Chapter 1

Vectors

1.1 Vectors that You Know

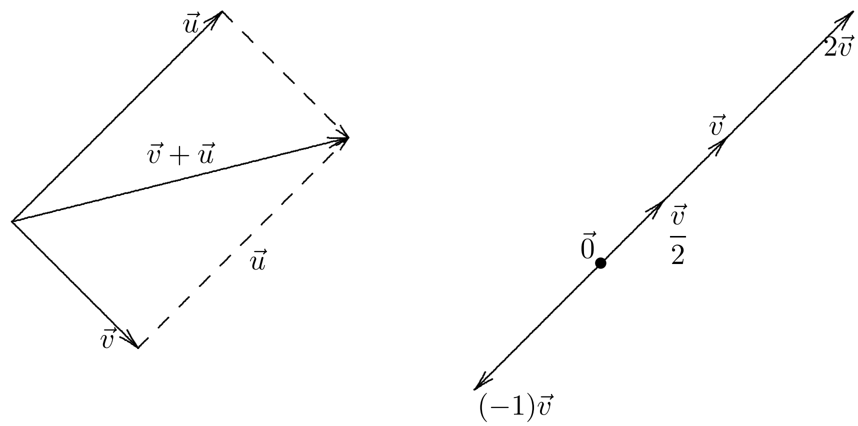
Vectors (and things made out of vectors or related to them) are the main subject matter of this book. Instead of starting with a precise mathematical definition of a vector, we give an informal, intuitive definition and discuss some examples.

> *Vectors are things that can be* **added to each other** *and* **multiplied by numbers***.*

This definition assumes that we all have intuitive notions of addition and multiplication, which can be carried over from ordinary numbers to other kinds of objects. This will be seen to be true for each type of example we discuss. A later chapter contains the unambiguous, formal definitions of "addition", "multiplication", and "vector" that historically arose out of such examples.

Example 1. Probably the most familiar vectors are those introduced in physics courses and characterized as *physical quantities with both magnitude and direction.* These include forces, velocities, electric fields, temperature gradients, magnetic moments. They are customarily drawn as arrows.

Within this geometrical conception of vectors as arrows, there is a well known construction for *adding* two vectors, shown in the drawing below. Initially the two arrows are thought of as rooted at the same point (the *origin*, or *zero vector*). One way of describing the sum is to slide one of the vectors (say $\vec{u}$) along the other one ($\vec{v}$), without rotating it or changing its length, so that the tail of $\vec{u}$ is at the head of $\vec{v}$. Then the sum $\vec{v}+\vec{u}$ is the arrow with its tail at the tail of $\vec{v}$ and its head at the head of $\vec{u}$, so that the three vectors form a triangle. Entirely equivalent to this "triangle rule" is the "parallelogram rule": Draw the parallelogram with $\vec{u}$ (in its original position) and $\vec{v}$ as adjacent sides. Then $\vec{v}+\vec{u}$ is the arrow pointing along the diagonal of the parallelogram starting from the origin.



The other half of the drawing shows how one *multiplies* a vector by a number geometrically: Simply put, the vector $\vec{v}$ is stretched by the numerical factor r. The word "stretching" is appropriate if r is a positive number greater than 1. If $0 \leq r < 1$, then "shrinking" is a better description. If r is negative, the arrow is reflected (so that it points in the opposite direction from the origin) in addition to a stretching or shrinking.

In this discussion we have tacitly assumed that r is a *real* number. Types of vectors that can be multiplied by *complex* numbers also arise. In this book we will deal mostly with real vector spaces. When complex vector spaces come up, we will call special attention to them; otherwise, all numbers are assumed to be real.

Example 2. Vectors are encountered in elementary courses also in the form of *n-tuples of real numbers*. A 2-tuple is a *pair*, such as $(2,3)$, and the vector space of all such pairs is called $\mathbf{R}^2$. Similarly, $\mathbf{R}^3$ consists of all strings of three numbers, and so on. Two pairs are added "componentwise", or "slot by slot":

$$(2,3)+\left(\tfrac{1}{2},-1\right)=\left(\tfrac{5}{2},2\right).$$

More formally, we can write down the definition of the sum of two vectors in any $\mathbf{R}^n$:

$$(a_1,a_2,\ldots,a_n)+(b_1,b_2,\ldots,b_n)\equiv(a_1+b_1,a_2+b_2,\ldots,a_n+b_n).$$

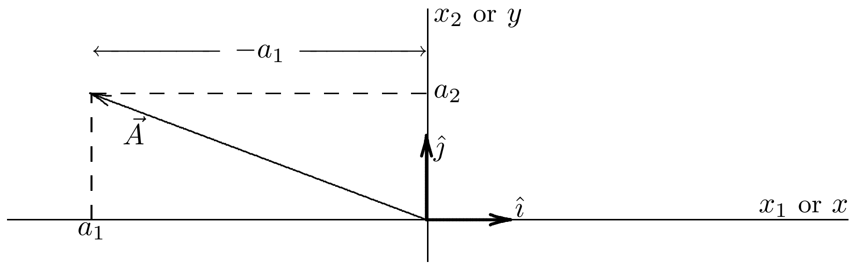
Similarly, multiplication is defined componentwise:

$$3(1,3,1)=(3,9,3);$$

$$r(a_1,a_2,\ldots,a_n)\equiv(ra_1,ra_2,\ldots,ra_n).$$

(The symbol "$\equiv$" means "equal by definition".)

Example 1 can be related to Example 2 by introducing a coordinate system into the geometrical space of Example 1:



Each physical vector is identified with a string of three numbers, (x_1, x_2, x_3) or (x, y, z). For example, the y component a_2 of the vector $\vec{A}$ is its projection onto the y axis, measured as a multiple of the *unit vector* or *basis vector* $\hat{\jmath}$ along that axis.

Indeed, we understand this sitation so well that we often think of Examples 1 and 2 as being the same thing! (Technically, one says that these two spaces of vectors are *isomorphic.*) Note, however, that the correspondence between them depends on the coordinate system. Introduce a rotated set of axes, and the same physical vector will correspond to a different string of numbers.

Important Notational Remark: There are several notations for vectors and their coordinates, and for basis vectors, which have grown up in connection with various applications. It is important to be able to deal with all of them and to tolerate occasional inconsistencies in notation. For example, a two-dimensional vector may be written in the various ways

$$(3,1) = 3\hat{\imath} + \hat{\jmath} = 3(1,0) + (0,1) = 3\hat{e}_1 + \hat{e}_2\,,$$

or, more generally,

$$\vec{A} = (a_1, a_2) = (A_1, A_2) = a_1\hat{e}_1 + a_2\hat{e}_2 = A_x\hat{\imath} + A_y\hat{\jmath},$$

etc. It is an unpleasant but unavoidable fact that sometimes numerical subscripts must be used to distinguish *different vectors* from each other, as well as, or instead of, to distinguish the *different coordinates* of the *same* vector. Since this issue will arise frequently in exercises and examples in the rest of this book, we shall belabor it a bit here. When we need to discuss two vectors in a two-dimensional space, we might call the vectors $\vec{x}$ and $\vec{y}$ and write them out as

$$\vec{x} = (x_1, x_2), \qquad \vec{y} = (y_1, y_2).$$

On another occasion, however, we may call the vectors $\vec{x}_1$ and $\vec{x}_2$ and write them out as

$$\vec{x}_1 = (x_1, y_1), \qquad \vec{x}_2 = (x_2, y_2).$$

Yes, this can be confusing, but an attempt to stick to a consistent notation would be misguided. Both notational conventions are used in the Real World, each has advantages under certain circumstances, and you must be prepared to handle whichever notation arises in any particular problem. In applications a vector (x, y, z) representing a point in physical space is often denoted by $\vec{r}$ (rather than $\vec{x}$), and we shall follow that convention when it seems best.

Remark on Terminology: Strictly speaking, a *coordinate* of the vector $(3, 1) = 3\hat{\imath} + \hat{\jmath}$ is one of the *numbers*, 3 and 1, that appear in its expansion as a linear combination of basis vectors; whereas a *component* of the vector is another *vector*, the part of the given vector that points along one of the basis vectors: the component of $(3, 1)$ along $\hat{\imath}$ is $3\hat{\imath}$ or $(3, 0)$. However, "component" is frequently used also to mean the same thing as "coordinate"; indeed, "coordinate" sounds strange in nongeometrical contexts (such as the example in the next paragraph).

It should be noted that n-tuples also occur in contexts where the directions involved are not in physical space. For example, in an economic or business application the numerical components a_j , b_j , … $(j = 1, \ldots, n)$ of some vectors $\vec{a}$, $\vec{b}$, … may be the prices, production quantities, and so on of n commodities. For instance, suppose a shirt factory produces per month

b_1 T-shirts of size S,
b_2 T-shirts of size M,
b_3 T-shirts of size L,
b_4 T-shirts of size XL,

and the company prices size-S shirts at p_1 dollars each, etc. Then the production level of the factory is summarized by the vector $\vec{b}$; the vector $3\vec{b}$ is what would be produced by three identical such factories; and

$$\sum_{j=1}^{4} p_j b_j \equiv \vec{p} \cdot \vec{b} \tag{1}$$

is the total revenue brought in by selling all the shirts produced.

In the expression (1) we recognize the familiar *vector dot product*, which also appears in such formulas from physics as $W = \vec{F} \cdot \vec{x}$ for the work done by a force that moves a body through a displacement $\vec{x}$. This operation of multiplying two vectors to get a number is not part of the definition of a general vector space; it is "extra structure" that exists in some concrete vector spaces but not others. It will be treated further in Chapter 6. (See also the discussions of row vs. column vectors in Sections 2.4, 4.5, and elsewhere.)

To most students the word "vector" already calls to mind one or the other of the two types of vectors just discussed. The next two examples are equally familiar mathematical objects, but possibly you have never thought of them as vectors.

Example 3: *Polynomials.* Consider the *power functions*

$$\{1, t, t^2, t^3, \dots\}.$$

These can be added together with numerical coefficients: e.g., $3t^3 - t + 2$. (Such a thing is called a *linear combination* of the vectors you start with.) We are used to manipulating these coefficients just like the components of ordinary vectors: We know how to add them by combining terms,

$$(t^2 + 3) + (t - 5) = t^2 + t - 2,$$

and we also know that the result can't be simplified any further. Note that the powers are playing the same role as the unit vectors $\{\hat{\imath}, \hat{\jmath}\}$ along the coordinate axes in Example 2. In each case we have a certain list of vectors from which all the vectors in the space can be built up by linear combination (and none of the vectors in the list can be left out); such a list is called a *basis.* The crucial difference is that in the vector space of polynomials the basis list is infinite (although any particular polynomial contains only finitely many terms).

Example 4: *Solutions of homogeneous linear differential equations.* Consider the ordinary differential equation

$$\frac{d^2y}{dt^2} + 4y = 0.$$

One of the first things taught in a course on differential equations is that this type of equation is most easily solved by means of complex numbers. The exponential of an imaginary number is defined as

$$e^{i\theta} = \cos\theta + i\sin\theta, \tag{2}$$

and the resulting exponential function of a complex variable can be shown to possess the familiar algebraic and calculus properties of the exponential of a real variable. Therefore,

$$y = f(t) = A\,e^{2it} + B\,e^{-2it}, \tag{3}$$

with A and B arbitrary complex numbers, is a solution of the differential equation. Furthermore, every complex-valued solution is of this form. The space of all these is a vector space. (We get new solutions by adding old ones together and by multiplying old ones by numbers. This is called the "principle of superposition" for homogeneous linear equations.) In fact, the formula sets up an identification between these vectors and the pairs of *complex* coefficients: $f \leftrightarrow (A, B)$. This is another example of an *isomorphism*, and the two exponential functions are another example of a basis.

A different isomorphism is given by the equally valid formula

$$f(t) = C\cos(2t) + D\sin(2t). \tag{4}$$

Formula (2) and the formulas

$$e^{-i\theta} = \cos\theta - i\sin\theta,$$

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}, \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \tag{5}$$

that follow from it enable us to write C and D in terms of A and B or vice versa. (If this is not already familiar to you, you should do Exercise 1.1.12 now!) The passage from (3) to (4) is very much like the effect of a *rotation of axes* in Examples 1 and 2. If we are interested only in *real-valued* solutions, we'll prefer the trigonometric basis in (4) and require C and D to be real instead of complex. On the other hand, the exponential basis in (3) is very useful in finding the solutions in the first place. (Just substitute $y = e^{rt}$ and solve for r. The reason this trick works is that the derivative of an exponential function is a numerical multiple of the function itself. Such a function is an *eigenvector* of the operation of differentiation — see Chapter 8.)

Many of the facts about linear ordinary differential equations taught in differential-equation courses are actually general principles of linear algebra (vector-space theory), not special to ordinary differential equations. Understanding this makes it easy to see how to extend those methods to *partial* differential equations, for instance.

A central theme of this book — and of all mathematics — is the "up and back down" process demonstrated in the broadening of the vector concept from Examples 1 and 2 to include Examples 3 and 4. Experience with the comparatively elementary concepts of geometrical and numerical vectors leads to the formulation of a more abstract and general concept of a vector space (whose details we have postponed to Chapter 3). One then recognizes vectors in many other concrete situations, such as differential equations, and the abstract concept and the theorems derived from it provide powerful new tools for understanding those subjects and solving problems within them.

Exercises

1.1.1 Let $\vec{u} = (1, 2)$ and $\vec{v} = (1, -1)$. On graph paper, sketch $\vec{u}$ and $\vec{v}$ as arrows rooted at the origin. Then, from the geometrical definitions of the vector operations, sketch (and label)

(a) $\vec{u} + \vec{v}$, (b) $\vec{u} - \vec{v}$,

(c) $3\vec{v}$, (d) $\frac{1}{2}\vec{u} - 2\vec{v}$.

(A protractor, as well as a ruler, will be helpful.) Finally, check how well your results agree with the arithmetical definitions of these vectors. (Correct any gross errors in the sketch, but accept minor inaccuracies as the inevitable result of the approximate nature of the sketching process.)

1.1.2 Let $\vec{x} = (1, 2, 3)$, $\vec{y} = (1, -1, 1)$, $\vec{z} = (-2, 2, -2)$. Calculate

(a) $\vec{x} + \vec{y}$, (b) $\vec{z} - \vec{x}$,

(c) $3\vec{x} - 4\vec{y} + \vec{z}$, (d) $\vec{z} + 2\vec{y}$.

1.1.3 For the vectors defined in Exercise 1.1.2, calculate

(a) $\vec{x} \cdot \vec{y}$, (b) $\vec{x} \cdot \vec{z}$, (c) $\vec{z} \cdot \vec{y}$.

1.1.4 Compare parts (a) and (b) of Exercise 1.1.3, and formulate the general theorem that the comparison illustrates.

1.1.5 For the vectors defined in Exercise 1.1.2, write out and simplify the formula $\vec{v} \equiv a\vec{x} + b\vec{y} + c\vec{z}$, where a, b, and c are arbitrary numbers. Then show that:

(a) Any change in b can be compensated by a change in c (so that $\vec{v}$ is unchanged).

(b) The coefficient a cannot be changed without changing $\vec{v}$, regardless of what happens to the other two coefficients.

1.1.6 Simplify:

(a) $3(t^2 + 3t + 2) - 10(t^3 + t^2 - 10) + 4(t - 1)^2 - t + 5$

(b) $-6(\cos(2t) + 3\sin(2t)) + 5\cos(2t) + 3e^{2it}$

1.1.7 Let $\vec{x} = (1, 1)$ and $\vec{y} = (0, 2)$.

(a) *(trivial)* Express $\vec{x}$ as a linear combination of the basis vectors $\hat{\imath}$ and $\hat{\jmath}$.

(b) Express $\hat{\imath}$ as a linear combination of $\vec{x}$ and $\vec{y}$.

1.1.8

(a) Simplify $(t-3)^3 + 5(t-3)^2 - 10(t-3) + 1$ into the standard form for polynomials (a linear combination of powers of t).

(b) Express $t^2 + 2$ as a linear combination of the powers of $(t - 3)$.

1.1.9

(a) Express $3\cos t - 2\sin t$ as a linear combination of e^{it} and e^{-it}.

(b) Express $e^{3it} - 2e^{-3it}$ as a linear combination of the trigonometric functions $\sin(3t)$ and $\cos(3t)$.

1.1.10 Let $\vec{u} = (10, 5, 0.1)$ represent the amount of flour, sugar, and baking powder required to produce a dozen bagels, and let $\vec{v} = (20, 7, 0.5)$ be the corresponding vector for a loaf of bread. Provide in words an interpretation for $60\vec{u} + 200\vec{v}$.

1.1.11 Whizbang Supersystems Inc. manufactures three models of computer, the Nerdstation 1000, 3000, and 5000. Each is priced according to its name (the cheapest costs \$1000, etc.)

(a) Show how the total revenue of the company can be expressed as a dot product of two vectors. (Define notation clearly.)

(b) The fixed cost of operating the Whizbang factory is \$100,000 per year. The production costs of Nerdstations is expressed by the vector $\vec{c} = (500, 1000, 1500)$. (That is, it costs \$500 to make one Model 1000, etc.) Find a formula in vector notation for the company's annual profit.

1.1.12 (a) Prove (5) from (2) (assuming the well known facts $\cos(-\theta) = \cos\theta$, $\sin(-\theta) = -\sin\theta$).

(b) Using (2) and (5), work out the equations expressing A and B in (3) in terms of C and D in (4), and the equations expressing C and D in terms of A and B.

1.2 Lines and Planes

When dealing with $\mathbf{R}^3$ or with physical space modeled by $\mathbf{R}^3$ (see the first two examples in Sec. 1.1), it is convenient to represent the different coordinates of a vector by different letters, thereby minimizing the use of subscripts:

$$\vec{r} = (x, y, z).$$

Often one deals with problems that are essentially two-dimensional; then the z component can be dropped:

$$\vec{r} = (x, y) \in \mathbf{R}^2.$$

Once the *origin* of a coordinate system in physical space has been fixed, each point in space can be identified with the vector with head at the point and tail at the origin.

As you know, a *(straight) line* in $\mathbf{R}^2$ can be defined by an equation of the form

$$ax + by = c, \tag{1}$$

where a, b, and c are constants. (Obviously, the equation is not unique, since all three constants can be multiplied by a nonzero number without changing the set of solutions (x, y).) Let us call this the *equation form* of a line. (If b is not zero, (1) can be rearranged into the "functional" form, $y = mx + d$.)

There is, however, another, equally good, way of representing a line. Introduce a new variable, t, which ranges through all the real numbers $(-\infty < t < \infty)$. Consider (for example) the functions

$$x = 3t - 2, \qquad y = 2t.$$

Plotting the points (x, y) on graph paper for various values of t, one easily sees that they form a line. We can put the coordinates together in vectorial form:

$$\vec{r}(t) = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3t - 2 \\ 2t \end{pmatrix}. \tag{$*$}$$

Using the definitions of vector addition and multiplication, we can rewrite this as

$$\vec{r} = t \begin{pmatrix} 3 \\ 2 \end{pmatrix} + \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

Thus each point on the line can be obtained by adding a fixed, or constant, vector to an arbitrary multiple of another fixed vector. The general case is

$$\vec{r} = t\vec{u} + \vec{r}_0 \tag{2}$$

(where vectors $\vec{u}$ and $\vec{r}_0$ are fixed, and t and $\vec{r}$ are variables). This is called the *parametric form* of a line (t being the “parameter” involved).

Notational remark: In the foregoing paragraph we have begun the practice of writing vectors as *columns* of numbers instead of rows. This made the equation $(*)$ easier to read by separating the two coordinate formulas visually. Later (Secs. 2.4 and 3.2) we will encounter a more important reason for using columns, when a distinction will be made between two kinds of vectors, one written as columns and one as rows. On the other hand, column vectors are difficult to typeset and use up a lot of paper, so it is quite common to revert to the row notation when there is no danger of confusion.

Example 1. What is the equation form of the line $(*)$?

Solution *(Method 1)*: Solve one of the coordinate equations for the parameter, and substitute into the other coordinate equation. Since $t = \frac{1}{2}y$, we have

$$x = \tfrac{3}{2}y - 2,$$

which can trivially be rearranged into the form (1).

Before giving a second method for solving this problem, we remark that if $\vec{r}_0$ in (2) is the *zero vector*,

$$\vec{0} \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

then the line passes through the origin in $\mathbf{R}^2$. (On the other hand, if $\vec{r}_0 \neq \vec{0}$, the line may pass through the origin anyhow. See Exercise 1.2.11.) In this case the equation form is

$$0 = ax + by = \vec{a} \cdot \vec{r}.$$

That is, the vector $\vec{a} \equiv (a, b)$ is perpendicular* to all the vectors $\vec{r}$ making up the line. In the more general situation (1), the corresponding statement is that $\vec{a}$ *is perpendicular to the vector joining any two points on the line.*

PROOF: Let $\vec{r}_1 = (x_1, y_1)$ and $\vec{r}_2 = (x_2, y_2)$ be two points on the line. Then their coordinates satisfy

$$ax_1 + by_1 = c, \qquad ax_2 + by_2 = c.$$

Subtract these two equations to get

$$a(x_1 - x_2) + b(y_1 - y_2) = 0.$$

That is, $\vec{r}_1 - \vec{r}_2$ (the vector with head at $\vec{r}_1$ and tail at $\vec{r}_2$) is perpendicular to $\vec{a}$.

On the other hand, in the notation of (2) the difference between two vectors $\vec{r}_1$ and $\vec{r}_2$ on the line is always a multiple of $\vec{u}$. Thus $\vec{u}$ is a vector tangent to the line, and $\vec{u}$ *must be perpendicular to* $\vec{a}$. It follows (take the dot product of equation (2) with $\vec{a}$) that $\vec{a} \cdot \vec{r} = \vec{a} \cdot \vec{r}_0$, and therefore the c in (1) equals $\vec{a} \cdot \vec{r}_0$.

Theorem 1: In $\mathbf{R}^2$ the vectors

$$\begin{pmatrix} a \\ b \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} b \\ -a \end{pmatrix}$$

---

* The words "perpendicular", "normal", and "orthogonal" are all equivalent when applied to vectors, lines, or planes. ("Orthogonal matrix" means something else, however, as will be explained in due course.)

are perpendicular. (One is obtained by rotating the other through a right angle.)

This is the vectorial version of the familiar theorem, "Two lines are perpendicular if and only if the product of their slopes is -1."

SOLUTION TO EXAMPLE 1 *(Method 2)*: Since $\vec{u} = (3, 2)$, a vector perpendicular to it is $\vec{a} = (2, -3)$.[†] Then

$$\vec{a} \cdot \vec{r}_0 = (2, -3) \cdot (-2, 0) = -4,$$

so an equation for the line is

$$2x - 3y = -4.$$

(This is equivalent to the result of the first method.)

Example 2. Find a parametric representation of the line $x + 7y = 2$.

SOLUTION: Let $y = t$. Then $x = -7t + 2$. So the line is

$$\vec{r} = \begin{pmatrix} -7t + 2 \\ t \end{pmatrix} = t \begin{pmatrix} -7 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

Note that there are many other correct answers, because our initial step was rather arbitrary. We could have taken $x = t$, or $y = 2t + 5$. The result would have been the same line, but with a different labeling of points by numbers t.

The parametric representation (2) makes sense in space of any dimension, not just in $\mathbf{R}^2$. (The same cannot be said for (1); we return to that problem later.)

Example 3. Find a parametric form of the line in $\mathbf{R}^3$ through the points $(3, 2, 4)$ and $(1, 1, -1)$.

SOLUTION: The vector difference between the two points, $(2, 1, 5)$, is parallel to the line and is therefore a suitable $\vec{u}$. The line passes through $(1, 1, -1)$, so that is a suitable $\vec{r}_0$. Therefore, we can write

$$\vec{r} = t(2, 1, 5) + (1, 1, -1),$$

or

$$x = 2t + 1, \qquad y = t + 1, \qquad z = 5t - 1.$$

---

† $(-2, 3)$ would do equally well in the role of $\vec{a}$. But you must use the same $\vec{a}$ on both sides of the equation $\vec{a} \cdot \vec{r} = \vec{a} \cdot \vec{r}_0$.

To check, we see that we recover the two original points by taking $t = 1$ and $t = 0$.

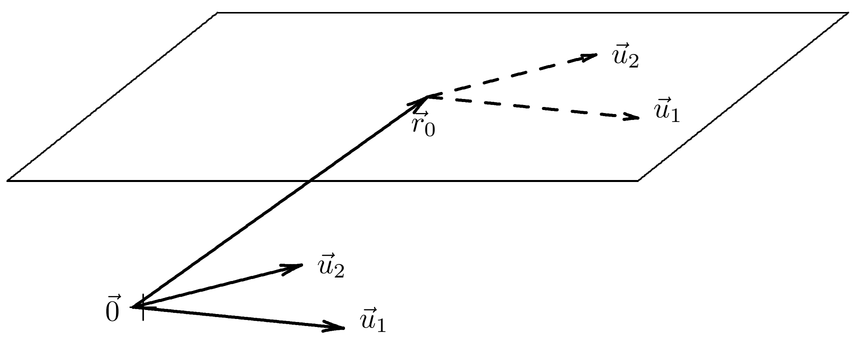
The analogue of (1) for *planes* in $\mathbf{R}^3$ is

$$ax + by + cz = d. \tag{3}$$

Planes also have parametric representations (which may be less familiar from elementary courses than the previous material in this section). Since a plane is a two-dimensional entity, one needs *two* independently varying parameters in order to sweep out all the points on it. It is not hard to guess that the correct analogue of (2) is

$$\vec{r} = s\vec{u}_1 + t\vec{u}_2 + \vec{r}_0\,, \tag{4}$$

where $\vec{r}_0$ is a point in the plane, and $\vec{u}_1$ and $\vec{u}_2$ are vectors tangent to the plane. We will demonstrate this with a sketch and with several examples.



Example 4. Find the parametric form of the plane

$$x - 3y + z = 0.$$

Solution: Choose $y = s$ and $z = t$ and solve for x as $3s - t$. Then

$$\vec{r} = \begin{pmatrix} 3s - t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

Example 5. Find the equation form of the plane

$$\vec{r} = s(5, 0, -1) + t(2, 2, 0).$$

SOLUTION: By analogy with the previous discussion of lines in $\mathbf{R}^2$, it is clear that the equation must be of the form $\vec{n} \cdot \vec{r} = 0$, where $\vec{n}$ is a vector perpendicular to the given $\vec{u}_1 = (5, 0, -1)$ and $\vec{u}_2 = (2, 2, 0)$. The main problem is finding such a vector.

Method 1: With the notation of (3) in mind, let $\vec{n} = (a, b, c)$. Write out the conditions of perpendicularity:

$$\vec{n} \cdot \vec{u}_1 = 5a - c = 0, \qquad \vec{n} \cdot \vec{u}_2 = 2a + 2b = 0.$$

Since *any* perpendicular vector will do, we can choose one coordinate arbitrarily (but not 0), say $c = 5$ (to avoid fractions). Then $a = 1$ and $b = -1$:

$$\vec{n} = (1, -1, 5).$$

In other words, the plane is $x - y + 5z = 0$.

Method 2: The analogue of Theorem 1 for three dimensions is provided by the *vector cross product*:

$$\vec{u} \times \vec{v} \equiv \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \equiv \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix}.$$

(The determinant notation here will be explained in Sec. 2.5, in case you are not familiar with it.) The cross product of two vectors is always perpendicular to both of them. If $\vec{u}$ and $\vec{v}$ are not parallel to each other, then $\vec{u} \times \vec{v}$ is not zero. In our plane problem, therefore, $\vec{u}_1 \times \vec{u}_2$ is suitable for use as $\vec{n}$. We calculate

$$\vec{u}_1 \times \vec{u}_2 = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ 5 & 0 & -1 \\ 2 & 2 & 0 \end{vmatrix} = \begin{pmatrix} 2 \\ -2 \\ 10 \end{pmatrix}.$$

This is not the same vector $\vec{n}$ we got with the other method, but that is not a problem: The two vectors are proportional, and (3) can be multiplied by any constant (on both sides, of course!) without changing the plane.

Example 6. Find a parametric form of the plane passing through the points $(1, 0, 0)$, $(2, 2, 2)$, and $(-1, 0, 6)$.

SOLUTION: The difference between any two points on the plane is a vector parallel to the plane, therefore a suitable candidate for $\vec{u}_1$ or $\vec{u}_2$. We arbitrarily choose to subtract the first given vector from each of the other two:

$$\vec{u}_1 = (1, 2, 2), \qquad \vec{u}_2 = (-2, 0, 6).$$

Let us also choose the first vector as $\vec{r}_0$. Then

$$\vec{r} = s\begin{pmatrix}1\\2\\2\end{pmatrix} + t\begin{pmatrix}-2\\0\\6\end{pmatrix} + \begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}s-2t+1\\2s\\2s+6t\end{pmatrix}.$$

Example 7. Find the equation form of the plane in the previous example.

SOLUTION: Reasoning as we did for lines, we see that the equation is $\vec{n}\cdot\vec{r} = d$, where $\vec{n}$ is perpendicular to the differences between any points in the plane (a clumsy vectorial way of saying "perpendicular to the plane"), and $d = \vec{n}\cdot\vec{r}_0$ for any point $\vec{r}_0$ in the plane. Therefore, it is easy to get these ingredients from the parametric form found in the previous example. By either of the methods in Example 5 we find that $\vec{n} = (6, -5, 2)$ is a vector perpendicular to $\vec{u}_1$ and $\vec{u}_2$. Using the $\vec{r}_0$ in Example 6, we get $\vec{n}\cdot\vec{r}_0 = 6$. Thus the equation of the plane is

$$6x - 5y + 2z = 6.$$

We summarize the principles used in solving these examples in the following theorem.

Theorem 2: In $\mathbf{R}^3$, the parametrized plane (4)

$$\vec{r} = s\vec{u}_1 + t\vec{u}_2 + \vec{r}_0$$

consists precisely of all the points satisfying the equation (3)

$$\vec{n}\cdot\vec{r} = d,$$

where $\vec{n}$ is any nonzero vector perpendicular to $\vec{u}_1$ and $\vec{u}_2$, and $d = \vec{n}\cdot\vec{r}_0$. (Here it is understood that $\vec{u}_1$ and $\vec{u}_2$ are not zero and are not parallel to each other.)

PROOF: We must show that the set of points (4) is the same as the set of points satisfying the equation (3). Therefore, we check that each set is a subset of the other one.

(i) If $\vec{r}$ is of the form (4), and $\vec{n}$ and d are as described, then taking the dot product of (4) with $\vec{n}$ indeed yields $\vec{n}\cdot\vec{r} = d$.

(ii) If $\vec{r}_1$ and $\vec{r}_2$ are two solutions of (3), then $\vec{r}_1 - \vec{r}_2$ is perpendicular to $\vec{n}$. (The proof of this is the same as that of the corresponding statement for lines, given earlier.) Since $\vec{r}_0$ solves (3), it follows that every solution $\vec{r}$ can be written as

$$\vec{r} = \vec{r}_0 + (\vec{r} - \vec{r}_0) = \vec{r}_0 + \vec{u},$$

where $\vec{u} \cdot \vec{n} = 0$. As in Example 4, it is easy to see that there will always be a two-parameter family of vectors $\vec{u}$ satisfying this perpendicularity condition. Thus the solution space of (3) can be represented in the form (4) for *some* $\vec{u}_1$ and $\vec{u}_2$. It is geometrically obvious that these can be chosen to be the same $\vec{u}_1$ and $\vec{u}_2$ that we started from; a formal proof of this requires some concepts that will be developed in later chapters.

We have shown that (1) and (2) are alternative descriptions of lines in two-dimensional space, and that (3) and (4) are alternative descriptions of planes in three-dimensional space. We observed that lines in $\mathbf{R}^3$ also have the parametrical representation (2). It is natural to ask whether a line in three dimensions has an equation form, analogous to (1) and (3). Since a line has only one free parameter, it is clear that there must be *two* conditions relating the three coordinates of a point on the line. In fact, since the intersection of two planes is a line (unless the planes are parallel), a line will be defined by *two* equations of the type (3). We leave further investigation of this situation to the exercises.

Exercises

1.2.1 Express in parametric form ($\vec{r} = t\vec{u} + \vec{r}_0$):

(a) The line through the origin in $\mathbf{R}^3$ parallel to the vector $(1, 0, 1)$.

(b) The line in $\mathbf{R}^2$ through the points $(1, 0)$ and $(0, -1)$.

(c) The line through $(2, 3)$ parallel to the vector $(1, 2)$.

(d) The line through the points $(1, 5, 7)$ and $(2, 10, 14)$.

1.2.2 Find an equation of the form $ax + by = c$ for these lines in $\mathbf{R}^2$.

(a) The line through the points $(4, 7)$ and $(2, -1)$.

(b) The line with parametric equation $\vec{r} = t(1, 1) + (4, -1)$.

(c) The line through the origin parallel to $(5, 1)$.

(d) The line with parametric equation $\vec{r} = t(0, 1) + (-2, -1)$.

1.2.3 Express in parametric form the plane defined by the equation

$$9x - 3y + z = 2.$$

1.2.4 Let $\vec{n} = (-1, 0, 2)$. The equation $\vec{n} \cdot \vec{r} = 5$ defines a plane in $\mathbf{R}^3$. Express the plane in the parametrized form $\vec{r} = s\vec{u}_1 + t\vec{u}_2 + \vec{r}_0$.

1.2.5 Find an equation $(ax + by + cz = d)$ for:

(a) The plane with parametric form $\vec{r} = s(1, 0, -1) + t(2, 1, 2)$.

(b) The plane through the origin perpendicular to $(2, 2, 1)$.

1.2.6 Find both a parametric form and an equation form for the plane passing through $(1, 0, 0)$, $(1, 1, 1)$, and the origin.

1.2.7 Express in equation form the plane whose parametric form is

$$\vec{r} = s(1, -1, 2) + t(2, 2, 2) + (5, 0, 2).$$

1.2.8 Find an equation and a parametric representation for the plane passing through the points $(1, 0, 1)$, $(2, 3, 1)$, $(5, 4, 5)$.

1.2.9 Find an equation and a parametric representation for the plane perpendicular to $\vec{n} = (3, 1, 1)$ and passing through $(4, 7, -1)$.

1.2.10 Prove Theorem 1 (including the statement in parentheses). Hint: Besides perpendicularity, what does "rotation through a right angle" entail?

1.2.11 Suppose that the line $\vec{r} = t\vec{u} + \vec{r}_0$ passes through the origin, but nevertheless $\vec{r}_0$ is not $\vec{0}$. How are $\vec{u}$ and $\vec{r}_0$ related in this situation?

1.2.12 Give an example of a plane $\vec{r} = s\vec{u}_1 + t\vec{u}_2 + \vec{r}_0$ such that (1) the plane passes through the origin, and (2) $\vec{r}_0$ is not $\vec{0}$ and is not parallel to either $\vec{u}_1$ or $\vec{u}_2$.

1.2.13 The analogue of Theorem 2 for a *line in* $\mathbf{R}^2$ was not stated formally as a theorem; you have to search through the expository text to find all the relevant conclusions. Summarize them formally, in analogy with Theorem 2.

1.2.14 In the notation of Theorem 2, show that if $\vec{n}$ has length 1, then $|d|$ is the distance of the plane from the origin.

1.2.15 The intersection of the planes

$$x + y + z = 2 \qquad \text{and} \qquad x - 2y + z = 0$$

is a line in $\mathbf{R}^3$. Find a parametric representation of this line. (*Suggestion:* Set the coordinate z equal to the parameter t.)

1.2.16 Characterize the line $\vec{r} = t(1, 0, -1)$ by a pair of equations, $\vec{n}_1 \cdot \vec{r} = 0 = \vec{n}_2 \cdot \vec{r}$. (Find two distinct planes containing the line. There are many different correct answers!)

1.2.17 Let $\vec{x} = (1, 0)$ and $\vec{y} = (2, 1)$. On a piece of graph paper plot and label the points $t\vec{x} + (1 - t)\vec{y}$ for $t = -0.5$, 0, 0.2, 0.5, 0.9, 1, and 1.2. From this example, formulate a general principle. (What is special about the points corresponding to $0 \le t \le 1$? These points are called *convex combinations* of the two given vectors.)

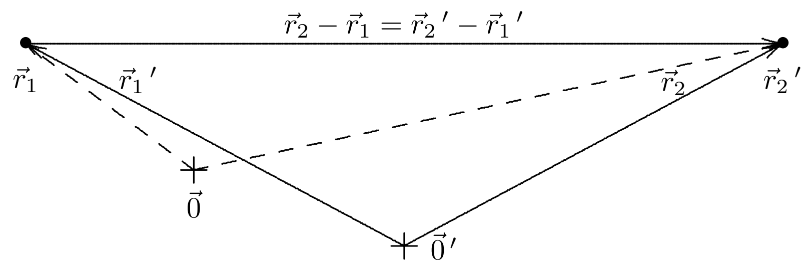
1.3 Points: A Deeper Look

In the first section of this book we listed various types of physical vectors: forces, velocities, and so on. Conspicuously absent from that list of vectorial physical quantities was *position*, the most fundamental of them all. The reason is that there is a special subtlety to the concept of a position vector, which would have made it a misleading example then.

At the beginning of the second section it was observed that once the origin of coordinates is fixed, a point in space (in other words, a possible value for a position variable) can be associated with a vector with tail at the origin. The rest of that section, therefore, was written as if points and vectors are basically the same thing.

Another kind of vector with the same physical units (e.g., meters) as position is *displacement*, the difference between two positions. If two particles are located at positions $\vec{r}_1$ and $\vec{r}_2$ then the second particle is displaced from the first by the vector $\vec{r}_2 - \vec{r}_1$. More generally, the vector difference can be interpreted as the displacement, or relative position, of one geometrical point relative to another point, even when the points are not occupied by physical particles. In the parametric equation of a line, $\vec{r} = t\vec{u} + \vec{r}_0$, $\vec{u}$ and $t\vec{u}$ are displacement vectors, while $\vec{r}_0$ and $\vec{r}$ are position vectors in the absolute sense.

This distinction is not just conceptual; there is a real mathematical difference between the two kinds of vectors. When we use $\mathbf{R}^3$ to model physical space, the location of the origin is chosen arbitrarily. (Of course, in any particular application or calculation, some choices may be more convenient or natural than others.) If the origin is changed, then *the position vectors representing the points* $\vec{r}_1$ *and* $\vec{r}_2$ *change*. However, the displacement vector $\vec{r}_2 - \vec{r}_1$ stays the same.



Similarly, vectors of velocity, force, etc. do not depend on the choice of origin. Vectors of each such type belong to their own space, which is not the same as the physical space. The origin of velocity space is the condition of being at rest, which has nothing to do with being located at a particular point. In fact, in pictorial representations one usually thinks of the origins of the spaces of possible velocities, forces, etc. of a particle as being located *at the location of the particle concerned*, not at the origin of coordinates. That is, such vectors should be drawn with their tails at the particle (as indicated in one of the sketches in the next section). This is important for visualizing the effects of such velocities and forces in generating displacements from that point (see the next section).

In some books, position vectors are called *bound vectors*, and vectors that are independent of the coordinate origin are called *free vectors*.

The distinction between two types of vectors has implications for the algebraic operations on vectors. The sum of a position vector and a displacement is a new position vector. The sum of two displacements is another displacement. But to add two position vectors (or to add two points in space to each other) is meaningless.

A very similar situation arises in computer programming in connection with *pointers*, which label locations in the computer's memory.* The sketch

* B. W. Kernighan and D. M. Ritchie, *The C Programming Language*, Prentice–Hall, Englewood Cliffs, 1978, Chapter 5.

shows a model of a computer memory.

1	2	$\cdots$	$2^{32} - 1$

"A pointer is a variable that contains the address of another variable." At one level it is just a number. However, the computer language needs to make a rigid distinction between pointers and ordinary integers. A pointer is like a position vector; the difference between two pointers is an integer (which can be used, for instance, as a subscript of an array) and is like a displacement vector. It is therefore legal to subtract two pointers (getting an integer) or to add or subtract an integer from a pointer (getting a new pointer), but attempting to add two pointers should produce an error message. Pointers, and only they, depend on "origin": The pointers locating an array (subscripted variable) in memory may change as the program is revised, or executed under different circumstances, but the integer subscript that indexes a particular element of the array will stay the same.

Returning to physical vectors, observe that the change of origin is unlike other changes of coordinates, such as a rotation of axes around the origin, or a change in the units in which the coordinates are measured from meters to feet. Coordinate (or basis) changes of the latter type (which will be studied extensively in Chapter 4) change the *numerical representation* of a vector as a string of numbers, but leave the vector itself, as an abstract object, unchanged (and there is no difference between free and bound vectors in these respects). The discussion above shows that for *points* there is another necessary level of abstraction: The representation of points by vectors is somewhat arbitrary (depending on the choice of origin), just as the representation of physical vectors by n-tuples of numbers is arbitrary (depending on the direction and scale of the axes).

There are physical situations where the space of possible positions is not "flat" (such as the space-time of general relativity, or the space of possible orientations of a rigid body). In such cases the points cannot be regarded as vectors at all. Nevertheless, velocities (more generally, tangents to curves — see the next section) are still correctly modeled by vectors. This is part of the subject called *differential geometry*, for which the material in this book is an important prerequisite.

Length and distance

This is a good opportunity to review some elementary terminology and notation that didn't make their way into the previous sections.

Numbers (real or complex) are called *scalars* when it is desired to distinguish them from vectors. (This terminology carries over to *functions* that take scalar or vector values, respectively.) The operation of multiplying a vector by a number is called *scalar multiplication* to distinguish it from other kinds of multiplication involving vectors (such as the dot and cross products).

In the spaces $\mathbf{R}^n$ and the spaces of physical quantities modeled by them, each vector has a *length* defined by

$$\|\vec{v}\| = \sqrt{\vec{v}\cdot\vec{v}} = \sqrt{\sum_{j=1}^{n} |v_j|^2}\,.$$

(The notation $|\vec{v}|$ is also used.) This scalar tells the "size" of the vector, discarding the directional information.

The *distance* between two vectors, $\vec{u}$ and $\vec{v}$, is the length of their difference vector, $\|\vec{u}-\vec{v}\|$. When the two vectors represent points, their difference is a displacement vector and hence the distance is independent of the choice of origin.

Exercises

1.3.1 Which of the following operations make sense, and which kind of vector is the result in each case?

(a) Subtraction of one position vector from another.

(b) Subtraction of a displacement vector from a position vector.

(c) Subtraction of one displacement vector from another.

(d) Subtraction of a position vector from a displacement vector.

1.3.2 Let $\vec{r}_1 = 2\hat{\imath}+\hat{\jmath}-3\hat{k}$, $\vec{r}_2 = \hat{\imath}+\hat{\jmath}+\hat{k}$. Calculate $\|\vec{r}_1\|$, $\|\vec{r}_2\|$, and $\|\vec{r}_2-\vec{r}_1\|$. (Remember that $2\hat{\imath}+\hat{\jmath}-3\hat{k}$ means the same thing as $(2,1,-3)$.)

1.3.3 Redefine coordinates in $\mathbf{R}^3$ by

$$x' = x - 1, \quad y' = y + 2, \quad z' = z.$$

Calculate the primed coordinates of the vectors $\vec{r}_1$ and $\vec{r}_2$ of the previous exercise, and verify that $\vec{r}_2 - \vec{r}_1$ and $\|\vec{r}_2 - \vec{r}_1\|$ are the same when calculated in the primed coordinates as in the original coordinates, although the numerical values of $\|\vec{r}_1\|$ and $\|\vec{r}_2\|$ change.

1.4 Curves and Tangent Vectors

Vectors are associated not only with straight lines but also with curved lines. A curve can be described by a vector-valued function of a real variable,

$$f\colon \mathbf{R} \to \mathbf{R}^p.$$

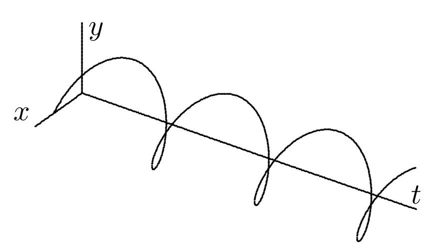
Or, to look at the situation from the other way around, a function from $\mathbf{R}$ into $\mathbf{R}^p$ can be represented geometrically by a curve. (The notation $f\colon \mathbf{R} \to \mathbf{R}^p$ means that f is a function that takes elements of $\mathbf{R}$ as input and yields elements of $\mathbf{R}^p$ as output.)

In fact, there are two different ways in which we can visualize such a function as a curve.

1. We can *graph* the function in a space of dimension $p+1$. For example, if $p = 2$ and

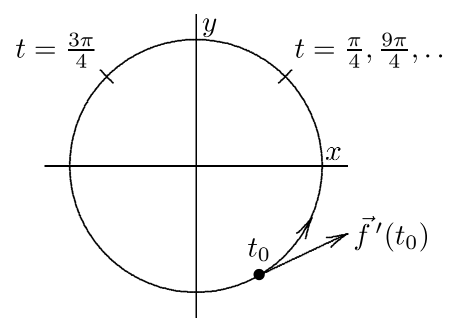
$$\vec{f}(t) = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \text{where} \qquad x = \cos t, \quad y = \sin t,$$

then (for a certain orientation of the axes) the graph looks like this:



This graph is a *helix*. With modern computer software it is not hard to produce a genuinely three-dimensional image of the curve, which can be rotated on the computer screen to reveal the curve's geometrical nature more clearly than a single two-dimensional projection on the printed page can do. But if we insist on visualizing functions this way as p increases, we will quickly run out of dimensions.

2. We can represent the function as a *parametrized curve* in p-dimensional space. That is, for each value of the independent variable, t, we plot the point $\vec{f}(t)$ in $\mathbf{R}^p$. For the previous example the curve is a circle:



We can think of each point as being labeled by the value of t that maps into it, but note that there could be more than one such value. A 3-dimensional example is

$$\vec{g}(t) = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad \text{where} \qquad x = \cos t, \qquad y = \sin t, \qquad z = t.$$

The curve in this case is the same helix as before, with the t-axis relabeled as z-axis. Only its interpretation has changed, and it is important to understand the conceptual difference. In one case there are 3 variables, in the other there are 4 (three dependent and one independent).

We can define the *derivative* of a vector-valued function by taking the ordinary derivative of each of its coordinates:

$$\vec{f}\,'(t) = \begin{pmatrix} f_1'(t) \\ f_2'(t) \\ \vdots \\ f_p'(t) \end{pmatrix}.$$

(A more profound definition will come later.) For our circle,

$$\vec{f}\,'(t) = \begin{pmatrix} \dfrac{dx}{dt} \\ \dfrac{dy}{dt} \end{pmatrix} = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}.$$

If t has the interpretation of time and $\vec{r} = \vec{f}(t)$ that of position, then $\vec{f}\,'(t)$ is the *velocity* at time t. A geometrical interpretation of $\vec{f}\,'$ is as a *tangent vector* to the parametrized curve, with length proportional to the "speed" at which the curve is traced out by the given parametrization. Customarily one thinks of the vector $\vec{f}\,'(t_0)$ as being attached to the corresponding point, $\vec{r}_0 \equiv \vec{f}(t_0)$, on the curve; see the sketch of the circle above. (Note that this point then becomes the *origin* as far as addition, etc., of vectors of this sort is concerned.) Just as the derivative of an ordinary real-valued function is used to construct the tangent line to the graph of the function, the parametric equation of the *tangent line* to the *graph* of $\vec{f}$ at t_0 is

$$\vec{r} = \vec{f}(t_0) + (t - t_0)\vec{f}\,'(t_0). \tag{1}$$

This line in $\mathbf{R}^{p+1}$ can be thought of as the "best straight-line approximation" to the graph in the small neighborhood near $(t_0, \vec{r}_0)$. One can also think of (1) as a parametric equation of a line in $\mathbf{R}^p$ (through the point $\vec{f}(t_0)$, along the tangent vector $\vec{f}\,'(t_0)$); this is the *tangent line* to the *parametrized curve.*

The numerical significance of the tangent vector is this: When $\vec{f}\,'(t_0)$ is multiplied by a small number $dt \equiv t - t_0$, the result is a vector $d\vec{r}$ that tells approximately how $\vec{f}(t)$ is displaced from $\vec{r}_0$. This is an approximation because the curve is being approximated by its tangent line at $\vec{r}_0$ (or, because the graph of $\vec{f}$ is being approximated by *its* tangent line at $(t_0, \vec{r}_0)$).

If $\vec{f}\,'(t_0)$ happens to be the zero vector, then (1) does not define a line. However, it is still true that (1) tells approximately how $\vec{f}(t)$ changes as t moves slightly away from t_0. (That is, $\vec{f}(t)$ is approximately constant in that case!) The fact that (1) is not a line does not necessarily mean that the curve, as a geometrical object in $\mathbf{R}^p$, does not have a tangent line at that point; see Exercises 1.4.4 and 1.4.5.

Later we will see how to generalize all these considerations when the *independent* variable of the function is also multidimensional (see Secs. 2.4 and 3.3–5). A different kind of generalization is to functions whose values $\vec{f}(t)$ lie not in $\mathbf{R}^p$ but in some more general space of vectors, such as those in Examples 1, 3, and 4 of Section 1.1. We'll return to this topic in Sec. 6.3 after building up enough background concepts.

Exercises

1.4.1 Calculate the derivative $\vec{g}\,'(t)$ for the helical curve in the text. Use it to find a parametric representation for the tangent line to the curve at the point where $t = \frac{\pi}{3}$.

1.4.2 Construct the tangent line at $t = \frac{\pi}{3}$ to the circular curve in the text ($x = \cos t,\ y = \sin t$). What is the relationship between this line and the one in the previous exercise?

1.4.3 A particle is forced to move along the trajectory $\vec{h}(t) = (t^2, 1{+}3t, e^{2t})$. At time $t = 2$ the particle is released from the curved track, and therefore moves off along the tangent line at the constant velocity $\vec{h}'(2)$. Where is the particle at time $t = 3$?

1.4.4 Consider the curve $\beta(t) = (t^5, t^3)$ in $\mathbf{R}^2$.

(a) Show that at the point where $t = 0$, the equations of this section define a tangent vector but not a tangent line.

(b) Find a reparametrization of the curve (define a new variable $\tau = \rho(t)$ via some increasing function ρ) that enables the tangent line at the origin to be constructed in the usual way.

1.4.5 Consider the curve $\beta(t) = (t^2, t^3)$ in $\mathbf{R}^2$.

(a) Show that at the point where $t = 0$, the equations of this section define a tangent vector but not a tangent line.

(b) Show that this curve does not have a tangent line at the origin.

1.4.6 Consider the differential equation $\dfrac{d^2y}{dt^2} + \omega^2 y = 0$, where ω is a parameter (independent of t), with initial data $y(0) = 1$, $y'(0) = -2$.

(a) Find the solution, $y(t)$. (Assume that ω is real and positive.) As ω varies, the solution moves along a curve in an infinite-dimensional space of functions. (Think of each function $y(t)$ as a single point on this curve. Keep in mind that the parameter along the curve is ω, not t.)

(b) Find the derivative of the solution with respect to ω at $\omega = 2$. This function plays the role of tangent vector to the curve of solutions.

(c) Use the result of (b) to construct an approximation to the function y when $\omega = 2.15$. This is a point on the tangent line to the curve of solutions at the point labeled by $\omega = 2$.

(d) Appraise the accuracy of the approximation you got in (c). (You can use a computer or a graphing calculator to plot the exact and approximate solutions as functions of t.) Notice the difference between what happens at small $|t|$ and at large $|t|$.