Chapter 3

Vector Spaces and Linear Functions

3.1 The Definition of a Vector Space

We began by saying that, loosely, vectors are any things that can be added to each other and multiplied by numbers. Two things were implicitly assumed then:

(1) The sum or product is always defined ("makes sense", "exists"). (We say that the space is *closed* under the addition and scalar-multiplication operations.)

(2) The operations satisfy algebraic properties that make them worthy of their names.

These are the things that have to be made explicit in the formal, precise definition:

Definition: A *(real) vector space* is a set $\mathcal{V}$ of elements $\vec{x}$, $\vec{y}$, $\vec{z}$, ... equipped with

(A) an operation of *addition*, such that the *sum*, $\vec{x}+\vec{y}$, of any two elements of $\mathcal{V}$ is an element of $\mathcal{V}$, and

(B) an operation of *scalar multiplication*, such that the *product*, $r\vec{x}$, of any real number with any element of $\mathcal{V}$ is an element of $\mathcal{V}$;

such that, moreover, the following conditions hold for all real numbers r, s, ... and all elements $\vec{x}$, $\vec{y}$, ... of $\mathcal{V}$:

1. $(\vec{x}+\vec{y})+\vec{z}=\vec{x}+(\vec{y}+\vec{z})$ (*associativity of addition*)

2. $\vec{x}+\vec{y}=\vec{y}+\vec{x}$ (*commutativity of addition*)

3. There is an element $\vec{0}$ in $\mathcal{V}$ such that $\vec{x}+\vec{0}=\vec{x}$ for all $\vec{x}$. (*existence of an additive identity*)

4. $\vec{x}+(-1)\vec{x}=\vec{0}$ (*existence of additive inverses*)

5. $1\vec{x}=\vec{x}$ (*existence of a multiplicative identity*)

6. $r(s\vec{x}) = (rs)\vec{x}$ (*associativity of multiplication*)

7. $r\vec{x} + s\vec{x} = (r+s)\vec{x}$ (*distributivity*)

8. $r\vec{x} + r\vec{y} = r(\vec{x} + \vec{y})$ (*another kind of distributivity*)

In summary, the usual algebraic manipulations with addition and numerical multiplication are valid. (But note that it is generally meaningless to *divide* one vector by another.)

"Vector" will be synonymous with "element of $\mathcal{V}$" — provided that $\mathcal{V}$ is the only vector space under discussion, so that there's no ambiguity. The real numbers (in their role as potential multipliers of vectors) are called "scalars".

Some vector spaces are based on *complex* scalars instead of real ones. Everything we say about real vector spaces (except their spatial visualization) will also be true of complex spaces, until we reach the topics of inner products and existence of eigenvectors in later chapters.

Examples

At the beginning of the book we looked at several examples of vectors:

- directed physical quantities such as velocities
- n-tuples of numbers
- polynomials
- solutions of homogeneous linear differential equations.

In all cases the operations of addition and scalar multiplication were natural and familiar; it is easy (although a bit tedious) to verify that the conditions (A), (B), 1–8 are satisfied in each case. Since then we discovered a new example: the space of all $m \times n$ matrices is a vector space.

Example 1. Prove that $\mathbf{R}^3$ is a vector space. (The space of n-tuples of scalars is called $\mathbf{R}^n$ or $\mathbf{C}^n$, depending on whether the scalars are real or complex.)

Solution: In this case the space $\mathcal{V} = \mathbf{R}^3$ comprises all triples of (real) numbers, (x_1, x_2, x_3). The definitions of addition and scalar multiplication for these objects were stated in Sec. 1.1, and it is obvious that the resulting sums and products are again objects of the same type, so (A) and (B) do not require any additional proof. So, we proceed to the 8 "axioms" in the definition of a vector space. We'll start with the seventh one, $r\vec{x} + s\vec{x} =$

$(r+s)\vec{x}$, just to have an example that involves both addition and scalar multiplication. Let us write it out in full detail, for an arbitrary vector $\vec{x}$ and arbitrary scalars r and s:

$$r(x_1, x_2, x_3) + s(x_1, x_2, x_3) = (r+s)(x_1, x_2, x_3).$$

Using the definitions of scalar multiplication and addition, and the ordinary distributive law for addition and multiplication of numbers, we can see that the left side of this equation really is equal to the right side:

$$\begin{aligned} r(x_1, x_2, x_3) + s(x_1, x_2, x_3) &= (rx_1, rx_2, rx_3) + (sx_1, sx_2, sx_3) \\ &= (rx_1 + sx_1, rx_2 + sx_2, rx_3 + sx_3) \\ &= \big((r+s)x_1, (r+s)x_2, (r+s)x_3\big) \\ &= (r+s)(x_1, x_2, x_3). \end{aligned}$$

The pattern is the same for each of the other axioms: it expresses some familiar property of numbers, together with the definitions of the two vector operations. We will not write out the verifications in full detail (except as homework exercises). The only axiom that requires extra comment is the third, the existence of the zero vector. In that case we need to point out that the zero vector is the string of numbers whose elements are all zeros:

$$\vec{0} = (0, 0, 0).$$

Example 2. Prove that the space $\mathcal{P}$ of all polynomials is a vector space.

SOLUTION: The hardest part of this proof is dealing with a notation general enough to handle all possible polynomials at once. A polynomial (in one variable, t) is a function of the form

$$\begin{aligned} p(t) &= a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \\ &= \sum_{j=0}^{n} a_j t^j. \end{aligned}$$

The integer n will be different for different polynomials.* To see that addition is well-defined, let

$$p(t) = \sum_{j=0}^{n} a_j t^j \quad \text{and} \quad q(t) = \sum_{j=0}^{m} b_j t^j$$

---

* For any particular polynomial, each a_j is a particular number, but t remains a variable. Some of the a_j may be 0, but normally it is tacitly understood that the leading coefficient, a_n , is not zero (except in the special case $n = 0$, where there would be nothing left in the expression if $a_0 = 0$ were omitted).

be two polynomials. Let N be the larger of n and m ($N \equiv \max(n, m)$), and whenever j is so big that a_j or b_j did not occur in the corresponding original polynomial, define that coefficient to be 0. Then by the rules of algebra,

$$(p+q)(t) \equiv p(t) + q(t) = \sum_{j=0}^{N} (a_j + b_j)t^j$$

— which, of course, is a polynomial, so (A) is true. Similarly but more simply, scalar multiplication is just multiplication of each coefficient in the polynomial by the scalar in question. We turn now to the verification of the 8 axioms; for variety, let's start with the commutative law this time.

$$\begin{aligned}(q+p)(t) &= \sum_{j=0}^{N} (b_j + a_j)t^j \\ &= \sum_{j=0}^{N} (a_j + b_j)t^j \\ &= (p+q)(t).\end{aligned}$$

Again it is fairly easy to see that all of the other axioms follow immediately from the corresponding properties of real numbers. The zero vector in this case is the zero function,

$$\begin{aligned}p(t) &= 0 \quad \text{for all } t \\ &= a_0 \quad \text{with } a_0 = 0.\end{aligned}$$

We can get new examples from old by restricting them; the new, smaller spaces are called *subspaces* (see Chapter 5). For example, instead of considering the space of *all* polynomials, let's look just at the space $\mathcal{P}_2$ of all *quadratic polynomials*,

$$p(t) = at^2 + bt + c.$$

By "quadratic" we mean "of degree *less than or* equal to 2." You can prove that this is a vector space by going through the argument in Example 2 without the complications of summation signs and arbitrary degrees.† Notice

† As we'll emphasize in Chapter 5, a better way to prove that such a set is a vector space is just to check that conditions (A) and (B) are satisfied — that is, addition and multiplication don't take us out of the set concerned — since the 8 axioms are then inherited automatically from the larger space.

that the polynomials with $a = 0$ *must* be included, else the set would not be a vector space:

$$(3t^2 + 1) + (-3t^2 + 2t + 2) = [0t^2+]2t + 3$$

(thus the set of polynomials of degree *exactly* 2 is not closed under addition).

Sometimes one has reason to enlarge a vector space, instead of restricting it. Another of our old examples was the space of functions

$$C \cos t + D \sin t$$

that satisfy a certain differential equation. This is a small part of a much larger vector space containing *all* (real-valued) functions of a real variable t (varying, say, from $-\infty$ to ∞). This space includes many "crazy" functions whose graphs are impossible to draw — for instance, there are functions that are not continuous for any value of t. It is useful to restrict attention to smaller spaces of "nice" functions that satisfy certain conditions; for example,

$$\mathcal{C}(-\infty, \infty) \equiv \text{ the space of all } \textit{continuous} \text{ functions of } t;$$

$$\mathcal{C}^1(-\infty, \infty) \equiv \text{ the space of all functions whose } \textit{derivatives} \text{ are continuous.}$$

$\mathcal{C}^1$ is a subset of $\mathcal{C}$, and our original space of solutions of the differential equation is a subset of $\mathcal{C}^1$. One of the most important modern applications of linear algebra is to solve many problems involving differential equations, etc., by regarding *functions as vectors*. Notice that the variable t in such examples is playing the same role as the index j in a vector space of n-tuples, $\vec{x} = (x_1, \ldots, x_j, \ldots, x_n)$.

Earlier we saw that the differential equation provided an example of the usefulness of looking at *complex* scalars. Another complex vector space, of importance to physicists, is the Hilbert space $\mathcal{H}$ of possible states of a quantum-mechanical system (e.g., all states of the electron in a hydrogen atom). (In fact, the complex numbers enter here in a more fundamental way than in the differential-equation case.)

Linear algebra is the study of what all these various vector spaces have in common. We emphasize that this subject is the foundation, the heart and soul, of modern applied mathematics.

Exercises

3.1.1 Complete Example 1 ("$\mathbf{R}^3$ is a vector space.") by verifying the other 7 axioms.

3.1.2 Complete Example 2 ("The polynomials form a vector space.") by verifying the other 7 axioms.

3.1.3 Prove that $\mathcal{C}(-\infty, \infty)$ (the space of all continuous functions of a real variable) is a vector space.

3.2 Linear Functions

Definition: A function on a vector space, $\mathcal{D}$, taking values in another vector space, $\mathcal{W}$, is *linear* if it "commutes" with the vector operations:

(1) $\vec{L}(\vec{x} + \vec{y}) = \vec{L}(\vec{x}) + \vec{L}(\vec{y})$ for all $\vec{x}$ and $\vec{y}$ in $\mathcal{D}$,

(2) $\vec{L}(r\vec{x}) = r\vec{L}(\vec{x})$ for all real r and all $\vec{x}$ in $\mathcal{D}$.

(Here we wrote "$\vec{L}$" to emphasize that the *value* of the function is a vector. From now on, however, the arrow will usually be omitted.)

From a practical point of view, linear functions are more important than vector spaces themselves. Typically, in an application one starts with some function or operation that satisfies the linearity conditions (for example, the differential operator $\frac{d^2}{dt^2} + 4$ in Example 4 of Sec. 1.1); then, to study that operator or solve equations involving it, one looks for vector spaces $\mathcal{D}$ and $\mathcal{W}$ in which that function acts. The precise choice of the domain $\mathcal{D}$ may be made for technical convenience — it may not be completely dictated by the original problem.

Remarks on terminology: Linear functions are also called *linear transformations*, *linear mappings*, or *linear operators*. ("Operator" is usually limited to cases where $\mathcal{D}$ and $\mathcal{W}$ are the same space.) If L is real-valued ($\mathcal{W} = \mathbf{R}$) and linear, it is called a linear *functional* (see Example 7). These alternative terminologies are especially common and useful in situations like Examples 2–7 below, where the vectors themselves are already functions; referring to both types of object as "functions" could be confusing. Another possible source of confusion when the elements of $\mathcal{D}$ and $\mathcal{W}$ are functions is that the argument, f, and the value, $L(f)$, of the linear transformation L

have arguments of their own that sometimes need to be written explicitly; to help with this, several slightly fussy remarks about notation are included in Example 2.

Example 1. Let $\mathcal{D}$ be $\mathbf{R}^2$ with typical element $\vec{x} \equiv \binom{x_1}{x_2}$ and let $\mathcal{W}$ be $\mathbf{R}^2$ with typical element $\vec{y} \equiv \binom{y_1}{y_2}$. Then the linear system

$$\begin{aligned} y_1 &= 3x_1 + x_2 \\ y_2 &= -x_1 \end{aligned}$$

defines a linear function,

$$\vec{y} = L(\vec{x}) = A\vec{x},$$

specified by the matrix $A = \begin{pmatrix} 3 & 1 \\ -1 & 0 \end{pmatrix}$. (More generally, every $m \times n$ matrix defines a linear function on $\mathbf{R}^n$ (as $\mathcal{D}$) into $\mathbf{R}^m$ (as $\mathcal{W}$); this general theory will be developed later in this section.)

PROOF: Let us verify that the conditions of linearity are satisfied. Let us consider two elements of $\mathcal{D}$,

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{and} \quad \vec{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

(We use "$\vec{z}$" because "$\vec{y}$" is already taken in the notation of this example.) Then

$$\begin{aligned} L(\vec{x} + \vec{z}) &= \begin{pmatrix} 3(x_1 + z_1) + (x_2 + z_2) \\ -(x_1 + z_1) \end{pmatrix} \\ &= \begin{pmatrix} 3x_1 + x_2 \\ -x_1 \end{pmatrix} + \begin{pmatrix} 3z_1 + z_2 \\ -z_1 \end{pmatrix} = L(\vec{x}) + L(\vec{z}); \end{aligned}$$

$$L(r\vec{x}) = \begin{pmatrix} 3rx_1 + rx_2 \\ -rx_1 \end{pmatrix} = r \begin{pmatrix} 3x_1 + x_2 \\ -x_1 \end{pmatrix} = rL(\vec{x}).$$

Example 2. Let $\mathcal{D} = \mathcal{P}_2$ (the quadratic polynomials) and let $\mathcal{W}$ be either $\mathcal{P}_2$ or $\mathcal{P}_1$ (the linear polynomials). Let L be the operation of differentiation:

$$[L(p)](t) \equiv p'(t).$$

[Example: If $p(t) = 3t^2 - 2t$, then $[L(p)](t) = 6t - 2$, and $[L(p)](4) = 22$.] Since our vectors are functions in this case, L is a *function on functions*, taking a function as input and producing a new function as output. This

L is linear on $\mathcal{D}$ into $\mathcal{W}$ — see Exercise 3.2.8. Here we have been careful to write "$[L(p)](t)$" and "$[L(p)](4)$" to make it clear that a function $L(p)$ (a polynomial in $\mathcal{W}$) has been defined or constructed, which is now to be evaluated at a point t or 4 (a number, generic or specific) to yield, finally, another number. In practice, shorter notations of the types $L(p)(t)$ and $Lp(t)$ are common. Would it make sense to group the symbols in the other order? The notation $L[p(t)]$ is indeed sometimes used when "$p(t)$" is thought of as a pattern of symbols, rather than the value of a function at a point t. For instance, our concrete example above could have been written more briefly as

$$L(3t^2 - 2t) = 6t - 2.$$

But such notations are potentially ambiguous: Does $L[p(4)]$ mean $p'(4)$ (which is 22), or $\frac{d[p(4)]}{dt} = \frac{d(40)}{dt}$ (which is 0)?

Example 3. Again let L be d/dt. Then L is linear as a function on $\mathcal{C}^1(-\infty, \infty)$ into $\mathcal{C}^0(-\infty, \infty) \equiv \mathcal{C}(-\infty, \infty)$. For this case, the two conditions in the definition of linearity are basic theorems of elementary calculus, which we take to be common knowledge.

Example 4. Let T be the operation *on functions* of *multiplication by* t:

$$[T(f)](t) \equiv t\, f(t).$$

[Example: If $f(t) = 3t + 1$, then $[T(f)](t) = 3t^2 + t$.] It is easy to see that T is a linear function on $\mathcal{P}_2$ into $\mathcal{P}_3$, and also on $\mathcal{C}(-\infty, \infty)$ into itself. (See Exercise 3.2.9.) More generally, for any fixed function c in $\mathcal{C}(-\infty, \infty)$ (say $c(t) = e^{-t^2}$) we can define a function M_c from $\mathcal{C}(-\infty, \infty)$ into itself by

$$M_c(f) \equiv cf; \quad \text{that is, } [M_c(f)](t) = e^{-t^2} f(t).$$

Example 5. Let us generalize Examples 2 and 4: d/dt is linear on $\mathcal{P}_n$ into $\mathcal{P}_{n-1}$ (or into $\mathcal{P}_n$, for that matter). T is linear on $\mathcal{P}_n$ into $\mathcal{P}_{n+1}$.

Example 6. A more advanced physical example: The electrostatic field due to a distribution of electric charges in empty space is a linear function of the charge density. This fact is called "the principle of superposition" in courses on electromagnetism. (Without becoming technical about multiple integrals and partial differential equations, we can't be precise now about which vector spaces are involved here as $\mathcal{D}$ and $\mathcal{W}$.)

Example 7. To approximate a definite integral numerically, one uses a *quadrature rule*, such as Simpson's rule or the trapezoidal rule. This is a formula of the type

$$\int_a^b f(t)\,dt \approx Q(f) \equiv \sum_{k=1}^{n} c_k f(t_k),$$

where the t_k are certain points in the interval $[a, b]$ and the c_k are certain coefficients (independent of f). Like the definite integral itself, a quadrature rule is a linear mapping from $\mathcal{C}[a, b]$ into $\mathbf{R}$. (See Sec. 7.1 for more on this.)

Remark: The two equations in the definition of linearity can be summarized in one equation:

$$L(r\vec{x} + \vec{y}) = rL(\vec{x}) + L(\vec{y}) \qquad (\dagger)$$

for all scalars r and all vectors $\vec{x}$ and $\vec{y}$. (To prove that a function is linear, you may either verify (†) or verify the conditions (1) and (2) in the definition.) Furthermore, by induction, if L is linear, then it can be pushed through any linear combination of vectors:

$$L\left(\sum_{j=1}^{N} r_j \vec{x}_j\right) = \sum_{j=1}^{N} r_j L(\vec{x}_j). \qquad (\ddagger)$$

This last equation expresses the full power of linearity — the way in which it is used. See Exercises 3.2.26–27 for further discussion.

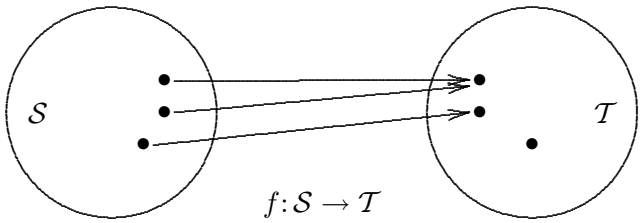
Domain and codomain

This is a good place to insert some remarks about the concept of a *function* in general (when the arguments (input; independent variables) and values (output; dependent variables) are not necessarily numbers). Let $\mathcal{S}$ and $\mathcal{T}$ be sets or "spaces" of any kind. (In our case, they are the vector spaces $\mathcal{D}$ and $\mathcal{W}$.) A *function f on $\mathcal{S}$ into $\mathcal{T}$* is an assignment of an element $y \equiv f(x)$ in $\mathcal{T}$ to each element x of $\mathcal{S}$. Such a situation is symbolized by the shorthand

$$f\colon \mathcal{S} \to \mathcal{T}.$$

Note that "into $\mathcal{T}$" means "taking values in $\mathcal{T}$". It does not mean that *all* of $\mathcal{T}$ is "reached" in this way. (On the other hand, it is ordinarily

understood that $f(x)$ *is* defined for *every* x in $\mathcal{S}$. However, in the sections of this book that treat nonlinear functions, we occasionally write things like "$f\colon \mathbf{R}^3 \to \mathbf{R}^2$" when we really have in mind functions f that might be defined only in certain regions of $\mathbf{R}^3$; the excuse for this sloppiness is avoiding complicated and pedantic statements.)



The elements of $\mathcal{S}$ or $\mathcal{T}$ may be functions themselves. For example, d/dt maps functions into functions, while

$$\int_0^3 \ldots \, dt$$

maps functions into numbers (it's a linear functional).

With this generalized understanding of functions, we don't really need the concept of "function of several variables" anymore. A function $f(x, y, z)$ of three real variables becomes $f(\vec{x})$, a function of *one* variable ranging through $\mathbf{R}^3$ (or a subset thereof).

Since our primary concern is with vector spaces and linear functions, we shall now revert to the notation L, $\mathcal{D}$, $\mathcal{W}$ instead of f, $\mathcal{S}$, $\mathcal{T}$, although many of the definitions and remarks that follow apply to functions in general. Thus, let L be a function on a set $\mathcal{D}$ into a set $\mathcal{W}$. That is, if $x \in \mathcal{D}$, then $L(x)$ is defined and is in $\mathcal{W}$. In shorthand, $L\colon \mathcal{D} \to \mathcal{W}$.

Definition: The set $\mathcal{D}$ (consisting of all the vectors (or other objects) on which L is defined) is the *domain* of L, sometimes denoted $\operatorname{dom} L$.

REMARK: When L is *linear*, part of the definition of "linear" is that $\mathcal{D}$ is a vector space (and not just a fragment of one, such as the unit square in $\mathbf{R}^2$).

Definition: The set $\mathcal{W}$ in which L takes its values is the *codomain* of L.

The crucial point is that this definition does not say that every point y in $\mathcal{W}$ *actually occurs* as a value $L(x)$. Therefore, we need to define a separate concept, that of "range", defined in Sec. 5.2.

Notice that if a set $\mathcal{W}$ satisfies the definition of codomain of L, then any larger set (containing $\mathcal{W}$ as a subset) also satisfies the definition. This means that there is a slight arbitrariness in the specification of the codomain of a function. For instance in Example 2 above we observed that the operation of differentiation, with $\mathcal{P}_2$ (the quadratic polynomials) as domain, is linear, and that the result of differentiating any quadratic polynomial is a first-degree polynomial. Therefore, we could regard that operator as having codomain $\mathcal{P}_1$ and write $L\colon \mathcal{P}_2 \to \mathcal{P}_1$; but we could also say that $\mathcal{P}_2$ is the codomain and write $L\colon \mathcal{P}_2 \to \mathcal{P}_2$. In what follows we'll see that this makes a difference for matrix representations of L; in the first case it will be represented by a 2×3 matrix, in the second case by a 3×3 matrix.

In most situations, however, there is a "natural" codomain consisting of all objects of the "type" that could reasonably be considered as candidates for values of the function. For instance, a linear function $L\colon \mathbf{R}^2 \to \mathbf{R}^2$ is defined by the equations

$$y_1 = 2x_1 + x_2\,,$$
$$y_2 = 4x_1 + 2x_2\,.$$

It is easy to see that the vector $\vec{y} = \binom{1}{0}$, for instance, is not $L(\vec{x})$ for any $\vec{x}$. Nevertheless, one says that the codomain of the function is $\mathbf{R}^2$, to emphasize that the values of the function are 2-component vectors, not numbers, or 3-component vectors, or functions. This is the most common type of appearance of the codomain concept.

Linear functions from $\mathbf{R}^n$ into $\mathbf{R}^p$

Theorem 1: Any $p \times n$ matrix, A, defines a linear function L on $\mathbf{R}^n$ into $\mathbf{R}^p$:

$$y_j = \sum_{k=1}^{n} A_{jk} x_k \quad (j = 1, 2, \ldots, p);$$

$$\vec{y} = A\vec{x} \quad (\vec{x} \in \mathbf{R}^n,\ \vec{y} \in \mathbf{R}^p).$$

This theorem says that we get a linear function from any formula which is "linear" in the elementary sense: it consists entirely of terms each of which is of first degree in just one of the independent variables, like $y = 3x_1 + x_2$. Conversely, formulas which are "nonlinear" in the elementary sense, such as $y = {x_1}^2 - 3x_2$, define functions which are nonlinear in the vector-space sense — unless such a formula happens to be equivalent to a linear one, as in the case

$$y = \frac{x_1({x_2}^2 + 1)}{{x_2}^2 + 1}\,.$$

This is made clear by the next theorem.

Theorem 2: Conversely, *every* linear function $L\colon \mathbf{R}^n \to \mathbf{R}^p$ can be expressed in terms of a matrix A as in Theorem 1. In fact, there is a rule for finding A: The kth column of A is the result of applying L to the vector

$$\hat{e}_k \equiv \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

which has a 1 in the kth row and 0 everywhere else.

REMARKS: The definition of $\hat{e}_k$ is summarized by

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

Here δ_{jk} is the *Kronecker delta symbol*, defined by

$$\delta_{jk} = \begin{cases} 1 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases}$$

The set of vectors $\{\hat{e}_k\}_{k=1}^n$ is called the *natural basis* of $\mathbf{R}^n$. The rule in Theorem 2 is summarized thus:

The kth-column rule: The kth column of the matrix of L is the image, under L, of the kth element of the natural basis.

PROOF OF THEOREMS: It's easy to verify (generalizing Example 1) that linear formulas give linear functions (Theorem 1), but the converse (Theorem 2) is more profound. Given that

$$L\Big(\sum_{k=1}^{N} r_k \vec{x}_k\Big) = \sum_{k=1}^{N} r_k L(\vec{x}_k)$$

for all vectors $\vec{x}_k \in \mathbf{R}^n$, we want to show:

For any $\vec{x} \equiv \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ in $\mathbf{R}^n$,

$$y_j \equiv [L(\vec{x})]_j = \sum_{k=1}^{n} A_{jk} x_k \quad (j = 1, \ldots, p),$$

where $A_{jk} = [L(\hat{e}_k)]_j$ for each j and k.

Well, $\vec{x} = \sum_{k=1}^{n} x_k \hat{e}_k$ implies that

$$L(\vec{x}) = L\left(\sum_k x_k \hat{e}_k\right) = \sum_{k=1}^{n} x_k L(\hat{e}_k)$$

(by linearity). Therefore,

$$[L(\vec{x})]_j = \left[\sum_k x_k L(\hat{e}_k)\right]_j = \sum_{k=1}^{n} x_k [L(\hat{e}_k)]_j$$

(by definition of vector addition). But this is just $\sum_{k=1}^{n} A_{jk} x_k$, Q.E.D.

Observe that a linear function is completely determined by its action on just a few vectors (here, the n vectors $\hat{e}_k$). We'll have much more about this in Chapter 4.

Remark: If p, the dimension of the codomain, is 1, then the matrix has only one row — it is a *row vector*. This is the "philosophical" difference between row vectors and column vectors: Row vectors represent linear functionals mapping vectors into numbers ($L\colon \mathbf{R}^n \to \mathbf{R}$), while columns represent the fundamental vectors themselves. (A column vector can also be regarded as a mapping of row vectors into numbers. And, of course, it is also the matrix of a mapping of numbers into vectors, $L\colon \mathbf{R} \to \mathbf{R}^p$.)

Affine functions

In elementary algebra, a function of the form

$$y = f(x) = ax + b$$

(with a and b real numbers, and x and y real variables) is called "linear". But it is *not* linear by our definition:

$$f(rx+z) = a(rx+z)+b = (ar)x + az + b,$$

but

$$rf(x)+f(z) = r(ax+b)+az+b = (ar)x+az+(1+r)b.$$

An analogous function in higher dimensions is (for example)

$$\begin{aligned} y_1 &= x_1 - 3x_2 + 6 \\ y_2 &= x_1 + 4x_2 + 12 \end{aligned}$$

The general pattern is

$$\vec{y} = F(\vec{x}) = A\vec{x} + \vec{b},$$

where $\vec{x}$ is in $\mathbf{R}^n$, $\vec{y}$ and $\vec{b}$ are in $\mathbf{R}^p$, and A is a $p \times n$ matrix.

Such functions are called *affine functions*. (In older books they're called "nonhomogeneous linear functions".) In abstract vector spaces the definition is the following.

Definition: A function $F\colon \mathcal{D} \to \mathcal{W}$ is *affine* if

$$\vec{F}(\vec{x}) = \vec{L}(\vec{x}) + \vec{b} \quad (\text{for all } \vec{x} \in \mathcal{D}),$$

where L is a *linear* function on $\mathcal{D}$ into $\mathcal{W}$ and $\vec{b}$ is a (fixed) vector in $\mathcal{W}$.

Technically, the affine functions include the linear functions, since $\vec{b}$ could be $\vec{0}$. However, occasionally the term "affine" may be used specifically to emphasize that the function in question is *not* linear ($\vec{b} \neq \vec{0}$). Analogy: The real numbers are complex numbers; but sometimes we say "z is complex" to mean that z is *not* real. Writers and readers of mathematics have to be careful about such ambiguities, interpreting terms sensibly in each context; the alternative would be an intolerably pedantic style.

Physical examples: The group of *Euclidean motions* (rotations and translations) in $\mathbf{R}^2$ (or in a physical plane) consists of the functions

$$F_{\theta,\vec{b}}(\vec{r}) = R_\theta \vec{r} + \vec{b}, \qquad R_\theta \equiv \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Note that $F_{\theta,\vec{0}}(\vec{r}) = R_\theta \vec{r}$, a pure rotation (which is linear). (See Exercises 3.2.17–18.) In three-dimensional space, three angles are required to specify

a rotation; otherwise, the Euclidean group can be described in the same way as in dimension 2. (The detailed description of all the possible rotations in dimension 3 and how they are related to each other is rather subtle, requiring the notions of *Euler angles* and *Lie algebra*, which are beyond the scope of this course. Introductions to these matters appear in advanced physics courses on classical and quantum mechanics.) The group of *Galilean transformations* consists of translations in *space-time* together with (linear) transformations to moving frames of reference:

$$(\vec{r}\,', t') = G_{\vec{v},\vec{b},\beta}(\vec{r}, t); \qquad \begin{aligned} \vec{r}\,' &= \vec{r} - \vec{v}t + \vec{b}, \\ t' &= t + \beta. \end{aligned}$$

These are affine functions on $\mathbf{R}^4$ into itself. The 3-vector parameter $\vec{b}$ and the scalar parameter β can be grouped together into a 4-vector parameter. In special relativity the Galilean transformations are replaced by the celebrated *Lorentz transformations*, in which the t' equation contains an $\vec{r}$ term.

Geometry and physics

Algebra and arithmetic are powerful and straightforward, but to make full use of vectors and linear functions on them in physical applications, it is also important to visualize what the linear functions are doing geometrically — that is, how they move vectors (and lines and planes) around in space. You should learn to recognize certain simple and frequently occurring motions as they are represented by matrices. We have just mentioned rotations. Here are some other examples; you might find it helpful to draw a sketch in each case of one or more typical vectors $\vec{r} = \binom{x}{y}$ and the resulting vectors $L\vec{r}$.

The matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ represents a *reflection* of the y coordinate (also called a reflection through the *horizontal* axis). This linear function changes the sign of y while leaving x unchanged. Visualize that: The vector is flipped "upside down".

The linear function with matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ interchanges x with y. By trying this out on a few examples, such as

$$\binom{1}{0}, \qquad \binom{0}{2}, \qquad \binom{1}{1}, \qquad \binom{1}{-2},$$

you should be able to see that this operation has the geometrical description of a *reflection through the diagonal line* $x = y$.

Finally, $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is not a reflection, but rather a *rotation through a right angle*. This is a special case of rotations as discussed above, but you should also see it by considering a variety of example vectors, as in the previous case.

More examples of matrices with simple geometrical interpretations are given in Exercises 3.2.16–24.

Operations on functions (Building new linear functions from old)

The following remarks are mostly reformulations of observations already made in Chapter 2.

Addition: The *sum* of two functions is defined by

$$[f + g](\vec{x}) \equiv f(\vec{x}) + g(\vec{x}).$$

The sum of linear functions is linear. Addition of linear functions corresponds to addition of their matrices: If $L(\vec{x}) = A\vec{x}$ and $K(\vec{x}) = B\vec{x}$, then $[L + K](\vec{x}) = (A + B)\vec{x}$.

Scalar multiplication: Similar remarks apply.

Composition: If we have functions $K{:}\mathcal{U} \to \mathcal{V}$ and $L{:}\mathcal{V} \to \mathcal{W}$, then their *composition*, $L \circ K$, is a function on $\mathcal{U}$ into $\mathcal{W}$ defined by

$$[L \circ K](\vec{x}) \equiv L\big(K(\vec{x})\big).$$

The composition of linear functions is linear. Composition of linear functions corresponds to multiplication of their matrices: If $L(\vec{y}) = A\vec{y}$ and $K(\vec{x}) = B\vec{x}$, then $[L \circ K](\vec{x}) = A(B\vec{x}) = (AB)\vec{x}$.

An example involving infinite-dimensional vector spaces: The most general second-order linear differential operator,

$$Af \equiv a(t)\,\frac{d^2 f}{dt^2} + b(t)\,\frac{df}{dt} + c(t)f(t),$$

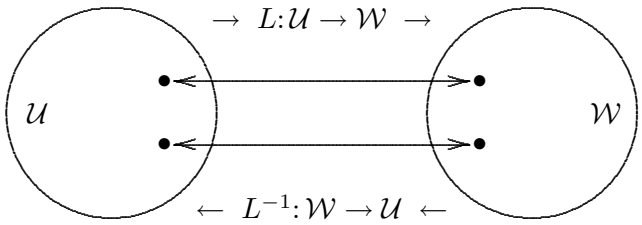
$$\big(A{:}\,\mathcal{C}^2(-\infty,\infty) \to \mathcal{C}^0(-\infty,\infty)\big),$$

is built up by composition and addition from the operator d/dt and the operators of multiplication by the fixed functions a, b, c.

Inversion: Given $L{:}\mathcal{U} \to \mathcal{W}$, there may exist a function $L^{-1}{:}\mathcal{W} \to \mathcal{U}$ such that $L \circ L^{-1} = L^{-1} \circ L =$ the identity function. That is,

$$L^{-1}\big(L(\vec{x})\big) = \vec{x} \quad \text{for all } \vec{x} \in \mathcal{U}, \quad \text{and} \quad L\big(L^{-1}(\vec{y})\big) = \vec{y} \quad \text{for all } \vec{y} \in \mathcal{W}.$$

The inverse of a linear function L is linear, and its matrix is the inverse of the matrix of L.



Exercises

3.2.1 The equations

$$u = x - 2y, \quad v = x + 3y, \quad w = x - 6y$$

define a linear function $L{:}\,\mathbf{R}^n \to \mathbf{R}^p$.

(a) What are [the numerical values of] n and p?

(b) What is the matrix that represents L?

3.2.2 Prove or disprove that each of these functions from $\mathbf{R}^2$ into itself is linear.

(a) $f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 3x_2 \\ 3x_1 - x_2^2 \end{pmatrix}$

(b) $f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 3x_2 \\ 3x_1 - 2x_2 \end{pmatrix}$

(c) $f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + 4y \\ 3x - y \end{pmatrix}$

3.2.3 Find the matrices representing these linear transformations of $\mathbf{R}^3$.

(a) $L(\vec{x}) = (x_1, x_2 + 2x_1, x_3)$

(b) $L(\vec{r}) = (x, z, y - 3x)$

(c) $L(\vec{r}) = (z, y, x + y + z)$

(d) $L(\vec{x}) = (x_1 + x_2, x_2 - x_1, x_3 + x_1 - x_2)$

3.2.4 Referring to Example 5 of Sec. 2.2, and letting

$$\vec{x} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \qquad \vec{z} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix},$$

calculate $B\vec{x}$ and $A(\vec{B}\vec{x})$ and compare the latter with $(AB)\vec{x}$; calculate $A\vec{z}$ and $B(\vec{A}\vec{z})$ and compare the latter with $(BA)\vec{z}$.

3.2.5 Suppose that

$$L\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad L\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

and that L is a linear function on all of $\mathbf{R}^2$. Find the matrix representing L. HINT: Start by expressing the natural basis vectors $\hat{e}_1$ and $\hat{e}_2$ as linear combinations of $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

3.2.6 Suppose that $f\colon \mathbf{R}^3 \to \mathbf{R}^2$ is affine. Find the matrix A and vector $\vec{b}$ such that $f(\vec{x}) = A\vec{x} + \vec{b}$, given that

$$f(\vec{0}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad f(\hat{e}_1) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad f(\hat{e}_2) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad f(\hat{e}_3) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

3.2.7 Show that under a change of origin in $\mathbf{R}^n$ as discussed in Sec. 1.3, an affine function $f\colon \mathbf{R}^n \to \mathbf{R}^n$ remains an affine function, but a linear function usually does not remain linear.

3.2.8 Verify that the operator of differentiation, acting as a function from $\mathcal{P}_2$ to $\mathcal{P}_1$, is linear. (Work out the derivatives of two polynomials, $a_1t^2 + b_1t + c_1$ and $a_2t^2 + b_2t + c_2$, and of their sum.)

3.2.9 Prove that the operation of multiplication by t (the argument variable of the functions) is a linear function from $\mathcal{P}_n$ to $\mathcal{P}_{n+1}$.

3.2.10 Define $A\colon \mathcal{C}^2(-\infty, \infty) \to \mathcal{C}(-\infty, \infty)$ by $[A(f)](t) \equiv f''(t) + 2f'(t) - e^t f(t)$. Prove or disprove that A is a linear operator.

3.2.11 Define $K\colon \mathcal{C}[0,1] \to \mathcal{C}[0,1]$ by

$$K(f)(t) = t^2 f(t) + \int_0^t f(\tilde{t})\, d\tilde{t}.$$

Show that K is linear (as a function of f).

3.2.12 For $u \in \mathcal{C}(-\infty, \infty)$, let's define $Bu \in \mathcal{C}(-\infty, \infty)$ by

$$Bu(x) = \int_0^x tu(t)\,dt.$$

Show that B is a linear function.

3.2.13 Here are two mappings defined on some space of continuously differentiable functions. Determine whether each is linear or nonlinear.

(a) $L(f) = \dfrac{d^2 f}{dt^2} + 4tf$ (b) $L(f) = \dfrac{df}{dt} - 3f^3$

3.2.14 State a reasonable choice of *domain* and *codomain* for the mapppings (differential operators) L in the previous exercise. (There is more than one possible correct answer.)

3.2.15 Let $D\colon \mathcal{C}^1 \to \mathcal{C}$ be the differentiation operator, and let $T\colon \mathcal{C} \to \mathcal{C}$ be the operation of multiplication of a function ($f(t)$) by its variable, t. (The domain of the functions f may be $\mathbf{R}$ or any fixed subinterval of $\mathbf{R}$.)

(a) Check that $C \equiv DT - TD$ makes sense as a linear function from $\mathcal{C}^1$ into $\mathcal{C}$, provided that the domain of each operator is properly interpreted. (C is called the *commutator* of the two original operators.)

(b) Calculate $C(f)(t)$ for $f(t) = t^2 + 3t$.

(c) Calculate $C(f)(t)$ for $f = \sin$.

(d) Calculate $C(f)(t)$ for an arbitrary function $f \in \mathcal{C}^1$.

3.2.16 Describe geometrically (using words like "rotation" and "reflection") the transformations of $\mathbf{R}^3$ represented by these matrices.

(a) $A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ (b) $B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ (c) $C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$

3.2.17 The matrix $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ describes a rotation of $\mathbf{R}^2$ counterclockwise through the angle θ.

(a) Calculate and compare $R_\theta R_\phi$ and $R_{\theta+\phi}$.

(b) What is R_θ^{-1}?

3.2.18 Justify the description of R_θ (see the previous exercise) as a *rotation through angle* θ by showing the following:

(a) $\|R_\theta \vec{r}\| = \|\vec{r}\|$ for all $\vec{r}$ (i.e., R_θ preserves the lengths of vectors).

(b) $\vec{r} \cdot R_\theta \vec{r} = \|\vec{r}\|^2 \cos\theta$.

(c) $\det R_\theta$ is positive (i.e., R_θ preserves the *orientation* of $\mathbf{R}^2$).

3.2.19 The matrix $X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ represents a 90° counterclockwise rotation around the x axis in $\mathbf{R}^3$.

(a) Find the corresponding matrices Y and Z for rotations around the other axes.

(b) Calculate XYX^{-1}. Rotate a book in your hands to convince yourself that the result is correct.

(c) Deduce geometrically (i.e., doing as little matrix multiplication as possible) the five analogous products,

$$YZY^{-1}, \quad ZXZ^{-1}, \quad YXY^{-1}, \quad ZYZ^{-1}, \quad XZX^{-1}.$$

3.2.20 Find the matrix of the linear function $A\colon \mathbf{R}^3 \to \mathbf{R}^3$ if:

(a) A is rotation by 90° around the axis $\hat{e}_3$.

(b) A is rotation by 180° around the axis $\hat{e}_2$.

(c) A is the reflection in the plane containing $\hat{e}_2$ and $\hat{e}_3$.

3.2.21 Let $\vec{a}_0$ be a fixed vector in $\mathbf{R}^3$. Which of these transformations in $\mathbf{R}^3$ are linear?

(a) $A(\vec{r}) = a_0$ (for all $\vec{r}$) (b) $A(\vec{r}) = \|\vec{r}\|\,\vec{a}_0$

(c) $A(\vec{r}) = \vec{r} + \vec{a}_0$ (d) $A(\vec{r}) = (\vec{a}_0 \cdot \vec{r})\vec{r}$

(e) $A(\vec{r}) = (\vec{a}_0 \cdot \vec{r})\vec{a}_0$ (f) $A(\vec{r}) = \vec{a}_0 \times \vec{r}$

3.2.22 Find the matrix of the linear operator of *reflection in the plane* perpendicular to the unit vector $\vec{a}_0 = \alpha_1\hat{e}_1 + \alpha_2\hat{e}_2 + \alpha_3\hat{e}_3$. HINT: Use (e) of the previous exercise to extract the component of a vector parallel to $\vec{a}_0$. Under reflection, this component changes sign while the rest of the vector is unchanged.

3.2.23 The linear function $A\colon \mathbf{R}^2 \to \mathbf{R}^2$ with matrix $\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$ is an example of a *shear transformation*. Describe it geometrically. (What happens to points on a typical horizontal line, $y = y_0$?)

3.2.24 If $0 < k < n$, let P be the operator from $\mathbf{R}^n$ to $\mathbf{R}^k$ that maps

$$(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) \qquad \text{to} \qquad (x_1, \ldots, x_k).$$

(P is called a *projection operator*.)

(a) Show that P is linear.

(b) Illustrate the action of P by a sketch, in the case where $k = 1$ and $n = 2$.

3.2.25 Which of these functionals on $\mathcal{C}[0, 1]$ (mappings from $\mathcal{C}[0, 1]$ to $\mathbf{R}$) are linear functionals?

(a) $\displaystyle\int_0^1 t|p(t)|\,dt$ (b) $\displaystyle\max_{0\le t\le 1} p(t)$

(c) $\displaystyle\int_0^1 p^2(t)\,dt$ (d) $\displaystyle\int_0^1 p(t)\sin^2 t\,dt.$

3.2.26 To prove that the condition (†) is equivalent to the definition of linearity [conditions (1) and (2)], one must show that each implies the other. The proof that (1) and (2) imply (†) is quite easy. What is wrong with the following alleged proof that (†) implies (1) and (2)?

> First, setting $r = 1$ in (†) yields $L(\vec{x} + \vec{y}) = L(\vec{x}) + L(\vec{y})$, which is (1). Second, setting $\vec{y} = \vec{0}$ in (†) yields $L(r\vec{x}) = rL(\vec{x}) + \vec{0} = rL(\vec{x})$, which is (2).

Insert the third step needed for a correct proof.

3.2.27

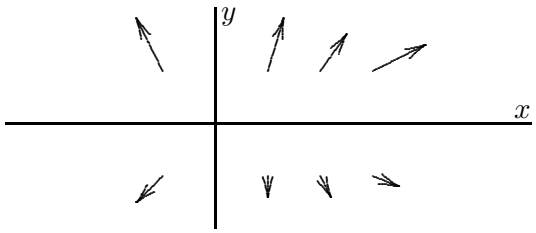
(a) Prove by mathematical induction that (‡) follows from (†) [or from (1) and (2)].

(b) Since (†), (1), and (2) are trivial special cases of (‡), it follows that (‡) is *equivalent* to the definition of linearity. Discuss the advantages or disadvantages of using (‡) as the "official" definition of linearity. [HINT: Sometimes an important mathematical concept arises as the hypothesis of a theorem; sometimes it arises as the conclusion of a theorem.]

3.3 Nonlinear Functions

In Sec. 1.4 we looked at nonlinear functions from $\mathbf{R}$ into $\mathbf{R}^p$ and observed that they represent curves, either in $\mathbf{R}^p$ or in $\mathbf{R}^{p+1}$. In Sec. 2.4 we looked at nonlinear functions from $\mathbf{R}^n$ into $\mathbf{R}$ and their geometrical representations by either graphs or level surfaces. Now we want to consider the general situation, functions from $\mathbf{R}^n$ (or some subset thereof)* into $\mathbf{R}^p$. As in the special cases discussed earlier, there are several conceptual or geometrical points of view that one can take toward such a function.

1a. If $n = 2$ or 3 and $p = n$, we may visualize the function as an attachment of a vector to each point (a *vector field*).



1b. The codomain may not have the same dimension as the domain, or directions in one may not have any natural identification with directions in the other. Then it is hard to draw a picture, but the principle is the same. EXAMPLES: $\vec{f}(\vec{r})$ at each point $\vec{r}$ could be a symmetric matrix (representing electric polarizability, or stress or strain, or the index of refraction of a crystal). In the quantum mechanics of an electron, $\vec{f}(\vec{r})$ can be in $\mathbf{C}^2$. Or consider the location hit by an artillery shell as a function of the elevation and azimuth of the gun: $\vec{r} = f(\theta, \phi) \in \mathbf{R}^2$. (In this case we'd be more likely to sketch the codomain than the domain.)

2a. If $n < p$, then $\vec{f}\colon \mathbf{R}^n \hookrightarrow \mathbf{R}^p$ may yield a *parametrized submanifold*. That is, the range of $\vec{f}$ is a geometrical set that is n-dimensional, in

* In this section we shall write $f\colon \mathbf{R}^n \hookrightarrow \mathbf{R}^p$ (rather than $\ldots \mathbf{R}^n \to \ldots$) to indicate that the possible values of the independent variable of f belong to $\mathbf{R}^n$ but the domain of f is not necessarily all of $\mathbf{R}^n$. Later on, however, we shall usually not bother to make this fine distinction.

the sense that it takes n numbers (coordinates) to specify a point on it, but it may (or may not) be "curved" so that it won't fit into a flat space of dimension less than p. If $n = 2$ and $p = 3$, the submanifold is a patch of surface floating in 3-dimensional space. For example, the formulas

$$\begin{aligned} x &= R\sin\theta\cos\phi \\ y &= R\sin\theta\sin\phi \qquad \left(R = \text{constant}, \quad 0 < \theta < \tfrac{\pi}{4}\,, \quad 0 < \phi < 2\pi\right) \\ z &= R\cos\theta \end{aligned}$$

embed a patch of $\mathbf{R}^2$ into $\mathbf{R}^3$ as a *spherical cap* (a fragment of a sphere, with a circular boundary).

2b. If $n = p$, $\vec{f}$ may be a *coordinate transformation*. For example, replacing the constant R by a variable r in the equations above gives the standard definition of spherical coordinates in $\mathbf{R}^3$.

Limits and continuity

Not every function is tame enough to work with. Ideally, we would like a function to be *differentiable*, so that it can be approximated locally by a linear function, as described in the next section. Failing that, we would like a function to be *continuous*, so that a small error in the input variable $\vec{x}$ will cause only a small error in the output, $f(\vec{x})$. In an abstract sense, "most" functions satisfy neither of these conditions. However, most functions encountered in practice — functions for which we can write down a formula in terms of familiar, elementary functions, or solutions of equations expressed in terms of such functions — are *piecewise* continuous (and differentiable): They are continuous everywhere in their domains except possibly for some isolated points or sets (lines, surfaces, etc.) of lower dimension than the domain.† One needs to be able to recognize the places where continuity fails, and to have a suitable language for describing what can happen there.

† It can be argued that any discontinuity that arises in a function representing a physical quantity, such as temperature or index of refraction, is an artifact of a mathematical idealization of the problem. When one looks at a discontinuity on a sufficiently small scale, either (1) the "true" function has a rapid but smooth variation there, or (2) the function becomes meaningless because the mathematical model has broken down (for example, because real matter is made of molecules, not continuum substances). This observation does not relieve us of the need to deal somehow with the discontinuities that are present in the model itself, even if only as a reason for rejecting the model for certain purposes.

The fundamental notion involved here is that of a *limit*. Recall that in single-variable calculus the definition of

$$\lim_{x \to a} f(x) = b$$

is

> For every $\epsilon > 0$ there is a $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|f(x) - b| < \epsilon$.

The absolute values here are our way of measuring *distances* between numbers. To extend the definition to vector functions $f\colon \mathbf{R}^n \hookrightarrow \mathbf{R}^p$, therefore, we need the distance function on vectors, which is defined in terms of the dot product:

$$\begin{aligned} \|\vec{x} - \vec{y}\| &\equiv \sqrt{(\vec{x} - \vec{y}) \cdot (\vec{x} - \vec{y})} \\ &\equiv \sqrt{\sum_{j=1}^{m} (x_j - y_j)^2} \quad \text{for } \vec{x} \text{ and } \vec{y} \in \mathbf{R}^m. \end{aligned} \tag{$*$}$$

(Of course, if either n or p equals 1, the distance function reverts to the elementary one.)

Remark: This distance is the length of the vector $\vec{x} - \vec{y}$. So the more fundamental concept is the *length* of a vector in $\mathbf{R}^m$,

$$\|\vec{x}\| \equiv \sqrt{\sum_{j=1}^{m} x_j^2}\,.$$

Sometimes (for instance, in Exercise 3.3.2) it is convenient or necessary to use some other measure of the size of a vector. For example, we might define

$$|\vec{x}| \equiv \max_{1 \le j \le m} |x_j|.$$

As far as limits are concerned, it doesn't really matter which definition of the size of $\vec{x} - \vec{y}$ is used, because if $\|\vec{x} - \vec{y}\|$ is small, then $|\vec{x} - \vec{y}|$ is also small, and vice versa. To see this, note that

$$\sum_{j=1}^{m} x_j^2 \le m \max_{1 \le j \le m} x_j^2 = m \left(\max_{1 \le j \le m} x_j \right)^2,$$

so that

$$\|\vec{x}\| \leq \sqrt{m}\,|\vec{x}|,$$

and also

$$\left(\max_{1\leq j\leq m} x_j\right)^2 \leq \sum_{j=1}^{m} x_j^2,$$

so that

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

Definition: $\lim_{\vec{x}\to\vec{a}} f(\vec{x}) = \vec{b}$ means that

For every $\epsilon > 0$ there is a $\delta > 0$ such that $0 < \|\vec{x} - \vec{a}\| < \delta$ implies $\|f(\vec{x}) - \vec{b}\| < \epsilon$.

Such a limit statement makes two separate assertions: that the limit exists, and that it is equal to the particular vector $\vec{b}$. Often only the first of these is important, and the relevant statement begins "There exists a $\vec{b} \in \mathbf{R}^p$ such that ... ."

Definition: f is *continuous at* $\vec{x} \in \mathbf{R}^n$ if

$$\lim_{\vec{z}\to\vec{x}} f(\vec{z}) = f(\vec{x}).$$

This statement makes three assertions: (1) $f(\vec{x})$ is defined; (2) the limit exists; and (3) these two vectors are equal.

Often one says just "f is continuous" to mean that f is continuous at *every* point in some tacitly understood set — usually all of $\mathbf{R}^n$, or at least the entire domain where f is defined.

Examples. Consider $n = 2$ with notation $\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, and $p = 1$.

(1) $f(x, y) = \dfrac{1}{x - y}$ is continuous everywhere except on the line $x = y$. On that line the function is not even defined. Any attempt to define the function there by a supplementary condition will leave the function discontinuous.

(2) $f(x, y) = \dfrac{x^2 - y^2}{x - y}$ is technically undefined (and hence discontinuous) on the line $x = y$. However, this discontinuity is *removable*. Everywhere else, the formula simplifies to $f(x, y) = x + y$, and this function

extends smoothly to the problematical line. In most applications this extended definition of f is perfectly appropriate. Algebraic simplification without attention to possible division by 0 will get you into trouble perhaps one percent of the time, so the phenomenon should be understood and kept in the back of your mind. (Beware of complicated computer calculations involving symbolic algebra programs.)

(3) $f(x, y) = |x - y|$ is defined and continuous everywhere. (On the line $x = y$ it is not differentiable.) The graph of the function consists of two planes meeting on the line $x = y$. The gradient of f is a function $g\colon \mathbf{R}^2 \hookrightarrow \mathbf{R}^2$ given by

$$g(x, y) = \begin{cases} (1, -1) & \text{if } x > y, \\ (-1, 1) & \text{if } x < y; \end{cases}$$

g is continuous (and differentiable) everywhere except on our favorite line, where it is not defined.

From the large number of elementary consequences of limits and continuity we list just the following. (If the functions are not everywhere defined, it is to be understood that their domains match up so that the compound function has a nontrivial domain of definition.)

Theorem 1:

(1) Linear combinations $\big(r_1 f(\vec{x}) + r_2 g(\vec{x})\big)$ and products $\big(f(\vec{x})g(\vec{x})\big)$ of continuous functions are continuous.

(2) The composition $f \circ g$,

$$(f \circ g)(\vec{x}) \equiv f(g(\vec{x})),$$

of two continuous functions ($g\colon \mathbf{R}^n \hookrightarrow \mathbf{R}^m$, $f\colon \mathbf{R}^m \hookrightarrow \mathbf{R}^p$) is continuous. More generally,

$$\lim_{\vec{x} \to \vec{a}} f(g(\vec{x})) = f\left(\lim_{\vec{x} \to \vec{a}} g(\vec{x})\right)$$

if f is continuous and the limit on the right exists and is in the domain of f. This last equation may be regarded as the operational significance, or practical power, of continuity: A continuous function "commutes" with taking limits and therefore enables limits like that on the left to be evaluated easily.

Theorem 2: $f\colon \mathbf{R}^n \hookrightarrow \mathbf{R}^p$ is continuous if and only if its p coordinate functions $f_j\colon \mathbf{R}^n \hookrightarrow \mathbf{R}$,

$$y_j \equiv \big(f(\mathbf{x})\big)_j = f_j(x_1, \ldots, x_n),$$

are continuous.

PROOF: This follows from the definition of the distance function $(*)$ in the vector spaces. If the f_j are continuous at $\vec{x} = (x_1, \ldots, x_n)$, then for every ϵ there is a δ_j such that $\|\vec{z} - \vec{x}\| < \delta_j$ guarantees $|f_j(\vec{z}) - f_j(\vec{x})| < \epsilon/\sqrt{p}$. Choose δ to be the smallest of the δ_j; then $\|\vec{z} - \vec{x}\| < \delta$ implies $\|f(\vec{z}) - f(\vec{x})\| < \epsilon$. Conversely, if $\|\vec{z} - \vec{x}\| < \delta$ guarantees $\|f(\vec{z}) - f(\vec{x})\|$ to be small, then it certainly guarantees the individual terms $(f_j(\vec{z}) - f_j(\vec{x}))^2$ in $(*)$ to be small.

REMARK: When a function is not everywhere defined, it is convenient and customary to consider only points in the function's domain in the role of $\vec{z}$ in the definition of a limit. This makes it possible to discuss limits and continuity at points $\vec{x}$ in the boundary of the domain. Furthermore, if $\vec{x}$ is isolated from the domain (neither in the domain, nor even in its boundary), then the question of continuity at $\vec{x}$ does not even arise. EXAMPLES: The largest natural domain of $f(x, y) = \sqrt{x - y}$ is the set where $x \geq y$. (In this discussion we are not admitting complex numbers as values of the square root.) In particular, f is defined and continuous on the line $x = y$. One would not say that f is discontinuous in the region $x < y$, although it is undefined there. The largest natural domain of $g(x, y) = 1/\sqrt{x - y}$ is $x > y$. On the boundary $x = y$, g is discontinuous (and undefined). For $x < y$, g is undefined, and again one would not ask whether it is continuous there.

Exercises

3.3.1 For functions of the following types, indicate whether $f \circ g$, $g \circ f$, both, or neither can be defined.

(a) $f\colon \mathbf{R}^3 \to \mathbf{R}$, $\quad g\colon \mathbf{R} \to \mathbf{R}^2$.

(b) $f\colon \mathbf{R}^3 \to \mathbf{R}$, $\quad g\colon \mathbf{R}^2 \to \mathbf{R}$.

(c) $f\colon \mathbf{R}^3 \to \mathbf{R}^3$, $\quad g\colon \mathbf{R}^3 \to \mathbf{R}^2$.

(d) $f\colon \mathbf{R}^3 \to \mathbf{C}$, $\quad g\colon \mathbf{R}^2 \to \mathbf{R}^3$.

3.3.2 Prove that any linear function $f: \mathbf{R}^n \to \mathbf{R}^p$ is continuous. (This is not necessarily true of a linear function whose domain is an infinite-dimensional vector space such as $\mathcal{C}^2(-\infty, \infty)$.) Recommended steps:

(a) Show that because f is linear, if f is continuous at $\vec{x} = \vec{0}$ then it is continuous at any $\vec{x}$.

(b) Combine the definition of continuity with the definition of limit above it in the text (with $\vec{a}$ and $\vec{b}$ equal to $\vec{0}$, because of (a)) to get a precise ϵ-δ statement of the theorem to be proved.

(c) THE MAIN STEP: In the notation of the Remark above the definition of limit, show that $|f(\vec{x})|$ is less than some constant times $|\vec{x}|$. (The constant depends on the matrix A representing the linear function; it is independent of $\vec{x}$.)

(d) Appeal to the Remark itself to show that (c) proves what we wanted to prove (the statement in (b)).

3.3.3 Discuss the pitfalls of solving the equation

$$(\sin x - \sin y)z = (x^2 - y^2)$$

for z by computer. [Consider (1) guaranteeing uniqueness of the solution, and (2) writing a program that will execute properly (not crash) for all numerical values of the input parameters, x and y.]

3.3.4 Prove part (2) of Theorem 1.

3.3.5 Describe and sketch the surface defined parametrically by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \sin v \cos w \\ \sin v \sin w \\ \sqrt{2} \cos v \end{pmatrix}, \qquad 0 \le v \le \pi, \quad 0 \le w < 2\pi.$$

3.3.6 Describe and sketch the surface defined parametrically by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (2 + \frac{1}{2}\cos\theta)\cos\phi \\ (2 + \frac{1}{2}\cos\theta)\sin\phi \\ \frac{1}{2}\sin\theta \end{pmatrix}, \qquad 0 \le \theta \le 2\pi, \quad 0 \le \phi \le 2\pi.$$

3.3.7 Sketch these vector fields in $\mathbf{R}^2$. (Pick at least 4 diverse points $\vec{r}$ and draw the arrow $f(\vec{r})$ with its tail at $\vec{r}$. Choose the field of view as $-1 \le x \le 2$, $-1 \le y \le 2$.)

(a) $F(x, y) = \begin{pmatrix} x \\ y \end{pmatrix}$ (b) $F(x, y) = \begin{pmatrix} y \\ x \end{pmatrix}$ (c) $F(x, y) = \begin{pmatrix} -y \\ x \end{pmatrix}$

In the remaining exercises, (a) **identify** n **and** p **in the characterization**

$$f\colon \mathbf{R}^n \hookrightarrow \mathbf{R}^p;$$

(b) **state the largest natural domain** $\mathcal{D}$ **(a subset of** $\mathbf{R}^n$**) of the function** f**; and** (c) **find all points (in** $\mathcal{D}$ **or on the boundary of** $\mathcal{D}$**) where the function fails to be continuous.**

3.3.8 $f(x,y) = \begin{pmatrix} \dfrac{1}{x^2+y^2} \\ \dfrac{}{\sqrt{|x^2-y^2|}} \end{pmatrix}$

3.3.9 $f(x,y,z) = \begin{pmatrix} \dfrac{1}{x-y} \\ \dfrac{x^2-z^2}{x-z} \\ \dfrac{1}{\sqrt{y-z}} \end{pmatrix}$

3.3.10 $f(u,v) = \begin{pmatrix} \tan(u+v) \\ \tanh(u+2v) \\ \ln(u+3v) \end{pmatrix}$ $\quad \left(\tanh z \equiv \dfrac{\sinh z}{\cosh z} = \dfrac{e^z - e^{-z}}{e^z + e^{-z}}\right)$

3.4 Differentials

In an earlier section (2.4) we reviewed the relation between the partial derivatives of a function $f\colon \mathbf{R}^2 \to \mathbf{R}$ and (a) the tangent plane to the graph of f at the point concerned; (b) directional derivatives of f; (c) the gradient vector, ∇f. Linear algebra makes these connections easier to understand and points the way to the generalization to higher dimensions ($f\colon \mathbf{R}^n \to \mathbf{R}^p$).

First, let's review the fundamental idea of the differential calculus of functions $f\colon \mathbf{R}^1 \to \mathbf{R}^1$ from a point of view that lends itself to a multivariable generalization (and a geometrical interpretation). Let us pose the problem of approximating $f(x)$ for x near some fixed x_0 by an *affine* function,

$$\hat{f}(x) = y_0 + m(x - x_0), \tag{1}$$

where y_0 and m are constants. Which such function is the "best" approximation to f near x_0? A reasonable criterion is that

$$f(x) - \hat{f}(x) = |x - x_0|\epsilon, \tag{2}$$

where ϵ is a function of x which approaches 0 as $x \to x_0$. In other words,

$$\lim_{x \to x_0} \frac{f(x) - y_0 - m(x - x_0)}{|x - x_0|} = 0. \tag{3}$$

If you succeed in finding numbers y_0 and m with this property, then any other choice of y_0 or m will *not* work; the error would not get small. This justifies the claim that $\hat{f}$ is the *best* approximation, and shows that it is unique, if it exists. It is clear that the right choice for y_0 is $f(x_0)$ (if f is continuous at x_0; otherwise (2) can't be satisfied). Then rearrangement of (3) leads to

$$m = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \equiv f'(x_0) \equiv \left.\frac{dy}{dx}\right|_{x_0}. \tag{4}$$

This is the standard definition of the derivative. Thus an affine approximation $\hat{f}$ at x_0 exists if and only if f is *differentiable* at x_0, and the 1×1 matrix of the linear part of $\hat{f}$ is $f'(x_0)$. Also, $\hat{f}$ is the function whose graph is the tangent line to the graph of f at x_0. (Note that, indeed, any other straight line would be a *worse approximation* (near x_0) to the curve.)

In the traditional notation of *increments* and *differentials*,

$$\Delta y \equiv f(x) - y_0\,, \qquad dy \equiv \hat{f}(x) - y_0\,, \qquad \Delta x = dx \equiv x - x_0\,,$$

(1) becomes

$$dy = m\,dx = \frac{dy}{dx}\,dx, \tag{5}$$

and (2) becomes

$$\Delta y = dy + |x - x_0|\epsilon = \frac{dy}{dx}\,\Delta x + o(\Delta x), \tag{6}$$

where "$o(\Delta x)$" is a notation for any quantity which approaches 0 *faster* than Δx itself does. (Typically this term will behave like $(\Delta x)^2$, and for that reason is also written "$O\big((\Delta x)^2\big)$". It is the thing previously called $|x - x_0|\epsilon$.) Note that all we really do by introducing the Δ–d notation is to perform a coordinate transformation which moves the origin from $(0, 0)$ to (x_0, y_0).

To understand the notation that is about to be introduced to handle the multivariable case, it is important to appreciate three things about the differential notation in single-variable calculus:

- Δy and dy are *functions* of dx, although that argument variable is almost never indicated explicitly in the notation. (The differential, or Leibnitz, notation is part of that long and pragmatically expedient tradition in which the logical distinction between a function and the variable that is the value of that function is swept under the rug.)
- The function *dy depends on the base point* x_0 as a parameter. (Each x_0 has its own tangent line.) Again, one usually doesn't indicate x_0 in the notation, but sometimes clarity calls for writing something like $d_{x_0}y$ or $d_{x_0}f$ (which, we reiterate, is a function of another variable, dx).
- The differential dy or df is not the same thing as the derivative, $\frac{dy}{dx}$ or f'. Nevertheless, the latter is the numerical coefficient, or 1×1 matrix, that specifies the former as a linear function from $\mathbf{R}$ to $\mathbf{R}$. So, the two are so closely related, in a one-to-one fashion, that they are "effectively" the same. Nevertheless, dy as a physical quantity has the same units or dimensions as y (length, for example), while the units of $\frac{dy}{dx}$ are those of y divided by those of x (for example, length over time, or velocity, if x is a time).

Our task is to generalize this framework of thought to multivariable functions,* $\vec{f}\colon \mathbf{R}^n \to \mathbf{R}^p$. We seek an affine function

$$\hat{f}(\vec{x}) = \vec{y}_0 + L(\vec{x} - \vec{x}_0) \tag{1'}$$

that is the best approximation to $\vec{f}(\vec{x})$ near $\vec{x}_0$. Here $L \equiv d_{\vec{x}_0}\vec{f}$ is a *linear* function from $\mathbf{R}^n$ to $\mathbf{R}^p$, called the *differential* of $\vec{f}$ at $\vec{x}_0$. The condition for the best approximation is that

$$\vec{f}(\vec{x}) - \hat{f}(\vec{x}) = \|\vec{x} - \vec{x}_0\|\,\vec{\epsilon}, \tag{2'}$$

where $\vec{\epsilon}$ is a function of $\vec{x}$ and $\|\vec{\epsilon}\| \to 0$ as $\vec{x} \to \vec{x}_0$. As before, $\vec{y}_0$ must be $\vec{f}(\vec{x}_0)$ if this is to hold. Then (2′) can be rewritten as

$$\lim_{\vec{x}\to\vec{x}_0} \frac{\vec{f}(\vec{x}) - \vec{f}(\vec{x}_0) - L(\vec{x} - \vec{x}_0)}{\|\vec{x} - \vec{x}_0\|} = \vec{0}. \tag{3'}$$

* This discussion also applies to functions $\vec{f}\colon \mathbf{R}^n \hookrightarrow \mathbf{R}^p$ that are not everywhere defined, so long as $\vec{x}_0$ is in the interior of the domain of $\vec{f}$ (that is, for some $\delta > 0$, $\|\vec{x} - \vec{x}_0\| < \delta$ implies that $\vec{f}(\vec{x})$ is defined).

If such an L exists, it is unique. If it exists, $\vec{f}$ is called *differentiable* at $\vec{x}_0$.

SPECIAL CASE 1: $f\colon \mathbf{R} \to \mathbf{R}^p$. As x varies, $\vec{f}(x)$ traces a curve in p-dimensional space. The derivative vector $\vec{f}\,'(x_0)$ is the $p \times 1$ matrix of L. Recall (Sec. 1.4) that this vector is tangent to the curve, with length proportional to the speed of the parametrization, and that when multiplied by a number $dx \equiv x - x_0$, it yields a vector $d\vec{y}$ which tells approximately how $\vec{f}(x)$ is displaced from $\vec{y}_0$.

SPECIAL CASE 2: $f\colon \mathbf{R}^n \to \mathbf{R}$. The graph of f is an n-dimensional "surface" in $(n+1)$-dimensional space. The graph of $\hat{f}$ is the n-dimensional *hyperplane* tangent to the surface at $(\vec{x}_0, y_0)$. (See Sec. 2.4.) For a given vector $d\vec{x} \equiv \vec{x} - \vec{x}_0$, the number $dy \equiv L(d\vec{x})$ tells approximately how much $f(\vec{x})$ differs from $y_0 \equiv f(\vec{x}_0)$, the graph being approximated by its tangent plane. Recall that this number dy is also called the *directional derivative* of f along $d\vec{x}$ at $\vec{x}_0$. When $d\vec{x}$ points along a coordinate axis and $\|d\vec{x}\| = 1$, then $L(d\vec{x})$ is a partial derivative. (To see this, think of what $(3')$ becomes when $\vec{x} - \vec{x}_0 = h\vec{e}_j$.) Thus the gradient $\nabla f(x_0, y_0)$ is the $1 \times n$ matrix of L. Writing out $L(d\vec{x})$ as a matrix product, we have [equivalent to $(1')$]

$$dy = \sum_{j=1}^{n} \frac{\partial y}{\partial x_j}\, dx_j \equiv \nabla y \cdot d\vec{x},$$

the classical expression for the differential of a function of several variables.

GENERAL CASE: $f\colon \mathbf{R}^n \to \mathbf{R}^p$. The matrix of L is called the *Jacobian matrix* or *derivative matrix*:

$$(4')_1 \qquad \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1} & \cdots & \frac{\partial f_p}{\partial x_n} \end{pmatrix} \equiv \vec{f}\,'(\vec{x}_0) \equiv \left.\frac{d\vec{y}}{d\vec{x}}\right|_{\vec{x}_0} \equiv J_{\vec{x}_0}\vec{f}.$$

Here all the partial derivatives are evaluated at $\vec{x}_0$. (As indicated, a variety of notations for this matrix are in use. The notation $\frac{d\vec{y}}{d\vec{x}}$ is nonstandard but sems to be the best analogue of the single-variable derivative notation. To make matters worse, some notations and terms that we prefer to reserve for the *determinant* of the Jacobian matrix (see Sec. 7.3) are sometimes used for the matrix itself; be alert when reading other authors.)

One can *define* the partial derivatives as the elements of the Jacobian matrix (it being defined as the matrix which represents the linear piece in

the best affine approximation to $\vec{f}$), then prove that they are given by the classical formula

$$(4')_2 \qquad \left.\frac{\partial f_j}{\partial x_k}\right|_{\vec{x}} = \lim_{h\to 0}\frac{1}{h}\big[f_j(x_1,\dots,x_k+h,\dots)-f_j(x_1,\dots,x_k,\dots)\big].$$

To review,

$$\hat{f}(\vec{x}) \equiv \vec{f}(\vec{x}_0) + L(\vec{x}-\vec{x}_0)$$

is the *best affine approximation* to $\vec{f}(\vec{x})$ for $\vec{x}$ in the neighborhood of $\vec{x}_0$. The condition of giving the best approximation uniquely defines the linear function L, and its matrix can be calculated by $(4')$.

IMPORTANT REMARK: In applied work, the best affine approximation is usually called the *first-order approximation*. It is thought of as comprising the first two terms in a Taylor series approximation to the function (cf. Sec. 8.2).

If $\vec{u}$ is a vector in $\mathbf{R}^n$, then $L(\vec{u})$ is a vector in $\mathbf{R}^p$. It is the *directional derivative* of the vector-valued function f along $\vec{u}$ at $\vec{x}_0$. It tells how f changes as one moves away from $\vec{x}_0$ in the direction of $\vec{u}$ at a "speed" equal to the length of $\vec{u}$. (As in the scalar case, nothing is really lost by considering only vectors $\vec{u}$ of unit length.)

APPLICATION:

Intrepid Roger Rapidrudder measured the velocity vector of the air at his plane to be $\vec{v} = (200, 300, -500)$ feet per second, and the partial derivatives of the velocity to be (in (feet/second)/foot)

$$\nabla v_1 = (2,3,5), \qquad \nabla v_2 = (-5,3,2), \qquad \nabla v_3 = (10,-1,0);$$

that is, $d\vec{v}$ has the matrix

$$\begin{pmatrix} 2 & 3 & 5 \\ -5 & 3 & 2 \\ 10 & -1 & 0 \end{pmatrix}.$$

(The x_1 axis points east, x_2 axis points north, x_3 axis points up.) "Gee," Roger said, "I'd be a lot less likely to crash if I were 200 feet east of here, where, according to my *best affine approximation*, the velocity vector is approximately ... " [*Let's finish the sentence.*]

The displacement to the new point is

$$\Delta\vec{x} \equiv \vec{x} - \vec{x}_0 = \begin{pmatrix} 200 \\ 0 \\ 0 \end{pmatrix}.$$

Therefore, the air velocity at the new point is

$$\begin{aligned} \vec{v}(\vec{x}) &\approx \vec{v}(\vec{x}_0) + d_{\vec{x}_0}\vec{v} = \vec{v}(\vec{x}_0) + \frac{d\vec{v}}{d\vec{x}}\,(\Delta\vec{x}) \\ &= \begin{pmatrix} 200 \\ 300 \\ -500 \end{pmatrix} + \begin{pmatrix} 2 & 3 & 5 \\ -5 & 3 & 2 \\ 10 & -1 & 0 \end{pmatrix} \begin{pmatrix} 200 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 600 \\ -700 \\ +1500 \end{pmatrix}. \end{aligned}$$

The positive sign on the vertical component indicates that the dangerous downdraft has indeed been eliminated. (One might remark, however, that the air must not be very turbulent if Roger trusts an affine (first-order) approximation over a distance of 200 feet! Such are the hazards of dressing up simple numerical exercises as serious applications. A more realistic, if more abstract, application will be presented very soon.)

In science and engineering the generic mathematical notations $d_{\vec{x}}\vec{v}$, $J_{\vec{x}}\vec{v}$, and $\frac{d\vec{v}}{d\vec{x}}$ are not usually used in connection with vector fields like the one in the foregoing example. Instead, the notation $\nabla\vec{v}$ and the term *gradient of the vector field* are used. In numerical terms the gradient $\nabla\vec{v}$ is a matrix (the same thing as $J_{\vec{x}}\vec{v}$), whose rows are the gradients of the scalar functions ∇v_1 , etc. It is, therefore, a "two-index" object in the sense that a vector is a one-index object. One index (the column index of the matrix, the index attached to "∇") labels the coordinate with respect to which one is differentiating, while the other (the row index, the index attached to "$\vec{v}$") labels the vector component that is being differentiated. Of special importance is the fact that a quantity of the form $\vec{u}\cdot\nabla\vec{v}$ is again a vector, the directional derivative of $\vec{v}$ along $\vec{u}$. In this dot product it is the ∇ index, *not* the $\vec{v}$ index, that participates:

$$[\vec{u}\cdot\nabla\vec{v}]_j = \sum_{k=1}^{3} u_k \frac{\partial v_j}{\partial x_k} \neq \sum_{k=1}^{3} u_k \frac{\partial v_k}{\partial x_j}.$$

Back to the general theory: Although it is hard to visualize the graphs of functions when there are several independent and dependent variables, it is worth pausing to consider how the present situation generalizes the

previously discussed low-dimensional cases, which we *can* visualize. (Low-dimensional analogies are helpful in understanding high-dimensional problems.) Note first that the graph of L is an n-dimensional subspace of $\mathbf{R}^{n+p}$. ("Above" each point $\vec{x} \in \mathbf{R}^n$ sits the point $\big(\vec{x}, L(\vec{x})\big)$ on the graph.) Therefore, the graph of $\hat{f}$ is an n-dimensional affine subspace of $\mathbf{R}^{n+p}$. It is *tangent* at the point $\big(\vec{x}_0, f(\vec{x}_0)\big)$ to the graph of f, which is an n-dimensional curved *hypersurface* in $\mathbf{R}^{n+p}$.

Thus, of the three things mentioned in the first paragraph of this section, the analogue of the tangent plane is the graph of the affine approximation $\hat{f}$; the analogue of the gradient vector is the differential (linear approximation) df, or the Jacobian matrix Jf that represents it; and the analogue of the directional derivative is the result of applying df to a vector. For physical vector fields $\vec{f}\colon \mathbf{R}^3 \hookrightarrow \mathbf{R}^3$, the last two are often actually called "gradient" and "directional derivative" and given corresponding notations, $\nabla \vec{f}$ and $\vec{u} \cdot \nabla \vec{f}$.

The notation of increments and differentials can be used in the multi-dimensional theory, too. Define

$$\Delta\vec{y} \equiv \vec{f}(\vec{x}) - \vec{y}_0\,, \qquad d\vec{y} \equiv \hat{f}(\vec{x}) - \vec{y}_0\,, \qquad \Delta\vec{x} = d\vec{x} \equiv \vec{x} - \vec{x}_0\,.$$

Then $(1')$ becomes

$$\begin{aligned} d\vec{y} &= L(d\vec{x}) \equiv d_{\vec{x}_0} f \\ &\equiv \frac{d\vec{y}}{d\vec{x}}\, d\vec{x} \qquad \text{(a matrix product)}, \end{aligned} \tag{5'}$$

or, in coordinates,

$$dy_j = \sum_{k=1}^{n} \frac{\partial y_j}{\partial x_k}\, dx_k \qquad (j = 1,\ \ldots,\ p). \tag{5''}$$

And $(2')$ becomes

$$\begin{aligned} \Delta\vec{y} &= d\vec{y} + \|\Delta\vec{x}\|\,\epsilon \\ &= \frac{d\vec{y}}{d\vec{x}}\, d\vec{x} + o(\Delta\vec{x}). \end{aligned} \tag{6'}$$

Here "$o(\Delta\vec{x})$" stands for an *error term* or *remainder* which vanishes *faster than $\Delta\vec{x}$ itself does* as $\Delta\vec{x}$ approaches 0. (The precise statement of this condition is $(3')$; $o(\Delta\vec{x})$ is the numerator of that expression.) For most

functions (more precisely, those whose *second-order* derivatives exist), the error term is of the type "$O\big(\|\Delta\vec{x}\,\|^2\big)$"; that is, it vanishes so fast that its norm is bounded by some constant times the quadratic quantity $\|\Delta\vec{x}\,\|^2$ when $\Delta\vec{x}$ is in a neighborhood of $\vec{0}$.

SUGGESTION: On one piece of paper, write out the six fundamental formulas (1)–(6) of single-variable calculus, and next to each write the analogous (primed) formula of multivariable calculus. If you understand all those formulas, then you understand this part of the course.

REMARK: Hoping not to be too fussy about it, we point out that there are good reasons to maintain the distinction between a differential, $d_{\vec{x}}f$, and the matrix of the differential, $J_{\vec{x}}f = \frac{df}{dx}$. Not only is there the ever-present distinction between a linear function and the matrix that represents it (a distinction which will become much more important in the next chapter when we consider changing coordinate systems). In the context of an application there is also a difference in physical units (dimensions). The value of df at a point is a physical quantity of the same type as the values of f itself. But the physical dimensions of Jf are those of f divided by those of $\vec{x}$, because it is only after Jf is multiplied by the column vector $\Delta\vec{x}$ that it becomes a quantity of the type of f. [EXAMPLE: If $f(t)$ is a velocity function (with units of meters per second), then $f'(t_0)$ is an acceleration (with units of meters per second per second), but $df \equiv f'(t_0)\,dt$ is a velocity again (albeit an "infinitesimally small" one). This point of physical bookkeeping underlines the metaphysical distinction between the slope of a tangent line and the tangent line itself.] In other words, when the differential of a function is written in conventional matrix notation, the word "differential" and the d notation should not be applied to the matrix of the differential, but only to the whole combination of the matrix and its argument vector; in contrast, for the differential itself, as a function, when there is no danger of confusion we are free (as usual in applied calculus) to blur the distinction between the function and the value of the function, and to either express or suppress the argument variable according to the convenience of the moment.

An understanding of the foregoing remark should make the following theorem appear less paradoxical. We leave the proof of the theorem to the exercises.

Theorem: The differential of a linear function $L\colon \mathbf{R}^n \to \mathbf{R}^p$ at any point is the function L itself:

$$d_{\vec{x}_0}L \text{ [evaluated at } \vec{x} - \vec{x}_0] = L(\vec{x} - \vec{x}_0).$$

A better example

The true utility of the notion of best linear approximation comes not in numerical approximations of the Roger Rapidrudder variety, but in more theoretical contexts. Here is a rather concrete situation where thinking in those terms produces a major simplification in comparison to a traditional calculation in terms of partial derivatives.

Let $M = \{m_{jk}\}$ be an $n \times n$ matrix, and $M^p = Q = \{q_{jk}\}$ one of its powers. (Here p is some fixed positive integer, such as 3.) We can write $Q = f(M)$; f is a matrix-valued function of a matrix variable, or, equivalently, a function from $\mathbf{R}^{n^2} \to \mathbf{R}^{n^2}$. Each of the n^2 matrix elements of Q is a function of all the n^2 matrix elements of M. We may set ourselves the project of calculating all the partial derivatives, such as $\partial q_{12}/\partial m_{11}$.

Let us see how we would do that by elementary methods, then see whether our more sophisticated new understanding suggests a better way of proceeding. To take a simple case, let $p = 2$ and $n = 2$. Then we can assign the matrix elements individual letters:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad M^2 = Q = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

A short calculation gives

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{pmatrix}.$$

We therefore have

$$\frac{\partial A}{\partial a} = 2a, \qquad \frac{\partial B}{\partial a} = b, \qquad \frac{\partial C}{\partial a} = c, \qquad \frac{\partial D}{\partial a} = 0,$$

and 12 other equally unilluminating formulas. Clearly, the calculation would be much messier if either p or n were greater.

Now let's consider the best linear approximation to the function $Q = f(M) = M^2$ as a whole. Look at

$$\begin{aligned} f(M + \Delta M) &= (M + \Delta M)(M + \Delta M) \\ &= M^2 + (\Delta M)M + M(\Delta M) + (\Delta M)^2 \\ &= f(M) + (\Delta M)M + M(\Delta M) + O\big((\Delta M)^2\big). \end{aligned}$$

The differential of f is the linear (first-degree) part of this expression:

$$d_M f = (\Delta M)M + M(\Delta M) \equiv (dM)M + M(dM).$$

To see the significance of this result, consider the special case

$$dM = \begin{pmatrix} h & 0 \\ 0 & 0 \end{pmatrix},$$

for which

$$\begin{aligned} d_M f &= h \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} + h \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} \\ &= h \begin{pmatrix} 2a & b \\ c & 0 \end{pmatrix}. \end{aligned}$$

In the language of directional derivatives, this is the directional derivative of M^2 along dM. In the language of partial derivatives, it is h times the matrix of partial derivatives of M^2 with respect to a. Comparing, we see that these are indeed the same 4 partial derivatives we found earlier, but calculated now in a simpler, more systematic, and more instructive way.

In particular, it is easy to generalize this calculation to higher dimensions ($n \times n$ matrices instead of 2×2), without getting into a notational morass. The generalization to higher powers p is slightly less trivial, but still quite tractable: For $f(M) = M^3$, for example, we have

$$d_M f = (dM)M^2 + M(dM)M + M^2(dM),$$

into which a simple special case for dM can be inserted easily.

Applications and examples

We repeat that any *first-order approximation* in science or engineering is an application of the theory of this section.

Example 1. *Electric field of a point charge.* A charged particle in three-dimensional space produces an electrical field that points directly away from the charge with magnitude proportional to the inverse square of the distance from the charge. Thus, if the origin of Cartesian coordinates is placed at the charge, and units are chosen so that the overall constant factor equals 1, then the electric field is

$$\vec{E}(\vec{r}) = \frac{\hat{r}}{r^2},$$

where $r = \|\vec{r}\| = \sqrt{x^2+y^2+z^2}$ is the distance and $\hat{r} = \vec{r}/\|\vec{r}\|$ is the unit vector pointing radially outward. Thus

$$\vec{E}(\vec{r}) = \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \begin{pmatrix} \frac{x}{(x^2+y^2+z^2)^{3/2}} \\ \frac{y}{(x^2+y^2+z^2)^{3/2}} \\ \frac{z}{(x^2+y^2+z^2)^{3/2}} \end{pmatrix}.$$

For most purposes it is easier to deal with this function in spherical coordinates (r, θ, ϕ) (which we will see frequently in the future). Today, however, for the sake of an example we will work directly in Cartesian coordinates.

Let us find the best linear (more properly, affine) approximation to $\vec{E}(\vec{r})$ at a point $\vec{r}_0$ (*not* $\vec{0}$, since $\vec{E}$ is obviously not differentiable there). For concreteness, consider

$$\vec{r}_0 = \hat{\imath} + .2\,\hat{\jmath} = \begin{pmatrix} 1 \\ 0.2 \\ 0 \end{pmatrix}.$$

According to the general formalism,

$$\vec{E}(\vec{r}_0 + \Delta\vec{r}) \approx \vec{E}(\vec{r}_0) + d_{\vec{r}_0}\vec{E},$$

where $d_{\vec{r}_0}\vec{E}$ is evaluated at

$$\Delta\vec{r} \equiv \vec{r} - \vec{r}_0 = \begin{pmatrix} x-1 \\ y-0.2 \\ z \end{pmatrix}.$$

The matrix of $d_{\vec{r}_0}\vec{E}$ (alias $\nabla\vec{E}(\vec{r}_0)$) is

$$J_{\vec{r}_0}\vec{E} = \begin{pmatrix} \frac{\partial E_x}{\partial x} & \frac{\partial E_x}{\partial y} & \cdots \\ \frac{\partial E_y}{\partial x} & \frac{\partial E_y}{\partial y} & \cdots \\ \cdots & & \end{pmatrix}.$$

The partial differentiations look rather tedious, but they can be greatly shortened by noting the symmetry of the problem. We calculate

$$\begin{aligned} \frac{\partial E_x}{\partial x} &= (x^2+y^2+z^2)^{-3/2} - \frac{3}{2}(2x^2)(x^2+y^2+z^2)^{-5/2} \\ &= (-2x^2+y^2+z^2)(x^2+y^2+z^2)^{-5/2}. \end{aligned}$$

The other diagonal elements, $\frac{\partial E_y}{\partial y}$ and $\frac{\partial E_z}{\partial z}$, can be obtained immediately by interchanging the role of x in this equation with that of y or z. The off-diagonal elements are easier:

$$\frac{\partial E_x}{\partial y} = -3xy(x^2+y^2+z^2)^{-5/2},$$

and the other 5 formulas are obvious by symmetry.

We note in passing that this Jacobian matrix is symmetric ($\frac{\partial E_x}{\partial y} = \frac{\partial E_y}{\partial x}$, etc.) and that the sum of the three diagonal elements is zero. Both of these are well known properties of the electric field at a point where no charge is located. Mathematically they arise because $\vec{E}$ is the gradient of a scalar function that satisfies Laplace's equation; see Chapter 7.

We can now evaluate $\vec{E}$ and its Jacobian matrix at $\vec{r}_0$:

$$r = \sqrt{1^2 + .2^2} \approx 1.02,$$

$$\vec{E}(\vec{r}_0) \approx \begin{pmatrix} .94 \\ .19 \\ .00 \end{pmatrix},$$

$$J_{\vec{r}_0}\vec{E} \approx \begin{pmatrix} -1.78 & -.54 & .00 \\ -.54 & .83 & .00 \\ .00 & .00 & .94 \end{pmatrix}.$$

The approximate-equality signs so far merely indicate that we have gone to low-precision floating-point arithmetic. Now we construct the linear approximation:

$$\vec{E}(\vec{r}) \approx \begin{pmatrix} .94 \\ .19 \\ .00 \end{pmatrix} + \begin{pmatrix} -1.78 & -.54 & .00 \\ -.54 & .83 & .00 \\ .00 & .00 & .94 \end{pmatrix} \begin{pmatrix} x - 1 \\ y - 0.2 \\ z \end{pmatrix}.$$

We could, of course, work out the matrix multiplication and algebraically simplify the expression; but for many purposes it is best to leave the formula in a form that exhibits explicitly its structure as a relation between small quantities (increments), $\Delta\vec{E} \approx d\vec{E}$ and $\Delta\vec{r}$.

WARNING: The most frequent student mistakes in calculating concrete linear approximations are the same as in their freshman analogue, the tangent-line problem. First, remember that in the formula

$$\vec{f}(\vec{x}) \approx \vec{f}(\vec{x}_0) + d_{\vec{x}_0}\vec{f} \equiv \vec{f}(\vec{x}_0) + \left.\frac{d\vec{f}}{d\vec{x}}\right|_{\vec{x}_0} (\vec{x} - \vec{x}_0),$$

the $\vec{x}$ variable in the Jacobian matrix $\frac{d\vec{f}}{d\vec{x}}$ must be replaced by the particular numerical vector $\vec{x}_0$, whereas in the vector $(\vec{x} - \vec{x}_0)$ that the matrix acts upon, the $\vec{x}$ remains a variable. Second, remember that in the latter case the components of $\vec{x}_0$ must be subtracted.

Example 2. *Polar coordinates.* The standard polar coordinate transformation in $\mathbf{R}^2$,

$$x = r\cos\theta,$$
$$y = r\sin\theta,$$

has the structure†

$$\vec{x} = f(\vec{u}), \qquad \vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \vec{u} = \begin{pmatrix} r \\ \theta \end{pmatrix}.$$

We can ask how x and y change as r and θ change slightly. Mathematically this is a problem of the sort we have been studying, although the physical interpretation of the function is quite different from the previous examples involving velocity and electric fields.

Let us find the best affine approximation to $\vec{x}$ when $\vec{u}$ varies in the neighborhood of the point

$$r_0 = \sqrt{2}, \quad \theta_0 = \frac{\pi}{4}; \qquad \vec{x}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Since we did not give a complete proof of the connection between affine approximations and partial derivatives [i.e., formulas (4′)], let us rederive that connection in this special case. As usual, we shift the origin of our coordinates to make the affine function truly linear:

$$\Delta\vec{u} = d\vec{u} = \vec{u} - \vec{u}_0\,, \qquad \Delta\vec{x} = \vec{x} - \vec{x}_0\,.$$

Then the "best" linear function must have the form

$$\Delta\vec{x} \approx d\vec{x} = \vec{a} + B\Delta\vec{u}$$

for some vector $\vec{a}$ and matrix B. Our claim is that

$$\vec{a} = 0, \qquad B = \left.\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix}\right|_{\vec{u}_0} = \left.\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}\right|_{\substack{r=\sqrt{2},\\ \theta=\frac{\pi}{4}}} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -1 \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix}.$$

The criterion for choosing $\vec{a}$ and B is that as $\Delta\vec{u}$ goes to 0, $\Delta\vec{x} - d\vec{x}$ should go to 0 "even faster".

† In this context the notation $\vec{r}$ for the Cartesian coordinate vector would probably cause more confusion than it cures, so we won't use it.

Continuity of f shows immediately that $\vec{a}$ must be 0. It remains to compare

$$d\vec{x} = B\,d\vec{u} = \begin{pmatrix} B_{11}\,dr + B_{12}\,d\theta \\ B_{21}\,dr + B_{22}\,d\theta \end{pmatrix}$$

with

$$\Delta\vec{x} = \begin{pmatrix} (r_0 + dr)\cos(\theta_0 + d\theta) \\ (r_0 + dr)\sin(\theta_0 + d\theta) \end{pmatrix} - \begin{pmatrix} r_0\cos\theta_0 \\ r_0\sin\theta_0 \end{pmatrix}.$$

Knowing about derivatives and Taylor approximations for single-variable functions, you could now complete this calculation quickly; but you were promised a rederivation from first principles, so let's go all the way. Note that

$$|dr|,\ |d\theta| \le \sqrt{(dr)^2 + (d\theta)^2} = \|\Delta\vec{u}\|.$$

Thus

$$\frac{\Delta\vec{x} - d\vec{x}}{\|\Delta\vec{u}\|} = \frac{1}{\|\Delta\vec{u}\|}\begin{pmatrix} (r_0 + dr)\cos(\theta_0 + d\theta) - r_0\cos\theta_0 - B_{11}\,dr - B_{12}\,d\theta \\ (r_0 + dr)\sin(\theta_0 + d\theta) - r_0\sin\theta_0 - B_{21}\,dr - B_{22}\,d\theta \end{pmatrix}$$

$$= \frac{dr}{\|\Delta\vec{u}\|}\begin{pmatrix} \frac{(r_0+dr)\cos(\theta_0+d\theta) - r\cos(\theta+d\theta)}{dr} - B_{11} \\ \frac{(r_0+dr)\sin(\theta_0+d\theta) - r\sin(\theta+d\theta)}{dr} - B_{21} \end{pmatrix}$$

$$+ \frac{d\theta}{\|\Delta\vec{u}\|}\begin{pmatrix} \frac{r\cos(\theta+d\theta) - r\cos\theta}{d\theta} - B_{12} \\ \frac{r\sin(\theta+d\theta) - r\sin\theta}{d\theta} - B_{22} \end{pmatrix}$$

(with appropriate modification if dr or $d\theta$ equals 0). As dr and $d\theta$ approach 0, the large fractions approach the expected partial derivatives by their classical definition $(4')_2$. Meanwhile, the factors $dr/\|\Delta\vec{u}\|$ and $d\theta/\|\Delta\vec{u}\|$ remain smaller than 1 (but don't approach 0). Therefore, if we choose the elements of B to be those partial derivatives,

$$B_{11} = \left.\frac{\partial x}{\partial r}\right|_{\vec{u}_0}, \quad \text{etc.},$$

and *only* in that case, we get

$$\frac{\Delta\vec{x} - d\vec{x}}{\|\Delta\vec{u}\|} \to 0.$$

That is,

$$\Delta\vec{x} - d\vec{x} = \|\Delta\vec{u}\|\vec{\epsilon} \quad \text{where } \|\vec{\epsilon}\| \to 0 \text{ as } d\vec{u} \to 0.$$

In the particular numerical case at hand, we get

$$\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \approx \begin{pmatrix} \frac{1}{\sqrt{2}} & -1 \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix}\begin{pmatrix} \Delta r \\ \Delta\theta \end{pmatrix}.$$

(Draw a sketch in the x–y plane to see that this is exactly right.)

Example 3. *Lagrangian mechanics.* Affine approximations and differentials can be constructed in infinite-dimensional spaces, although there are massive technical complications there that we shall need to sweep under the rug.

If a particle is subject to a force derived from a potential-energy function $V(x)$, then its position $x(t)$ satisfies the equation of motion (*Newton's second law*)

$$m\,\frac{d^2x}{dt^2} + \frac{dV}{dx} = 0.$$

(For notational simplicity we consider motion in only one dimension; m is the mass of the particle.) The *Lagrangian function* of the particle is defined as the difference of the kinetic and potential energies, and its *action* over a time interval (t_1, t_2) is the time integral of the Lagrangian over the course of the motion:

$$S(x; t_1, t_2) \equiv \int_{t_1}^{t_2} L\big(x(t), \tfrac{dx}{dt}\big)\,dt \equiv \int_{t_1}^{t_2} \left[\frac{m}{2}\left(\frac{dx}{dt}\right)^2 - V(x(t))\right] dt.$$

Thus x as a function belongs to the space $\mathcal{C}^2(t_1, t_2)$ of twice-differentiable functions on the interval $t_1 \le t \le t_2$, and S is a nonlinear function of the type $S{:}\,\mathcal{C}^2(t_1, t_2) \to \mathbf{R}$. We can ask how S changes when x is replaced by a "nearby" function $x(t) + \Delta x(t)$ that agrees with x at the initial and final times (hence $\Delta x(t_1) = 0 = \Delta x(t_2)$).

As in the calculation with a function on matrices earlier in this section, we can substitute $x + \Delta x$ into the formula for S and discard all the terms that are of higher than linear order. We get

$$S(x + \Delta x) = \int_{t_1}^{t_2} \left[\frac{m}{2}\left(\frac{dx}{dt}\right)^2 + m\,\frac{dx}{dt}\,\frac{d\Delta x}{dt} + \dots \right.$$
$$\left. - V(x(t)) - \frac{dV(x(t))}{dx}\,\Delta x(t) + \dots\right] dt$$

$$= \int_{t_1}^{t_2} \left\{ \left[\frac{m}{2}\left(\frac{dx}{dt}\right)^2 - V(x(t))\right] + \left[m\,\frac{dx(t)}{dt}\right]\frac{d\Delta x(t)}{dt} \right.$$
$$\left. - \left[\frac{dV}{dx}\right]\Delta x + O\big((\Delta x)^2\big)\right\} dt.$$

To see exactly how the $\frac{d\Delta x}{dt}$ term depends on Δx, we need to integrate that term by parts, getting

$$S(x+\Delta x) = S(x) - \int_{t_1}^{t_2} \left[m\,\frac{d^2x(t)}{dt^2} + \frac{dV(x(t))}{dx} \right] \Delta x(t)\,dt + O\big((\Delta x)^2\big).$$

(The condition $\Delta x(t_1) = 0 = \Delta x(t_2)$ is used to get rid of the endpoint terms in the integration by parts.)

In particular, $S(x; t_1, t_2)$ will be *stationary under infinitesimal variations* of the path $x(t)$ (i.e., $d_x S = 0$ for all functions Δx) provided that the quantity in the square brackets vanishes for that particular path:

$$m\,\frac{d^2x(t)}{dt^2} + \frac{dV(x(t))}{dx} = 0.$$

But this is precisely the equation of motion that we began from! Stationarity of the action can be taken as the fundamental axiom of mechanics, in place of Newton's law of motion, to which it is equivalent.

More generally, the vanishing of dS is a criterion for a possible maximum or minimum of S as a functional of the path x, just as vanishing of ∇f signals a possible extremum of a function $f\colon \mathbf{R}^n \to \mathbf{R}$. The formulation of mechanics in terms of stationarity of the action is historically called *the principle of least action*, because in most physical theories the stationary points do turn out to be minima.

Exercises

3.4.1 The function $f\colon \mathbf{R}^3 \to \mathbf{R}^2$ is defined by

$$\begin{pmatrix} u \\ v \end{pmatrix} = f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x^2 - y^2 \\ 2xy + z \end{pmatrix}. \qquad \text{Let} \quad \vec{x}_0 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

(a) Find the matrix ($J_{\vec{x}_0} f$ or $\frac{df}{d\vec{x}}$) representing the differential of this function, $d_{\vec{x}_0} f$.

(b) Use the result of (a) to find an approximate value for $f\begin{pmatrix} 1.1 \\ -.2 \\ -.9 \end{pmatrix}$.

3.4.2 Consider the function $h\colon \mathbf{R}^2 \to \mathbf{R}^2$ defined by $\begin{cases} u = \cos x + \sin y, \\ v = \sin x - \cos(2y). \end{cases}$

(a) Find the best affine approximation to h for (x, y) near (π, π).

(b) Use (a) to approximate u and v for $x = 1.1\pi$, $y = 0.8\pi$.

(c) Find a vector that points in the direction of fastest increase of u at (π, π).

3.4.3 This problem centers upon the formulas $\begin{cases} x = \cosh u \cos v, \\ y = \sinh u \sin v. \end{cases}$ Consider the function $F\colon \mathbf{R}^2 \to \mathbf{R}^2$, $\begin{pmatrix} x \\ y \end{pmatrix} = F(u, v)$, defined by these formulas.

(a) Let $\vec{u}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Calculate the differential, $d_{\vec{u}_0}F$ (represented by the Jacobian matrix, $\frac{dF}{du}\big|_{u=\vec{u}_0}$).

(b) Find the best affine approximation to F at $\vec{u}_0$.

3.4.4 Roger Rapidrudder measured the gradient vector of the air pressure at his cockpit to be $\nabla p = (2, 1, -1)$.

(a) Find the directional derivative of the pressure in the direction $\hat{u} = \frac{1}{\sqrt{2}}(1, 1, 0)$.

(b) A more pedantic author would write one of these two vectors (∇p and $\hat{u}$) as a column vector, not a row. Which one, and why?

3.4.5 An atom near the point $\vec{r} = \vec{r}_0 \equiv (1, 2, 3)$ is subject to an electric field $\vec{E} = \begin{pmatrix} x^2 - y^2 \\ x^2 + y^2 \\ z \end{pmatrix}$. Find the *best affine approximation* to $\vec{E}(\vec{r})$ for $\vec{r}$ near $\vec{r}_0$.

3.4.6 Roger Rapidrudder's altimeter measured the elevation z of the ground beneath him to be

$$z = f(x, y) = 6 - x^2 - y^2$$

(where x and y are the coordinates in the east and north directions.)

(a) Find the affine function that best approximates f near $(2, 0)$.

(b) Use your answer in (a) to *approximate* the elevation at $(2.1, -.2)$.

(c) Sketch the graphs of the true elevation and the affine approximation.

3.4.7 Roger Rapidrudder's microelectronic satellite navigation system gave his position as $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$, and his velocity meter read $\begin{pmatrix} 200 \\ 40 \\ -1 \end{pmatrix}$, at the moment when he entered the powerful magnetic field of the Iraqi anti-aircraft defense system. The CIA has determined that this field is

$$\vec{B}(\vec{r}) = \begin{pmatrix} e^{-z}\cos(\pi y) \\ e^{-z}\sin(\pi x) \\ x^2 + y^2 \end{pmatrix},$$

so Roger knew that the directional derivative of the field at his plane, in the direction of his motion, was ... what?

3.4.8 A solenoid in Roger's left engine creates a magnetic field

$$\vec{B}(\vec{r}) = \begin{pmatrix} y^2 - z^2 \\ z^2 - x^2 \\ x^2 - y^2 \end{pmatrix}.$$

(From now on in this problem column vectors are written as rows to save space.)

(a) Find the *best affine approximation* to $\vec{B}$ in the vicinity of the point $\vec{r}_0 = (1, 0, 1)$.

(b) Find the directional derivative of $\vec{B}$ at $\vec{r}_0$ in the direction of the unit vector $\vec{v} \equiv \frac{1}{\sqrt{10}}(1, -3, 0)$.

(c) Use your answer to (a) *or* (b) to estimate $\vec{B}(\vec{r})$ at $\vec{r} = (2, -3, 1)$, after noting that $\vec{B}(\vec{r}_0) = (-1, 0, 1)$. (*Don't* calculate the exact value!)

3.4.9 At a certain point in a gas, the pressure is measured as 760 torr, and the pressure gradient as $(-2, 2, 1)$ torr/cm.

(a) What is the directional derivative of the pressure in the direction of the vector $\vec{w} = (1, 0, 1)$?

(b) What is the best affine approximation to the pressure at a point 0.03 centimeters away along the direction of $\vec{w}$?

(c) Estimate (using the affine approximation) how far away from the point of measurement we can go along $\vec{w}$ before the pressure changes by 2%.

(d) Answer part (c) with "along $\vec{w}$" omitted. (I.e., give the answer for the direction in which the pressure is expected to change most rapidly.)

3.4.10 (a) Prove the theorem that $d_{\vec{x}_0} L = L$ if L is linear. What does this say about the relation between the matrix representing L and the partial derivatives of L?

(b) Formulate and prove the corresponding theorem about affine functions.

3.5 The Chain Rule

In Chapter 2 we looked at the chain rule for composite functions of the type $g \circ f \colon \mathbf{R} \xrightarrow{f} \mathbf{R}^m \xrightarrow{g} \mathbf{R}$. The formula for $(g \circ f)'$ contains one term for each component of the intermediate vector variable. If one or both of the other two variables is also multidimensional, the principle is the same; we get formulas like

$$\frac{\partial z_j}{\partial x_l} = \sum_{k=1}^{m} \frac{\partial z_j}{\partial y_k} \frac{\partial y_k}{\partial x_l} \, .$$

This looks suspiciously like matrix multiplication. Our task now is to understand these formulas from a geometrical or linear-algebraic point of view, using the concept of a differential.

We shall need to appeal to the fact that the value of a linear function (when it isn't zero) is of roughly the same size as its argument, as the argument gets very small or very large:

Lemma: If $L \colon \mathbf{R}^n \to \mathbf{R}^p$ is linear, then $\|L(\vec{x})\| \leq C \, \|\vec{x}\|$, where C is a constant (which can depend on L but not on $\vec{x}$). In other words, $L(\vec{x}) = O(\vec{x})$.

Proof: Use the matrix representation:

$$|L(\vec{x})_j| = \left| \sum_{k=1}^{n} A_{jk} x_k \right| \leq \left[\sum_{k=1}^{n} (A_{jk})^2 \right]^{\frac{1}{2}} \left[\sum_{k=1}^{n} x_k{}^2 \right]^{\frac{1}{2}} = \left[\sum_{k=1}^{n} (A_{jk})^2 \right]^{\frac{1}{2}} \|\vec{x}\|.$$

(In the second step we have used a well-known inequality for numerical sums. It is an instance of the *Cauchy–Schwarz inequality*, discussed in Sec. 6.1:

Regard the sum over k as a dot product, keeping j fixed.) Square these components, sum, and take the square root:

$$\|L(\vec{x})\| \le \left[\sum_{j=1}^{p}\sum_{k=1}^{n}(A_{jk})^2\right]^{\frac{1}{2}}\|\vec{x}\|.$$

As promised, the mess inside the brackets is independent of $\vec{x}$.

Corollary: For a differentiable function, $\vec{y} = f(\vec{x})$, with differential $d_{\vec{x}_0}f \equiv L$, we have

$$\Delta\vec{y} = L(\Delta\vec{x}) + o(\Delta\vec{x}) = O(\Delta\vec{x}).$$

(Recall that this means that

$$\frac{\|\Delta\vec{y}\|}{\|\Delta\vec{x}\|} \le C \quad \text{for some constant } C$$

at least when Δx is sufficiently small. An example of a function which *fails* to be of order $O(\Delta\vec{x})$ would be $\|\Delta\vec{x}\|^{\frac{1}{2}}$ (times some constant vector field, if you want to make it a vector-valued function).)

Armed with this technical tool, we can investigate the differential of a *composite function*, $g \circ f$, when f is a function from $\mathbf{R}^n$ to $\mathbf{R}^m$ and g is a function from $\mathbf{R}^m$ to $\mathbf{R}^p$. Then $g \circ f$ is a function from $\mathbf{R}^n$ to $\mathbf{R}^p$, defined by

$$\big(g \circ f\big)(\vec{x}) \equiv g\big(f(\vec{x})\big).$$

In other words,

$$\vec{y} = f(\vec{x}) \quad \text{and} \quad \vec{z} = g(\vec{y}) \quad \text{imply} \quad \vec{z} = \big(g \circ f\big)(\vec{x}).$$

Case 1: f and g are linear and are represented by matrices A_f and A_g. Then we know that

$$A_{g\circ f} = A_g A_f\,. \qquad (*)$$

Also, we saw in the previous section that the differential of a linear function is the function itself. Therefore, A_f and A_g are also the Jacobian matrices of f and g respectively, at any values of their arguments. That is,

$$(A_f)_{kl} = \frac{\partial y_k}{\partial x_l}\,, \qquad (A_g)_{jk} = \frac{\partial z_j}{\partial y_k}\,.$$

Thus $(*)$ is an instance of the chain rule:

$$z_j = \sum_{k=1}^{m} \frac{\partial z_j}{\partial y_k} \sum_{l=1}^{n} \frac{\partial y_k}{\partial x_l}\, x_l\,,$$

or

$$(A_{g\circ f})_{jl} = \frac{\partial z_j}{\partial x_l} = \sum_{k=1}^{m} \frac{\partial z_j}{\partial y_k}\, \frac{\partial y_k}{\partial x_l}\,.$$

CASE 2: f and g are affine (constant + linear):

$$\begin{aligned} \vec{y} = f(\vec{x}) &= \vec{y}_0 + L_f(\vec{x} - \vec{x}_0) \qquad \big(\vec{y}_0 = f(\vec{x}_0)\big), \\ \vec{z} = g(\vec{y}) &= \vec{z}_0 + L_g(\vec{y} - \vec{y}_0) \qquad \big(\vec{z}_0 = g(\vec{y}_0)\big). \end{aligned}$$

Thus

$$\begin{aligned} \vec{z} = g\big(f(\vec{x})\big) &= \vec{z}_0 + L_g\big(L_f(\vec{x} - \vec{x}_0)\big) \\ &= \big(g \circ f\big)(\vec{x}_0) + \big(L_g \circ L_f\big)(\vec{x} - \vec{x}_0). \end{aligned}$$

Which is to say that

$$\Delta\vec{z} = \big(L_g \circ L_f\big)(\Delta\vec{x}).$$

Again, written out in matrix language this is a special case of the chain rule. (The matrix elements of the Ls are partial derivatives.)

CASE 3: f and g are nonlinear but differentiable:

$$\begin{aligned} L_f \equiv d_{\vec{x}_0} f \quad &\text{is represented by} \quad f'(\vec{x}_0) \equiv \frac{d\vec{y}}{d\vec{x}} \equiv J_{\vec{x}_0} f, \\ L_g \equiv d_{\vec{y}_0} g \quad &\text{is represented by} \quad g'(\vec{y}_0) \equiv \frac{d\vec{z}}{d\vec{y}} \equiv J_{\vec{y}_0} g. \end{aligned}$$

That is,

$$\begin{aligned} \vec{y} = f(\vec{x}) &= \vec{y}_0 + L_f(\vec{x} - \vec{x}_0) + o(\vec{x} - \vec{x}_0) \qquad \big(\vec{y}_0 \equiv f(\vec{x}_0)\big), \\ \vec{z} = g(\vec{y}) &= \vec{z}_0 + L_g(\vec{y} - \vec{y}_0) + o(\vec{y} - \vec{y}_0) \qquad \big(\vec{z}_0 \equiv g(\vec{y}_0)\big), \end{aligned}$$

(We are deliberately being verbose here, writing equations over in several equivalent forms to give experience in seeing various notations.) This time, substitution of f into g will yield a derivation of the chain rule.

The formulas are the same as in the affine case except for the "o" terms, since the point of differentiability is that each function is equal to an affine

function plus something small. To simplify the calculation, let's use the increment notation ($\Delta\vec{y} \equiv \vec{y} - \vec{y}_0$, etc.). We get

$$\begin{aligned}\vec{z} &= \vec{z}_0 + L_g\big(L_f(\Delta\vec{x}) + o(\Delta\vec{x})\big) + o(\Delta\vec{y}) \\ &= \vec{z}_0 + \big(L_g \circ L_f\big)(\Delta\vec{x}) + L_g\big(o(\Delta\vec{x})\big) + o(\Delta\vec{y}).\end{aligned}$$

Now (by definition of "o")

$$\frac{o(\Delta\vec{x})}{\|\Delta\vec{x}\|} \to 0 \quad \text{as} \quad \vec{x} \to \vec{x}_0\,.$$

Therefore, by the lemma,

$$\frac{L_g\big(o(\Delta\vec{x})\big)}{\|\Delta\vec{x}\|} \to 0 \quad \text{as} \quad \vec{x} \to \vec{x}_0\,.$$

That is,

$$L_g\big(o(\Delta\vec{x})\big) = o(\Delta\vec{x}). \tag{$\dagger$}$$

But also,

$$\frac{o(\Delta\vec{y})}{\|\Delta\vec{y}\|} \to 0 \quad \text{as} \quad \vec{y} \to \vec{y}_0\,,$$

and (by the corollary) $\Delta\vec{y} = O(\Delta\vec{x})$. So we have

$$\frac{o(\Delta\vec{y})}{\|\Delta\vec{x}\|} = \frac{o(\Delta\vec{y})}{\|\Delta\vec{y}\|}\,\frac{\|\Delta\vec{y}\|}{\|\Delta\vec{x}\|} \to 0 \quad \text{as} \quad \vec{x} \to \vec{x}_0\,.$$

That is,

$$o(\Delta\vec{y}) = o(\Delta\vec{x}). \tag{$\ddagger$}$$

(Actually, we need a special argument to cover the possibility that $\|\Delta\vec{y}\| = 0$; but we are not trying to give a complete proof here, just the central idea.)

Putting ($\dagger$) and ($\ddagger$) together, we see that

$$\vec{z} \equiv \big(g \circ f\big)(\vec{x}) = \vec{z}_0 + \big(L_g \circ L_f\big)(\vec{x} - \vec{x}_0) + o(\vec{x} - \vec{x}_0), \tag{1}$$

where $\vec{z}_0 = \big(g \circ f\big)(\vec{x}_0)$. That is, $L_{g\circ f} = L_g \circ L_f$, or

$$d_{\vec{x}_0}(g \circ f) = \big[d_{f(\vec{x}_0)}g\big] \circ [d_{\vec{x}_0}f]\,. \tag{2}$$

That is,

$$\frac{d\vec{z}}{d\vec{x}} = \frac{d\vec{z}}{d\vec{y}}\,\frac{d\vec{y}}{d\vec{x}} \qquad \text{(a matrix product)}, \tag{3}$$

or $\big(g \circ f\big)'(\vec{x}) = g'\big(f(\vec{x})\big)\, f'(\vec{x})$. That is,

$$\frac{\partial z_j}{\partial x_l} = \sum_{k=1}^{m} \frac{\partial z_j}{\partial y_k}\, \frac{\partial y_k}{\partial x_l} \tag{4}$$

for all index pairs, $j \in \{1, 2, \ldots, p\}$ and $l \in \{1, 2, \ldots, n\}$.

In summary, we have here four equivalent versions of the chain rule.

(1) is about nonlinear functions.

(2) is about linear functions.

(3) is about matrices.

(4) is about numbers.

But they all mean the same thing.

Our emphasis on the fundamental meaning of differentiation and the chain rule in terms of affine approximations should not be misinterpreted as a disparagement of the classical formula (4). In practice it remains very useful. In a sense the Leibnitz notation in (4) (or (3)) does the correct matrix multiplication for you automatically (just as in one dimension the Leibnitz formula $\frac{dz}{dx} = \frac{dz}{dy}\,\frac{dy}{dx}$ is more transparent than $(g \circ f)'(x) = g'\big(f(x)\big)\, f'(x)\,$). Of course, there is a possible pitfall: the factor $\frac{d\vec{z}}{d\vec{y}}$ must be numerically evaluated at $f(\vec{x})$, not $\vec{x}$. Note also that careless throwing around of partial-derivative symbols without a clear understanding of what they mean in context can lead to false conclusions; see Exercises 3.5.7–8. (Traditionally, a course in thermodynamics is the training ground for learning these things through harsh experience.)

Example 1. Find the differential of the function

$$z(u, v) = \cos\frac{1}{uv} + \sin\sqrt{\frac{u}{v}}\,.$$

Solution: We shall do this in several seemingly different ways, which are all really the same.

Method 1: Calculate $\frac{\partial z}{\partial u}$ and $\frac{\partial z}{\partial v}$ as elementary (though messy) single-variable differentiation problems. Then the Jacobian matrix is the row vector $\big(\frac{\partial z}{\partial u}, \frac{\partial z}{\partial v}\big)$. Although this may be conceptually the most obvious method, it is also the most inefficient. We omit the boring details.

Method 2: Introduce auxiliary variables

$$x = \frac{1}{uv}\,, \quad y = \sqrt{\frac{u}{v}}\,.$$

Then $z = \cos x + \sin y$, and

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\,\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\,\frac{\partial y}{\partial u}\,, \qquad \frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\,\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\,\frac{\partial y}{\partial v}\,.$$

We calculate

$$\frac{\partial z}{\partial x} = -\sin x, \quad \frac{\partial z}{\partial y} = \cos y;$$

$$\frac{\partial x}{\partial u} = -\frac{1}{u^2 v}\,, \quad \frac{\partial x}{\partial v} = -\frac{1}{uv^2}\,, \quad \frac{\partial y}{\partial u} = \frac{1}{2\sqrt{uv}}\,, \quad \frac{\partial y}{\partial v} = -\frac{\sqrt{u}}{2v\sqrt{v}}\,.$$

Thus

$$\begin{aligned} dz &\equiv \frac{\partial z}{\partial u}\,du + \frac{\partial z}{\partial v}\,dv \\ &= \frac{1}{\sqrt{uv}}\left(\frac{1}{u\sqrt{uv}}\sin\frac{1}{uv} + \frac{1}{2}\cos\sqrt{\frac{u}{v}}\right)du \\ &\qquad + \frac{1}{v\sqrt{v}}\left(\frac{1}{u\sqrt{v}}\sin\frac{1}{uv} - \frac{\sqrt{u}}{2}\cos\sqrt{\frac{u}{v}}\right)dv. \end{aligned}$$

Method 3: Introduce vector variables

$$\vec{u} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Then

$$\frac{dz}{d\vec{u}} = \frac{dz}{d\vec{x}}\,\frac{d\vec{x}}{d\vec{u}}$$

where

$$\frac{dz}{d\vec{x}} = \nabla z = (-\sin x, \cos y), \qquad \frac{d\vec{x}}{d\vec{u}} = \begin{pmatrix} -\frac{1}{u^2 v} & -\frac{1}{uv^2} \\ \frac{1}{2\sqrt{uv}} & -\frac{\sqrt{u}}{2v\sqrt{v}} \end{pmatrix}.$$

Perform the matrix multiplication and evaluate $\vec{x}$ at the point $\vec{u}$:

$$\begin{aligned} \frac{dz}{d\vec{u}} = \Bigg(&\frac{1}{\sqrt{uv}}\left(\frac{1}{u\sqrt{uv}}\sin\frac{1}{uv} + \frac{1}{2}\cos\sqrt{\frac{u}{v}}\right), \\ &\frac{1}{v\sqrt{v}}\left(\frac{1}{u\sqrt{v}}\sin\frac{1}{uv} - \frac{\sqrt{u}}{2}\cos\sqrt{\frac{u}{v}}\right)\Bigg). \end{aligned}$$

Methods 4–5: Introduce auxiliary variables

$$\xi = uv, \quad \eta = \frac{u}{v},$$

and proceed as in Method 2 or 3, respectively.

Method 6: Write the answer as a product of *three* matrices, corresponding to the decomposition of the composite function into three steps,

$$\begin{pmatrix} u \\ v \end{pmatrix} \to \begin{pmatrix} \xi \\ \eta \end{pmatrix} \to \begin{pmatrix} x \\ y \end{pmatrix} \to z.$$

Example 2. Atoms near the point $\vec{r}_0 \equiv \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$ sit in an electric field

$$\vec{E} = \begin{pmatrix} x^2 - y \\ x^2 + y^2 \\ z^3 \end{pmatrix}.$$

(a) Find the first-order (best affine) approximation to $\vec{E}(\vec{r})$ for $\vec{r}$ near $\vec{r}_0$.

(b) Suppose that the index of refraction of a crystal depends on the electric field according to the law

$$n = 1 + 0.01{E_x}^2 + 0.04{E_y}^2 + 0.02{E_z}^2.$$

Use the multidimensional chain rule to find $\dfrac{\partial n}{\partial y}$ at $\vec{r}_0$.

SOLUTION: The Jacobian matrix is

$$\frac{d\vec{E}}{d\vec{r}} = \begin{pmatrix} 2x & -1 & 0 \\ 2x & 2y & 0 \\ 0 & 0 & 3z^2 \end{pmatrix} = \begin{pmatrix} 6 & -1 & 0 \\ 6 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix} \text{ at } \vec{r}_0.$$

Therefore,

$$\vec{E}(\vec{r}) \approx \vec{E}(\vec{r}_0) + \frac{d\vec{E}}{d\vec{r}}\,(\vec{r} - \vec{r}_0) = \begin{pmatrix} 7 \\ 13 \\ 1 \end{pmatrix} + \begin{pmatrix} 6 & -1 & 0 \\ 6 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x-3 \\ y-2 \\ z-1 \end{pmatrix}$$

is the first-order approximation asked for in (a). Therefore, the matrix solution to (b) is

$$\begin{aligned}\nabla n = \frac{dn}{d\vec{E}}\,\frac{d\vec{E}}{d\vec{r}} &= (0.02E_x, 0.08E_y, 0.04E_z)\big|_{\vec{r}_0}\,\frac{d\vec{E}}{d\vec{r}} \\ &= (0.14, 1.04, 0.04)\begin{pmatrix} 6 & -1 & 0 \\ 6 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix} = (\,*\,, 4.02,\, *\,),\end{aligned}$$

where the numbers $*$ are irrelevant to the question asked, and

$$4.02 = \frac{\partial n}{\partial y}\,.$$

Here is an alternative solution to (b), but in "classical" partial-derivative notation instead of vectors and matrices; it is really the same method:

$$\begin{aligned}\frac{\partial n}{\partial y} &= \frac{\partial n}{\partial E_x}\,\frac{\partial E_x}{\partial y} + \frac{\partial n}{\partial E_y}\,\frac{\partial E_y}{\partial y} + \frac{\partial n}{\partial E_z}\,\frac{\partial E_z}{\partial y} \\ &= 0.02E_x\,(-1) + 0.08E_y\,(2y) + 0.04E_z\,(0).\end{aligned}$$

When all this is evaluated at $(x, y, z) = (3, 2, 1)$, we again get 4.02.

Exercises

3.5.1 We have two nonlinear functions, $f\colon \mathbf{R}^3 \to \mathbf{R}^2$ and $g\colon \mathbf{R}^2 \to \mathbf{R}^2$. Suppose that

$$f\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad f'\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad g'\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}.$$

What is the derivative (Jacobian) matrix of $g \circ f$ at the point $(1, 2, 1)$?

3.5.2 Let $x = u^2 - \cos v$, $y = e^u + v^2$, $\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} \sin t \\ 2\cos t \end{pmatrix}$. Calculate $\dfrac{dx}{dt}$ and $\dfrac{dy}{dt}$ when $t = \pi$.

3.5.3 The velocity field in a gas is given by

$$\vec{u}(\vec{r}) \equiv u_x\,\hat{\imath}+u_y\,\hat{\jmath}+u_z\,\hat{k} = (x^2-z^2)\,\hat{\imath}+(y^2+x^2)\,\hat{\jmath}+z^2\,\hat{k} \equiv \begin{pmatrix} x^2-z^2 \\ y^2+x^2 \\ z^2 \end{pmatrix}.$$

(a) Calculate the matrix representing $d_{\vec{r}_0}\vec{u}$ at $\vec{r}_0 = 2\,\hat{\imath}+3\,\hat{\jmath}+4\,\hat{k}$.

(b) Calculate $d\vec{u}/dt$ as measured along the path

$$\vec{r}(t) = (2+t)\,\hat{\imath}+(3-t)\,\hat{\jmath}+(4+t^2)\,\hat{k}$$

at $t=0$. (Note that $\vec{r}(0) = \vec{r}_0$.)

(c) Estimate, using the "best affine approximation" based on the result of (a), the velocity field at the point with coordinates $(2.1, 3.2, 3.9)$.

(d) Find the direction of fastest increase of the x component of $\vec{u}$ at $\vec{r}_0$.

(e) Find the direction of fastest increase of the length of $\vec{u}$ at $\vec{r}_0$. HINT: Compose with $f(\vec{u}) \equiv \|\vec{u}\| = \sqrt{{u_x}^2+{u_y}^2+{u_z}^2}$.

3.5.4 Continuing Exercise 3.4.6: If air pressure depends upon height as $P = 700e^{-z}$ and Roger is flying east at 500 feet per second, find the rate of change (with time) of the pressure directly below him (at the ground) as he passes over the point $(2,0)$.

3.5.5 Suppose a force field in space is given by

$$\vec{F}(\vec{r}) = x\cos\frac{\pi z}{4}\,\hat{\imath}+y\sin\frac{\pi z}{4}\,\hat{\jmath}+xz\,\hat{k}.$$

(a) Calculate [the matrix representing] $d_{\vec{r}_0}\vec{F}$ at $\vec{r}_0 = (0,1,1)$. (Think of $\vec{F}$ as a column vector.)

(b) A particle follows the helical path

$$x = \cos 2t, \qquad y = \sin 2t, \qquad z = \frac{4t}{\pi}.$$

Calculate $d\vec{F}/dt$ at $t = \pi/4$ (as measured by an observer travelling with the particle).

(c) *(for students who know some advanced physics)* Originally the exercise's author intended to call this force field an "electric field". Why might that have made a misleading or ambiguous question? HINT: What does your electromagnetism textbook say about how electric and magnetic fields change under Galilean or Lorentz transformations?

3.5.6 After Roger measured the gradient of the pressure in Exercise 3.4.4, his navigator needed to convert all the measurements into the parabolic-cylindrical coordinate system (u, v, w), where

$$x = 2uv, \qquad y = u^2 - v^2, \qquad z = w.$$

The plane was located at $(u, v, w) = (1, 2, 0)$ when the pressure gradient was measured. Find the partial derivatives of p with respect to u, v, and w at that point.

3.5.7 The temperature in a square plate is given by the equation $T = x+3y$. An ant crawled on the plate along the parabolic path $y = x^2$ until it reached the point $(2, 4)$, whereupon it found itself fried. What is wrong with the following argument?

$$\begin{aligned}\frac{\partial T}{\partial x} &= \frac{\partial T}{\partial x}\frac{\partial x}{\partial x} + \frac{\partial T}{\partial y}\frac{\partial y}{\partial x} \\ &= \frac{\partial T}{\partial x} + \frac{\partial T}{\partial y}\frac{\partial y}{\partial x}.\end{aligned}$$

We can cancel the two equal terms, concluding that

$$\frac{\partial T}{\partial y}\frac{\partial y}{\partial x} = 0.$$

Substituting the given functions and numbers, we learn that

$$12 = 0.$$

3.5.8. An ordinary differential equation $\dfrac{dy}{dx} = g(x, y)$ is called *exact* if it can be rewritten in the form

$$X(x, y)\,dx + Y(x, y)\,dy = 0$$

and the left-hand side can be recognized as the differential $d\psi$ of a function $\psi(x, y)$ (i.e., $X = \frac{\partial \psi}{\partial x}$ and $Y = \frac{\partial \psi}{\partial y}$). In that case each

level curve $\psi(x, y) = C$ is the graph of a solution of the differential equation. (With luck, the equation $\psi(x, y) = C$ can be solved for y as a function of x.)

(a) A common student error is to try to find the function ψ in this way:

$$\text{Given:} \quad \frac{dy}{dx} = -\frac{xy^2 + 1}{x^2 y};$$

$$(xy^2 + 1)\,dx + x^2 y\,dy = 0;$$

$$\int (xy^2 + 1)\,dx + \int x^2 y\,dy = 0;$$

$$\frac{1}{2}\,x^2 y^2 + x + \frac{1}{2}\,x^2 y^2 = C;$$

therefore,

$$\psi(x, y) \equiv x^2 y^2 + x = C.$$

This answer is *wrong*; the equation is indeed exact, but the correct solution is

$$\psi(x, y) \equiv \frac{1}{2} x^2 y^2 + x = C.$$

Verify this (by implicit differentiation), and explain where the argument went astray.

(b) Find the correct ψ by solving the equations

$$\frac{\partial \psi}{\partial x} = xy^2 + 1, \qquad \frac{\partial \psi}{\partial y} = x^2 y,$$

recalling that the "constant" of integration in each case might be a function of the other variable.

(c) Show that the incorrect solution method demonstrated in (a) *will give the right answer by accident*, so long as the correct function ψ does not contain any term linear in one of the variables x and y. (This circumstance, especially together with (d), makes this fallacy especially hard to eradicate among our youth.)

(d) There is a large class of differential equations called *separable*, which are correctly solved in this way:

$$\text{Given:} \quad \frac{dy}{dx} = \frac{xy}{x^2 + 1};$$

$$\frac{dy}{y} = \frac{x\,dx}{x^2+1}\,;$$

$$\int \frac{dy}{y} = \int \frac{x\,dx}{x^2+1}\,;$$

therefore,

$$\ln|y| = \frac{1}{2}\ln(x^2+1) + C.$$

Explain why integrating each term with respect to its own variable is correct both here and in (b) whereas it was wrong in (a).

3.5.9 Find the differential (or the Jacobian matrix) of

$$w = F(u,v) \equiv \cos(5x+3y),$$

where $x = u^3 + v^3$ and $y = uv$.

3.5.10 Find the differential (or the Jacobian matrix) of

$$w = F(u,v) \equiv 5x^2y^2 - 3y + 2x + z,$$

$$x = u\cos v, \quad y = \sin(u/v), \quad z = u^2 - v^2.$$

3.5.11 Find the differential of $w = F(x,y)$ if

$$w = \frac{\cos^{-1}\frac{x^2-y^2}{x^2+y^2}}{\ln\frac{1}{x+1}}\,.$$

(Introduce auxiliary variables $u = \frac{x^2-y^2}{x^2+y^2}$, $v = \frac{1}{x+1}$. Feel free to use *Maple* or similar software, provided that your paper includes well written English commentary with nontrivial mathematical content.)

3.5.12 Find the differential of $w = f(x-y, x^2-y^2)$, where f is an arbitrary differentiable function.

3.5.13 Let f and g be arbitrary twice-differentiable functions of the type $\mathbf{R} \to \mathbf{R}$. Prove that

$$u(t,x) \equiv f(x-ct) + g(x+ct)$$

is always a solution of the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2\,\frac{\partial^2 u}{\partial x^2}\,.$$

3.5.14 Let $G(\vec{x}, \vec{y})$ be a function of two vector variables (i.e., $G\colon \mathbf{R}^{2n} \to \mathbf{R}$, where the $2n$ variables separate naturally into two n-dimensional parts). The *diagonal value* of G is the function $H\colon \mathbf{R}^n \to \mathbf{R}$ defined by $H(\vec{x}) = G(\vec{x}, \vec{x})$. Show that

$$\nabla H(\vec{x}) = \nabla_{\vec{x}} G(\vec{x}, \vec{x}) + \nabla_{\vec{y}} G(\vec{x}, \vec{x}),$$

where $\nabla_{\vec{x}} G(\vec{x}, \vec{y})$ is the gradient of G with respect to its first variable with the second held fixed, $\nabla_{\vec{y}} G(\vec{x}, \vec{y})$ is the gradient with respect to the second variable with the first fixed, and in the final formula it is understood that $\vec{y}$ is set equal to $\vec{x}$ *after* these differentiations are performed.

3.5.15 Continuing Exercise 3.4.7: What is the time derivative of the magnetic field, as Roger measures it in his moving plane? (Notice that arithmetically, this is a *simpler* problem than the original one. Explain why.)