Elementary Differential Geometry

Revised Second Edition

Introduction

This book presupposes a reasonable knowledge of elementary calculus and linear algebra. It is a working knowledge of the fundamentals that is actually required. The reader will, for example, frequently be called upon to *use* the chain rule for differentiation, but its proof need not concern us.

Calculus deals mostly with real-valued functions of one or more variables, linear algebra with functions (linear transformations) from one vector space to another. We shall need functions of these and other types, so we give here general definitions that cover all types.

A *set* S is a collection of objects that are called the *elements* of S. A set A is a *subset* of S provided each element of A is also an element of S.

A *function* f from a set D to a set R is a rule that assigns to each element x of D a unique element $f(x)$ of R. The element $f(x)$ is called the *value* of f at x. The set D is called the *domain* of f; the set R is sometimes called the *range* of f. If we wish to emphasize the domain and range of a function f, the notation $f\colon D \to R$ is used. Note that the function is denoted by a single letter, say f, while $f(x)$ is merely a value of f.

Many different terms are used for functions—mappings, transformations, correspondences, operators, and so on. A function can be described in various ways, the simplest case being an explicit formula such as

$$f(x) = 3x^2 + 1,$$

which we may also write as $x \to 3x^2 + 1$.

If both f_1 and f_2 are functions from D to R, then $f_1 = f_2$ means that $f_1(x) = f_2(x)$ for all x in D. This is not a definition, but a logical consequence of the definition of *function*.

Let $f\colon D \to R$ and $g\colon E \to S$ be functions. In general, the *image* of f is the subset of R consisting of all elements of the form $f(x)$; it is usually denoted by $f(D)$. If this image happens to be a subset of the domain E of g, it is possible to combine these two functions to get the *composite function* $g(f): D \to S$. By definition, $g(f)$ is the function whose value at each element x of D is the element $g(f(x))$ of S.

If $f: D \to R$ is a function and A is a subset of D, then the *restriction* of f to A is the function $f|A: A \to R$ defined by the same rule as f, but applied only to elements of A. This seems a rather minor change, but the function $f|A$ may have properties quite different from f itself.

Here are two vital properties that a function may possess. A function $f: D \to R$ is *one-to-one* provided that if x and y are any elements of D such that $x \neq y$, then $f(x) \neq f(y)$. A function $f: D \to R$ is *onto* (or *carries D onto R*) provided that for every element y of R there is at least one element x of D such that $f(x) = y$. In short, the image of f is the entire set R. For example, consider the following functions, each of which has the real numbers as both domain and range:

(1) The function $x \to x^3$ is both one-to-one and onto.
(2) The exponential function $x \to e^x$ is one-to-one, but not onto.
(3) The function $x \to x^3 + x^2$ is onto, but not one-to-one.
(4) The sine function $x \to \sin x$ is neither one-to-one nor onto.

If a function $f: D \to R$ is both one-to-one and onto, then for each element y of R there is one and only one element x such that $f(x) = y$. By defining $f^{-1}(y) = x$ for all x and y so related, we obtain a function $f^{-1}: R \to D$ called the *inverse* of f. Note that the function f^{-1} is also one-to-one and onto, and that *its* inverse function is the original function f.

Here is a short list of the main notations used throughout the book, in order of their appearance in Chapter 1:

$\mathbf{p}, \mathbf{q}$	points	(Section 1.1)
f, g	real-valued functions	(Section 1.1)
$\mathbf{v}, \mathbf{w}$	tangent vectors	(Section 1.2)
V, W	vector fields	(Section 1.2)
α, β	curves	(Section 1.4)
ϕ, ψ	differential forms	(Section 1.5)
F, G	mappings	(Section 1.7)

In Chapter 1 we define these concepts for Euclidean 3-space. (Extension to arbitrary dimensions is virtually automatic.) In Chapter 4 we show how these concepts can be adapted to a surface.

A few references are given to the brief bibliography at the end of the book; these are indicated by initials in square brackets.

Chapter 1
Calculus on Euclidean Space

As mentioned in the Preface, the purpose of this initial chapter is to establish the mathematical language used throughout the book. Much of what we do is simply a review of that part of elementary calculus dealing with differentiation of functions of three variables and with curves in space. Our definitions have been formulated so that they will apply smoothly to the later study of surfaces.

1.1 Euclidean Space

Three-dimensional space is often used in mathematics without being formally defined. Looking at the corner of a room, one can picture the familiar process by which rectangular coordinate axes are introduced and three numbers are measured to describe the position of each point. A precise definition that realizes this intuitive picture may be obtained by this device: instead of saying that three numbers *describe the position* of a point, we define them to *be* a point.

1.1 Definition *Euclidean 3-space* $\mathbf{R}^3$ is the set of all ordered triples of real numbers. Such a triple $\mathbf{p} = (p_1, p_2, p_3)$ is called a *point* of $\mathbf{R}^3$.

In linear algebra, it is shown that $\mathbf{R}^3$ is, in a natural way, a vector space over the real numbers. In fact, if $\mathbf{p} = (p_1, p_2, p_3)$ and $\mathbf{q} = (q_1, q_2, q_3)$ are points of $\mathbf{R}^3$, their *sum* is the point

$$\mathbf{p} + \mathbf{q} = (p_1 + q_1, p_2 + q_2, p_3 + q_3).$$

The *scalar multiple* of a point $\mathbf{p} = (p_1, p_2, p_3)$ by a number a is the point

$$a\mathbf{p} = (ap_1, ap_2, ap_3).$$

It is easy to check that these two operations satisfy the axioms for a vector space. The point $\mathbf{0} = (0, 0, 0)$ is called the *origin* of $\mathbf{R}^3$.

Differential calculus deals with another aspect of $\mathbf{R}^3$ starting with the notion of differentiable real-valued functions on $\mathbf{R}^3$. We recall some fundamentals.

1.2 Definition Let x, y, and z be the real-valued functions on $\mathbf{R}^3$ such that for each point $\mathbf{p} = (p_1, p_2, p_3)$

$$x(\mathbf{p}) = p_1, \quad y(\mathbf{p}) = p_2, \quad z(\mathbf{p}) = p_3.$$

These functions x, y, z are called the *natural coordinate functions* of $\mathbf{R}^3$. We shall also use index notation for these functions, writing

$$x_1 = x, \quad x_2 = y, \quad x_3 = z.$$

Thus the value of the function x_i on a point $\mathbf{p}$ is the number p_i, and so we have the identity $\mathbf{p} = (p_1, p_2, p_3) = (x_1(\mathbf{p}), x_2(\mathbf{p}), x_3(\mathbf{p}))$ for each point $\mathbf{p}$ of $\mathbf{R}^3$. Elementary calculus does not always make a sharp distinction between the *numbers* p_1, p_2, p_3 and the *functions* x_1, x_2, x_3. Indeed the analogous distinction on the real line may seem pedantic, but for higher-dimensional spaces such as $\mathbf{R}^3$, its absence leads to serious ambiguities. (Essentially the same distinction is being made when we denote a function on $\mathbf{R}^3$ by a single letter f, reserving $f(\mathbf{p})$ for its value at the point $\mathbf{p}$.)

We assume that the reader is familiar with partial differentiation and its basic properties, in particular the chain rule for differentiation of a composite function. We shall work mostly with first-order partial derivatives $\partial f/\partial x$, $\partial f/\partial y$, $\partial f/\partial z$ and second-order partial derivatives $\partial^2 f/\partial x^2$, $\partial^2 f/\partial x \partial y, \ldots$ In a few situations, third- and even fourth-order derivatives may occur, but to avoid worrying about exactly how many derivatives we can take in any given context, we establish the following definition.

1.3 Definition A real-valued function f on $\mathbf{R}^3$ is *differentiable* (or *infinitely differentiable*, or *smooth*, or *of class* C^∞) provided all partial derivatives of f, of all orders, exist and are continuous.

Differentiable real-valued functions f and g may be added and multiplied in a familiar way to yield functions that are again differentiable and real-valued. We simply add and multiply their values at each point—the formulas read

$$(f + g)(\mathbf{p}) = f(\mathbf{p}) + g(\mathbf{p}), \quad (fg)(\mathbf{p}) = f(\mathbf{p})g(\mathbf{p}).$$

The phrase "differentiable real-valued function" is unpleasantly long. Hence we make the convention that *unless the context indicates otherwise*, "function" shall mean "real-valued function," and (unless the issue is explicitly raised) the functions we deal with will be assumed to be differentiable. We do not intend to overwork this convention; for the sake of emphasis the words "differentiable" and "real-valued" will still appear fairly frequently.

Differentiation is always a *local* operation: To compute the value of the function $\partial f/\partial x$ at a point $\mathbf{p}$ of $\mathbf{R}^3$, it is sufficient to know the values of f at all points $\mathbf{q}$ of $\mathbf{R}^3$ that are sufficiently near $\mathbf{p}$. Thus, Definition 1.3 is unduly restrictive; the domain of f need not be the whole of $\mathbf{R}^3$, but need only be an *open set* of $\mathbf{R}^3$. By an *open set* $\mathscr{O}$ of $\mathbf{R}^3$ we mean a subset of $\mathbf{R}^3$ such that if a point $\mathbf{p}$ is in $\mathscr{O}$, then so is every other point of $\mathbf{R}^3$ that is sufficiently near $\mathbf{p}$. (A more precise definition is given in Chapter 2.) For example, the set of all points $\mathbf{p} = (p_1, p_2, p_3)$ in $\mathbf{R}^3$ such that $p_1 > 0$ is an open set, and the function $yz \log x$ defined on this set is certainly differentiable, even though its domain is not the whole of $\mathbf{R}^3$. Generally speaking, the results in this chapter remain valid if $\mathbf{R}^3$ is replaced by an arbitrary open set $\mathscr{O}$ of $\mathbf{R}^3$.

We are dealing with *three-dimensional* Euclidean space only because this is the dimension we use most often in later work. It would be just as easy to work with *Euclidean n-space* $\mathbf{R}^n$, for which the points are n-tuples $\mathbf{p} = (p_1, \ldots, p_n)$ and which has n natural coordinate functions $x_1, \ldots, x_n$. All the results in this chapter are valid for Euclidean spaces of arbitrary dimensions, although we shall rarely take advantage of this except in the case of the *Euclidean plane* $\mathbf{R}^2$. In particular, the results are valid for the *real line* $\mathbf{R}^1 = \mathbf{R}$. Many of the concepts introduced are designed to deal with higher dimensions, however, and are thus apt to be overelaborate when reduced to dimension 1.

Exercises

1. Let $f = x^2y$ and $g = y \sin z$ be functions on $\mathbf{R}^3$. Express the following functions in terms of x, y, z:

(a) fg^2. (b) $\dfrac{\partial f}{\partial x} g + \dfrac{\partial g}{\partial y} f$.

(c) $\dfrac{\partial^2 (fg)}{\partial y \partial z}$. (d) $\dfrac{\partial}{\partial y}(\sin f)$.

2. Find the value of the function $f = x^2y - y^2z$ at each point:
(a) (1, 1, 1). (b) (3, −1, ½).
(c) $(a, 1, 1 - a)$. (d) (t, t^2, t^3).

3. Express $\partial f/\partial x$ in terms of x, y, and z if
(a) $f = x\sin(xy) + y\cos(xz)$.
(b) $f = \sin g$, $g = e^h$, $h = x^2 + y^2 + z^2$.

4. If g_1, g_2, g_3, and h are real-valued functions on $\mathbf{R}^3$, then

$$f = h(g_1, g_2, g_3)$$

is the function such that

$$f(\mathbf{p}) = h(g_1(\mathbf{p}), g_2(\mathbf{p}), g_3(\mathbf{p})) \quad \text{for all } \mathbf{p}.\dagger$$

Express $\partial f/\partial x$ in terms of x, y, and z, if $h = x^2 - yz$ and
(a) $f = h(x + y, y^2, x + z)$. (b) $f = h(e^z, e^{x+y}, e^x)$.
(c) $f = h(x, -x, x)$.

1.2 Tangent Vectors

Intuitively, a vector in $\mathbf{R}^3$ is an oriented line segment, or "arrow." Vectors are used widely in physics and engineering to describe forces, velocities, angular momenta, and many other concepts. To obtain a definition that is both practical and precise, we shall describe an "arrow" in $\mathbf{R}^3$ by giving its starting point $\mathbf{p}$ and the change, or vector $\mathbf{v}$, necessary to reach its end point $\mathbf{p} + \mathbf{v}$. Strictly speaking, $\mathbf{v}$ is just a point of $\mathbf{R}^3$.

2.1 Definition‡ A *tangent vector* $\mathbf{v}_p$ to $\mathbf{R}^3$ consists of two points of $\mathbf{R}^3$: its *vector part* $\mathbf{v}$ and its *point of application* $\mathbf{p}$.

We shall always picture $\mathbf{v}_p$ as *the arrow from the point* $\mathbf{p}$ to the point $\mathbf{p} + \mathbf{v}$. For example, if $\mathbf{p} = (1, 1, 3)$ and $\mathbf{v} = (2, 3, 2)$, then $\mathbf{v}_p$ runs from (1, 1, 3) to (3, 4, 5) as in Fig. 1.1.

We emphasize that tangent vectors are equal, $\mathbf{v}_p = \mathbf{w}_q$, if and only if they have the same vector part, $\mathbf{v} = \mathbf{w}$, and the same point of application, $\mathbf{p} = \mathbf{q}$.

† A consequence is the identity $f = f(x, y, z)$.

‡ The term "tangent" in this definition will acquire a more direct geometric meaning in Chapter 4.

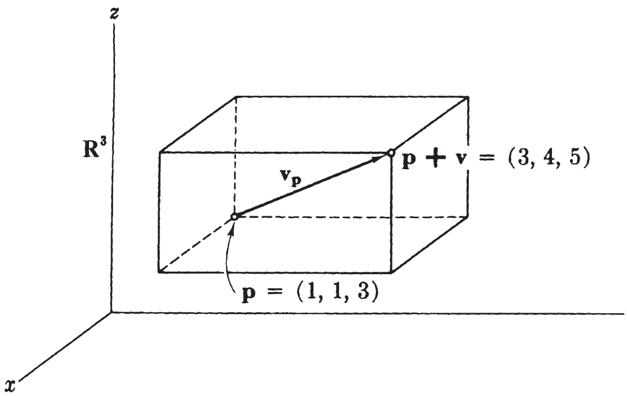


FIG. 1.1

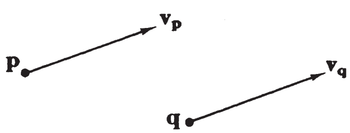


FIG. 1.2

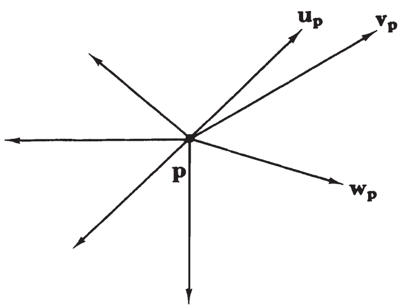


FIG. 1.3

Tangent vectors $\mathbf{v}_p$ and $\mathbf{v}_q$ with the same vector part, but different points of application, are said to be *parallel* (Fig. 1.2). It is essential to recognize that $\mathbf{v}_p$ and $\mathbf{v}_q$ are different tangent vectors if $\mathbf{p} \neq \mathbf{q}$. In physics the concept of moment of a force shows this clearly enough: The same force $\mathbf{v}$ applied at different points $\mathbf{p}$ and $\mathbf{q}$ of a rigid body can produce quite different rotational effects.

2.2 Definition Let $\mathbf{p}$ be a point of $\mathbf{R}^3$. The set $T_p(\mathbf{R}^3)$ consisting of all tangent vectors that have $\mathbf{p}$ as point of application is called the *tangent space* of $\mathbf{R}^3$ at $\mathbf{p}$ (Fig. 1.3).

We emphasize that $\mathbf{R}^3$ has a different tangent space at each and every one of its points.

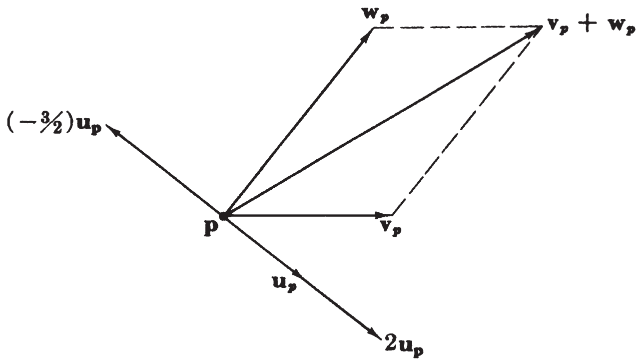


FIG. 1.4

Since all the tangent vectors in a given tangent space have the same point of application, we can borrow the vector addition and scalar multiplication of $\mathbf{R}^3$ to turn $T_p(\mathbf{R}^3)$ into a vector space. Explicitly, we define $\mathbf{v}_p + \mathbf{w}_p$ to be $(\mathbf{v} + \mathbf{w})_p$. and if c is a number we define $c(\mathbf{v}_p)$ to be $(c\mathbf{v})_p$. This is just the usual "parallelogram law" for addition of vectors, and scalar multiplication by c merely stretches a tangent vector by the factor c—reversing its direction if $c < 0$ (Fig. 1.4).

These operations on the tangent space $T_p(\mathbf{R}^3)$ make it a vector space isomorphic to $\mathbf{R}^3$ itself. Indeed, it follows immediately from the definitions above that for a fixed point $\mathbf{p}$, the function $\mathbf{v} \to \mathbf{v}_p$ is a linear isomorphism from $\mathbf{R}^3$ to $T_p(\mathbf{R}^3)$—that is, a linear transformation that is one-to-one and onto.

A standard concept in physics and engineering is that of a force field. The gravitational force field of the earth, for example, assigns to each point of space a force (vector) directed at the center of the earth.

2.3 Definition A *vector field* V on $\mathbf{R}^3$ is a function that assigns to each point $\mathbf{p}$ of $\mathbf{R}^3$ a tangent vector $V(\mathbf{p})$ to $\mathbf{R}^3$ at $\mathbf{p}$.

Roughly speaking, a vector field is just a big collection of arrows, one at each point of $\mathbf{R}^3$.

There is a natural algebra of vector fields. To describe it, we first reexamine the familiar notion of addition of real-valued functions f and g. It is possible to add f and g because it is possible to add their values at each point. The same is true of vector fields V and W. At each point $\mathbf{p}$, the values $V(\mathbf{p})$ and $W(\mathbf{p})$ are in the same vector space—the tangent space $T_p(\mathbf{R}^3)$—hence we can add $V(\mathbf{p})$ and $W(\mathbf{p})$. Consequently, we can add V and W by adding their values at each point. The formula for this addition is thus the same as for addition of functions,

$$(V + W)(\mathbf{p}) = V(\mathbf{p}) + W(\mathbf{p}).$$

This scheme occurs over and over again. We shall call it the *pointwise principle:* If a certain operation can be performed on the values of two functions at each point, then that operation can be extended to the functions themselves; simply apply it to their values at each point.

For example, we invoke the pointwise principle to extend the operation of *scalar multiplication* (on the tangent spaces of $\mathbf{R}^3$). If f is a real-valued function on $\mathbf{R}^3$ and V is a vector field on $\mathbf{R}^3$, then fV is defined to be the vector field on $\mathbf{R}^3$ such that

$$(fV)(\mathbf{p}) = f(\mathbf{p})V(\mathbf{p}) \quad \text{for all } \mathbf{p}.$$

Our aim now is to determine in a concrete way just what vector fields look like. For this purpose we introduce three special vector fields that will serve as a "basis" for all vector fields.

2.4 Definition Let U_1, U_2, and U_3 be the vector fields on $\mathbf{R}^3$ such that

$$U_1(\mathbf{p}) = (1, 0, 0)_p$$
$$U_2(\mathbf{p}) = (0, 1, 0)_p$$
$$U_3(\mathbf{p}) = (0, 0, 1)_p$$

for each point $\mathbf{p}$ of $\mathbf{R}^3$ (Fig. 1.5). We call U_1, U_2, U_3—collectively—the *natural frame field* on $\mathbf{R}^3$.

Thus, U_i ($i = 1, 2, 3$) is the unit vector field in the positive x_i direction.

2.5 Lemma If V is a vector field on $\mathbf{R}^3$, there are three uniquely determined real-valued functions, v_1, v_2, v_3 on $\mathbf{R}^3$ such that

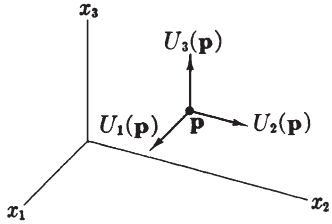


FIG. 1.5

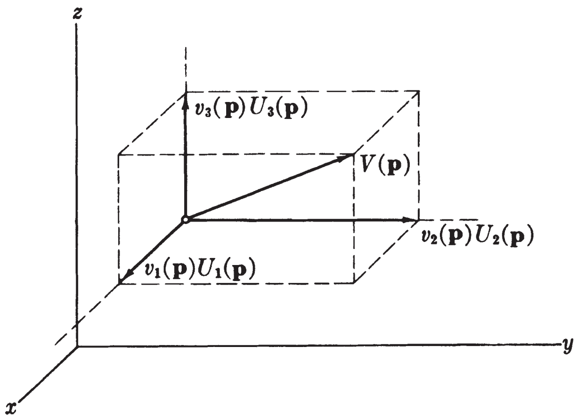


FIG. 1.6

$$V = v_1U_1 + v_2U_2 + v_3U_3.$$

The functions v_1, v_2, v_3 are called the *Euclidean coordinate functions of V.*

Proof. By definition, the vector field V assigns to each point $\mathbf{p}$ a tangent vector $V(\mathbf{p})$ at $\mathbf{p}$. Thus, the vector part of $V(\mathbf{p})$ depends on $\mathbf{p}$, so we write it $(v_1(\mathbf{p}), v_2(\mathbf{p}), v_3(\mathbf{p}))$. (This defines v_1, v_2, and v_3 as real-valued *functions* on $\mathbf{R}^3$.) Hence

$$\begin{aligned} V(\mathbf{p}) &= (v_1(\mathbf{p}), v_2(\mathbf{p}), v_3(\mathbf{p}))_p \\ &= v_1(\mathbf{p})(1,0,0)_p + v_2(\mathbf{p})(0,1,0)_p + v_3(\mathbf{p})(0,0,1)_p \\ &= v_1(\mathbf{p})U_1(\mathbf{p}) + v_2(\mathbf{p})U_2(\mathbf{p}) + v_3(\mathbf{p})U_3(\mathbf{p}) \end{aligned}$$

for each point $\mathbf{p}$ (Fig. 1.6). By our (pointwise principle) definitions, this means that the vector fields V and $\sum v_iU_i$ have the same (tangent vector) value at each point. Hence $V = \sum v_iU_i$. ◆

This last sentence uses two of our standard conventions: $\sum v_iU_i$ means sum over $i = 1, 2, 3$; the symbol (◆) indicates the end of a proof.

The tangent-vector identity $(a_1, a_2, a_3)_p = \sum a_iU_i(\mathbf{p})$ appearing in this proof will be used very often.

Computations involving vector fields may always be expressed in terms of their Euclidean coordinate functions. For example, addition and multiplication by a function, are expressed in terms of coordinates by

$$\sum v_iU_i + \sum w_iU_i = \sum (v_i + w_i)U_i,$$

$$f\left(\sum v_iU_i\right) = \sum (fv_i)U_i.$$

Since this is differential calculus, we shall naturally require that the various objects we deal with be differentiable. A vector field V is *differentiable* provided its *E*uclidean coordinate functions are differentiable (in the sense of Definition 1.3). From now on, we shall understand "vector field" to mean "differentiable vector field."

Exercises

1. Let $\mathbf{v} = (-2, 1, -1)$ and $\mathbf{w} = (0, 1, 3)$.
(a) At an arbitrary point $\mathbf{p}$, express the tangent vector $3\mathbf{v}_p - 2\mathbf{w}_p$ as a linear combination of $U_1(\mathbf{p})$, $U_2(\mathbf{p})$, $U_3(\mathbf{p})$.
(b) For $\mathbf{p} = (1, 1, 0)$, make an accurate sketch showing the four tangent vectors $\mathbf{v}_p$, $\mathbf{w}_p$, $-2\mathbf{v}_p$, and $\mathbf{v}_p + \mathbf{w}_p$.

2. Let $V = xU_1 + yU_2$ and $W = 2x^2U_2 - U_3$. Compute the vector field $W - xV$, and find its value at the point $\mathbf{p} = (-1, 0, 2)$.

3. In each case, express the given vector field V in the standard form $\sum v_iU_i$.
(a) $2z^2U_1 = 7V + xyU_3$.
(b) $V(\mathbf{p}) = (p_1, p_3 - p_1, 0)_p$ for all $\mathbf{p}$.
(c) $V = 2(xU_1 + yU_2) - x(U_1 - y^2U_3)$.
(d) At each point $\mathbf{p}$, $V(\mathbf{p})$ is the vector from the point (p_1, p_2, p_3) to the point $(1 + p_1, p_2p_3, p_2)$.
(e) At each point $\mathbf{p}$, $V(\mathbf{p})$ is the vector from $\mathbf{p}$ to the origin.

4. If $V = y^2U_1 - x^2U_3$ and $W = x^2U_1 - zU_2$, find functions f and g such that the vector field $fV + gW$ can be expressed in terms of U_2 and U_3 only.

5. Let $V_1 = U_1 - xU_3$, $V_2 = U_2$, and $V_3 = xU_1 + U_3$.
(a) Prove that the vectors $V_1(\mathbf{p})$, $V_2(\mathbf{p})$, $V_3(\mathbf{p})$ are linearly independent at each point of $\mathbf{R}^3$.
(b) Express the vector field $xU_1 + yU_2 + zU_3$ as a linear combination of V_1, V_2, V_3.

1.3 Directional Derivatives

Associated with each tangent vector $\mathbf{v}_p$ to $\mathbf{R}^3$ is the straight line $t \to \mathbf{p} + t\mathbf{v}$ (see Example 4.2). If f is a differentiable function on $\mathbf{R}^3$, then $t \to f(\mathbf{p} + t\mathbf{v})$ is an ordinary differentiable function on the real line. Evidently the derivative of this function at $t = 0$ tells the initial rate of change of f as $\mathbf{p}$ moves in the $\mathbf{v}$ direction

3.1 Definition Let f be a differentiable real-valued function on $\mathbf{R}^3$, and let $\mathbf{v}_p$ be a tangent vector to $\mathbf{R}^3$. Then the number

$$\mathbf{v}_p[f] = \frac{d}{dt}(f(\mathbf{p} + t\mathbf{v}))|_{t=0}$$

is called the *derivative of f with respect* to $\mathbf{v}_p$.

This definition appears in elementary calculus with the additional restriction that $\mathbf{v}_p$ be a unit vector. Even though we do not impose this restriction, we shall nevertheless refer to $\mathbf{v}_p[f]$ as a *directional derivative.*

For example, we compute $\mathbf{v}_p[f]$ for the function $f = x^2yz$, with $\mathbf{p} = (1, 1, 0)$ and $\mathbf{v} = (1, 0, -3)$. Then

$$\mathbf{p} + t\mathbf{v} = (1, 1, 0) + t(1, 0, -3) = (1 + t, 1, -3t)$$

describes the line through $\mathbf{p}$ in the $\mathbf{v}$ direction. Evaluating f along this line, we get

$$f(\mathbf{p} + t\mathbf{v}) = (1 + t)^2 \cdot 1 \cdot (-3t) = -3t - 6t^2 - 3t^3.$$

Now,

$$\frac{d}{dt}(f(\mathbf{p} + t\mathbf{v})) = -3 - 12t - 9t^2;$$

hence at $t = 0$, we find $\mathbf{v}_p[f] = -3$. Thus, in particular, the function f is initially decreasing as $\mathbf{p}$ moves in the $\mathbf{v}$ direction.

The following lemma shows how to compute $\mathbf{v}_p[f]$ in general, in terms of the partial derivatives of f at the point $\mathbf{p}$.

3.2 Lemma If $\mathbf{v}_p = (v_1, v_2, v_3)_p$ is a tangent vector to $\mathbf{R}^3$, then

$$\mathbf{v}_p[f] = \sum v_i \frac{\partial f}{\partial x_i}(\mathbf{p}).$$

Proof. Let $\mathbf{p} = (p_1, p_2, p_3)$; then

$$\mathbf{p} + t\mathbf{v} = (p_1 + tv_1, p_2 + tv_2, p_3 + tv_3).$$

We use the chain rule to compute the derivative at $t = 0$ of the function

$$f(\mathbf{p} + t\mathbf{v}) = f(p_1 + tv_1, p_2 + tv_2, p_3 + tv_3).$$

Since

$$\frac{d}{dt}(p_i + tv_i) = v_i,$$

we obtain

$$\mathbf{v}_p[f] = \frac{d}{dt}(f(\mathbf{p}+t\mathbf{v}))|_{t=0} = \sum \frac{\partial f}{\partial x_i}(\mathbf{p})v_i. \qquad \blacklozenge$$

Using this lemma, we recompute $\mathbf{v}_p[f]$ for the example above. Since $f = x^2yz$, we have

$$\frac{\partial f}{\partial x} = 2xyz, \quad \frac{\partial f}{\partial y} = x^2z, \quad \frac{\partial f}{\partial z} = x^2y.$$

Thus, at the point $\mathbf{p} = (1, 1, 0)$,

$$\frac{\partial f}{\partial x}(\mathbf{p}) = 0, \quad \frac{\partial f}{\partial y}(\mathbf{p}) = 0, \quad \text{and} \quad \frac{\partial f}{\partial z}(\mathbf{p}) = 1.$$

Then by the lemma,

$$\mathbf{v}_p[f] = 0 + 0 + (-3)1 = -3,$$

as before.

The main properties of this notion of derivative are as follows.

3.3 Theorem Let f and g be functions on $\mathbf{R}^3$, $\mathbf{v}_p$ and $\mathbf{w}_p$ tangent vectors, a and b numbers. Then

(1) $(a\mathbf{v}_p + b\mathbf{w}_p)[f] = a\mathbf{v}_p[f] + b\mathbf{w}_p[f]$.
(2) $\mathbf{v}_p[af + bg] = a\mathbf{v}_p[f] + b\mathbf{v}_p[g]$.
(3) $\mathbf{v}_p[fg] = \mathbf{v}_p[f]\cdot g(\mathbf{p}) + f(\mathbf{p})\cdot \mathbf{v}_p[g]$.

Proof. All three properties may be deduced easily from the preceding lemma. For example, we prove (3). By the lemma, if $\mathbf{v} = (v_1, v_2, v_3)$, then

$$\mathbf{v}_p[fg] = \sum v_i \frac{\partial (fg)}{\partial x_i}(\mathbf{p}).$$

But

$$\frac{\partial (fg)}{\partial x_i} = \frac{\partial f}{\partial x_i}\cdot g + f\cdot \frac{\partial g}{\partial x_i}.$$

Hence

$$\begin{aligned}\mathbf{v}_p[fg] &= \sum v_i\left(\frac{\partial f}{\partial x_i}(\mathbf{p})\cdot g(\mathbf{p}) + f(\mathbf{p})\cdot\frac{\partial g}{\partial x_i}(\mathbf{p})\right)\\ &= \left(\sum v_i \frac{\partial f}{\partial x_i}(\mathbf{p})\right)g(\mathbf{p}) + f(\mathbf{p})\left(\sum v_i \frac{\partial g}{\partial x_i}(\mathbf{p})\right)\\ &= \mathbf{v}_p[f]\cdot g(\mathbf{p}) + f(\mathbf{p})\cdot \mathbf{v}_p[g]. \qquad \blacklozenge\end{aligned}$$

The first two properties in the preceding theorem may be summarized by saying that $\mathbf{v}_p[f]$ is *linear* in $\mathbf{v}_p$ and in f. The third property, as its proof makes clear, is essentially just the usual Leibniz rule for differentiation of a product. *No matter what form differentiation may take, it will always have suitable* linear *and* Leibnizian *properties.*

We now use the pointwise principle to define the *operation of a vector field V on a function f.* The result is the real-valued function $V[f]$ whose value at each point $\mathbf{p}$ is the number $V(\mathbf{p})[f]$, that is, the derivative of f with respect to the tangent vector $V(\mathbf{p})$ at $\mathbf{p}$. This process should be no surprise, since for a function f on the real line, one begins by defining the derivative of f *at a point*—then the derivative *function* df/dx is the function whose value at each point is the derivative at that point. Evidently, the definition of $V[f]$ is strictly analogous. In particular, if U_1, U_2, U_3 is the natural frame field on $\mathbf{R}^3$, then $U_i[f] = \partial f/\partial x_i$. This is an immediate consequence of Lemma 3.2. For example, $U_1(\mathbf{p}) = (1, 0, 0)_p$; hence

$$U_1(\mathbf{p})[f] = \frac{d}{dt}(f(p_1 + t,\, p_2,\, p_3))\Big|_{t=0},$$

which is precisely the definition of $(\partial f/\partial x_1)(\mathbf{p})$. This is true for all points $\mathbf{p} = (p_1, p_2, p_3)$; hence $U_1[f] = \partial f/\partial x_1$.

We shall use this notion of directional derivative more in the case of vector fields than for individual tangent vectors.

3.4 Corollary If V and W are vector fields on $\mathbf{R}^3$ and f, g, h are real-valued functions, then

(1) $(fV + gW)[h] = fV[h] + gW[h]$.
(2) $V[af + bg] = aV[f] + bV[g]$, for all real numbers a and b.
(3) $V[fg] = V[f]\cdot g + f\cdot V[g]$.

Proof. The pointwise principle guarantees that to derive these properties from Theorem 3.3 we need only be careful about the placement of parentheses. For example, we prove the third formula. By definition, the value of the function $V[fg]$ at $\mathbf{p}$ is $V(\mathbf{p})[fg]$. But by Theorem 3.3 this is

$$\begin{aligned} V(\mathbf{p})[f]\cdot g(\mathbf{p}) + f(\mathbf{p})\cdot V(\mathbf{p})[g] &= V[f](\mathbf{p})\cdot g(\mathbf{p}) + f(\mathbf{p})\cdot V[g](\mathbf{p}) \\ &= (V[f]\cdot g + f\cdot V[g])(\mathbf{p}). \end{aligned}$$ ◆

If the use of parentheses here seems extravagant, we remind the reader that a meticulous proof of Leibniz's formula

$$\frac{d}{dx}(fg) = \frac{df}{dx}\cdot g + f\cdot\frac{dg}{dx}$$

must involve the same shifting of parentheses.

Note that the linearity of $V[f]$ in V and f is for *functions* as "scalars" in the first formula in Corollary 3.4 but only for *numbers* as "scalars" in the second. This stems from the fact that fV signifies merely multiplication, but $V[f]$ is differentiation.

The identity $U_i[f] = \partial f/\partial x_i$ makes it a simple matter to carry out explicit computations. For example, if $V = xU_1 - y^2U_3$ and $f = x^2y + z^3$, then

$$\begin{aligned} V[f] &= xU_1[x^2y] + xU_1[z^3] - y^2U_3[x^2y] - y^2U_3[z^3] \\ &= x(2xy) + 0 - 0 - y^2(3z^2) = 2x^2y - 3y^2z^2. \end{aligned}$$

3.5 Remark Since the subscript notation $\mathbf{v}_p$ for a tangent vector is somewhat cumbersome, from now on we shall frequently omit the point of application $\mathbf{p}$ from the notation. This can cause no confusion, since $\mathbf{v}$ and $\mathbf{w}$ will always denote tangent vectors, and $\mathbf{p}$ and $\mathbf{q}$ points of $\mathbf{R}^3$. In many situations (for example, Definition 3.1) the point of application is crucial, and will be indicated by using either the old notation $\mathbf{v}_p$ or the phrase "a tangent vector $\mathbf{v}$ to $\mathbf{R}^3$ at $\mathbf{p}$."

Exercises

1. Let $\mathbf{v}_p$ be the tangent vector to $\mathbf{R}^3$ with $\mathbf{v} = (2, -1, 3)$ and $\mathbf{p} = (2, 0, -1)$. Working directly from the definition, compute the directional derivative $\mathbf{v}_p[f]$, where

(a) $f = y^2z$. (b) $f = x^7$.
(c) $f = e^x \cos y$.

2. Compute the derivatives in Exercise 1 using Lemma 3.2.

3. Let $V = y^2U_1 - xU_3$, and let $f = xy$, $g = z^3$. Compute the functions

(a) $V[f]$. (b) $V[g]$.
(c) $V[fg]$. (d) $fV[g] - gV[f]$.
(e) $V[f^2 + g^2]$. (f) $V[V[f]]$.

4. Prove the identity $V = \sum V[x_i]U_i$, where x_1, x_2, x_3 are the natural coordinate functions. (*Hint:* Evaluate $V = \sum v_iU_i$ on x_j.)

5. If $V[f] = W[f]$ for every function f on $\mathbf{R}^3$, prove that $V = W$.

1.4 Curves in $\mathbf{R}^3$

Let I be an open interval in the real line R. We shall interpret this liberally to include not only the usual finite open interval $a < t < b$ (a, b real numbers), but also the infinite types $a < t$ (a half-line to $+\infty$), $t < b$ (a half-line to $-\infty$), and also the whole real line.

One can picture a curve in $\mathbf{R}^3$ as a trip taken by a moving point α. At each "time" t in some open interval, α is located at the point

$$\alpha(t) = (\alpha_1(t), \alpha_2(t), \alpha_3(t))$$

in $\mathbf{R}^3$. In rigorous terms then, α is a function from I to $\mathbf{R}^3$, and the real-valued functions α_1, α_2, α_3 are its *Euclidean coordinate functions.* Thus we write $\alpha = (\alpha_1, \alpha, \alpha_3)$, meaning, of course, that

$$\alpha(t) = (\alpha_1(t), \alpha_2(t), \alpha_3(t)) \quad \text{for all } t \text{ in } I.$$

We define the function α to be *differentiable* provided its (real-valued) coordinate functions are differentiable in the usual sense.

4.1 Definition A *curve* in $\mathbf{R}^3$ is a differentiable function $\alpha\colon I \to \mathbf{R}^3$ from an open interval I into $\mathbf{R}^3$.

We shall give several examples of curves, which will be used in Chapter 2 to experiment with results on the geometry of curves.

4.2 Example (1) ***Straight line.*** A line is the simplest type of curve in Euclidean space; its coordinate functions are linear (in the sense $t \to at + b$, not in the homogeneous sense $t \to at$). Explicitly, the curve $\alpha\colon \mathbf{R} \to \mathbf{R}^3$ such that

$$\alpha(t) = \mathbf{p} + t\mathbf{q} = (p_1 + tq_1,\ p_2 + tq_2,\ p_3 + tq_3) \quad (\mathbf{q} \neq 0)$$

is the *straight line* through the point $\mathbf{p} = \alpha(0)$ in the $\mathbf{q}$ direction.

(2) ***Helix.*** (Fig. 1.7). The curve $t \to (a\cos t, a\sin t, 0)$ travels around a circle of radius $a > 0$ in the xy plane of $\mathbf{R}^3$. If we allow this curve to rise (or fall) at a constant rate, we obtain a *helix* $\alpha\colon \mathbf{R} \to \mathbf{R}^3$, given by the formula

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

where $a > 0$, $b \neq 0$.

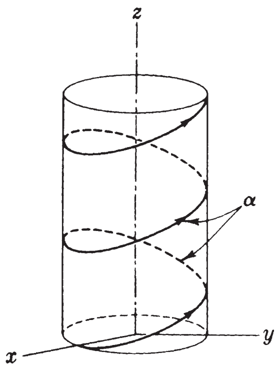


FIG. 1.7

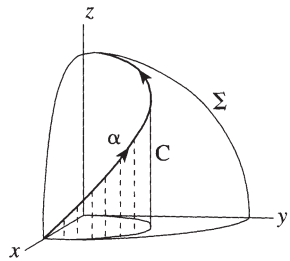


FIG. 1.8

(3) The curve

$$\alpha(t) = \left(1 + \cos t, \sin t, 2 \sin\frac{t}{2}\right) \quad \text{for all } t$$

has a noteworthy property: Let C be the cylinder in $\mathbf{R}^3$ over the circle in the xy plane with center at (1, 0, 0) and radius 1. Then α perpetually travels the route sliced from C by the sphere Σ with radius 2 and center at the origin. A segment of this route is shown in Fig. 1.8.

(4) The curve α: $\mathbf{R} \to \mathbf{R}^3$ such that

$$\alpha(t) = \left(e^t, e^{-t}, \sqrt{2}t\right)$$

shares with the helix in (2) the property of rising constantly. However, it lies over the hyperbola $xy = 1$ in the xy plane instead of over a circle.

(5) The 3-*curve* α: $\mathbf{R} \to \mathbf{R}^3$ is defined by

$$\alpha(t) = (3t - t^3, 3t^2, 3t + t^3).$$

If the coordinate functions of a curve are simple enough, its shape in $\mathbf{R}^3$ can be found, at least approximately, by plotting a few points. We could get a reasonable picture of curve α for $0 \leqq t \leqq 1$ by computing $\alpha(t)$ for $t = 0, 1/10, 1/2, 9/10, 1$.

If we visualize a curve α in $\mathbf{R}^3$ as a moving point, then at every time t there is a tangent vector at the point $\alpha(t)$ that gives the instantaneous velocity of α at that time. ◆

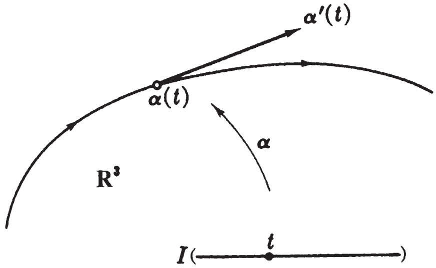


FIG. 1.9

4.3 Definition Let $\alpha\colon I \to \mathbf{R}^3$ be a curve in $\mathbf{R}^3$ with $\alpha = (\alpha_1, \alpha_2, \alpha_3)$. For each number t in I, the *velocity vector of* α *at* t is the tangent vector

$$\alpha'(t) = \left(\frac{d\alpha_1}{dt}(t), \frac{d\alpha_2}{dt}(t), \frac{d\alpha_3}{dt}(t)\right)_{\alpha(t)}$$

at the point $\alpha(t)$ in $\mathbf{R}^3$ (Fig. 1.9).

This definition can be interpreted geometrically as follows. The derivative at t of a real-valued function f on $\mathbf{R}$ is given by

$$\frac{df}{dt}(t) = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}.$$

This formula still makes sense if f is replaced by a curve $\alpha = (\alpha_1, \alpha_2, \alpha_3)$. In fact,

$$\frac{1}{\Delta t}(\alpha(t + \Delta t) - \alpha(t)) = \left(\frac{\alpha_1(t + \Delta t) - \alpha_1(t)}{\Delta t}, \frac{\alpha_2(t + \Delta t) - \alpha_2(t)}{\Delta t}, \frac{\alpha_3(t + \Delta t) - \alpha_3(t)}{\Delta t}\right).$$

This is the vector from $\alpha(t)$ to $\alpha(t + \Delta t)$, scalar multiplied by $1/\Delta t$ (Fig. 1.10).

Now, as Δt gets smaller, $\alpha(t + \Delta t)$ approaches $\alpha(t)$, and in the limit as $\Delta t \to 0$, we get a vector *tangent* to the curve α at the point $\alpha(t)$, namely,

$$\left(\frac{d\alpha_1}{dt}(t), \frac{d\alpha_2}{dt}(t), \frac{d\alpha_3}{dt}(t)\right).$$

As the figure suggests, the point of application of this vector must be the point $\alpha(t)$. Thus the standard limit operation for derivatives gives rise to our definition of the velocity of a curve.

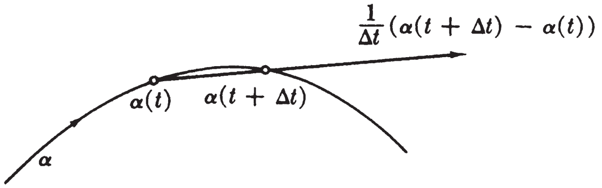


FIG. 1.10

An application of the identity

$$(v_1, v_2, v_3)_p = \sum v_i U_i(\mathbf{p})$$

to the velocity vector $\alpha'(t)$ at t yields the alternative formula

$$\alpha'(t) = \sum \frac{d\alpha_i}{dt}(t) U_i(\alpha(t)).$$

For example, the velocity of the straight line $\alpha(t) = \mathbf{p} + t\mathbf{q}$ is

$$\alpha'(t) = (q_1, q_2, q_3)_{\alpha(t)} = \mathbf{q}_{\alpha(t)}.$$

The fact that α is straight is reflected in the fact that all its velocity vectors are parallel; only the point of application changes as t changes.

For the helix

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

the velocity is

$$\alpha'(t) = (-a \sin t, a \cos t, b)_{\alpha(t)}.$$

The fact that the helix rises constantly is shown by the constancy of the z coordinate of $\alpha'(t)$.

Given any curve, it is easy to construct new curves that follow the same route.

4.4 Definition Let $\alpha: I \to \mathbf{R}^3$ be a curve. If $h: J \to I$ is a differentiable function on an open interval J, then the composite function

$$\beta = \alpha(h): J \to \mathbf{R}^3$$

is a curve called a *reparametrization* of α by h.

For each $s \in J$, the new curve β is at the point $\beta(s) = \alpha(h(s))$ reached by α at $h(s)$ in I (Fig. 1.11). Thus β represents a different trip over at least part of the route of α.

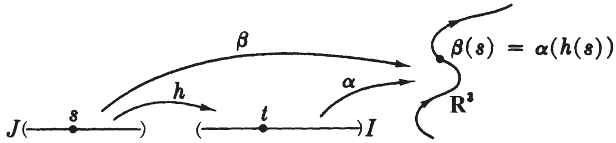


FIG. 1.11

To compute the coordinates of β, simply substitute $t = h(s)$ into the coordinates $\alpha_1(t)$, $\alpha_2(t)$, $\alpha_3(t)$ of α. For example, suppose

$$\alpha(t) = \left(\sqrt{t},\, t\sqrt{t},\, 1-t\right) \quad on \quad I\colon 0 < t < 4.$$

If $h(s) = s^2$ on J: $0 < s < 2$, then the reparametrized curve is

$$\beta(s) = \alpha(h(s)) = \alpha(s^2) = (s,\ s^3,\ 1 - s^2).$$

The following lemma relates the velocities of a curve and of a reparametrization.

4.5 Lemma If β is the reparametrization of α by h, then

$$\beta'(s) = (dh/ds)(s)\alpha'(h(s)).$$

Proof. If $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, then

$$\beta(s) = \alpha(h(s)) = (\alpha_1(h(s)), \alpha_2(h(s)), \alpha_3(h(s)).$$

Using the "prime" notation for derivatives, the chain rule for a composition of real-valued functions f and g reads $(g(f))' = g'(f)\cdot f'$. Thus, in the case at hand,

$$\alpha_i(h)'(s) = \alpha_i'(h(s))\cdot h'(s).$$

By the definition of velocity, this yields

$$\begin{aligned}\beta'(s) &= \alpha(h)'(s)\\ &= (\alpha_1'(h(s))\cdot h'(s), \alpha_2'(h(s))\cdot h'(s), \alpha_3'(h(s))\cdot h'(s))\\ &= h'(s)\alpha'(h(s)).\end{aligned}$$ ◆

According to this lemma, to obtain the velocity of a reparametrization of α by h, first reparametrize α' by h, then scalar multiply by the derivative of h.

Since velocities are tangent vectors, we can take the derivative of a function with respect to a velocity.

4.6 Lemma Let α be a curve in $\mathbf{R}^3$ and let f be a differentiable function on $\mathbf{R}^3$. Then

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

Proof. Since

$$\alpha' = \left(\frac{d\alpha_1}{dt}, \frac{d\alpha_2}{dt}, \frac{d\alpha_3}{dt}\right)_\alpha,$$

we conclude from Lemma 3.2 that

$$\alpha'(t)[f] = \sum \frac{\partial f}{\partial x_i}(\alpha(t))\frac{d\alpha_i}{dt}(t).$$

But the composite function $f(\alpha)$ may be written $f(\alpha_1, \alpha_2, \alpha_3)$, and the chain rule then gives exactly the same result for the derivative of $f(\alpha)$. ◆

By definition, $\alpha'(t)[f]$ is the rate of change of f along the line through $\alpha(t)$ in the $\alpha'(t)$ direction. (If $\alpha'(t) \neq 0$, this is the tangent line to α at $\alpha(t)$; see Exercise 9.) The lemma shows that this rate of change is the same as that of f along the curve α itself.

Since a curve $\alpha\colon I \to \mathbf{R}^3$ is a function, it makes sense to say that α is one-to-one; that is, $\alpha(t) = \alpha(t_1)$ only if $t = t_1$. Another special property of curves is periodicity: A curve $\alpha\colon \mathbf{R} \to \mathbf{R}^3$ is *periodic* if there is a number $p > 0$ such that $\alpha(t + p) = \alpha(t)$ for all t—and the smallest such number p is then called the *period* of α.

From the viewpoint of calculus, the most important condition on a curve α is that it be *regular*, that is, have all velocity vectors different from zero. Such a curve can have no corners or cusps.

The following remarks about curves (offered without proof) describe another familiar way to formulate the concept of "curve." If f is a differentiable real-valued function on $\mathbf{R}^2$, let

$$C\colon f = a$$

be the set of all points $\mathbf{p}$ in $\mathbf{R}^2$ such that $f(\mathbf{p}) = a$. Now, if the partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ are never simultaneously zero at any point of C, then C consists of one or more separate "components," which we shall call *Curves.*† For example, $C\colon x^2 + y^2 = r^2$ is the circle of radius r centered at the

† The capital C distinguishes this notion from a (parametrized) curve $\alpha\colon I \to \mathbf{R}^2$.

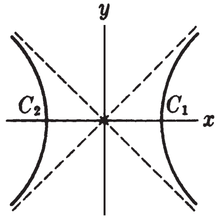


FIG. 1.12

origin of $\mathbf{R}^2$, and the hyperbola C: $x^2 - y^2 = r^2$ splits into two Curves ("branches") C_1 and C_2 as shown in Fig. 1.12.

Every Curve C is the route of many regular curves, called *parametrizations* of C. For example, the curve

$$\alpha(t) = (r \cos t, r \sin t)$$

is a well-known periodic parametrization of the circle given above, and for $r > 0$ the one-to-one curve

$$\beta(t) = (r \cosh t, r \sinh t)$$

parametrizes the branch $x > 0$ of the hyperbola.

Exercises

1. Compute the velocity vector of the curve in Example 4.2(3) for arbitrary t and for $t = 0$, $t = \pi/2$, $t = \pi$, visualizing those on Fig. 1.8.

2. Find the unique curve such that $\alpha(0) = (1, 0, 5)$ and $\alpha'(t) = (t^2, t, e^t)$.

3. Find the coordinate functions of the curve $\beta = \alpha(h)$, where α is the curve in Example 4.2(3) and $h(s) = \cos^{-1}(s)$ on J: $0 < s < 1$.

4. Reparametrize the curve α in Example 4.2(4) using $h(s) = \log s$ on J: $s > 0$. Check the equation in Lemma 4.5 in this case by calculating each side separately.

5. Find the equation of the straight line through the points $(1, -3, -1)$ and $(6, 2, 1)$. Does this line meet the line through the points $(-1, 1, 0)$ and $(-5, -1, -1)$?

6. Deduce from Lemma 4.6 that in the definition of directional derivative (Def. 3.1) the straight line $t \to \mathbf{p} + t\mathbf{v}$ can be replaced by any curve α with *initial velocity* $\mathbf{v}_p$, that is, such that $\alpha(0) = \mathbf{p}$ and $\alpha'(0) = \mathbf{v}_p$.

7. (*Continuation.*)
(a) Show that the curves with coordinate functions

$$(t, 1+t^2, t), \quad (\sin t, \cos t, t), \quad (\sinh t, \cosh t, t)$$

all have the same initial velocity $\mathbf{v}_p$.
(b) If $f = x^2 - y^2 + z^2$, compute $\mathbf{v}_p[f]$ by calculating $d(f(\alpha))/dt$ at $t = 0$, using each of three curves in (a).

8. Sketch the following Curves in $\mathbf{R}^2$, and find parametrizations for each.
(a) C: $4x^2 + y^2 = 1$, (b) C: $3x + 4y = 1$,
(c) C: $y = e^x$.

9. For a fixed t, the *tangent line* to a regular curve α at the point $\alpha(t)$ is the straight line $u \to \alpha(t) + u\alpha'(t)$, where we delete the point of application of $\alpha'(t)$. Find the tangent line to the helix $\alpha(t) = (2\cos t, 2\sin t, t)$ at the points $\alpha(0)$ and $\alpha(\pi/4)$.

1.5 1-Forms

If f is a real-valued function on $\mathbf{R}^3$, then in elementary calculus the differential of f is usually defined as

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz.$$

It is not always made clear exactly what this formal expression means. In this section we give a rigorous treatment using the notion of 1-form, and forms tend to appear at crucial moments in later work.

5.1 Definition A *1-form* ϕ on $\mathbf{R}^3$ is a real-valued function on the set of all tangent vectors to $\mathbf{R}^3$ such that ϕ is linear at each point, that is,

$$\phi(a\mathbf{v} + b\mathbf{w}) = a\phi(\mathbf{v}) + b\phi(\mathbf{w})$$

for any numbers a, b and tangent vectors $\mathbf{v}$, $\mathbf{w}$ at the same point of $\mathbf{R}^3$.

We emphasize that for every tangent vector $\mathbf{v}$, a 1-form ϕ defines a real number $\phi(\mathbf{v})$; and for each point $\mathbf{p}$ in $\mathbf{R}^3$, the resulting function ϕ_p: $T_p(\mathbf{R}^3) \to \mathbf{R}$ is linear. Thus at each point $\mathbf{p}$, ϕ_p is an element of the *dual space* of $T_p(\mathbf{R}^3)$. In this sense the notion of 1-form is dual to that of vector field.

The sum of 1-forms ϕ and ψ is defined in the usual pointwise fashion:

$$(\phi + \psi)(\mathbf{v}) = \phi(\mathbf{v}) + \psi(\mathbf{v}) \quad \text{for all tangent vectors } \mathbf{v}.$$

Similarly, if f is a real-valued function on $\mathbf{R}^3$ and ϕ is a 1-form, then $f\phi$ is the 1-form such that

$$(f\phi)(\mathbf{v}_p) = f(\mathbf{p})\phi(\mathbf{v}_p)$$

for all tangent vectors $\mathbf{v}_p$.

There is also a natural way to *evaluate a 1-form* ϕ *on a vector field* V to obtain a real-valued function $\phi(V)$: At each point $\mathbf{p}$ the value of $\phi(V)$ is the number $\phi(V(\mathbf{p}))$. Thus a 1-form may also be viewed as a machine that converts vector fields into real-valued functions. If $\phi(V)$ is differentiable whenever V is, we say that ϕ is *differentiable*. As with vector fields, we shall always assume that the 1-forms we deal with are differentiable.

A routine check of definitions shows that $\phi(V)$ is linear in both ϕ and V; that is,

$$\phi(fV + gW) = f\phi(V) + g\phi(W)$$

and

$$(f\phi + g\psi)(V) = f\phi(V) + g\psi(V),$$

where f and g are functions.

Using the notion of directional derivative, we now define a most important way to convert functions into 1-forms.

5.2 Definition If f is a differentiable real-valued function on $\mathbf{R}^3$, the *differential* df of f is the 1-form such that

$$df(\mathbf{v}_p) = \mathbf{v}_p[f] \quad \text{for all tangent vectors } \mathbf{v}_p.$$

In fact, df is a 1-form, since by definition it is a real-valued function on tangent vectors, and by (1) of Theorem 3.3 it is linear at each point $\mathbf{p}$. Clearly, df knows all rates of change of f in all directions on $\mathbf{R}^3$, so it is not surprising that differentials are fundamental to the calculus on $\mathbf{R}^3$.

Our task now is to show that these rather abstract definitions lead to familiar results when expressed in terms of coordinates.

5.3 Example 1-Forms on $\mathbf{R}^3$. (1) The differentials dx_1, dx_2, dx_3 of the natural coordinate functions. Using Lemma 3.2 we find

$$dx_i(\mathbf{v}_p) = \mathbf{v}_p[x_i] = \sum_j v_j \frac{\partial x_i}{\partial x_j}(\mathbf{p}) = \sum_j v_j \delta_{ij} = v_i,$$

where δ_{ij} is the Kronecker delta (0 if $i \neq j$, 1 if $i = j$). Thus *the value of* dx_i *on an arbitrary tangent vector* $\mathbf{v}_p$ *is the* ith *coordinate* v_i *of its vector part*—and does not depend on the point of application $\mathbf{p}$.

(2) The 1-form $\psi = f_1 dx_1 + f_2 dx_2 + f_3 dx_3$. Since dx_i is a 1-form, our definitions show that ψ is also a 1-form for any functions f_1, f_2, f_3. The value of ψ on an arbitrary tangent vector $\mathbf{v}_p$ is

$$\psi(\mathbf{v}_p) = \left(\sum f_i dx_i\right)(\mathbf{v}_p) = \sum f_i(\mathbf{p}) dx_i(\mathbf{v}_p) = \sum f_i(\mathbf{p}) v_i,$$

The first of these examples shows that the 1-forms dx_1, dx_2, dx_3 are the analogues for tangent vectors of the natural coordinate functions x_1, x_2, x_3 for points. Alternatively, we can view dx_1, dx_2, dx_3 as the "duals" of the natural unit vector fields U_1, U_2, U_3. In fact, it follows immediately from (1) above that the function $dx_i(U_j)$ has the constant value δ_{ij}.

We now show that every 1-form can be written in the concrete manner given in (2) above.

5.4 Lemma If ϕ is a 1-form on $\mathbf{R}^3$, then $\phi = \sum f_i dx_i$, where $f_i = \phi(U_i)$. These functions f_1, f_2, f_3 are called the *Euclidean coordinate functions* of ϕ.

Proof. By definition, a 1-form is a function on tangent vectors; thus ϕ and $\sum f_i dx_i$ are equal if and only if they have the same value on every tangent vector $\mathbf{v}_p = \sum v_i U_i(\mathbf{p})$. In (2) of Example 5.3 we saw that

$$\left(\sum f_i dx_i\right)(\mathbf{v}_p) = \sum f_i(\mathbf{p}) v_i.$$

On the other hand,

$$\phi(\mathbf{v}_p) = \phi\left(\sum v_i U_i(\mathbf{p})\right) = \sum v_i \phi(U_i(\mathbf{p})) = \sum v_i f_i(\mathbf{p})$$

since $f_i = \phi(U_i)$. Thus ϕ and $\sum f_i dx_i$ do have the same value on every tangent vector. ◆

This lemma shows that a 1-form on $\mathbf{R}^3$ is nothing more than an expression $f\,dx + g\,dy + h\,dz$, and such expressions are now rigorously defined as functions on tangent vectors. Let us now show that the definition of differential of a function (Definition 5.2) agrees with the informal definition given at the start of this section.

5.5 Corollary If f is a differentiable function on $\mathbf{R}^3$, then

$$df = \sum \frac{\partial f}{\partial x_i} dx_i.$$

Proof. The value of $\sum(\partial f/\partial x_i)dx_i$ on an arbitrary tangent vector $\mathbf{v}_p$ is $\sum(\partial f/\partial x_i)\,(\mathbf{p})v_i$. By Lemma 3.2, $df(\mathbf{v}_p) = \mathbf{v}_p[f]$ is the same. Thus the 1-forms df and $\sum(\partial f/\partial x_i)\,dx_i$ are equal. ◆

Using either this result or the definition of d, it is immediate that

$$d(f+g) = df + dg.$$

Finally, we determine the effect of d on *products* of functions and on *compositions* of functions.

5.6 Lemma Let fg be the product of differentiable functions f and g on $\mathbf{R}^3$. Then

$$d(fg) = g\,df + f\,dg.$$

Proof. Using Corollary 5.5, we obtain

$$d(fg) = \sum \frac{\partial(fg)}{\partial x_i}dx_i = \sum\left(\frac{\partial f}{\partial x_i}g + f\frac{\partial g}{\partial x_i}\right)dx_i$$

$$= g\left(\sum\frac{\partial f}{\partial x_i}dx_i\right) + f\left(\sum\frac{\partial g}{\partial x_i}dx_i\right) = g\,df + f\,dg. \quad ◆$$

5.7 Lemma Let $f\colon \mathbf{R}^3 \to \mathbf{R}$ and $h\colon \mathbf{R} \to \mathbf{R}$ be differentiable functions, so the composite function $h(f)\colon \mathbf{R}^3 \to \mathbf{R}$ is also differentiable. Then

$$d(h(f)) = h'(f)\,df.$$

Proof. (The prime here is just the ordinary derivative, so $h'(f)$ is again a composite function, from $\mathbf{R}^3$ to $\mathbf{R}$.) The usual chain rule for a composite function such as $h(f)$ reads

$$\frac{\partial(h(f))}{\partial x_i} = h'(f)\frac{\partial f}{\partial x_i}.$$

Hence

$$d(h(f)) = \sum\frac{\partial(h(f))}{\partial x_i}dx_i = \sum h'(f)\frac{\partial f}{\partial x_i}dx_i = h'(f)\,df. \quad ◆$$

To compute df for a given function f it is almost always simpler to use these properties of d rather than substitute in the formula of Corollary 5.5. Then from df we immediately get the partial derivatives of f, and, in fact, *all its directional derivatives*. For example, suppose

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

Then by Lemmas 5.6 and 5.7,

$$\begin{aligned} df &= (2x\,dx)y + (x^2 - 1)dy + (2y\,dy)z + (y^2 + 2)dz \\ &= \underbrace{2xy}_{\partial f/\partial x}\,dx + \underbrace{(x^2 + 2yz - 1)}_{\partial f/\partial y}dy + \underbrace{(y^2 + 2)}_{\partial f/\partial z}dz. \end{aligned}$$

Now use the rules above to evaluate this expression on a tangent vector $\mathbf{v}_p$. The result is

$$\mathbf{v}_p[f] = df(\mathbf{v}_p) = 2p_1p_2v_1 + (p_1^2 + 2p_2p_3 - 1)v_2 + (p_2^2 + 1)v_3.$$

Exercises

1. Let $\mathbf{v} = (1, 2, -3)$ and $\mathbf{p} = (0, -2, 1)$. Evaluate the following 1-forms on the tangent vector $\mathbf{v}_p$.

(a) $y^2\,dx$. (b) $z\,dy - y\,dz$.
(c) $(z^2 - 1)dx - dy + x^2\,dz$.

2. If $\phi = \sum f_i dx_i$ and $V = \sum v_i U_i$, show that the 1-form ϕ evaluated on the vector field V is the function $\phi(V) = \sum f_i v_i$.

3. Evaluate the 1-form $\phi = x^2\,dx - y^2\,dz$ on the vector fields
$V = xU_1 + yU_2 + zU_3$,
$W = xy\,(U_1 - U_3) + yz\,(U_1 - U_2)$, and $(1/x)V + (1/y)W$.

4. Express the following differentials in terms of df:

(a) $d(f^5)$. (b) $d\left(\sqrt{f}\right)$, where $f > 0$.
(c) $d(\log(1 + f^2))$.

5. Express the differentials of the following functions in the standard form $\sum f_i\,dx_i$.

(a) $(x^2 + y^2 + z^2)^{1/2}$. (b) $\tan^{-1}(y/x)$.

6. In each case compute the differential of f and find the directional derivative $\mathbf{v}_p[f]$, for $\mathbf{v}_p$ as in Exercise 1.

(a) $f = xy^2 - yz^2$. (b) $f = xe^{yz}$.
(c) $f = \sin(xy)\cos(xz)$.

7. Which of the following are 1-forms? In each case ϕ is the function on tangent vectors such that the value of ϕ on $(v_1, v_2, v_3)_p$ is

(a) $v_1 - v_3$.
(b) $p_1 - p_3$.
(c) $v_1 p_3 + v_2 p_1$.
(d) $\mathbf{v}_p[x^2 + y^2]$.
(e) 0.
(f) $(p_1)^2$.

In case ϕ is a 1-form, express it as $\sum f_i\, dx_i$.

8. Prove Lemma 5.6 directly from the definition of d.

9. A 1-form ϕ is *zero* at a point $\mathbf{p}$ provided $\phi(\mathbf{v}_p) = 0$ for all tangent vectors at $\mathbf{p}$. A point at which its differential df is zero is called a *critical point* of the function f. Prove that $\mathbf{p}$ is a critical point of f if and only if

$$\frac{\partial f}{\partial x}(\mathbf{p}) = \frac{\partial f}{\partial y}(\mathbf{p}) = \frac{\partial f}{\partial z}(\mathbf{p}) = 0.$$

Find all critical points of $f = (1 - x^2)y + (1 - y^2)z$.

(*Hint:* Find the partial derivatives of f by computing df.)

10. (*Continuation.*) Prove that the local maxima and local minima of f are critical points of f. (f has a *local maximum* at $\mathbf{p}$ if $f(\mathbf{q}) \leqq f(\mathbf{p})$ for all $\mathbf{q}$ near $\mathbf{p}$.)

11. It is sometimes asserted that df is the linear approximation of Δf.

(a) Explain the sense in which $(df)(\mathbf{v}_p)$ is a linear approximation of $f(\mathbf{p} + \mathbf{v}) - f(\mathbf{p})$.

(b) Compute exact and approximate values of $f(0.9, 1.6, 1.2) - f(1, 1.5, 1)$, where $f = x^2y/z$.

1.6 Differential Forms

The 1-forms on $\mathbf{R}^3$ are part of a larger system called the *differential forms* on $\mathbf{R}^3$. We shall not give as rigorous an account of differential forms as we did of 1-forms since our use of the full system on $\mathbf{R}^3$ is limited. However, the *properties* established here are valid whenever differential forms are used.

Roughly speaking, a *differential form* on $\mathbf{R}^3$ is an expression obtained by adding and multiplying real-valued functions and the differentials dx_1, dx_2, dx_3 of the natural coordinate functions of $\mathbf{R}^3$. These two operations obey the usual associative and distributive laws; however, the multiplication is not commutative. Instead, it obeys the

$$\textit{alternation rule: } dx_i\, dx_j = -dx_j\, dx_i \quad (1 \leqq i, j \leqq 3).$$

This rule appears—although rather inconspicuously—in elementary calculus (see Exercise 9).

A consequence of the alternation rule is the fact that "repeats are zero," that is, $dx_i\, dx_i = 0$, since if $i = j$ the alternation rule reads

$$dx_i\, dx_i = -dx_i\, dx_i.$$

If each summand of a differential form contains p dx_i's ($p = 0, 1, 2, 3$), the form is called a *p-form*, and is said to have *degree p*. Thus, shifting to dx, dy, dz, we find

A 0-form is just a differentiable function f.

A 1-form is an expression $f\,dx + g\,dy + h\,dz$, just as in the preceding section.

A 2-form is an expression $f\,dx\,dy + g\,dx\,dz + h\,dy\,dz$.

A 3-form is an expression $f\,dx\,dy\,dz$.

We already know how to add 1-forms: simply add corresponding coefficient functions. Thus, in index notation,

$$\sum f_i\,dx_i + \sum g_i\,dx_i = \sum (f_i + g_i)dx_i.$$

The corresponding rule holds for 2-forms or 3-forms.

On three-dimensional Euclidean space, all p-forms with $p > 3$ are zero. This is a consequence of the alternation rule, for a product of more than three dx_i's must contain some dx_i twice, but repeats are zero, as noted above. For example, $dx\,dy\,dx\,dz = -dx\,dx\,dy\,dz = 0$, since $dx\,dx = 0$. As a reminder that the alternation rule is to be used, we denote this multiplication of forms by a *wedge* $\wedge$. (However, we do not bother with the wedge when only products of dx, dy, dz are involved.)

6.1 Example Computation of wedge products.

(1) Let

$$\phi = x\,dx - y\,dy \quad \text{and} \quad \psi = z\,dx + x\,dz.$$

Then

$$\begin{aligned}\phi \wedge \psi &= (x\,dx - y\,dy) \wedge (z\,dx + x\,dz)\\ &= xz\,dx\,dx + x^2\,dx\,dz - yz\,dy\,dx - yx\,dy\,dz.\end{aligned}$$

But $dx\,dx = 0$ and $dy\,dx = -dx\,dy$. Thus

$$\phi \wedge \psi = yz\,dx\,dy + x^2 dx\,dz - xy\,dy\,dz.$$

In general, the product of two 1-forms is a 2-form.

(2) Let ϕ and ψ be the 1-forms given above and let $\theta = z\,dy$. Then

$$\theta \wedge \phi \wedge \psi = yz^2\,dy\,dx\,dy + x^2z\,dy\,dx\,dz - xyz\,dy\,dy\,dz.$$

Since $dy\,dx\,dy$ and $dy\,dy\,dz$ each contain repeats, both are zero. Thus

$$\theta \wedge \phi \wedge \psi = -x^2z\,dx\,dy\,dz.$$

(3) Let ϕ be as above, and let η be the 2-form $y\,dx\,dz + x\,dy\,dz$. Omitting forms containing repeats, we find

$$\phi \wedge \eta = x^2\,dx\,dy\,dz - y^2\,dy\,dx\,dz = (x^2 + y^2)\,dx\,dy\,dz.$$

It should be clear from these examples that the wedge product of a p-form and a q-form is a $(p + q)$-form. Thus such a product is automatically zero whenever $p + q > 3$.

6.2 Lemma If ϕ and ψ are 1-forms, then

$$\phi \wedge \psi = -\psi \wedge \phi.$$

Proof. Write

$$\phi = \sum f_i\,dx_i, \quad \psi = \sum g_i\,dx_i.$$

Then by the alternation rule,

$$\phi \wedge \psi = \sum f_i g_j\,dx_i\,dx_j = -\sum g_j f_i\,dx_j\,dx_i = -\psi \wedge \phi. \qquad \blacklozenge$$

In the language of differential forms, the operator d of Definition 5.2 converts a 0-form f into a 1-form df. It is easy to generalize to an operator (also denoted by d) that converts a p-form η into a $(p + 1)$-form $d\eta$: One simply applies d (of Definition 5.2) to the coefficient functions of η. For example, here is the case $p = 1$.

6.3 Definition If $\phi = \sum f_i\,dx_i$ is a 1-form on $\mathbf{R}^3$, the *exterior derivative* of ϕ is the 2-form $d\phi = \sum df_i \wedge dx_i$.

If we expand the preceding definition using Corollary 5.5, we obtain the following interesting formula for the exterior derivative of

$$\phi = f_1dx_1 + f_2dx_2 + f_3dx_3:$$

$$d\phi = \left(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\right)dx_1dx_2 + \left(\frac{\partial f_3}{\partial x_1} - \frac{\partial f_1}{\partial x_3}\right)dx_1dx_3 + \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}\right)dx_2dx_3.$$

There is no need to memorize this formula; it is more reliable simply to apply the definition in each case. For example, suppose

$$\phi = xy\ dx + x^2\ dz.$$

Then

$$\begin{aligned} d\phi &= d(xy) \wedge dx + d(x^2) \wedge dz \\ &= (y\ dx + x\ dy) \wedge dx + (2x\ dx) \wedge dz \\ &= -x\ dx\ dy + 2x\ dx\ dz. \end{aligned}$$

It is easy to check that the general exterior derivative enjoys the same linearity property as the particular case in Definition 5.2; that is,

$$d(a\phi + b\psi) = a\ d\phi + b\ d\psi,$$

where ϕ and ψ are arbitrary forms and a and b are numbers.

The exterior derivative and the wedge product work together nicely:

6.4 Theorem Let f and g be functions, ϕ and ψ 1-forms. Then

(1) $d(fg) = df\ g + f\ dg$.
(2) $d(f\phi) = df \wedge \phi + f\ d\phi$.
(3) $d(\phi \wedge \psi) = d\phi \wedge \psi - \phi \wedge d\psi$.†

Proof. The first formula is just Lemma 5.6. We include it to show the family resemblance of all three formulas. The proof of (2) is a simpler version of that of (3), so we outline a proof of the latter—watching to see where the minus sign comes from.

It suffices to prove the formula when $\phi = f\ du$, $\psi = g\ dv$, where u and v are any of the coordinate functions x_1, x_2, x_3. In fact, every 1-form is a sum of such terms, so the general case will follow by the linearity of d and the algebra of wedge products.

For example, let us try the typical case $\phi = f\ dx$, $\psi = g\ dy$. Since repeats kill, there is no use writing down terms that are bound to be eliminated. Hence

$$d(\phi \wedge \psi) = d(fg\ dx\ dy) = \frac{\partial(fg)}{\partial z} dz\ dx\ dy = \left(f\frac{\partial g}{\partial z} + g\frac{\partial f}{\partial z} \right) dx\ dy\ dz. \quad (*)$$

† As usual, multiplication takes precedence over addition or subtraction, so this expression should be read as $(d\phi \wedge \psi) - (\phi \wedge d\psi)$.

Now,

$$d\phi \wedge \psi = d(f\ dx) \wedge g\ dy = \frac{\partial f}{\partial z}\ dz \wedge dx \wedge g\ dy = g \frac{\partial f}{\partial z}\ dx\ dy\ dz.$$

But

$$\phi \wedge d\psi = f\ dx \wedge d(g\ dy) = f\ dx \wedge \frac{\partial g}{\partial z}\ dz \wedge dy = -f \frac{\partial g}{\partial z}\ dx\ dy\ dz,$$

since $dx\ dz\ dy = -dx\ dy\ dz$. Thus we must *subtract* this last equation from its predecessor to get (∗). ◆

One way to remember the minus sign in equation (3) of the theorem is to treat d as if it were a 1-form. To reach ψ, d must change places with ϕ, hence the minus sign is consistent with the alternation rule in Lemma 6.2.

Differential forms, and the associated notions of wedge product and exterior derivative, provide the means of expressing quite complicated relations among the partial derivatives in a highly efficient way. The wedge product saves much useless labor by discarding, right at the start, terms that will eventually disappear. But the exterior derivative d is the key. Exercise 8 shows, for example, how it replaces all three of the differentiation operations of classical vector analysis.

Exercises

1. Let $\phi = yz\ dx + dz$, $\psi = \sin z\ dx + \cos z\ dy$, $\xi = dy + z\ dz$. Find the standard expressions (in terms of $dxdy, \ldots$) for

(a) $\phi \wedge \psi$, $\psi \wedge \xi$, $\xi \wedge \phi$. (b) $d\phi$, $d\psi$, $d\xi$.

2. Let $\phi = dx/y$ and $\psi = z\ dy$. Check the Leibnizian formula (3) of Theorem 6.4 in this case by computing each term separately.

3. For any function f show that $d(df) = 0$. Deduce that $d(f\ dg) = df \wedge dg$.

4. Simplify the following forms:

(a) $d(f\ dg + g\ df)$. (b) $d((f - g)\ (df + dg))$.
(c) $d(f\ dg \wedge g\ df)$. (d) $d(gf\ df) + d(f\ dg)$.

5. For any three 1-forms $\phi_i = \sum_j f_{ij} dx_j$ $(1 \leqq i \leqq 3)$, prove

$$\phi_1 \wedge \phi_2 \wedge \phi_3 = \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} dx_1\ dx_2\ dx_3.$$

6. If r, ϑ, z are the cylindrical coordinate functions on $\mathbf{R}^3$, then $x = r\cos\vartheta$, $y = r\sin\vartheta$, $z = z$. Compute the *volume element dx dy dz* of $\mathbf{R}^3$ in cylindrical coordinates. (That is, express $dx\,dy\,dz$ in terms of the functions r, ϑ, z, and their differentials.)

7. For a 2-form

$$\eta = f\,dx\,dy + g\,dx\,dz + h\,dy\,dz,$$

the *exterior derivative* $d\eta$ is defined to be the 3-form obtained by replacing f, g, and h by their differentials. Prove that for any 1-form ϕ, $d(d\phi) = 0$.

Exercises 3 and 7 show that $d^2 = 0$, that is, for any form ξ, $d(d\xi) = 0$. (If ξ is a 2-form, then $d(d\xi) = 0$, since its degree exceeds 3.)

8. Classical *vector analysis* avoids the use of differential forms on $\mathbf{R}^3$ by converting 1-forms and 2-forms into vector fields by means of the following one-to-one correspondences:

$$\sum f_i\,dx_i \overset{(1)}{\leftrightarrow} \sum f_i U_i \overset{(2)}{\leftrightarrow} f_3 dx_1 dx_2 - f_2 dx_1 dx_3 + f_1 dx_2 dx_3.$$

Vector analysis uses three basic operations based on partial differentiation:

Gradient of a function f:

$$\text{grad } f = \sum \frac{\partial f}{\partial x_i} U_i.$$

Curl of a vector field $V = \sum f_i U_i$:

$$\text{curl } V = \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}\right)U_1 + \left(\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}\right)U_2 + \left(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\right)U_3.$$

Divergence of a vector field $V = \sum f_i U_i$:

$$\text{div } V = \sum \frac{\partial f_i}{\partial x_i}.$$

Prove that all three operations may be expressed by exterior derivatives as follows:

(a) $df \overset{(1)}{\leftrightarrow} \text{grad } f$.

(b) If $\phi \overset{(1)}{\leftrightarrow} V$, then $d\phi \overset{(2)}{\leftrightarrow} \text{curl } V$.

(c) If $\eta \overset{(1)}{\leftrightarrow} V$, then $d\eta = (\text{div } V)\,dx\,dy\,dz$.

9. Let f and g be real-valued functions on $\mathbf{R}^2$. Prove that

$$df \wedge dg = \begin{vmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\ \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{vmatrix} dx\, dy.$$

This formula appears in elementary calculus; show that it implies the alternation rule.

1.7 Mappings

In this section we discuss functions from $\mathbf{R}^n$ to $\mathbf{R}^m$. If $n = 3$ and $m = 1$, then such a function is just a real-valued function on $\mathbf{R}^3$. If $n = 1$ and $m = 3$, it is a curve in $\mathbf{R}^3$. Although our results will necessarily be stated for arbitrary m and n, we are primarily interested in only three other cases:

$$\mathbf{R}^2 \to \mathbf{R}^2, \quad \mathbf{R}^2 \to \mathbf{R}^3, \quad \mathbf{R}^3 \to \mathbf{R}^3.$$

The fundamental observation about a function $F\colon \mathbf{R}^n \to \mathbf{R}^m$ is that it can be completely described by m real-valued functions on $\mathbf{R}^n$. (We saw this already in Section 4 for $n = 1$, $m = 3$.)

7.1 Definition Given a function $F\colon \mathbf{R}^n \to \mathbf{R}^m$, let $f_1, f_2, \ldots, f_m$, denote the real-valued functions on $\mathbf{R}^n$ such that

$$F(\mathbf{p}) = (f_1(\mathbf{p}), f_2(\mathbf{p}), \ldots, f_m(\mathbf{p}))$$

for all points $\mathbf{p}$ in $\mathbf{R}^n$. These functions are called the *Euclidean coordinate functions* of F, and we write $F = (f_1, f_2, \ldots, f_m)$.

The function F is *differentiable* provided its coordinate functions are differentiable in the usual sense. A differentiable function $F\colon \mathbf{R}^n \to \mathbf{R}^m$ is called a *mapping* from $\mathbf{R}^n$ to $\mathbf{R}^m$.

Note that the coordinate functions of F are the composite functions $f_i = x_i(F)$, where $x_1, \ldots, x_m$ are the coordinate functions of $\mathbf{R}^m$.

Mappings may be described in many different ways. For example, suppose $F\colon \mathbf{R}^3 \to \mathbf{R}^3$ is the mapping $F = (x^2, yz, xy)$. Thus

$$F(\mathbf{p}) = \left(x(\mathbf{p})^2, y(\mathbf{p})z(\mathbf{p}), x(\mathbf{p})y(\mathbf{p})\right) \quad \text{for all } \mathbf{p}.$$

Now, $\mathbf{p} = (p_1, p_2, p_3)$, and by definition of the coordinate functions,

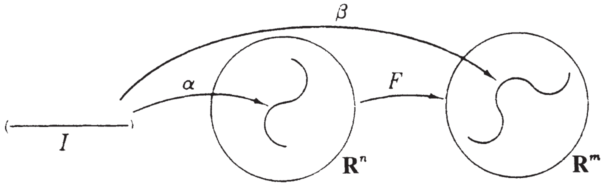


FIG. 1.13

$$x(\mathbf{p}) = p_1,\ y(\mathbf{p}) = p_2,\ z(\mathbf{p}) = p_3.$$

Hence we obtain the following *pointwise* formula for F:

$$F(p_1, p_2, p_3) = (p_1^2, p_2 p_3, p_1 p_2) \quad \text{for all } p_1, p_2, p_3.$$

Thus, for example,

$$F(1, -2, 0) = (1, 0, -2) \quad \text{and} \quad F(-3, 1, 3) = (9, 3, -3).$$

In principle, one could deduce the theory of curves from the general theory of mappings. But curves are reasonably simple, while a mapping, even in the case $\mathbf{R}^2 \to \mathbf{R}^2$, can be quite complicated. Hence we reverse this process and use curves, at every stage, to gain an understanding of mappings.

7.2 Definition If $\alpha\colon I \to \mathbf{R}^n$ is a curve in $\mathbf{R}^n$ and $F\colon \mathbf{R}^n \to \mathbf{R}^m$ is a mapping, then the composite function $\beta = F(\alpha)\colon I \to \mathbf{R}^m$ is a curve in $\mathbf{R}^m$ called the *image of* α *under* F (Fig. 1.13).

7.3 Example Mappings. (1) Consider the mapping $F\colon \mathbf{R}^3 \to \mathbf{R}^3$ such that

$$F = (x - y, x + y, 2z).$$

In pointwise terms then,

$$F(p_1, p_2, p_3) = (p_1 - p_2, p_1 + p_2, 2p_3) \quad \text{for all } p_1, p_2, p_3.$$

Only when a mapping is quite simple can one hope to get a good idea of its behavior by merely computing its values at some finite number of points. But this function *is* quite simple —it is a *linear* transformation from $\mathbf{R}^3$ to $\mathbf{R}^3$.

Thus by a well-known theorem of linear algebra, F is completely determined by its values at three (linearly independent) points, say the *unit points*

$$\mathbf{u}_1 = (1, 0, 0), \quad \mathbf{u}_2 = (0, 1, 0), \quad \mathbf{u}_3 = (0, 0, 1).$$

(2) The mapping F: $\mathbf{R}^2 \rightarrow \mathbf{R}^2$ such that $F(u, v) = (u^2 - v^2, 2uv)$. (Here u and v are the coordinate functions of $\mathbf{R}^2$.) To analyze this mapping, we examine its effect on the curve $\alpha(t) = (r \cos t, r \sin t)$, where $0 \leqq t \leqq 2\pi$. This curve takes one counterclockwise trip around the circle of radius r with center at the origin. The image curve is

$$\begin{aligned} \beta(t) &= F(\alpha(t)) \\ &= F(r \cos t, r \sin t) \\ &= (r^2 \cos^2 t - r^2 \sin^2 t, 2r^2 \cos t \sin t) \end{aligned}$$

with $0 \leqq t \leqq 2\pi$. Using the trigonometric identities

$$\cos 2t = \cos^2 t - \sin^2 t, \quad \sin 2t = 2 \sin t \cos t,$$

we find for $\beta = F(\alpha)$ the formula

$$\beta(t) = (r^2 \cos 2t, r^2 \sin 2t),$$

with $0 \leqq t \leqq 2\pi$. This curve takes *two* counterclockwise trips around the circle of radius r^2 centered at the origin (Fig. 1.14).

Thus the effect of F is to wrap the plane $\mathbf{R}^2$ smoothly around itself twice—leaving the origin fixed, since $F(0, 0) = (0, 0)$. In this process, each circle of radius r is wrapped twice around the circle of radius r^2.

Generally speaking, differential calculus deals with approximation of smooth objects by linear objects. The best-known case is the approximation of a differentiable real-valued function f near x by the linear function $\Delta x \rightarrow f'(x)\, \Delta x$, which gives the tangent line at x to the graph of f. Our goal now is to define an analogous linear approximation for a mapping F: $\mathbf{R}^n \rightarrow \mathbf{R}^m$ near a point $\mathbf{p}$ of $\mathbf{R}^n$.

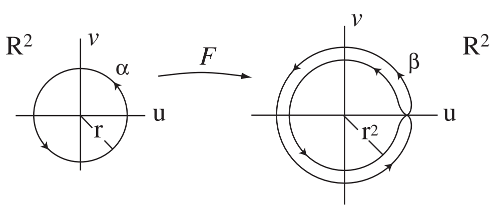


FIG. 1.14

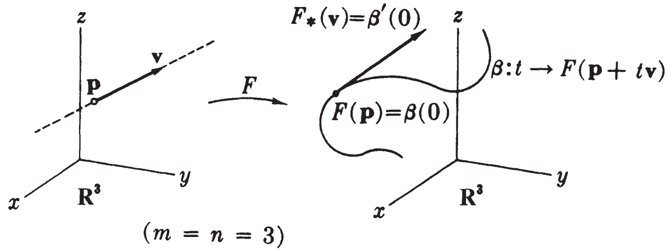


FIG. 1.15

Since $\mathbf{R}^n$ is filled by the radial lines $\alpha(t) = \mathbf{p} + t\mathbf{v}$ starting at $\mathbf{p}$, $\mathbf{R}^m$ is filled by their image curves $\beta(t) = F(\mathbf{p} + t\mathbf{v})$ starting at $F(\mathbf{p})$ (Fig. 1.15). So we approximate F near $\mathbf{p}$ by the map $F*$ that sends each initial velocity $\alpha'(0) = \mathbf{v}_p$ to the initial velocity $\beta'(0)$.

7.4 Definition Let F: $\mathbf{R}^n \to \mathbf{R}^m$ be a mapping. If $\mathbf{v}$ is a tangent vector to $\mathbf{R}^n$ at $\mathbf{p}$, let $F*(\mathbf{v})$ be the initial velocity of the curve $t \to F(\mathbf{p} + t\mathbf{v})$. The resulting function $F*$ sends tangent vectors to $\mathbf{R}^n$ to tangent vectors to $\mathbf{R}^m$, and is called the *tangent map* of F.

The tangent map can be described explicitly as follows.

7.5 Proposition Let $F = (f_1, f_2, \ldots, f_m)$ be a mapping from $\mathbf{R}^n$ to $\mathbf{R}^m$. If $\mathbf{v}$ is a tangent vector to $\mathbf{R}^n$ at $\mathbf{p}$, then

$$F*(\mathbf{v}) = (\mathbf{v}[f_1], \ldots, \mathbf{v}[f_m]) \quad \text{at } F(\mathbf{p}).$$

Proof. For definiteness, take $m = 3$. Then

$$\beta(t) = F(\mathbf{p} + t\mathbf{v}) = (f_1(\mathbf{p} + t\mathbf{v}), f_2(\mathbf{p} + t\mathbf{v}), f_3(\mathbf{p} + t\mathbf{v})).$$

By definition, $F*(\mathbf{v}) = \beta'(0)$. To get $\beta'(0)$, we take the derivatives, at $t = 0$, of the coordinate functions of β (Definition 4.3). But

$$\frac{d}{dt}(f_i(\mathbf{p} + t\mathbf{v}))\Big|_{t=0} = \mathbf{v}[f_i].$$

Thus

$$F*(\mathbf{v}) = (\mathbf{v}[f_1], \ \mathbf{v}[f_2], \ \mathbf{v}[f_3])\big|_{\beta(0)},$$

and $\beta(0) = F(\mathbf{p})$. ◆

Fix a point $\mathbf{p}$ of $\mathbf{R}^n$. The definition of tangent map shows that F_* sends tangent vectors at $\mathbf{p}$ to tangent vectors at $F(\mathbf{p})$. Thus for each $\mathbf{p}$ in $\mathbf{R}^n$, the function F_* gives rise to a function

$$F_{*p}\colon T_p(\mathbf{R}^n) \to T_{F(p)}(\mathbf{R}^m)$$

called the tangent map of F *at* $\mathbf{p}$. (Compare the analogous situation in elementary calculus where a function $f\colon \mathbf{R} \to \mathbf{R}$ has a derivative function $f'\colon \mathbf{R} \to \mathbf{R}$ that at each point t of $\mathbf{R}$ gives the derivative of f at t.)

7.6 Corollary If $F\colon \mathbf{R}^n \to \mathbf{R}^m$ is a mapping, then at each point $\mathbf{p}$ of $\mathbf{R}^n$ the tangent map $F_{*p}\colon T_p(\mathbf{R}^n) \to T_{F(p)}(\mathbf{R}^m)$ is a linear transformation.

Proof. We must show that for tangent vectors $\mathbf{v}$ and $\mathbf{w}$ at $\mathbf{p}$ and numbers a, b,

$$F_*(a\mathbf{v} + b\mathbf{w}) = aF_*(\mathbf{v}) + bF_*(\mathbf{w}).$$

This follows immediately from the preceding proposition by using the linearity in assertion (1) of Theorem 3.3. ◆

In fact, *the tangent map* F_{*p} at $\mathbf{p}$ *is the linear transformation that best approximates* F *near* $\mathbf{p}$. This idea is fully developed in advanced calculus, where it is used to prove Theorem 7.10.

Another consequence of the proposition is that *mappings preserve velocities of curves*. Explicitly:

7.7 Corollary Let $F\colon \mathbf{R}^n \to \mathbf{R}^m$ be a mapping. If $\beta = F(\alpha)$ is the image of a curve α in $\mathbf{R}^n$, then $\beta' = F_*(\alpha')$.

Proof. Again, set $m = 3$. If $F = (f_1, f_2, f_3)$, then

$$\beta = F(\alpha) = (f_1(\alpha), f_2(\alpha), f_3(\alpha)).$$

Hence Theorem 7.5 gives

$$F_*(\alpha') = (\alpha'[f_1], \alpha'[f_2], \alpha'[f_3]).$$

But by Lemma 4.6,

$$\alpha'[f_i] = \frac{df_i(\alpha)}{dt}.$$

Hence

$$F_*(\alpha'(t)) = \left(\frac{df_1(\alpha)}{dt}(t), \frac{df_2(\alpha)}{dt}(t), \frac{df_3(\alpha)}{dt}(t)\right)_{\beta(t)} = \beta'(t).$$ ◆

Let $\{U_j\}$ $(1 \le j \le n)$ and $\{\overline{U}_i\}$ $(1 \le i \le m)$ be the natural frame fields of $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively (Def. 2.4). Then:

7.8 Corollary *If* $F = (f_1, \ldots, f_m)$ *is a mapping from* $\mathbf{R}^n$ *to* $\mathbf{R}^m$, *then*

$$F_*(U_j(\mathbf{p})) = \sum_{i=1}^{m} \frac{\partial f_i}{\partial x_j}(\mathbf{p})\overline{U}_i(F(\mathbf{p})) \quad (1 \le j \le n).$$

Proof. This follows directly from Proposition 7.5, since $U_j[f_i] = \dfrac{\partial f_i}{\partial x_j}$. ◆

The matrix appearing in the preceding formula,

$$\left(\frac{\partial f_i}{\partial x_j}(\mathbf{p})\right)_{1 \le i \le m,\, 1 \le j \le n}$$

is called the *Jacobian matrix* of F at $\mathbf{p}$. (When $m = n = 1$; it reduces to a single number: the derivative of F at $\mathbf{p}$.)

Just as the derivative of a function is used to gain information about the function, the tangent map F_* can be used in the study of a mapping F.

7.9 Definition A mapping F: $\mathbf{R}^n \to \mathbf{R}^m$ is *regular* provided that at every point $\mathbf{p}$ of $\mathbf{R}^n$ the tangent map F_{*p} is one-to-one.

Since tangent maps are linear transformations, standard results of linear algebra show that the following conditions are equivalent:

(1) F_{*p} is one-to-one.
(2) $F_*(\mathbf{v}_p) = 0$ implies $\mathbf{v}_p = 0$.
(3) The Jacobian matrix of F at $\mathbf{p}$ has rank n, the dimension of the domain $\mathbf{R}^n$ of F.

The following noteworthy property of linear transformations $T: V \to W$ will be useful in dealing with tangent maps. If the vector spaces V and W have the same dimension, then T is one-to-one if and only if it is onto, so either property is equivalent to T being a linear isomorphism.

A mapping that has a (differentiable) inverse mapping is called a *diffeomorphism*. The results of this section all remain valid when Euclidean spaces $\mathbf{R}^n$ are replaced by open sets of Euclidean spaces, so we can speak of a diffeomorphism from one open set to another.

We state without proof one of the fundamental results of advanced calculus.

7.10 Theorem Let $F: \mathbf{R}^n \to \mathbf{R}^n$ be a mapping between Euclidean spaces of the same dimension. If F_{*p} is one-to-one at a point $\mathbf{p}$, there is an open set $\mathscr{U}$ containing $\mathbf{p}$ such that F restricted to $\mathscr{U}$ is a diffeomorphism of $\mathscr{U}$ onto an open set $\mathscr{V}$.

This is called the *inverse function theorem* since it asserts that the restricted mapping $\mathscr{U} \to \mathscr{V}$ has a differentiable inverse mapping $\mathscr{V} \to \mathscr{U}$. Exercise 6 gives a suggestion of its importance.

Exercises

In the first four exercises F denotes the mapping $F(u, v) = (u^2 - v^2, 2uv)$ in Example 7.3.

1. Find all points $\mathbf{p}$ such that

(a) $F(\mathbf{p}) = (0, 0)$. (b) $F(\mathbf{p}) = (8, 6)$.

(c) $F(\mathbf{p}) = \mathbf{p}$.

2. (a) Sketch the horizontal line $v = 1$ and its image under F (a parabola).

(b) Do the same for the vertical $u = 1$.

(c) Describe the image of the unit square $0 \leqq u, v \leqq 1$ under F.

3. Let $\mathbf{v} = (v_1, v_2)$ be a tangent vector to $\mathbf{R}^2$ at $\mathbf{p} = (p_1, p_2)$. Apply Definition 7.4 directly to express $F_*(\mathbf{v})$ in terms of the coordinates of $\mathbf{v}$ and $\mathbf{p}$.

4. Find a formula for the Jacobian matrix of F at all points, and deduce that F_{*p} is a linear isomorphism at every point of $\mathbf{R}^2$ except the origin.

5. If $F: \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation, prove that $F_*(\mathbf{v}_p) = F(\mathbf{v})_{F(p)}$.

6. (a) Give an example to demonstrate that a one-to-one and onto mapping need not be a diffeomorphism. (*Hint:* Take $m = n = 1$.)

(b) Prove that if a one-to-one and onto mapping F: $\mathbf{R}^n \to \mathbf{R}^n$ is regular, then it is a diffeomorphism.

7. Prove that a mapping F: $\mathbf{R}^n \to \mathbf{R}^m$ preserves directional derivatives in this sense: If $\mathbf{v}_p$ is a tangent vector to $\mathbf{R}^n$ and g is a differentiable function on $\mathbf{R}^m$, then $F_*(\mathbf{v}_p)[g] = \mathbf{v}_p[g(F)]$.

8. In the definition of tangent map (Def. 7.4), the straight line $t \to \mathbf{p} + t\mathbf{v}$ can be replaced by any curve α with initial velocity $\mathbf{v}_p$.

9. Let F: $\mathbf{R}^n \to \mathbf{R}^m$ and G: $\mathbf{R}^m \to \mathbf{R}^p$ be mappings. Prove:

(a) Their composition GF: $\mathbf{R}^n \to \mathbf{R}^p$ is a (differentiable) mapping. (Take $m = p = 2$ for simplicity.)

(b) $(GF)_* = G_*F_*$. (*Hint:* Use the preceding exercise.)

This concise formula is the general *chain rule*. Unless dimensions are small, it becomes formidable when expressed in terms of Jacobian matrices.

(c) If F is a diffeomorphism, then so is its inverse mapping F^{-1}.

10. Show (in two ways) that the map F: $\mathbf{R}^2 \to \mathbf{R}^2$ such that $F(u, v) = (ve^u, 2u)$ is a diffeomorphism:

(a) Prove that it is one-to-one, onto, and regular;

(b) Find a formula for its inverse F^{-1}: $\mathbf{R}^2 \to \mathbf{R}^2$ and observe that F^{-1} is differentiable. Verify the formula by checking that both $F\,F^{-1}$ and $F^{-1}\,F$ are identity maps.

1.8 Summary

Starting from the familiar notion of real-valued functions and using linear algebra at every stage, we have constructed a variety of mathematical objects. The basic notion of tangent vector led to vector fields, which dualized to 1-forms—which in turn led to arbitrary differential forms. The notions of curve and differentiable function were generalized to that of a mapping F: $\mathbf{R}^n \to \mathbf{R}^m$.

Then, starting from the usual notion of the derivative of a real-valued function, we proceeded to construct appropriate differentiation operations for these objects: the directional derivative of a function, the exterior derivative of a form, the velocity of a curve, the tangent map of a mapping. These operations all reduced to (ordinary or partial) derivatives of real-valued coordinate functions, but it is noteworthy that in most cases the *definitions* of these operations did not involve coordinates. (This could be achieved in all cases.) Generally speaking, these differentiation operations all exhibited in one form or another the characteristic linear and Leibnizian properties of ordinary differentiation.

Most of these concepts are probably already familiar to the reader, at least in special cases. But we now have careful definitions and a catalogue of basic properties that will enable us to begin the exploration of differential geometry.

Chapter 2
Frame Fields

Roughly speaking, geometry begins with the measurement of distances and angles. We shall see that the geometry of Euclidean space can be derived from the *dot product*, the natural inner product on Euclidean space.

Much of this chapter is devoted to the geometry of curves in $\mathbf{R}^3$. We emphasize this topic not only because of its intrinsic importance, but also because the basic method used to investigate curves has proved effective throughout differential geometry. A curve in $\mathbf{R}^3$ is studied by assigning at each point a certain *frame*—that is, set of three orthogonal unit vectors. The rate of change of these vectors along the curve is then expressed in terms of the vectors themselves by the celebrated *Frenet formulas* (Theorem 3.2). In a real sense, the theory of curves in $\mathbf{R}^3$ is merely a corollary of these fundamental formulas.

Later on we shall use this "method of moving frames" to study a *surface* in $\mathbf{R}^3$. The general idea is to think of a surface as a kind of two-dimensional curve and follow the Frenet approach as closely as possible. To carry out this scheme we shall need the generalization (Theorem 7.2) of the Frenet formulas devised by E. Cartan. It was Cartan who, in the early 1900s, first realized the full power of this method not only in differential geometry but also in a variety of related fields.

2.1 Dot Product

We begin by reviewing some basic facts about the natural inner product on the vector space $\mathbf{R}^3$.

1.1 Definition The dot product of points $\mathbf{p} = (p_1, p_2, p_3)$ and $\mathbf{q} = (q_1, q_2, q_3)$ in $\mathbf{R}^3$ is the number

$$\mathbf{p} \bullet \mathbf{q} = p_1q_1 + p_2q_2 + p_3q_3.$$

The dot product is an inner product since it has the following three properties:

(1) Bilinearity:

$$(a\mathbf{p} + b\mathbf{q}) \bullet \mathbf{r} = a\mathbf{p} \bullet \mathbf{r} + b\mathbf{q} \bullet \mathbf{r},$$

$$\mathbf{r} \bullet (a\mathbf{p} + b\mathbf{q}) = a\mathbf{r} \bullet \mathbf{p} + b\mathbf{r} \bullet \mathbf{q}.$$

(2) Symmetry: $\mathbf{p} \bullet \mathbf{q} = \mathbf{q} \bullet \mathbf{p}$.

(3) Positive definiteness: $\mathbf{p} \bullet \mathbf{p} \geqq 0$, and $\mathbf{p} \bullet \mathbf{p} = 0$ if and only if $\mathbf{p} = 0$.

(Here $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{r}$ are arbitrary points of $\mathbf{R}^3$, and a and b are numbers.)

The *norm* of a point $\mathbf{p} = (p_1, p_2, p_3)$ is the number

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

The norm is thus a real-valued function on $\mathbf{R}^3$; it has the fundamental properties $\|\mathbf{p} + \mathbf{q}\| \leqq \|\mathbf{p}\| + \|\mathbf{q}\|$ and $\|a\mathbf{p}\| = |a|\,\|\mathbf{p}\|$, where $|a|$ is the absolute value of the number a.

In terms of the norm we get a compact version of the usual distance formula in $\mathbf{R}^3$.

1.2 Definition If $\mathbf{p}$ and $\mathbf{q}$ are points of $\mathbf{R}^3$, the *Euclidean distance* from $\mathbf{p}$ to $\mathbf{q}$ is the number

$$d(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|.$$

In fact, since

$$\mathbf{p} - \mathbf{q} = (p_1 - q_1, p_2 - q_2, p_3 - q_3),$$

expansion of the norm gives the well-known formula (Fig. 2.1)

$$d(\mathbf{p}, \mathbf{q}) = \left((p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2\right)^{1/2}.$$

Euclidean distance may be used to give a more precise definition of open sets (Chapter 1, Section 1). First, if $\mathbf{p}$ is a point of $\mathbf{R}^3$ and $\varepsilon > 0$ is a number, the ε *neighborhood* $\mathcal{N}_\varepsilon$ of $\mathbf{p}$ in $\mathbf{R}^3$ is the set of all points $\mathbf{q}$ of $\mathbf{R}^3$ such that $d(\mathbf{p}, \mathbf{q}) < \varepsilon$. Then a subset $\mathcal{O}$ of $\mathbf{R}^3$ is *open* provided that each point of $\mathcal{O}$ has an ε neighborhood that is entirely contained in $\mathcal{O}$. In short, all points near enough to a point of an open set are also in the set. This definition is valid with $\mathbf{R}^3$ replaced by $\mathbf{R}^n$—or indeed any set furnished with a reasonable distance function.

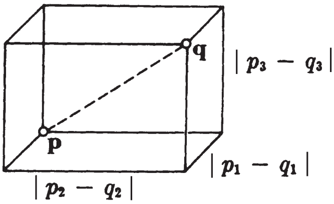


FIG. 2.1

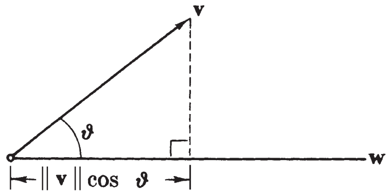


FIG. 2.2

We saw in Chapter 1 that for each point **p** of $\mathbf{R}^3$ there is a *canonical isomorphism* $\mathbf{v} \to \mathbf{v}_p$ from $\mathbf{R}^3$ onto the tangent space $\mathbf{T}_p(\mathbf{R}^3)$ at **p**. These isomorphisms lie at the heart of Euclidean geometry—using them, the dot product on $\mathbf{R}^3$ itself may be transferred to each of its tangent spaces.

1.3 Definition The *dot product* of tangent vectors $\mathbf{v}_p$ and $\mathbf{w}_p$ at the same point of $\mathbf{R}^3$ is the number $\mathbf{v}_p \bullet \mathbf{w}_p = \mathbf{v} \bullet \mathbf{w}$.

For example, $(1, 0, -1)_p \bullet (3, -3, 7)_p = 1(3) + 0(-3) + (-1)7 = -4$. Evidently this definition provides a dot product on each tangent space $T_p(\mathbf{R}^3)$ with the same properties as the original dot product on $\mathbf{R}^3$. In particular, each tangent vector $\mathbf{v}_p$ to $\mathbf{R}^3$ has *norm* (or *length*) $\| \mathbf{v}_p \| = \| \mathbf{v} \|$.

A fundamental result of linear algebra is the Schwarz inequality $| \mathbf{v} \bullet \mathbf{w} | \leqq \| \mathbf{v} \| \, \| \mathbf{w} \|$. This permits us to define the cosine of the angle ϑ between **v** and **w** by the equation (Fig. 2.2).

$$\mathbf{v} \bullet \mathbf{w} = \|\mathbf{v}\| \, \|\mathbf{w}\| \cos \vartheta.$$

Thus the dot product of two vectors is the product of their lengths times the cosine of the angle between them. (The angle ϑ is not uniquely determined unless further restrictions are imposed, say $0 \leqq \vartheta \leqq \pi$.)

In particular, if $\vartheta = \pi/2$, then $\mathbf{v} \bullet \mathbf{w} = 0$. Thus we shall define two vectors to be *orthogonal* provided their dot product is zero. A vector of length 1 is called a *unit vector.*

1.4 Definition A set $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ of three mutually orthogonal unit vectors tangent to $\mathbf{R}^3$ at **p** is called a *frame* at the point **p**.

Thus $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is a frame if and only if

$$\mathbf{e}_1 \bullet \mathbf{e}_1 = \mathbf{e}_2 \bullet \mathbf{e}_2 = \mathbf{e}_3 \bullet \mathbf{e}_3 = 1,$$

$$\mathbf{e}_1 \bullet \mathbf{e}_2 = \mathbf{e}_1 \bullet \mathbf{e}_3 = \mathbf{e}_2 \bullet \mathbf{e}_3 = 0.$$

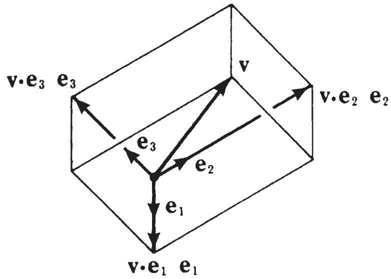


FIG. 2.3

By the symmetry of the dot product, the second row of equations is, of course, the same as

$$\mathbf{e}_2 \bullet \mathbf{e}_1 = \mathbf{e}_3 \bullet \mathbf{e}_1 = \mathbf{e}_3 \bullet \mathbf{e}_2 = 0.$$

Using index notation, all nine equations may be concisely expressed as $\mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij}$ for $1 \leqq i\ j \leqq 3$, where δ_{ij} is the Kronecker delta (0 if $i \neq j$, 1 if $i = j$). For example, at each point $\mathbf{p}$ of $\mathbf{R}^3$, the vectors $U_1(\mathbf{p})$, $U_2(\mathbf{p})$, $U_3(\mathbf{p})$ of Definition 2.4 in Chapter 1 constitute a frame at $\mathbf{p}$.

1.5 Theorem Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be a frame at a point $\mathbf{p}$ of $\mathbf{R}^3$. If $\mathbf{v}$ is any tangent vector to $\mathbf{R}^3$ at $\mathbf{p}$, then (Fig. 2.3)

$$\mathbf{v} = (\mathbf{v} \bullet \mathbf{e}_1)\mathbf{e}_1 + (\mathbf{v} \bullet \mathbf{e}_2)\mathbf{e}_2 + (\mathbf{v} \bullet \mathbf{e}_3)\mathbf{e}_3.$$

Proof. First we show that the vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are linearly independent. Suppose $\sum a_i\mathbf{e}_i = 0$. Then

$$0 = \left(\sum a_i\mathbf{e}_i\right) \bullet \mathbf{e}_j = \sum a_i\mathbf{e}_i \bullet \mathbf{e}_j = \sum a_i\delta_{ij} = a_j,$$

where all sums are over $i = 1, 2, 3$. Thus

$$a_1 = a_2 = a_3 = 0,$$

as required. Now, the tangent space $T_p(\mathbf{R}^3)$ has dimension 3, since it is linearly isomorphic to $\mathbf{R}^3$. Thus by a well-known theorem of linear algebra, the three independent vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ form a basis for $T_p(\mathbf{R}^3)$. Hence for each vector $\mathbf{v}$ there are three (unique) numbers c_1, c_2, c_3 such that

$$\mathbf{v} = \sum c_i\mathbf{e}_i.$$

But

$$\mathbf{v} \bullet \mathbf{e}_j = \left(\sum c_i\mathbf{e}_i\right) \bullet \mathbf{e}_j = \sum c_i\delta_{ij} = c_j,$$

and thus

$$\mathbf{v} = \sum (\mathbf{v} \bullet \mathbf{e}_j)\mathbf{e}_j. \qquad \blacklozenge$$

This result (valid in any inner-product space) is one of the great labor-saving devices in mathematics. For to find the coordinates of a vector **v** with respect to an *arbitrary* basis, one must in general solve a set of nonhomogeneous linear equations, a task that even in dimension 3 is not always entirely trivial. But the theorem shows that to find the coordinates of **v** with respect to a frame (that is, an *orthonormal* basis) it suffices merely to compute the three dot products $\mathbf{v} \bullet \mathbf{e}_1$, $\mathbf{v} \bullet \mathbf{e}_2$, $\mathbf{v} \bullet \mathbf{e}_3$. We call this process *orthonormal expansion* of **v** in terms of the frame $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$. In the special case of the natural frame $U_1(\mathbf{p})$, $U_2(\mathbf{p})$, $U_3(\mathbf{p})$, the identity

$$\mathbf{v} = (v_1, v_2, v_3) = \sum v_i U_i(\mathbf{p})$$

is an orthonormal expansion, and the dot product is defined in terms of these *Euclidean coordinates* by $\mathbf{v} \bullet \mathbf{w} = \sum v_i w_i$. If we use instead an arbitrary frame $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, then each vector **v** has new coordinates $a_i = \mathbf{v} \bullet \mathbf{e}_i$ relative to this frame, but *the dot product is still given by the same simple formula*

$$\mathbf{v} \bullet \mathbf{w} = \sum a_i b_i$$

since

$$\mathbf{v} \bullet \mathbf{w} = \left(\sum a_i \mathbf{e}_i\right) \bullet \left(\sum b_i \mathbf{e}_i\right) = \sum_{i,j} a_i b_j \mathbf{e}_i \bullet \mathbf{e}_j$$

$$= \sum_{i,j} a_i b_j \delta_{ij} = \sum a_i b_i.$$

When applied to more complicated geometric situations, the advantage of using frames becomes enormous, and this is why they appear so frequently throughout this book.

The notion of frame is very close to that of orthogonal matrix.

1.6 Definition Let $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ be a frame at a point **p** of $\mathbf{R}^3$. The 3×3 matrix A whose rows are the Euclidean coordinates of these three vectors is called the *attitude matrix* of the frame.

Explicitly, if

$$\mathbf{e}_1 = (a_{11}, a_{12}, a_{13})_p,$$

$$\mathbf{e}_2 = (a_{21}, a_{22}, a_{23})_p,$$

$$\mathbf{e}_3 = (a_{31}, a_{32}, a_{33})_p,$$

then

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Thus A does describe the "attitude" of the frame in $\mathbf{R}^3$, although not its point of application.

Evidently the rows of A are orthonormal, since

$$\sum_k a_{ik}a_{jk} = \mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij} \quad \text{for } 1 \leqq i,\ j \leqq 3.$$

By definition, this means that A is an *orthogonal* matrix.

In terms of matrix multiplication, these equations may be written $A\,{}^tA = I$, where I is the 3×3 identity matrix and ${}^tA$ is the *transpose* of A:

$${}^tA = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}.$$

It follows by a standard theorem of linear algebra that ${}^tAA = I$, so that ${}^tA = A^{-1}$, the *inverse* of A.

There is another product on $\mathbf{R}^3$, closely related to the wedge product of 1-forms and second in importance only to the dot product. We shall transfer it immediately to each tangent space of $\mathbf{R}^3$.

1.7 Definition If $\mathbf{v}$ and $\mathbf{w}$ are tangent vectors to $\mathbf{R}^3$ at the same point $\mathbf{p}$, then the *cross product* of $\mathbf{v}$ and $\mathbf{w}$ is the tangent vector

$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} U_1(\mathbf{p}) & U_2(\mathbf{p}) & U_3(\mathbf{p}) \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

This formal determinant is to be expanded along its first row. For example, if $\mathbf{v} = (1, 0, -1)_p$ and $\mathbf{w} = (2, 2, -7)_p$, then

$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} U_1(\mathbf{p}) & U_2(\mathbf{p}) & U_3(\mathbf{p}) \\ 1 & 0 & -1 \\ 2 & 2 & -7 \end{vmatrix}$$

$$= 2U_1(\mathbf{p}) + 5U_2(\mathbf{p}) + 2U_3(\mathbf{p}) = (2, 5, 2)_p.$$

Familiar properties of determinants show that the cross product $\mathbf{v} \times \mathbf{w}$ is *linear* in $\mathbf{v}$ and in $\mathbf{w}$, and satisfies the *alternation rule*

$$\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}.$$

Hence, in particular, $\mathbf{v} \times \mathbf{v} = 0$. The geometric usefulness of the cross product is based mostly on this fact:

1.8 Lemma The cross product $\mathbf{v} \times \mathbf{w}$ is orthogonal to both $\mathbf{v}$ and $\mathbf{w}$, and has length such that

$$\| \mathbf{v} \times \mathbf{w} \|^2 = (\mathbf{v} \bullet \mathbf{v})(\mathbf{w} \bullet \mathbf{w}) - (\mathbf{v} \bullet \mathbf{w})^2.$$

Proof. Let $\mathbf{v} \times \mathbf{w} = \sum c_i U_i(\mathbf{p})$. Then the dot product $\mathbf{v} \bullet (\mathbf{v} \times \mathbf{w})$ is just $\sum v_i c_i$. But by the definition of cross product, the Euclidean coordinates c_1, c_2, c_3 of $\mathbf{v} \times \mathbf{w}$ are such that

$$\mathbf{v} \bullet (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} v_1 & v_2 & v_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

This determinant is zero, since two of its rows are the same; thus $\mathbf{v} \times \mathbf{w}$ is orthogonal to $\mathbf{v}$, and similarly, to $\mathbf{w}$.

Rather than use tricks to prove the length formula, we give a brute-force computation. Now,

$$\begin{aligned}(\mathbf{v} \bullet \mathbf{v})(\mathbf{w} \bullet \mathbf{w}) - (\mathbf{v} \bullet \mathbf{w})^2 &= \left(\sum v_i^2\right)\left(\sum w_j^2\right) - \left(\sum v_i w_i\right)^2 \\ &= \sum_{i,j} v_i^2 w_j^2 - \left\{\sum v_i^2 w_i^2 + 2\sum_{i<j} v_i w_i v_j w_j\right\} \\ &= \sum_{i \neq j} v_i^2 w_j^2 - 2\sum_{i<j} v_i w_i v_j w_j.\end{aligned}$$

On the other hand,

$$\begin{aligned}\| \mathbf{v} \times \mathbf{w} \|^2 &= (\mathbf{v} \times \mathbf{w}) \bullet (\mathbf{v} \times \mathbf{w}) = \sum c_i^2 \\ &= (v_2 w_3 - v_3 w_2)^2 + (v_3 w_1 - v_1 w_3)^2 + (v_1 w_2 - v_2 w_1)^2,\end{aligned}$$

and expanding these squares gives the same result as above. ◆

A more intuitive description of the length of a cross product is

$$\| \mathbf{v} \times \mathbf{w} \| = \| \mathbf{v} \| \| \mathbf{w} \| \sin \vartheta,$$

where $0 \leqq \vartheta \leqq \pi$ is the smaller of the two angles from $\mathbf{v}$ to $\mathbf{w}$. The direction of $\mathbf{v} \times \mathbf{w}$ on the line orthogonal to $\mathbf{v}$ and $\mathbf{w}$ is given, for practical purposes, by this “right-hand rule”: If the fingers of the right hand point in the

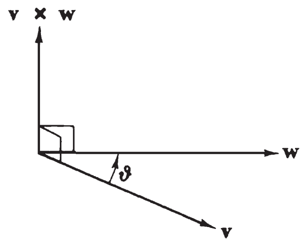


FIG. 2.4

direction of the shortest rotation of **v** to **w**, then the thumb points in the direction of $\mathbf{v} \times \mathbf{w}$ (Fig. 2.4).

Combining the dot and cross product, we get the *triple scalar product*, which assigns to any three vectors **u**, **v**, **w** the number $\mathbf{u} \bullet \mathbf{v} \times \mathbf{w}$ (Exercise 4). Parentheses are unnecessary: $\mathbf{u} \bullet (\mathbf{v} \times \mathbf{w})$ is the only possible meaning.

Exercises

1. Let $\mathbf{v} = (1, 2, -1)$ and $\mathbf{w} = (-1, 0, 3)$ be tangent vectors at a point of $\mathbf{R}^3$. Compute:

(a) $\mathbf{v} \bullet \mathbf{w}$. (b) $\mathbf{v} \times \mathbf{w}$.
(c) $\mathbf{v}/\|\mathbf{v}\|$, $\mathbf{w}/\|\mathbf{w}\|$. (d) $\|\mathbf{v} \times \mathbf{w}\|$.
(e) the cosine of the angle between **v** and **w**.

2. Prove that Euclidean distance has the properties
(a) $d(\mathbf{p}, \mathbf{q}) \geqq 0$; $d(\mathbf{p}, \mathbf{q}) = 0$ if and only if $\mathbf{p} = \mathbf{q}$,
(b) $d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p})$,
(c) $d(\mathbf{p}, \mathbf{q}) + d(\mathbf{q}, \mathbf{r}) \geqq d(\mathbf{p}, \mathbf{r})$, for any points **p**, **q**, **r** in $\mathbf{R}^3$.

3. Prove that the tangent vectors

$$\mathbf{e}_1 = \frac{(1, 2, 1)}{\sqrt{6}}, \quad \mathbf{e}_2 = \frac{(-2, 0, 2)}{\sqrt{8}}, \quad \mathbf{e}_3 = \frac{(1, -1, 1)}{\sqrt{3}}$$

constitute a frame. Express $\mathbf{v} = (6, 1, -1)$ as a linear combination of these vectors. (Check the result by direct computation.)

4. Let $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, $\mathbf{w} = (w_1, w_2, w_3)$. Prove that

(a) $$\mathbf{u} \bullet \mathbf{v} \times \mathbf{w} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

(b) $\mathbf{u} \bullet \mathbf{v} \times \mathbf{w} \neq 0$ if and only if **u**, **v**, and **w** are linearly independent.
(c) If any two vectors in $\mathbf{u} \bullet \mathbf{v} \times \mathbf{w}$ are reversed, the product changes sign.
(d) $\mathbf{u} \bullet \mathbf{v} \times \mathbf{w} = \mathbf{u} \times \mathbf{v} \bullet \mathbf{w}$.

5. Prove that $\mathbf{v} \times \mathbf{w} \neq 0$ if and only if **v** and **w** are linearly independent, and show that $\| \mathbf{v} \times \mathbf{w} \|$ is the area of the parallelogram with sides **v** and **w**.

6. If $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is a frame, show that

$$\mathbf{e}_1 \bullet \mathbf{e}_2 \times \mathbf{e}_3 = \pm 1.$$

Deduce that any 3×3 orthogonal matrix has determinant ±1.

7. If **u** is a unit vector, then the *component* of **v** in the **u** direction is

$$(\mathbf{v} \bullet \mathbf{u})\mathbf{u} = \| \mathbf{v} \| \cos \vartheta \, \mathbf{u}.$$

Show that **v** has a unique expression $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, where $\mathbf{v}_1 \bullet \mathbf{v}_2 = 0$ and $\mathbf{v}_1$ is the component of **v** in the **u** direction.

8. Prove: The volume of the parallelepiped with sides **u**, **v**, **w** is $\pm \mathbf{u} \bullet \mathbf{v} \times \mathbf{w}$ (Fig. 2.5). (*Hint:* Use the indicated unit vector $\mathbf{e} = \mathbf{v} \times \mathbf{w} / \| \mathbf{v} \times \mathbf{w} \|$.)

9. Prove, using ε-neighborhoods, that each of the following subsets of $\mathbf{R}^3$ is open:
(a) All points **p** such that $\| \mathbf{p} \| < 1$.
(b) All **p** such that $p_3 > 0$. (*Hint:* $| p_i - q_i | \leqq d(\mathbf{p}, \mathbf{q})$.)

10. In each case, let S be the set of all points **p** that satisfy the given condition. Describe S, and decide whether it is *open.*
(a) $p_1^2 + p_2^2 + p_3^2 = 1$.
(b) $p_3 \neq 0$.
(c) $p_1 = p_2 \neq p_3$.
(d) $p_1^2 + p_2^2 < 9$.

11. If f is a differentiable function on $\mathbf{R}^3$, show that the gradient

$$\nabla f = \sum \frac{\partial f}{\partial x_i} U_i$$

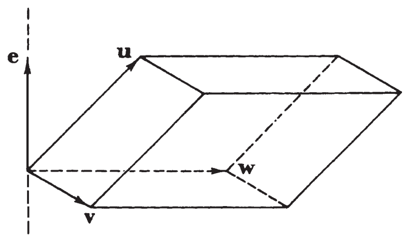


FIG. 2.5

(Ex. 8 of Sec. 1.6) has the following properties:

(a) $\mathbf{v}[f] = (df)(\mathbf{v}) = \mathbf{v} \bullet (\nabla f)(\mathbf{p})$ for any tangent vector at $\mathbf{p}$.

(b) The norm $\| (\nabla f)(\mathbf{p}) \| = \left[\sum (\partial f/\partial x_i)^2(\mathbf{p})\right]^{1/2}$ of (∇f) $(\mathbf{p})$ is the maximum of the directional derivatives $\mathbf{u}[f]$ for all *unit* vectors at $\mathbf{p}$. Furthermore, if $(\nabla f)(\mathbf{p}) \neq 0$, the unit vector for which the maximum occurs is

$$(\nabla f)(\mathbf{p})/\| (\nabla f)(\mathbf{p}) \|.$$

The notations grad f, curl V, and div V (in the exercise referred to) are often replaced by ∇f, $\nabla \times V$, and $\nabla \bullet V$, respectively.

12. *Angle functions.* Let f and g be differentiable real-valued functions on an interval I. Suppose that $f^2 + g^2 = 1$ and that ϑ_0 is a number such that $f(0) = \cos \vartheta_0$, $g(0) = \sin \vartheta_0$. If ϑ is the function such that

$$\vartheta(t) = \vartheta_0 + \int_0^t (fg' - gf')\,du,$$

prove that

$$f = \cos \vartheta, \quad g = \sin \vartheta.$$

Hint: We want $(f - \cos \vartheta)^2 + (g - \sin \vartheta)^2 = 0$, so show that its derivative is zero.

The point of this exercise is that ϑ is a differentiable function, unambiguously defined on the whole interval I.

2.2 Curves

We begin the geometric study of curves by reviewing some familiar definitions. Let $\alpha\colon I \to \mathbf{R}^3$ be a curve. In Chapter 1, Section 4, we defined the velocity vector $\alpha'(t)$ of α at t. Now we define the *speed* of α at t to be the length $v(t) = \| \alpha'(t) \|$ of the velocity vector. Thus speed is a real-valued function on the interval I. In terms of Euclidean coordinates $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, we have

$$\alpha'(t) = \left(\frac{d\alpha_1}{dt}(t), \frac{d\alpha_2}{dt}(t), \frac{d\alpha_3}{dt}(t)\right).$$

Hence the speed function v of α is given by the usual formula

$$v = \| \alpha' \| = \left(\left(\frac{d\alpha_1}{dt}\right)^2 + \left(\frac{d\alpha_2}{dt}\right)^2 + \left(\frac{d\alpha_3}{dt}\right)^2\right)^{1/2}.$$

In physics, the distance traveled by a moving point is determined by integrating its speed with respect to time. Thus we define the *arc length* of α from $t = a$ to $t = b$ to be the number

$$\int_a^b \| \alpha'(t) \| \, dt.$$

Substituting the formula for $\| \alpha' \|$ given above, we get the usual formula for arc length. This length involves only the restriction of α (defined on some open interval) to the *closed* interval $[a, b]$: $a \leqq t \leqq b$. Such a restriction $\sigma\colon [a, b] \to \mathbf{R}^3$ is called a *curve segment*, and its length is denoted by $L(\sigma)$. Note that the velocity of σ is well defined at the endpoints a and b of $[a, b]$.

Sometimes one is interested only in the route followed by a curve and not in the particular speed at which it traverses its route. One way to ignore the speed of a curve α is to reparametrize to a curve β that has *unit speed* $\| \beta' \| = 1$. Then β represents a "standard trip" along the route of α.

2.1 Theorem If α is a regular curve in $\mathbf{R}^3$, then there exists a reparametrization β of α such that β has unit speed.

Proof. Fix a number a in the domain I of $\alpha\colon I \to \mathbf{R}^3$, and consider the *arc length function*

$$s(t) = \int_a^b \| \alpha'(u) \| \, du.$$

(The resulting reparametrization is said to be *based at* $t = a$.) Thus the derivative ds/dt of the function $s = s(t)$ is the speed function $v = \| \alpha' \|$ of α. Since α is regular, by definition α' is never zero; hence $ds/dt > 0$. By a standard theorem of calculus, the function s has an inverse function $t = t(s)$, whose derivative dt/ds at $s = s(t)$ is the reciprocal of ds/dt at $t = t(s)$. In particular, $dt/ds > 0$.

Now let β be the reparametrization $\beta(s) = \alpha(t(s))$ of α. We assert that β has unit speed. In fact, by Lemma 4.5 of Chapter 1,

$$\beta'(s) = \frac{dt}{ds}(s)\alpha'(t(s)).$$

Hence, by the preceding remarks, the speed of β is

$$\| \beta'(s) \| = \frac{dt}{ds}(s) \| \alpha'(t(s)) \| = \frac{dt}{ds}(s)\frac{ds}{dt}(t(s)) = 1. \qquad \blacklozenge$$

We shall use the notation of this proof frequently in later work. The unit-speed curve β is sometimes said to have *arc-length parametrization*, since the arc length of β from $s = a$ to $s = b$ $(a < b)$ is just $b - a$.

For example, consider the helix α in Example 4.2 of Chapter 1. Since $\alpha(t) = (a\cos t, a\sin t, bt)$, the velocity α' is given by the formula

$$\alpha'(t) = (-a \sin t, a \cos t, b).$$

Hence

$$\| \alpha'(t) \|^2 = \alpha'(t) \bullet \alpha'(t) = a^2 \sin^2 t + a^2 \cos^2 t + b^2 = a^2 + b^2.$$

Thus α has *constant* speed $c = \| \alpha' \| = (a^2 + b^2)^{1/2}$. If we measure arc length from $t = 0$, then

$$s(t) = \int_0^t c \, du = ct.$$

Hence, $t(s) = s/c$. Substituting in the formula for α, we get the unit-speed reparametrization

$$\beta(s) = \alpha\left(\frac{s}{c}\right) = \left(a \cos \frac{s}{c}, a \sin \frac{s}{c}, \frac{bs}{c}\right).$$

It is easy to check directly that $\| \beta'(s) \| = 1$ for all s.

A reparametrization $\alpha(h)$ of a curve α is *orientation-preserving* if $h' \geq 0$ and *orientation-reversing* if $h' \leq 0$. In the latter case, $\alpha(h)$ still follows the route of α but in the opposite direction. By definition, a unit-speed reparametrization is always orientation-preserving since $ds/dt > 0$ for a regular curve.

In the *theory* of curves we will frequently reparametrize regular curves to obtain unit speed; however, it is rarely possible to do this in practice. The problem is basic calculus: Even when the coordinate functions of the curve are rather simple, the speed function cannot usually be integrated explicitly—at least in terms of familiar functions.

The general notion of vector field (Definition 2.3 of Chapter 1) can be adapted to curves as follows.

2.2 Definition A *vector field Y on curve* $\alpha: I \to \mathbf{R}^3$ is a function that assigns to each number t in I a tangent vector $Y(t)$ to $\mathbf{R}^3$ at the point $\alpha(t)$.

We have already met such vector fields: For any curve α, its velocity α' evidently satisfies this definition. Note that unlike α', arbitrary vector fields on α need not be tangent to α, but may point in any direction (Fig. 2.6).

The properties of vector fields on curves are analogous to those of vector fields on $\mathbf{R}^3$. For example, if Y is a vector field on $\alpha: I \to \mathbf{R}^3$, then for each t in I we can write

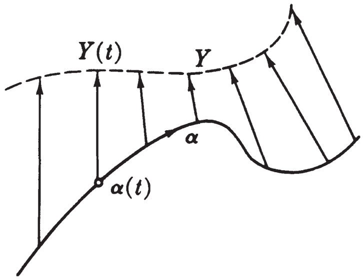


FIG. 2.6

$$Y(t) = (y_1(t), y_2(t), y_3(t))_{\alpha(t)} = \sum y_i(t)U_i(\alpha(t)).$$

We have thus defined real-valued functions y_1, y_2, y_3 on I, called the *Euclidean coordinate functions* of Y. These will always be assumed to be differentiable. Note that the composite function $t \to U_i(\alpha(t))$ is a vector field on α. Where it seems safe to do so, we shall often write merely U_i instead of $U_i(\alpha(t))$.

The operations of addition, scalar multiplication, dot product, and cross product of vector fields (on the same curve) are all defined in the usual pointwise fashion. Thus if

$$Y(t) = t^2U_1 - tU_3, \quad Z(t) = (1 - t^2)U_2 + tU_3,$$

and $f(t) = (t + 1)/t$, we obtain the vector fields

$$(Y + Z)(t) = t^2U_1 + (1 - t^2)U_2,$$

$$(fY)(t) = t(t + 1)U_1 - (t + 1)U_3,$$

$$(Y \times Z)(t) = \begin{vmatrix} U_1 & U_2 & U_3 \\ t^2 & 0 & -t \\ 0 & 1 - t^2 & t \end{vmatrix}$$

$$= t(1 - t^2)U_1 - t^3U_2 + t^2(1 - t^2)U_3$$

and the real-valued function

$$(Y \bullet Z)(t) = -t^2.$$

To differentiate a vector field on α *one simply differentiates its Euclidean coordinate functions*, thus obtaining a new vector field on α. Explicitly, if $Y = \sum y_iU_i$, then $Y' = \sum \frac{dy_i}{dt}U_i$. Thus, for Y as above, we get

$$Y' = 2tU_1 - U_3, \quad Y'' = 2U_1, \quad \text{and} \quad Y''' = 0.$$

In particular, the derivative α'' of the velocity α' of α is called the *acceleration* of α. Thus if $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, the acceleration α'' is the vector field

$$\alpha'' = \left(\frac{d^2\alpha_1}{dt^2}, \frac{d^2\alpha_2}{dt^2}, \frac{d^2\alpha_3}{dt^2}\right)_\alpha$$

on α. By contrast with velocity, acceleration is generally not tangent to the curve.

As we mentioned earlier, in whatever form it appears, differentiation always has suitable linearity and Leibnizian properties. In the case of vector fields on a curve, it is easy to prove the linearity property

$$(aY + bZ)' = aY' + bZ'$$

(a and b numbers) and the Leibnizian properties

$$(fY)' = \frac{df}{dt}Y + fY' \quad \text{and} \quad (Y \bullet Z)' = Y' \bullet Z + Y \bullet Z'.$$

If the function $Y \bullet Z$ is constant, the last formula shows that

$$Y' \bullet Z + Y \bullet Z' = 0.$$

This observation will be used frequently in later work. In particular, if Y has constant length $\| Y \|$, then Y and Y' are orthogonal at each point, since $\| Y \|^2 = Y \bullet Y$ constant implies $2Y \bullet Y' = 0$.

Recall that tangent vectors are parallel if they have the same vector parts. We say that a vector field Y on a curve is *parallel* provided all its (tangent vector) values are parallel. In this case, if the common vector part is (c_1, c_2, c_3), then

$$Y(t) = (c_1, c_2, c_3)_{\alpha(t)} = \sum c_i U_i \quad \text{for all } t.$$

Thus parallelism for a vector field is equivalent to the constancy of its Euclidean coordinate functions.

Vanishing of derivatives is always important in calculus; here are three simple cases.

2.3 Lemma (1) A curve α is constant if and only if its velocity is zero, $\alpha' = 0$.

(2) A nonconstant curve α is a straight line if and only if its acceleration is zero, $\alpha'' = 0$.

(3) A vector field Y on a curve is parallel if and only if its derivative is zero, $Y' = 0$.

Proof. In each case it suffices to look at the Euclidean coordinate functions. For example, we shall prove (2). If $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, then

$$\alpha'' = \left(\frac{d^2\alpha_1}{dt^2}, \frac{d^2\alpha_2}{dt^2}, \frac{d^2\alpha_3}{dt^2} \right).$$

Thus $\alpha'' = 0$ if and only if each $d^2\alpha_i/dt^2 = 0$. By elementary calculus, this is equivalent to the existence of constants p_i and q_i such that

$$\alpha_i(t) = p_i + tq_i, \quad \text{for } i = 1, 2, 3.$$

Thus $\alpha(t) = \mathbf{p} + t\mathbf{q}$, and α is a straight line as defined in Example 4.2 of Chapter 1. (Note that nonconstancy implies $\mathbf{q} \neq 0$.) ◆

Exercises

1. For the curve $\alpha(t) = (2t, t^2, t^3/3)$,
(a) find the velocity, speed, and acceleration for arbitrary t, and at $t = 1$;
(b) find the arc length function $s = s(t)$ (based at $t = 0$), and determine the arc length of α from $t = -1$ to $t = +1$.

2. Show that a curve has constant speed if and only if its acceleration is everywhere orthogonal to its velocity.

3. Show that the curve $\alpha(t) = (\cosh t, \sinh t, t)$ has arc length function $s(t) = \sqrt{2} \sinh t$, and find a unit-speed reparametrization of α.

4. Consider the curve $\alpha(t) = (2t, t^2, \log t)$ on I: $t > 0$. Show that this curve passes through the points $\mathbf{p} = (2, 1, 0)$ and $\mathbf{q} = (4, 4, \log 2)$, and find its arc length between these points.

5. Suppose that β_1 and β_2 are unit-speed reparametrizations of the same curve α. Show that there is a number s_0 such that $\beta_2(s) = \beta_1(s + s_0)$ for all s. What is the geometric significance of s_0?

6. Let Y be a vector field on the helix $\alpha(t) = (\cos t, \sin t, t)$. In each of the following cases, express Y in the form $\sum y_i U_i$:
(a) $Y(t)$ is the vector from $\alpha(t)$ to the origin of $\mathbf{R}^3$.
(b) $Y(t) = \alpha'(t) - \alpha''(t)$.
(c) $Y(t)$ has unit length and is orthogonal to both $\alpha'(t)$ and $\alpha''(t)$.
(d) $Y(t)$ is the vector from $\alpha(t)$ to $\alpha(t + \pi)$.

7. A reparametrization $\alpha(h)$: $[c, d] \to \mathbf{R}^3$ of a curve segment α: $[a, b] \to \mathbf{R}^3$ is *monotone* provided either

(i) $h' \geq 0$, $h(c) = a$, $h(d) = b$ or (ii) $h' \leqq 0$, $h(c) = b$, $h(d) = a$.

Prove that monotone reparametrization does not change arc length.

8. Let Y be a vector field on a curve α. If $\alpha(h)$ is a reparametrization of α, show that the reparametrization $Y(h)$ is a vector field on $\alpha(h)$, and prove the chain rule $Y(h)' = h'Y'(h)$.

9. (*Numerical integration.*) The curve segments

$$\alpha(t) = (\sin t,\ t^2 \cos t,\ \sin 2t), \quad \beta(t) = (t^2 \sin t, t^2, t^2(1 + \cos t)),$$

defined on $0 \leqq t \leqq \pi$, run from the origin 0 to $(0, \pi^2, 0)$. Which is shorter? (See Integration in the Appendix.)

10. Let $\alpha, \beta\colon I \to \mathbf{R}^3$ be curves such that $\alpha'(t)$ and $\beta'(t)$ are parallel (same Euclidean coordinates) at each t. Prove that α and β are *parallel* in the sense that there is a point $\mathbf{p}$ in $\mathbf{R}^3$ such that $\beta(t) = \alpha(t) + \mathbf{p}$ for all t.

11. Prove that a straight line is the shortest distance between two points in $\mathbf{R}^3$. Use the following scheme; let $\alpha\colon [a, b] \to \mathbf{R}^3$ be an arbitrary curve segment from $\mathbf{p} = \alpha(a)$ to $\mathbf{q} = \alpha(b)$. Let $\mathbf{u} = (\mathbf{q} - \mathbf{p})/\| \mathbf{q} - \mathbf{p} \|$.

(a) If σ is a straight line segment from $\mathbf{p}$ to $\mathbf{q}$, say

$$\sigma(t) = (1 - t)\mathbf{p} + t\mathbf{q} \quad (0 \leqq t \leqq 1),$$

show that $L(\sigma) = d(\mathbf{p}, \mathbf{q})$.

(b) From $\| \alpha' \| \geqq \alpha' \bullet \mathbf{u}$, deduce $L(\alpha) \geqq d(\mathbf{p}, \mathbf{q})$, where $L(\alpha)$ is the length of α and d is Euclidean distance.

(c) Furthermore, show that if $L(\alpha) = d(\mathbf{p}, \mathbf{q})$, then (but for parametrization) α is a straight line segment. (*Hint:* write $\alpha' = (\alpha' \bullet \mathbf{u})\mathbf{u} + Y$, where $Y \bullet \mathbf{u} = 0$.)

2.3 The Frenet Formulas

We now derive mathematical measurements of the turning and twisting of a curve in $\mathbf{R}^3$. Throughout this section we deal only with *unit-speed* curves; in the next we extend the results to arbitrary regular curves.

Let $\beta\colon I \to \mathbf{R}^3$ be a unit-speed curve, so $\| \beta'(s) \| = 1$ for each s in I. Then $T = \beta'$ is called the *unit tangent* vector field on β. Since T has constant length 1, its derivative $T' = \beta''$ measures the way the curve is turning in $\mathbf{R}^3$. We call T' the *curvature* vector field of β. Differentiation of $T \bullet T = 1$ gives $2T' \bullet T = 0$, so T' is always orthogonal to T, that is, *normal* to β.

The length of the curvature vector field T' gives a numerical measurement of the turning of β. The real-valued function κ such that $\kappa(s) = \| T'(s) \|$ for

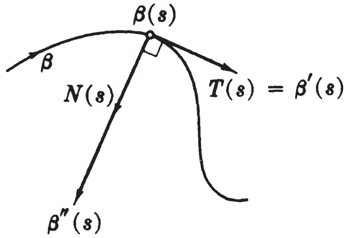


FIG. 2.7

all s in I is called the *curvature* function of β. Thus $\kappa \geqq 0$, and the larger κ is, the sharper the turning of β.

To carry this analysis further, we impose the restriction that κ is never zero so $\kappa > 0$. The unit-vector field $N = T'/\kappa$ on β then tells the *direction* in which β is turning at each point. N is called the *principal normal* vector field of β (Fig. 2.7). The vector field $B = T \times N$ on β is called the *binormal* vector field of β.

3.1 Lemma Let β be a unit-speed curve in $\mathbf{R}^3$ with $\kappa > 0$. Then the three vector fields T, N, and B on β are unit vector fields that are mutually orthogonal at each point. We call T, N, B the *Frenet frame field* on β.

Proof. By definition $\| T \| = 1$. Since $\kappa = \| T' \| > 0$,

$$\|N\| = (1/\kappa)\,\|T'\| = 1.$$

We saw above that T and N are orthogonal—that is, $T \bullet N = 0$. Then by applying Lemma 1.8 at each point, we conclude that $\| B \| = 1$, and B is orthogonal to both T and N. ◆

In summary, we have $T = \beta'$, $N = T'/\kappa$, and $B = T \times N$, satisfying $T \bullet T = N \bullet N = B \bullet B = 1$, with all other dot products zero.

The key to the successful study of the geometry of a curve β is to use its Frenet frame field T, N, B whenever possible, instead of the natural frame field U_1, U_2, U_3. The Frenet frame field of β is full of information about β, whereas the natural frame field contains none at all.

The first and most important use of this idea is to express the *derivatives* T', N', B' *in terms of* T, N, B. Since $T = \beta'$, we have $T' = \beta'' = \kappa N$. Next consider B'. We claim that B' is, at each point, a scalar multiple of N. To prove this, it suffices by orthonormal expansion to show that $B' \bullet B = 0$ and $B' \bullet T = 0$. The former holds since B is a unit vector. To prove the latter, differentiate $B \bullet T = 0$, obtaining $B' \bullet T + B \bullet T' = 0$; then

$$B' \bullet T = -B \bullet T' = -B \bullet \kappa N = 0.$$

Thus we can now define the *torsion* function τ of the curve β to be the real-valued function on the interval I such that $B' = -\tau N$. (The minus sign is traditional.) By contrast with curvature, there is no restriction on the values of τ—it may be positive, negative, or zero at various points of I. We shall presently show that τ does measure the torsion, or twisting, of the curve β.

3.2 Theorem (Frenet formulas). If β: $I \to \mathbf{R}^3$ is a unit-speed curve with curvature $\kappa > 0$ and torsion τ, then

$$\begin{aligned} T' &= \kappa N, \\ N' &= -\kappa T + \tau B, \\ B' &= -\tau N. \end{aligned}$$

Proof. As we saw above, the first and third formulas are essentially just the definitions of curvature and torsion. To prove the second, we use orthonormal expansion to express N' in terms of T, N, B:

$$N' = N' \bullet T\, T + N' \bullet N\, N + N' \bullet B\, B.$$

These coefficients are easily found. Differentiating $N \bullet T = 0$, we get $N' \bullet T + N \bullet T' = 0$; hence

$$N' \bullet T = -N \bullet T' = -N \bullet \kappa N = -\kappa.$$

As usual, $N' \bullet N = 0$, since N is a unit vector field. Finally,

$$N' \bullet B = -N \bullet B' = -N \bullet (-\tau N) = \tau. \qquad \blacklozenge$$

3.3 Example We compute the Frenet frame T, N, B and the curvature and torsion functions of the unit-speed helix

$$\beta(s) = \left(a \cos \frac{s}{c}, a \sin \frac{s}{c}, \frac{bs}{c} \right),$$

where $c = (a^2 + b^2)^{1/2}$ and $a > 0$. Now

$$T(s) = \beta'(s) = \left(-\frac{a}{c} \sin \frac{s}{c}, \frac{a}{c} \cos \frac{s}{c}, \frac{b}{c} \right).$$

Hence

$$T'(s) = \left(-\frac{a}{c^2} \cos \frac{s}{c}, -\frac{a}{c^2} \sin \frac{s}{c}, 0 \right).$$

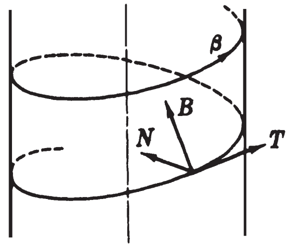


FIG. 2.8

Thus

$$\kappa(s) = \|T'(s)\| = \frac{a}{c^2} = \frac{a}{a^2 + b^2} > 0.$$

Since $T' = \kappa N$, we get

$$N(s) = \left(-\cos\frac{s}{c}, -\sin\frac{s}{c}, 0 \right).$$

Note that regardless of what values a and b have, N always points straight in toward the axis of the cylinder on which β lies (Fig. 2.8).

Applying the definition of cross product to $B = T \times N$ gives

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

It remains to compute torsion. Now,

$$B'(s) = \left(\frac{b}{c^2}\cos\frac{s}{c}, \frac{b}{c^2}\sin\frac{s}{c}, 0 \right),$$

and by definition, $B' = -\tau N$. Comparing the formulas for B' and N, we conclude that

$$\tau(s) = \frac{b}{c^2} = \frac{b}{a^2 + b^2}.$$

So the torsion of the helix is also constant.

Note that when the parameter b is zero, the helix reduces to a circle of radius a. The curvature of this circle is $\kappa = 1/a$ (so the smaller the radius, the larger the curvature), and the torsion is identically zero.

This example is a very special one—in general (as the examples in the exercises show) neither the curvature nor the torsion functions of a curve need be constant.

3.4 Remark We have emphasized all along the distinction between a tangent vector and a point of $\mathbf{R}^3$. However, Euclidean space has, as we have seen, the remarkable property that given a point $\mathbf{p}$, there is a natural one-to-one correspondence between points (v_1, v_2, v_3) and tangent vectors $(v_1, v_2, v_3)_p$ at $\mathbf{p}$. Thus one can transform points into tangent vectors (and vice versa) by means of this canonical isomorphism. In the next two sections particularly, it will often be convenient to switch quietly from one to the other without change of notation. Since *corresponding objects have the same Euclidean coordinates*, this switching can have no effect on scalar multiplication, addition, dot products, differentiation, or any other operation defined in terms of Euclidean coordinates.

Thus a vector field $Y = (y_1, y_2, y_3)_\beta$ on a curve β becomes itself a curve (y_1, y_2, y_3) in $\mathbf{R}^3$. In particular, if Y is parallel, its Euclidean coordinate functions are constant, so Y is identified with a single point of $\mathbf{R}^3$.

A *plane* in $\mathbf{R}^3$ can be described as the union of all the perpendiculars to a given line at a given point. In vector language then, the *plane through* $\mathbf{p}$ *orthogonal to* $\mathbf{q} \neq 0$ consists of all points $\mathbf{r}$ in $\mathbf{R}^3$ such that $(\mathbf{r} - \mathbf{p}) \bullet \mathbf{q} = 0$. By the remark above, we may picture $\mathbf{q}$ as a tangent vector at $\mathbf{p}$ as shown in Fig. 2.9.

We can now give an informative approximation of a given curve near an arbitrary point on the curve. The goal is to show how curvature and torsion influence the shape of the curve. To derive this approximation we use a Taylor approximation of the curve—and express this in terms of the Frenet frame at the selected point.

For simplicity, we shall consider the unit-speed curve $\beta = (\beta_1, \beta_2, \beta_3)$ near the point $\beta(0)$. For s small, each coordinate $\beta_i(s)$ is closely approximated by the initial terms of its Taylor series:

$$\beta_i(s) \sim \beta_i(0) + \frac{d\beta_i}{ds}(0)s + \frac{d^2\beta_i}{ds^2}(0)\frac{s^2}{2} + \frac{d^3\beta_i}{ds^3}(0)\frac{s^3}{6}.$$

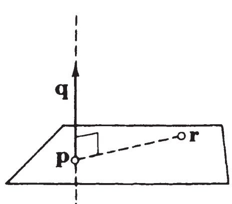


FIG. 2.9

Hence

$$\beta(s) \sim \beta(0) + s\beta'(0) + \frac{s^2}{2}\beta''(0) + \frac{s^3}{6}\beta'''(0).$$

But $\beta'(0) = T_0$, and $\beta''(0) = \kappa_0 N_0$, where the subscript indicates evaluation at $s = 0$, and we assume $\kappa_0 \neq 0$. Now

$$\beta''' = (\kappa N)' = \frac{d\kappa}{ds}N + \kappa N'.$$

Thus by the Frenet formula for N', we get

$$\beta'''(0) = -\kappa_0^2 T_0 + \frac{d\kappa}{ds}(0)N_0 + \kappa_0\tau_0 B_0.$$

Finally, substitute these derivatives into the approximation of $\beta(s)$ given above, and keep only the dominant term in each component (that is, the one containing the smallest power of s). The result is

$$\beta(s) \sim \beta(0) + sT_0 + \kappa_0\frac{s^2}{2}N_0 + \kappa_0\tau_0\frac{s^3}{6}B_0.$$

Denoting the right side by $\hat{\beta}(s)$, we obtain a curve $\hat{\beta}$ called the *Frenet approximation* of β near $s = 0$. We emphasize that β has a different Frenet approximation near each of its points; if 0 is replaced by an arbitrary number s_0, then s is replaced by $s - s_0$, as usual in Taylor expansions.

Let us now examine the Frenet approximation given above. The first term in the expression for $\hat{\beta}$ is just the point $\beta(0)$. The first two terms give the *tangent line* $s \to \beta(0) + sT_0$ of β at $\beta(0)$—the best linear approximation of β near $\beta(0)$. The first three terms give the parabola

$$s \to \beta(0) + sT_0 + \kappa_0(s^2/2)N_0,$$

which is the best quadratic approximation of β near $\beta(0)$. Note that this parabola lies in the plane through $\beta(0)$ orthogonal to B_0, the *osculating plane* of β at $\beta(0)$. This parabola has the same shape as the parabola $y = \kappa_0 x^2/2$ in the xy plane, and is completely determined by the curvature κ_0 of β at $s = 0$.

Finally, the torsion τ_0, which appears in the last and smallest term of $\hat{\beta}$, controls the motion of β orthogonal to its osculating plane at $\beta(0)$, as shown in Fig. 2.10.

On the basis of this discussion, it is a reasonable guess that *if a unit-speed curve has curvature identically zero, then it is a straight line*. In fact, this follows immediately from (2) of Lemma 2.3, since $\kappa = \| T' \| = \| \beta'' \|$, so that $\kappa = 0$ if and only if $\beta'' = 0$. Thus curvature does measure deviation from straightness.

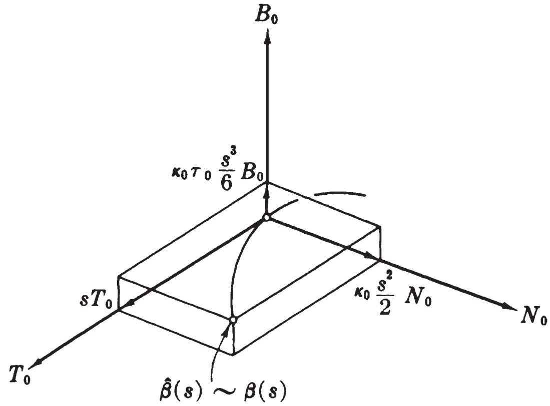


FIG. 2.10

A *plane curve* in $\mathbf{R}^3$ is a curve that lies in a single plane of $\mathbf{R}^3$. Evidently a plane curve does not twist in as interesting a way as even the simple helix in Example 3.3. The discussion above shows that for s small the curve β tends to stay in its osculating plane at $\beta(0)$; it is $\tau_0 \neq 0$ that causes β to twist out of the osculating plane. Thus if the torsion of β is identically zero, we may well suspect that β never leaves this plane.

3.5 Corollary Let β be a unit-speed curve in $\mathbf{R}^3$ with $\kappa > 0$. Then β is a plane curve if and only if $\tau = 0$.

Proof. Suppose β is a plane curve. Then by the remarks above, there exist points $\mathbf{p}$ and $\mathbf{q}$ such that $(\beta(s) - \mathbf{p}) \bullet \mathbf{q} = 0$ for all s. Differentiation yields

$$\beta'(s) \bullet \mathbf{q} = \beta''(s) \bullet \mathbf{q} = 0 \quad \text{for all } s.$$

Thus $\mathbf{q}$ is always orthogonal to $T = \beta'$ and $N = \beta''/\kappa$. But B is also orthogonal to T and N, so, since B has unit length, $B = \pm\mathbf{q}/\|\mathbf{q}\|$. Thus $B' = 0$, and by definition $\tau = 0$ (Fig. 2.11).

Conversely, suppose $\tau = 0$. Thus $B' = 0$; that is, B is parallel and may thus be identified (by Remark 3.4) with a *point* of $\mathbf{R}^3$. We assert that β lies in the plane through $\beta(0)$ orthogonal to B. To prove this, consider the real-valued function

$$f(s) = (\beta(s) - \beta(0)) \bullet B \quad \text{for all } s.$$

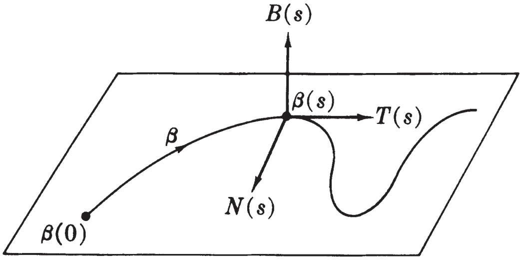


FIG. 2.11

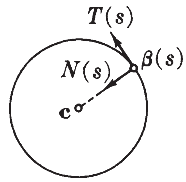


FIG. 2.12

Then

$$\frac{df}{ds} = \beta' \bullet B = T \bullet B = 0.$$

But obviously, $f(0) = 0$, so f is identically zero. Thus

$$(\beta(s) - \beta(0)) \bullet B \quad \text{for all } s,$$

which shows that β lies entirely in this plane orthogonal to the (parallel) binormal of β. ◆

We saw at the end of Example 3.3 that a circle of radius a has curvature $1/a$ and torsion zero. Furthermore, the formula given there for the principal normal shows that for a circle, N always points toward its center. This suggests how to prove the following converse.

3.6 Lemma If β is a unit-speed curve with constant curvature $\kappa > 0$ and torsion zero, then β is part of a circle of radius $1/\kappa$.

Proof. Since $\tau = 0$, β is a plane curve. What we must now show is that every point of β is at distance $1/\kappa$ from some fixed point—which will thus be the center of the circle. Consider the curve $\gamma = \beta + (1/\kappa)N$. Using the hypothesis on β, and (as usual) a Frenet formula, we find

$$\gamma' = \beta' + \frac{1}{\kappa}N' = T + \frac{1}{\kappa}(-\kappa T) = 0.$$

Hence the curve γ is constant; that is, $\beta(s) + (1/\kappa)N(s)$ has the same value, say $\mathbf{c}$, for all s (see Fig. 2.12). But the distance from $\mathbf{c}$ to $\beta(s)$ is

$$d(\mathbf{c}, \beta(s)) = \|\mathbf{c} - \beta(s)\| = \left\|\frac{1}{\kappa}N(s)\right\| = \frac{1}{\kappa}.$$ ◆

In principle, every geometric problem about curves can be solved by means of the Frenet formulas. In simple cases it may be just enough to record the data of the problem in convenient form, differentiate, and use the Frenet formulas. For example, suppose β is a unit-speed curve that lies entirely in the sphere Σ of radius a centered at the origin of $\mathbf{R}^3$. To stay in the sphere, β must curve; in fact it is a reasonable guess that the minimum possible curvature occurs when β is on a great circle of Σ. Such a circle has radius a, so we conjecture that *a spherical curve β has curvature $\kappa \geqq 1/a$, where a is the radius of its sphere*.

To prove this, observe that since every point of Σ has distance a from the origin, we have $\beta \bullet \beta = a^2$. Differentiation yields $2\beta' \bullet \beta = 0$, that is, $\beta \bullet T = 0$. Another differentiation gives $\beta' \bullet T + \beta \bullet T' = 0$, and by using a Frenet formula we get $T \bullet T + \kappa\beta \bullet N = 0$; hence

$$\kappa\beta \bullet N = -1.$$

By the Schwarz inequality,

$$|\beta \bullet N| \leqq \|\beta\| \, \|N\| = a,$$

and since $\kappa \geqq 0$ we obtain the required result:

$$\kappa = |\kappa| = \frac{1}{|\beta \bullet N|} \geqq \frac{1}{a}.$$

Continuation of this procedure leads to a necessary and sufficient condition (expressed in terms of curvature and torsion) for a curve to be *spherical*, that is, lie on some sphere in $\mathbf{R}^3$ (Exercise 10).

Exercises

1. Compute the *Frenet apparatus* κ, τ, T, N, B of the unit-speed curve $\beta(s) = (\tfrac{4}{5}\cos s,\ 1 - \sin s,\ -\tfrac{3}{5}\cos s)$. Show that this curve is a circle; find its center and radius.

2. Consider the curve

$$\beta(s) = \left(\frac{(1+s)^{3/2}}{3}, \frac{(1-s)^{3/2}}{3}, \frac{s}{\sqrt{2}} \right)$$

defined on I: $-1 < s < 1$. Show that β has unit speed, and compute its Frenet apparatus.

3. For the helix in Example 3.3, check the Frenet formulas by direct substitution of the computed values of κ, τ, T, N, B.

4. Prove that

$$\begin{aligned} T &= N \times B = -B \times N, \\ N &= B \times T = -T \times B, \\ B &= T \times N = -N \times T. \end{aligned}$$

(A formal proof uses properties of the cross product established in the Exercises of Section 1—but one can recall these formulas by using the right-hand rule given at the end of that section.)

5. If A is the vector field $\tau T + \kappa B$ on a unit-speed curve β, show that the Frenet formulas become

$$\begin{aligned} T' &= A \times T, \\ N' &= A \times N, \\ B' &= A \times B. \end{aligned}$$

6. A unit-speed parametrization of a circle may be written

$$\gamma(s) = \mathbf{c} + r \cos\frac{s}{r}\,\mathbf{e}_1 + r \sin\frac{s}{r}\,\mathbf{e}_2,$$

where $\mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij}$.

If β is a unit-speed curve with $\kappa(0) > 0$, prove that there is one and only one circle γ that approximates β near $\beta(0)$ in the sense that

$$\gamma(0) = \beta(0), \quad \gamma'(0) = \beta'(0), \quad \text{and} \quad \gamma''(0) = \beta''(0).$$

Show that γ lies in the osculating plane of β at $\beta(0)$ and find its center $\mathbf{c}$ and radius r (see Fig. 2.13). The circle γ is called the *osculating circle* and $\mathbf{c}$ the *center of curvature* of β at $\beta(0)$. (The same results hold when 0 is replaced by any number s.)

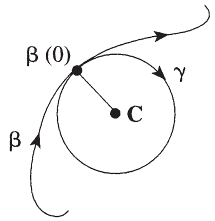


FIG. 2.13

7. If α and a reparametrization $\overline{\alpha} = \alpha(h)$ are both unit-speed curves, show that

(a) $h(s) = \pm s + s_0$ for some number s_0;

(b) $\overline{T} = \pm T(h)$,

$\overline{N} = N(h), \quad \overline{\kappa} = \kappa(h), \quad \overline{\tau} = \tau(h)$,

$\overline{B} = \pm B(h)$,

where the sign ($\pm$) is the same as that in (a), and we assume $\kappa > 0$. Thus even in the orientation-reversing case, the principal normals N and $\overline{N}$ still point in the same direction.

8. *Curves in the plane.* For a unit-speed curve $\beta(s) = (x(s), y(s))$ in $\mathbf{R}^2$, the *unit tangent* is $T = \beta' = (x', y')$ as usual, but the *unit normal* N is defined by rotating T through +90°, so $N = (-y', x')$. Thus T' *and* N are collinear, and the *plane curvature* $\tilde{\kappa}$ of β is defined by the Frenet equation $T' = \tilde{\kappa}N$.

(a) Prove that $\tilde{\kappa} = T' \bullet N$ and $N' = -\tilde{\kappa}T$.

(b) The *slope angle* $\varphi(s)$ of β is the differentiable function such that

$$T = (\cos\varphi, \sin\varphi) = \cos\varphi\, U_x + \sin\varphi\, U_y.$$

(The existence of φ derives from Ex. 12 of Sec. 1.) Show that $\tilde{\kappa} = \varphi'$.

(c) Find the curvature $\tilde{\kappa}$ of the following plane curves.

(i) $(r\cos\frac{t}{r}, r\sin\frac{t}{r})$, counterclockwise circle.

(ii) $(r\cos(-\frac{t}{r}), r\sin(-\frac{t}{r}))$, clockwise circle.

(d) Show that if $\tilde{\kappa}$ does not change sign, then $|\tilde{\kappa}|$ is the usual $\mathbf{R}^3$ curvature κ. (For such comparisons we can always regard $\mathbf{R}^2$ as, say, the xy plane in $\mathbf{R}^3$.)

9. Let $\tilde{\beta}$ be the Frenet approximation of a unit-speed curve β with $\tau \neq 0$ near $s = 0$.

If, say, the B_0 component of $\tilde{\beta}$ is removed, the resulting curve is the *orthogonal projection* of $\tilde{\beta}$ in the T_0N_0 plane. It is the view of $\beta \approx \tilde{\beta}$ that one gets by looking toward $\beta(0) = \tilde{\beta}(0)$ directly along the vector B_0.

Sketch the general shape of the orthogonal projections of $\tilde{\beta}$ near $s = 0$ in each of the planes T_0N_0 (*osculating plane*), T_0B_0 (*rectifying plane*), and N_0B_0 (*normal plane*). These views of $\beta \approx \tilde{\beta}$ can be confirmed experimentally using a bent piece of wire. For computer views, see Exercise 15 of Section 4.

10. *Spherical curves.* Let α be a unit-speed curve with $\kappa > 0$, $\tau \neq 0$.

(a) If α lies on a sphere of center $\mathbf{c}$ and radius r, show that

$$\alpha - \mathbf{c} = -\rho N - \rho'\sigma B,$$

where $\rho = 1/\kappa$ and $\sigma = 1/\tau$. Thus $r^2 = \rho^2 + (\rho'\sigma)^2$.

(b) Conversely, if $\rho^2 + (\rho'\sigma)^2$ has constant value r^2 and $\rho' \neq 0$, show that α lies on a sphere of radius r.
(*Hint:* For (b), show that the "center curve" $\gamma = \alpha + \rho N + \rho'\sigma B$—suggested by (a)—is constant.)

11. Let β, $\bar{\beta}$: $I \to \mathbf{R}^3$ be unit-speed curves with nonvanishing curvature and torsion. If $T = \bar{T}$, then β and $\bar{\beta}$ are parallel (Ex. 10 of Sec. 2). If $B = \bar{B}$, prove that $\bar{\beta}$ is parallel to either β or the curve $s \to -\beta(s)$.

2.4 Arbitrary-Speed Curves

It is a simple matter to adapt the results of the previous section to the study of a regular curve α: $I \to \mathbf{R}^3$ that does not necessarily have unit speed. We merely transfer to α the Frenet apparatus of a unit-speed reparametrization $\bar{\alpha}$ of α. Explicitly, if s is an arc length function for α as in Theorem 2.1, then

$$\alpha(t) = \bar{\alpha}(s(t)) \quad \text{for all } t,$$

or, in functional notation, $\alpha = \bar{\alpha}(s)$, as suggested by Fig. 2.14. Now if $\bar{\kappa} > 0$, $\bar{\tau}$, $\bar{T}$, $\bar{N}$, and $\bar{B}$ are defined for $\bar{\alpha}$ as in Section 3, we define for α the

curvature function: $\kappa = \bar{\kappa}(s)$,
torsion function: $\tau = \bar{\tau}(s)$,
unit tangent vector field: $T = \bar{T}(s)$,
principal normal vector field: $N = \bar{N}(s)$,
binormal vector field: $B = \bar{B}(s)$.

In general κ and $\bar{\kappa}$ are different functions, defined on different intervals. But *they give exactly the same description of the turning of the common route of α and $\bar{\alpha}$*, since at any point $\alpha(t) = \bar{\alpha}(s(t))$ the numbers $\kappa(t)$ and $\bar{\kappa}(s(t))$ are

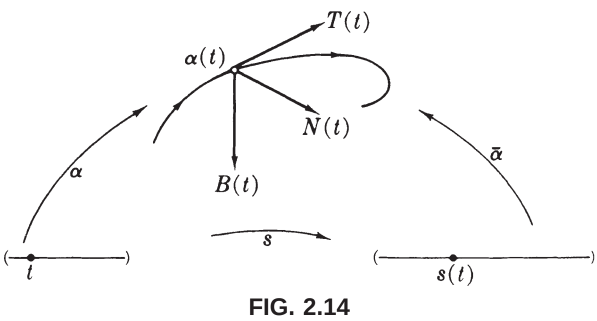


FIG. 2.14

by definition the same. Similarly with the rest of the Frenet apparatus; since only a change of parametrization is involved, its fundamental geometric meaning is the same as before. In particular, T, N, B is again a frame field on α linked to the shape of α as indicated in the discussion of Frenet approximations.

For purely theoretical work, this simple transference is often all that is needed. Data about α converts into data about the unit-speed reparametrization $\bar{\alpha}$; results about $\bar{\alpha}$ convert to results about α. For example, if α is a regular curve with $\tau = 0$, then by the definition above $\bar{\alpha}$ has $\bar{\tau} = 0$; by Corollary 3.5, $\bar{\alpha}$ is a plane curve, so obviously α is too.

However, for explicit numerical computations—and occasionally for the theory as well—this transference is impractical, since it is rarely possible to find explicit formulas for $\bar{\alpha}$. (For example, try to find a unit-speed parametrization for the curve $\alpha(t) = (t, t^2, t^3)$.)

The Frenet formulas are valid only for unit-speed curves; they tell the rate of change of the frame field T, N, B *with respect to arc length.* However, the speed v of the curve is the proper correction factor in the general case.

4.1 Lemma If α is a regular curve in $\mathbf{R}^3$ with $\kappa > 0$, then

$$\begin{aligned} T' &= \kappa v N, \\ N' &= -\kappa v T + \tau v B, \\ B' &= -\tau v N. \end{aligned}$$

Proof. Let $\bar{\alpha}$ be a unit-speed reparametrization of α. Then by definition, $T = \bar{T}(s)$, where s is an arc length function for α. The chain rule as applied to differentiation of vector fields (Exercise 7 of Section 2) gives

$$T' = \bar{T}'(s)\frac{ds}{dt}.$$

By the usual Frenet equations, $\bar{T}' = \bar{\kappa}\bar{N}$. Substituting the function s in this equation yields

$$\bar{T}'(s) = \bar{\kappa}(s)\bar{N}(s) = \kappa N$$

by the definition of κ and N in the arbitrary-speed case. Since ds/dt is the speed function v of α, these two equations combine to yield $T' = \kappa v N$. The formulas for N' and B' are derived in the same way. ◆

There is a commonly used notation for the calculus that completely ignores change of parametrization. For example, the same letter would designate both a curve α and its unit-speed parametrization $\bar{\alpha}$, and similarly with the

Frenet apparatus of these two curves. Differences in derivatives are handled by writing, say, dT/dt for T', but dT/ds for either $\bar{T}'$ or its reparametrization $\bar{T}'(s)$. If these conventions were used, the proof above would combine the chain rule $dT/dt = (dT/ds)\,(ds/dt)$ and the Frenet formula $dT/ds = \kappa N$ to give $dT/dt = \kappa v N$.

Only for a *constant-speed* curve is acceleration always orthogonal to velocity, since $\beta' \bullet \beta'$ constant is equivalent to $(\beta' \bullet \beta')' = 2\beta' \bullet \beta'' = 0$. In the general case, we analyze velocity and acceleration by expressing them in terms of the Frenet frame field.

4.2 Lemma If α is a regular curve with speed function v, then the velocity and acceleration of α are given by (Fig. 2.15.)

$$\alpha' = vT, \quad \alpha'' = \frac{dv}{dt}T + \kappa v^2 N.$$

Proof. Since $\alpha = \bar{\alpha}(s)$, where s is the arc length function of α, we find, using Lemma 4.5 of Chapter 1, that

$$\alpha' = \bar{\alpha}'(s)\frac{ds}{dt} = v\bar{T}(s) = vT.$$

Then a second differentiation yields

$$\alpha'' = \frac{dv}{dt}T + vT' = \frac{dv}{dt}T + \kappa v^2 N,$$

where we use Lemma 4.1. ◆

The formula $\alpha' = vT$ is to be expected since α' and T are each tangent to the curve and T has a unit length, while $\|\alpha'\| = v$. The formula for acceleration is more interesting. By definition, α'' is the rate of change of the

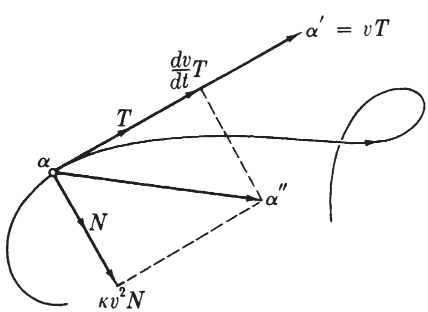


FIG. 2.15

velocity α', and in general both the length and the direction of α' are changing. The *tangential component* $(dv/dt)T$ of α'' measures the rate of change of the length of α' (that is, of the speed of α). The *normal component* $\kappa v^2 N$ measures the rate of change of the direction of α'. Newton's laws of motion show that these components may be experienced as forces. For example, in a car that is speeding up or slowing down on a straight road, the only force one feels is due to $(dv/dt)T$. If one takes an unbanked curve at speed v, the resulting sideways force is due to $\kappa v^2 N$. Here κ measures how sharply the *road* turns; the effect of speed is given by v^2, so 60 miles per hour is four times as unsettling as 30.

We now find effectively computable expressions for the Frenet apparatus.

4.3 Theorem Let α be a regular curve in $\mathbf{R}^3$. Then

$$T = \alpha'/\| \alpha' \|,$$

$$N = B \times T, \qquad \kappa = \| \alpha' \times \alpha'' \| / \| \alpha' \|^3,$$

$$B = \alpha' \times \alpha'' / \| \alpha' \times \alpha'' \|, \quad \tau = (\alpha' \times \alpha'') \bullet \alpha''' / \| \alpha' \times \alpha'' \|^2.$$

Proof. Since $v = \| \alpha' \| > 0$, the formula $T = \alpha'/\|\alpha'\|$ is equivalent to $\alpha' = vT$. From the preceding lemma we get

$$\alpha' \times \alpha'' = (vT) \times \left(\frac{dv}{dt} T + \kappa v^2 N \right)$$

$$= v \frac{dv}{dt} T \times T + \kappa v^3 T \times N = \kappa v^3 B,$$

since $T \times T = 0$. Taking norms we find

$$\| \alpha' \times \alpha'' \| = \| \kappa v^3 B \| = \kappa v^3$$

because $\| B \| = 1$, $\kappa \geqq 0$, and $v > 0$. Indeed, *this equation shows that for* regular *curves*, $\| \alpha' \times \alpha'' \| > 0$ *is equivalent to the usual condition* $\kappa > 0$. (Thus for $\kappa > 0$, α' and α'' are linearly independent and determine the osculating plane at each point, as do T and N.) Then

$$B = \frac{\alpha' \times \alpha''}{\kappa v^3} = \frac{\alpha' \times \alpha''}{\| \alpha' \times \alpha'' \|}.$$

Since $N = B \times T$ is true for any Frenet frame field (Exercise 4 of Section 3), only the formula for torsion remains to be proved.

To find the dot product $(\alpha' \times \alpha'') \bullet \alpha'''$ we express everything in terms of T, N, B. We already know that $\alpha' \times \alpha'' = \kappa v^3 B$. Thus, since $0 = T \bullet B = N \bullet B$, *we need only find the B component of* α'''. But

$$\alpha''' = \left(\frac{dv}{dt}T + \kappa v^2 N\right)' = \kappa v^2 N' + \cdots$$
$$= \kappa v^3 \tau B + \cdots,$$

where we use Lemma 4.1. Consequently, $(\alpha' \times \alpha'') \bullet \alpha''' = \kappa^2 v^6 \tau$, and since $\| \alpha' \times \alpha'' \| = \kappa v^3$, we have the required formula for τ. ◆

The triple scalar product in this formula for τ could (by Exercise 4 of Section 1) also be written $\alpha' \bullet \alpha'' \times \alpha'''$. But we need $\alpha' \times \alpha''$ anyway, so it is more efficient to find $(\alpha' \times \alpha'') \bullet \alpha'''$.

4.4 Example We compute the Frenet apparatus of the 3-*curve*

$$\alpha(t) = (3t - t^3, 3t^2, 3t + t^3).$$

The derivatives are

$$\alpha'(t) = 3(1 - t^2, 2t, 1 + t^2),$$
$$\alpha''(t) = 6(-t, 1, t),$$
$$\alpha'''(t) = 6(-1, 0, 1).$$

Now,

$$\alpha'(t) \bullet \alpha'(t) = 18(1 + 2t^2 + t^4),$$

so

$$v(t) = \| \alpha'(t) \| = \sqrt{18}(1 + t^2).$$

Applying the definition of cross product yields

$$\alpha'(t) \times \alpha''(t) = 18\begin{vmatrix} U_1 & U_2 & U_3 \\ 1 - t^2 & 2t & 1 + t^2 \\ -t & 1 & t \end{vmatrix} = 18(-1 + t^2, -2t, 1 + t^2).$$

Dotting this vector with itself, we get

$$(18)^2\left[(-1 + t^2)^2 + 4t^2 + (1 + t^2)^2\right] = 2(18)^2(1 + t^2)^2.$$

Hence

$$\| \alpha'(t) \times \alpha''(t) \| = 18\sqrt{2}(1 + t^2).$$

The expressions above for $\alpha' \times \alpha''$ and α''' yield

$$(\alpha' \times \alpha'') \bullet \alpha''' = 6 \bullet 18 \bullet 2.$$

It remains only to substitute this data into the formulas in Theorem 4.3, with N being computed by another cross product. The final results are

$$T = \frac{(1-t^2, 2t, 1+t^2)}{\sqrt{2}(1+t^2)},$$

$$N = \frac{(-2t, 1-t^2, 0)}{1+t^2},$$

$$B = \frac{(-1+t^2, -2t, 1+t^2)}{\sqrt{2}(1+t^2)},$$

$$\kappa = \tau = \frac{1}{3(1+t^2)^2}.$$

Alternatively, we could use the identity in Lemma 1.8 to compute $\| \alpha' \times \alpha'' \|$ and express

$$(\alpha' \times \alpha'') \bullet \alpha''' = \alpha' \bullet (\alpha'' \times \alpha''')$$

as a determinant by Exercise 4 of Section 1.

To summarize, we now have the Frenet apparatus for an arbitrary regular curve α, namely, its curvature, torsion, and Frenet frame field. This apparatus satisfies the extended Frenet formulas with speed factor v and can be computed by Theorem 4.3. If $v = 1$, that is, if α is a unit-speed curve, the results of Section 3 are recovered.

Let us consider some applications of the Frenet formulas. There are a number of natural ways in which a given curve β gives rise to a new curve $\tilde{\beta}$ whose geometric properties illuminate some aspect of the behavior of β.

For example, the *spherical image* of a unit-speed curve β is the curve $\sigma \approx T$ with the same Euclidean coordinates as $T = \beta'$. Geometrically, σ is gotten by moving each $T(s)$ to the origin of $\mathbf{R}^3$, as suggested in Fig. 2.16. Thus σ lies on the unit sphere Σ, and the *motion* of σ represents the *turning* of β.

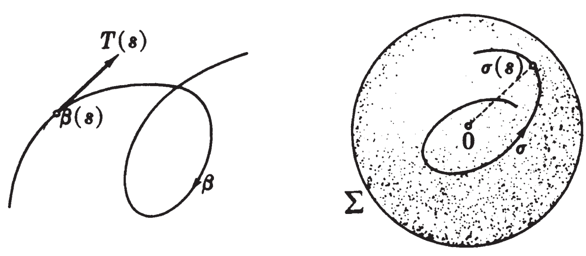


FIG. 2.16

For instance, if β is the helix in Example 3.3, the formula there for T shows that

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

So the spherical image of a helix lies on the circle cut from $\sum$ by the plane $z = b/c$.

Although the original curve β has unit speed, we cannot expect that σ does also. In fact, $\sigma = T$ implies $\sigma' = T' = \kappa N$, so the *speed* of σ equals the *curvature* κ of β. Thus to compute the curvature of σ, we must use the extended Frenet formulas in Theorem 4.3. From

$$\sigma'' = (\kappa N)' = \frac{d\kappa}{ds}N + \kappa N' = -\kappa^2 T + \frac{d\kappa}{ds}N + \kappa\tau B,$$

we get

$$\sigma' \times \sigma'' = -\kappa^3 N \times T + \kappa^2 \tau N \times B = \kappa^2(\kappa B + \tau T).$$

By Theorem 4.3 the curvature of the spherical image σ is

$$\kappa_\sigma = \frac{\|\sigma' \times \sigma''\|}{v^3} = \frac{\sqrt{\kappa^2 + \tau^2}}{\kappa} = \left(1 + \left(\frac{\tau}{\kappa}\right)^2\right)^{1/2} \geq 1$$

and thus depends only on the ratio of torsion to curvature for the original curve β.

Here is a closely related application in which this ratio τ/κ turns out to be decisive.

4.5 Definition A regular curve α in $\mathbf{R}^3$ is a *cylindrical helix* provided the unit tangent T of α has constant angle ϑ with some fixed unit vector $\mathbf{u}$; that is, $T(t) \bullet \mathbf{u} = \cos\vartheta$ for all t.

This condition is not altered by reparametrization, so for theoretical purposes we need only deal with a cylindrical helix β that has unit speed. So suppose β is a unit-speed curve with $T \bullet \mathbf{u} = \cos\vartheta$. If we pick a reference point, say $\beta(0)$, on β, then the real-valued function

$$h(s) = (\beta(s) - \beta(0)) \bullet \mathbf{u}$$

tells how far $\beta(s)$ has "risen" in the $\mathbf{u}$ direction since leaving $\beta(0)$ (Fig. 2.17). But

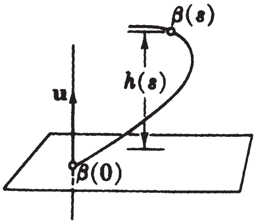


FIG. 2.17

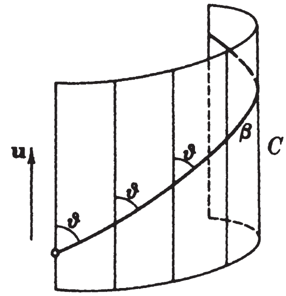


FIG. 2.18

$$\frac{dh}{ds} = \beta' \bullet \mathbf{u} = T \bullet \mathbf{u} = \cos \vartheta,$$

so β is rising at a constant rate *relative to arc length*, and $h(s) = s \cos \vartheta$. If we shift to an arbitrary parametrization, this formula becomes

$$h(t) = s(t) \cos \vartheta,$$

where s is the arc length function.

By drawing a line through each point of β in the $\mathbf{u}$ direction, we construct a cylinder C on which β moves in such a way as to cut each such line at constant angle ϑ, as in Fig. 2.18. In the special case when this cylinder is circular, β is evidently a helix of the type defined in Example 3.3.

It turns out to be quite easy to identify cylindrical helices.

4.6 Theorem A regular curve α with $\kappa > 0$ is a cylindrical helix if and only if the ratio τ/κ is constant.

Proof. It suffices to consider the case where α has unit speed. If α is a cylindrical helix with $T \bullet \mathbf{u} = \cos \vartheta$, then

$$0 = (T \bullet \mathbf{u})' = T' \bullet \mathbf{u} = \kappa N \bullet \mathbf{u}.$$

Since $\kappa > 0$, we conclude that $N \bullet \mathbf{u} = 0$. Thus for each s, $\mathbf{u}$ lies in the plane determined by $T(s)$ and $B(s)$. Orthonormal expansion yields

$$\mathbf{u} = \cos \vartheta \, T + \sin \vartheta \, B.$$

As usual we differentiate and apply Frenet formulas to obtain

$$0 = (\kappa \cos \vartheta - \tau \sin \vartheta)N.$$

Hence $\tau \sin \vartheta = \kappa \cos \vartheta$, so that τ/κ has constant value $\cot \vartheta$.

Conversely, suppose that τ/κ is constant. Choose an angle ϑ such that $\cot \vartheta = \tau/\kappa$. If

$$U = \cos \vartheta \, T + \sin \vartheta \, B,$$

we find

$$U' = (\kappa \cos \vartheta - \tau \sin \vartheta)N = 0.$$

This parallel vector field U then determines (as in Remark 3.4) a unit vector $\mathbf{u}$ such that $T \bullet \mathbf{u} = \cos \vartheta$, so α is a cylindrical helix. ◆

In Exercise 9 this information about cylindrical helices is used to show that *circular* helices are characterized by constancy of curvature and torsion (see also Corollary 5.5 of Chapter 3).

Simple hypotheses on a regular curve in $\mathbf{R}^3$ thus have the following effects ($\Leftrightarrow$ means "if and only if"):

$\kappa = 0$	$\Leftrightarrow$	straight line,
$\tau = 0$	$\Leftrightarrow$	plane curve,
κ const > 0 and $\tau = 0$	$\Leftrightarrow$	circle,
κ const > 0 and τ const > 0	$\Leftrightarrow$	circular helix,
τ/κ const $\neq 0$	$\Leftrightarrow$	cylindrical helix.

Exercises

Computer commands that produce the Frenet apparatus, κ, τ, T, N, B, of a curve are given in the Appendix. Their use is optional in the following exercises.

1. For the curve $\alpha(t) = (2t, t^2, t^3/3)$,
(a) Compute the Frenet apparatus.
(b) Sketch the curve for $-4 \leqq t \leqq 4$, showing T, N, B at $t = 2$.
(c) Find the limiting values of T, N, and B as $t \to -\infty$ and $t \to \infty$.

2. Express the curvature and torsion of the curve $\alpha(t) = (\cosh t, \sinh t, t)$ in terms of arc length s measured from $t = 0$.

3. The curve $\alpha(t) = (t\cos t,\ t\sin t,\ t)$ lies on a double cone and passes through the vertex at $t = 0$.

(a) Find the Frenet apparatus of α at $t = 0$.

(b) Sketch the curve for $-2\pi \leqq t \leqq 2\pi$, showing T, N, B at $t = 0$.

4. Show that the curvature of a regular curve in $\mathbf{R}^3$ is given by

$$\kappa^2 v^4 = \| \alpha'' \|^2 - (dv/dt)^2.$$

5. If α is a curve with constant speed $c > 0$, show that

$$T = \alpha'/c, \quad N = \alpha''/\| \alpha'' \|, \quad B = \alpha' \times \alpha''/(c\| \alpha'' \|),$$

$$\kappa = \frac{\| \alpha'' \|}{c^2}, \quad \tau = \frac{\alpha' \times \alpha'' \bullet \alpha'''}{c^2 \| \alpha'' \|^2},$$

where for N, B, τ, we assume α'' never zero, that is, $\kappa > 0$.

6. (a) If α is a cylindrical helix, prove that its unit vector $\mathbf{u}$ (Thm. 4.5) is

$$\mathbf{u} = \frac{\tau}{\sqrt{\kappa^2 + \tau^2}} T + \frac{\kappa}{\sqrt{\kappa^2 + \tau^2}} B,$$

and the coefficients here are $\cos\vartheta$ and $\sin\vartheta$ (for ϑ as in Def. 4.5).

(b) Check (a) for the cylindrical helix in Example 4.2 of Chapter 1.

7. Let $\alpha\colon I \to \mathbf{R}^3$ be a cylindrical helix with unit vector $\mathbf{u}$. For $t_0 \in I$, the curve

$$\gamma(t) = \alpha(t) - ((\alpha(t) - \alpha(t_0)) \bullet \mathbf{u})\,\mathbf{u}$$

is called a *cross-sectional curve* of the cylinder on which α lies. Prove:

(a) γ lies in the plane through $\alpha(t_0)$ orthogonal to $\mathbf{u}$.

(b) The curvature of γ is $\kappa/\sin^2\vartheta$, where κ is the curvature of α.

8. Verify that the following curves are cylindrical helices and, for each, find the unit vector $\mathbf{u}$, angle ϑ, and cross-sectional curve σ.

(a) The curve in Exercise 1. (b) The curve in Example 4.4.

(c) The curve in Exercise 2.

9. If α is a curve with $\kappa > 0$ and τ both constant, show that α is a circular helix.

10. (a) Prove that a curve is a cylindrical helix if and only if its spherical image is part of a circle.

(b) Sketch the spherical image of the cylindrical helix in Exercise 1. Is it a complete circle? Find its center.

11. If α is a curve with $\kappa > 0$, its *central curve* $\alpha^* = \alpha + (1/\kappa)N$ consists of all centers of curvature of α (Ex. 6 of Sec. 3). For nonzero numbers a and b, let β_{ab} be the helix in Example 3.3.

(a) Show that the central curve of β_{ab} is the helix $\beta_{\hat{a}b}$, where $\hat{a} = -b^2/a$.

(b) Deduce that the central curve of $\beta_{\hat{a}b}$ is the original helix β_{ab}.

(c) (*Computer graphics.*) Plot three complete turns of the mutually central helices $\beta_{2,1}$ and $\beta_{-1/2,1}$ in the same figure.

12. If $\alpha(t) = (x(t), y(t))$ is a regular curve in $\mathbf{R}^2$, show that its plane curvature (Ex. 8 of Sec. 3) is given by

$$\tilde{\kappa} = \frac{\alpha'' \bullet J(\alpha')}{v^3} = \frac{x'y'' - x''y'}{(x'^2 + y'^2)^{3/2}},$$

where J is the rotation operator $J(a, b) = (-b, a)$.

13. (*Continuation.*) For a plane curve α with $\tilde{\kappa} \neq 0$, the central curve $\alpha^* = \alpha + (1/\tilde{\kappa})N$ is called the *evolute* of α. Thus α^* gives a direct pointwise description of the turning of α.

(a) Show that

$$\alpha^* = \alpha + \frac{\alpha' \bullet \alpha'}{\alpha'' \bullet J(\alpha')} J(\alpha').$$

(b) Find a formula for the line segment λ_t from $\alpha(t)$ to $\alpha^*(t)$. This segment is the radius (line) of the approximating circle to α near $\alpha(t)$ (Ex. 6 of Sec. 3)

(c) Prove that λ_t is normal to α at $\alpha(t)$ and tangent to α^* at $\alpha^*(t)$. (*Hint:* It can be assumed that α has unit speed.)

14. (*Continuation, Computer graphics.*) In each case, plot the given plane curve and its evolute on the same figure, showing some of the construction lines λ_t.

(a) The ellipse $a(t) = (2\cos t, \sin t)$.

(b) The cycloid $\alpha(t) = (t + \sin t, 1 + \cos t)$ for $-2\pi \leq t \leq 2\pi$. (Here the evolute bears an unexpected relation to the original curve.)

15. (*Computer continuation of Ex. 9 of Sec. 3.*)

(a) Write the commands that, given a regular curve α with $\kappa(0) > 0$, plot—on a small interval $-\varepsilon \leqq t \leqq \varepsilon$—the orthogonal projection of α into the osculating, rectifying, and normal planes at $\alpha(0)$. Show the projections as curves in $\mathbf{R}^2$.

(b) Test (a) on the curves (3), (4), (5) in Example 4.2 of Chapter 1 and those in Example 4.3 of Chapter 3. Compare results.

The following exercise shows that the condition $\kappa > 0$ cannot be avoided in a detailed study of the geometry of curves in $\mathbf{R}^3$ for even if κ is zero at only a single point, the geometric character of the curve can change radically at that point. (This difficulty does not arise for curves in the plane.)

16. It is shown in advanced calculus that the function

$$f(t) = \begin{cases} 0 & \text{if } t \le 0, \\ e^{-1/t^2} & \text{if } t > 0. \end{cases}$$

is infinitely differentiable (has continuous derivatives of all orders). Thus

$$\alpha(t) = (t,\ f(t),\ f(-t))$$

is a well-defined differentiable curve.

(a) Sketch α on an interval $-a \leqq t \leqq a$.
(b) Show that the curvature of α is zero only at $t = 0$.
(c) What are the osculating planes of α for $t < 0$ and $t > 0$?

In the following exercise, a global geometric invariant of curves is gotten by integrating a local invariant.

17. The *total curvature* of a unit-speed curve $\alpha\colon I \to \mathbf{R}^3$ is $\int_I \kappa(s)\,ds$. If α is merely regular, the formula becomes $\int_I \kappa(t)v(t)\,dt$. Find the total curvature of the following curves:

(a) The curve in Example 4.4.
(b) The helix in Example 3.3.
(c) The curve in Exercise 2.
(d) The ellipse $\alpha(t) = (a\cos t,\ b\sin t)$ on $0 \leqq t \leqq 2\pi$.

18. One definition of convexity for a smoothly closed plane curve is that its curvature κ is positive (hence its plane curvature $\tilde{\kappa}$ is either always positive or always negative). Prove that a convex closed plane curve has total curvature 2π. (*Hint:* Consider its spherical image.)

A theorem of Fenchel asserts that every regular closed curve α in $\mathbf{R}^3$ has total curvature $\geqq 2\pi$. Surprisingly, this has an easy proof in terms of surface theory (see Sec. 8 of Ch. 6).

19. (*Computer*.)

(a) Plot the curve

$$\tau(t) = (4\cos 2t + 2\cos t,\ 4\sin 2t - 2\sin t,\ \sin 3t) \quad \text{on } 0 \leqq t \leqq 2\pi.$$

Even looking at this curve from different viewpoints may not make its crossing pattern clear, but Exercise 21 of Section 5.4 will show that τ is a *trefoil knot.*

(Intuitively, a simple closed curve in $\mathbf{R}^3$ is a *knot* provided it cannot be continuously deformed—always remaining simply closed—until it becomes a circle.)

The Fary-Milnor theorem asserts that every knot has total curvature strictly greater than 4π. Show:

(b) The plane curve obtained from τ by removing the z-component $\sin 3t$ has total curvature exactly 4π. (This curve is not simply closed, and hence is not a knot.)

(c) τ can be deformed to a knot that has (numerically estimated) total curvature less than 4.01π.

20. (*Computer.*)

(a) Write a command that, given an arbitrary regular curve, returns the test function in Exercise 10 of Section 3 whose constancy implies that the curve lies on a sphere. (Plotting this function provides a good test for constancy and does not require simplifying it.) (*Hint:* To allow for arbitrary parametrization, replace derivatives $f'(s)$ by $f'(t)v(t)$, where $v(t) = ds/dt$.)

(b) In each case, decide whether the curve lies on a sphere, and if so, find its radius and center:

(i) $\alpha(t) = (2\sin t, \sin 2t, 2\sin^2 t)$;

(ii) $\beta(t) = (\cos^2 t, \sin 2t, 2\sin t)$;

(iii) $\gamma(t) = (\cos t, 1 + \sin t, 2\sin \frac{t}{2})$.

21. Prove that the cubic curve $\gamma(t) = (at, bt^2, ct^3)$, $abc \neq 0$, is a cylindrical helix if and only if $3ac = \pm 2b^2$. (Computer optional.)

2.5 Covariant Derivatives

In Chapter 1 the definition of a new object (curve, differential form, mapping, . . .) was usually followed by an appropriate notion of *derivative* of that object. To see how to define the derivative of a vector field on a Euclidean space, we mimic the definition of the derivative $\mathbf{v}[f]$ of a function f relative to a tangent vector $\mathbf{v}$ at a point $\mathbf{p}$ (Definition 3.1 of Chapter 1). In fact, replacing f by a vector field W on $\mathbf{R}^3$ gives a vector field $t \to W(\mathbf{p} + t\mathbf{v})$ on the curve $t \to \mathbf{p} + t\mathbf{v}$. The derivative of such a vector field was defined in Section 2. Then the derivative of W with respect to $\mathbf{v}$ will be the derivative of $t \to W(\mathbf{p} + t\mathbf{v})$ at $t = 0$.

5.1 Definition Let W be a vector field on $\mathbf{R}^3$, and let $\mathbf{v}$ be a tangent vector field to $\mathbf{R}^3$ at the point $\mathbf{p}$. Then the *covariant derivative* of W with respect to $\mathbf{v}$ is the tangent vector

$$\nabla_{\mathbf{v}} W = W(\mathbf{p} + t\mathbf{v})'(0)$$

at the point $\mathbf{p}$.

Evidently $\nabla_v W$ *measures the initial rate of change of* $W(\mathbf{p})$ *as* $\mathbf{p}$ *moves in the* $\mathbf{v}$ *direction.* (The term "covariant" derives from the generalization of this notion discussed in Chapter 7.)

For example, suppose $W = x^2U_1 + yzU_3$, and

$$\mathbf{v} = (-1, 0, 2) \quad \text{at} \quad \mathbf{p} = (2, 1, 0).$$

Then

$$\mathbf{p} + t\mathbf{v} = (2 - t, 1, 2t),$$

so

$$W(\mathbf{p} + t\mathbf{v}) = (2 - t)^2U_1 + 2tU_3,$$

where strictly speaking U_1 and U_3 are also evaluated at $\mathbf{p} + t\mathbf{v}$. Thus,

$$\nabla_v W = W(\mathbf{p} + t\mathbf{v})'(0) = -4U_1(\mathbf{p}) + 2U_3(\mathbf{p}).$$

5.2 Lemma If $W = \sum w_iU_i$ is a vector field on $\mathbf{R}^3$, and $\mathbf{v}$ is a tangent vector at $\mathbf{p}$, then

$$\nabla_v W = \sum \mathbf{v}[w_i]U_i(\mathbf{p}).$$

Proof. We have

$$W(\mathbf{p}+t\mathbf{v}) = \sum w_i(\mathbf{p}+t\mathbf{v})U_i(\mathbf{p}+t\mathbf{v})$$

for the restriction of W to the curve $t \to \mathbf{p} + t\mathbf{v}$. To differentiate such a vector field (at $t = 0$), one simply differentiates its Euclidean coordinates (at $t = 0$). But by the definition of directional derivative (Definition 3.1 of Chapter 1), the derivative of $w_i(\mathbf{p} + t\mathbf{v})$ at $t = 0$ is precisely $\mathbf{v}[w_i]$. Thus

$$\nabla_v W = W(\mathbf{p} + t\mathbf{v})'(0) = \sum \mathbf{v}[w_i]U_i(\mathbf{p}). \qquad \blacklozenge$$

In short, *to apply* ∇_v *to a vector field, apply* $\mathbf{v}$ *to its Euclidean coordinates.* Thus the following linearity and Leibnizian properties of covariant derivative follow easily from the corresponding properties (Theorem 3.3 of Chapter 1) of directional derivatives.

5.3 Theorem Let $\mathbf{v}$ and $\mathbf{w}$ be tangent vectors to $\mathbf{R}^3$ at $\mathbf{p}$, and let Y and Z be vector fields on $\mathbf{R}^3$. Then for numbers a, b and functions f,

(1) $\nabla_{av+bw} Y = a\nabla_v Y + b\nabla_w Y$.
(2) $\nabla_v(aY + bZ) = a\nabla_v Y + b\nabla_v Z$.
(3) $\nabla_v(fY) = \mathbf{v}[f]Y(\mathbf{p}) + f(\mathbf{p})\nabla_v Y$.
(4) $\mathbf{v}[Y \bullet Z] = \nabla_v Y \bullet Z(\mathbf{p}) + Y(\mathbf{p}) \bullet \nabla_v Z$.

Proof. For example, let us prove (4). If

$$Y = \sum y_i U_i \quad \text{and} \quad Z = \sum z_i U_i\,,$$

then

$$Y \bullet Z = \sum y_i z_i.$$

Hence by Theorem 3.3 of Chapter 1,

$$\mathbf{v}[Y \bullet Z] = \mathbf{v}\left[\sum y_i z_i\right] = \sum \mathbf{v}[y_i] z_i(\mathbf{p}) + \sum y_i(\mathbf{p}) \mathbf{v}[z_i].$$

But by the preceding lemma,

$$\nabla_v Y = \sum \mathbf{v}[y_i] U_i(\mathbf{p}) \quad \text{and} \quad \nabla_v Z = \sum \mathbf{v}[z_i] U_i(\mathbf{p}).$$

Thus the two sums displayed above are precisely $\nabla_v Y \bullet Z(\mathbf{p})$ and $Y(\mathbf{p}) \bullet \nabla_v Z$. ◆

Using the pointwise principle (Chapter 1, Section 2), we can take the covariant derivative of a vector field W with respect to a *vector field* V, rather than a single tangent vector $\mathbf{v}$. The result is the *vector field* $\nabla_V W$ whose value at each point $\mathbf{p}$ is $\nabla_{V(p)} W$. Thus $\nabla_V W$ consists of all the covariant derivatives of W with respect to the vectors of V. It follows immediately from the lemma above that if $W = \sum w_i U_i$, then

$$\nabla_V W = \sum V[w_i] U_i.$$

Coordinate computations are easy using the basic identity $U_i[f] = \partial f/\partial x_i$. For example, suppose $V = (y - x)U_1 + xyU_3$ and (as in the example above) $W = x^2 U_1 + yzU_3$. Then

$$V[x^2] = (y - x)U_1[x^2] = 2x(y - x),$$
$$V[yz] = xyU_3[yz] = xy^2.$$

Hence

$$\nabla_V W = 2x(y - x)U_1 + xy^2 U_3.$$

For the covariant derivative $\nabla_V W$ as expressed entirely in terms of vector fields, the properties in the preceding theorem take the following form.

5.4 Corollary Let V, W, Y, and Z be vector fields on $\mathbf{R}^3$. Then

(1) $\nabla_{fV+gW} Y = f\nabla_V Y + g\nabla_W Y$, for all functions f and g.

(2) $\nabla_V(aY + bZ) = a\nabla_V Y + b\nabla_V Z$, for all numbers a and b.

(3) $\nabla_V(fY) = V[f]Y + f\nabla_V Y$, for all functions f.
(4) $V[Y \bullet Z] = \nabla_V Y \bullet Z + Y \bullet \nabla_V Z$.

We shall omit the proof, which is an exercise in the use of parentheses based on the (pointwise principle) definition $(\nabla_V Y)(\mathbf{p}) = \nabla_{V(p)} Y$.

Note that $\nabla_V Y$ does not behave symmetrically with respect to V and Y. This is to be expected, since it is Y that is being differentiated, while the role of V is merely algebraic. In particular, $\nabla_{fV} Y$ is $f\nabla_V Y$, but $\nabla_V(fY)$ is not $f\nabla_V Y$: There is an extra term arising from the differentiation of f by V.

Exercises

1. Consider the tangent vector $\mathbf{v} = (1, -1, 2)$ at the point $\mathbf{p} = (1, 3, -1)$. Compute $\nabla_{\mathbf{v}} W$ directly from the definition, where

(a) $W = x^2 U_1 + yU_2$. (b) $W = xU_1 + x^2 U_2 - z^2 U_3$.

2. Let $V = -yU_1 + xU_3$ and $W = \cos x U_1 + \sin x U_2$. Express the following covariant derivatives in terms of U_1, U_2, U_3:

(a) $\nabla_V W$. (b) $\nabla_V V$.
(c) $\nabla_V(z^2 W)$. (d) $\nabla_W(V)$.
(e) $\nabla_V(\nabla_V W)$. (f) $\nabla_V(xV - zW)$.

3. If W is a vector field with constant length $\|W\|$, prove that for any vector field V, the covariant derivative $\nabla_V W$ is everywhere orthogonal to W.

4. Let X be the special vector field $\sum x_i U_i$, where x_1, x_2, x_3 are the natural coordinate functions of $\mathbf{R}^3$. Prove that $\nabla_V X = V$ for every vector field V.

5. Let W be a vector field defined on a region containing a regular curve α. Then $t \to W(\alpha(t))$ is a vector field on α called the *restriction* of W to α and denoted by W_α.

(a) Prove that $\nabla_{\alpha'(t)} W = (W_\alpha)'(t)$.
(b) Deduce that the straight line in Definition 5.1 may be replaced by *any* curve with initial velocity $\mathbf{v}$. Thus the derivative Y' of a vector field Y on a curve α is (almost) $\nabla_{\alpha'} Y$.

2.6 Frame Fields

When the Frenet formulas were discovered (by Frenet in 1847, and independently by Serret in 1851), the theory of *surfaces* in $\mathbf{R}^3$ was already a richly developed branch of geometry. The success of the Frenet approach to curves led Darboux (around 1880) to adapt this "method of moving frames" to the study of surfaces. Then, as we mentioned earlier, it was Cartan who brought the method to full generality. His essential idea was very simple: To each point of the object under study (a curve, a surface, Euclidean space itself, . . .) assign a frame; then using orthonormal expansion express the rate of change of the frame in terms of the frame itself. This, of course, is just what the Frenet formulas do in the case of a curve.

In the next three sections we shall carry out this scheme for the Euclidean space $\mathbf{R}^3$. We shall see that geometry of curves and surfaces in $\mathbf{R}^3$ is not merely an analogue, but actually a *corollary*, of these basic results. Since the main application (to surface theory) comes only in Chapter 6, these sections may be postponed, and read later as a preliminary to that chapter.

By means of the pointwise principle (Chapter 1, Section 2) we can automatically extend operations on individual tangent vectors to operations on vector fields. For example, if V and W are vector fields on $\mathbf{R}^3$, then the *dot product* $V \bullet W$ of V and W is the (differentiable) real-valued function on $\mathbf{R}$ whose value at each point $\mathbf{p}$ is $V(\mathbf{p}) \bullet W(\mathbf{p})$. The *norm* $\|\mathbf{V}\|$ of V is the real-valued function on $\mathbf{R}^3$ whose value at $\mathbf{p}$ is $\|V(\mathbf{p})\|$. Thus $\|V\| = (V \bullet V)^{1/2}$. By contrast with $V \bullet W$, the norm function $\|V\|$ need not be differentiable at points for which $V(\mathbf{p}) = 0$, since the square-root function is badly behaved at 0.

In Chapter 1 we called the three vector fields U_1, U_2, U_3 the natural frame field on $\mathbf{R}^3$. Here is a simple but crucial generalization.

6.1 Definition Vector fields E_1, E_2, E_3 on $\mathbf{R}^3$ constitute a *frame field* on $\mathbf{R}^3$ provided

$$E_i \bullet E_j = \delta_{ij} \quad (1 \le i, j \le 3),$$

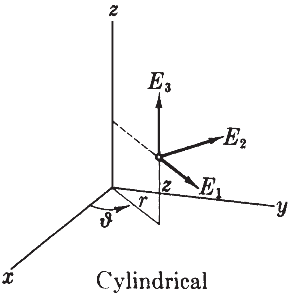
where δ_{ij} is the Kronecker delta.

Thus at each point $\mathbf{p}$ the vectors $E_1(\mathbf{p})$, $E_2(\mathbf{p})$, $E_3(\mathbf{p})$ do in fact form a frame (Definition 1.4) since they have unit length and are mutually orthogonal.

In elementary calculus, frame fields are usually derived from coordinate systems, as in the following cases.

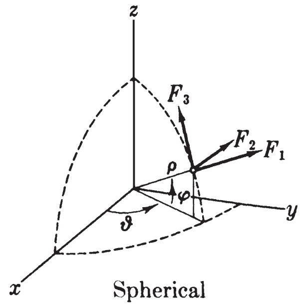
6.2 Example (1) *The cylindrical frame field* (Fig. 2.19). Let r, ϑ, z be the usual cylindrical coordinate functions on $\mathbf{R}^3$. We shall pick a unit vector field in the direction in which each coordinate increases (when the other two are held constant). For r, this is evidently

$$E_1 = \cos\vartheta\, U_1 + \sin\vartheta\, U_2,$$



Cylindrical

FIG. 2.19



Spherical

FIG. 2.20

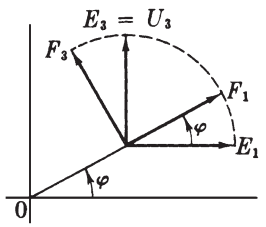


FIG. 2.21

pointing straight out from the z axis. Then

$$E_2 = -\sin\vartheta\, U_1 + \cos\vartheta\, U_2$$

points in the direction of increasing ϑ as in Fig. 2.19. Finally, the direction of increase of z is, of course, straight up, so

$$E_3 = U_3.$$

It is easy to check that $E_i \bullet E_j = \delta_{ij}$, so this is a frame field (defined on all of $\mathbf{R}^3$ except the z axis). We call it the *cylindrical frame field* on $\mathbf{R}^3$.

(2) *The spherical frame field on* $\mathbf{R}^3$ (Fig. 2.20). In a similar way, a frame field F_1, F_2, F_3 can be derived from the spherical coordinate functions ρ, ϑ, φ on $\mathbf{R}^3$. As indicated in the figure, we shall measure φ up from the xy plane rather than (as is usually done) down from the z axis.

Let E_1, E_2, E_3 be the cylindrical frame field. For spherical coordinates, the unit vector field F_2 in the direction of increasing ϑ is the same as above, so $F_2 = E_2$. The unit vector field F_1, in the direction of increasing ρ, points straight out from the origin; hence it can be expressed as

$$F_1 = \cos\varphi\; E_1 + \sin\varphi\; E_3$$

(Fig. 2.21). Similarly, the vector field for increasing φ is

$$F_3 = -\sin\varphi\, E_1 + \cos\varphi\, E_3.$$

Thus the formulas for E_1, E_2, E_3 in (1) yield

$$F_1 = \cos\varphi(\cos\vartheta\, U_1 + \sin\vartheta\, U_2) + \sin\varphi\, U_3,$$

$$F_2 = -\sin\vartheta\, U_1 + \cos\vartheta\, U_2,$$

$$F_3 = -\sin\varphi(\cos\vartheta\, U_1 + \sin\vartheta\, U_2) + \cos\varphi\, U_3.$$

By repeated use of the identity $\sin^2 + \cos^2 = 1$, we check that F_1, F_2, F_3 is a frame field—the *spherical frame field* on $\mathbf{R}^3$. (Its actual domain of definition is $\mathbf{R}^3$ minus the z axis, as in the cylindrical case.)

The following useful result is an immediate consequence of orthonormal expansion.

6.3 Lemma Let E_1, E_2, E_3 be a frame field on $\mathbf{R}^3$.

(1) If V is a vector field on $\mathbf{R}^3$, then $V = \sum f_i E_i$, where the functions $f_i = V \bullet E_i$ are called the *coordinate functions* of V with respect to E_1, E_2, E_3.

(2) If $V = \sum f_i E_i$ and $W = \sum g_i E_i$, then $V \bullet W = \sum f_i g_i$. In particular, $\| V \| = (\sum f_i^2)^{1/2}$.

Thus a given vector field V has a different set of coordinate functions with respect to each choice of a frame field E_1, E_2, E_3. The *Euclidean* coordinate functions (Lemma 2.5 of Chapter 1), of course, come from the natural frame field U_1, U_2, U_3. In Chapter 1, we used this natural frame field exclusively, but now we shall gradually shift to arbitrary frame fields. The reason is clear: In studying curves and surfaces in $\mathbf{R}^3$, we shall then be able to choose a frame field *specifically adapted to the problem at hand.* Not only does this simplify computations, but it gives a clearer understanding of geometry than if we had insisted on using the same frame field in every situation.

Exercises

1. If V and W are vector fields on $\mathbf{R}^3$ that are linearly independent at each point, show that

$$E_1 = \frac{V}{\| V \|}, \quad E_2 = \frac{\tilde{W}}{\| \tilde{W} \|}, \quad E_3 = E_1 \times E_2$$

is a frame field, where $\tilde{W} = W - (W \bullet E_1)E_1$.

2. Express each of the following vector fields (i) in terms of the cylindrical frame field (with coefficients in terms of r, ϑ, z) and (ii) in terms of the spherical frame field (with coefficients in terms of ρ, ϑ, φ):

(a) U_1. (b) $\cos\vartheta\, U_1 + \sin\vartheta\, U_2 + U_3$.
(c) $xU_1 + yU_2 + zU_3$.

3. Find a frame field E_1, E_2, E_3 such that

$$E_1 = \cos x\, U_1 + \sin x \cos z\, U_2 + \sin x \sin z\, U_3.$$

2.7 Connection Forms

Once more we state the essential point: The power of the Frenet formulas stems not from the fact that they tell what the derivatives T', N', B' are, but from the fact that *they express these derivatives in terms of* T, N, B—and thereby define curvature and torsion. We shall now do the same thing with an arbitrary frame field E_1, E_2, E_3 on $\mathbf{R}^3$; namely, *express the covariant derivatives of these vector fields in terms of the vector fields themselves.* We begin with the covariant derivative with respect to an arbitrary tangent vector $\mathbf{v}$ at a point $\mathbf{p}$. Then

$$\nabla_v E_1 = c_{11}E_1(\mathbf{p}) + c_{12}E_2(\mathbf{p}) + c_{13}E_3(\mathbf{p}),$$
$$\nabla_v E_2 = c_{21}E_1(\mathbf{p}) + c_{22}E_2(\mathbf{p}) + c_{23}E_3(\mathbf{p}),$$
$$\nabla_v E_3 = c_{31}E_1(\mathbf{p}) + c_{32}E_2(\mathbf{p}) + c_{33}E_3(\mathbf{p}),$$

where by orthonormal expansion the coefficients of these equations are

$$c_{ij} = \nabla_v E_i \bullet E_j(\mathbf{p}) \quad \text{for } 1 \leqq i,\ j \leqq 3.$$

These coefficients c_{ij}, depend on the particular tangent vector $\mathbf{v}$, so a better notation for them is

$$\omega_{ij}(\mathbf{v}) = \nabla_v E_i \bullet E_j(\mathbf{p}), \quad (1 \leqq i,\ j \leqq 3).$$

Thus for each choice of i and j, ω_{ij} is a real-valued function defined on all tangent vectors. But we have met that kind of function before.

7.1 Lemma Let E_1, E_2, E_3 be a frame field on $\mathbf{R}^3$. For each tangent vector $\mathbf{v}$ to $\mathbf{R}^3$ at the point $\mathbf{p}$, let

$$\omega_{ij}(\mathbf{v}) = \nabla_v E_i \bullet E_j(\mathbf{p}), \quad (1 \leqq i,\ j \leqq 3).$$

Then each ω_{ij} is a 1-form, and $\omega_{ij} = -\omega_{ji}$. These 1-forms are called the *connection forms* of the frame field E_1, E_2, E_3.

Proof. By definition, ω_{ij} is a real-valued function on tangent vectors, so to verify that ω_{ij} is a 1-form (Def. 5.1 of Ch. 1), it suffices to check the linearity condition. Using Theorem 5.3, we get

$$\begin{aligned}\omega_{ij}(a\mathbf{v} + b\mathbf{w}) &= \nabla_{av+bw}E_i \bullet E_j(\mathbf{p}) \\ &= (a\nabla_v E_i + b\nabla_w E_i) \bullet E_j(\mathbf{p}) \\ &= a\nabla_v E_i \bullet E_j(\mathbf{p}) + b\nabla_w E_i \bullet E_j(\mathbf{p}) \\ &= a\omega_{ij}(\mathbf{v}) + b\omega_{ij}(\mathbf{w}).\end{aligned}$$

To prove that $\omega_{ij} = -\omega_{ji}$ we must show that $\omega_{ij}(\mathbf{v}) = -\omega_{ji}(\mathbf{v})$ for every tangent vector $\mathbf{v}$. By definition of frame field, $E_i \bullet E_j = \delta_{ij}$, and since each Kronecker delta has constant value 0 or 1, the Leibnizian formula (4) of Theorem 5.3 yields

$$0 = \mathbf{v}[E_i \bullet E_j] = \nabla_v E_i \bullet E_j(\mathbf{p}) + E_i(\mathbf{p}) \bullet \nabla_v E_j.$$

By the symmetry of the dot product, the two vectors in this last term may be reversed, so we have found that $0 = \omega_{ij}(\mathbf{v}) + \omega_{ji}(\mathbf{v})$. ◆

The geometric significance of the connection forms is no mystery. The definition $\omega_{ij}(\mathbf{v}) = \nabla_v E_i \bullet E_j(\mathbf{p})$ shows that $\omega_{ij}(\mathbf{v})$ *is the initial rate at which* E_i *rotates toward* E_j *as* $\mathbf{p}$ *moves in the* $\mathbf{v}$ *direction.* Thus the 1-forms ω_{ij} contain this information for *all* tangent vectors to $\mathbf{R}^3$.

The following basic result is little more than a rephrasing of the definition of connection forms.

7.2 Theorem Let ω_{ij} $(1 \leqq i, j \leqq 3)$ be the connection forms of a frame field E_1, E_2, E_3 on $\mathbf{R}^3$. Then for any vector field V on $\mathbf{R}^3$,

$$\nabla_V E_i = \sum_j \omega_{ij}(V)E_j, \quad (1 \leqq i \leqq 3).$$

We call these the *connection equations* of the frame field E_1, E_2, E_3.

Proof. For fixed i, both sides of this equation are vector fields. Thus we must show that at each point $\mathbf{p}$,

$$\nabla_{V(p)}E_i = \sum_j \omega_{ij}(V(\mathbf{p}))E_j(\mathbf{p}).$$

But as we have already seen, the very definition of connection form makes this equation a consequence of orthonormal expansion. ◆

When $i = j$, the skew-symmetry condition $\omega_{ij} = -\omega_{ji}$ becomes $\omega_{ii} = -\omega_{ii}$; thus

$$\omega_{11} = \omega_{22} = \omega_{33} = 0.$$

Hence this condition has the effect of reducing the nine 1-forms ω_{ij} for $1 \leqq i, j \leqq 3$ to essentially only three, say ω_{12}, ω_{13}, ω_{23}. It is perhaps best to regard the connection forms ω_{ij} as the entries of a skew-symmetric matrix of 1-forms,

$$\omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{pmatrix} = \begin{pmatrix} 0 & \omega_{12} & \omega_{13} \\ -\omega_{12} & 0 & \omega_{23} \\ -\omega_{13} & -\omega_{23} & 0 \end{pmatrix}.$$

Thus in expanded form, the connection equations (Theorem 7.2) become

$$\begin{aligned} \nabla_V E_1 &= \qquad\qquad\qquad \omega_{12}(V)E_2 + \omega_{13}(V)E_3, \\ \nabla_V E_2 &= -\omega_{12}(V)E_1 \qquad\qquad\qquad + \omega_{23}(V)E_3, \\ \nabla_V E_3 &= -\omega_{13}(V)E_1 - \omega_{23}(V)E_2. \end{aligned} \qquad (*)$$

showing an obvious relation to the Frenet formulas

$$\begin{aligned} T' &= \qquad \kappa N, \\ N' &= -\kappa N \qquad\quad + \tau B, \\ B' &= \qquad -\tau N. \end{aligned}$$

The absence from the Frenet formulas of terms corresponding to $\omega_{13}(V)E_3$ and $-\omega_{13}(V)E_1$ is a consequence of the special way the Frenet frame field is fitted to its curve. Having gotten $T(\sim E_1)$, we chose $N(\sim E_2)$ so that the derivative T' would be a scalar multiple of N alone and not involve $B(\sim E_3)$.

Another difference between the Frenet formulas and the equations above stems from the fact that $\mathbf{R}^3$ has three dimensions, while a curve has but one. The coefficients—curvature κ and torsion τ—in the Frenet formulas measure the rate of change of the frame field T, N, B only along its curve, that is, in the direction of T alone. But the coefficients in the connection equations must be able to make this measurement for E_1, E_2, E_3 with respect to *arbitrary* vector fields in $\mathbf{R}^3$. This is why the connection forms are 1-forms and not just functions.

These formal differences aside, a more fundamental distinction stands out. It is because a Frenet frame field is specially fitted to its curve that the Frenet formulas give information about that curve. Since the frame field E_1, E_2, E_3 used above is completely arbitrary, the connection equations give no direct information about $\mathbf{R}^3$, but only information about the "rate of rotation" of that particular frame field. This is not a weakness, but a strength, since as indicated earlier, if we can fit a frame field to a geometric problem arising in $\mathbf{R}^3$, then the connection equations will give direct information about that problem. Thus, these equations play a fundamental role in all the differential geometry of $\mathbf{R}^3$. For example, the Frenet formulas can be deduced from them (Exercise 8).

Given an arbitrary frame field E_1, E_2, E_3 on $\mathbf{R}^3$, it is fairly easy to find an explicit formula for its connection forms. First use orthonormal expansion to express the vector fields E_1, E_2, E_3 in terms of the natural frame field U_1, U_2, U_3 on $\mathbf{R}^3$:

$$E_1 = a_{11}U_1 + a_{12}U_2 + a_{13}U_3,$$
$$E_2 = a_{21}U_1 + a_{22}U_2 + a_{23}U_3,$$
$$E_3 = a_{31}U_1 + a_{32}U_2 + a_{33}U_3.$$

Here each $a_{ij} = E_i \bullet U_j$ is a real-valued function on $\mathbf{R}^3$. The matrix

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

with these functions as entries is called the *attitude matrix* of the frame field E_1, E_2, E_3. In fact, at each point $\mathbf{p}$, the numerical matrix

$$A(\mathbf{p}) = (a_{ij}(\mathbf{p}))$$

is exactly the attitude matrix of the frame $E_1(\mathbf{p})$, $E_2(\mathbf{p})$, $E_3(\mathbf{p})$ as in Definition 1.6. Since attitude matrices are orthogonal, the transpose ${}^tA$ of A is equal to its inverse A^{-1}.

Define the differential of $A = (a_{ij})$ to be $dA = (da_{ij})$, so dA is a matrix whose entries are 1-forms. We can now give a simple expression for the connection forms in terms of the attitude matrix.

7.3 Theorem If $A = (a_{ij})$ is the attitude matrix and $\omega = (\omega_{ij})$ the matrix of connection forms of a frame field E_1, E_2, E_3, then

$$\omega = dA\,{}^tA \quad \text{(matrix multiplication)},$$

or equivalently,

$$\omega_{ij} = \sum_k a_{jk}\, da_{ik} \quad \text{for} \quad 1 \leqq i,\ j \leqq 3.$$

Since the proof is routine, it may be more informative to illustrate the result by an example. For the cylindrical frame field in Example 6.2, we found the attitude matrix

$$A = \begin{pmatrix} \cos\vartheta & \sin\vartheta & 0 \\ -\sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus

$$\omega = dA\,{}^tA = \begin{pmatrix} -\sin\vartheta\, d\vartheta & \cos\vartheta\, d\vartheta & 0 \\ -\cos\vartheta\, d\vartheta & -\sin\vartheta\, d\vartheta & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\vartheta & -\sin\vartheta & 0 \\ \sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & d\vartheta & 0 \\ -d\vartheta & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Since $\omega_{12} = d\vartheta$ is the only nonzero connection form (except, of course, $\omega_{21} = -\omega_{12}$), the connection equations (∗) reduce to

$$\begin{aligned} \nabla_V E_1 &= d\vartheta(V)E_2 = V[\vartheta]E_2, \\ \nabla_V E_2 &= -d\vartheta(V)E_1 = -V[\vartheta]E_1, \\ \nabla_V E_3 &= 0. \end{aligned}$$

These equations have immediate geometrical significance. Because V is arbitrary, the third equation says that the vector field E_3 is parallel. We knew this already since in the cylindrical frame field, E_3 is just U_3.

The first two equations tell us that the covariant derivatives of E_1 and E_2 with respect to a vector field V depend only on the rate of change of the angle ϑ in the V direction.

For example, the definition of ϑ shows that $V[\vartheta] = 0$ whenever V is a vector field that at each point is tangent to a plane through the z axis. Thus for a vector field of this type the connection equations above predict that $\nabla_V E_1 = \nabla_V E_2 = 0$. In fact, it is clear from Fig. 2.19 that E_1 and E_2 do remain parallel on any plane through the z axis.

Exercises

1. For any function f, show that the vector fields

$$E_1 = (\sin f\ U_1 + U_2 - \cos f\ U_3)/\sqrt{2},$$
$$E_2 = (\sin f\ U_1 - U_2 - \cos f\ U_3)/\sqrt{2},$$
$$E_3 = \cos f\ U_1 + \sin f\ U_3$$

form a frame field, and find its connection forms.

2. Find the connection forms of the natural frame field U_1, U_2, U_3.

3. For any function f, show that

$$A = \begin{pmatrix} \cos^2 f & \cos f \sin f & \sin f \\ \sin f \cos f & \sin^2 f & -\cos f \\ -\sin f & \cos f & 0 \end{pmatrix}$$

is the attitude matrix of a frame field, and compute its connection forms.

4. Prove that the connection forms of the spherical frame field are

$$\omega_{12} = \cos\varphi\ d\vartheta, \quad \omega_{13} = d\varphi, \quad \omega_{23} = \sin\varphi\ d\vartheta.$$

5. If E_1, E_2, E_3 is a frame field and $W = \sum f_i E_i$, prove the *covariant derivative formula:*

$$\nabla_V W = \sum_j \left\{ V[f_j] + \sum_i f_i \omega_{ij}(V) \right\} E_j.$$

6. Let E_1, E_2, E_3 be the cylindrical frame field. If V is a vector field such that $V[r] = r$ and $V[\vartheta] = 1$, compute $\nabla_V (r \cos\vartheta E_1 + r \sin\vartheta E_3)$.

7. (*Computer.*) (a) Write a computer command that, given the attitude matrix A of a frame field on $\mathbf{R}^3$, returns the matrix $\omega = dA\ {}^tA$ of its connection forms. (*Hint:* For *Maple*, use the differential operator d from the package *difforms*. For *Mathematica*, use the total differential Dt.) (b) Test part (a) on the cylindrical frame field and on the spherical frame field (Ex. 4).

8. Let β be a unit-speed curve in $\mathbf{R}^3$ with $\kappa > 0$, and suppose that E_1, E_2, E_3 is a frame field on $\mathbf{R}^3$ such that the restriction of these vector fields to β gives the Frenet-frame field T, N, B of β. Prove that

$$\omega_{12}(T) = \kappa, \quad \omega_{13}(T) = 0, \quad \omega_{23}(T) = \tau.$$

Then deduce the Frenet formulas from the connection equations. (*Hint:* Ex. 5 of Sec. 5.)

2.8 The Structural Equations

We have seen that 1-forms—the connection forms—give the simplest description of the rate of rotation of a frame field. Furthermore, the frame field itself can be described in terms of 1-forms.

8.1 Definition If E_1, E_2, E_3 is a frame field on $\mathbf{R}^3$, then the *dual* 1-*forms* $\theta_1, \theta_2, \theta_3$ of the frame field are the 1-forms such that

$$\theta_i(\mathbf{v}) = \mathbf{v} \bullet E_i(\mathbf{p})$$

for each tangent vector $\mathbf{v}$ to $\mathbf{R}^3$ at $\mathbf{p}$.

Note that θ_i is linear on the tangent vectors at each point; hence it *is* a 1-form. In particular, $\theta_i(E_j) = \delta_{ij}$, so readers familiar with the notion of dual vector spaces will recognize that at each point, $\theta_1, \theta_2, \theta_3$ gives the dual basis of E_1, E_2, E_3.

In the case of the natural frame field U_1, U_2, U_3, the dual forms are just dx_1, dx_2, dx_3. In fact, from Example 5.3 of Chapter 1 we get

$$dx_i(\mathbf{v}) = v_i = \mathbf{v} \bullet U_i(\mathbf{p})$$

for each tangent vector $\mathbf{v}$; hence $dx_i = \theta_i$.

Using dual forms, the orthonormal expansion formula in Lemma 6.3 may be written $V = \sum \theta_i(V)E_i$. In the characteristic fashion of duality, this formula becomes the following lemma.

8.2 Lemma Let $\theta_1, \theta_2, \theta_3$ be the dual 1-forms of a frame field E_1, E_2, E_3. Then any 1-form ϕ on $\mathbf{R}^3$ has a unique expression

$$\phi = \sum \phi(E_i)\theta_i.$$

Proof. Two 1-forms are the same if they have the same value on any vector field V. But

$$\begin{aligned}\left(\sum \phi(E_i)\theta_i\right)(V) &= \sum \phi(E_i)\theta_i(V)\\ &= \phi\left(\sum \theta_i(V)E_i\right) = \phi(V).\end{aligned}$$

◆

Thus ϕ is expressed in terms of dual forms of E_1, E_2, E_3 by evaluating it on E_1, E_2, E_3. This useful fact is the generalization to arbitrary frame fields of Lemma 5.4 of Chapter 1.

We compared a frame field E_1, E_2, E_3 to the natural frame field by means of its attitude matrix $A = (a_{ij})$, for which

$$E_i = \sum a_{ij} U_j \quad (1 \leq i \leq 3).$$

The dual formulation is just

$$\theta_i = \sum a_{ij} dx_j$$

with *the same coefficients.* In fact, by the preceding lemma,

$$\theta_i = \sum \theta_i(U_j) dx_j.$$

But

$$\theta_i(U_j) = E_i \bullet U_j = \left(\sum a_{ik} U_k\right) \bullet U_j = \sum a_{ik} \delta_{kj} = a_{ij}.$$

These formulas for E_i and θ_i show plainly that θ_1, θ_2, θ_3 is merely the dual description of the frame field E_1, E_2, E_3.

In calculus, when a new function appears on the scene, it is natural to ask what its derivative is. Similarly with 1-forms—having associated with each frame field its dual forms and connection forms, it is reasonable to ask what their exterior derivatives are. The answer is given by two neat sets of equations discovered by Cartan.

8.3 Theorem (Cartan structural equations.) Let E_1, E_2, E_3 be a frame field on $\mathbf{R}^3$ with dual forms θ_1, θ_2, θ_3 and connection forms ω_{ij} $(1 \leqq i, j \leqq 3)$. The exterior derivatives of these forms satisfy

(1) the *first structural equations:*

$$d\theta_i = \sum_j \omega_{ij} \wedge \theta_j \quad (1 \leqq i \leqq 3);$$

(2) the s*econd structural equations:*

$$d\omega_{ij} = \sum_k \omega_{ik} \wedge \omega_{kj} \quad (1 \leqq i,\ j \leqq 3).$$

Because θ_i is the dual of E_i, the first structural equations may be easily recognized as the dual of the connection equations. Only later experience will show that the second structural equations mean that $\mathbf{R}^3$ is flat—roughly speaking, in the same sense that the plane $\mathbf{R}^2$ is flat.

The most efficient proof of the structural equations requires some preliminary remarks. In the Cartan approach, the fundamental objects are not individual forms, but rather *matrices whose entries are forms.* We have already seen that the simplest description of the connection forms ω_{ij} of a frame field is as a single skew-symmetric matrix ω with entries ω_{ij}. Then, for example, ω is expressed in terms of the attitude matrix A of the frame field by the matrix equation $\omega = dA\,{}^tA$. (Here, as always, to apply d to a matrix, apply it to each entry of the matrix.)

Similarly, the dual forms of a frame field can be described by a single $n \times 1$ matrix θ with entries θ_i. If ξ is the $n \times 1$ matrix whose entries are the natural coordinates x_i of $\mathbf{R}^3$, then

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \quad \text{and} \quad d\xi = \begin{pmatrix} dx_1 \\ dx_2 \\ dx_3 \end{pmatrix},$$

so the formula $\theta_i = \sum a_{ij}\, dx_j$ above can be written as

$$\theta = A\, d\xi.$$

For such matrices of forms, matrix multiplication is defined as usual, but of course when *entries* are multiplied it is by the wedge product.

The proof of Theorem 8.3 is now quite simple. Recall that since the attitude matrix A is orthogonal, ${}^tAA$ is the identity matrix I, which can be inserted in any matrix formula without effect.

Proof of the First Structural Equation. Since $d^2 = 0$, we evidently have $d(d\xi) = 0$, so

$$d\theta = d(A\, d\xi) = dA \cdot d\xi = dA\,{}^tA \cdot A\, d\xi = \omega\theta.$$

Expressed in terms of entries, this is indeed the version in (1) of Theorem 8.3.

Proof of the Second Structural Equation. For functions f and g.

$$d(df\; g) = d(g\, df) = dg \wedge df = -df \wedge dg.$$

Thus, using the transpose rule ${}^t(AB) = {}^tB\,{}^tA$, we get

$$d\omega = d(dA\,{}^tA) = -dA \cdot d({}^tA) = -dA\,{}^tA \cdot A^t(dA) = -\omega{}^t\omega = \omega\omega,$$

where the last step uses the skew-symmetry of ω. Again, in terms of entries, this is the version in (2) of Theorem 8.3. ◆

8.4 Example Structural equations for the spherical frame field (Example 6.2). The dual forms and connection forms are

$$\begin{aligned} \theta_1 &= d\rho, & \omega_{12} &= \cos\varphi\, d\vartheta, \\ \theta_2 &= \rho\cos\varphi\, d\vartheta, & \omega_{13} &= d\varphi, \\ \theta_3 &= \rho\, d\varphi, & \omega_{23} &= \sin\varphi\, d\vartheta. \end{aligned}$$

Let us check, say, the first structural equation

$$d\theta_3 = \sum \omega_{3j} \wedge \theta_j = \omega_{31} \wedge \theta_1 + \omega_{32} \wedge \theta_2.$$

Using the skew-symmetry $\omega_{ij} = -\omega_{ji}$ and the general properties of forms developed in Chapter 1, we get

$$\omega_{31} \wedge \theta_1 = -d\varphi \wedge d\rho = d\rho \wedge d\varphi,$$

$$\omega_{32} \wedge \theta_2 = (-\sin\varphi\, d\vartheta) \wedge (\rho\cos\varphi\, d\vartheta) = 0$$

(the latter since $d\vartheta \wedge d\vartheta = 0$). The sum of these terms is, correctly,

$$d\theta_3 = d(\rho\, d\varphi) = d\rho \wedge d\varphi.$$

Second structural equations involve only one wedge product. For example, since $\omega_{11} = \omega_{22} = 0$,

$$d\omega_{12} = \sum \omega_{1k} \wedge \omega_{k2} = \omega_{13} \wedge \omega_{32}.$$

In this case,

$$\omega_{13} \wedge \omega_{32} = d\varphi \wedge (-\sin\varphi\, d\vartheta) = -\sin\varphi\, d\varphi \wedge d\vartheta.$$

which is the same as

$$d\omega_{12} = d(\cos\varphi\, d\vartheta) = d(\cos\varphi) \wedge d\vartheta = -\sin\varphi\, d\varphi \wedge d\vartheta.$$

To derive the expressions given above for the dual 1-forms, first compute dx_1, dx_2, dx_3 by differentiating the well-known equations

$$\begin{aligned} x_1 &= \rho\cos\varphi\cos\vartheta, \\ x_2 &= \rho\cos\varphi\sin\vartheta, \\ x_3 &= \rho\sin\vartheta. \end{aligned}$$

Then substitute in the formula $\theta_i = \sum a_{ij}\, dx_j$, where $A = (a_{ij})$ is the attitude matrix from Example 6.2. This result, somewhat disguised, is derived in elementary calculus by a familiar plausibility argument: If at each point the spherical coordinates ρ, ϑ, φ are altered by increments $d\rho$, $d\vartheta$, $d\varphi$, then the sides of the resulting infinitesimal box (Fig. 2.22) are $d\rho$, $\rho\cos\varphi\, d\vartheta$, $\rho\, d\varphi$. These are exactly the formulas for θ_1, θ_2, θ_3.

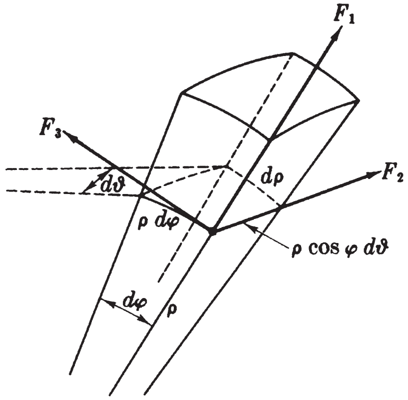


FIG. 2.22

The structural equations provide a powerful method for dealing with geometrical problems in $\mathbf{R}^3$: Select a frame field well adapted to the problem at hand; find its dual 1-forms and connection forms; apply the structural equations; interpret the results. We will use this method later to study the geometry of surfaces in $\mathbf{R}^3$.

Exercises

1. For a 1-form $\phi = \sum f_i\theta_i$, prove

$$d\phi = \sum_j \left\{ df_j + \sum_i f_i\omega_{ij} \right\} \wedge \theta_j.$$

(Compare Ex. 5 of Sec. 7.)

2. Check all the structural equations of the spherical frame field.

3. For the cylindrical frame field E_1, E_2, E_3.
(a) Starting from the basic cylindrical equations $x = r\cos\vartheta$, $y = r\sin\vartheta$, $z = z$, show that the dual 1-forms are

$$\theta_1 = dr, \quad \theta_2 = r\,d\vartheta, \quad \theta_3 = dz.$$

(b) Deduce that $E_1[r] = 1$, $E_2[\vartheta] = 1/r$, $E_3[z] = 1$ and that the other six possibilities $E_1[\vartheta]$, . . . are all zero.
(c) For a function $f(r, \vartheta, z)$, show that

$$E_1[f] = \frac{\partial f}{\partial r}, \quad E_2[f] = \frac{1}{r}\frac{\partial f}{\partial \vartheta}, \quad E_3[f] = \frac{\partial f}{\partial z}.$$

4. Frame fields on $\mathbf{R}^2$. Given a frame field E_1, E_2 on $\mathbf{R}^2$ there is an angle function ψ such that

$$E_1 = \cos\psi\, U_1 + \sin\psi\, U_2,$$
$$E_2 = -\sin\psi\, U_1 + \cos\psi\, U_2.$$

(a) Express the connection form and dual 1-forms in terms of ψ and the natural coordinates x, y.
(b) What are the structural equations in this case? Check that the results in part (a) satisfy these equations.

(*Hint:* Defining $E_3 = U_3$ gives a frame field on $\mathbf{R}^3$.)

2.9 Summary

We have accomplished the aims set at the beginning of this chapter. The idea of a moving frame has been expressed rigorously as a *frame field*—either on a curve in $\mathbf{R}^3$ or on an open set of $\mathbf{R}^3$ itself. In the case of a curve, we used only the Frenet frame field T, N, B of the curve. Expressing the derivatives of these vector fields in terms of the vector fields themselves, we discovered the *curvature* and *torsion* of the curve. It is already clear that curvature and torsion tell a lot about the geometry of a curve; we shall find in Chapter 3 that they tell everything. In the case of an open set of $\mathbf{R}^3$, we dealt with an arbitrary frame field E_1, E_2, E_3. Cartan's generalization (Theorem 7.2) of the Frenet formulas followed the same pattern of expressing the (covariant) derivatives of these vector fields in terms of the vector fields themselves. Omitting the vector field V from the notation in Theorem 7.2, we have

Cartan	*Frenet*
$\nabla E_1 = \omega_{12}E_2 + \omega_{13}E_3,$	$T' = \kappa N,$
$\nabla E_2 = -\omega_{12}E_1 + \omega_{23}E_3,$	$N' = -\kappa T + \tau B,$
$\nabla E_3 = -\omega_{13}E_1 - \omega_{23}E_2,$	$B' = -\tau N.$

Cartan's equations are not conspicuously more complicated than Frenet's, because the notion of 1-form is available for the coefficients ω_{ij}, the connection forms.

Chapter 3
Euclidean Geometry

We recall some familiar features of plane geometry. First of all, two triangles are *congruent* if there is a rigid motion of the plane that carries one triangle exactly onto the other. Corresponding angles of congruent triangles are equal, corresponding sides have the same length, the areas enclosed are equal, and so on. Indeed, any geometric property of a given triangle is automatically shared by every congruent triangle. Conversely, there are a number of simple ways in which one can decide whether two given triangles are congruent—for example, if for each the same three numbers occur as lengths of sides.

In this chapter we shall investigate the rigid motions (isometries) of Euclidean space, and see how these remarks about triangles can be extended to other geometric objects.

3.1 Isometries of $\mathbf{R}^3$

An isometry, or rigid motion, of Euclidean space is a mapping that preserves the Euclidean distance d between points (Definition 1.2, Chapter 2).

1.1 Definition An *isometry* of $\mathbf{R}^3$ is a mapping $F\colon \mathbf{R}^3 \to \mathbf{R}^3$ such that

$$d(F(\mathbf{p}), F(\mathbf{q})) = d(\mathbf{p}, \mathbf{q})$$

for all points $\mathbf{p}$, $\mathbf{q}$ in $\mathbf{R}^3$.

1.2 Example (1) ***Translations***. Fix a point $\mathbf{a}$ in $\mathbf{R}^3$ and let T be the mapping that adds $\mathbf{a}$ to every point of $\mathbf{R}^3$. Thus $T(\mathbf{p}) = \mathbf{p} + \mathbf{a}$ for all

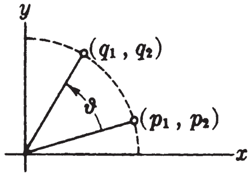


FIG. 3.1

points $\mathbf{p}$. T is called *translation* by $\mathbf{a}$. It is easy to see that T is an isometry, since

$$\begin{aligned} d(T(\mathbf{p}),\ T(\mathbf{q})) &= d(\mathbf{p} + \mathbf{a}, \mathbf{q} + \mathbf{a}) \\ &= \|(\mathbf{p} + \mathbf{a}) - (\mathbf{q} + \mathbf{a})\| \\ &= \|\mathbf{p} - \mathbf{q}\| = d(\mathbf{p}, \mathbf{q}). \end{aligned}$$

(2) ***Rotation around a coordinate axis.*** A rotation of the xy plane through an angle ϑ carries the point (p_1, p_2) to the point (q_1, q_2) with coordinates (Fig. 3.1)

$$\begin{aligned} q_1 &= p_1 \cos \vartheta - p_2 \sin \vartheta, \\ q_2 &= p_1 \sin \vartheta + p_2 \cos \vartheta. \end{aligned}$$

Thus a *rotation* C of three-dimensional Euclidean space $\mathbf{R}^3$ around the z axis, through an angle ϑ, has the formula

$$C(\mathbf{p}) = C(p_1, p_2, p_3) = (p_1 \cos \vartheta - p_2 \sin \vartheta, p_1 \sin \vartheta + p_2 \cos \vartheta, p_3).$$

Evidently, the mapping C is a linear transformation. A straightforward computation shows that C preserves Euclidean distance, so it is an isometry.

Recall that if F and G are mappings of $\mathbf{R}^3$, the composite function GF is a mapping of $\mathbf{R}^3$ obtained by applying first F, then G.

1.3 Lemma If F and G are isometries of $\mathbf{R}^3$, then the composite mapping GF is also an isometry of $\mathbf{R}^3$.

Proof. Since G is an isometry, the distance from $G(F(\mathbf{p}))$ to $G(F(\mathbf{q}))$ is $d(F(\mathbf{p}), F(\mathbf{q}))$. But since F is an isometry, this distance equals $d(\mathbf{p}, \mathbf{q})$. Thus GF preserves distance; hence it is an isometry. ◆

In short, a composition of isometries is again an isometry.

We also recall that if $F\colon \mathbf{R}^3 \to \mathbf{R}^3$ is both one-to-one and onto, then F has a unique inverse function $F^{-1}\colon \mathbf{R}^3 \to \mathbf{R}^3$, which sends each point $F(\mathbf{p})$ back to $\mathbf{p}$. The relationship between F and F^{-1} is best described by the formulas

$$FF^{-1} = I, \quad F^{-1}F = I,$$

where I is the *identity mapping* of $\mathbf{R}^3$, that is, the mapping such that $I(\mathbf{p}) = \mathbf{p}$ for all $\mathbf{p}$.

Translations of $\mathbf{R}^3$ (as defined in Example 1.2) are the simplest type of isometry.

1.4 Lemma (1) If S and T are translations, then $ST = TS$ is also a translation.

(2) If T is translation by $\mathbf{a}$, then T has an inverse T^{-1}, which is translation by $-\mathbf{a}$.

(3) Given any two points $\mathbf{p}$ and $\mathbf{q}$ of $\mathbf{R}^3$, there exists a unique translation T such that $T(\mathbf{p}) = \mathbf{q}$.

Proof. To prove (3), for example, note that translation by $\mathbf{q} - \mathbf{p}$ certainly carries $\mathbf{p}$ to $\mathbf{q}$. This is the only possibility, since if T is translation by $\mathbf{a}$ and $T(\mathbf{p}) = \mathbf{q}$, then $\mathbf{p} + \mathbf{a} = \mathbf{q}$; hence $\mathbf{a} = \mathbf{q} - \mathbf{p}$. ◆

A useful special case of (3) is that if T is a translation such that for some one point $T(\mathbf{p}) = \mathbf{p}$, then $T = I$.

The rotation in Example 1.2 is an example of an *orthogonal transformation* of $\mathbf{R}^3$, that is, a linear transformation $C\colon \mathbf{R}^3 \to \mathbf{R}^3$ that preserves dot products in the sense that

$$C(\mathbf{p}) \cdot C(\mathbf{q}) = \mathbf{p} \cdot \mathbf{q} \quad \text{for all } \mathbf{p}, \mathbf{q}.$$

1.5 Lemma If $C\colon \mathbf{R}^3 \to \mathbf{R}^3$ is an orthogonal transformation, then C is an isometry of $\mathbf{R}^3$.

Proof. First we show that C preserves norms. By definition, $\|\mathbf{p}\|^2 = \mathbf{p} \cdot \mathbf{p}$; hence

$$\|C(\mathbf{p})\|^2 = C(\mathbf{p}) \cdot C(\mathbf{p}) = \mathbf{p} \cdot \mathbf{p} = \|\mathbf{p}\|^2.$$

Thus $\| C(\mathbf{p}) \| = \| \mathbf{p} \|$ for all points $\mathbf{p}$. Since C is linear, it follows easily that C is an isometry:

$$d(C(\mathbf{p}), C(\mathbf{q})) = \|C(\mathbf{p}) - C(\mathbf{q})\| = \|C(\mathbf{p} - \mathbf{q})\| = \|\mathbf{p} - \mathbf{q}\|$$
$$= d(\mathbf{p}, \mathbf{q}) \quad \text{for all } \mathbf{p}, \mathbf{q}.$$ ◆

Our goal now is Theorem 1.7, which asserts that every isometry can be expressed as an orthogonal transformation followed by a translation. The main part of the proof is the following converse of Lemma 1.5.

1.6 Lemma If F is an isometry of $\mathbf{R}^3$ such that $F(\mathbf{0}) = \mathbf{0}$, then F is an orthogonal transformation.

Proof. First we show that F preserves dot products; then we show that F is a linear transformation. Note that by definition of Euclidean distance, the norm $\| \mathbf{p} \|$ of a point $\mathbf{p}$ is just the Euclidean distance $d(\mathbf{0}, \mathbf{p})$ from the origin to $\mathbf{p}$. By hypothesis, F preserves Euclidean distance, and $F(\mathbf{0}) = \mathbf{0}$; hence

$$\|F(\mathbf{p})\| = d(\mathbf{0}, F(\mathbf{p})) = d(F(\mathbf{0}), F(\mathbf{p})) = d(\mathbf{0}, \mathbf{p}) = \|\mathbf{p}\|.$$

Thus F preserves norms. Now by a standard trick ("polarization"), we shall deduce that it also preserves dot products. Since F is an isometry,

$$d(F(\mathbf{p}), F(\mathbf{q})) = d(\mathbf{p}, \mathbf{q})$$

for any pair of points. Hence

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

By the definition of norm, this implies

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

Hence

$$\|F(\mathbf{p})\|^2 - 2F(\mathbf{p}) \bullet F(\mathbf{q}) + \|F(\mathbf{q})\|^2 = \|\mathbf{p}\|^2 - 2\mathbf{p} \bullet \mathbf{q} + \|\mathbf{q}\|^2.$$

The norm terms here cancel, since F preserves norms, and we find

$$F(\mathbf{p}) \bullet F(\mathbf{q}) = \mathbf{p} \bullet \mathbf{q},$$

as required.

It remains to prove that F is linear. Let $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ be the unit points (1, 0, 0), (0, 1, 0), (0, 0, 1), respectively. Then we have the identity

$$\mathbf{p} = (p_1, p_2, p_3) = \sum p_i \mathbf{u}_i.$$

Also, the points $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ are orthonormal; that is, $\mathbf{u}_i \bullet \mathbf{u}_j = \delta_{ij}$.

We know that F preserves dot products, so $F(\mathbf{u}_1)$, $F(\mathbf{u}_2)$, $F(\mathbf{u}_3)$ must also be orthonormal. Thus orthonormal expansion gives

$$F(\mathbf{p}) = \sum F(\mathbf{p}) \bullet F(\mathbf{u}_i)\; F(\mathbf{u}_i).$$

But

$$F(\mathbf{p}) \bullet F(\mathbf{u}_i) = \mathbf{p} \bullet \mathbf{u}_i = p_i,$$

so

$$F(\mathbf{p}) = \sum p_i F(\mathbf{u}_i).$$

Using this identity, it is a simple matter to check the linearity condition

$$F(a\mathbf{p} + b\mathbf{q}) = aF(\mathbf{p}) + bF(\mathbf{q}). \qquad \blacklozenge$$

We now give a concrete description of an arbitrary isometry.

1.7 Theorem If F is an isometry of $\mathbf{R}^3$, then there exist a unique translation T and a unique orthogonal transformation C such that

$$F = TC.$$

Proof. Let T be translation by $F(\mathbf{0})$. Then Lemma 1.4 shows that T^{-1} is translation by $-F(\mathbf{0})$. But $T^{-1}F$ is an isometry, by Lemma 1.3, and furthermore,

$$(T^{-1}F)(\mathbf{0}) = T^{-1}(F(\mathbf{0})) = F(\mathbf{0}) - F(\mathbf{0}) = \mathbf{0}.$$

Thus by Lemma 1.6, $T^{-1}F$ is an orthogonal transformation, say $T^{-1}F = C$. Applying T on the left, we get $F = TC$.

To prove the required uniqueness, we suppose that F can also be expressed as $\overline{T}\,\overline{C}$, where $\overline{T}$ is a translation and $\overline{C}$ an orthogonal transformation. We must prove $\overline{T} = T$ and $\overline{C} = C$. Now $TC = \overline{T}\,\overline{C}$; hence $C = T^{-1}\overline{T}\,\overline{C}$. Since C and $\overline{C}$ are linear transformations, they of course send the origin to itself. It follows that $(T^{-1}\overline{T})(\mathbf{0}) = \mathbf{0}$. But since $T^{-1}\overline{T}$ is a translation, we conclude that $T^{-1}\overline{T} = I$; hence $\overline{T} = T$. Then the equation $TC = \overline{T}\,\overline{C}$ becomes $TC = T\overline{C}$. Applying T^{-1} gives $C = \overline{C}$ $\blacklozenge$

Thus every isometry of $\mathbf{R}^3$ *can be* uniquely *described as an orthogonal transformation followed by a translation.* When $F = TC$ as in Theorem 1.7, we call

C the *orthogonal part of F,* and T the *translation part of F.* Note that CT is generally not the same as TC (Exercise 1).

This decomposition theorem is the decisive fact about isometries of $\mathbf{R}^3$ (and its proof holds for $\mathbf{R}^n$ as well). We will use it to find an explicit formula for an arbitrary isometry.

First, recall from linear algebra that if C: $\mathbf{R}^3 \to \mathbf{R}^3$ is *any* linear transformation, its *matrix* (relative to the natural basis of $\mathbf{R}^3$) is the 3×3 matrix $\{c_{ij}\}$ such that

$$C(p_1, p_2, p_3) = \left(\sum c_{1j}p_j, \sum c_{2j}p_j, \sum c_{3j}p_j\right).$$

Thus, using the *column-vector* conventions, $\mathbf{q} = C(\mathbf{p})$ can be written as

$$\begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}.$$

By a standard result of linear algebra, *a linear transformation of* $\mathbf{R}^3$ *is orthogonal* (preserves dot products) *if and only if its matrix is orthogonal* (transpose equals inverse).

Returning to the decomposition $F = TC$ in Theorem 1.7, if T is translation by $\mathbf{a} = (a_1, a_2, a_3)$, then

$$F(\mathbf{p}) = TC(\mathbf{p}) = \mathbf{a} + C(\mathbf{p}).$$

Using the above formula for $C(\mathbf{p})$, we get

$$F(\mathbf{p}) = F(p_1, p_2, p_3) = \left(a_1 + \sum c_{1j}p_j, a_2 + \sum c_{2j}p_j, a_3 + \sum c_{3j}p_j\right).$$

Alternatively, using the column-vector conventions, $\mathbf{q} = F(\mathbf{p})$ means

$$\begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}.$$

Exercises

Throughout these exercises, A, B, and C denote orthogonal transformations (or their matrices), and T_a is translation by $\mathbf{a}$.

1. Prove that $CT_a = T_{C(a)}C$.

2. Given isometries $F = T_aA$ and $G = T_bB$, find the translation and orthogonal part of FG and GF.

3. Show that an isometry $F = T_aC$ has an inverse mapping F^{-1}, which is also an isometry. Find the translation and orthogonal parts of F^{-1}.

4. If

$$C = \begin{pmatrix} -2/3 & 2/3 & -1/3 \\ 2/3 & 1/3 & -2/3 \\ 1/3 & 2/3 & 2/3 \end{pmatrix} \quad \text{and} \quad \begin{cases} \mathbf{p} = (3, 1, -6), \\ \mathbf{q} = (1, 0, 3), \end{cases}$$

show that C is orthogonal; then compute $C(\mathbf{p})$ and $C(\mathbf{q})$, and check that $C(\mathbf{p}) \bullet C(\mathbf{q}) = \mathbf{p} \bullet \mathbf{q}$.

5. Let $F = T_aC$, where $\mathbf{a} = (1, 3, -1)$ and

$$C = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}.$$

If $\mathbf{p} = (2, -2, 8)$, find the coordinates of the point $\mathbf{q}$ for which

(a) $\mathbf{q} = F(\mathbf{p})$. (b) $\mathbf{q} = F^{-1}(\mathbf{p})$.

(c) $\mathbf{q} = (CT_a)(\mathbf{p})$.

6. In each case decide whether F is an isometry of $\mathbf{R}^3$. If so, find its translation and orthogonal parts.

(a) $F(\mathbf{p}) = -\mathbf{p}$. (b) $F(\mathbf{p}) = (\mathbf{p} \bullet \mathbf{a})\,\mathbf{a}$, where $\|\mathbf{a}\| = 1$.

(c) $F(\mathbf{p}) = (p_3 - 1, p_2 - 2, p_1 - 3)$. (d) $F(\mathbf{p}) = (p_1, p_2, 1)$.

A *group* G is a set furnished with an *operation* that assigns to each pair g_1, g_2 of elements of G an element g_1g_2, subject to these rules: (1) associative law: $(g_1g_2)g_3 = g_1(g_2g_3)$, (2) there is a unique *identity element* e such that $eg = ge = g$ for all g in G, and (3) inverses: For each g in G there is an element g^{-1} in G such that $gg^{-1} = g^{-1}g = e$.

Groups occur naturally in many parts of geometry, and we shall mention a few in subsequent exercises. Basic properties of groups may be found in a variety of elementary textbooks.

7. Prove that the set $\mathscr{E}(3)$ of all isometries of $\mathbf{R}^3$ forms a group—with composition of functions as the operation. $\mathscr{E}(3)$ is called the *Euclidean group* of order 3.

A subset H of a group G is a *subgroup* of G provided (1) if g_1 and g_2 are in H, then so is g_1g_2, (2) is g is in H, so is g^{-1}, and hence (3) the identity element e of G is in H. A subgroup H of G is automatically a group.

8. Prove that the set $\mathscr{T}(3)$ of all translations of $\mathbf{R}^3$ and the set $O(3)$ of all orthogonal transformations of $\mathbf{R}^3$ are each subgroups of the Euclidean group $\mathscr{E}(3)$. $O(3)$ is called the *orthogonal group* of order 3. Which isometries of $\mathbf{R}^3$ are in both these subgroups?

It is easy to check that the results of this section, though stated for $\mathbf{R}^3$, remain valid for Euclidean spaces $\mathbf{R}^n$ of any dimension.

9. (a) Give an explicit description of an arbitrary 2×2 orthogonal matrix C. (*Hint:* Use an angle and a sign.)
(b) Give a formula for an arbitrary isometry F of $\mathbf{R} = \mathbf{R}^1$.

3.2 The Tangent Map of an Isometry

In Chapter 1 we showed that an arbitrary mapping F: $\mathbf{R}^3 \rightarrow \mathbf{R}^3$ has a tangent map F_* that carries each tangent vector $\mathbf{v}$ at $\mathbf{p}$ to a tangent vector $F_*(\mathbf{v})$ at $F(\mathbf{p})$. If F is an isometry, its tangent map is remarkably simple. (Since the distinction between tangent vector and point is crucial here, we temporarily restore the point of application to the notation.)

2.1 Theorem Let F be an isometry of $\mathbf{R}^3$ with orthogonal part C. Then

$$F_*(\mathbf{v}_p) = C(\mathbf{v})_{F(p)}$$

for all tangent vectors $\mathbf{v}_p$ to $\mathbf{R}^3$.

Verbally: To get $F_*(\mathbf{v}_p)$, first shift the tangent vector $\mathbf{v}_p$ to the canonically corresponding point $\mathbf{v}$ of $\mathbf{R}^3$, then apply the orthogonal part C of F, and finally shift this point $C(\mathbf{v})$ to the canonically corresponding tangent vector at $F(\mathbf{p})$ (Fig. 3.2). *Thus all tangent vectors at all points* $\mathbf{p}$ *of* $\mathbf{R}^3$ *are "rotated" in exactly the same way by* F_**—only the new point of application* $F(\mathbf{p})$ *depends on* $\mathbf{p}$.

Proof. Write $F = TC$ as in Theorem 1.7. Let T be translation by $\mathbf{a}$, so $F(\mathbf{p}) = \mathbf{a} + C(\mathbf{p})$. If $\mathbf{v}_p$ is a tangent vector to $\mathbf{R}^3$, then by Definition 7.4 of Chapter 1, $F_*(\mathbf{v}_p)$ is the initial velocity of the curve $t \rightarrow F(\mathbf{p} + t\mathbf{v})$. But using the linearity of C, we obtain

$$\begin{aligned} F(\mathbf{p} + t\mathbf{v}) &= TC(\mathbf{p} + t\mathbf{v}) = T(C(\mathbf{p}) + tC(\mathbf{v})) = \mathbf{a} + C(\mathbf{p}) + tC(\mathbf{v}) \\ &= F(\mathbf{p}) + tC(\mathbf{v}). \end{aligned}$$

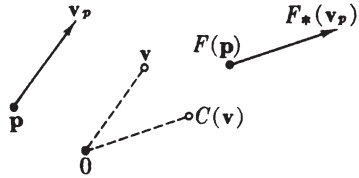


FIG. 3.2

Thus $F_*(\mathbf{v}_p)$ is the initial velocity of the curve $t \to F(\mathbf{p}) + tC(\mathbf{v})$, which is precisely the tangent vector $C(\mathbf{v})_{F(p)}$. ◆

Expressed in terms of Euclidean coordinates, this result becomes

$$F_*\left(\sum_j v_j U_j\right) = \sum_{i,j} c_{ij} v_j \overline{U}_i,$$

where $C = (c_{ij})$ is the orthogonal part of the isometry F, and if U_i is evaluated at $\mathbf{p}$, then $\overline{U}_i$ is evaluated at $F(\mathbf{p})$.

2.2 Corollary Isometries preserve dot products of tangent vectors. That is, if $\mathbf{v}_p$ and $\mathbf{w}_p$ are tangent vectors to $\mathbf{R}^3$ at the same point, and F is an isometry, then

$$F_*(\mathbf{v}_p) \bullet F_*(\mathbf{w}_p) = \mathbf{v}_p \bullet \mathbf{w}_p.$$

Proof. Let C be the orthogonal part of F, and recall that C, being an orthogonal transformation, preserves dot products in $\mathbf{R}^3$. By Theorem 2.1,

$$F_*(\mathbf{v}_p) \bullet F_*(\mathbf{w}_p) = C(\mathbf{v})_{F(p)} \bullet C(\mathbf{w})_{F(p)} = C(\mathbf{v}) \bullet C(\mathbf{w})$$
$$= \mathbf{v} \bullet \mathbf{w} = \mathbf{v}_p \bullet \mathbf{w}_p$$

where we have twice used Definition 1.3 of Chapter 2 (dot products of tangent vectors). ◆

Since dot products are preserved, it follows automatically that derived concepts such as norm and orthogonality are preserved. Explicitly, if F is an isometry, then $\| F_*(\mathbf{v}) \| = \| \mathbf{v} \|$, and if $\mathbf{v}$ and $\mathbf{w}$ are orthogonal, so are $F_*(\mathbf{v})$ and $F_*(\mathbf{w})$. Thus frames are also preserved: if $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is a frame at some point $\mathbf{p}$ of $\mathbf{R}^3$ and F is an isometry, then $F_*(\mathbf{e}_1), F_*(\mathbf{e}_2), F_*(\mathbf{e}_3)$ is a frame at $F(\mathbf{p})$. (A direct proof is easy: $\mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij}$, so by Corollary 2.2, $F_*(\mathbf{e}_i) \bullet F_*(\mathbf{e}_j) = \mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij}$.)

Assertion (3) of Lemma 1.4 shows how two *points* uniquely determine a translation. We now show that two *frames* uniquely determine an isometry.

2.3 Theorem Given any two frames on $\mathbf{R}^3$, say $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ at the point $\mathbf{p}$ and $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$ at the point $\mathbf{q}$, there exists a unique isometry F of $\mathbf{R}^3$ such that $F_*(\mathbf{e}_i) = \mathbf{f}_i$ for $1 \leqq i \leqq 3$.

Proof. First we show that there is such an isometry. Let $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3$, and $\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \hat{\mathbf{f}}_3$ be the points of $\mathbf{R}^3$ canonically corresponding to the vectors in the two frames. Let C be the unique linear transformation of $\mathbf{R}^3$ such that $C(\hat{\mathbf{e}}_i) = \hat{\mathbf{f}}_i$ for $1 \leqq i \leqq 3$. It is easy to check that C is orthogonal. Then let T be a translation by the point $\mathbf{q} - C(\mathbf{p})$. Now we assert that the isometry $F = TC$ carries the $\mathbf{e}$ frame to the $\mathbf{f}$ frame. First note that

$$F(\mathbf{p}) = T(C(\mathbf{p})) = \mathbf{q} - C(\mathbf{p}) + C(\mathbf{p}) = \mathbf{q}.$$

Then using Theorem 2.1 we get

$$F_*(\mathbf{e}_i) = C(\hat{\mathbf{e}}_i)_{F(p)} = (\hat{\mathbf{f}}_i)_{F(p)} = (\hat{\mathbf{f}}_i)_q = \mathbf{f}_i$$

for $1 \leqq i \leqq 3$.

To prove uniqueness, we observe that by Theorem 2.1 this choice of C is the *only* possibility for the orthogonal part of the required isometry. The translation part is then completely determined also, since it must carry $C(\mathbf{p})$ to $\mathbf{q}$. Thus the isometry $F = TC$ is uniquely determined. ◆

To compute the isometry in the theorem, recall that the attitude matrix A of the $\mathbf{e}$ frame has the Euclidean coordinates of $\mathbf{e}_i$ as its ith row: a_{i1}, a_{i2}, a_{i3}. The attitude matrix B of the $\mathbf{f}$ frame is similar. We claim that C in the theorem (or strictly speaking, its matrix) is ${}^tBA$. To verify this it suffices to check that ${}^tBA(\mathbf{e}_i) = \mathbf{f}_i$, since this uniquely characterizes C. For $i = 1$ we find, using the column-vector conventions,

$$\begin{aligned} {}^tBA\begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} &= {}^tB\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} \\ &= {}^tB\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{21} & b_{31} \\ b_{12} & b_{22} & b_{32} \\ b_{13} & b_{23} & b_{33} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} b_{11} \\ b_{12} \\ b_{13} \end{pmatrix}, \end{aligned}$$

that is, ${}^tBA(\mathbf{e}_1) = \mathbf{f}_1$. The cases $i = 2, 3$ are similar; hence $C = {}^tBA$. As noted above, T is then necessarily translated by $\mathbf{q} - C(\mathbf{p})$.

Exercises

1. If T is a translation, show that for every tangent vector $\mathbf{v}$ the vector $T(\mathbf{v})$ is parallel to $\mathbf{v}$ (same Euclidean coordinates).

2. Prove the general formulas $(GF)_* = G_*F_*$ and $(F^{-1})_* = (F_*)^{-1}$ in the special case where F and G are isometries of $\mathbf{R}^3$.

3. Given the frame

$$\mathbf{e}_1 = (2, 2, 1)/3, \qquad \mathbf{e}_2 = (-2, 1, 2)/3, \qquad \mathbf{e}_3 = (1, -2, 2)/3$$

at $\mathbf{p} = (0, 1, 0)$ and the frame

$$\mathbf{f}_1 = (1, 0, 1)/\sqrt{2}, \qquad \mathbf{f}_2 = (0, 1, 0), \qquad \mathbf{f}_3 = (1, 0, -1)/\sqrt{2}$$

at $\mathbf{q} = (3, -1, 1)$, find a and C such that the isometry $F = T_aC$ carries the $\mathbf{e}$ frame to the $\mathbf{f}$ frame.

4. (a) Prove that an isometry $F = TC$ carries the plane through $\mathbf{p}$ orthogonal to $\mathbf{q} \neq 0$ to the plane through $F(\mathbf{p})$ orthogonal to $C(\mathbf{q})$.
(b) If P is the plane through (1/2, −1, 0) orthogonal to (0, 1, 0) find an isometry $F = TC$ such that $F(P)$ is the plane through (1, −2, 1) orthogonal to (1, 0, −1).

5. (*Computer*.)
(a) Verify that both sets of vectors in Exercise 3 form frames by showing that $A^tA = I$ for their attitude matrices.
(b) Find the matrix C that carries each $\mathbf{e}_i$ to $\mathbf{f}_i$, and check this for $i = 1, 2, 3$.

3.3 Orientation

We now come to one of the most interesting and elusive ideas in geometry. Intuitively, it is *orientation* that distinguishes between a right-handed glove and a left-handed glove in ordinary space. To handle this concept mathematically, we replace gloves by frames and separate all the frames on $\mathbf{R}^3$ into two classes as follows. Recall that associated with each frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ at a point of $\mathbf{R}^3$ is its attitude matrix A. According to the exercises for Section 1 of Chapter 2,

$$\mathbf{e}_1 \bullet \mathbf{e}_2 \times \mathbf{e}_3 = \det A = \pm 1.$$

When this number is +1, we shall say that the frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is *positively oriented* (or right-handed); when it is −1, the frame is *negatively oriented* (or left-handed).

We omit the easy proof of the following facts.

3.1 Remark (1) At each point of $\mathbf{R}^3$ the frame assigned by the natural frame field U_1, U_2, U_3 is positively oriented.

(2) A frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is positively oriented if and only if $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$. Thus the orientation of a frame can be determined, for practical purposes, by the "right-hand rule" given at the end of Section 1 of Chapter 2. Pictorially, the frame (P) in Fig. 3.3 is positively oriented, whereas the frame (N) is negatively oriented. In particular, *Frenet frames are always positively oriented*, since by definition, $B = T \times N$.

(3) For a positively oriented frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, the cross products are

$$\mathbf{e}_1 = \mathbf{e}_2 \times \mathbf{e}_3 = -\mathbf{e}_3 \times \mathbf{e}_2,$$

$$\mathbf{e}_2 = \mathbf{e}_3 \times \mathbf{e}_1 = -\mathbf{e}_1 \times \mathbf{e}_3,$$

$$\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2 = -\mathbf{e}_2 \times \mathbf{e}_1.$$

For a negatively oriented frame, reverse the vectors in each cross product. (One need not memorize these formulas—the right-hand rule will give them all correctly.)

Having attached a sign to each frame on $\mathbf{R}^3$, we next attach a sign to each isometry F of $\mathbf{R}^3$. In Chapter 2 we proved the well-known fact that the determinant of an orthogonal matrix is either +1 or −1. Thus if C is the orthogonal part of the isometry F, we define the *sign* of F to be the determinant of C, with notation

$$\operatorname{sgn} F = \det C.$$

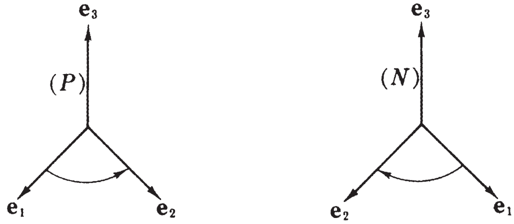


FIG. 3.3

We know that the tangent map of an isometry carries frames to frames. The following result tells what happens to their orientations.

3.2 Lemma If $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ is a frame at some point of $\mathbf{R}^3$ and F is an isometry, then

$$F_*(\mathbf{e}_1) \bullet F_*(\mathbf{e}_2) \times F_*(\mathbf{e}_3) = (\operatorname{sgn} F)\, \mathbf{e}_1 \bullet \mathbf{e}_2 \times \mathbf{e}_3.$$

Proof. If $\mathbf{e}_j = \sum a_{jk} U_k$, then by the coordinate form of Theorem 2.1 we have

$$F_*(\mathbf{e}_j) = \sum_{i,k} c_{ik} a_{jk} \overline{U}_i,$$

where $C = (c_{ij})$ is the orthogonal part of F. Thus the attitude matrix of the frame $F_*(\mathbf{e}_1)$, $F_*(\mathbf{e}_2)$, $F_*(\mathbf{e}_3)$ is the matrix

$$\left(\sum_k c_{ik} a_{jk}\right) = \left(\sum_k c_{ik}\, {}^t a_{kj}\right) = C\, {}^t A.$$

But the triple scalar product of a frame is the determinant of its attitude matrix, and by definition, $\operatorname{sgn} F = \det C$. Consequently,

$$\begin{aligned} F_*(\mathbf{e}_1) \bullet F_*(\mathbf{e}_2) \times F_*(\mathbf{e}_3) &= \det\,(C\, {}^t A) \\ &= \det C \cdot \det {}^t A = \det C \cdot \det A \\ &= (\operatorname{sgn} F)\mathbf{e}_1 \cdot \mathbf{e}_2 \times \mathbf{e}_3. \end{aligned}$$ ◆

This lemma shows that if $\operatorname{sgn} F = +1$, then F_* carries positively oriented frames to positively oriented frames and carries negatively oriented frames to negatively oriented frames. On the other hand, if $\operatorname{sgn} F = -1$, positive goes to negative and negative to positive.

3.3 Definition An isometry F of $\mathbf{R}^3$ is said to be

orientation-preserving if $\operatorname{sgn} F = \det C = +1$,

orientation-reversing if $\operatorname{sgn} F = \det C = -1$,

where C is the orthogonal part of F.

3.4 Example (1) ***Translations.*** All translations are orientation-preserving. Geometrically this is clear, and in fact the orthogonal part of a translation T is just the identity mapping I, so $\operatorname{sgn} T = \det I = +1$.

(2) ***Rotations.*** Consider the orthogonal transformation C given in Example 1.2, which rotates $\mathbf{R}^3$ through angle θ around the z axis. Its matrix is

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Hence $\operatorname{sgn} C = \det C = +1$, so C is orientation-preserving (see Exercise 4).

(3) ***Reflections.*** One can (literally) see reversal of orientation by using a mirror. Suppose the yz plane of $\mathbf{R}^3$ is the mirror. If one looks toward that plane, the point $\mathbf{p} = (p_1, p_2, p_3)$ appears to be located at the point

$$R(\mathbf{p}) = (-p_1, p_2, p_3)$$

(Fig. 3.4). The mapping R so defined is called *reflection* in the yz plane. Evidently it is an orthogonal transformation, with matrix

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus R is an orientation-reversing isometry, as confirmed by the experimental fact that the mirror image of a right hand is a left hand.

Both dot and cross product were originally defined in terms of *Euclidean* coordinates. We have seen that the dot product is given by the same formula,

$$\mathbf{v} \bullet \mathbf{w} = \left(\sum v_i \mathbf{e}_i\right) \bullet \left(\sum w_i \mathbf{e}_i\right) = \sum v_i w_i,$$

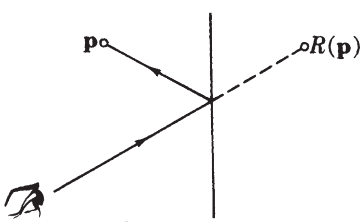


FIG. 3.4

no matter what frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is used to get coordinates for $\mathbf{v}$ and $\mathbf{w}$. Almost the same result holds for cross products, but orientation is now involved.

3.5 Lemma Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ be a frame at a point of $\mathbf{R}^3$. If $\mathbf{v} = \sum v_i \mathbf{e}_i$ and $\mathbf{w} = \sum w_i \mathbf{e}_i$, then

$$\mathbf{v} \times \mathbf{w} = \varepsilon \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix},$$

where $\varepsilon = \mathbf{e}_1 \bullet \mathbf{e}_2 \times \mathbf{e}_3 = \pm 1$.

Proof. It suffices merely to expand the cross product

$$\mathbf{v} \times \mathbf{w} = (v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3) \times (w_1\mathbf{e}_1 + w_2\mathbf{e}_2 + w_3\mathbf{e}_3)$$

using the formulas (3) of Remark 3.1. For example, if the frame is positively oriented, for the $\mathbf{e}_1$ component of $\mathbf{v} \times \mathbf{w}$ we get

$$v_2\mathbf{e}_2 \times w_3\mathbf{e}_3 + v_3\mathbf{e}_3 \times w_2\mathbf{e}_2 = (v_2w_3 - v_3w_2)\mathbf{e}_1.$$

Since $\varepsilon = 1$ in this case, we get the same result by expanding the determinant in the statement of this lemma. ◆

It follows immediately that the effect of an isometry on cross products also involves orientation.

3.6 Theorem Let $\mathbf{v}$ and $\mathbf{w}$ be tangent vectors to $\mathbf{R}^3$ at $\mathbf{p}$. If F is an isometry of $\mathbf{R}^3$, then

$$F_*(\mathbf{v} \times \mathbf{w}) = (\operatorname{sgn} F) F_*(\mathbf{v}) \times F_*(\mathbf{w}).$$

Proof. Write $\mathbf{v} = \sum v_i U_i(\mathbf{p})$ *and* $\mathbf{w} = \sum w_i U_i(\mathbf{p})$. Now let

$$\mathbf{e}_i = F_*(U_i(\mathbf{p})).$$

Since F_* is linear,

$$F_*(\mathbf{v}) = \sum v_i \mathbf{e}_i \quad \text{and} \quad F_*(\mathbf{w}) = \sum w_i \mathbf{e}_i.$$

A straightforward computation using Lemma 3.5 shows that

$$F_*(\mathbf{v}) \times F_*(\mathbf{w}) = \varepsilon F_*(\mathbf{v} \times \mathbf{w}),$$

where

$$\varepsilon = \mathbf{e}_1 \bullet \mathbf{e}_2 \times \mathbf{e}_3 = F_*(U_1(\mathbf{p})) \bullet F_*(U_2(\mathbf{p})) \times F_*(U_3(\mathbf{p})).$$

But U_1, U_2, U_3 is positively oriented, so by Lemma 3.2, $\varepsilon = \operatorname{sgn} F$. ◆

Exercises

1. Prove

$$\operatorname{sgn}(FG) = \operatorname{sgn} F \cdot \operatorname{sgn} G = \operatorname{sgn}(GF).$$

Deduce that $\operatorname{sgn} F = \operatorname{sgn}(F^{-1})$.

2. If H_0 is an orientation-reversing isometry of $\mathbf{R}^3$, show that *every* orientation-reversing isometry has a unique expression H_0F, where F is orientation-preserving.

3. Let $\mathbf{v} = (3, 1, -1)$ and $\mathbf{w} = (-3, -3, 1)$ be tangent vectors at some point. If C is the orthogonal transformation given in Exercise 4 of Section 1, check the formula

$$C_*(\mathbf{v} \times \mathbf{w}) = (\operatorname{sgn} C) C_*(\mathbf{v}) \times C_*(\mathbf{w}).$$

4. A *rotation* is an orthogonal transformation C such that $\det C = +1$. Prove that C does, in fact, rotate $\mathbf{R}^3$ around an axis. Explicitly, given a rotation C, show that there exists a number ϑ and points $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ with $\mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij}$ such that (Fig. 3.5)

$$C(\mathbf{e}_1) = \cos\vartheta\, \mathbf{e}_1 + \sin\vartheta\, \mathbf{e}_2,$$

$$C(\mathbf{e}_2) = -\sin\vartheta\, \mathbf{e}_1 + \cos\vartheta\, \mathbf{e}_2,$$

$$C(\mathbf{e}_3) = \mathbf{e}_3.$$

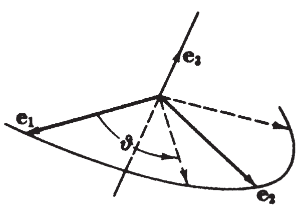


FIG. 3.5

(*Hint:* The fact that the dimension of $\mathbf{R}^3$ is odd means that C has an eigenvalue +1, so there is a point $\mathbf{p} \neq \mathbf{0}$ such that $C(\mathbf{p}) = \mathbf{p}$.)

5. Let $\mathbf{a}$ be a point of $\mathbf{R}^3$ such that $\| \mathbf{a} \| = 1$. Prove that the formula

$$C(\mathbf{p}) = \mathbf{a} \times \mathbf{p} + (\mathbf{p} \bullet \mathbf{a})\, \mathbf{a}$$

defines an orthogonal transformation. Describe its general effect on $\mathbf{R}^3$.

6. Prove
(a) The set $O^+(3)$ of all rotations of $\mathbf{R}^3$ is a subgroup of the orthogonal group $O(3)$ (see Ex. 8 of Sec. 3.1).
(b) The set $\mathscr{E}^+(3)$ of all orientation-preserving isometries of $\mathbf{R}^3$ is a subgroup of the Euclidean group $\mathscr{E}(3)$.

3.4 Euclidean Geometry

In the discussion at the beginning of this chapter, we recalled a fundamental feature of plane geometry: If there is an isometry carrying one triangle onto another, then the two (congruent) triangles have exactly the same geometric properties. A close examination of this statement will show that it does not admit a proof—it is, in fact, just the definition of "geometric property of a triangle." More generally, *Euclidean geometry* can be defined as the totality of concepts that are preserved by isometries of Euclidean space. For example, Corollary 2.2 shows that the notion of dot product on tangent vectors belongs to Euclidean geometry. Similarly, Theorem 3.6 shows that the cross product is preserved by isometries (except possibly for sign).

This famous definition of Euclidean geometry is somewhat generous, however. In practice, the label "Euclidean geometry" is usually attached only to those concepts that are preserved by isometries, but *not* by arbitrary mappings, or even the more restrictive class of mappings (diffeomorphisms) that possess inverse mappings. An example should make this distinction clearer. If $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ is a curve in $\mathbf{R}^3$, then the various derivatives

$$\alpha' = \left(\frac{d\alpha_1}{dt}, \frac{d\alpha_2}{dt}, \frac{d\alpha_3}{dt}\right), \qquad \alpha'' = \left(\frac{d^2\alpha_1}{dt^2}, \frac{d^2\alpha_2}{dt^2}, \frac{d^2\alpha_3}{dt^2}\right), \ \ldots$$

look pretty much alike. Now, Theorem 7.8 of Chapter 1 asserts that *velocity is preserved by arbitrary mappings* F: $\mathbf{R}^3 \to \mathbf{R}^3$, that is, if $\beta = F(\alpha)$, then $\beta' = F_*(\alpha')$. But it is easy to see that *acceleration is not preserved by arbitrary mappings*. For example, if $\alpha(t) = (t, 0, 0)$ and $F = (x^2, y, z)$, then $\alpha'' = 0$; hence $F_*(\alpha'') = 0$. But $\beta = F(\alpha)$ has the formula $\beta(t) = (t^2, 0, 0)$, so $\beta'' = 2U_1$. Thus in this case, $\beta = F(\alpha)$, but $\beta'' \neq F*(\alpha'')$. We shall see in a moment, however, that acceleration is preserved by *isometries*.

For this reason, the notion of velocity belongs to the *calculus* of Euclidean space, while the notion of acceleration belongs to Euclidean *geometry*. In this section we examine some of the concepts introduced in Chapter 2 and prove that they are, in fact, preserved by isometries. (We leave largely to the reader the easier task of showing that they are not preserved by diffeomorphisms.)

Recall the notion of vector field on a curve (Definition 2.2 of Chapter 2). If Y is a vector field on α: $I \to \mathbf{R}^3$ and F: $\mathbf{R}^3 \to \mathbf{R}^3$ is any mapping, then $\bar{Y} = F*(Y)$ is a vector field on the image curve $\bar{\alpha} = F(\alpha)$. In fact, for each t in I, $Y(t)$ is a tangent vector to $\mathbf{R}^3$ at the point $\alpha(t)$. But then $\bar{Y}(t) = F*(Y(t))$ is a tangent vector to $\mathbf{R}^3$ at the point $F(\alpha(t)) = \bar{\alpha}(t)$.

(These relationships are illustrated in Fig. 3.6.) Isometries preserve the *derivatives* of such vector fields.

4.1 Corollary Let Y be a vector field on a curve α in $\mathbf{R}^3$, and let F be an isometry of $\mathbf{R}^3$. Then $\bar{Y} = F*(Y)$ is a vector field on $\bar{\alpha} = F(\alpha)$, and

$$\bar{Y}' = F*(Y').$$

Proof. To differentiate a vector field $Y = \sum y_j U_j$, one simply differentiates its Euclidean coordinate functions, so

$$Y' = \sum \frac{dy_j}{dt} U_j.$$

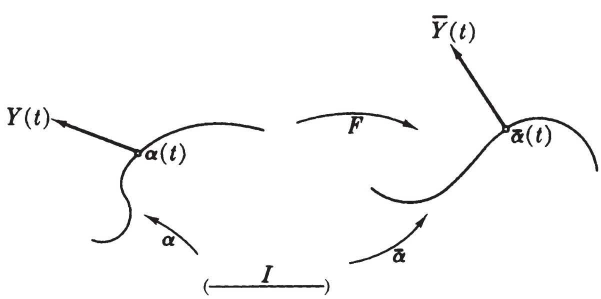


FIG. 3.6

Thus by the coordinate version of Theorem 2.1, we get

$$F_*(Y') = \sum c_{ij} \frac{dy_j}{dt} \overline{U}_i.$$

On the other hand,

$$\overline{Y} = F_*(Y) = \sum c_{ij} y_j \overline{U}_i.$$

But each c_{ij} is constant, being by definition an entry in the matrix of the orthogonal part of the isometry F. Hence

$$\overline{Y}' = \sum \frac{d}{dt}(c_{ij} y_j)\overline{U}_i = \sum c_{ij} \frac{dy_j}{dt} \overline{U}_i.$$

Thus the vector fields $F_*(Y')$ and $\overline{Y}'$ are the same. ◆

We claimed earlier that isometries preserve acceleration: If $\overline{\alpha} = F(\alpha)$, where F is an isometry, then $\overline{\alpha}'' = F_*(\alpha'')$. This is an immediate consequence of the preceding result, for if we set $Y = \alpha'$, then by Theorem 7.8 of Chapter 1, $\overline{Y} = \overline{\alpha}'$; hence

$$\overline{\alpha}'' = \overline{Y}' = F_*(Y') = F_*(\alpha'').$$

Now we show that the Frenet apparatus of a curve is preserved by isometries. This is certainly to be expected on intuitive grounds, since a rigid motion ought to carry one curve into another that turns and twists in exactly the same way. And this is what happens *when the isometry is orientation-preserving.*

4.2 Theorem Let β be a unit-speed curve in $\mathbf{R}^3$ with positive curvature, and let $\overline{\beta} = F(\beta)$ be the image curve of β under an isometry F of $\mathbf{R}^3$. Then

$$\overline{\kappa} = \kappa, \qquad \overline{T} = F_*(T),$$
$$\overline{\tau} = (\operatorname{sgn} F)\tau, \qquad \overline{N} = F_*(N),$$
$$\overline{B} = (\operatorname{sgn} F)F_*(B),$$

where $\operatorname{sgn} F = \pm 1$ is the sign of the isometry F.

Proof. Note that $\overline{\beta}$ is also a unit-speed curve, since

$$\|\overline{\beta}'\| = \|F_*(\beta')\| = \|\beta'\| = 1.$$

Thus the definitions in Section 3 of Chapter 2 apply to both β and $\bar{\beta}$, so

$$\bar{T} = \bar{\beta}' = F_*(\beta') = F_*(T).$$

Since F_* preserves both acceleration and norms, it follows from the definition of curvature that

$$\bar{\kappa} = \|\bar{\beta}''\| = \|F_*(\beta'')\| = \|\ \beta''\ \| = \kappa.$$

To get the full Frenet frame, we now use the hypothesis $\kappa > 0$ (which implies $\bar{\kappa} > 0$, since $\bar{\kappa} = \kappa$). By definition, $N = \beta''/\kappa$; hence using preceding facts, we find

$$\bar{N} = \frac{\bar{\beta}''}{\bar{\kappa}} = \frac{F_*(\beta'')}{\kappa} = F_*\left(\frac{\beta''}{\kappa}\right) = F_*(N).$$

It remains only to prove the interesting cases B and τ. Since the definition $B = T \times N$ involves a cross product, we use Theorem 3.6 to get

$$\bar{B} = \bar{T} \times \bar{N} = F_*(T) \times F_*(N) = (\operatorname{sgn} F)F_*(T \times N) = (\operatorname{sgn} F)F_*(B).$$

The definition of torsion is essentially $\tau = -B' \bullet N = B \bullet N'$. Thus, using the results above for B and N, we get

$$\bar{\tau} = \bar{B} \bullet \bar{N}' = (\operatorname{sgn} F)F_*(B) \bullet F_*(N') = (\operatorname{sgn} F)B \bullet N' = (\operatorname{sgn} F)\tau. \qquad \blacklozenge$$

The presence of sgn F in the formula for the torsion of $F(\beta)$ shows that the torsion of a curve gives a more subtle description of the curve than has been apparent so far. *The sign of τ measures the orientation of the twisting of the curve.* If F is orientation-reversing, the formula $\bar{\tau} = -\tau$ proves that the twisting of the image of curve $F(\beta)$ is exactly opposite to that of β itself.

A simple example will illustrate this reversal.

4.3 Example Let β be the unit-speed helix

$$\beta(s) = \left(\cos\frac{s}{c}, \sin\frac{s}{c}, \frac{s}{c}\right),$$

gotten from Example 3.3 of Chapter 2 by setting $a = b = 1$; hence $c = \sqrt{2}$. We know from the general formulas for helices that $\kappa = \tau = 1/2$. Now let R be reflection in the xy plane, so R is the isometry $R(x, y, z) = (x, y, -z)$. Thus the image curve $\bar{\beta} = \mathbf{R}(\beta)$ is the mirror image

$$\bar{\beta}(s) = \left(\cos\frac{s}{c}, \sin\frac{s}{c}, -\frac{s}{c}\right)$$

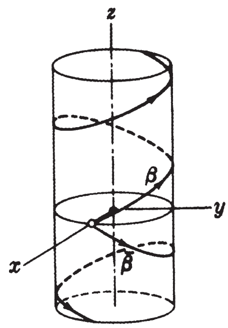


FIG. 3.7

of the original curve. One can see in Fig. 3.7 that the mirror has its usual effect: β and $\bar{\beta}$ twist in opposite ways—if β is "right-handed," then $\bar{\beta}$ is "left-handed." (The fact that β is going up and $\bar{\beta}$ down is, in itself, irrelevant.) Formally: The reflection R is orientation-reversing; hence the theorem predicts $\bar{\kappa} = \kappa = \frac{1}{2}$ and $\bar{\tau} = -\tau = -\frac{1}{2}$. Since $\bar{\beta}$ is just the helix gotten in Example 3.3 of Chapter 2 by taking $a = 1$ and $b = -1$, this may be checked by the general formulas there.

Exercises

1. Let $F = TC$ be an isometry of $\mathbf{R}^3$, β a unit speed curve in $\mathbf{R}^3$. Prove
(a) If β is a cylindrical helix, then $F(\beta)$ is a cylindrical helix.
(b) If β has spherical image σ, then $F(\beta)$ has spherical image $C(\sigma)$.

2. Let $Y = (t, 1 - t^2, 1 + t^2)$ be a vector field on the helix

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

and let C be the orthogonal transformation

$$C = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

Compute $\bar{\alpha} = C(\alpha)$ and $\bar{Y} = C_*(Y)$, and check that

$$C_*(Y') = \bar{Y}', \quad C_*(\alpha'') = \bar{\alpha}'', \quad Y' \bullet \alpha'' = \bar{Y}' \bullet \bar{\alpha}''.$$

3. Sketch the triangles in $\mathbf{R}^2$ that have vertices

$$\Delta_1: \ (3,1), \ (7,1), \ (7,4), \quad \Delta_2: \ (2,0), \ (2,5), \ (-2/5, 16/5).$$

Show that these triangles are congruent by exhibiting an isometry $F = TC$ that carries Δ_1 to Δ_2. (*Hint*: the orthogonal part C is not altered if the triangles are translated.)

4. If $F: \mathbf{R}^3 \to \mathbf{R}^3$ is a diffeomorphism such that F_* preserves dot products, show that F is an isometry. (*Hint:* Show that F preserves lengths of curve segments and deduce that F^{-1} does also.)

5. Let F be an isometry of $\mathbf{R}^3$. For each vector field V let $\overline{V}$ be the vector field such that $F_*(V(\mathbf{p})) = \overline{V}(F(\mathbf{p}))$ for all $\mathbf{p}$. Prove that isometries preserve covariant derivatives; that is, show $\overline{\nabla_V W} = \nabla_{\overline{V}} \overline{W}$.

3.5 Congruence of Curves

In the case of curves in $\mathbf{R}^3$, the general notion of congruence takes the following form.

5.1 Definition Two curves $\alpha, \beta: I \to \mathbf{E}^3$ are *congruent* provided there exists an isometry F of $\mathbf{R}^3$ such that $\beta = F(\alpha)$; that is, $\beta(t) = F(\alpha(t))$ for all t in I.

Intuitively speaking, congruent curves are the same except for position in space. They represent *trips at the same speed along routes of the same shape.* For example, the helix $\alpha(t) = (\cos t, \sin t, t)$ spirals around the z axis in exactly the same way the helix $\beta(t) = (t, \cos t, \sin t)$ spirals around the x axis. Evidently these two curves are congruent, since if F is the isometry such that

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

then $F(\alpha) = \beta$.

To decide whether given curves α and β are congruent, it is hardly practical to try all the isometries of $\mathbf{R}^3$ to see whether there is one that carries α to β. What we want is a description of the shape of a unit-speed curve so accurate that if α and β have the same description, then they must be congruent. The proper description, as the reader will doubtless suspect, is given by curvature and torsion. To prove this we need one preliminary result.

Curves whose congruence is established by a translation are said to be *parallel.* Thus, curves $\alpha, \beta: I \to \mathbf{E}^3$ are parallel if and only if there is a point

p in $\mathbf{R}^3$ such that $\beta(s) = \alpha(s) + \mathbf{p}$ for all s in I, or, in functional notation, $\beta = \alpha + \mathbf{p}$.

5.2 Lemma Two curves $\alpha, \beta\colon I \to \mathbf{R}^3$ are parallel if their velocity vectors $\alpha'(s)$ and $\beta'(s)$ are parallel for each s in I. In this case, if $\alpha(s_0) = \beta(s_0)$ for some one s_0 in I, then $\alpha = \beta$.

Proof. By definition, if $\alpha'(s)$ and $\beta'(s)$ are parallel, they have the same Euclidean coordinates. Thus

$$\frac{d\alpha_i}{ds}(s) = \frac{d\beta_i}{ds}(s) \quad \text{for } 1 \leqq i \leqq 3,$$

where α_i and β_i are the Euclidean coordinate functions of α and β. But by elementary calculus, the equation $d\alpha_i/ds = d\beta_i/ds$ implies that there is a constant p_i such that $\beta_i = \alpha_i + p_i$. Hence $\beta = \alpha + \mathbf{p}$. Furthermore, if $\alpha(s_0) = \beta(s_0)$, we deduce that $\mathbf{p} = \mathbf{0}$; hence $\alpha = \beta$. ◆

5.3 Theorem If $\alpha, \beta\colon I \to \mathbf{R}^3$ are unit-speed curves such that $\kappa_\alpha = \kappa_\beta$ and $\tau_\alpha = \pm\tau_\beta$, then α and β are congruent.

Proof. There are two main steps:

(1) Replace α by a suitably chosen congruent curve $F(\alpha)$.

(2) Show that $F(\alpha) = \beta$ (Fig. 3.8).

Our guide for the choice in (1) is Theorem 4.2. Fix a number, say 0, in the interval I. If $\tau_\alpha = \tau_\beta$, then let F be the (orientation-preserving) isometry that carries the Frenet frame $T_\alpha(0)$, $N_\alpha(0)$, $B_\alpha(0)$ of α at $\alpha(0)$ to the

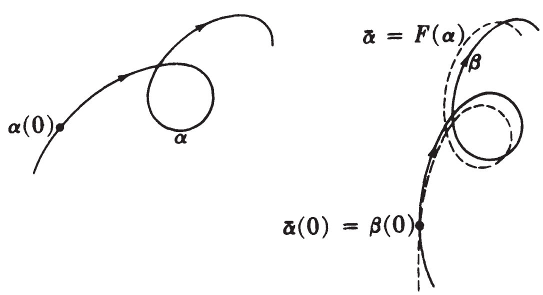


FIG. 3.8

Frenet frame $T_\beta(0)$, $N_\beta(0)$, $B_\beta(0)$, of β at $\beta(0)$. (The existence of this isometry is guaranteed by Theorem 2.3.) Denote the Frenet apparatus of $\bar{\alpha} = F(\alpha)$ by $\bar{\kappa}$, $\bar{\tau}$, $\bar{T}$, $\bar{N}$, $\bar{B}$; then it follows immediately from Theorem 4.2 and the information above that

$$\begin{aligned} \bar{\alpha}(0) &= \beta(0), & \bar{T}(0) &= T_\beta(0), \\ \bar{\kappa} &= \kappa_\beta, & \bar{N}(0) &= N_\beta(0), \\ \bar{\tau} &= \tau_\beta, & \bar{B}(0) &= B_\beta(0). \end{aligned} \tag{‡}$$

On the other hand, if $\tau_\alpha = -\tau_\beta$, we choose F to be the (orientation-reversing) isometry that carries $T_\alpha(0)$, $N_\alpha(0)$, $B_\alpha(0)$ at $\alpha(0)$ to the frame $T_\beta(0)$, $N_\beta(0)$, $B_\beta(0)$ at $\beta(0)$. (Frenet frames are positively oriented; hence this last frame is negatively oriented: This is why F is orientation-reversing.) Then it follows from Theorem 4.2 that the equations (‡) hold also for $\bar{\alpha} = F(\alpha)$ and β. For example,

$$\bar{B}(0) = -F_*(B_\alpha(0)) = B_\beta(0).$$

For step (2) of the proof, we shall show $\bar{T} = T_\beta$; that is, the unit tangents of $\bar{\alpha} = F(\alpha)$ and β are parallel at each point. Since $\bar{\alpha}(0) = \beta(0)$, it will follow from Lemma 5.2 that $F(\alpha) = \beta$. On the interval I, consider the real-valued function $f = \bar{T} \bullet T_\beta + \bar{N} \bullet N_\beta + \bar{B} \bullet B_\beta$. Since these are *unit* vector fields, the Schwarz inequality (Sec. 1, Ch. 2) shows that

$$\bar{T} \bullet T_\beta \leqq 1;$$

furthermore, $\bar{T} \bullet T_\beta = 1$ if and only if $\bar{T} = T_\beta$. Similar remarks hold for the other two terms in f. Thus it *suffices to show that f has constant value* 3. By (‡), $f(0) = 3$. Now consider

$$f' = \bar{T}' \bullet T_\beta + \bar{T} \bullet T_\beta' + \bar{N}' \bullet N_\beta + \bar{N} \bullet N_\beta' + \bar{B}' \bullet B_\beta + \bar{B} \bullet B_\beta'$$

A simple computation completes the proof. Substitute the Frenet formulas in this expression and use the equations $\bar{\kappa} = \kappa_\beta$, $\bar{\tau} = \tau_\beta$ from (‡). The resulting eight terms cancel in pairs, so $f' = 0$, and f has, indeed, constant value 3. ◆

Thus, *a unit-speed curve is determined but for position in* $\mathbf{R}^3$ *by its curvature and torsion.*

Actually the proof of Theorem 5.3 does more than establish that α and β are congruent; it shows how to compute *explicitly* an isometry carrying α to β. We illustrate this in a special case.

5.4 Example Consider the unit-speed curves α, β: $\mathbf{R} \to \mathbf{R}^3$ such that

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

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

where $c = \sqrt{2}$. Obviously, these curves are congruent by means of a reflection—they are the helices considered in Example 4.3—but we shall ignore this in order to describe a general method for computing the required isometry. According to Example 3.3 of Chapter 2, α and β have the same curvature, $\kappa_\alpha = 1/2 = \kappa_\beta$; but torsions of opposite sign, $\tau_\alpha = 1/2 = -\tau_\beta$. Thus the theorem predicts congruence by means of an orientation-reversing isometry F. From its proof we see that F must carry the Frenet frame

$$T_\alpha(0) = (0, a, a),$$

$$N_\alpha(0) = (-1, 0, 0),$$

$$B_\alpha(0) = (0, -a, a),$$

where $a = 1/\sqrt{2}$, to the frame

$$T_\beta(0) = (0, a, -a),$$

$$N_\beta(0) = (-1, 0, 0),$$

$$-B_\beta(0) = (0, -a, -a),$$

where the minus sign will produce orientation reversal. (These explicit formulas also come from Example 3.3 of Chapter 2.) By the remark following Theorem 2.3, the isometry F has orthogonal part $C = {}^tBA$, where A and B are the attitude matrices of the two frames above. Thus

$$C = \begin{pmatrix} 0 & -1 & 0 \\ a & 0 & -a \\ -a & 0 & -a \end{pmatrix} \begin{pmatrix} 0 & a & a \\ -1 & 0 & 0 \\ 0 & -a & a \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

since $a = 1/\sqrt{2}$. These two frames have the same point of application $\alpha(0) = \beta(0) = (1, 0, 0)$. But C does not move this point, so the translation part of F is just the identity map. Thus we have (correctly) found that the reflection $F = C$ carries α to β.

From the viewpoint of Euclidean geometry, two curves in $\mathbf{R}^3$ are "the same" if they differ only by an isometry of $\mathbf{R}^3$. What, for example, is a helix?

It is not just a curve that spirals around the z axis as in Example 3.3 of Chapter 2, but any curve congruent to one of these special helices. One can give general formulas, but the best characterization follows.

5.5 Corollary Let α be a unit speed curve in $\mathbf{R}^3$. Then α is a helix if and only if both its curvature and torsion are nonzero constants.

Proof. For any numbers $a > 0$ and $b \neq 0$, let $\beta_{a,b}$ be the special helix given in Example 3.3 of Chapter 2. If α is congruent to $\beta_{a,b}$, then (changing the sign of b if necessary) we can assume the isometry is orientation-preserving. Thus, α has curvature and torsion

$$\kappa = \frac{a}{a^2 + b^2}, \qquad \tau = \frac{b}{a^2 + b^2}.$$

Conversely, suppose α has constant nonzero κ and τ. Solving the preceding equations, we get

$$a = \frac{\kappa}{\kappa^2 + \tau^2}, \qquad b = \frac{\tau}{\kappa^2 + \tau^2}.$$

Thus α and $\beta_{a,b}$ have the same curvature and torsion; hence they are congruent. ◆

Our results so far demand unit speed, but it is easy to weaken this restriction.

5.6 Corollary Let $\alpha, \beta: I \to \mathbf{R}^3$ arbitrary-speed curves. If

$$v_\alpha = v_\beta > 0, \quad \kappa_\alpha = \kappa_\beta > 0, \quad \text{and} \quad \tau_\alpha = \pm\tau_\beta,$$

then the curves α and β are congruent.

The proof is immediate, for the data ensures that the unit speed parametrizations of α and β have the same curvature and torsion—hence they are congruent. But then the original curves are congruent under the same isometry since their speeds are the same.

The theory of curves we have presented applies only to regular curves with positive curvature $\kappa > 0$, because only for such curves is it possible to define the Frenet frame field. However, an arbitrary curve α in $\mathbf{R}^3$ can be studied by means of an *arbitrary* frame field on α, that is, three unit-vector fields E_1, E_2, E_3 on α that are orthogonal at each point.

At a critical point later on, we will need this generalization of the congruence theorem (5.3):

5.7 Theorem Let α, β: $I \to \mathbf{R}^3$ be curves defined on the same interval. Let E_1, E_2, E_3 be a frame field on α, and F_1, F_2, F_3 a frame field on β. If

(1) $\alpha' \bullet E_i = \beta' \bullet F_i \quad (1 \leqq i \leqq 3)$,

(2) $E_i' \bullet E_j = F_i' \bullet F_j \quad (1 \leqq i, j \leqq 3)$,

then α and β are congruent.

Explicitly, for any t_0 in I, if $\mathbf{F}$ is the unique Euclidean isometry that sends each $E_i(t_0)$ to $F_i(t_0)$, then $\mathbf{F}(\alpha) = \beta$.

Proof. Let $\mathbf{F}$ be the specified isometry. Since $\mathbf{F}_*$ preserves dot products, it follows that the vector fields $\bar{E}_i = F_*(E_i)$ for $1 \leq i \leq 3$ form a frame field on $\bar{\alpha} = \mathbf{F}(\alpha)$. And since $\mathbf{F}_*$ preserves velocities of curves and derivatives of vector fields, by using condition (1) in the theorem, we find

$$\bar{\alpha}(t_0) = \beta(t_0) \quad \text{and} \quad \bar{\alpha}' \bullet \bar{E}_i = \beta' \bullet F_i \quad \text{for } 1 \leqq i \leqq 3. \tag{$*$}$$

Similarly, from condition (2), we get

$$\bar{E}_i(t_0) = F_i(t_0) \quad \text{and} \quad \bar{E}_i' \bullet \bar{E}_j = F_i' \cdot F_j \quad \text{for } 1 \leqq i, j \leqq 3. \tag{$**$}$$

In view of this last equation, orthonormal expansion yields

$$\bar{E}_i' = \sum_j a_{ij} \bar{E}_j \quad \text{and} \quad F_i' = \sum_j a_{ij} F_j,$$

with the *same* coefficient functions a_{ij}. Note that $a_{ij} + a_{ji} = 0$; hence $a_{ii} = 0$. (*Proof:* Differentiate $\bar{E}_i \bullet \bar{E}_j = \delta_{ij}$.)

Now let $f = \sum \bar{E}_j \bullet F_i$. We prove $f = 3$ as before: $f(t_0) = 3$, and

$$f' = \sum (\bar{E}_i' \bullet F_i + \bar{E}_i \bullet F_i') = \sum_{i,j} (a_{ij} + a_{ji}) \bar{E}_j \bullet F_i = 0.$$

Thus each $\bar{E}_i \bullet F_i = 1$, that is, $\bar{E}_i$ and F_i are parallel at each point. By ($*$) the same is true for

$$\bar{\alpha}' = \sum (\bar{\alpha}' \bullet \bar{E}_i) \bar{E}_i \quad \text{and} \quad \beta' = \sum (\beta' \bullet F_i) F_i.$$

Since $\alpha(t_0) = \beta(t_0)$, Lemma 5.2 gives the required result, $F(\alpha) = \bar{\alpha} = \beta$. ◆

5.8 Remark Existence theorem for curves in $\mathbf{R}^3$. Curvature and torsion tell whether two unit-speed curves are isometric, but they do more than that:

Given any two continuous functions $\kappa > 0$ and τ on an interval I, there exists a unit-speed curve α: $I \to \mathbf{R}^3$ that has these functions as its curvature and torsion. (As we know, any two such curves are congruent.) Thus the natural description of curves in $\mathbf{R}^3$ is devoid of geometry, consisting of a pair of real-valued functions.

The proof of the existence theorem requires advanced methods, so we have preferred to illustrate it by the corresponding result for plane curves (Exercises 7–10). Though simpler, this 2-dimensional version has the advantage that plane curvature $\tilde{\kappa}$ is not required to be positive.

Exercises

1. Given a curve $\alpha = (\alpha_1, \alpha_2, \alpha_3)$: $I \to \mathbf{R}^3$, prove that β: $I \to \mathbf{R}^3$ is congruent to α if and only if β can be written as

$$\beta(t) = \mathbf{p} + \alpha_1(t)\mathbf{e}_1 + \alpha_2(t)\mathbf{e}_2 + \alpha_3(t)\mathbf{e}_3,$$

where $\mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij}$.

2. Let E_1, E_2, E_3, be a frame field on $\mathbf{R}^3$ with dual forms θ_i and connection forms ω_{ij}. Prove that two curves α, β: $I \to \mathbf{R}^3$ are congruent if $\theta_i(\alpha') = \theta_i(\beta')$ and $\omega_{ij}(\alpha') = \omega_{ij}(\beta')$ for $1 \leqq i, j \leqq 3$ (*Hint:* Use Thm. 5.7.)

3. Show that the curve

$$\beta(t) = \left(t + \sqrt{3}\,\sin t,\ 2\cos t,\ \sqrt{3}t - \sin t\right)$$

is a helix by finding its curvature and torsion. Find a helix of the form $\alpha(t) = (a\cos t, a\sin t, bt)$ and an isometry F such that $F(\alpha) = \beta$.

4. (*Computer; see Appendix.*) (a) Show that the curves

$$\alpha(t) = \left(t + t^2, t - t^2, 1 + \sqrt{2}t^3\right), \quad \beta(t) = \left(t^2 + t^3, 1 - \sqrt{2}t, t^2 - t^3\right),$$

defined on the entire real line, have the same speed, curvature, and torsion. (b) Find formulas for T and C such that the isometry $F = TC$ carries α to β and verify explicitly that $F(\alpha) = \beta$. (*Hint:* Use Ex. 5 of Sec. 2.)

5. (*Computer optional.*) Is the following curve a helix? Prove your answer.

$$c(t) = (-2\cos t + 2\sin t + 2t, 2\cos t + \sin t + 4t, \cos t + 2\sin t - 4t).$$

6. Congruence of curves.

(a) Prove that curves α, β: $I \to \mathbf{R}^2$ are congruent if $\tilde{\kappa}_\alpha = \tilde{\kappa}_\beta$ and they have the same speed.

(b) Show that the space curves

$$\alpha(t) = \left(\sqrt{2}t, t^2, 0\right) \quad \text{and} \quad \beta(t) = (-t, t, t^2)$$

are congruent. Find an isometry that carries α to β.

7. Given a continuous function f on an interval I, prove—using ordinary integration of functions—that there exists a unit-speed curve $\beta(s)$ in $\mathbf{R}^2$ for which $f(s)$ is the plane curvature. (*Hint:* Reverse the logic in Ex. 8 of Sec. 2.3.)

8. Show that $\beta(s) = (x(s), y(s))$ in the preceding exercise is given by the solutions of the differential equations

$$x'(s) = \cos\varphi(s), \quad y'(s) = \sin\varphi(s), \quad \varphi'(s) = f(s),$$

with initial conditions $x(0) = y(0) = \varphi(0) = 0$. (These initial conditions suffice, since any other β differs at most by a Euclidean isometry and a reparametrization $s \to s + c$.)

Explicit integration is rarely possible; the following exercises use numerical integration.

9. (*Numerical integration, computer graphics.*) Write computer commands that (a) given $f(s)$, produce a numerical description of the solution curve $\beta(s)$ in the preceding exercise, and (b) given $f(s)$, plot the solution curve.

10. (*Continuation.*) Plot unit-speed plane curves with the given plane curvature function f on at least the given interval.

(a) $f(s) = 1 + e^s$, on $-6 \le s \le 3$.
(b) $f(s) = 2 + 3\cos 3s$, on $0 \le s \le 2\pi$.
(c) $f(s) = 3 - 2s^2 + s^3$, on $-2.5 \le s \le 3.5$.

Adjust scales on axes as needed.

3.6 Summary

The basic result of this chapter is that an arbitrary isometry of Euclidean space can be uniquely expressed as an orthogonal transformation followed by a translation. A consequence is that the tangent map of an isometry F is, at every point, essentially just the orthogonal part of F. Then it is a routine matter to test the concepts introduced earlier to see which belong to *Euclidean geometry*, that is, which are preserved by isometries of Euclidean space.

Finally, we proved an analogue for curves of the various criteria for congruence of triangles in plane geometry; namely, we showed that a necessary and sufficient condition for two curves in $\mathbf{R}^3$ to be congruent is that they have the same curvature and torsion (and speed). Furthermore, the sufficiency proof shows how to find the required isometry explicitly.