all closed curves, $\alpha(a) = \alpha(b)$.

The preceding theorem can be interpreted roughly as follows: The "boundary" of the curve segment α from $\mathbf{p}$ to $\mathbf{q}$ is $\mathbf{q} - \mathbf{p}$, where the purely formal minus sign indicates that $\mathbf{p}$ is the starting point of α. Then the integral of df over α equals the "integral" of f over the boundary of α, namely, $f(\mathbf{q}) - f(\mathbf{p})$. This interpretation will be justified by the analogous theorem (6.5) in dimension 2.

If we consider a closed rectangle R: $a \leqq u \leqq b$, $c \leqq v \leqq d$ in $\mathbf{R}^2$ as a 2-dimensional interval, then a 2-*segment* is a differentiable map $\mathbf{x}\colon R \to M$ of R into M (Fig. 4.36). As for a 1-segment, differentiability means that $\mathbf{x}$ can be extended over a larger open set containing R. Although we use the patch notation $\mathbf{x}$, we do not assume that $\mathbf{x}$ is either regular or one-to-one—but the partial velocities $\mathbf{x}_u$ and $\mathbf{x}_v$ are still well defined.

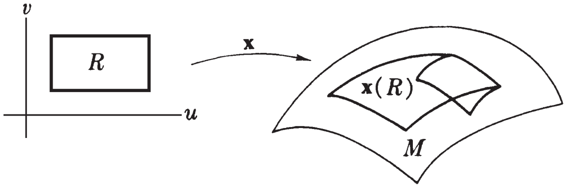


FIG. 4.36

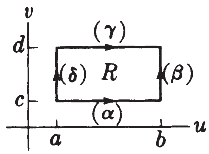


FIG. 4.37

If η is a 2-form on M, then the pullback $\mathbf{x}^*\eta$ of η has, using Example 4.7, the expression $h\, du\, dv$, where

$$h = (\mathbf{x}^*\eta)(U_1, U_2) = \eta(\mathbf{x}_*U_1, \mathbf{x}_*U_2) = \eta(\mathbf{x}_u, \mathbf{x}_v).$$

Then strict analogy with Definition 6.1 yields:

6.3 Definition Let η be a 2-form on M, and let $\mathbf{x}$: $R \to M$ be a 2-segment in M. The *integral of* η *over* $\mathbf{x}$ is

$$\iint_{\mathbf{x}} \eta = \iint_R x^*\eta = \int_a^b \int_c^d \eta(\mathbf{x}_u, \mathbf{x}_v)du\, dv.$$

The physical applications of this integral are no less rich than those of Definition 6.1; however, we proceed directly toward the analogue of Theorem 6.2.

6.4 Definition Let $\mathbf{x}$: $R \to M$ be a 2-segment in M, with R the closed rectangle $a \leqq u \leqq b$, $c \leqq v \leqq d$ (Fig. 4.37). The *edge curves* of $\mathbf{x}$ are the curve segments α, β, γ, δ such that

$$\alpha(u) = \mathbf{x}(u, c),$$

$$\beta(v) = \mathbf{x}(b, v),$$

$$\gamma(u) = \mathbf{x}(u, d),$$

$$\delta(v) = \mathbf{x}(a, v).$$

Then the *boundary* $\partial\mathbf{x}$ of the 2-segment $\mathbf{x}$ is the formal expression

$$\partial\mathbf{x} = \alpha + \beta - \gamma - \delta.$$

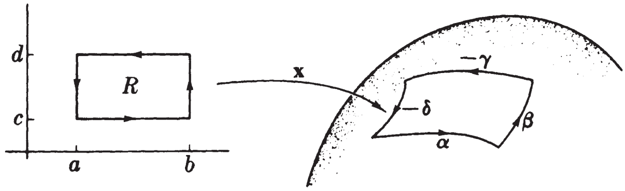


FIG. 4.38

These four curve segments are gotten by restricting the function $\mathbf{x}$: $R \to M$ to the four line segments that make up the boundary of the rectangle R (Fig. 4.37). The formal minus signs before γ and δ signal that the parametrizations of γ and δ would have to be reversed to give a consistent trip around the rim of $\mathbf{x}(R)$ (Fig. 4.38).

Then if ϕ is a 1-form on M, the *integral of* ϕ *over the boundary* $\partial\mathbf{x}$ *of* $\mathbf{x}$ is defined to be

$$\int_{\partial\mathbf{x}} \phi = \int_\alpha \phi + \int_\beta \phi - \int_\gamma \phi - \int_\delta \phi.$$

The 2-dimensional analogue of Theorem 6.2 is

6.5 Theorem (Stokes' theorem) If ϕ is a 1-form on M, and $\mathbf{x}$: $R \to M$ is a 2-segment, then

$$\iint_{\mathbf{x}} d\phi = \int_{\partial\mathbf{x}} \phi.$$

Proof. We work on the double integral and show that it turns into the integral of ϕ over the boundary of $\mathbf{x}$. Combining Definitions 6.3 and 4.4 gives

$$\iint_R d\phi(\mathbf{x}_u, \mathbf{x}_v)\, du\, dv = \iint_R \left(\frac{\partial}{\partial u}(\phi(\mathbf{x}_v) - \frac{\partial}{\partial v}(\phi(\mathbf{x}_u))) \right) du\, dv.$$

Let $f = \phi(\mathbf{x}_u)$ and $g = \phi(\mathbf{x}_v)$. Then the equation above becomes

$$\iint_{\mathbf{x}} d\phi = \iint_R \frac{\partial g}{\partial u} du\, dv - \iint_R \frac{\partial f}{\partial v} du\, dv. \tag{1}$$

Now we treat these double integrals as iterated integrals. Suppose the rectangle R is given, as usual, by the inequalities $a \leqq u \leqq b$, $c \leqq v \leqq d$. Then integrating first with respect to u, we find

$$\iint_R \frac{\partial g}{\partial u} du\, dv = \int_c^b I(v) dv, \quad \text{where } I(v) = \int_a^b \frac{\partial g}{\partial u}(u, v)\, du.$$

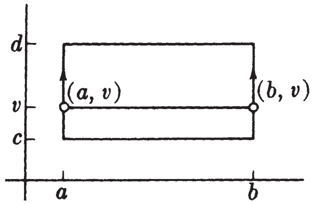


FIG. 4.39

In the partial integral defining $I(v)$, v is constant, so the integrand is just the ordinary derivative with respect to u. Thus, the fundamental theorem of calculus applies to give $I(v) = g(b, v) - g(a, v)$ (Fig. 4.39). Hence

$$\iint_R \frac{\partial g}{\partial u} du\, dv = \int_c^d g(b,v)\, dv - \int_c^d g(a,v)\, dv. \tag{2}$$

Consider the first integral on the right. By definition, $g(b, v) = \phi(\mathbf{x}_v(b, v))$. But $\mathbf{x}_v(b, v)$ is exactly the velocity $\beta'(v)$ of the "right side" curve β in $\partial\mathbf{x}$. Hence by Definition 6.1,

$$\int_c^d g(b,v)\, dv = \int_c^d \phi(\beta'(v))\, dv = \int_\beta \phi.$$

A similar argument shows that the other integral in (2) is $\int_\delta \phi$. Thus

$$\iint_R \frac{\partial g}{\partial u} du\, dv = \int_\beta \phi - \int_\delta \phi. \tag{3}$$

In the same way—but integrating first with respect to v—we find

$$\iint_R \frac{\partial f}{\partial v} du\, dv = \int_\gamma \phi - \int_\alpha \phi. \tag{4}$$

Assembling the information in (1), (3), and (4) gives the required result,

$$\iint_{\mathbf{x}} d\phi = \left(\int_\beta \phi - \int_\delta \phi\right) - \left(\int_\gamma \phi - \int_\alpha \phi\right) = \int_{\partial\mathbf{x}} \phi. \qquad \blacklozenge$$

Stokes' theorem ranks as one of the most useful results in all mathematics. Alternative formulations and extensive applications can be found in texts on advanced calculus and applied mathematics. We will use it to study the geometry and topology of surfaces.

The line integral $\int_\alpha \phi$ is not particularly sensitive to reparametrization of the curve segment α. All that matters is the overall direction in which the route of α is traversed, as indicated by what the reparametrization does to end points.

6.6 Lemma Let $\alpha(h)$: $[a, b] \to M$ be a reparametrization of a curve segment α: $[c, d] \to M$ by h: $[a, b] \to [c, d]$. For any 1-form ϕ on M,

(1) If h is *orientation-preserving*, that is, if $h(a) = c$ and $h(b) = d$, then

$$\int_{\alpha(h)} \phi = \int_{\alpha} \phi.$$

(2) If h is *orientation-reversing*, that is, if $h(a) = d$ and $h(b) = c$, then

$$\int_{\alpha(h)} \phi = -\int_{\alpha} \phi.$$

Proof. The velocity of $\alpha(h)$ is $\alpha(h)' = dh/dt\ \alpha'(h)$, so

$$\int_{\alpha(h)} \phi = \int_a^b \phi\Big(\alpha(h)'\Big)\, dt = \int_a^b \phi(\alpha'(h))\frac{dh}{dt}\, dt.$$

Now we apply the theorem on change of variables in an integral to the last integral above. If h is orientation-preserving, then

$$\int_{\alpha(h)} \phi = \int_c^d \phi(\alpha')\, du = \int_{\alpha} \phi,$$

while in the orientation-reversing case,

$$\int_{\alpha(h)} \phi = \int_d^c \phi(\alpha')\, du = -\int_c^d \phi(\alpha')\, du = -\int_{\alpha} \phi. \quad \blacklozenge$$

This lemma provides a concrete interpretation to the formal minus signs in the boundary $\partial\mathbf{x} = \alpha + \beta - \gamma - \delta$ of a 2-segment. For any curve

$$\xi: [t_0, t_1] \to M,$$

let $-\xi$ be any orientation-*reversing* reparametrization of ξ, for instance, $(-\xi)(t) = \xi(t_0 + t_1 - t)$. Then by the lemma,

$$\int_{-\xi} \phi = -\int_{\xi} \phi.$$

Thus the formula for $\int_{\partial\mathbf{x}} \phi$ just before Theorem 6.5 can be rewritten as

$$\int_{\partial\mathbf{x}} \phi = \int_{\alpha} \phi + \int_{\beta} \phi + \int_{-\gamma} \phi + \int_{-\delta} \phi.$$

Exercises

1. If α is a curve in $\mathbf{R}^2$ and ϕ is a 1-form, prove this computational rule for finding $\phi(\alpha')dt$: Substitute $u = \alpha_1$ and $v = \alpha_2$ into the coordinate expression $\phi = f(u, v)\, du + g(u, v)\, dv$.

2. Let α: $[-1, 1] \to \mathbf{R}^2$ be the curve segment given by $\alpha(t) = (t, t^2)$.

(a) If $\phi = v^2\, du + 2uv\, dv$, compute $\int_{\alpha} \phi$.

(b) Find a function f such that $df = \phi$ and check Theorem 6.2 in this case.

3. Let ϕ be a 1-form on a surface M. Show:

(a) If ϕ is closed, then $\int_{\partial \mathbf{x}} \phi = 0$ for every 2-segment $\mathbf{x}$ in M.

(b) If ϕ is exact, then more generally,

$$\int_{\alpha} \phi = \sum_i \int_{\alpha_i} \phi = 0$$

for any closed, piecewise smooth curve whose smooth segments are $\alpha_1, \ldots, \alpha_k$ (hence α_k ends at the start of α_1).

4. The 1-form

$$\psi = \frac{u\,dv - v\,du}{u^2 + v^2}$$

is well-defined on the plane $\mathbf{R}^2$ with the origin $\mathbf{0}$ removed. Show:

(a) ψ is closed but not exact on $\mathbf{R}^2 - \mathbf{0}$. (*Hint:* Integrate around the unit circle and use Ex. 3.)

(b) The restriction of ψ to, say, the right half-plane $u > 0$ is exact.

5. (a) Show that every curve α in $\mathbf{R}^2$ that does not pass through the origin has an (orientation-preserving) reparametrization in the polar form

$$\alpha(t) = \big(r(t)\cos\vartheta(t),\, r(t)\sin\vartheta(t)\big).$$

(*Hint:* Use Ex. 12 of Sec. 2.1.)

If the curve $\alpha\colon [a, b] \to \mathbf{R}^2 - \mathbf{0}$ is closed, prove:

(b) $\operatorname{wind}(\alpha) = \dfrac{\vartheta(b) - \vartheta(a)}{2\pi}$ is an integer.

This integer, called the *winding number of* α *about* $\mathbf{0}$, represents the total algebraic number of times α has gone around the origin in the counterclockwise direction. (Note that $\operatorname{wind}(\alpha) = \operatorname{wind}(\alpha/\|\alpha\|)$.)

(c) If ψ is the 1-form in Exercise 4, then $\operatorname{wind}(\alpha) = \dfrac{1}{2\pi}\displaystyle\int_{\alpha} \psi$.

(d) If $\alpha = (f, g)$, then

$$\operatorname{wind}(\alpha) = \frac{1}{2\pi}\int_a^b \frac{fg' - gf'}{f^2 + g^2}\,dt = \frac{1}{2\pi}\int_a^b \frac{\det(\alpha(t), \alpha'(t))}{\alpha(t)\bullet\alpha(t)}\,dt.$$

(The determinant is of the 2×2 matrix whose rows are $\alpha(t)$ and $\alpha'(t)$.)

6. (*Continuation, by computer.*) For a point $\mathbf{p} \in \mathbf{R}^2$ not on a closed curve α, the *winding number of* α *about* $\mathbf{p}$ is defined to be $\operatorname{wind}(\alpha - \mathbf{p})$.

(a) Write commands that given a closed curve α in $\mathbf{R}^2$ and a point $\mathbf{p}$ not on α, return the winding number of α about $\mathbf{p}$. (*Hint:* Use either integral in (d) of the preceding exercise.)

(b) In each case, plot the curve α (observing its orientation) and estimate the winding numbers about the indicated points $\mathbf{p}$. Then calculate these numbers, using the integral in (d) above. (Numerical integration is efficient here, since the result is known to be an integer.)

(i) lemniscate, $\alpha(t) = (2\sin t, \sin 2t)$; $\mathbf{p} = (1, 0), (0, 1), (-1, 0)$.

(ii) limaçon, $\beta(t) = (3\sin t + 1)(\cos t, \sin t)$; $\mathbf{p} = (0, 1), (0, 3), (0, 5)$.

7. Let F: $M \to N$ be a mapping. Prove:

(a) If α is a curve segment in M, and ϕ is a 1-form on N, then $\int_\alpha F^*\phi = \int_{F(\alpha)} \phi$.

(b) If $\mathbf{x}$ is 2-segment in M, and η is a 2-form on N, then $\iint_{\mathbf{x}} F^*\eta = \iint_{F(\mathbf{x})} \eta$.

8. Let $\mathbf{x}$ be a patch in a surface M. For a curve segment

$$\alpha(t) = \mathbf{x}(a_1(t), a_2(t)), \quad a \leqq t \leqq b,$$

in $\mathbf{x}(R)$, show that

$$\int_\alpha \phi = \int_a^b \left(\phi(\mathbf{x}_u)\frac{da_1}{dt} + \phi(\mathbf{x}_v)\frac{da_2}{dt} \right) dt,$$

where $\mathbf{x}_u$ and $\mathbf{x}_v$ are evaluated on (a_1, a_2). (This generalizes Ex. 1, which is recovered by using the identity patch $\mathbf{x}(u, v) = (u, v)$ in $\mathbf{R}^2$.)

9. Let $\mathbf{x}$ be the usual parametrization of the torus T (Ex. 2.5). For integers m and n, let α be the closed curve

$$\alpha(t) = \mathbf{x}(mt, nt) \quad (0 \leqq t \leqq 2\pi).$$

Find:

(a) $\int_\alpha \xi$, where ξ is the 1-form such that $\xi(\mathbf{x}_u) = 1$ and $\xi(\mathbf{x}_v) = 0$.

(b) $\int_\alpha \eta$, where η is the 1-form such that $\eta(\mathbf{x}_u) = 0$ and $\eta(\mathbf{x}_v) = 1$.

For an arbitrary closed curve γ in T, $\int_\gamma \xi/(2\pi)$ is an integer that counts the total (algebraic) number of times γ goes around the torus in the general direction of the parallels, and $\int_\gamma \eta/(2\pi)$ gives a similar count for the meridians. (This suggests the informal notation $\xi = d\vartheta$, $\eta = d\varphi$, but see Ex. 7 of Sec. 7.)

10. Let $\mathbf{x}$: $R \to M$ be a 2-segment defined on the unit square R: $0 \leqq u, v \leqq 1$. If ϕ is the 1-form on M such that

$$\phi(\mathbf{x}_u) = u + v \quad \text{and} \quad \phi(\mathbf{x}_v) = uv,$$

compute $\iint_{\mathbf{x}} d\phi$ and $\int_{\partial\mathbf{x}} \phi$ separately, and check the results by Stokes' theorem. (*Hint:* $\mathbf{x}^* d\phi = d(\mathbf{x}^*\phi)$.)

11. Same as Exercise 10, except that $\boldsymbol{R}$: $0 \leqq u \leqq \pi/2$, $0 \leqq v \leqq \pi$, and $\phi(\mathbf{x}_u) = u\cos v$, $\phi(\mathbf{x}_v) = v\sin u$.

The following exercise is a 2-dimensional analogue of Lemma 6.6. However, with future applications in mind, we generalize 2-segments $\mathbf{x}$: $R \to M$ by replacing the rectangle R by any compact region $\mathscr{R}$ in $\mathbf{R}^2$ whose boundary consists of smooth curve segments. (Compactness ensures that integrals over $\mathscr{R}$ will be finite.)

12. (*Effect of change of variables.*) Let $\mathbf{x}$: $\mathscr{S} \to M$ be a differentiable mapping and let (U, V): $\mathscr{R} \to \mathscr{S}$ be a one-to-one regular map whose Jacobian determinant

$$J(u, v) = \det\begin{bmatrix} \dfrac{\partial U}{\partial u} & \dfrac{\partial U}{\partial v} \\[2mm] \dfrac{\partial V}{\partial u} & \dfrac{\partial V}{\partial v} \end{bmatrix}$$

is always positive (*orientation-preserving* case) or always negative (*orientation-reversing* case). Then let $\mathbf{y}$: $\mathscr{R} \to M$ be given by $\mathbf{y}(u, v) = \mathbf{x}(U(u, v), V(u, v))$.

(a) For a 2-form η on M, use

$$\mathbf{x}_u = \mathbf{y}_U \frac{\partial U}{\partial u} + \mathbf{y}_V \frac{\partial V}{\partial u}, \quad \mathbf{x}_v = \mathbf{y}_U \frac{\partial U}{\partial v} + \mathbf{y}_V \frac{\partial V}{\partial v}$$

to prove

$$\eta(\mathbf{x}_u, \mathbf{x}_v) = J(u, v)\, \eta(\mathbf{y}_U(U, V), \mathbf{y}_V(U, V)).$$

(b) Deduce that $\iint_{\mathbf{x}} \eta = \iint_{\mathbf{y}} \eta$ in the orientation-preserving case, and minus this in the orientation-reversing case.

(*Hint:* The formula for change of variables in a double integral involves the absolute value of a Jacobian determinant.)

13. The *classical Stokes' theorem* asserts, typically, that if $\mathbf{x}$: $D \to \mathbf{R}^3$ is a 2-segment and V is a vector field on $\mathbf{R}^3$, then

$$\int_{\partial\mathbf{x}} V \bullet ds = \iint_{\mathbf{x}} U \bullet (\nabla \times V)\, dA,$$

where $\nabla \times V = \operatorname{curl} V$. Interpret this as a special case of Theorem 6.5. (*Hint:* Assume $dA \approx \sqrt{EG - F^2}\, du\, dv$†, and use Ex. 8 of Sec. 1.6.)

† This geometric result will be clarified later on. Note that Theorem 6.5 does not involve geometry.

4.7 Topological Properties of Surfaces

Topological properties are the most basic a surface can have. In this section we discuss four such properties, phrasing the definitions in terms most efficient for geometry.

7.1 Definition A surface is *connected* provided that for any two points **p** and **q** of M there is a curve segment in M from **p** to **q**. (See also Exercise 9.)

Thus a connected surface M is all in one piece, since one can travel from any point in M to any other without leaving M. Most of the surfaces we have met so far have been connected; the surface M: $z^2 - x^2 - y^2 = 1$ (a hyperboloid of two sheets) is not. Connectedness is a mild and natural condition that is sometimes included in the definition of surface.

The general definition of *compactness* is expressed in terms of open coverings. An *open covering* of a set A is a collection of open sets that *covers A* in the sense that each point of A is in at least one of the sets.

A subset A in a space S (for us, either a Euclidean space or a surface) is *compact* provided that given any open covering of A some finite number of the sets already covers A. In elementary calculus it is usually proved that a closed interval I: $a \leqq t \leqq b$ in **R** is compact, and this result extends to higher dimensions. In particular, any closed rectangle R: $a \leqq u \leqq b$, $c \leqq v \leqq d$ in $\mathbf{R}^2$ is compact.

We will need this abstract definition at a few crucial points, but in surface theory the following concrete criterion is more useful.

7.2 Lemma A surface M is compact if and only if it can be covered by the images of a finite number of 2-segments in M.

Proof. Suppose M is compact. For each point **p** in M, by using a coordinate patch containing **p**, we can construct a 2-segment whose image contains a neighborhood of **p**. The definition of compactness shows that a finite number of these neighborhoods already covers M; hence the corresponding 2-segments cover M.

The converse is an exercise in finiteness. First we show that the image $\mathbf{x}(R)$ of a single 2-segment is compact. Recall that the definition of differentiability for 2-segments allows us to assume that **x** has been smoothly extended over an open set containing the (closed) rectangle R.

Let $\{\mathscr{U}_\alpha\}$ be an open covering of M. For each point **r** in R, one of these sets, say $\mathscr{U}_\mathbf{r}$, contains $\mathbf{x}(\mathbf{r})$. Being differentiable, **x** is also continuous, so **r** has a neighborhood $\mathcal{N}_r$ that is carried into $\mathcal{U}_r$ by $\mathbf{x}$. For all $\mathbf{r}$, these neighborhoods $\mathcal{N}_r$ form an open covering of R. As mentioned above, the rectangle R is compact, so some finite number of these neighborhoods cover R. But this means that the corresponding original sets $\mathcal{U}_r$ (finite in number!) cover $\mathbf{x}(\boldsymbol{R})$; hence it is compact.

Now suppose that M is covered by the images of a finite number of 2-segments $\mathbf{x}_1, \ldots, \mathbf{x}_k$. If $\{\mathcal{U}_\alpha\}$ is an open covering of M, then we have just shown that a finite number of these sets suffice to cover the image of each $\mathbf{x}_i$. Collecting these sets for $i = 1, 2, \ldots, k$ produces a finite number of sets $\mathcal{U}_a$ that cover M. Thus M is compact. ◆

It follows at once from this proof that *a region R in M is compact if it is composed of the images of finitely many 2-segments in M*. For example, spheres are compact, since if the formula for the geographical patch (Example 2.2) is applied on the closed rectangle.

$$R: -\pi \leqq u \leqq \pi, \quad -\pi/2 \leqq v \leqq \pi/2,$$

then this single 2-segment covers the entire sphere†. Similarly, the torus (Example 2.5) is compact, as is every surface of revolution whose profile curve is closed.

The following lemma generalizes this fundamental fact: A continuous real-valued function on a closed rectangle R in the plane takes on a maximum at some point of R.

7.3 Lemma A continuous function f on a compact region $\mathcal{R}$ in a surface M takes on a maximum at some point of M.

Proof. We show this in the only case we need: where $\mathcal{R}$ consists of the images $\mathbf{x}_1(R_1), \ldots, \mathbf{x}_k(R_k)$ of a finite number of 2-segments (for example, where $\mathcal{R}$ is an entire surface).

Since each $\mathbf{x}_i$ is continuous, the composite functions $f(\mathbf{x}_i)$: $R_i \to \mathbf{R}$ are all continuous. By the remark above, for each index i, there is a point (u_i, v_i) in R_i where $f(\mathbf{x}_i)$ is a maximum. Let $f(\mathbf{x}_j(u_j, v_j))$ be the largest of these finite number k of maximum values; then evidently $f(\mathbf{p}) \leqq f(\mathbf{x}_j(u_j, v_j))$ for all $\mathbf{p}$ in M. ◆

† Amazingly, *every* compact surface can be expressed as the image of a single 2-segment. See Ch. 1 of [Ma].

This lemma is useful in proving noncompactness. For example, a cylinder C such as $x^2 + y^2 = r^2$ is not compact since the coordinate function z is unbounded on C.

Finite size alone does not produce compactness. For example, the open disk $\mathscr{D}$: $x^2 + y^2 < 1$ in the xy plane is itself a surface. Although $\mathscr{D}$ is bounded and has finite area π, it is not compact since the function $f = (1 - x^2 - y^2)^{-1}$ is continuous on $\mathscr{D}$ and does not have a maximum.

In general, a compact surface cannot have open edges, as $\mathscr{D}$ does, but must be smoothly closed up everywhere and finite in size—like a sphere or torus.

Roughly speaking, an orientable surface is one that is not twisted. Of the many equivalent definitions of orientability for surfaces, the following is perhaps the simplest.

7.4 Definition A surface M is *orientable* if there exists a differentiable (or merely continuous) 2-form μ on M that is nonzero at every point of M.

Recall that a 2-form is zero at a point $\mathbf{p}$ if it is zero on every pair of tangent vectors at $\mathbf{p}$—or equivalently, on one linearly independent pair. Thus the plane $\mathbf{R}^2$ is orientable since *du dv* is a nonvanishing 2-form. This definition of orientability is somewhat mysterious, so for a surface M in $\mathbf{R}^3$ we give a more intuitive description in terms of Euclidean geometry. A *unit normal U* on M is a differentiable Euclidean vector field on M that has unit length and is everywhere normal to M.

7.5 Proposition A surface $M \subset \mathbf{R}^3$ is orientable if and only if there exists a unit normal vector field on M. If M is connected as well as orientable, there are exactly two unit normals, $\pm U$.

Proof. We use the cross product of $\mathbf{R}^3$ to convert normal vector fields into 2-forms, and vice versa.

Let U be a unit normal on M. If $\mathbf{v}$ and $\mathbf{w}$ are tangent vectors to M at $\mathbf{p}$, define

$$\mu(\mathbf{v}, \mathbf{w}) = U(\mathbf{p}) \bullet \mathbf{v} \times \mathbf{w}.$$

Standard properties of the cross product show that μ is a 2-form on M. When $\mathbf{v}$ and $\mathbf{w}$ are linearly independent, so are all three vectors, so $\mu(\mathbf{v}, \mathbf{w}) \neq 0$. Thus μ is nonvanishing, which proves that M is orientable.

Conversely, suppose M is orientable, with μ a nonvanishing 2-form. Again, nonvanishing implies that if $\mathbf{v}$, $\mathbf{w}$ are linearly independent vectors at a point $\mathbf{p}$ of M, then $\mu(\mathbf{v}, \mathbf{w}) \neq 0$. Now define

$$Z(\mathbf{p}) = \frac{\mathbf{v} \times \mathbf{w}}{\mu(\mathbf{v}, \mathbf{w})}.$$

This formula is independent of the choice of **v**, **w**. Explicitly, for any other such pair **v**′, **w**′, it follows from Lemma 4.2 and the analogous formula for cross products that

$$\frac{\mathbf{v}' \times \mathbf{w}'}{\mu(\mathbf{v}', \mathbf{w}')} = \frac{\mathbf{v} \times \mathbf{w}}{\mu(\mathbf{v}, \mathbf{w})}.$$

Properties of the cross product show that $Z(p)$ is nonzero and normal to M. The formula for cross product shows that Z is differentiable. Thus $U = Z/\|Z\|$ is the required unit normal.

If U is a unit normal on M, then so is $-U$. To show that there are no others, let V also be a unit normal. At each point these (differentiable) unit vector fields are collinear, so the only values for the dot product $V \bullet U$ are $+1$ and -1. On a connected surface, a nonvanishing differentiable function cannot change sign (Exercise 4), hence *either* $V \bullet U = +1$ everywhere, so $V = U$, *or* $V \bullet U = -1$ everywhere, so $V = -U$. ◆

For example, all spheres, cylinders, surfaces of revolution, and quadric surfaces are orientable. It follows from Lemma 3.8 that every surface in $\mathbf{R}^3$ that can be defined implicitly is orientable.

However, nonorientable surfaces do exist in $\mathbf{R}^3$. The simplest example is the famous Möbius band B, made from a strip of paper by giving it a half twist, then gluing its ends together (Fig. 4.40). B is nonorientable since it cannot have a (differentiable) unit normal. To see this let γ be a closed curve, as in Fig. 4.40, that runs once around the band with $\gamma(0) = \gamma(1) = \mathbf{p}$. Now suppose a unit normal vector U at $\gamma(0)$ moves continuously around γ. As the figure shows, the twist in B forces the contradiction

$$U(\mathbf{p}) = U(\gamma(1)) = -U(\gamma(0)) = -U(\mathbf{p}).$$

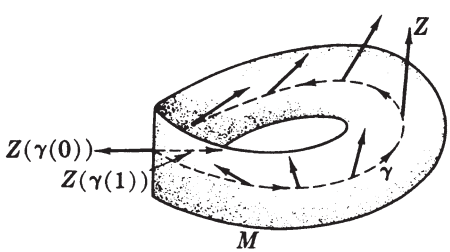


FIG. 4.40

The last of the topological properties we consider will let us express formally the intuitive idea that a plane is simpler than a cylinder, and a sphere is simpler than a torus. The key is that in the cylinder or a torus there are closed curves that cannot be continuously shrunk down to a point.

7.6 Definition A closed curve α in M is *homotopic to a constant* provided there is a 2-segment $\mathbf{x}$: $R \to M$ (called a *homotopy*) defined on

$$R: a \leqq u \leqq b, 0 \leqq v \leqq 1$$

such that α is the base curve of $\mathbf{x}$ and the other three edge curves are constant at $\mathbf{p} = \alpha(a) = \alpha(b)$.

A curve such as α for which $\alpha(a) = \alpha(b)$ holds but not necessarily $\alpha'(a) = \alpha'(b)$ is often called a *loop at* $\mathbf{p}$. Since the sides β and δ of $\mathbf{x}$ are constant at $\mathbf{p}$, for every $0 \leqq v \leqq 1$ the u-parameter curve $\alpha_v(u) = \mathbf{x}(u, v)$ is also a loop at $\mathbf{p}$. As v varies from 0 to 1, the loop α_v varies continuously from $\alpha_0 = \alpha$ to the curve $\alpha_1 = \gamma$, which is constant at $\mathbf{p}$.

It is easy to show that in the plane every loop is homotopic to a constant. Given a loop α: $[a, b] \to \mathbf{R}^2$ at $\mathbf{p}$ in $\mathbf{R}^2$, use scalar multiplication in $\mathbf{R}^2$ to define $\mathbf{x}(u, v) = v\alpha(a) + (1 - v)\alpha(u)$. Then

$$\mathbf{x}(u, 0) = \alpha(u) \text{ and } \mathbf{x}(u, 1) = \mathbf{p} \quad \text{for all } a \leqq u \leqq b, \text{ and}$$

$$\mathbf{x}(a, v) = \mathbf{x}(b, v) = \mathbf{p} \qquad \text{for all } 0 \leqq v \leqq 1.$$

Hence $\mathbf{x}$ is a homotopy from α to a constant.

7.7 Definition A surface M is *simply connected* provided it is connected and every loop in M is homotopic to a constant.

(This definition is valid for any manifold—and more generally.) The preceding homotopy shows that the plane $\mathbf{R}^2$ is simply connected, and the same formula works for any Euclidean space.

The 2-sphere Σ is also simply connected. Consider the following scheme of proof. Let α be a loop in Σ at, say, the north pole of Σ. Pick a point $\mathbf{q}$ not on α. For simplicity, suppose $\mathbf{q}$ is the south pole. Now let $\mathbf{x}$ be the homotopy under which each point of α moves due north along a great circle, reaching $\mathbf{p}$ in unit time. This $\mathbf{x}$ is a homotopy of α to a constant, as required.

But there is a difficulty here: finding the point $\mathbf{q}$. In our usual case, where α is differentiable, techniques from advanced calculus will show that there is always a point $\mathbf{q}$ not on α. However, if α is merely continuous, it may actually fill the entire sphere. In this case, topological methods can be used to deform α slightly, making it no longer space-filling; then the scheme above is valid.

To show that a given surface is simply connected, we can always try to construct the necessary homotopies; however, to show that a surface is *not* simply connected, indirect means are usually required. One of the most effective derives from integration. Recall that a differential form ϕ is *closed* if $d\phi = 0$.

7.8 Lemma Let ϕ be a closed 1-form on a surface M. If a loop α in M is homotopic to a constant, then $\int_\alpha \phi = 0$.

Proof. Suppose $\mathbf{x}$ is a homotopy showing that a loop α is homotopic to a constant, say $\mathbf{p}$. Now we apply Stokes' theorem (Theorem 6.5). The integral over a constant curve is zero, and $d\phi = 0$, hence

$$0 = \iint_{\mathbf{x}} d\phi = \int_\alpha \phi. \qquad \blacklozenge$$

Now suppose we remove a single point, say the origin $\mathbf{0}$, from the (simply connected) plane $\mathbf{R}^2$. The loop α: $[0, 2\pi] \to C$ given by $\alpha(t) = (\cos t, \sin t)$ circles once around the missing point. It seems obvious that α cannot be shrunk down to a point in the punctured plane $P = \mathbf{R}^2 - 0$. The preceding lemma provides an easy way to prove it.

Exercise 4 of the preceding section shows that the 1-form

$$\psi = \frac{x\,dy - y\,dx}{x^2 + y^2}$$

on P is closed and that its integral around α is 2π (these are easy computations). By the lemma, α is not homotopic to a constant; hence P is not simply connected.

As noted earlier, since $d^2 = 0$, exact forms are always closed. However, closed forms need not be exact. For example, if the closed 1-form ψ were exact, it would follow from Stokes' theorem (6.5) that $\int_\alpha \psi = 0$. However, in an important special case, closed 1-forms *are* exact.

7.9 Lemma (Poincaré) On a simply connected surface, every closed 1-form is exact.

Proof. First we show that *the integral of a closed 1-form is path independent*, that is, if α and β are curve segments from $\mathbf{p}$ to $\mathbf{q}$, then $\int_\alpha \phi = \int_\beta \phi$. In fact, if $-\beta$ is an orientation-reversing reparametrization of β, then $\alpha + (-\beta) = \alpha - \beta$ is a loop. By simple connectedness, it is homotopic to a constant. Thus, using Lemma 6.6,

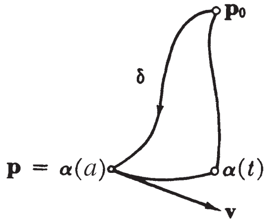


FIG. 4.41

$$0 = \int_{\alpha-\beta} \phi = \int_\alpha \phi + \int_{-\beta} \phi = \int_\alpha \phi - \int_\beta \phi.$$

Now suppose ϕ is a closed 1-form on a simply connected surface M. Pick a point $\mathbf{p}_0$ and define $f(\mathbf{p}) = \int_\delta \phi$ for any curve segment δ from $\mathbf{p}_0$ to $\mathbf{p}$. Path independence makes f a well-defined function on M.

To show that $df = \phi$, we must show that $df(\mathbf{v}) = \phi(\mathbf{v})$ for every tangent vector $\mathbf{v}$ at a point $\mathbf{p}$. This is equivalent to $\mathbf{v}[f] = \phi(\mathbf{v})$.

Let α: $[a, b] \to M$ be a curve with initial velocity $\alpha'(a) = \mathbf{v}$. Then $\delta + \alpha|[a, t]$ is a curve from $\mathbf{p}_0$ to $\alpha(t)$ (Fig. 4.41), so by the definition of f,

$$f(\alpha(t)) = \int_{\delta+\alpha|[a,\,t]} \phi = f(\mathbf{p}) + \int_a^t \phi(\alpha'(u))\, du.$$

Taking the derivative with respect to t gives

$$\alpha'(t)[f] = (f(\alpha))'(t) = \phi(\alpha'(t)).$$

When $t = 0$ this becomes $\mathbf{v}[f] = \phi(\mathbf{v})$, as required. ◆

Among the four properties we have discussed there are two direct implications—both yielding orientability.

7.10 Theorem A compact surface in $\mathbf{R}^3$ is orientable.

This is an easy consequence of the following nontrivial topological theorem, a 2-dimensional version of the Jordan Curve Theorem. *If M is a compact surface in $\mathbf{R}^3$, then M separates $\mathbf{R}^3$ into two nonempty open sets: an* exterior *(the points that can escape to infinity) and an* interior *(the points trapped inside M)*. So we need only pick the unit normal vector at each $\mathbf{p}$ in M that points into, say, the exterior and apply Proposition 7.5. Thus orientation is at stake when in elementary calculus the "outward unit normal" is assigned to a surface in $\mathbf{R}^3$.

7.11 Theorem A simply connected surface is orientable.

We defer the proof until Section 4 of Chapter 8. (Notice, for future reference, that this theorem does not mention $\mathbf{R}^3$.)

A final note: Because the properties discussed in this section are topological, they can be defined solely in terms of open sets and continuous functions. However, differentiable versions are usually more practical for use in geometry.

Exercises

1. Which of the following surfaces are compact and which are connected?

(a) A sphere with one point removed.

(b) The region $z > 0$ in M: $z = xy$.

(c) A torus with the curve $\alpha(t) = \mathbf{x}(t, t)$ removed. (See Ex. 2 of Sec. 4.3.)

(d) The surface in Fig. 4.8.

(e) M: $x^2 + y^4 + z^6 = 1$.

2. Let F be a mapping of a surface M *onto* a surface N. Prove:

(a) If M is connected, then N is connected.

(b) If M is compact, then N is compact. (Try both the covering definition and the criterion in Lem. 7.2.)

3. Let F: $M \to N$ be a regular mapping. Prove that if N is orientable, then M is orientable.

4. Let f be a differentiable real-valued function on a connected surface. Prove:

(a) If $df = 0$, then f is constant.

(b) If f is never zero then either $f > 0$ or $f < 0$.

5. Of the four basic types of surfaces of revolution (see Ex. 11 of Sec. 5)—plane, sphere, cylinder, torus—which are,

(a) connected? (b) compact?

(c) orientable? (d) simply connected?

A closed curve α in M is *freely* homotopic to a constant if the conditions on $\mathbf{x}$ in Definition 7.6 are weakened to $\beta = \delta$ with only γ required to be constant (Fig. 4.42). Then the v constant curves α_v are loops that move along the curve $\beta = \delta$ as they shrink to $\mathbf{p}$.

6. (a) If a loop α is freely homotopic to a constant via $\mathbf{x}$, show that for any 1-form ϕ,

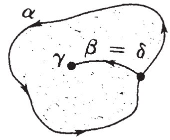


FIG. 4.42

$$\int_{\alpha} \phi = \iint_{\mathbf{x}} d\phi.$$

(b) If closed curves α and β in $\mathbf{R}^2 - \mathbf{p}$ are freely homotopic in $\mathbf{R}^2 - \mathbf{p}$, show that they have the same winding number about $\mathbf{p}$.

(c) A *smooth disk* $\mathcal{D}$ in a surface is the image of the unit disk $x^2 + y^2 \leq 1$ in $\mathbf{R}^2$ under a one-to-one regular map F. Show that the 2-segment

$$\mathbf{x}(u, v) = F(u \cos v, u \sin v), \quad (0 \leqq u \leqq 1, 0 \leqq v \leqq 2\pi),$$

fills $\mathcal{D}$ and is a free homotopy of the (closed) boundary curve $v \to \mathbf{x}(1, v)$ to a constant.

7. Let ϕ be a closed 1-form and α a closed curve.

(a) Show that $\int_{\alpha} \phi = 0$ if either ϕ is exact or α is freely homotopic to a constant.

(b) Deduce that on a torus T the meridians and parallels are not freely homotopic to constants, and the closed 1-forms ξ and η of Exercise 9 of Section 6 are not exact.

8. (*Counterexamples.*) Give examples to show that the following are false:

(a) Converses of (a) and (b) of Exercise 2.

(b) Exercise 3 with F not regular.

(c) Converse of Exercise 3.

9. (a) If $\mathbf{p}$ is a point of a surface S, show that the set of all points of S that can be connected to $\mathbf{p}$ by a (piecewise smooth) curve in S is an open set of S. (*Hint:* Each point of a surface has a neighborhood that is connected in the sense of Def. 7.1.)

(b) Same as (a) but with *can* replaced by *cannot*.

(c) For a surface M, show that Definition 7.1 ("path-connectedness") is equivalent to the general topological definition of connectedness, namely: If $\mathcal{U}$ and $M - \mathcal{U}$ are open sets of M, and $\mathcal{U}$ contains at least one point, then $M = \mathcal{U}$. (Use the corresponding property for a closed interval in $\mathbf{R}$, a standard result of analysis.)

10. The Hausdorff axiom asserts that distinct points $\mathbf{p} \neq \mathbf{q}$ have disjoint neighborhoods. Prove:

(a) $\mathbf{R}^3$ obeys the Hausdorff axiom. (The same proof works for all $\mathbf{R}^n$.)
(b) A surface M in $\mathbf{R}^3$ obeys the Hausdorff axiom.

11. If $\mathscr{R}$ is a compact region in a surface M, prove that $\mathscr{R}$ is a closed set of M, that is, $M - \mathscr{R}$ is an open set. (*Hint*: To show that $M - \mathscr{R}$ is open, use the preceding exercise and the fact that a finite intersection of neighborhoods of $\mathbf{p}$ is again a neighborhood of $\mathbf{p}$.)

12. Let M and N be surfaces in $\mathbf{R}^3$ such that M is contained in N.
(a) If M is compact and N is connected, prove that $M = N$. (*Hint:* Show that M is both closed and open in N.)
(b) Give examples to show that (a) fails if either M is not compact or N is not connected.
(c) Deduce from (a) that if F: $M \to N$ is a local diffeomorphism with M compact and N connected, then $F(M) = N$.

4.8 Manifolds

Surfaces in $\mathbf{R}^3$ are a matter of everyday experience, so it is reasonable to investigate them mathematically. But examining this concept with a critical eye, we may well ask whether there could not be surfaces in $\mathbf{R}^4$, or in $\mathbf{R}^n$—or even surfaces that are not contained in any Euclidean space. To devise a definition for such a surface, we must rely not on our direct experience of the real world, but on our mathematical experience of surfaces in $\mathbf{R}^3$. Thus we shall strip away from Definition 1.2 every feature that involves $\mathbf{R}^3$. What is left will be just a surface.

To begin with, a surface will be a set M: a collection of any objects whatsoever. We call these objects the *points* of M, but as examples below will show, they definitely need not be the usual points of some Euclidean space. An *abstract patch in M* will now be just a one-to-one function $\mathbf{x}$: $D \to M$ from an open set D of $\mathbf{R}^2$ into the set M.

To get a workable definition of surface we must find a way to define what it means for functions involving M to be differentiable. The key to this problem turns out to be the smooth overlap condition in Corollary 3.3. To *prove* this condition is now a logical impossibility since $\mathbf{R}^3$ is gone, so in the usual fashion of mathematics, we make it an axiom.

8.1 Definition A *surface* is a set M furnished with a collection $\mathscr{P}$ of abstract patches in M satisfying

(1) *The covering axiom:* The images of the patches in the collection $\mathscr{P}$ cover M.

(2) *The smooth overlap axiom:* For any patches **x**, **y** in $\mathscr{P}$, the composite functions $\mathbf{y}^{-1}\mathbf{x}$ and $\mathbf{x}^{-1}\mathbf{y}$ are Euclidean differentiable—and defined on open sets of $\mathbf{R}^2$.

This definition generalizes Definition 1.2: A surface in $\mathbf{R}^3$ *is* a surface in this sense. However, there is a technical gap in this definition that requires attention. First, for any patch $\mathbf{x}: D \to M$ in a surface, define a set $\mathbf{x}(\mathscr{U})$ to be open provided $\mathscr{U}$ is open in $D \subset \mathbf{R}^2$. Then the *open sets* of M are all unions of such sets. (This is consistent with the case $M \subset \mathbf{R}^3$, since there **x** and $\mathbf{x}^{-1}$ are continuous.)

Examples like that in Exercise 11 show that for the open sets to behave properly we must add another axiom to the definition of surface.

(3) *The Hausdorff axiom:* For any points $\mathbf{p} \neq \mathbf{q}$ in M there exist disjoint (that is, nonoverlapping) patches **x** and **y** with **p** in $\mathbf{x}(D)$ and **q** in $\mathbf{y}(E)$.

Here is an example of an important surface that, as we will soon see, cannot be found in $\mathbf{R}^3$.

8.2 Example *The projective plane P*. Starting from the unit sphere Σ in $\mathbf{R}^3$ we construct P by identifying *antipodal points* in Σ, that is, by considering **p** and $-\mathbf{p}$ to be the same point of P (Fig. 4.43). Formally, this means that the set P consists of all antipodal pairs $\{\mathbf{p}, -\mathbf{p}\}$ of points in Σ. Order is not relevant here; that is, $\{\mathbf{p}, -\mathbf{p}\} = \{-\mathbf{p}, \mathbf{p}\}$. (Working with the projective plane often involves looking back and forth between antipodal points.)

There are two important mappings associated with P: the *antipodal map* $A(\mathbf{p}) = -\mathbf{p}$ on Σ and the *projection* $F(\mathbf{p}) = \{\mathbf{p}, -\mathbf{p}\}$ of Σ onto P. Note that $FA = F$.

Call a patch **x** in Σ "small" if it is contained in a single open hemisphere. Then the composite function $F\mathbf{x}$ is one-to-one, and is thus an abstract patch. The collection of all such abstract patches makes P a surface. In fact, the covering condition (1) is clear, and the Hausdorff axiom derives from the corresponding property for Euclidean spaces. The smooth overlap axiom (2) can be checked as follows.

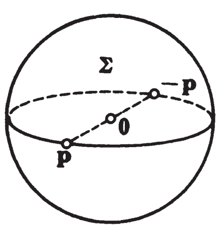


FIG. 4.43

Suppose $F\mathbf{x}$ and $F\mathbf{y}$ overlap in Σ; that is, their images have points in common. If $\mathbf{x}$ and $\mathbf{y}$ overlap in Σ, then $(F\mathbf{y})^{-1}F\mathbf{x} = \mathbf{y}^{-1}\mathbf{x}$, which by Corollary 3.3 is differentiable and defined on an open set. On the other hand, if $\mathbf{x}$ and $\mathbf{y}$ do not overlap, replace $\mathbf{y}$ by $A\mathbf{y}$. Then $\mathbf{x}$ and $A\mathbf{y}$ *do* overlap, so the previous argument applies.

To emphasize the distinction between a surface in $\mathbf{R}^3$ and the general notion of surface defined above, we sometimes call the latter an *abstract surface*.

To get as many patches as possible in an abstract surface M we always understand that its patch collection $\mathscr{P}$ has been enlarged to include all the abstract patches in M that overlap smoothly with those originally in $\mathscr{P}$. We emphasize that abstract surfaces M_1 and M_2 with the *same* set of points are nevertheless different surfaces if their (enlarged) patch collections $\mathscr{P}_1$ and $\mathscr{P}_2$ are different.

There is essentially only one problem to solve in establishing calculus on an abstract surface M, and that is to define the *velocity* of a curve in M. In the old definition a velocity vector was a tangent vector to $\mathbf{R}^3$, so something new is needed. For everything else—differentiable functions, curves themselves, tangent vectors, vector fields, differentiable forms, and so on—the *definitions* and *theorems* given for surfaces in $\mathbf{R}^3$ apply without change. It is necessary to tinker with a few *proofs*, but no serious problems arise.

It makes little difference what we define velocity to be—provided the new definition produces the same essential behavior as before. The most efficient choice is based on the directional derivative property in Lemma 4.6, Chapter 1.

8.3 Definition Let $\alpha\colon I \to M$ be a curve in an abstract surface M. For each t in I the *velocity vector* $\alpha'(t)$ is the function such that

$$\alpha'(t)[f] = \frac{d(f\alpha)}{dt}(t)$$

for every differentiable real-valued function f on M.

Thus $\alpha'(t)$ is a real-valued function whose domain is the set $\mathscr{F}$ of all real-valued functions on M. This is all we need to generalize the calculus on surfaces in $\mathbf{R}^3$ to the case of an abstract surface.

We now have a calculus for $\mathbf{R}^n$ (Chapter 1) and another one for surfaces. These are strictly analogous, but analogies in mathematics, though useful initially, can be annoying in the long run. What we need is a single calculus of which these two will be special cases. The most general object on which calculus can be conducted is called a *manifold*. It is simply an abstract surface of arbitrary dimension n.

8.4 Definition An *n-dimensional manifold M* is a set furnished with a collection $\mathscr{P}$ of *abstract patches* (one-to-one functions $\mathbf{x}: D \to M$, D an open set in $\mathbf{R}^n$) satisfying

(1) The covering property: The images of the patches in the collection $\mathscr{P}$ cover M.

(2) The smooth overlap property: For any patches $\mathbf{x}$, $\mathbf{y}$ in $\mathscr{P}$, the composite functions $\mathbf{y}^{-1}\mathbf{x}$ and $\mathbf{x}^{-1}\mathbf{y}$ are Euclidean differentiable—and defined on open sets of $\mathbf{R}^n$.

(3) The Hausdorff property: For any points $\mathbf{p} \neq \mathbf{q}$ in M there are disjoint patches $\mathbf{x}$ and $\mathbf{y}$ with $\mathbf{p}$ in $\mathbf{x}(D)$ and $\mathbf{q}$ in $\mathbf{y}(E)$.

Thus a surface (Definition 8.1) is just a 2-dimensional manifold. As before, Euclidean n-space $\mathbf{R}^n$ is an n-dimensional manifold whose (initial) patch collection consists only of the identity map.

To keep this definition as close as possible to that of a surface in $\mathbf{R}^3$, we have deviated somewhat from the standard definition of manifold in which it is the inverse functions $\mathbf{x}^{-1}$: $\mathbf{x}(D) \to D$ that are axiomatized.

The calculus of an arbitrary n-dimensional manifold is defined in the same way as for $n = 2$. Usually we need only replace $i = 1, 2$ by $i = 1, 2, \ldots, n$. Differential forms on an n-dimensional manifold have the same general properties as in the case $n = 2$, which we have explored in Sections 4, 5, and 6. But there are p-forms for $0 \leqq p \leqq n$, so when n is large, the algebra becomes more complicated.

Wherever calculus appears in mathematics and its applications, manifolds will be found, and higher dimensional manifolds turn out to be important in problems—both pure and applied—that initially seem to involve only dimensions 2 or 3. For example, here is a 4-dimensional manifold that has already appeared, implicitly at least, in this chapter.

8.5 Example *The tangent bundle of a surface*. For a surface M, let $T(M)$ be the set of all tangent vectors to M at all points of M. (For simplicity we assume M is a surface in $\mathbf{R}^3$, but it could just as well be an abstract surface or, indeed, a manifold of any dimension.) Since M has dimension 2 and each tangent space $T_p(M)$ has dimension 2, we anticipate that $T(M)$ will have dimension 4.

To get a natural patch collection for $T(M)$ we derive from each patch $\mathbf{x}$ in M an abstract patch $\tilde{x}$ in $T(M)$. Given $\mathbf{x}$: $D \to M$, let $\tilde{D}$ be the open set in $\mathbf{R}^4$

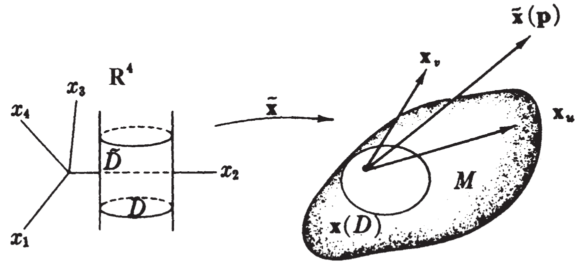


FIG. 4.44

consisting of all points (p_1, p_2, p_3, p_4) for which (p_1, p_2) is in D. Then let $\tilde{\mathbf{x}}$ be the function $\tilde{D} \to T(M)$ given by

$$\tilde{\mathbf{x}}(p_1, p_2, p_3, p_4) = p_3\mathbf{x}_u(p_1, p_2) + p_4\mathbf{x}_v(p_1, p_2).$$

(In Fig. 4.44 we identify $\mathbf{R}^2$ with the x_1x_2 plane of $\mathbf{R}^4$ and deal as best we can with dimension 4.)

Using Exercise 3 of Section 3 and the proof of Lemma 3.6, it is not difficult to check that each such function $\tilde{\mathbf{x}}$ is one-to-one and that the collection of all such patches satisfies the conditions in Definition 8.4. Thus $T(M)$ is a 4-dimensional manifold, called the *tangent bundle* of M.

Exercises

1. Prove that a surface M is nonorientable if there is a smoothly closed curve α: $[a, b] \to M$ and a tangent vector field Y on α such that

(i) Y and α' are linearly independent at every point, and

(ii) $Y(b) = -Y(a)$.

(*Hint:* Assume M is orientable and deduce a contradiction.)

2. Establish the following properties of the projective plane P.

(a) If F: $\Sigma \to P$ is the projection, then each tangent vector to P is the image under $F*$ of exactly two tangent vectors to Σ, these of the form $\mathbf{v}_p$ and $-\mathbf{v}_{-p}$.

(b) P is compact, connected, and nonorientable—hence P is not diffeomorphic to any surface in $\mathbf{R}^3$.

3. (a) Prove that the tangent bundle $T(M)$ of a surface is a manifold. (If $\mathbf{x}$ and $\mathbf{y}$ are overlapping patches in M, find an explicit formula for $\tilde{\mathbf{y}}^{-1}\tilde{\mathbf{x}}$.)

(b) If M is the image of a single patch $\mathbf{x}$: $\mathbf{R}^2 \to M$, show that the tangent bundle of M is diffeomorphic to $\mathbf{R}^4$.

4. A surface M in $\mathbf{R}^3$ is *closed* if it is a closed set of $\mathbf{R}^3$, that is, $\mathbf{R}^3 - M$ is an open set. (Confusingly, closed surface has sometimes been defined by analogy with closed curve to mean *compact surface*.)

(a) Prove that every surface given in the implicit form M: $g = c$ is closed.
(b) Prove that a compact surface in $\mathbf{R}^3$ is closed. (*Hint:* Use the scheme of Ex. 11 of Sec. 7.)
(c) Give an example of a closed surface in $\mathbf{R}^3$ that is not compact.

5. A surface M in $\mathbf{R}^3$ is *bounded* provided there is a number $R > 0$ such that $\|\mathbf{p}\| \leqq R$ for all $\mathbf{p} \in M$.

(a) Prove that a compact surface $M \subset \mathbf{R}^3$ is bounded.
(b) Give an example of a surface in $\mathbf{R}^3$ that is bounded but not compact. It follows from this exercise and Exercise 4 that if $M \subset \mathbf{R}^3$ is compact, it is closed and bounded. The converse is true but is more difficult to prove (see [Mu]).

6. Let $\hat{M}$ be the set of all the unit normal vectors on a surface M in $\mathbf{R}^3$. (So there are two points of $\hat{M}$ for each point of M.) For each patch $\mathbf{x}$ in M we define two patches in $\hat{M}$, namely,

$$\mathbf{x}_{\pm}(u, v) = \pm \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}.$$

(a) Show that the set of all such patches makes $\hat{M}$ a surface, called the *orientation covering surface* of M.
(b) Describe the orientation covering surface of the unit sphere $\Sigma \subset \mathbf{R}^3$ and of the torus $T \subset \mathbf{R}^3$.
(c) If a connected surface M in $\mathbf{R}^3$ is orientable, show that $\hat{M}$ consists of two diffeomorphic copies of M. (*Hint:* M has smooth unit normals $\pm U$.)
(d) The natural map $\pm \boldsymbol{U} \to \mathbf{p}$ of $\hat{M}$ onto M is regular.

7. (*Continuation.*) If a connected surface M in $\mathbf{R}^3$ is nonorientable, show that $\hat{M}$ is (i) connected and (ii) orientable. (Thus nonorientability can be cured by doubling. For an example, see Ex. 10.)
(*Hints:* (i) If α: $[a, b] \to M$ is a curve, then any unit normal vector at $\alpha(a)$ can be moved differentiably along α as a unit normal $U_\alpha(t)$—thus giving a curve in $\hat{M}$. Assume this fact: if M is nonorientable, there exists a loop α in M with $U_{\alpha(b)} = -U_{\alpha(a)}$.
(ii) Any patch $\mathbf{z}$ determines a unit vector field $U_\mathbf{z} = \mathbf{z}_u \times \mathbf{z}_v / \|\mathbf{z}_u \times \mathbf{z}_v\|$. If patches $\mathbf{x}$ and $\mathbf{y}$ in M meet, but $U_\mathbf{y} = -U_\mathbf{x}$ on the overlap, then $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ do not meet.)

8. A Möbius band B can be constructed as a ruled surface by

$$\mathbf{x}(u, v) = \beta(u) + v\delta(u), \quad \text{with, say, } -1/3 < v < 1/3,$$

where $\beta(u) = (\cos u, \sin u, 0)$ and

$$\delta(u) = \left(\cos\frac{u}{2}\cos u, \cos\frac{u}{2}\sin u, \sin\frac{u}{2}\right).$$

(The ruling makes only a half turn as it traverses the circle β.)

Show that $\mathbf{x}$ is one-to-one and regular, with unit normal

$$U(u, v) = \frac{\mathbf{x}_u(u, v) \times \delta(u)}{\|\mathbf{x}_u(u, v)\|}.$$

9. (*Continuation, computer graphics.*)
(a) Plot the Möbius band B.
(b) Plot a surface that represents B with its central circle β removed (for clarity, remove a small band around β). Is this surface connected? orientable?

Although a point of $\hat{M}$ is a *vector* in $\mathbf{R}^3$, not a point of $\mathbf{R}^3$, in favorable cases $\hat{M}$ can be turned into a surface in $\mathbf{R}^3$ by mapping each U_p in $\hat{M}$ to the point $\mathbf{p} + \varepsilon U_p$ (Euclidean coordinates) in $\mathbf{R}^3$ for some small $\varepsilon > 0$. (For $\varepsilon = 1$, the point would be the tip of the "arrow" U_p.)

10. (*Continuation.*)
(a) Using the scheme above, plot the orientation covering surface $\hat{B}$ of the Möbius band B. (Take $\varepsilon = 1/4$.)
(b) By inspection, is $\hat{B}$ connected? orientable? How is $\hat{B}$ related to the surface in (b) of the preceding exercise?

11. (*Plane with two origins.*) Let Z consist of all ordered pairs of real numbers and one additional point $\mathbf{0}^*$. Let $\mathbf{x}$ and $\mathbf{y}$ be the functions from $\mathbf{R}^2$ to Z such that

$$\mathbf{x}(u, v) = \mathbf{y}(u, v) = (u, v) \quad \text{if} \quad (u, v) \neq (0, 0),$$

but

$$\mathbf{x}(0, 0) = \mathbf{0} = (0, 0) \quad \text{and} \quad \mathbf{y}(0, 0) = \mathbf{0}^*.$$

(a) Show that the abstract patches $\mathbf{x}$ and $\mathbf{y}$ constitute a patch collection that satisfies the first two conditions in Definition 8.1, but not the Hausdorff axiom. Without the Hausdorff axiom, strange things can happen. For example, prove:
(b) A convergent sequence in Z can have two limits.
(c) The function F: $Z \to Z$ that reverses $\mathbf{0}$ and $\mathbf{0}^*$, leaving all other points fixed, is a differentiable mapping.

12. (a) Given a one-to-one function H from a manifold M onto an arbitrary set A, prove there is a unique way to make A a manifold so that H becomes a diffeomorphism. (*Hint*: Diffeomorphisms move patches to patches.)

(b) In each of the following cases, find natural choices of H and M that make the set a manifold.

(i) The set of all 2×2 real symmetric matrices.

(ii) The set of all circles in $\mathbf{R}^2$.

(iii) The set of all great circles on a sphere Σ.

(iv) The set of all (finite) closed intervals in $\mathbf{R}$.

13. (*Integral curves.*) A curve α in M is an *integral curve* of a vector field V on M provided $\alpha'(t) = V(\alpha(t))$ for all t. Thus an integral curve has at each point the velocity prescribed by V. If $\alpha(0) = \mathbf{p}$, we say that α *starts at* $\mathbf{p}$.

(a) In $\mathbf{R}^2$, show that the curve $\alpha(t) = (u(t), v(t))$ is an integral curve of $V = f_1U_1 + f_2U_2$ starting at $(a, b) \in \mathbf{R}^2$ if and only if

$$u' = f_1(u, v), \quad u(0) = a,$$
$$v' = f_2(u, v), \quad v(0) = b.$$

The theory of differential equations guarantees that there is a unique solution for α.

(b) Find an explicit formula for the integral curve of $V = -u^2U_1 + uvU_2$ on $\mathbf{R}^2$ that starts at the point $(1, -1)$. (The differential equations involved can be solved by elementary methods since one of them is particularly simple.)

(c) Sketch (by hand or by computer) the integral curve α on suitable intervals $A < t < -1$ and $-1 < t < B$.

14. (*Continuation.*) Show that every vector field V on a surface M has an integral curve β starting at any given point $\mathbf{p}$. Specifically, if $\mathbf{x}$: $D \to M$ is a patch with $\mathbf{x}(a, b) = \mathbf{p}$, and $\overline{V}$ is the vector field on D such that $\mathbf{x}_*(\overline{V}) = V$, show that $\beta(t) = \mathbf{x}(u(t), v(t))$, where $(u(t), v(t))$ is the integral curve of $\overline{V}$ starting at (a, b).

15. (*Cartesian products.*) For any sets A and B the Cartesian product $A \times B$ consists of all ordered pairs (a, b), with a in A and b in B. If $\mathbf{x}$: $D \to M$ and $\mathbf{y}$: $E \to N$ are patches in surfaces M and N, define $\mathbf{x} \times \mathbf{y}$: $D \times E \to M \times N$ by

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

Show that $\mathbf{x} \times \mathbf{y}$ is an abstract patch and that the collection $\mathscr{P}$ of all such patches makes $M \times N$ a 4-dimensional manifold. $M \times N$ is called the *Cartesian product* of M and N.

The same scheme works for any two manifolds. It derives from the way the x axis and y axis produce the xy plane; indeed, $\mathbf{R}\times\mathbf{R}$ is precisely $\mathbf{R}^2$.

16. If M is an abstract surface, a *proper imbedding* of M into $\mathbf{R}^3$ is a one-to-one regular mapping F: $M \to \mathbf{R}^3$ such that the inverse function F^{-1}: $F(M) \to M$ is continuous. Prove that the image $F(M)$ of a proper imbedding is a surface in $\mathbf{R}^3$ (Def. 1.2) and that it is diffeomorphic to M.

If F: $M \to \mathbf{R}^3$ is merely regular, then F is an *immersion* of M into $\mathbf{R}^3$, and the image $F(M)$ is often called an "immersed surface," even though it can cut across itself and hence not satisfy Definition 1.2.

4.9 Summary

The discovery of calculus made it possible to study arbitrary curved surfaces M in $\mathbf{R}^3$. Initially this was done mostly in terms of the natural Euclidean coordinates $\{x, y, z\}$ of $\mathbf{R}^3$. However, it gradually became clear that in many contexts, coordinates $\{u, v\}$ *in the surface itself* were more efficient. Thus a two-dimensional calculus was developed for surfaces, one that remains valid even if the surface is not contained in $\mathbf{R}^3$.

Along with the Euclidean spaces, such surfaces are prime examples of the general notion of *manifold*. The calculus of any manifold involves differentiable functions, vector fields, differential forms, mappings—and various operations of differentiation and integration. These features are all preserved in a suitable sense by diffeomorphisms—indeed, this criterion gives a formal definition of *manifold theory*.