Chapter 6

Inner Products and Differential Vector Calculus

6.1 Inner Products, Norms, and Metrics

In previous chapters we have frequently made use of the familiar dot product in $\mathbf{R}^3$, especially in connection with the applications of linear algebra in multivariable calculus. However, this operation of multiplying two vectors to get a number is definitely not part of the general definition of a vector space (Sec. 3.1) and therefore has played no role in the development of the theory.

In fact, so far our vector spaces have carried no concepts of

- length
- distance
- angle
- orthogonality (perpendicularity).

An *inner product* (which is an *extra* operation besides addition and scalar multiplication) provides these, and also leads to definitions of

- limits
- infinite sums or linear combinations
- continuous functions, differentiable functions, etc.

The prototype of an inner product is the dot product in $\mathbf{R}^n$,

$$\vec{x} \cdot \vec{y} \equiv \sum_{j=1}^{n} x_j y_j \,.$$

It has these properties:

RI1. $\vec{x} \cdot \vec{x} > 0$ — except that $\vec{0} \cdot \vec{0} = 0$. (*positivity*)

RI2. $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$. (*symmetry*)

RI3. $(r\vec{x} + \vec{y}) \cdot \vec{z} = r(\vec{x} \cdot \vec{z}) + \vec{y} \cdot \vec{z}$, and hence
$\vec{x} \cdot (r\vec{y} + \vec{z}) = r(\vec{x} \cdot \vec{y}) + \vec{x} \cdot \vec{z}$. (*bilinearity*)

These conditions make up the **definition** of an inner product in the abstract sense.

REMARK: An inner product is sometimes called a *scalar product*. Note, however, that this is an entirely different operation from *scalar multiplication*. Scalar multiplication combines a number and a vector to form a new vector (is a function of the type $\mathbf{R}\times\mathcal{V}\to\mathcal{V}$), while an inner product combines two vectors to yield a number (is a function of the type $\mathcal{V}\times\mathcal{V}\to\mathbf{R}$).

The dot product defines a *norm*

$$\|\vec{x}\| \equiv \sqrt{\vec{x}\cdot\vec{x}}$$

(commonly called the *length* of the vector), with the properties

N1. $\|\vec{x}\| > 0$ — except that $\|\vec{0}\| = 0$. (*reality* and *positivity*)

N2. $\|r\vec{x}\| = |r|\,\|\vec{x}\|$. (*homogeneity*)

N3. $\|\vec{x}+\vec{y}\| \le \|\vec{x}\| + \|\vec{y}\|$. (*triangle inequality*)

The norm, in turn, defines a *metric*, or *distance* between two vectors:

$$\mathrm{d}(\vec{x},\vec{y}) \equiv \|\vec{x}-\vec{y}\|.$$

It has these properties:

M1. $\mathrm{d}(\vec{x},\vec{y}) > 0$ — except that $\mathrm{d}(\vec{x},\vec{x}) = 0$. (*positivity*)

M2. $\mathrm{d}(\vec{x},\vec{y}) = \mathrm{d}(\vec{y},\vec{x})$. (*symmetry*)

M3. $\mathrm{d}(\vec{x},\vec{z}) \le \mathrm{d}(\vec{x},\vec{y}) + \mathrm{d}(\vec{y},\vec{z})$. (*triangle inequality*)

If the scalars for a vector space are complex numbers instead of real numbers, an interesting complication arises in deciding how to define inner products. What, indeed, should be the dot product and the norm in the elementary complex vector space, $\mathbf{C}^n$?

First try: $\vec{x}\cdot\vec{y} \equiv \sum_{j=1}^{n} x_j y_j\,; \quad \|\vec{x}\| = \sqrt{\vec{x}\cdot\vec{x}}\,.$ $\qquad * * *$ **WRONG** $* * *$

This violates positivity. If x_j is a complex number, $(x_j)^2$ is not necessarily positive, or even real. This would lead to weird definitions of length and distance.

Second try: $\vec{x}\cdot\vec{y} \equiv \sum_{j=1}^{n} x_j \overline{y_j}\,; \quad \|\vec{x}\|^2 \equiv \vec{x}\cdot\vec{x} = \sum_{j=1}^{n} |x_j|^2.$

(The bar indicates the complex conjugate: $\overline{x+iy} \equiv x-iy$ (for real x and y).) This time $\|\vec{x}\|$ is positive (except when $\vec{x}=\vec{0}$, when it's zero). Therefore, $\|\vec{x}\|$ and $\mathrm{d}(\vec{x},\vec{y}) \equiv \|\vec{x}-\vec{y}\|$ have the normal properties of length and distance.

Therefore, the definition involving the complex conjugation gives a much more useful dot product than the more obvious first try. On the other hand, the new dot product is not symmetric, nor is it linear in the second variable. These properties must be given up to keep positivity. For the complex dot product we have a revised list of properties:

CI1. $\vec{x} \cdot \vec{x} > 0$ — except that $\vec{0} \cdot \vec{0} = 0$. (*positivity*)

CI2. $\vec{x} \cdot \vec{y} = \overline{\vec{y} \cdot \vec{x}}$. (*Hermitian symmetry*)

CI3. $(c\vec{x} + \vec{y}) \cdot \vec{z} = c(\vec{x} \cdot \vec{z}) + \vec{y} \cdot \vec{z}$ and hence
$\vec{x} \cdot (c\vec{y} + \vec{z}) = \overline{c}(\vec{x} \cdot \vec{y}) + \vec{x} \cdot \vec{z}$. (*sesquilinearity*)

("Sesqui-" means "$1\frac{1}{2}$ times". The inner product is linear in the first variable and almost linear in the second one.)

Inner products, etc., in general vector spaces

Just like addition and scalar multiplication, the concepts of inner product, norm, and metric can be generalized from $\mathbf{R}^n$ to other spaces.

Definition: A *metric* (or *distance function*) is any real-valued function, $\mathrm{d}(x, y)$, on a set of any kind, that satisfies M1–M3.

Definition: A *norm* is any real-valued function, $\|\vec{x}\|$, on a vector space, that satisfies N1–N3.

Every norm defines a metric by

$$\mathrm{d}(\vec{x}, \vec{y}) \equiv \|\vec{x} - \vec{y}\|,$$

but there are metrics that do not come from norms.

Definition: An *inner product* on a vector space is any scalar-valued function of two vector variables that satisfies RI1–RI3 if the vector space is real, or CI1–CI3 if the space is complex. The notation $\langle \vec{x}, \vec{y} \rangle$ (instead of $\vec{x} \cdot \vec{y}$) is often used for inner products to emphasize that something other than the usual dot product may be involved.

Every inner product defines a norm by

$$\|\vec{x}\| \equiv \sqrt{\langle \vec{x}, \vec{x} \rangle}\,,$$

but there are norms which do not come from inner products — for example,

$$\|\vec{x}\|_1 \equiv \sum_{j=1}^{n} |x_j|$$

on the vector space $\mathbf{R}^n$. In Sec. 3.3 we had occasion to deal in passing with a "length function" that we designated $|\vec{x}|$, which is actually another (non-inner-product) norm with the official notation

$$\|\vec{x}\|_\infty \equiv \max_{1\le j\le n} |x_j|.$$

In $\mathbf{R}^n$ (but *not in infinite-dimensional vector spaces*) all norms are "equivalent". This does not mean that they are numerically equal ($\|\vec{x}\|_1 = \|\vec{x}\|_\infty = \|\vec{x}\|$) — that is obviously false — but rather means that they define the same concepts of "limit", "convergence", and "continuity", because when one norm is small the others must also be small. We showed this in Sec. 3.3 for $\|\vec{x}\|_\infty$ and $\|\vec{x}\|$; for $\|\vec{x}\|_1$, see Exercise 6.1.10.

NOTE: On an abstract vector space the inner product (if there is one at all) is **not** *defined* by a formula like

$$\vec{x}\cdot\vec{y} \equiv \sum_{j=1}^{n} x_j y_j\,.$$

Trying to use such a formula to prove theorems about abstract inner products is missing the point! Such a formula does not make any sense until a basis is introduced into the space. For a good choice of basis, the above formula *may* turn out to be true, but it is not fundamental. The inner product is just a function that takes pairs of vectors into numbers. (Even on $\mathbf{R}^n$ one sometimes finds it useful to consider some inner product other than the usual one, as we shall see later.)

Complex inner products are extremely important in quantum mechanics. Unfortunately, there the inner product is always defined to be linear in the *right* variable, with the complex conjugation on the *left*:

$$\langle \vec{x}, c\vec{y}\rangle = c\langle \vec{x}, \vec{y}\rangle = \langle \overline{c}\,\vec{x}, \vec{y}\rangle;$$

for example, in $\mathbf{C}^n$

$$\vec{x}\cdot\vec{y} \equiv \sum_{j=1}^{n} \overline{x_j}\, y_j$$

unlike the definition stated above (the one standard among mathematicians).

Examples of inner products *other* than the usual dot product are easy to construct on finite-dimensional vector spaces (see below); but the main reason for considering inner products in general is that *infinite-dimensional* vector spaces have useful inner products. For example, if f and g are functions in $\mathcal{C}(-1,1)$ (the space of continuous functions — let's say complex-valued ones — on the interval $-1 < t < 1$), then one defines

$$\langle f, g \rangle \equiv \int_{-1}^{1} f(t)\,\overline{g(t)}\,dt$$

— so that

$$\|f\|^2 = \int_{-1}^{1} |f(t)|^2\,dt.$$

This inner product is used in connection with Fourier series (cf. Secs. 4.3 and 5.3).

Another example: Let $\mathcal{P}$ be the space of all (real-valued) polynomials, restricted as functions to the domain $0 \leq t < \infty$. The inner product definition

$$\langle f, g \rangle = \int_0^\infty f(t)\,g(t)\,dt \qquad\qquad * * * \textbf{ WRONG } * * *$$

is no good — the integral will never converge. But the inner product

$$\langle f, g \rangle \equiv \int_0^\infty f(t)\,g(t)\,te^{-t}\,dt$$

is well-defined and useful — see Exercise 6.2.5.

One might wonder whether inner products on finite-dimensional spaces other than the standard dot product have any real use. Are they just toy examples to prepare the way for infinite-dimensional problems? We shall see that "strange" inner products can indeed arise naturally, but also that they are always in some sense equivalent to the dot product.

As a first example, let's use the notation

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbf{R}^2 \quad \text{(with } \vec{y} \text{ similar)}, \qquad A = \begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix},$$

and consider the definition

$$\langle \vec{x}, \vec{y} \rangle = \vec{x} \cdot (A\vec{y}) = 5x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2\,.$$

It is easy to see that this function is bilinear and symmetric. It is true, but less obvious, that it satisfies the positivity property, RI1. (Note, in particular, that the presence of minus signs in the off-diagonal terms does not necessarily spoil positivity, nor does their absence guarantee it.) A systematic way of testing such an expression for positivity will emerge from our study of eigenvectors in a later chapter; we shall see in a moment by an indirect method that this particular example is positive.

Therefore, we have constructed an inner product. It might be useful in a problem involving the matrix A. More likely, however, it might arise from the ordinary dot product by a change of basis, as we shall now see.

We turn to what at first sight appears to be a completely new example problem. Consider the basis

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad \vec{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

for $\mathbf{R}^2$, and use the notation

$$\vec{x} = r_1\vec{v}_1 + r_2\vec{v}_2\,, \qquad \vec{y} = s_1\vec{v}_1 + s_2\vec{v}_2\,.$$

The usual inner product in $\mathbf{R}^2$ is $\vec{x}\cdot\vec{y} = x_1y_1 + x_2y_2$. The problem is to calculate $\vec{x}\cdot\vec{y}$ in terms of the r and s coefficients. The most efficient way to do this is to start by computing the dot products of the basis vectors:

$$\vec{v}_1\cdot\vec{v}_1 = 5, \qquad \vec{v}_1\cdot\vec{v}_2 = \vec{v}_2\cdot\vec{v}_1 = -1, \qquad \vec{v}_2\cdot\vec{v}_2 = 2.$$

Then by bilinearity we calculate

$$\begin{aligned}(r_1\vec{v}_1 + r_2\vec{v}_2)\cdot(s_1\vec{v}_1 + s_2\vec{v}_2) &= 5r_1s_1 - r_1s_2 - r_2s_1 + 2r_2s_2 \\ &= (r_1, r_2)\begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \\ &= \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}\cdot\begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} s_1 \\ s_2 \end{pmatrix}.\end{aligned}$$

(In passing, we are pointing out here that the dot product with a column vector is the same thing as the ordinary matrix product with a row vector on the left.)

The result of the calculation is what would be called $\langle\vec{r},\vec{s}\rangle$ in the notation of the previous example. So, that strange inner product is simply the standard inner product in disguise. Insofar as the new basis $\{\vec{v}_1, \vec{v}_2\}$ is a useful one to work in, the formula for the dot product in that representation of $\mathbf{R}^2$ will be useful. On the other hand, often one works in the opposite direction: starting with an arbitrary inner product, one constructs a basis (called an *orthonormal basis*) in which it will look like the standard dot product. This will be the subject of our next section.

Elementary theorems about inner products

All inner products satisfy

Theorem (Cauchy–Schwarz Inequality):

$$|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \, \|\vec{y}\|,$$

with equality if and only if $\vec{x}$ and $\vec{y}$ are dependent (parallel).

Proof: We consider the case of a real vector space; the proof for the complex case is similar but more complicated. Since the inner product has the positivity property RI1, for any real number r we must have

$$\begin{aligned} 0 &\leq \langle (\vec{x} - r\vec{y}), (\vec{x} - r\vec{y}) \rangle \\ &= \|\vec{x}\|^2 - 2r\langle \vec{x}, \vec{y} \rangle + r^2 \|\vec{y}\|^2, \end{aligned}$$

with equality only when $\vec{x} = r\vec{y}$. As a function of r, therefore, the right-hand side of this inequality has no real roots when $\vec{x}$ and $\vec{y}$ are independent. (Its graph is a parabola that lies entirely above the horizontal axis.) Consequently, its discriminant ($b^2 - 4ac$ in the traditional notation) must be negative. That is,

$$\langle \vec{x}, \vec{y} \rangle^2 < \|\vec{x}\|^2 \|\vec{y}\|^2.$$

This is equivalent to the inequality to be proved. On the other hand, when the vectors *are* dependent ($\vec{x} = r\vec{y}$ for a particular r), it is easy to see that

$$\langle \vec{x}, \vec{y} \rangle^2 = r^2 \|\vec{y}\|^2 = \|\vec{x}\|^2 \|\vec{y}\|^2.$$

Another important property is

Theorem: All norms satisfy the inequality

$$\big| \|\vec{x}\| - \|\vec{y}\| \big| \leq \|\vec{x} - \vec{y}\|.$$

Special case: Absolute values of numbers satisfy

$$\big| |a| - |b| \big| \leq |a - b|.$$

Proof: Write the triangle property (N3) as

$$\|\vec{u} + \vec{w}\| \leq \|\vec{u}\| + \|\vec{w}\|.$$

Let $\vec{u} = \vec{x} - \vec{y}$, $\vec{w} = \vec{y}$:

$$\|\vec{x}\| \leq \|\vec{x} - \vec{y}\| + \|\vec{y}\|.$$

Therefore,

$$\|\vec{x}\| - \|\vec{y}\| \leq \|\vec{x} - \vec{y}\|.$$

Interchanging the roles of $\vec{x}$ and $\vec{y}$, we see that also

$$\|\vec{y}\| - \|\vec{x}\| \leq \|\vec{x} - \vec{y}\|.$$

The two inequalities together are equivalent to the theorem to be proved.

Exercises

6.1.1 Which of these expressions defines an inner product on $\mathbf{R}^2$? Explain what's wrong with each of the others. Here $\vec{x}_1 = (x_1, y_1)$ and $\vec{x}_2 = (x_2, y_2)$ are vectors in $\mathbf{R}^2$.

(a) $x_1 x_2 - y_1 y_2$

(b) $10 x_1 x_2 + 20 y_1 y_2$

(c) $x_1 y_2$

(d) ${x_1}^2 {x_2}^2 + {y_1}^2 {y_2}^2$

6.1.2 Reinterpret Exercise 6.1.1 in the alternative notation, $\vec{x} = (x_1, x_2)$ and $\vec{y} = (y_1, y_2)$. Which of the expressions (if any) are inner products now?

6.1.3 Prove that the norm

$$\|\vec{x}\| \equiv \sqrt{\langle \vec{x}, \vec{x} \rangle}$$

defined by any inner product does indeed satisfy the triangle inequality, N3.

6.1.4 With $\vec{x} = (x_1, x_2, x_3)$ and $\vec{y} = (y_1, y_2, y_3)$, explain whether or not each of these satisfies the definition of an inner product on $\mathbf{R}^3$.

(a) $\langle \vec{x}, \vec{y} \rangle = x_1 y_1 + x_1 y_2 + x_2 y_1 + 1.1 x_2 y_2 + 25 x_3 y_3$

(b) $\langle \vec{x}, \vec{y} \rangle = {x_1}^2 {y_1}^2 + {x_2}^2 {y_2}^2 + {x_3}^2 {y_3}^2$

6.1.5 Tell whether each of these formulas defines an inner product on the function space $\mathcal{C}(0, 2)$. (Explain your answers briefly.) The scalars in this problem are real, not complex.

(a) $\langle f, g\rangle = \int_0^2 f(t)\, g(t)\, (t^2+1)\, dt$

(b) $\langle f, g\rangle = \int_0^2 f(t)\, g(t)\, (t-1)\, dt$

(c) $\langle f, g\rangle = \int_0^2 f(t)^2\, g(t)^4\, dt$

6.1.6 Tell whether each of these formulas defines an inner product on the subspace of bounded functions in the space $\mathcal{C}(0,\infty)$. (Explain your answers briefly.) The scalars in this problem are real, not complex. The boundedness condition should help you decide whether the integrals converge (cf. Exercise 5.3.1).

(a) $\langle f, g\rangle = \int_0^\infty f(t)\, g(t)\, \frac{1}{t^2+1}\, dt$

(b) $\langle f, g\rangle = \int_0^\infty f(t)\, g(t)\, (t^2+1)\, dt$

(c) $\langle f, g\rangle = \int_0^2 f(t)\, g(t)\, dt$

6.1.7 Define the "superdistance" between any two points in $\mathbf{R}^2$ to be the square of the usual distance.

(a) Show that the superdistance satisfies the three properties in the definition of a metric.

(b) Show that the superdistance does not arise from any norm by the formula

$$\mathrm{d}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|.$$

Hint: Show that N2 must be violated.

6.1.8 Prove that if $\|\cdots\|$ is indeed a norm, then

$$\mathrm{d}(\vec{x}, \vec{y}) \equiv \|\vec{x} - \vec{y}\|$$

is a metric.

6.1.9 Use the Cauchy–Schwarz inequality to give an alternative proof of Exercise 3.3.2, working entirely with the usual norm $\|\vec{x}\|$ instead of introducing the alternative norm $|\vec{x}|$.

6.1.10 Show that the norm $\|\vec{x}\|_1$ on $\mathbf{R}^n$ is *equivalent* to the usual norm $\|\vec{x}\|$, in the sense that there are numbers C_1 and C_2 such that

$$\|\vec{x}\|_1 \le C_1\|\vec{x}\| \quad \text{and} \quad \|\vec{x}\| \le C_2\|\vec{x}\|_1 \quad \text{for all } \vec{x} \in \mathbf{R}^n.$$

(C_1 and C_2 are independent of $\vec{x}$ but may depend on n.)

6.2 Orthogonality

For two vectors, the opposite extreme from being parallel (dependent) is being perpendicular. Unlike parallelism, perpendicularity requires an inner product for its definition.

Definitions:

1. $\vec{x}$ and $\vec{y}$ are *orthogonal* (written $\vec{x} \perp \vec{y}$) if $\vec{x}\cdot\vec{y} = 0$. (The words *perpendicular* and *normal* are also used, especially in the most geometrical contexts.)
2. A *set* $\{\vec{x}_1, \ldots, \vec{x}_l\}$ is *orthogonal* (abbreviated *OG*) if

$$\vec{x}_j \cdot \vec{x}_k = 0 \quad \text{whenever } j \ne k.$$

3. A set is *orthonormal* (abbreviated *ON*) if it is orthogonal and also every $\vec{x}_j$ in it is a unit vector. That is,

$$\vec{x}_j \cdot \vec{x}_k = \delta_{jk} \equiv \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \ne k. \end{cases}$$

EXAMPLE: The natural basis in $\mathbf{R}^n$ is ON:

$$\hat{e}_j \cdot \hat{e}_k = \delta_{jk}\,.$$

As in this example, we shall usually indicate unit vectors, especially those in ON bases, with "hats" instead of arrows.

Theorem: An orthogonal set that does not contain the zero vector is linearly independent. Hence an orthogonal set in $\mathbf{R}^n$ with n elements (all nonzero) is a basis for $\mathbf{R}^n$, and similarly for any n-dimensional space.

PROOF: If $\sum_j c_j \vec{x}_j = \vec{0}$, then for each k,

$$0 = \left(\sum_j c_j \vec{x}_j \right) \cdot \vec{x}_k = \sum_j c_j (\vec{x}_j \cdot \vec{x}_k) = c_k \|\vec{x}_k\|^2.$$

Thus $c_k = 0$, since $\|\vec{x}_k\| \neq 0$.

Here is the main theorem about orthogonal bases:

Theorem: If $\{\vec{u}_1, \ldots, \vec{u}_n\}$ is an orthogonal basis, then the coordinates of an arbitrary vector, $\vec{x} = \sum_{j=1}^n c_j \vec{u}_j$, are

$$c_j = \frac{\vec{x} \cdot \vec{u}_j}{\|\vec{u}_j\|^2} \,.$$

In particular, if the basis is orthonormal, then

$$c_j = \vec{x} \cdot \hat{u}_j \,.$$

REMARKS: In a complex vector space with the "physics convention", $\vec{x} \cdot \vec{u}_j$ in these formulas must be replaced by $\vec{u}_j \cdot \vec{x}$. (Think of the case of $\mathbf{C}^n$ with the usual inner product: It is the coordinates of $\vec{u}_j$ that get conjugated.) In this discussion we are using the elementary dot-product notation, but the theorem is valid for any inner product on any vector space; in the angular-bracket notation, of course, $\vec{x} \cdot \vec{u}_j$ is replaced by $\langle \vec{x}, \vec{u}_j \rangle$.

The proof of this theorem is essentially contained in that of the preceding one. (Change xs to us and start with the sum equal to $\vec{x}$ instead of $\vec{0}$.)

The theorem tells us that orthogonal bases are very easy to calculate with. To expand a vector with respect to the basis, we don't need to solve a system of linear equations, as we have been doing heretofore; we need only do the arithmetic to evaluate the inner products $\vec{x} \cdot \vec{u}_j$.

Example: $\hat{u}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{pmatrix}, \quad \hat{u}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-i}{\sqrt{2}} \end{pmatrix}$ is an ON basis for $\mathbf{C}^2$.

Let

$$\vec{x} = \begin{pmatrix} 5i + 3 \\ 6 \end{pmatrix} = c_1 \hat{u}_1 + c_2 \hat{u}_2 \,.$$

Find c_1 and c_2 .

SOLUTION:

$$c_1 = \vec{x} \cdot \hat{u}_1 = (5i+3)\frac{1}{\sqrt{2}} - \frac{6i}{\sqrt{2}} = \frac{1}{\sqrt{2}}(3-i).$$

$$c_2 = \vec{x} \cdot \hat{u}_2 = (5i+3)\frac{1}{\sqrt{2}} + \frac{6i}{\sqrt{2}} = \frac{1}{\sqrt{2}}(3+11i).$$

CHECK: $$\frac{1}{\sqrt{2}}(3-i) \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} + \frac{1}{\sqrt{2}}(3+11i) \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} 6+10i \\ 1+11 \end{pmatrix} = \begin{pmatrix} 5i+3 \\ 6 \end{pmatrix} = \vec{x}.$$

The formula for the coefficients in a Fourier series is an infinite-dimensional analogue of the theorem. In Sec. 4.3 we stated without proof that every reasonably nice function on the domain $[0, \pi]$ has a convergent Fourier series,

$$f(x) = \sum_{n=1}^{\infty} c_n \sin(nx).$$

One can easily calculate that (for integers n and m)

$$\int_0^{\pi} \sin(nx) \sin(mx)\, dx = 0 \quad \text{if } m \neq n$$

and

$$\int_0^{\pi} \sin^2(nx)\, dx = \frac{\pi}{2}.$$

This means that the set $\{\sin(nx)\}$ $(n = 1, \ldots, \infty)$ is orthogonal (but not orthonormal) with respect to the inner product

$$\langle f, g \rangle \equiv \int_0^{\infty} f(x)\, \overline{g(x)}\, dx$$

defined on the "nice" functions with domain $[0, \pi]$. From the main theorem, therefore,

$$c_n = \frac{2}{\pi} \int_0^{\pi} f(x) \sin(nx)\, dx$$

is the formula for the Fourier coefficients.

The following is an immediate corollary of the main theorem above:

Theorem: Let $\{\hat{u}_1, \ldots, \hat{u}_n\}$ be an *orthonormal* basis (for a vector space $\mathcal{V}$ equipped with an inner product $\langle \cdot, \cdot \rangle$). Let M be the matrix representing a linear function $L\colon \mathcal{V} \to \mathcal{V}$ with respect to that basis. Then

$$M_{jk} = \langle L\hat{u}_k, \hat{u}_j \rangle.$$

REMARKS: If the vector space is real, or if it is complex and the "physics notation" is used, then this formula can be written in the less awkward form

$$M_{jk} = \langle \hat{u}_j, L\hat{u}_k \rangle.$$

Often this is condensed to $M_{jk} = \langle j|L|k \rangle$ when there is no ambiguity about which ON basis is in use. This notation, which emphasizes the symmetry between the left and right indices of the matrix, is standard in quantum physics and has crept into other subjects.

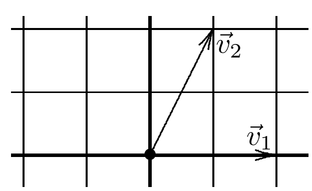
THE GRAM–SCHMIDT PROCESS

Theorem: Any linearly independent set can be replaced by an ON set with the same span. In particular, any basis can be replaced by an ON basis.

Algorithm: Let's work through this for an example. Suppose we have the two vectors in $\mathbf{R}^2$ drawn in the picture:

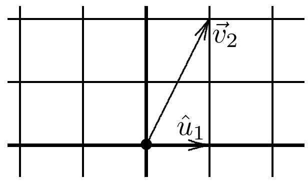
$$\vec{v}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \qquad \vec{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

This basis is very far from being orthonormal: The vectors are not orthogonal to each other, nor are they of unit length.



STEP 1: Pick one of the vectors and replace it by the unit vector in the same direction.

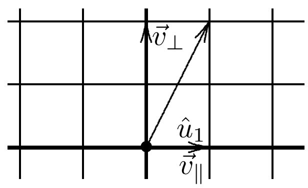
$$\hat{u}_1 \equiv \frac{\vec{v}_1}{\|\vec{v}_1\|} = \frac{1}{\sqrt{2^2 + 0^2}} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$



STEP 2, SUBSTEP A: Split the remaining vector into components parallel and perpendicular to $\hat{u}_1$.

$$\vec{v}_2 = \vec{v}_{\parallel} + \vec{v}_{\perp} .$$

Replace $\vec{v}_2$ by $\vec{v}_{\perp}$.



How can we calculate these components? Well, we want $\vec{v}_{\parallel}$ to be a multiple of $\hat{u}_1$, and we also want the remainder, $\vec{v}_{\perp} \equiv \vec{v}_2 - \vec{v}_{\parallel}$, to be orthogonal to $\hat{u}_1$. We claim that the unique answer is

$$\vec{v}_{\parallel} \equiv (\vec{v}_2 \cdot \hat{u}_1)\hat{u}_1 .$$

This obviously satisfies the first condition; let's verify the second:

$$\begin{aligned} &\left[\vec{v}_2 - (\vec{v}_2 \cdot \hat{u}_1)\hat{u}_1\right] \cdot \hat{u}_1 \\ &\quad = (\vec{v}_2 \cdot \hat{u}_1) - (\vec{v}_2 \cdot \hat{u}_1) = 0. \end{aligned}$$

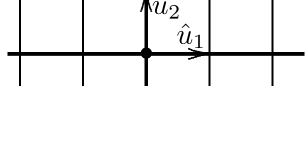
In our numerical example,

$$\vec{v}_{\perp} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - [1 \times 1 + 2 \times 0] \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} .$$

STEP 2, SUBSTEP B: Replace $\vec{v}_{\perp}$ by the unit vector in its direction, $\hat{u}_2$. In the example,

$$\hat{u}_2 \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix} .$$

(We have ended up with the natural basis, but of course that is an accidental consequence of the fact that $\hat{u}_1 = \vec{e}_1$ in this example. In general the basis would be rotated.)



If we had more than two vectors in the original set, we would now have to break the third vector up into a part "parallel" to the first two and a part perpendicular to the first two; normalize it to unit length; and continue in this way until we used up all the vectors in the original set. Let's be precise about how this works. Suppose we have a list of vectors, $\{\vec{v}_1, \dots, \vec{v}_n\}$, which are independent but not orthogonal. (As an exercise, think about what will happen if they are not independent.) To simplify notation, let's assume that in the Gram–Schmidt process we will go to work on the vectors in the order in which they are originally listed. Thus STEP 1 is to replace $\vec{v}_1$ by

$$\hat{u}_1 \equiv \frac{\vec{v}_1}{\|\vec{v}_1\|} \,;$$

and at STEP $k+1$ we will have already replaced $\{\vec{v}_1, \dots, \vec{v}_k\}$ by an ON set, $\{\hat{u}_1, \dots, \hat{u}_k\}$, whose span is the same as the span of $\{\vec{v}_1, \dots, \vec{v}_k\}$, and our problem is to construct $\hat{u}_{k+1}$.

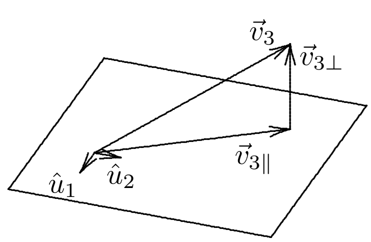
SUBSTEP A: Split $\vec{v}_{k+1}$ into parallel and perpendicular parts:

$$\vec{v}_{k+1} = \vec{v}_{k+1\,\|} + \vec{v}_{k+1\,\perp} \,.$$

The "parallel part" must lie in the span of $\{\hat{u}_1, \dots, \hat{u}_k\}$ and must also be chosen so that the remainder, $\vec{v}_{k+1\,\perp}$, is perpendicular to that subspace. The formula for the parallel part is

$$\vec{v}_{k+1\,\|} \equiv \sum_{j=1}^{k} (\vec{v}_{k+1} \cdot \hat{u}_j)\hat{u}_j \,.$$

Using the orthonormality of the $\hat{u}$s, it is easy to see that $\vec{v}_{k+1\,\perp} \cdot \hat{u}_j = 0$ for $j = 1, \dots, k$, as required. (Note, incidentally, that the right-hand side of the formula is the same expression that the main theorem above would give for $\vec{v}_{k+1}$ itself, if $\vec{v}_{k+1}$ were a member of the subspace spanned by the $\hat{u}$s. In that case, $\vec{v}_{k+1\,\perp}$ would be 0.)



Therefore, the formula for the perpendicular part (which is what we are most interested in) is

$$\vec{v}_{k+1\,\perp} = \vec{v}_{k+1} - \sum_{j=1}^{k} (\vec{v}_{k+1} \cdot \hat{u}_j)\hat{u}_j \,.$$

SUBSTEP B: Replace $\vec{v}_{k+1\,\perp}$ by its unit vector,

$$\hat{u}_{k+1} \equiv \frac{\vec{v}_{k+1\,\perp}}{\|\vec{v}_{k+1\,\perp}\|} \,.$$

Corollary (the Projection Theorem) and Definitions: Let $\mathcal{V}$ be a vector space with an inner product. If $\mathcal{W}$ is a finite-dimensional subspace of $\mathcal{V}$, then every $\vec{x} \in \mathcal{V}$ is the sum of a vector $\vec{x}_{\|}$ in $\mathcal{W}$ and a vector $\vec{x}_{\perp}$ orthogonal to $\mathcal{W}$. (Future quantum-mechanicians might note that the theorem is true for infinite-dimensional subspaces satisfying a technical condition beyond the scope of this course.) $\vec{x}_{\|}$ is called the *[orthogonal] projection* of $\vec{x}$ onto the subspace. It is calculated, given an ON basis for the subspace, by

$$\vec{x}_{\|} = \sum_{j=1}^{k} (\vec{x} \cdot \hat{u}_j)\hat{u}_j \,.$$

(If $\vec{x}$ happens to lie in $\mathcal{W}$, then this is a formula for $\vec{x}$ itself — the same one given by the main theorem above.) In this situation one says that $\mathcal{V}$ is the *orthogonal direct sum* of $\mathcal{W}$ and $\mathcal{W}^{\perp}$ (the *orthogonal complement* of $\mathcal{W}$, consisting of all vectors which are orthogonal to all elements of $\mathcal{W}$): $\mathcal{V} = \mathcal{W} \oplus \mathcal{W}^{\perp}$.

The Gram–Schmidt process can be applied to an infinite sequence of vectors. For example, the powers $\{1, t, t^2, t^3, \dots\}$ form a basis for the space of polynomials, $\mathcal{P}$. Regard the polynomials as functions on the interval $-1 \le t \le 1$, and equip the space with the inner product

$$\langle f, g \rangle \equiv \int_{-1}^{1} f(t)\, g(t)\, dt.$$

Then the power basis is not ON, but we can apply the GS procedure to it to get an ON basis. (In comparing with the general algorithm stated above, bear in mind that in this example we start the indexing with 0, not 1, so that the index will match the degree of the polynomial.)

Note first that

$$\|1\|^2 = \langle 1, 1\rangle = \int_{-1}^{1} 1\, dt = 2.$$

Therefore, the normalized vector in this first direction is $\hat{u}_0 = \sqrt{\frac{1}{2}}$. Project* the next function, t, onto this:

$$\langle t, \hat{u}_0\rangle = \sqrt{\frac{1}{2}} \int_{-1}^{1} t\, dt = 0.$$

Thus $t_{\|} = 0$, and $t_{\perp} = t$. This must be normalized:

$$\langle t, t\rangle = \int_{-1}^{1} t^2\, dt = \frac{2}{3}\,.$$

Thus $\hat{u}_1 = \sqrt{\frac{3}{2}}\, t$. Now project t^2 onto the plane spanned by $\{\hat{u}_0, \hat{u}_1\}$:

$$\langle t^2, \hat{u}_0\rangle = \sqrt{\frac{1}{2}} \int_{-1}^{1} t^2\, dt = \frac{2}{3}\sqrt{\frac{1}{2}}\,;$$

$$\langle t^2, \hat{u}_1\rangle = \sqrt{\frac{3}{2}} \int_{-1}^{1} t^3\, dt = 0.$$

Hence

$$(t^2)_{\|} = \langle t^2, \hat{u}_0\rangle \hat{u}_0 + \langle t^2, \hat{u}_1\rangle \hat{u}_1 = \frac{1}{3}\,;$$

$$(t^2)_{\perp} = t^2 - \frac{1}{3}\,.$$

Then

$$\|(t^2)_{\perp}\|^2 = \int_{-1}^{1} \left(t^4 - \frac{2}{3}\, t^2 + \frac{1}{9}\right) dt = \frac{2}{5} - \frac{4}{9} + \frac{2}{9} = \frac{8}{45}\,,$$

so $\hat{u}_2 = \sqrt{\frac{45}{8}}(t^2 - \frac{1}{3})$.

---

* The number $\langle t, \hat{u}_0\rangle$ is sometimes called the "scalar projection" of the vector t onto the vector $\hat{u}_0$. The resulting "vector projection" is this number times the vector $\hat{u}_0$. In this problem, where the vectors are functions and $\hat{u}_0$ is a *constant* function, the vector projection, $t_{\|}$, is just another number, smaller than the scalar projection by another factor $\frac{1}{\sqrt{2}}$. In the text this step is not written out, because it is immediately obvious that the vector will be zero.

In summary, the results of the first few steps are

$$\hat{u}_0 = \sqrt{\frac{1}{2}}, \qquad \hat{u}_1 = \sqrt{\frac{3}{2}}\,t, \qquad \hat{u}_2 = \sqrt{\frac{45}{8}}\left(t^2 - \frac{1}{3}\right).$$

One can show that

$$\hat{u}_n(t) = \sqrt{\frac{2n+1}{2}}\,\frac{1}{2^n n!}\,\frac{d^n}{dt^n}(t^2-1)^n.$$

(This is called *Rodrigues's formula*. Memorizing it is not encouraged; it's here only to point out that the entire infinite sequence of Gram–Schmidt steps can be solved at once if one works hard enough.)

The sequence of polynomials constructed in this way is called the *normalized Legendre polynomials*. The function $\hat{u}_n$ satisfies the differential equation

$$\left(1-t^2\right)\frac{d^2u}{dt^2} - 2t\,\frac{du}{dt} + n(n+1)u = 0,$$

which is very important in applied mathematics. (It arises in solving partial differential equations in spherical coordinates.) More precisely, the multiples of $\hat{u}_n$ are the only solutions of this *Legendre's equation* which do not blow up at the endpoints, $t = \pm 1$. (If n is not an integer, there are no such solutions at all.)

In fact, many important second-order differential equations have polynomial solutions which are orthogonal with respect to *some* inner product. Famous examples are named after Hermite (Exercise 6.2.6) and Laguerre (Example 4 and Exercise 6.2.5). Both of these are important in quantum mechanics: The solutions of the quantum theory of a harmonic oscillator are constructed from the Hermite polynomials, while the Laguerre polynomials show up in the theory of the hydrogen atom.

Gram–Schmidt examples

Example 1. Construct an ON basis for $\mathbf{R}^3$ that includes one vector parallel to $(1,1,1)$.

Solution: Apply Gram–Schmidt to any sequence of vectors with $\vec{v}_1 = (1,1,1)$. (If we have the bad luck to choose a vector that is linearly dependent on the vectors preceding it, we will simply waste a Gram–Schmidt step; see Exercise 6.2.15.) The first unit vector is

$$\hat{u}_1 = \frac{1}{\sqrt{3}}(1,1,1).$$

For lack of anything better, we take $\vec{v}_2 = (1,0,0)$, $\vec{v}_3 = (0,1,0)$. Then

$$\vec{v}_{2\parallel} = (\vec{v}_2 \cdot \hat{u}_1)\hat{u}_1 = \frac{1}{3}(1,1,1),$$

$$\vec{v}_{2\perp} = \frac{1}{3}(2,-1,-1);$$

$$\hat{u}_2 = \frac{1}{\sqrt{6}}(2,-1,-1).$$

(Note, incidentally, that the factor $\frac{1}{3}$ can be ignored in calculating the coefficient $\frac{1}{\sqrt{6}}$; it is guaranteed to cancel.) Continue to the next step:

$$\begin{aligned}\vec{v}_{3\parallel} &= (\vec{v}_3 \cdot \hat{u}_1)\hat{u}_1 + (\vec{v}_3 \cdot \hat{u}_2)\hat{u}_2 \\ &= \frac{1}{3}(1,1,1) - \frac{1}{6}(2,-1,-1) \\ &= \frac{1}{2}(0,1,1);\end{aligned}$$

$$\vec{v}_{3\perp} = \frac{1}{2}(0,1,-1);$$

$$\hat{u}_3 = \frac{1}{\sqrt{2}}(0,1,-1).$$

It is easy to check that $\{\hat{u}_j\}$ $(j = 1,2,3)$ is orthonormal.

Example 2. Find an orthonormal basis for the inner product in $\mathbf{R}^2$ used as an example in the previous section,

$$\langle \vec{x}, \vec{y} \rangle = 5x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2\,. \qquad (*)$$

Solution: The easy way is to know the answer beforehand: By the second part of the example in the previous section, we can take the image of the natural basis under the mapping

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$$

where

$$\vec{x} = r_1\vec{v}_1 + r_2\vec{v}_2\,, \qquad \vec{v}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

From previous discussions of basis changes, we know that the matrix of this mapping is the inverse of the matrix whose columns are the $\vec{v}_j$:

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}.$$

The columns of this matrix,

$$\hat{u}_1 = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad \hat{u}_2 = \frac{1}{3}\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

ought to be orthonormal in the inner product $(*)$ (*not* in the usual dot product). Check:

$$\begin{aligned} \langle \hat{u}_1, \hat{u}_1 \rangle &= \frac{1}{9}(5 - 2 - 2 + 8) = 1, \\ \langle \hat{u}_1, \hat{u}_2 \rangle &= \frac{1}{9}(5 - 2 + 1 - 4) = 0, \\ \langle \hat{u}_2, \hat{u}_2 \rangle &= \frac{1}{9}(5 + 1 + 1 + 2) = 1. \end{aligned}$$

Alternative solution: We can construct a basis for this example by the Gram–Schmidt method. The result will, of course, depend on the vectors chosen as input for the algorithm. If we pick $\vec{v}_1$ (in the notation of our general Gram–Schmidt discussion, not that of this example) to be $\binom{1}{0}$, we get

$$\|\vec{v}_1\|^2 = 5 - 0 - 0 + 0 = 5,$$

so

$$\hat{u}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Then if we pick $\vec{v}_2 = \binom{0}{1}$, we have

$$\vec{v}_{2\parallel} = \frac{1}{5}(0 - 1 - 0 + 0)\begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$\vec{v}_{2\perp} = \frac{1}{5}\begin{pmatrix} 1 \\ 5 \end{pmatrix},$$

$$\begin{aligned} \vec{u}_2 &= (5 - 5 - 5 + 50)^{-1/2}\begin{pmatrix} 1 \\ 5 \end{pmatrix} \\ &= \frac{1}{3\sqrt{5}}\begin{pmatrix} 1 \\ 5 \end{pmatrix}. \end{aligned}$$

You can check that $\hat{u}_1$ and $\hat{u}_2$ are orthonormal (with respect to the strange inner product).

Example 3. Suppose that the Rapidrudder polynomials are the normalized orthogonal polynomials defined by applying the Gram–Schmidt algorithm to the power functions, with respect to the inner product

$$\langle p, q \rangle = \int_1^4 p(t)q(t)\,dt.$$

Find the first two Rapidrudder polynomials.

SOLUTION: Call the orthonormal polynomials $\hat{u}_0(t)$, $\hat{u}_1(t)$, ... We calculate

$$\|1\|^2 = \int_1^4 1\,dt = 3,$$

so the first polynomial is

$$\hat{u}_0(t) = \frac{1}{\sqrt{3}}\,.$$

The projection of the next power function onto this is

$$t_\| = \langle \hat{u}_0, t \rangle \hat{u}_0 = \frac{1}{3}\int_1^4 t\,dt = \frac{1}{6}(16-1) = \frac{5}{2}\,.$$

Therefore,

$$t_\perp = t - \frac{5}{2}\,.$$

Then

$$\begin{aligned}\|t_\perp\|^2 &= \int_1^4 \left(t^2 - 5t + \frac{25}{4}\right)\,dt \\ &= \left[\frac{t^3}{3} - \frac{5t^2}{2} + \frac{25t}{4}\right]_1^4 \\ &= \frac{63 \cdot 4 - 75 \cdot 6 + 75 \cdot 3}{12} = \frac{27}{12} = \frac{9}{4}\,.\end{aligned}$$

Thus

$$\hat{u}_1(t) = \frac{2}{3}\left(t - \frac{5}{2}\right).$$

Example 4. The formula $\langle p, q \rangle = \int_0^\infty p(t)q(t)\,e^{-t}\,dt$ defines an inner product on the vector space of polynomials. Find the first three of the orthonormal polynomials associated with this inner product.

SOLUTION: Since we will be applying the Gram–Schmidt algorithm to the power functions, the formula

$$\int_0^\infty t^n\, e^{-t}\, dt = n!$$

will be used repeatedly; this formula can be found in many handbooks (or proved by repeated integration by parts). Let $v_0 = t^0$, etc., and let $\hat{u}_0$, etc., be the resulting orthogonal polynomials.

Step 1: We have $\|v_0\|^2 = \int_0^\infty e^{-t}\, dt = 0! = 1$. Thus

$$\hat{u}_0 = v_0 = 1$$

(that is, the *function* whose value is 1 at every value of t).

Step 2: We find $\langle \hat{u}_0, v_1\rangle = \int_0^\infty t\, e^{-t}\, dt = 1! = 1$, so $v_{1\|} = \langle \hat{u}_0, v_1\rangle \hat{u}_0 = 1$. Therefore, $v_{1\perp} = v_1 - v_{1\|} = t - 1$. Then we calculate

$$\|v_{1\perp}\|^2 = \int_0^\infty (t^2 - 2t + 1)\, e^{-t}\, dt = 2! - 2 + 1 = 1,$$

and conclude

$$\hat{u}_1 = t - 1.$$

Step 3: The relevant integrals are

$$\langle \hat{u}_0, v_2\rangle = \int_0^\infty t^2 e^{-t}\, dt = 2! = 2,$$
$$\langle \hat{u}_1, v_2\rangle = \int_0^\infty (t-1) t^2 e^{-t}\, dt = 3! - 2! = 4.$$

So

$$v_{2\|} = \langle \hat{u}_0, v_2\rangle \hat{u}_0 + \langle \hat{u}_1, v_2\rangle \hat{u}_1 = 2 + 4(t-1) = 4t - 2,$$

and hence $v_{2\perp} = t^2 - 4t + 2$. Then

$$\begin{aligned}\|v_{2\perp}\|^2 &= \int_0^\infty (t^2 - 4t + 2)^2 e^{-t}\, dt \\ &= \int_0^\infty (t^4 - 8t^3 + 16t^2 + 4t^2 - 16t + 4) e^{-t}\, dt \\ &= 4! - 8 \cdot 3! + 20 \cdot 2! - 16 + 4 = 24 - 48 + 40 - 12 = 4.\end{aligned}$$

Therefore,

$$\hat{u}_2 = \frac{v_{2\perp}}{\|v_{2\perp}\|} = \frac{1}{2}\,t^2 - 2t + 1.$$

REMARK: The polynomials obtained by this procedure (continued to arbitrarily high degree) are one type of normalized *Laguerre polynomials*. Other types of Laguerre polynomials are defined by inserting an extra factor t^m in the inner product formula. The Laguerre polynomials with $m = 1$ (Exercise 6.2.5) are the most important, because the partial differential equation that arises in the quantum theory of the hydrogen atom can be solved in terms of them.

Exercises

6.2.1 Let $\vec{v}_1 = (2, 1, 2)$, $\vec{v}_2 = (1, 1, 1)$. Find an orthonormal basis for $\mathbf{R}^3$ whose first element is proportional to $\vec{v}_1$ and whose second element lies in $\operatorname{span}\{\vec{v}_1, \vec{v}_2\}$.

6.2.2 Use the Gram–Schmidt method to find an orthonormal basis for the space spanned by the vectors

$$\vec{v}_1 = (1, 0, 1), \quad \vec{v}_2 = (0, 1, 0), \quad \vec{v}_3 = (1, 1, 1), \quad \vec{v}_4 = (1, 2, 3).$$

6.2.3 Recall that the Legendre polynomials form an orthogonal, but not orthonormal, basis for the space of polynomials on the interval $-1 \le t \le 1$ with the inner product $\langle p, q\rangle = \int_{-1}^{1} p(t)q(t)\,dt$. The first three Legendre polynomials are

$$p_0(t) = 1, \qquad p_1(t) = t, \qquad p_2(t) = t^2 - \frac{1}{3}\,.$$

Their norms are

$$\|p_0\| = \sqrt{2}\,, \qquad \|p_1\| = \sqrt{\frac{2}{3}}\,, \qquad \|p_2\| = \sqrt{\frac{8}{45}}\,.$$

Let $q(t) = t^2 - 5t + 6$. Write q as a linear combination of Legendre polynomials. (Use the method demonstrating the general theory of orthogonal bases. Optionally, you may point out another way that is quicker.)

6.2.4 Define an inner product on the space of polynomials by

$$\langle f, g\rangle = \int_0^2 e^t\, f(t) g(t)\, dt.$$

Find the first three orthogonal polynomials with respect to this inner product. (Apply the Gram–Schmidt process to the standard basis, $\{1, t, t^2, \ldots\}$.) SUGGESTION: Start by finding a formula or making a table for the integrals

$$\int_0^2 e^t\, t^n\, dt$$

for at least the first few values of n.

6.2.5 Regard the polynomials in the space $\mathcal{P}_2$ as functions on the interval $0 \le t < \infty$, and equip this space with the inner product

$$\langle p, q\rangle = \int_0^\infty p(t)\, q(t)\, te^{-t}\, dt.$$

Starting from the standard basis $\{1, t, t^2\}$, find an orthonormal basis by the Gram–Schmidt method. (These are the first three Laguerre polynomials.)

6.2.6 Consider the inner product $\langle p, q\rangle \equiv \int_{-\infty}^{\infty} p(t)\, q(t)\, e^{-t^2}\, dt$ on the vector space of all polynomials. When the Gram–Schmidt procedure is applied to the standard basis $\{1, t, t^2, \ldots\} \equiv \{v_0, v_1, v_2, \ldots\}$, the elements of the resulting orthonormal basis are called *Hermite polynomials*. Find the first 3 Hermite polynomials. FREE INFORMATION:

$$\int_{-\infty}^{\infty} e^{-t^2}\, dt = \sqrt{\pi}, \quad \int_{-\infty}^{\infty} t^2\, e^{-t^2}\, dt = \frac{\sqrt{\pi}}{2}, \quad \int_{-\infty}^{\infty} t^4\, e^{-t^2}\, dt = \frac{3\sqrt{\pi}}{4}.$$

6.2.7 Construct an orthonormal basis for $\mathbf{R}^2$ including one vector proportional to

$$\vec{v}_1 \equiv \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

6.2.8 Which statements are correct? Give counterexamples to justify the statements you exclude. (The vector space involved here is real, not complex.)

$$\vec{v} \text{ is sure to equal } \sum_{j=1}^{n} \langle \vec{v}, \vec{u}_j\rangle \vec{u}_j \text{ if } \{\vec{u}_j\} \text{ are}$$

(A) orthogonal.

(B) a basis for a vector space containing $\vec{v}$.

(C) orthonormal.

(D) both (B) and (C)

6.2.9 Consider the polynomial space $\mathcal{P}_2$, equipped with the inner product

$$\langle p, q\rangle \equiv \int_0^2 p(t)\, q(t)\, dt.$$

Replace the standard basis $\{1, t, t^2\} \equiv \{v_0, v_1, v_2\}$ by an orthonormal basis, by the Gram–Schmidt procedure.

6.2.10 By the Gram–Schmidt method, find an orthonormal basis for the span of $\{\vec{v}_1 = (1,1,0,0),\ \vec{v}_2 = (0,0,1,1),\ \vec{v}_3 = (0,1,0,1)\}$.

6.2.11 Consider

$$\langle f, g\rangle \equiv \int_0^1 x\, f(x)\, g(x)\, dx, \qquad (*)$$

which is defined for f and g elements of the space of real-valued, continuous functions on the unit interval, denoted $\mathcal{C}(0,1)$. Show that $(*)$ defines an inner product on $\mathcal{C}(0,1)$.

6.2.12 Verify that the vectors $\vec{a}$ and $\vec{b}$ are orthogonal, and extend them to an orthogonal basis for $\mathbf{R}^4$. (Calculator or computer advised.)

$$\vec{a} = (1,1,1,2), \quad \vec{b} = (1,2,3,-3).$$

6.2.13 Construct an orthonormal basis for the subspace spanned by the vectors

$$\vec{x}_1 = (1,1,-1,-2), \quad \vec{x}_2 = (5,8,-2,-3), \quad \vec{x}_3 = (3,9,3,8).$$

6.2.14 Construct an orthogonal basis for the subspace spanned by the vectors

$$\vec{x}_1 = (2,1,3,-1), \quad \vec{x}_2 = (7,4,3,-3), \quad \vec{x}_3 = (1,1,-6,0).$$

6.2.15 How would you modify the Gram–Schmidt construction if one of the vectors in the original list turns out to be linearly dependent on the vectors preceding it?

6.2.16 Let $\{\hat{v}_j\}$ and $\{\hat{w}_j\}$ $(j = 1, \ldots, n)$ be two orthonormal bases for an n-dimensional real vector space, $\mathcal{V}$. Let C be the change-of-basis matrix connecting these two bases, as in Sec. 4.4. Show that C is an *orthogonal matrix*; that is, the columns of C form an *orthonormal* set in $\mathbf{R}^n$.

6.2.17 If $\mathcal{S}$ is a sub*set* of $\mathbf{R}^n$, one defines $\mathcal{S}^\perp$ to be the set of all vectors $\vec{v}$ in $\mathbf{R}^n$ such that $\vec{x} \cdot \vec{v} = 0$ for all $\vec{x} \in \mathcal{S}$. (The dot indicates the standard inner product in $\mathbf{R}^n$.)

(a) Show that $\mathcal{S}^\perp$ is always a sub*space* of $\mathbf{R}^n$ (even if $\mathcal{S}$ isn't).

(b) Show that if $\vec{x}_0 \in \mathcal{S}$, then $\vec{x}_0 \in \left(\mathcal{S}^\perp\right)^\perp$. That is, $\mathcal{S}$ is a subset of the subspace $\left(\mathcal{S}^\perp\right)^\perp$. (It may help to draw a sketch of an $\mathcal{S}$, $\mathcal{S}^\perp$, and $\left(\mathcal{S}^\perp\right)^\perp$ in $\mathbf{R}^3$.)

6.2.18 In the situation of the preceding exercise: Show that if $\mathcal{S}$ *is* a subspace, then

$$\left(\mathcal{S}^\perp\right)^\perp = \mathcal{S}.$$

HINT: Appeal to the projection theorem.

6.3 The Geometry of Curves

This section has three purposes: First, we explore the conceptual question of what it means to differentiate a vector-valued function, and the role of an inner product in answering that question. Second, working in $\mathbf{R}^3$ with its standard dot product, we define the quantities that describe the geometry of a curve in space (alias a vector-valued function of a scalar variable): arc length, tangent and normal vectors, curvature and torsion. Third, we define integrals along a curve with respect to arc length, and we take a first look at line integrals of vector fields along curves, which will return in the next chapter.

DIFFERENTIATION IN ABSTRACT VECTOR SPACES

Let $\vec{f}(t)$ be a vector-valued function of a scalar variable, t. In Sec. 1.4 we studied such functions whose codomain was $\mathbf{R}^3$; in that case, we perceived no ambiguity in what it meant to differentiate a function — just differentiate each of its coordinates:

$$\vec{r} = \vec{f}(t) \equiv \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}; \qquad \frac{d\vec{r}}{dt} = \vec{f}\,'(t) = \begin{pmatrix} x'(t) \\ y'(t) \\ z'(t) \end{pmatrix}.$$

(We identified $\vec{f}\,'(t)$ as the *tangent vector* at t to the curve traversed by the moving point $\vec{r}$.) In a physical application, however, the vectors have an absolute, geometrical significance independent of the choice of coordinate system in the physical space. If we change coordinates (see Sec. 4.4, and also Sec. 1.3), the physical vector-valued function will be represented by a different list of real-valued functions in the roles of $x(t)$, $y(t)$, and $z(t)$. We really ought to ask whether the definition of the derivative in terms of derivatives of the coordinate functions gives the same physical answer in all coordinate systems.

If the codomain of a vector-valued function is a generic vector space, $\mathcal{W}$, there are two approaches that can be followed to come up with a proper definition of its derivative.* The first is to stick with the coordinate definition and to answer, in the affirmative, the question just asked. Let $\{\vec{v}_j\}$ be a basis for $\mathcal{W}$. For simplicity we consider here only the case where this basis is *fixed* — that is, the vectors $\vec{v}_j$ themselves do not depend on t. (There are situations, however, where t-dependent bases are natural and useful, as we shall see below.) We'll also assume that $\dim \mathcal{W} < \infty$. Then $\vec{f}(t)$ can be expressed as

$$\vec{f}(t) = \sum_{j=1}^{\dim \mathcal{W}} r_j(t)\vec{v}_j$$

and the most obvious definition of its derivative is

$$\vec{f}\,'(t) = \sum_{j=1}^{\dim \mathcal{W}} r_j'(t)\vec{v}_j\,.$$

For instance, the codomain might be the space of $m \times m$ matrices: For each t, $\vec{f}(t) = U(t)$ is a matrix. (Thus $\dim \mathcal{W} = m^2$.) If each matrix element is a differentiable function of t, then $U'(t)$ is defined. The usual *product*

* This subsection is, as its title says, abstract, and your instructor may even suggest that you skip over it. However, the formula below for the derivative of the inverse of a matrix has great practical value and should certainly be learned.

rule holds for the derivative of a product of matrices, as one can verify by writing out the definition of the matrix product in terms of matrix elements and differentiating each element of the product. From this we can deduce by implicit differentiation a very useful identity for invertible-matrix-valued functions:

Theorem: $\left(U^{-1}\right)' = -U^{-1}U'U^{-1}$ (if U^{-1} exists).

PROOF: We have $1 = UU^{-1}$. Differentiate:

$$0 = \frac{d}{dt}\left(UU^{-1}\right) = \frac{dU}{dt}\,U^{-1} + U\,\frac{dU^{-1}}{dt}\,.$$

Thus $U\left(U^{-1}\right)' = -U'U^{-1}$; multiply by U^{-1} on the left to get the formula in the theorem.

If we change to a different basis for $\mathcal{W}$, the new coordinates of each $\vec{f}(t)$ will be related to the old coordinates by a certain t-independent transformation matrix, as explained in Theorem 1 of Sec. 4.4:

$$s_k(t) = \sum_j C_{kj} r_j(t).$$

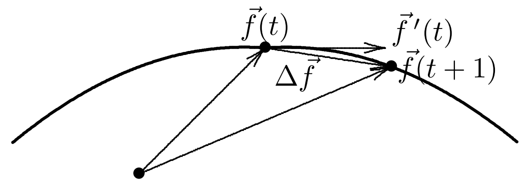
Then

$$s'_k(t) = \sum_j C_{kj} r'_j(t),$$

and it follows that the coordinates of the derivative vector $\vec{f}\,'(t)$ are transformed to the new coordinate system by the same matrix as those of the original vector, $\vec{f}(t)$. This can be taken to justify the claim that this approach to differentiation yields "the same physical answer in all coordinate systems."

One feels, however, that there ought to be a more directly physical, or geometrical, concept of a derivative, from which the coordinate formula with respect to any particular basis should follow. Indeed, this is what was achieved in Sec. 3.4, where the differential of a function from $\mathbf{R}^n$ to $\mathbf{R}^p$ was defined abstractly, and then the partial derivatives of the codomain coordinates with respect to the domain coordinates emerged as the coefficients of the matrix representation of the differential with respect to the natural bases. Right now we're concerned with the case $n = 1$, for which the usual definition of a derivative from calculus can be carried over:

$$\vec{f}\,'(t) \equiv \lim_{h\to 0} \frac{\vec{f}(t+h) - \vec{f}(t)}{h} \equiv \frac{\Delta \vec{f}}{\Delta t}\,.$$



This construction (and the whole treatment in Sec. 3.4) so far is confined to the $\mathbf{R}^p$ spaces, however, because it makes use of the standard definition of distance, $\|\vec{y}-\vec{x}\| \equiv \sqrt{(y_1-x_1)^2+\cdots}$, in those spaces in order to define the limit. It was observed in Sec. 6.1 that an inner product for any vector space $\mathcal{W}$ defines a distance function (metric), $\mathrm{d}(\vec{x},\vec{y}) \equiv \|\vec{x}-\vec{y}\|$. This enables one to define *limits* for functions into (or out of) $\mathcal{W}$, and hence to define continuity, derivatives, etc. The definitions are exact transcriptions of the usual ones for scalar functions, with norms ($\|\vec{x}-\vec{y}\|$) in place of absolute values ($|x-y|$). We have already done that (at the level of detail appropriate for this course) in Secs. 3.3 and 3.4 for the $\mathbf{R}^n$ spaces. In finite-dimensional spaces, whenever (fixed) bases are introduced, the definitions in terms of the inner product agree with the "naive" definitions in terms of coordinates with respect to the bases, discussed above. In finite-dimensional spaces, in fact, all inner products ultimately define the same limits, but in infinite-dimensional spaces this is definitely not true; that is why nonstandard inner products are more important in infinite dimensions.

Tangent vectors, unit tangent vectors, and arc length

Return now to a function $\vec{f}\colon \mathbf{R} \hookrightarrow \mathbf{R}^3$. (In fact, most of the discussion applies to any $\mathbf{R}^n$ in the role of $\mathbf{R}^3$.) We think of the $\mathbf{R}^3$ as representing physical space, or the space of possible values of some physical quantity such as velocity or force, with respect to an orthonormal basis; thus the dot product in $\mathbf{R}^3$ has a fundamental geometrical meaning.

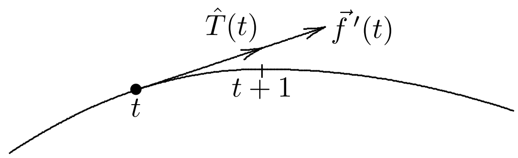
The term "curve" in common usage has two different meanings. Sometimes it refers to the *image* (or range) of the function $\vec{f}$, as a geometrical object sitting in $\mathbf{R}^3$. Sometimes, however, it refers, in effect, to the function $\vec{f}$ itself; that is, to the geometrical curve together with an assignment of a number t to each point on it. A curve in this second sense is called a *parametrized curve*. The differential and integral calculus of curves is largely concerned with extracting from a parametrized curve some quantities that characterize the geometry of the curve in the first sense — quantities that are independent of the particular parametrization used. To put it another way, a parametrization is a choice of a particular coordinate system on the one-dimensional object, the curve. We need to consider the freedom to change to a different coordinate system; in general, the transformation between the two coordinates will be a nonlinear function, say $t = \phi(u)$.

The tangent vector, $\vec{f}\,'(t)$, is not generally a vector of unit length. Also, the tangent vector clearly depends on the choice of parametrization, since by the chain rule

$$\frac{d\vec{f}}{du} = \frac{d\vec{f}}{dt}\,\frac{dt}{du} = \vec{f}\,'\big(\phi(u)\big)\phi'(u).$$

We can define a *unit tangent vector*,

$$\hat{T}(t) = \frac{\vec{f}\,'(t)}{\|\vec{f}\,'(t)\|}\,.$$



Consider now the construction of the unit tangent vector if the parameter u is used:

$$\hat{T}\big(t(u)\big) = \frac{(\vec{f}\circ\phi)'}{\|(\vec{f}\circ\phi)'\|}\,.$$

If $\phi'(u)$ is positive, it will cancel from the numerator and denominator, so the $\hat{T}$ at a particular physical point on the curve will be the same regardless of which parameter is used. If $\phi'(u)$ is negative, the direction of $\hat{T}$ will reverse. Clearly, $\hat{T}$ and $-\hat{T}$ are the only unit vectors proportional to the original tangent vector $\vec{f}\,'$. (Let us ignore for now the complications that result if either $\vec{f}\,'(t)$ or $\phi'(u)$ is zero; cf. Exercises 1.4.4 and 1.4.5.)

Instead of continuing to talk about all possible parametrizations, it is often convenient to introduce a special parametrization that is naturally associated with the geometry of the curve. Let a be an arbitrary value of t (in the domain of $\vec{f}$). Define the *arc length* s by

$$s = \sigma(t) \equiv \int_a^t \|\vec{f}\,'(\tilde{t})\|\,d\tilde{t}.$$

Then s is the "tape-measure distance" along the curve from $\vec{f}(a)$ to $\vec{f}(t)$. (For a detailed justification of this claim, see any textbook for third-semester calculus.) Furthermore,

$$\frac{ds}{dt} = \|\vec{f}\,'(t)\|, \qquad \hat{T} = \frac{d\vec{f}}{ds}\,.$$

It should be noted that for a concrete function $\vec{f}$, it is often impossible to evaluate the integral defining $\sigma(t)$ in terms of elementary functions, but there is no problem in calculating the derivatives needed to find ds/dt and $\hat{T}$.

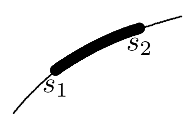
REMARK 1: In the equation $\hat{T} = d\vec{f}/ds$ we did not indicate an argument variable. It is understood that the equation holds at each point on the curve, and each such point may be labeled either by its parameter value t or by $s = \sigma(t)$. In the terminology of Sec. 7.3, we are adopting a "geometrical" point of view. From an "analytical" point of view, we should write, more pedantically,

$$\hat{T} \circ \sigma^{-1} = \frac{d}{ds}\big(\vec{f} \circ \sigma^{-1}\big).$$

REMARK 2: The parameter s is not uniquely defined. It depends upon the starting point a, and upon the direction in which the curve is traversed (the sign of $\hat{T}$). It could also change by a constant factor if the units in which t or $\|\vec{f}(t)\|$ is measured are changed.

The introduction of the arc-length parameter makes it possible to determine the physical length of the curve segment between $\vec{f}(t_1)$ and $\vec{f}(t_2)$ as simply the difference $|s_2 - s_1|$, where $s_j = \sigma(t_j)$. (In terms of t, the length is $\int_{t_1}^{t_2} \|\vec{f}\,'(t)\|\, dt$.) It also makes it possible to set up in a simple way the integrals for such quantities as the total amount of material in a wire when a density of material at each point is given. If the density at the point labeled by s is $\rho(s)$ grams per centimeter,† then the total mass of a wire segment is

$$\int_{s_1}^{s_2} \rho(s)\, ds.$$



In terms of an arbitrary parameter t this is

$$\int_{t_1}^{t_2} \rho\big(\sigma(t)\big)\, \|\vec{f}\,'(t)\|\, dt.$$

In the spirit of Remark 1, $\rho\big(\sigma(t)\big)$ is often written as $\rho(t)$ or simply as ρ.

† The units of density here are not grams per *cubic* centimeter, because it is understood that an integration over the (small) cross section of the wire has already been performed.

Normal vectors, binormal vectors, curvature, torsion

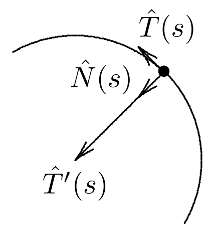
In the spirit of Remark 1, all the quantities to be treated here will be considered as functions of s or of t interchangeably, without notational distinction and without explicit indication of an independent variable.

The *curvature* of the curve $\vec{f}(t)$ at each point is defined by

$$\kappa \equiv \left\| \frac{d\hat{T}}{ds} \right\| .$$

At any point where $\kappa \neq 0$, the *(principal) normal vector* is defined to be

$$\hat{N} \equiv \frac{1}{\kappa} \frac{d\hat{T}}{ds} = \frac{d\hat{T}/dt}{\|d\hat{T}/dt\|} .$$



$\hat{N}$ is indeed normal (orthogonal) to $\hat{T}$ and hence to the curve, because the derivative of a unit vector with respect to a parameter is always normal to the original unit vector (Exercise 6.3.3). The curvature κ tells how fast the tangent vector is turning in the direction of the normal vector. On a sufficiently small scale near each point, therefore, the curve looks like a circular arc with radius $1/\kappa$ and center located off the curve in the direction indicated by $\hat{N}$.

Warning: Don't confuse this type of normal vector with the normal vectors to coordinate surfaces treated in Secs. 4.2 and 5.5. In dimension 2 the normal direction to a coordinate curve (determined by the gradient of the coordinate that is constant on that curve) points in the same direction as the $\hat{N}$ and the $d\hat{T}/ds$ of that curve, but need not have the same length or the same sign. In dimension 3 the $\hat{N}$ of a coordinate curve may not be perpendicular to a coordinate *surface* at all, because the curve might very well be turning within the surface. (See Exercise 6.3.8.)

We now come to the part of the construction that is special to dimension 3. Since $\hat{T}$ and $\hat{N}$ are orthogonal unit vectors,

$$\hat{B} \equiv \hat{T} \times \hat{N} \qquad (*)$$

is also a unit vector that is orthogonal to both of them, and $\{\hat{T}, \hat{N}, \hat{B}\}$ is a right-handed orthonormal basis, called the *Frenet–Serret basis* or the *moving trihedral*. $\hat{B}$ is called the *binormal vector*. Like the bases associated with curvilinear coordinate systems (Sec. 4.2), this basis differs from point to point; but it is associated with a single curve rather than an entire coordinate system. It is "adapted to the geometry of the curve". The coordinates of a vector (say, a force) with respect to this basis may have greater physical significance than those with respect to an arbitrary Cartesian basis. For example, if $\vec{f}(t)$ is the trajectory of a particle, then the $\hat{T}$-component of the acceleration vector $\vec{f}\,''(t)$ describes the change of speed of the particle, while the $\hat{N}$-component describes the change of direction. (The centripetal acceleration in circular motion is a familiar special case of this normal component of acceleration.) The $\hat{B}$-component of the acceleration is 0, because the definition of $\hat{N}$ forces $\hat{N}$ to lie in the plane of $\hat{T}$ and $\vec{f}\,''$.

It can be shown (see the theorem below) that $d\hat{B}/ds$ is parallel to $\hat{N}$. The *torsion* τ is defined by

$$|\tau| = \left\| \frac{d\hat{B}}{ds} \right\|$$

with the sign of τ chosen so that

$$\frac{d\hat{B}}{ds} = -\tau\hat{N}.$$

The torsion describes the extent to which the curve is wandering out of its original plane. In particular, if the curve lies entirely in a plane, then that plane is parallel to the $\hat{T}$-$\hat{N}$ plane, and hence $\tau = 0$.

Example. Calculate the Frenet–Serret basis and the curvature and torsion of the helix

$$x = R\cos(\omega t),$$
$$y = R\sin(\omega t),$$
$$z = Rbt \qquad (b << \omega).$$

Interpret the resulting κ and τ geometrically. (See Sec. 1.4 for a graph of a helix.)

SOLUTION: The tangent vector $\vec{f}\,'(t)$ is $(-\omega R\sin(\omega t), \omega R\cos(\omega t), Rb)$. Its length is $R\sqrt{\omega^2 + b^2} = ds/dt$, so

$$\hat{T} = \frac{1}{\sqrt{\omega^2 + b^2}} \begin{pmatrix} -\omega\sin(\omega t) \\ \omega\cos(\omega t) \\ b \end{pmatrix}.$$

Thus $\hat{T}' = (\omega^2 + b^2)^{-1/2}(-\omega^2\cos(\omega t), -\omega^2\sin(\omega t), 0)$ and hence

$$\hat{N} = \begin{pmatrix} -\cos(\omega t) \\ -\sin(\omega t) \\ 0 \end{pmatrix}.$$

Furthermore, $d\hat{T}/ds = (d\hat{T}/dt)/(ds/dt)$, so

$$\kappa = \|d\hat{T}/ds\| = \frac{\omega^2}{R(\omega^2 + b^2)}.$$

When b is small, $1/\kappa$ is approximately equal to R, the radius of the helix, as expected. Next, we calculate the cross product $(*)$ to be

$$\hat{B} = \frac{1}{\sqrt{\omega^2 + b^2}} \begin{pmatrix} b\sin(\omega t) \\ -b\cos(\omega t) \\ \omega \end{pmatrix}.$$

Then $-\tau\hat{N} = d\hat{B}/ds = (d\hat{B}/dt)/(ds/dt)$, so a short calculation yields

$$\tau = \frac{\omega b}{R(\omega^2 + b^2)}.$$

When b is small, we have $\tau \approx \frac{b}{R\omega}$. From the original equation of the curve, we see that one complete turn around the helix corresponds to an interval $\frac{2\pi}{\omega}$ in t, and hence that the distance between successive coils of the helix is $\frac{2\pi Rb}{\omega} \approx 2\pi R^2\tau$. This shows that τ measures how strongly the curve fails to lie in a plane.

Theorem (Frenet–Serret formulas):

$$\frac{d\hat{T}}{ds} = \kappa\hat{N}, \tag{1}$$

$$\frac{d\hat{N}}{ds} = \tau\hat{B} - \kappa\hat{T}, \tag{2}$$

$$\frac{d\hat{B}}{ds} = -\tau\hat{N}. \tag{3}$$

Proof: (1) is the definition of κ and $\hat{N}$. Differentiating $(*)$, and using (1) and the properties of the cross product, we get

$$\begin{aligned} \frac{d\hat{B}}{ds} &= \frac{d\hat{T}}{ds} \times \hat{N} + \hat{T} \times \frac{d\hat{N}}{ds} \\ &= \vec{0} + \text{term orthogonal to } \hat{T}. \end{aligned}$$

But $d\hat{B}/ds$ is also orthogonal to $\hat{B}$, since $\hat{B}$ is a unit vector. Therefore, $d\hat{B}/ds$ must point along $\hat{N}$. Once this fact is established, (3) is merely the definition of τ. Finally, $(*)$ implies that

$$\hat{N} = \hat{B} \times \hat{T}, \qquad \hat{T} = \hat{N} \times \hat{B}.$$

(These equations are just the cyclic permutations of the relationship among the three elements of an orthonormal basis, just like those for $\{\hat{\imath}, \hat{\jmath}, \hat{k}\}$.) It follows that

$$\begin{aligned} \frac{d\hat{N}}{ds} &= \frac{d\hat{B}}{ds} \times \hat{T} + \hat{B} \times \frac{d\hat{T}}{ds} \\ &= -\tau \hat{N} \times \hat{T} + \kappa \hat{B} \times \hat{N} \\ &= +\tau \hat{B} - \kappa \hat{T}. \end{aligned}$$

Line integrals

The integrals along curves that play the most conspicuous roles in physics are not arc-length integrals of scalar integrands, but *line integrals of vector fields*. The "line" in the problem is actually an arbitrary curve segment, C, in the space $\mathbf{R}^n$; it has a definite orientation (an "arrow" drawn on it to define a positive direction). Then the integral of the field $\vec{F}(\vec{x})$ along C,

$$I = \int_C \vec{F} \cdot d\vec{r},$$

is defined (in ways we shall soon discuss). In elementary physics the foremost application of these integrals is to the *work* done by a force field $\vec{F}$ on a particle that moves along the path C. For vividness we'll use that language in reviewing the meaning of the line integral. The main point is that there are two complementary approaches, and a third formulation that subsumes both of them.

Both approaches go back to the elementary concept of work for a particle moving in one dimension under a constant force: The work done by the force is the product of the force times the distance moved. More precisely, it is the force times the *displacement*, resulting in a positive number if the motion is in the same direction as the force, but a negative number if the motion is in the opposite direction.

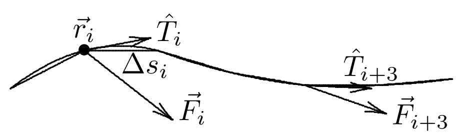
One way of moving this definition to three dimensions is the following: Let the displacement be a vector, $\Delta\vec{r}$. Let $\hat{T}$ be the unit vector in that direction (with the same sign):

$$\hat{T} = \frac{\Delta\vec{r}}{\Delta s} \quad \text{where } \Delta s = \|\Delta\vec{r}\|.$$

Then the work done is the component of the force in the direction of the motion, which is $\vec{F}\cdot\hat{T}$, times the distance moved, which is Δs. Now suppose that the particle moves along a curve instead of a straight line segment, and that the force is not necessarily constant. We think of chopping up the curve into tiny pieces that are approximately straight and applying the foregoing analysis to each piece. Then the total work done is

$$I = \lim \sum_j (\vec{F}(\vec{r}_j)\cdot\hat{T}_j)\Delta s_j \,,$$

where Δs_j is the length of the jth segment, $\hat{T}_j$ is the tangent vector to that segment, and $\vec{r}_j$ is an arbitrary point on the segment, so that $\vec{F}(\vec{r}_j)\cdot\hat{T}_j$ is the component of the force field parallel to the motion on that piece of the curve. As in all fundamental definitions of integrals, the limit is that where all the lengths Δs_j in the Riemann sum go to zero.



But this is just an instance of the scalar integrals with respect to arc length discussed earlier in this section:

$$\int_C \vec{F}\cdot d\vec{r} = \int_C (\vec{F}\cdot\hat{T})\,ds; \tag{G}$$

the integrand, instead of being the mass density of a wire, is the parallel component of a force. We call this the "geometric" definition of a line integral; it is the closest to physical intuition.

Example 1: Evaluate the work done when a particle moves under the force $\vec{F}(\vec{r}) = x\,\hat{\imath} + x\,\hat{\jmath}$ from the point $(0,0,0)$ to $(1,0,0)$ and then from there to $(1,2,0)$.

Solution: Along the first half of the path, we can take the arc length parameter to be $s = x$, and the unit tangent vector is $\hat{T} = \hat{\imath}$. So that part of the integral is

$$\int_{C_1} \vec{F} \cdot d\vec{r} = \int_0^1 (x\,\hat{\imath} + x\,\hat{\jmath}) \cdot \hat{\imath}\, dx = \int_0^1 x\, dx = \frac{1}{2}\,.$$

Along the second half, we have $s = y$ and $\hat{T} = \hat{\jmath}$, and hence

$$\int_{C_2} \vec{F} \cdot d\vec{r} = \int_0^2 x\, dy = \int_0^2 dy = 2,$$

since $x = 1$ everywhere on that segment. So the total work is $I = 2.5$.

Example 2: Evaluate the work done when a particle moves under the force $\vec{F}(\vec{r}) = x\,\hat{\imath} + x\,\hat{\jmath}$ from the point $\vec{r}_{\rm i} = (0, 0, 0)$ to $\vec{r}_{\rm f} = (1, 2, 0)$ along a straight line.

Solution: The unit tangent vector to the path is

$$\hat{T} = \frac{\vec{r}_{\rm f} - \vec{r}_{\rm i}}{\|\vec{r}_{\rm f} - \vec{r}_{\rm i}\|} = \frac{1}{\sqrt{5}}(\hat{\imath} + 2\hat{\jmath}).$$

The length of the path is $\sqrt{5}$, and the arc length parametrization is

$$x(s) = \frac{s}{\sqrt{5}}\,, \qquad y(s) = \frac{2s}{\sqrt{5}}$$

as s varies from 0 to $\sqrt{5}$. Therefore,

$$\begin{aligned} I &= \int_{s_{\rm i}}^{s_{\rm f}} \left(\frac{s}{\sqrt{5}}\,\hat{\imath} + \frac{s}{\sqrt{5}}\,\hat{\jmath} \right) \cdot \frac{1}{\sqrt{5}}(\hat{\imath} + 2\hat{\jmath})\, ds \\ &= \int_0^{\sqrt{5}} \left(\frac{s}{5} + \frac{2s}{5} \right) ds = \frac{3}{10}\, s^2 \Big|_0^{\sqrt{5}} = 1.5. \end{aligned}$$

Example 3: Evaluate $I = \int_C \vec{F} \cdot d\vec{r}$ when $\vec{F}(\vec{r}) = (x^2 z e^{2y}, y x^3, z)$ and C is the circle $x^2 + y^2 = 1$, $z = 0$.

Solution: Since the circle has unit radius, we can take the standard polar coordinate as the arc length parameter: $s = \theta$. The unit tangent vector to the circle is

$$\hat{T} = \hat{\theta} = -\sin\theta\,\hat{\imath} + \cos\theta\,\hat{\jmath}$$

(see Sec. 4.2). Thus

$$I = \int_0^{2\pi} \vec{F} \cdot \hat{T}\, d\theta = \int_0^{2\pi} (0 + \sin\theta \cos^4\theta + 0)\, d\theta = -\,\frac{1}{5}\,\cos^5\theta \bigg|_0^{2\pi} = 0.$$

(The first and third terms in $\vec{F} \cdot \hat{T}$ vanish because the factor z in F_x and F_z vanishes on C; the third term would have vanished anyway because the z component of $\vec{F}$ is never parallel to the motion (in other words, z does not change with s and therefore $dz = 0$).)

The other way of understanding work (and other line integrals) in three dimensions is to think separately about the work done in three perpendicular directions. Again, start with a vectorial displacement $\Delta\vec{r} = (\Delta x, \Delta y, \Delta z)$ and a constant force. The total work is the sum

$$F_x \Delta x + F_y \Delta y + F_z \Delta z.$$

Again, the case of a general motion through a general force is the limit of a Riemann sum of such contributions from "infinitesimal" displacements along the curve:

$$\int_C \vec{F} \cdot d\vec{r} = \int_C (F_x\, dx + F_y\, dy + F_z\, dz)\,. \tag{A}$$

We shall call this the "algebraic" formulation of the line integral. The concrete interpretation of this expression can be somewhat subtle, because, for instance, the integrand F_x is a function not only of x but also of y and z — but y and z are themselves in some sense functions of x, because (x, y, z) are the coordinates of a point on the definite curve C. Therefore, we will postpone all examples and problems of evaluating line integrals in this second way to Chapter 7 (although most readers will already be familiar with the subject from third-semester calculus and physics courses). Note, however, that Example 1 above comes out right if the formula (A) is applied naively, all the careful discussion of arc lengths and tangent vectors being shoved under the rug. (The usual intuitive argument is that dy and $dz = 0$ on C_1 because y and z are not changing there, so only the dx part of the integral needs to be considered; similarly, only the dy part contributes on C_2. Later we will see how to justify this more convincingly.)

Still a third physical approach to work is to define it as the integral over *time* of the integrand $\vec{F} \cdot \vec{v}$, where $\vec{v}$ is the velocity of the particle. If the particle follows the path $\vec{r} = \vec{g}(t)$, then $\vec{v}(t) = \vec{g}\,'(t)$, the tangent vector. Thus

$$\int_C \vec{F} \cdot d\vec{r} = \int_{t_\mathrm{i}}^{t_\mathrm{f}} \vec{F}\big(\vec{g}(t)\big) \cdot \vec{g}\,'(t)\, dt, \tag{P}$$

where $\vec{g}(t_\mathrm{i})$ and $\vec{g}(t_\mathrm{f})$ are the initial and final points of the path C. The magical fact is that this integral depends only on the spatial curve C, not on the details of how the particle traverses the curve in time. This is what allows work and energy to be defined in mechanics. (Here it is assumed that the force field $\vec{F}(\vec{r})$ does not depend on t.) To see how this happens, return momentarily to the definition of I in terms of arc length, formula (G). Substitute the concrete formulas for arc length and unit tangent vector in terms of a parametrization from earlier in this section, taking the time as the parameter (so that what was called $\vec{f}$ in the earlier formulas is now $\vec{g}$):

$$I = \int_C \vec{F} \cdot \hat{T}\, ds = \int_{t_\mathrm{i}}^{t_\mathrm{f}} \vec{F} \cdot \frac{\vec{g}\,'}{\|\vec{g}\,'\|}\, \|\vec{g}\,'\|\, dt = \int_{t_\mathrm{i}}^{t_\mathrm{f}} \vec{F} \cdot \vec{g}\,'\, dt.$$

The factor $\|\vec{g}\,'\|$ cancels out, leaving the parametric formula (P). The point is that this calculation applies to *any* parametrization of the curve, not just the one that corresponds to the physical motion. Thus (P) is a general prescription for evaluating line integrals, and one is *free to choose whichever parametrization of the curve is most convenient for calculations.*

In fact, the parametric definition is the most general, subsuming all the previous definitions. If we take t in (P) to be the arc length s, then (P) collapses back into the geometric definition (G). At the other extreme, we might choose $t = x$ in evaluating the part of the integral involving F_x, choose $t = y$ in evaluating the F_y term, etc.; then (P) becomes the algebraic definition (A). We shall have much more to say about (A) and (P) and their relation to each other in Secs. 7.4 and 7.6.

Savvy readers will undoubtedly have thought that the solution given above for Exercise 2 was unnecessarily complicated. Indeed it was; the reason for discussing the problem in that way was to emphasize the geometrical meaning of the line integral and to give practice in reasoning about arc length, *not* to demonstrate the most efficient way of solving that problem.

Alternative solution to Example 2: The most convenient parameter is probably $t = x$. We see that as x varies from 0 to 1, y varies (linearly) from 0 to 2. Thus $y = 2x = 2t$, and in vector form the parametrization is

$$\vec{g}(t) = \begin{pmatrix} t \\ 2t \\ 0 \end{pmatrix}; \qquad \vec{g}\,'(t) = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}.$$

Thus

$$I = \int_0^1 (x\,\hat{\imath} + x\,\hat{\jmath}) \cdot (\hat{\imath} + 2\hat{\jmath})\,dt = \int_0^1 (t + 2t)\,dt = \frac{3}{2}\ t^2\Big|_0^1 = 1.5.$$

Similarly, in Example 1 it is most convenient to choose $t = x$ on C_1 and $t = y$ on C_2; then formula (P) gives the answer quickly, without having to worry about *unit* tangent vectors (although in this particular case the tangent vectors $\vec{g}\,'$ turn out to be of unit length anyway). Likewise, in Example 3 the angle *is* the arc length, so there is not much difference between (G) and (P) in this case; however, if the radius of the circle had not been 1, using (P) would have saved a small bit of time. (Try Exercise 6.3.11 both ways!)

Generally speaking, the geometrical formula (G) will be the nicest way to evaluate the line integral if the curve is very simple and symmetrical, or if $\vec{F} \cdot \hat{T}$ is constant on the curve and the total arc length of the curve is known. (In the latter case the integral is just the product of the two given numbers.) For more complicated problems the parametric formula (P) is likely to be easier to apply. If the curve consists of straight line segments the algebraic formula (A) is easy to apply directly; in other cases (A) is often used as a starting point for rederiving (P) in the particular case at hand, by means of the chain rule; this procedure will be described in more detail in Sec. 7.6.

Exercises

6.3.1 A matrix U is called *orthogonal* if $U^{-1} = U^{\mathrm{t}}$ (see Sec. 8.2). Show that if $U(t)$ is orthogonal (for each t), then $A = U'(t)U(t)^{-1}$ is antisymmetric ($A^{\mathrm{t}} = -A$). In particular, if $U(0) = 1$ (the unit matrix) in this situation, then $U'(0)$ is antisymmetric.

6.3.2 *(Example of Exercise 6.3.1)* Let

$$U(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}.$$

Calculate $U'(t)$, $U'(t)U(t)^{-1}$, $U(0)$, and $U'(0)$.

6.3.3 Suppose that $\|\vec{f}(t)\|$ is a constant (independent of t). Show that $\vec{f}(t) \cdot \vec{f}\,'(t) = 0$ for all t. *Hint:* $\|\vec{f}(t)\|^2 = \vec{f}(t) \cdot \vec{f}(t)$. This can be differentiated by the product rule (see Exercise 6.3.4).

6.3.4 Show that the product rule (Leibnitz rule) applies to dot products of vector-valued functions:

$$\frac{d}{dt}\big(\vec{f}(t) \cdot \vec{g}(t)\big) = \vec{f}\,'(t) \cdot \vec{g}(t) + \vec{f}(t) \cdot \vec{g}\,'(t).$$

6.3.5 Show that the product rule (Leibnitz rule) applies to cross products of vector-valued functions:

$$\frac{d}{dt}\big(\vec{f}(t) \times \vec{g}(t)\big) = \vec{f}\,'(t) \times \vec{g}(t) + \vec{f}(t) \times \vec{g}\,'(t).$$

6.3.6 Find $\hat{T}$, $\hat{N}$, $\hat{B}$, κ, τ for the curve $\vec{f}(t) = \begin{pmatrix} a \\ bt \\ t^2 \end{pmatrix}$.

6.3.7 Find $\hat{T}$, $\hat{N}$, $\hat{B}$, κ, τ for the curve $\vec{f}(t) = \begin{pmatrix} at \\ bt^2 \\ t^3 \end{pmatrix}$.

6.3.8 Find the normal vector $\hat{N}$ to a circle of constant colatitude θ on a sphere of radius 1. (Either do this for a general θ, or choose a value of θ other than 0, $\frac{\pi}{2}$, or π.) Compare with the normal vectors to the coordinate surfaces, ∇r, $\nabla \theta$, and $\nabla \phi$ (Exercise 5.5.15). Sketch all four vectors. You should find that the four directions are all different.

In the remaining exercises, evaluate $I = \int_C \vec{A}(\vec{r}) \cdot d\vec{r}$ along the given curve in $\mathbf{R}^2$, when

$$A(x,y) = y\,\hat{\imath} - x\,\hat{\jmath}.$$

6.3.9 C is the rectangle with corners

$$(1,1), \quad (5,1), \quad (5,3), \quad (1,3)$$

(traversed in that order).

6.3.10 C is the line segment from $(0,0)$ to $(10,10)$.

6.3.11 C is the circle $x^2 + y^2 = 4$, travelled counterclockwise.

6.3.12 C is the parabolic path

$$y = 1 - x^2, \qquad -1 \le x \le 1.$$

(Use the parametric formula (P) with $x = t$.)

6.4 Nabla: The Vector Differential Operations

The standard notation for the calculus of vector fields in three-dimensional space, making heavy use of the formalism of dot and cross products, is essentially due to J. Willard Gibbs, the most prominent American theoretical physicist and mathematician of the turn of the twentieth century. The main operations can be expressed formally in terms of a “vector-valued differential operator”,

$$\nabla \equiv \hat\imath \,\frac{\partial}{\partial x} + \hat\jmath \,\frac{\partial}{\partial y} + \hat k \,\frac{\partial}{\partial z}\,.$$

This thing is pronounced “del” or, less ambiguously, “nabla”.* The operations are:

Gradient: $$\nabla f \equiv \frac{\partial f}{\partial x}\,\hat\imath + \frac{\partial f}{\partial y}\,\hat\jmath + \frac{\partial f}{\partial z}\,\hat k \equiv \operatorname{grad} f$$

Divergence: $$\nabla\cdot\vec A \equiv \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \equiv \operatorname{div}\vec A$$

Curl: $$\nabla\times\vec A \equiv \begin{vmatrix} \hat\imath & \hat\jmath & \hat k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} \equiv \operatorname{curl}\vec A$$
$$= \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)\hat\imath + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)\hat\jmath + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)\hat k$$

Laplacian: $$\nabla^2 f \equiv \nabla\cdot(\nabla f) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \equiv \Delta f \equiv \operatorname{Lap} f$$

* Nineteenth-century British mathematical physicists named the symbol after a certain kind of Assyrian harp.

Let's immediately get straight what kinds of animals these are:

$$\begin{aligned} \text{grad}\colon & \text{scalar} \to \text{vector} \\ \text{div}\colon & \text{vector} \to \text{scalar} \\ \text{curl}\colon & \text{vector} \to \text{vector} \\ \text{Lap}\colon & \text{scalar} \to \text{scalar} \end{aligned}$$

(That is, the gradient operation acts on a function of the type $\mathbf{R}^3 \to \mathbf{R}$ to produce a function of the type $\mathbf{R}^3 \to \mathbf{R}^3$, and so on.) EXAMPLE: If

$$\vec{F}(\vec{r}) = x\,\hat{\imath} + y\,\hat{\jmath} + (x^2 + y^2)\,\hat{k},$$

then

$$\nabla \cdot \vec{F} = \frac{\partial}{\partial x}(x) + \frac{\partial}{\partial y}(y) + \frac{\partial}{\partial z}(x^2 + y^2) = 1 + 1 + 0 = 2,$$

and

$$\nabla \times \vec{F} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x & y & x^2 + y^2 \end{vmatrix} = \hat{\imath}(2y - 0) + \hat{\jmath}(0 - 2x) + \hat{k}(0 - 0) = 2y\,\hat{\imath} - 2x\,\hat{\jmath}.$$

Although we concentrate on dimension 3, many things in vector calculus generalize to arbitrary dimension. (Those that do not are those that involve the curl or the cross-product symbol ($\times$) or the normal vector to a surface.) Occasionally we shall comment on what things reduce to in dimension 2.

RELATIONS BETWEEN ∇ AND THE JACOBIAN MATRIX

A vector field $\vec{A}$ is a function from $\mathbf{R}^3$ to $\mathbf{R}^3$. It therefore has a Jacobian matrix, $\vec{A}\,'$, whose elements are the partial derivatives of the components of $\vec{A}$:

$$\vec{A}\,' \equiv \begin{pmatrix} \frac{\partial A_x}{\partial x} & \frac{\partial A_x}{\partial y} & \cdots \\ \frac{\partial A_y}{\partial x} & \cdots & \\ \cdots & & \end{pmatrix}.$$

We notice that $\nabla \cdot \vec{A}$ is the sum of the diagonal elements of this matrix, also known as the *trace*, $\operatorname{tr} \vec{A}\,'$. We'll encounter the trace again in Sec. 8.2, observing that it is invariant under linear changes of coordinate system. That is the reason why the divergence is invariant under rotations — that is, it is the *same* scalar function regardless of which directions the orthogonal coordinate axes point in.

The components of the curl are related to the *antisymmetric part* of the Jacobian matrix:

$$\begin{aligned}\vec{A}' - (\vec{A}')^{\mathrm{t}} &= \begin{pmatrix} 0 & \frac{\partial A_x}{\partial y} - \frac{\partial A_y}{\partial x} & \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \\ \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} & 0 & \frac{\partial A_y}{\partial z} - \frac{\partial A_z}{\partial y} \\ \frac{\partial A_z}{\partial x} - \frac{\partial A_x}{\partial z} & \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -(\nabla \times \vec{A})_z & (\nabla \times \vec{A})_y \\ (\nabla \times \vec{A})_z & 0 & -(\nabla \times \vec{A})_x \\ -(\nabla \times \vec{A})_y & (\nabla \times \vec{A})_x & 0 \end{pmatrix}.\end{aligned}$$

In dimensions other than 3, it is the antisymmetrized Jacobian matrix which is significant (in deciding whether $\vec{A}$ is a gradient, for instance — see the end of Sec. 7.5). It is only in dimension 3 that the antisymmetric matrix can be related to a vector, $\nabla \times \vec{A}$. (In dimension n, an antisymmetric matrix has $\frac{1}{2}\,n(n-1)$ independent components; a vector has, of course, n components. Only when $n = 3$ are these numbers the same.) We'll look at this special three-dimensional connection in more detail in the next chapter.

Although it rarely occurs in the "classical" vector calculus associated with electromagnetism, the *gradient of a vector field* also can be defined, as already mentioned in Sec. 3.4. In a particular orthonormal coordinate system, this is done simply by taking the ordinary gradient of each scalar component of the vector field:

$$\nabla \vec{A} = \begin{pmatrix} \nabla A_1 \\ \nabla A_2 \\ \nabla A_3 \end{pmatrix} = \left(\frac{\partial A_j}{\partial x_k} \right) \qquad (j = 1,\, 2,\, 3, \quad k = 1,\, 2,\, 3).$$

Obviously, this operation "adds an index" to the field; the gradient of a vector-valued function is a matrix-valued function. Indeed, this gradient is simply another notation for the entire Jacobian matrix of $\vec{A}$. The dot product $\vec{B} \cdot \nabla \vec{A}$ is the directional derivative of $\vec{A}$ along $\vec{B}$, or, equivalently, the first-order change in $\vec{A}(\vec{r})$ under displacement of $\vec{r}$ by $\vec{B}$, which is $d\vec{A}(\vec{B})$ in the "modern" notation of Sec. 3.4. (For stylistic reasons people prefer in the present context a notation more consistent with classical vector calculus.) Here $\vec{B}$, as well as $\vec{A}$, may be a function of $\vec{r}$; $\vec{B}$ and $\vec{A}$ should be evaluated at the same point. To emphasize its interpretation as a directional derivative, $\vec{B} \cdot \nabla \vec{A}$ is often written $(\vec{B} \cdot \nabla)\vec{A}$.

The gradient (or the directional derivative) of a vector field is frequently used in applications in continuum mechanics. It also appears in identities describing the reduction of the vector operations acting on dot and cross products of fields, such as items 2(d) and 2(e) in the list below.

Similarly, the *Laplacian of a vector field* is calculated by taking the ordinary Laplacian of each component, to get a new vector field:

$$\nabla^2 \vec{A} = \nabla^2 A_x \,\hat{\imath} + \nabla^2 A_y \,\hat{\jmath} + \nabla^2 A_z \,\hat{k}.$$

A survey of identities satisfied by ∇

1. *Consequences of the linearity of differentiation:*

   (a) $$\nabla \cdot (r\vec{A} + \vec{B}) = r\nabla \cdot \vec{A} + \nabla \cdot \vec{B},$$

   and similarly for the gradient, curl, and Laplacian of a linear combination. (In other words, all of these are *linear* operators.)

2. *Consequences of the product rule:*

   (a) $$\nabla \cdot (f\vec{A}) = (\nabla f) \cdot \vec{A} + f\,\nabla \cdot \vec{A},$$

   (b) $$\nabla \times (f\vec{A}) = (\nabla f) \times \vec{A} + f\,\nabla \times \vec{A},$$

   (c) $$\nabla \cdot (\vec{A} \times \vec{B}) = \vec{B} \cdot (\nabla \times \vec{A}) - \vec{A} \cdot (\nabla \times \vec{B})$$

   (d) $$\nabla \times (\vec{A} \times \vec{B}) = (\vec{B} \cdot \nabla)\vec{A} - (\vec{A} \cdot \nabla)\vec{B} - \vec{B}(\nabla \cdot \vec{A}) + \vec{A}(\nabla \cdot \vec{B}).$$

   (e) $$\nabla(\vec{A} \cdot \vec{B}) = (\vec{B} \cdot \nabla)\vec{A} + (\vec{A} \cdot \nabla)\vec{B} + \vec{B} \times (\nabla \times \vec{A}) + \vec{A} \times (\nabla \times \vec{B}).$$

Remarks: Such identities can be derived or verified by writing everything out in components. When in doubt, write it out (when you don't have the opportunity to look it up in a reliable book). Note that algebraic vector identities, such as

$$(\vec{A} \times \vec{B}) \cdot \vec{C} = -(\vec{B} \times \vec{A}) \cdot \vec{C},$$

can be applied with ∇ in the role of one of the vectors if *and only if* care is taken to keep ∇ acting on the same functions on both sides of the identity. For example,

$$(\nabla \times \vec{A}) \cdot \vec{B} \neq -(\vec{A} \times \nabla) \cdot \vec{B},$$

since (in the normal interpretation of these symbols) ∇ is acting only on $\vec{A}$ in the left side of the equation and only on $\vec{B}$ in the right side. Here's an example where a standard vector identity does give a true differential-operator identity:

3. *Consequence of the identity* $\vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \cdot \vec{C})\vec{B} - (\vec{A} \cdot \vec{B})\vec{C}$ *:*

   (a) $$\nabla \times (\nabla \times \vec{A}) = \nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A},$$

   where the Laplacian of a vector field is defined by taking the Laplacians of its Cartesian components. We shall show in a moment that identity (2d) is also a consequence of this identity together with "care" (see above) to apply the product rule so as to effect the differentiation of both $\vec{A}$ and $\vec{B}$ in both terms of the result.

4. *Consequences of* $\frac{\partial^2 f}{\partial x\, \partial y} = \frac{\partial^2 f}{\partial y\, \partial x}$ *:*

   (a) $$\nabla \times \nabla f = 0,$$

   (b) $$\nabla \cdot (\nabla \times \vec{A}) = 0.$$

   These two identities are trivial to prove but extremely important.

Example. Prove identity 2(d).

SOLUTION: We use the identities for the cross product listed in the theorem at the end of Sec. 2.5. Identity (5) gives

$$\nabla \times (\vec{A} \times \vec{B}) = (\nabla \cdot \vec{B})\vec{A} - (\nabla \cdot \vec{A})\vec{B}, \qquad * * * \textbf{ CAREFUL! } * * *$$

where it must be understood that ∇ continues to act on all the quantities to its right. However, the standard interpretation of the symbols in this formula would be that the action of ∇ is confined inside the parentheses. To get a correct formula, therefore, we must apply the product rule for differentiation and write separately the terms where ∇ acts on the vector field outside the parentheses; those terms are vectorial directional derivatives:

$$\nabla \times (\vec{A} \times \vec{B}) = (\nabla \cdot \vec{B})\vec{A} + (\vec{B} \cdot \nabla)\vec{A} - (\nabla \cdot \vec{A})\vec{B} - (\vec{A} \cdot \nabla)\vec{B}.$$

This is equivalent to the formula to be proved.

ALTERNATIVE SOLUTION: Perhaps you are not confident in the manipulations in the first proof. Then write out each side of the identity in elementary terms:

$$\begin{aligned}
\nabla \times (\vec{A} \times \vec{B}) &= \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ (\vec{A} \times \vec{B})_x & (\vec{A} \times \vec{B})_y & (\vec{A} \times \vec{B})_z \end{vmatrix} \\
&= \hat{\imath} \left(\frac{\partial}{\partial y}(A_x B_y - A_y B_x) - \frac{\partial}{\partial z}(A_z B_x - A_x B_z) \right) \\
&\quad + \hat{\jmath} \left(\frac{\partial}{\partial z}(A_y B_z - A_z B_y) - \frac{\partial}{\partial x}(A_x B_y - A_y B_x) \right) \\
&\quad + \hat{k} \left(\frac{\partial}{\partial x}(A_z B_x - A_x B_z) - \frac{\partial}{\partial y}(A_y B_z - A_z B_y) \right) \\
&= \hat{\imath} \left(\frac{\partial A_x}{\partial y} B_y + A_x \frac{\partial B_y}{\partial y} - \frac{\partial A_y}{\partial y} B_x - A_y \frac{\partial B_x}{\partial y} \right. \\
&\qquad \left. - \frac{\partial A_z}{\partial z} B_x - A_z \frac{\partial B_x}{\partial z} + \frac{\partial A_x}{\partial z} B_z + A_x \frac{\partial B_z}{\partial z} \right) \\
&\quad + \text{similar } \hat{\jmath} \text{ and } \hat{k} \text{ terms,}
\end{aligned}$$

$$\begin{aligned}
&(B\cdot\nabla)\vec{A} - (\vec{A} \cdot \nabla)\vec{B} - \vec{B}(\nabla \cdot \vec{A}) + \vec{A}(\nabla \cdot \vec{B}) \\
&\quad = B_x \frac{\partial}{\partial x}\vec{A} + B_y \frac{\partial}{\partial y}\vec{A} + B_z \frac{\partial}{\partial z}\vec{A} - A_x \frac{\partial}{\partial x}\vec{B} - A_y \frac{\partial}{\partial y}\vec{B} - A_z \frac{\partial}{\partial z}\vec{B} \\
&\qquad - \left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \right) \vec{B} + \left(\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} \right) \vec{A} \\
&\quad = \hat{\imath} \left(B_x \frac{\partial A_x}{\partial x} + B_y \frac{\partial A_x}{\partial y} + B_z \frac{\partial A_x}{\partial z} - A_x \frac{\partial B_x}{\partial x} - A_y \frac{\partial B_x}{\partial y} - A_z \frac{\partial B_x}{\partial z} \right. \\
&\qquad \left. - \left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \right) B_x + \left(\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} \right) A_x \right) \\
&\qquad + \text{similar } \hat{\jmath} \text{ and } \hat{k} \text{ terms.}
\end{aligned}$$

In the second expression there is a cancellation of all the terms in which all three indices are x. The remaining 8 terms match up with those in the first expression. If you are still not confident, write out the $\hat{\jmath}$ and $\hat{k}$ terms, too. (After a little experience, it will be obvious that $\frac{2}{3}$ of the terms in calculations like these are always exactly the same as the remaining $\frac{1}{3}$ with the *cyclic* substitution

$$x \longrightarrow y \longrightarrow z \longrightarrow x$$

carried out once and twice. Recognizing this saves a tremendous amount of writing.)

Exercises

6.4.1 Consider the vector field $\vec{A}(\vec{r}) = x^2\,\hat{\imath} + y^2\,\hat{\jmath} + z^2\,\hat{k}$. Calculate each of these:

(a) $\nabla\vec{A}$ (gradient) (b) $\nabla\cdot\vec{A}$ (divergence) (c) $\nabla\times\vec{A}$ (curl)

6.4.2 The motion of a planetary atmosphere is governed by the equation[†]

$$\frac{d\vec{V}}{dt} + 2\Omega\hat{k}\times\vec{V} = -\frac{1}{\rho}\nabla p + B\vec{r} + \nu\nabla^2\vec{V}.$$

Here $\hat{k}$ is the unit vector parallel to the rotation axis, which we take to be the z axis. Write out the x, y, and z components of this equation (i.e., the three scalar equations that it summarizes). [The basic vectors in the theory are $\hat{k}$ (rotation axis), $\vec{V}$ (velocity field), and $\vec{r}$ (position). The basic scalars are t (time), Ω (angular speed of rotation), ρ (density), p (pressure, relative to an adiabat), B (buoyancy coefficient), and ν (kinematic viscosity).]

6.4.3 Let $\vec{B}$ be a constant vector field (i.e., the same vector $\vec{B}$ is attached to every point $\vec{r}$ in $\mathbf{R}^3$). Define $\vec{A}(\vec{r}) = \vec{B}\times\vec{r}$.

(a) Show that $\nabla\times\vec{A} = 2\vec{B}$.

(b) Calculate $\nabla\cdot\vec{A}$.

6.4.4 Here $\vec{A}$ is a vector field and f is a scalar function. One of these identities is correct, one is sheer nonsense, and one is sensible but numerically wrong. Label which is which, and correct the one with the numerical error (rederiving it if necessary).

(a) $\nabla\cdot(f\vec{A}) = (\nabla f)\cdot\vec{A} + f\,(\nabla\times\vec{A})$

(b) $\nabla\cdot(f\vec{A}) = (\nabla f)\cdot\vec{A} - 2f\,(\nabla\cdot\vec{A})$

(c) $\nabla\cdot(\nabla\times\vec{A}) = 0$

6.4.5 Tell which of the following "identities" is true, which is meaningful but generally false, and which is nonsensical. (There is one of each.)

† K. Zhang and G. Schubert, Penetrative convection and zonal flow on Jupiter, *Science* **273**, 941–943 (1996).

Explain your judgments briefly. Then prove the true one (by writing out the definitions of the "∇" operations).

(a) $\nabla \times \vec{B} + \nabla \cdot \vec{B} = \nabla \times (\nabla \times \vec{B})$

(b) $\nabla \cdot (g\vec{B}) = (\nabla g) \cdot \vec{B} + g(\nabla \cdot \vec{B})$

(c) $\nabla \cdot (\vec{A} \times \vec{B}) = \vec{A} \cdot (\nabla \times \vec{B})$

In the remaining exercises, prove the cited identity from the list in the text.

6.4.6 2(b)

6.4.7 2(c)

6.4.8 2(e)

6.4.9 3(a)

6.4.10 4(a)

6.4.11 4(b)