Chapter 7

Linearity

We began this book by learning how to systematically solve systems of linear algebraic equations. This "elementary" problem formed our launching pad for developing the fundamentals of linear algebra. In its initial form, matrices and vectors were the primary focus of our study, but the theory was developed in a sufficiently general and abstract form that it can be immediately used in many other useful situations — particularly infinite-dimensional function spaces. Indeed, applied mathematics deals, not just with algebraic equations, but also with differential equations, difference equations, integral equations, stochastic systems, differential delay equations, control systems, and many other types — only a few of which, unfortunately, can be adequately developed in this introductory text. It is now time to assemble what we have learned about linear algebraic systems and place the results in a suitably general framework that will lead to insight into the key principles that govern all linear systems arising in mathematics and its applications.

The most basic underlying object of linear systems theory is the vector space, and we have already seen that the elements of vector spaces can be vectors, or functions, or even vector-valued functions. The seminal ideas of span, linear independence, basis, and dimension are equally applicable and equally vital in more general contexts, particularly function spaces. Just as vectors in Euclidean space are prototypes for elements of general vector spaces, matrices are also prototypes for more general objects, known as *linear functions*. Linear functions are also known as linear maps or, when one is dealing with function spaces, linear operators, and include linear differential operators, linear integral operators, function evaluation, and many other basic operations. Linear operators on infinite-dimensional function spaces are the basic objects of quantum mechanics. Each quantum mechanical observable (mass, energy, momentum) is formulated as a linear operator on an infinite-dimensional Hilbert space — the space of wave functions or states of the system, [**54**]. It is remarkable that quantum mechanics is an entirely linear theory, whereas classical and relativistic mechanics are inherently nonlinear. The holy grail of modern physics — the unification of general relativity and quantum mechanics — is to resolve the apparent incompatibility of the microscopic linear and macroscopic nonlinear physical regimes.

In geometry, linear functions are interpreted as linear transformations of space (or space-time), and, as such, lie at the foundations of motion of bodies, such as satellites and planets; computer graphics and games; video, animation, and movies; and the mathematical formulation of symmetry. Many familiar geometrical transformations, including rotations, scalings and stretches, reflections, projections, shears, and screw motions, are linear. But including translational motions requires a slight extension of linearity, known as an affine transformation. The basic geometry of linear and affine transformations will be developed in Section 7.2.

Linear functions form the simplest class of functions on vector spaces, and must be thoroughly understood before any serious progress can be made in the vastly more complicated nonlinear world. Indeed, nonlinear functions are often approximated by linear functions, generalizing the calculus approximation of a scalar function by its tangent line. This linearization process is applied to nonlinear functions of several variables studied in

P. J. Olver, C. Shakiban, *Applied Linear Algebra*, Undergraduate Texts in Mathematics,
https://doi.org/10.1007/978-3-319-91041-3_7

multivariable calculus, as well as the nonlinear systems arising in physics and mechanics, which can often be well approximated by linear differential equations.

A linear system is just an equation formed by a linear function. The most basic linear system is a system of linear algebraic equations. Linear systems theory includes linear differential equations, linear boundary value problems, linear integral equations, and so on, all in a common conceptual framework. The fundamental ideas of linear superposition and the relation between the solutions to inhomogeneous and homogeneous systems are universally applicable to all linear systems. You have no doubt encountered many of these concepts in your study of elementary ordinary differential equations. In this text, they have already appeared in our discussion of the solutions to linear algebraic systems. The final section introduces the notion of the adjoint of a linear map between inner product spaces, generalizing the transpose operation on matrices, the notion of a positive definite linear operator, and the characterization of the solution to such a linear system by a minimization principle. The full import of these fundamental concepts will appear in the context of linear boundary value problems and partial differential equations, [**61**].

7.1 Linear Functions

We begin our study of linear functions with the basic definition. For simplicity, we shall concentrate on real linear functions between real vector spaces. Extending the concepts and constructions to complex linear functions on complex vector spaces is not difficult, and will be dealt with in due course.

Definition 7.1. Let V and W be real vector spaces. A function $L: V \to W$ is called *linear* if it obeys two basic rules:

$$L[\mathbf{v}+\mathbf{w}] = L[\mathbf{v}] + L[\mathbf{w}], \qquad\qquad L[c\,\mathbf{v}] = c\,L[\mathbf{v}], \tag{7.1}$$

for all $\mathbf{v}, \mathbf{w} \in V$ and all scalars c. We will call V the *domain* and W the *codomain*† for L.

In particular, setting $c = 0$ in the second condition implies that a linear function always maps the zero element $\mathbf{0} \in V$ to the zero element‡ $\mathbf{0} \in W$, so

$$L[\mathbf{0}] = \mathbf{0}. \tag{7.2}$$

We can readily combine the two defining conditions (7.1) into a single rule

$$L[c\,\mathbf{v} + d\,\mathbf{w}] = c\,L[\mathbf{v}] + d\,L[\mathbf{w}], \qquad \text{for all} \qquad \mathbf{v}, \mathbf{w} \in V, \quad c, d \in \mathbb{R}, \tag{7.3}$$

that characterizes linearity of a function L. An easy induction proves that a linear function respects linear combinations, so

$$L[c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k] = c_1\,L[\mathbf{v}_1] + \cdots + c_k\,L[\mathbf{v}_k] \tag{7.4}$$

for all $c_1, \ldots, c_k \in \mathbb{R}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$.

The interchangeable terms *linear map*, *linear operator*, and, when $V = W$, *linear transformation* are all commonly used as alternatives to "linear function", depending on the

† The terms "range" and "target" are also sometimes used for the codomain. However, some authors use "range" to mean the image of L. An alternative name for domain is "source".

‡ We will use the same notation for these two zero elements even though they may belong to different vector spaces. The reader should be able to determine where each lives from the context.

circumstances and taste of the author. The term "linear operator" is particularly useful when the underlying vector space is a function space, so as to avoid confusing the two different uses of the word "function". As usual, we will often refer to the elements of a vector space as "vectors", even though they might be functions or matrices or something else, depending on the context.

Example 7.2. The simplest linear function is the zero function $\mathrm{O}[\mathbf{v}] \equiv \mathbf{0}$, which maps every element $\mathbf{v} \in V$ to the zero vector in W. Note that, in view of (7.2), this is the *only* constant linear function; a nonzero constant function is *not*, despite its evident simplicity, linear. Another simple but important linear function is the identity function $\mathrm{I} = \mathrm{I}_V : V \to V$, which maps V to itself and leaves every vector unchanged: $\mathrm{I}[\mathbf{v}] = \mathbf{v}$. Slightly more generally, the operation of scalar multiplication $M_a[\mathbf{v}] = a\,\mathbf{v}$ by a scalar $a \in \mathbb{R}$ defines a linear function from V to itself, with $M_0 = \mathrm{O}$, the zero function from V to itself, and $M_1 = \mathrm{I}$, the identity function on V, appearing as special cases.

Example 7.3. Suppose $V = \mathbb{R}$. We claim that every linear function $L : \mathbb{R} \to \mathbb{R}$ has the form

$$y = L[x] = a\,x,$$

for some constant a. Therefore, the only scalar linear functions are those whose graph is a straight line passing through the origin. To prove this, we write $x \in \mathbb{R}$ as a scalar product $x = x \cdot 1$. Then, by the second property in (7.1),

$$L[x] = L[x \cdot 1] = x \cdot L[1] = a\,x, \qquad \text{where} \qquad a = L[1],$$

as claimed.

Warning. Even though the graph of the function

$$y = a\,x + b, \tag{7.5}$$

is a straight line, it is *not* a linear function — unless $b = 0$, so the line goes through the origin. The proper mathematical name for a function of the form (7.5) is an *affine function*; see Definition 7.21 below.

Example 7.4. Let $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$. Let A be an $m \times n$ matrix. Then the function $L[\mathbf{v}] = A\,\mathbf{v}$ given by matrix multiplication is easily seen to be a linear function. Indeed, the requirements (7.1) reduce to the basic distributivity and scalar multiplication properties of matrix multiplication:

$$A(\mathbf{v} + \mathbf{w}) = A\,\mathbf{v} + A\,\mathbf{w}, \qquad A(c\,\mathbf{v}) = c\,A\,\mathbf{v}, \qquad \text{for all} \qquad \mathbf{v}, \mathbf{w} \in \mathbb{R}^n, \quad c \in \mathbb{R}.$$

In fact, *every* linear function between two Euclidean spaces has this form.

Theorem 7.5. Every linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ is given by matrix multiplication, $L[\mathbf{v}] = A\,\mathbf{v}$, where A is an $m \times n$ matrix.

Warning. Pay attention to the order of m and n. While A has size $m \times n$, the linear function L goes *from* $\mathbb{R}^n$ *to* $\mathbb{R}^m$.

Proof: The key idea is to look at what the linear function does to the basis vectors. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis of $\mathbb{R}^n$, as in (2.17), and let $\widehat{\mathbf{e}}_1, \ldots, \widehat{\mathbf{e}}_m$ be the standard

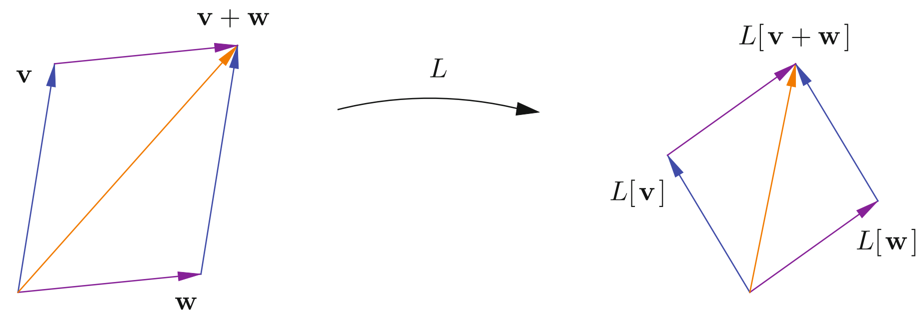


Figure 7.1. Linear Function on Euclidean Space.

basis of $\mathbb{R}^m$. (We temporarily place hats on the latter to avoid confusing the two.) Since $L[\mathbf{e}_j] \in \mathbb{R}^m$, we can write it as a linear combination of the latter basis vectors:

$$L[\mathbf{e}_j] = \mathbf{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} = a_{1j}\,\widehat{\mathbf{e}}_1 + a_{2j}\,\widehat{\mathbf{e}}_2 + \;\cdots\; + a_{mj}\,\widehat{\mathbf{e}}_m, \qquad j = 1, \ldots, n. \tag{7.6}$$

Let us construct the $m \times n$ matrix

$$A = (\, \mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_n \,) = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \tag{7.7}$$

whose columns are the image vectors (7.6). Using (7.4), we then compute the effect of L on a general vector $\mathbf{v} = (\, v_1, v_2, \ldots, v_n \,)^T \in \mathbb{R}^n$:

$$L[\mathbf{v}] = L[v_1\,\mathbf{e}_1 + \cdots + v_n\,\mathbf{e}_n] = v_1\,L[\mathbf{e}_1] + \cdots + v_n\,L[\mathbf{e}_n] = v_1\,\mathbf{a}_1 + \cdots + v_n\,\mathbf{a}_n = A\,\mathbf{v}.$$

The final equality follows from our basic formula (2.13) connecting matrix multiplication and linear combinations. We conclude that the vector $L[\mathbf{v}]$ coincides with the vector $A\,\mathbf{v}$ obtained by multiplying $\mathbf{v}$ by the coefficient matrix A. *Q.E.D.*

The proof of Theorem 7.5 shows us how to construct the matrix representative of a given linear function $L\colon \mathbb{R}^n \to \mathbb{R}^m$. We merely assemble the image column vectors $\mathbf{a}_1 = L[\mathbf{e}_1], \ldots, \mathbf{a}_n = L[\mathbf{e}_n]$ into an $m \times n$ matrix A.

The two basic linearity conditions (7.1) have a simple geometrical interpretation. Since vector addition is the same as completing the parallelogram sketched in Figure 7.1, the first linearity condition requires that L map parallelograms to parallelograms. The second linearity condition says that if we stretch a vector by a factor c, then its image under L must also be stretched by the same amount. Thus, one can often detect linearity by simply looking at the geometry of the function.

Example 7.6. As a specific example, consider the function $R_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ that rotates the vectors in the plane around the origin by a specified angle θ. This geometric transformation clearly preserves parallelograms — see Figure 7.2. It also respects stretching of vectors, and hence defines a linear function. In order to find its matrix representative,

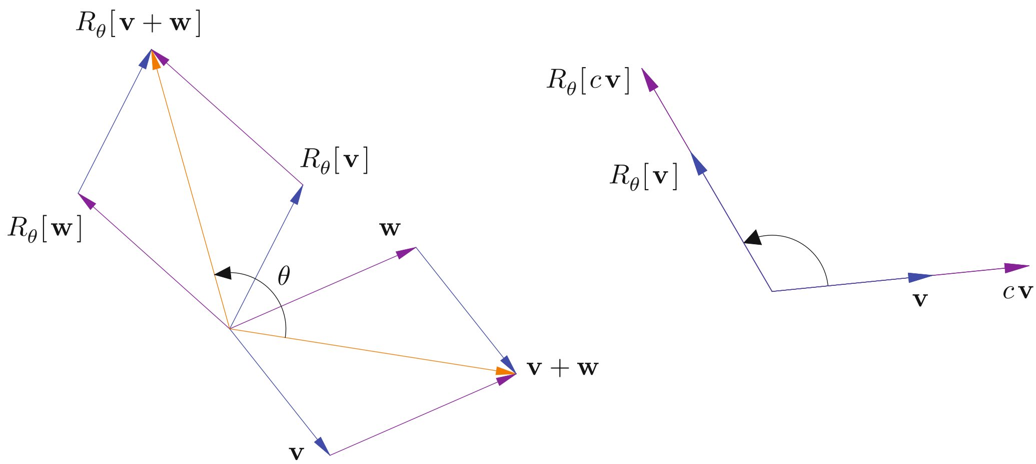


Figure 7.2. Linearity of Rotations.

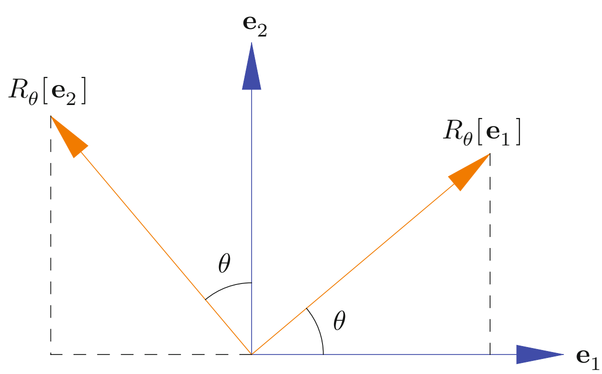


Figure 7.3. Rotation in $\mathbb{R}^2$.

we need to find out where the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2$ are mapped. Referring to Figure 7.3, and keeping in mind that the rotated vectors also have unit length, we have

$$R_\theta[\mathbf{e}_1] = (\cos\theta)\,\mathbf{e}_1 + (\sin\theta)\,\mathbf{e}_2 = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix},$$

$$R_\theta[\mathbf{e}_2] = -(\sin\theta)\,\mathbf{e}_1 + (\cos\theta)\,\mathbf{e}_2 = \begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}.$$

According to the general recipe (7.7), we assemble these two column vectors to obtain the matrix form of the rotation transformation, and so

$$R_\theta[\mathbf{v}] = A_\theta\,\mathbf{v}, \qquad \text{where} \qquad A_\theta = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}. \tag{7.8}$$

Therefore, rotating a vector $\mathbf{v} = \begin{pmatrix}x\\ y\end{pmatrix}$ through angle θ produces the vector

$$\widehat{\mathbf{v}} = R_\theta[\mathbf{v}] = A_\theta\,\mathbf{v} = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}x\cos\theta - y\sin\theta\\ x\sin\theta + y\cos\theta\end{pmatrix}$$

with coordinates $\widehat{x} = x\cos\theta - y\sin\theta$, $\widehat{y} = x\sin\theta + y\cos\theta$. These formulas can be proved directly, but, in fact, are a consequence of the underlying linearity of rotations.

Exercises

7.1.1. Which of the following functions $F\colon \mathbb{R}^3 \to \mathbb{R}$ are linear? (a) $F(x,y,z) = x$, (b) $F(x,y,z) = y - 2$, (c) $F(x,y,z) = x + y + 3$, (d) $F(x,y,z) = x - y - z$, (e) $F(x,y,z) = x\,y\,z$, (f) $F(x,y,z) = x^2 - y^2 + z^2$, (g) $F(x,y,z) = e^{x-y+z}$.

7.1.2. Explain why the following functions $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ are not linear.

(a) $\begin{pmatrix} x+2 \\ x+y \end{pmatrix}$, (b) $\begin{pmatrix} x^2 \\ y^2 \end{pmatrix}$, (c) $\begin{pmatrix} |\,y\,| \\ |\,x\,| \end{pmatrix}$, (d) $\begin{pmatrix} \sin(x+y) \\ x-y \end{pmatrix}$, (e) $\begin{pmatrix} x+e^y \\ 2x+y \end{pmatrix}$.

7.1.3. Which of the following functions $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ are linear?

(a) $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x-y \\ x+y \end{pmatrix}$, (b) $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y+1 \\ x-y-1 \end{pmatrix}$, (c) $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\,y \\ x-y \end{pmatrix}$,

(d) $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3\,y \\ 2\,x \end{pmatrix}$, (e) $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2+y^2 \\ x^2-y^2 \end{pmatrix}$, (f) $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y-3\,x \\ x \end{pmatrix}$.

7.1.4. Explain why the translation function $T\colon \mathbb{R}^2 \to \mathbb{R}^2$, defined by $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+a \\ y+b \end{pmatrix}$ for $a, b \in \mathbb{R}$, is almost never linear. Precisely when is it linear?

7.1.5. Find a matrix representation for the following linear transformations on $\mathbb{R}^3$: (a) counterclockwise rotation by 90° around the z-axis; (b) clockwise rotation by 60° around the x-axis; (c) reflection through the (x,y)-plane; (d) counterclockwise rotation by 120° around the line $x = y = z$; (e) rotation by 180° around the line $x = y = z$; (f) orthogonal projection onto the $x\,y$-plane; (g) orthogonal projection onto the plane $x - y + 2\,z = 0$.

7.1.6. Find a linear function $L\colon \mathbb{R}^2 \to \mathbb{R}$ such that $L\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2$ and $L\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 3$. Is it unique?

7.1.7. Find a linear function $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $L\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ and $L\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$.

7.1.8. Under what conditions does there exist a linear function $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $L\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}$ and $L\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}$? Under what conditions is L uniquely defined? In the latter case, write down the matrix representation of L.

7.1.9. Can you construct a linear function $L\colon \mathbb{R}^3 \to \mathbb{R}$ such that $L\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = 1$, $L\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 4$, and $L\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = -2$? If yes, find one. If not, explain why not.

♢ 7.1.10. Given $\mathbf{a} = (\,a, b, c\,)^T \in \mathbb{R}^3$, prove that the cross product map $L_{\mathbf{a}}[\mathbf{v}] = \mathbf{a} \times \mathbf{v}$, as defined in (4.2), is linear, and find its matrix representative.

7.1.11. Is the Euclidean norm function $N(\mathbf{v}) = \|\,\mathbf{v}\,\|$, for $\mathbf{v} \in \mathbb{R}^n$, linear?

7.1.12. Let V be a vector space. Prove that every linear function $L\colon \mathbb{R} \to V$ has the form $L[x] = x\,\mathbf{b}$, where $x \in \mathbb{R}$, for some $\mathbf{b} \in V$.

7.1.13. *True or false*: The quadratic form $Q(\mathbf{v}) = \mathbf{v}^T K\,\mathbf{v}$ defined by a symmetric $n \times n$ matrix K defines a linear function $Q\colon \mathbb{R}^n \to \mathbb{R}$.

♢ 7.1.14. (a) Prove that L is linear if and only if it satisfies (7.3).
(b) Use induction to prove that L satisfies (7.4).

7.1.15. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $B = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$ be 2×2 matrices. For each of the following functions, prove that $L\colon \mathcal{M}_{2\times 2} \to \mathcal{M}_{2\times 2}$ defines a linear map, and then find its matrix representative with respect to the standard basis $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ of $\mathcal{M}_{2\times 2}$:

$$(a)\ L[X] = AX, \qquad (b)\ R[X] = XB, \qquad (c)\ K[X] = AXB.$$

7.1.16. The domain space of the following functions is the space of $n \times n$ real matrices A. Which are linear? What is the codomain space in each case? (a) $L[A] = 3A$; (b) $L[A] = \mathrm{I} - A$; (c) $L[A] = A^T$; (d) $L[A] = A^{-1}$; (e) $L[A] = \det A$; (f) $L[A] = \operatorname{tr} A$; (g) $L[A] = (a_{11}, \ldots, a_{nn})^T$, i.e., the vector of diagonal entries of A; (h) $L[A] = A\mathbf{v}$, where $\mathbf{v} \in \mathbb{R}^n$; (i) $L[A] = \mathbf{v}^T A \mathbf{v}$, where $\mathbf{v} \in \mathbb{R}^n$.

♢ 7.1.17. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis of V and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be any vectors in W. Show that there is a unique linear function $L\colon V \to W$ such that $L[\mathbf{v}_i] = \mathbf{w}_i$, $i = 1, \ldots, n$.

♡ 7.1.18. *Bilinear functions*: Let V, W, Z be vector spaces. A function that takes any pair of vectors $\mathbf{v} \in V$ and $\mathbf{w} \in W$ to a vector $\mathbf{z} = B[\mathbf{v}, \mathbf{w}] \in Z$ is called *bilinear* if, for each fixed $\mathbf{w}$, it is a linear function of $\mathbf{v}$, so $B[c\mathbf{v} + d\tilde{\mathbf{v}}, \mathbf{w}] = c\,B[\mathbf{v}, \mathbf{w}] + d\,B[\tilde{\mathbf{v}}, \mathbf{w}]$, and, for each fixed $\mathbf{v}$, it is a linear function of $\mathbf{w}$, so $B[\mathbf{v}, c\mathbf{w} + d\tilde{\mathbf{w}}] = c\,B[\mathbf{v}, \mathbf{w}] + d\,B[\mathbf{v}, \tilde{\mathbf{w}}]$. Thus, $B\colon V \times W \to Z$ defines a function on the Cartesian product space $V \times W$, as defined in Exercise 2.1.13. (a) Show that $B[\mathbf{v}, \mathbf{w}] = v_1 w_1 - 2 v_2 w_2$ is a bilinear function from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$. (b) Show that $B[\mathbf{v}, \mathbf{w}] = 2 v_1 w_2 - 3 v_2 w_3$ is a bilinear function from $\mathbb{R}^2 \times \mathbb{R}^3$ to $\mathbb{R}$. (c) Show that if V is an inner product space, then $B[\mathbf{v}, \mathbf{w}] = \langle \mathbf{v}, \mathbf{w} \rangle$ defines a bilinear function $B\colon V \times V \to \mathbb{R}$. (d) Show that if A is any $m \times n$ matrix, then $B[\mathbf{v}, \mathbf{w}] = \mathbf{v}^T A \mathbf{w}$ defines a bilinear function $B\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$. (e) Show that every bilinear function $B\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ arises in this way. (f) Show that a vector-valued function $B\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^k$ defines a bilinear function if and only if each of its components $B_i\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \ldots, k$, is a bilinear function. (g) *True or false*: A bilinear function $B\colon V \times W \to Z$ defines a linear function on the Cartesian product space.

Linear Operators

So far, we have concentrated on linear functions on Euclidean space, and discovered that they are all represented by matrices. For function spaces, there is a much wider variety of linear operators available, and a complete classification is out of the question. Let us look at some of the main representative examples that arise in applications.

Example 7.7. (a) Recall that $\mathrm{C}^0[a, b]$ denotes the vector space consisting of all continuous functions on the interval $[a, b]$. Evaluation of the function at a point, namely $L[f] = f(x_0)$, defines a linear operator $L\colon \mathrm{C}^0[a, b] \to \mathbb{R}$, because

$$L[cf + dg] = (cf + dg)(x_0) = c\,f(x_0) + d\,g(x_0) = c\,L[f] + d\,L[g]$$

for any functions $f, g \in \mathrm{C}^0[a, b]$ and scalars (constants) c, d.

(b) Another real-valued linear function is the integration operator

$$I[f] = \int_a^b f(x)\,dx, \tag{7.9}$$

that maps $I\colon \mathrm{C}^0[a, b] \to \mathbb{R}$. Linearity of I is an immediate consequence of the basic integration identity

$$\int_a^b \left[\, c f(x) + d g(x) \,\right] dx = c \int_a^b f(x)\,dx \; + \; d \int_a^b g(x)\,dx,$$

which is valid for arbitrary integrable — which includes continuous — functions f, g and constants c, d.

(c) We have already seen that multiplication of functions by a constant, $M_c[\,f(x)\,] = c\,f(x)$, defines a linear map $M_c\colon \mathrm{C}^0[\,a,b\,] \to \mathrm{C}^0[\,a,b\,]$; the particular case $c = 1$ reduces to the identity transformation $\mathrm{I} = M_1$. More generally, if $a(x) \in \mathrm{C}^0[\,a,b\,]$ is a given continuous function, then the operation $M_a[\,f(x)\,] = a(x)\,f(x)$ of multiplication by a also defines a linear transformation $M_a\colon \mathrm{C}^0[\,a,b\,] \to \mathrm{C}^0[\,a,b\,]$.

(d) Another important linear transformation is the indefinite integral

$$J[\,f\,] = g, \qquad \text{where} \qquad g(x) = \int_a^x f(y)\,dy. \tag{7.10}$$

According to the Fundamental Theorem of Calculus, [**2**, **78**], the integral of a continuous function is continuously differentiable. Therefore, $J\colon \mathrm{C}^0[\,a,b\,] \to \mathrm{C}^1[\,a,b\,]$ defines a linear operator from the space of continuous functions to the space of continuously differentiable functions.

(e) Conversely, differentiation of functions is also a linear operation. To be precise, since not every continuous function can be differentiated, we take the domain space to be the vector space $\mathrm{C}^1[\,a,b\,]$ of continuously differentiable functions on the interval $[\,a,b\,]$. The derivative operator

$$D[\,f\,] = f' \tag{7.11}$$

defines a linear operator $D\colon \mathrm{C}^1[\,a,b\,] \to \mathrm{C}^0[\,a,b\,]$. This follows from the elementary differentiation formula

$$D[\,c\,f + d\,g\,] = (c\,f + d\,g)' = c\,f' + d\,g' = c\,D[\,f\,] + d\,D[\,g\,],$$

valid whenever c, d are constant.

Exercises

7.1.19. Which of the following define linear operators on the vector space $\mathrm{C}^1(\mathbb{R})$ of continuously differentiable scalar functions? What is the codomain?
(a) $L[\,f\,] = f(0) + f(1)$, (b) $L[\,f\,] = f(0)\,f(1)$, (c) $L[\,f\,] = f'(1)$, (d) $L[\,f\,] = f'(3) - f(2)$, (e) $L[\,f\,] = x^2\,f(x)$, (f) $L[\,f\,] = f(x+2)$, (g) $L[\,f\,] = f(x) + 2$, (h) $L[\,f\,] = f'(2\,x)$, (i) $L[\,f\,] = f'(x^2)$, (j) $L[\,f\,] = f(x)\sin x - f'(x)\cos x$, (k) $L[\,f\,] = 2\log f(0)$, (l) $L[\,f\,] = \displaystyle\int_0^1 e^y\,f(y)\,dy$, (m) $L[\,f\,] = \displaystyle\int_0^1 |\,f(y)\,|\,dy$, (n) $L[\,f\,] = \displaystyle\int_{x-1}^{x+1} f(y)\,dy$, (o) $L[\,f\,] = \displaystyle\int_x^{x^2} \frac{f(y)}{y}\,dy$, (p) $L[\,f\,] = \displaystyle\int_0^{f(x)} y\,dy$, (q) $L[\,f\,] = \displaystyle\int_0^x y^2\,f'(y)\,dy$, (r) $L[\,f\,] = \displaystyle\int_{-1}^1 [\,f(y) - f(0)\,]\,dy$, (s) $L[\,f\,] = \displaystyle\int_{-1}^x [\,f(y) - y\,]\,dy$.

7.1.20. *True or false*: The average or mean $A[\,f\,] = \dfrac{1}{b-a}\displaystyle\int_a^b f(x)\,dx$ of a function on the interval $[\,a,b\,]$ defines a linear operator $A\colon \mathrm{C}^0[\,a,b\,] \to \mathbb{R}$.

7.1.21. Prove that multiplication $M_h[\,f(x)\,] = h(x)\,f(x)$ by a given function $h \in \mathrm{C}^n[\,a,b\,]$ defines a linear operator $M_h\colon \mathrm{C}^n[\,a,b\,] \to \mathrm{C}^n[\,a,b\,]$. Which result from calculus do you need to complete the proof?

7.1.22. Show that if $w(x)$ is any continuous function, then the weighted integral

$$I_w[f] = \int_a^b f(x)\,w(x)\,dx \text{ defines a linear operator } I_w\colon \mathrm{C}^0[a,b] \to \mathbb{R}.$$

7.1.23. (a) Show that the partial derivatives $\partial_x[f] = \dfrac{\partial f}{\partial x}$ and $\partial_y[f] = \dfrac{\partial f}{\partial y}$ both define linear operators on the space of continuously differentiable functions $f(x,y)$.

(b) For which values of a,b,c,d is the map $L[f] = a\,\dfrac{\partial f}{\partial x} + b\,\dfrac{\partial f}{\partial y} + c\,f + d$ linear?

7.1.24. Prove that the Laplacian operator $\Delta[f] = \dfrac{\partial^2 f}{\partial x^2} + \dfrac{\partial^2 f}{\partial y^2}$ defines a linear function on the vector space of twice continuously differentiable functions $f(x,y)$.

7.1.25. Show that the gradient $G[f] = \nabla f$ defines a linear operator from the space of continuously differentiable scalar-valued functions $f\colon \mathbb{R}^2 \to \mathbb{R}$ to the space of continuous vector fields $\mathbf{v}\colon \mathbb{R}^2 \to \mathbb{R}^2$.

7.1.26. Prove that, on $\mathbb{R}^3$, the gradient, curl, and divergence all define linear operators. Be precise in your description of the domain space and the codomain space in each case.

The Space of Linear Functions

Given two vector spaces V, W, we use $\mathcal{L}(V,W)$ to denote the set of all[†] linear functions $L\colon V \to W$. We claim that $\mathcal{L}(V,W)$ is itself a vector space. We add linear functions $L, M \in \mathcal{L}(V,W)$ in the same way we add general functions:

$$(L+M)[\mathbf{v}] = L[\mathbf{v}] + M[\mathbf{v}].$$

You should check that $L+M$ satisfies the linear function axioms (7.1), provided that L and M do. Similarly, multiplication of a linear function by a scalar $c \in \mathbb{R}$ is defined so that $(c\,L)[\mathbf{v}] = c\,L[\mathbf{v}]$, again producing a linear function. The zero element of $\mathcal{L}(V,W)$ is the zero function $\mathrm{O}[\mathbf{v}] \equiv \mathbf{0}$. The verification that $\mathcal{L}(V,W)$ satisfies the basic vector space axioms of Definition 2.1 is left to the reader.

In particular, if $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, then Theorem 7.5 implies that we can identify $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with the space $\mathcal{M}_{m\times n}$ of all $m \times n$ matrices. Addition of linear functions corresponds to matrix addition, while scalar multiplication coincides with the usual scalar multiplication of matrices. (Why?) Therefore, the space of all $m \times n$ matrices is a vector space — a fact we already knew. The *standard basis* for $\mathcal{M}_{m\times n}$ is given by the $m\,n$ matrices E_{ij}, $1 \le i \le m$, $1 \le j \le n$, which have a single 1 in the (i,j) position and zeros everywhere else. Therefore, the dimension of $\mathcal{M}_{m\times n}$ is $m\,n$. Note that E_{ij} corresponds to the specific linear transformation that maps $E_{ij}[\mathbf{e}_j] = \widehat{\mathbf{e}}_i$, while $E_{ij}[\mathbf{e}_k] = \mathbf{0}$ whenever $k \ne j$.

Example 7.8. The space of linear transformations of the plane, $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$, is identified with the space $\mathcal{M}_{2\times 2}$ of 2×2 matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The standard basis of $\mathcal{M}_{2\times 2}$ consists

[†] In infinite-dimensional situations, one usually imposes additional restrictions, e.g., continuity or boundedness of the linear operators. We shall relegate these more subtle distinctions to a more advanced treatment of the subject. See [**50**, **67**] for a full discussion of the rather sophisticated analytical details, which play an important role in serious quantum mechanical applications.

of the $4 = 2 \cdot 2$ matrices

$$E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Indeed, we can uniquely write any other matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\,E_{11} + b\,E_{12} + c\,E_{21} + d\,E_{22},$$

as a linear combination of these four basis matrices. Of course, as with any vector space, this is but one of many other possible bases of $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$.

Dual Spaces

A particularly important case is that in which the codomain of the linear functions is $\mathbb{R}$.

Definition 7.9. The *dual space* to a vector space V is the vector space $V^* = \mathcal{L}(V, \mathbb{R})$ consisting of all real-valued linear functions $\ell : V \to \mathbb{R}$.

If $V = \mathbb{R}^n$, then, by Theorem 7.5, every linear function $\ell : \mathbb{R}^n \to \mathbb{R}$ is given by multiplication by a $1 \times n$ matrix, i.e., a row vector. Explicitly,

$$\ell[\mathbf{v}] = \mathbf{a}\,\mathbf{v} = a_1 v_1 + \;\cdots\; + a_n v_n, \qquad \text{where} \qquad \mathbf{a} = (\,a_1\;\, a_2\; \ldots\; a_n\,), \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

Therefore, we can identify the dual space $(\mathbb{R}^n)^*$ with the space of *row* vectors with n entries. In light of this observation, the distinction between row vectors and column vectors is now seen to be much more sophisticated than mere semantics or notation. Row vectors should more properly be viewed as real-valued linear functions — the *dual* objects to column vectors.

The *standard dual basis* $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ of $(\mathbb{R}^n)^*$ consists of the standard row basis vectors; namely, $\boldsymbol{\varepsilon}_j$ is the row vector with 1 in the j^{th} slot and zeros elsewhere. The j^{th} dual basis element defines the linear function

$$E_j[\mathbf{v}] = \boldsymbol{\varepsilon}_j\,\mathbf{v} = v_j,$$

which picks off the j^{th} coordinate of $\mathbf{v}$ — with respect to the original basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Thus, the dimensions of $V = \mathbb{R}^n$ and its dual $V^* = (\mathbb{R}^n)^*$ are both equal to n.

An inner product structure provides a mechanism for identifying a vector space and its dual. However, it should be borne in mind that this identification will depend upon the choice of inner product.

Theorem 7.10. Let V be a finite-dimensional real inner product space. Then every linear function $\ell : V \to \mathbb{R}$ is given by taking the inner product with a fixed vector $\mathbf{a} \in V$:

$$\ell[\mathbf{v}] = \langle\, \mathbf{a}\,, \mathbf{v}\,\rangle. \tag{7.12}$$

Proof: Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis of V. If we write $\mathbf{v} = y_1 \mathbf{v}_1 + \cdots + y_n \mathbf{v}_n$, then, by linearity,

$$\ell[\mathbf{v}] = y_1\,\ell[\mathbf{v}_1] + \;\cdots\; + y_n\,\ell[\mathbf{v}_n] = b_1\,y_1 + \;\cdots\; + b_n\,y_n, \qquad \text{where} \qquad b_i = \ell[\mathbf{u}_i]. \tag{7.13}$$

On the other hand, if we write $\mathbf{a} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$, then

$$\langle \mathbf{a}, \mathbf{v} \rangle = \sum_{i,j=1}^{n} x_j\, y_i \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \sum_{i,j=1}^{n} g_{ij} x_j y_i, \tag{7.14}$$

where $G = (g_{ij})$ is the $n \times n$ Gram matrix with entries $g_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$. Equality of (7.13, 14) requires that $G\mathbf{x} = \mathbf{b}$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, $\mathbf{b} = (b_1, b_2, \ldots, b_n)^T$. Invertibility of G as guaranteed by Theorem 3.34, allows us to solve for $\mathbf{x} = G^{-1}\mathbf{b}$ and thereby construct the desired vector $\mathbf{a}$. In particular, if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an orthonormal basis, then $G = \mathrm{I}$ and hence $\mathbf{a} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$. *Q.E.D.*

Remark. For the particular case in which $V = \mathbb{R}^n$ is endowed with the standard dot product, Theorem 7.10 identifies a row vector representing a linear function with the corresponding column vector obtained by transposition $\mathbf{a} \mapsto \mathbf{a}^T$. Thus, the naïve identification of a row and a column vector is, in fact, an indication of a much more subtle phenomenon that relies on the identification of $\mathbb{R}^n$ with its dual based on the Euclidean inner product. Alternative inner products will lead to alternative, more complicated, identifications of row and column vectors; see Exercise 7.1.31 for details.

Important. Theorem 7.10 is *not* true if V is infinite-dimensional. This fact will have important repercussions for the analysis of the differential equations of continuum mechanics, which will lead us immediately into the much deeper waters of generalized function theory, as described in [**61**].

Exercises

7.1.27. Write down a basis for and dimension of the linear function spaces (*a*) $\mathcal{L}(\mathbb{R}^3, \mathbb{R})$, (*b*) $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$, (*c*) $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$, (*d*) $\mathcal{L}(\mathcal{P}^{(3)}, \mathbb{R})$, (*e*) $\mathcal{L}(\mathcal{P}^{(2)}, \mathbb{R}^2)$, (*f*) $\mathcal{L}(\mathcal{P}^{(2)}, \mathcal{P}^{(2)})$. Here $\mathcal{P}^{(n)}$ is the space of polynomials of degree $\leq n$.

7.1.28. *True or false*: The set of linear transformations $L: \mathbb{R}^2 \to \mathbb{R}^2$ such that $L\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$ is a subspace of $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$. If true, what is its dimension?

7.1.29. *True or false*: The set of linear transformations $L: \mathbb{R}^3 \to \mathbb{R}^3$ such that $L\begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}0\\1\\0\end{pmatrix}$ is a subspace of $\mathcal{L}(\mathbb{R}^3, \mathbb{R}^3)$. If true, what is its dimension?

7.1.30. Consider the linear function $L: \mathbb{R}^3 \to \mathbb{R}$ defined by $L(x, y, z) = 3x - y + 2z$. Write down the vector $\mathbf{a} \in \mathbb{R}^3$ such that $L[\mathbf{v}] = \langle \mathbf{a}, \mathbf{v} \rangle$ when the inner product is (*a*) the Euclidean dot product; (*b*) the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + 2 v_2 w_2 + 3 v_3 w_3$; (*c*) the inner product defined by the positive definite matrix $K = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$.

◇ 7.1.31. Let $\mathbb{R}^n$ be equipped with the inner product $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T K \mathbf{w}$. Let $L[\mathbf{v}] = \mathbf{r}\,\mathbf{v}$ where $\mathbf{r}$ is a row vector of size $1 \times n$. (*a*) Find a formula for the column vector $\mathbf{a}$ such that (7.12) holds for the linear function $L: \mathbb{R}^n \to \mathbb{R}$. (*b*) Illustrate your result when $\mathbf{r} = (2, -1)$, using (*i*) the dot product (*ii*) the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = 3 v_1 w_1 + 2 v_2 w_2$, (*iii*) the inner product induced by $K = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$.

♡ 7.1.32. *Dual Bases*: Given a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V, the *dual basis* $\ell_1, \ldots, \ell_n$ of V^* consists of the linear functions uniquely defined by the requirements $\ell_i(\mathbf{v}_j) = \begin{cases} 1 & i = j, \\ 0, & i \neq j. \end{cases}$
(a) Show that $\ell_i[\mathbf{v}] = x_i$ gives the i^{th} coordinate of a vector $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ with respect to the given basis. (b) Prove that the dual basis is indeed a basis for the dual vector space. (c) Prove that if $V = \mathbb{R}^n$ and $A = (\,\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_n\,)$ is the $n \times n$ matrix whose columns are the basis vectors, then the rows of the inverse matrix A^{-1} can be identified as the corresponding dual basis of $(\mathbb{R}^n)^*$.

7.1.33. Use Exercise 7.1.32(c) to find the dual basis for: (a) $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$; (b) $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$; (c) $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $\mathbf{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$; (d) $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 0 \\ -3 \\ 1 \end{pmatrix}$, $\mathbf{v}_3 = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$; (e) $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$, $\mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$, $\mathbf{v}_4 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 2 \end{pmatrix}$.

7.1.34. Let $\mathcal{P}^{(2)}$ denote the space of quadratic polynomials equipped with the L^2 inner product $\langle\, p\,, q\,\rangle = \int_0^1 p(x)\, q(x)\, dx$. Find the polynomial q that represents the following linear functions, i.e., such that $L[p] = \langle\, q\,, p\,\rangle$: (a) $L[p] = p(0)$, (b) $L[p] = \frac{1}{2}\, p'(1)$, (c) $L[p] = \int_0^1 p(x)\, dx$, (d) $L[p] = \int_{-1}^1 p(x)\, dx$.

7.1.35. Find the dual basis, as defined in Exercise 7.1.32, for the monomial basis of $\mathcal{P}^{(2)}$ with respect to the L^2 inner product $\langle\, p\,, q\,\rangle = \int_0^1 p(x)\, q(x)\, dx$.

7.1.36. Write out a proof of Theorem 7.10 that does not rely on finding an orthonormal basis.

Composition

Besides adding and multiplying by scalars, one can also compose linear functions.

Lemma 7.11. Let V, W, Z be vector spaces. If $L\colon V \to W$ and $M\colon W \to Z$ are linear functions, then the composite function $M \circ L\colon V \to Z$, defined by $(M \circ L)[\mathbf{v}] = M[L[\mathbf{v}]]$ is also linear.

Proof: This is straightforward:

$$\begin{aligned}(M \circ L)[c\,\mathbf{v} + d\,\mathbf{w}] &= M[L[c\,\mathbf{v} + d\,\mathbf{w}]] = M[c\,L[\mathbf{v}] + d\,L[\mathbf{w}]] \\ &= c\,M[L[\mathbf{v}]] + d\,M[L[\mathbf{w}]] = c\,(M \circ L)[\mathbf{v}] + d\,(M \circ L)[\mathbf{w}],\end{aligned}$$

where we used, successively, the linearity of L and then of M. *Q.E.D.*

For example, if $L[\mathbf{v}] = A\mathbf{v}$ maps $\mathbb{R}^n$ to $\mathbb{R}^m$, and $M[\mathbf{w}] = B\mathbf{w}$ maps $\mathbb{R}^m$ to $\mathbb{R}^l$, so that A is an $m \times n$ matrix and B is a $l \times m$ matrix, then

$$(M \circ L)[\mathbf{v}] = M[L[\mathbf{v}]] = B(A\mathbf{v}) = (B\,A)\,\mathbf{v},$$

and hence the composition $M \circ L\colon \mathbb{R}^n \to \mathbb{R}^l$ corresponds to the $l \times n$ product matrix BA. In other words, on Euclidean space, *composition of linear functions is the same as matrix multiplication*. And, like matrix multiplication, composition of (linear) functions is not, in general, commutative.

Example 7.12. Composing two rotations results in another rotation: $R_\varphi \circ R_\theta = R_{\varphi+\theta}$. In other words, if we first rotate by angle θ and then by angle φ, the net effect is rotation by angle $\varphi+\theta$. On the matrix level of (7.8), this implies that

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = A_\varphi\, A_\theta = A_{\varphi+\theta} = \begin{pmatrix} \cos(\varphi+\theta) & -\sin(\varphi+\theta) \\ \sin(\varphi+\theta) & \cos(\varphi+\theta) \end{pmatrix}.$$

Multiplying out the left-hand side, we deduce the well-known trigonometric addition formulas

$$\cos(\varphi+\theta) = \cos\varphi\,\cos\theta - \sin\varphi\,\sin\theta, \qquad \sin(\varphi+\theta) = \cos\varphi\,\sin\theta + \sin\varphi\,\cos\theta.$$

In fact, this constitutes a *bona fide* proof of these two trigonometric identities!

Example 7.13. One can build up more sophisticated linear operators on function space by adding and composing simpler ones. In particular, higher order derivative operators are obtained by composing the derivative operator D, defined in (7.11), with itself. For example,

$$D^2[\,f\,] = D \circ D[\,f\,] = D[\,f'\,] = f''$$

defines the second derivative operator. One needs to exercise due care about the domain of definition, since not every function is differentiable. In general, the k^{th} order derivative

$$D^k[\,f\,] = f^{(k)}(x) \quad \text{defines a linear operator} \quad D^k\colon \mathrm{C}^n[\,a,b\,] \longrightarrow \mathrm{C}^{n-k}[\,a,b\,] \quad \text{for all } n \geq k,$$

obtained by composing D with itself k times.

If we further compose D^k with the linear operation of multiplication by a given function $a(x)$ we obtain the linear operator $(a\,D^k)[\,f\,] = a(x)\,f^{(k)}(x)$. Finally, a general *linear ordinary differential operator* of order n,

$$L = a_n(x)\,D^n + a_{n-1}(x)\,D^{n-1} + \;\cdots\; + a_1(x)\,D + a_0(x), \tag{7.15}$$

is obtained by summing such operators. If the coefficient functions $a_0(x), \ldots, a_n(x)$ are continuous, then

$$L[\,u\,] = a_n(x)\,\frac{d^n u}{dx^n} + a_{n-1}(x)\,\frac{d^{n-1}u}{dx^{n-1}} + \;\cdots\; + a_1(x)\,\frac{du}{dx} + a_0(x)u \tag{7.16}$$

defines a linear operator from $\mathrm{C}^n[\,a,b\,]$ to $\mathrm{C}^0[\,a,b\,]$. The most important case — but certainly not the only one arising in applications — is when the coefficients $a_i(x) = c_i$ are all constant.

Exercises

7.1.37. For each of the following pairs of linear functions $S, T\colon \mathbb{R}^2 \to \mathbb{R}^2$, describe the compositions $S \circ T$ and $T \circ S$. Do the functions commute?
(*a*) $S =$ counterclockwise rotation by 60°; $T =$ clockwise rotation by 120°;
(*b*) $S =$ reflection in the line $y = x$; $T =$ rotation by 180°;
(*c*) $S =$ reflection in the x-axis; $T =$ reflection in the y-axis;
(*d*) $S =$ reflection in the line $y = x$; $T =$ reflection in the line $y = 2x$;
(*e*) $S =$ orthogonal projection on the x-axis; $T =$ orthogonal projection on the y-axis;
(*f*) $S =$ orthogonal projection on the x-axis; $T =$ orthogonal projection on the line $y = x$;
(*g*) $S =$ orthogonal projection on the x-axis; $T =$ rotation by 180°;
(*h*) $S =$ orthogonal projection on the x-axis; $T =$ counterclockwise rotation by 90°;
(*i*) $S =$ orthogonal projection on the line $y = -2x$; $T =$ reflection in the line $y = x$.

7.1.38. Find a matrix representative for the linear functions (a) $L:\mathbb{R}^2 \to \mathbb{R}^2$ that maps $\mathbf{e}_1$ to $\begin{pmatrix} 1 \\ -3 \end{pmatrix}$ and $\mathbf{e}_2$ to $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$; (b) $M:\mathbb{R}^2 \to \mathbb{R}^2$ that takes $\mathbf{e}_1$ to $\begin{pmatrix} -1 \\ -3 \end{pmatrix}$ and $\mathbf{e}_2$ to $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$; and (c) $N:\mathbb{R}^2 \to \mathbb{R}^2$ that takes $\begin{pmatrix} 1 \\ -3 \end{pmatrix}$ to $\begin{pmatrix} -1 \\ -3 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ to $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$. (d) Explain why $M = N \circ L$. (e) Verify part (d) by multiplying the matrix representatives.

7.1.39. On the vector space $\mathbb{R}^3$, let R denote counterclockwise rotation around the x axis by 90° and S counterclockwise rotation around the z-axis by 90°. (a) Find matrix representatives for R and S. (b) Show that $R \circ S \neq S \circ R$. Explain what happens to the standard basis vectors under the two compositions. (c) Give an experimental demonstration of the noncommutativity of R and S by physically rotating a solid object, e.g., this book, in the prescribed manners.

7.1.40. Let P denote orthogonal projection of $\mathbb{R}^3$ onto the plane $V = \{ z = x + y \}$ and Q denote orthogonal projection onto the plane $W = \{ z = x - y \}$. Is the composition $R = Q \circ P$ the same as orthogonal projection onto the line $L = V \cap W$? Verify your conclusion by computing the matrix representatives of P, Q, and R.

7.1.41. (a) Write the linear operator $L[f(x)] = f'(b)$ as a composition of two linear functions. Do your linear functions commute? (b) For which values of a, b, c, d, e is $L[f(x)] = a\, f'(b) + c\, f(d) + e$ a linear function?

7.1.42. Let $L = x\,D + 1$, and $M = D - x$ be differential operators. Find $L \circ M$ and $M \circ L$. Do the differential operators commute?

7.1.43. Show that the space of constant coefficient linear differential operators of order $\leq n$ is a vector space. Determine its dimension by exhibiting a basis.

7.1.44. (a) Explain why the differential operator $L = D \circ M_a \circ D$ obtained by composing the linear operators of differentiation $D[f(x)] = f'(x)$ and multiplication $M_a[f(x)] = a(x)\, f(x)$ by a given function $a(x)$ defines a linear operator. (b) Re-express L as a linear differential operator of the form (7.16).

♢ 7.1.45. (a) Show that composition of linear functions is associative: $(L \circ M) \circ N = L \circ (M \circ N)$. Be precise about the domain and codomain spaces involved. (b) How do you know the result is a linear function? (c) Explain why this proves associativity of matrix multiplication.

7.1.46. Show that if $p(x, y)$ is any polynomial, then $L = p(\partial_x, \partial_y)$ defines a linear, constant coefficient partial differential operator. For example, if $p(x, y) = x^2 + y^2$, then $L = \partial_x^2 + \partial_y^2$ is the Laplacian operator $\Delta[f] = \dfrac{\partial^2 f}{\partial x^2} + \dfrac{\partial^2 f}{\partial y^2}$.

♡ 7.1.47. The *commutator* of two linear transformations $L, M: V \to V$ on a vector space V is

$$K = [L, M] = L \circ M - M \circ L. \tag{7.17}$$

(a) Prove that the commutator K is a linear transformation on V. (b) Explain why Exercise 1.2.30 is a special case. (c) Prove that L and M commute if and only if $[L, M] = \mathrm{O}$. (d) Compute the commutators of the linear transformations defined by the following pairs of matrices:
(i) $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 1 & 2 \end{pmatrix}$, (ii) $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, (iii) $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.
(e) Prove that the *Jacobi identity*

$$\big[[L, M], N \big] + \big[[N, L], M \big] + \big[[M, N], L \big] = \mathrm{O} \tag{7.18}$$

is valid for any three linear transformations. (f) Verify the Jacobi identity for the first three matrices in part (c). (g) Prove that the commutator $B[L, M] = [L, M]$ defines a bilinear map $B: \mathcal{L}(V, V) \times \mathcal{L}(V, V) \to \mathcal{L}(V, V)$ on the Cartesian product space, cf. Exercise 7.1.18.

♢ 7.1.48. (a) In (one-dimensional) quantum mechanics, the differentiation operator $P[\,f(x)\,] = f'(x)$ represents the *momentum* of a particle, while the operator $Q[\,f(x)\,] = x\,f(x)$ of multiplication by the function x represents its *position*. Prove that the position and momentum operators satisfy the *Heisenberg Commutation Relations* $[\,P, Q\,] = P \circ Q - Q \circ P = \mathrm{I}$. (b) Prove that there are no matrices P, Q that satisfy the Heisenberg Commutation Relations. *Hint*: Use Exercise 1.2.31.

Remark. The noncommutativity of quantum mechanical observables lies at the heart of the Uncertainty Principle. The result in part (b) is one of the main reasons why quantum mechanics must be an intrinsically infinite-dimensional theory.

♡ 7.1.49. Let $\mathcal{D}^{(1)}$ denote the set of all first order linear differential operators $L = p(x)\,D + q(x)$ where p, q are polynomials. (a) Prove that $\mathcal{D}^{(1)}$ is a vector space. Is it finite-dimensional or infinite-dimensional? (b) Prove that the commutator (7.17) of $L, M \in \mathcal{D}^{(1)}$ is a first order differential operator $[\,L, M\,] \in \mathcal{D}^{(1)}$ by writing out an explicit formula. (c) Verify the Jacobi identity (7.18) for the first order operators $L = D$, $M = x\,D + 1$, and $N = x^2\,D + 2x$.

7.1.50. Do the conclusions of Exercise 7.1.49(a–b) hold for the space $\mathcal{D}^{(2)}$ of second order differential operators $L = p(x)\,D^2 + q(x)\,D + r(x)$, where p, q, r are polynomials?

Inverses

The inverse of a linear function is defined in direct analogy with the Definition 1.13 of the inverse of a (square) matrix.

Definition 7.14. Let $L\colon V \to W$ be a linear function. If $M\colon W \to V$ is a function such that both compositions

$$L \circ M = \mathrm{I}_W, \qquad M \circ L = \mathrm{I}_V, \tag{7.19}$$

are equal to the identity function, then we call M the *inverse* of L and write $M = L^{-1}$.

The two conditions (7.19) require

$$L[\,M[\,\mathbf{w}\,]\,] = \mathbf{w} \quad \text{for all} \quad \mathbf{w} \in W, \qquad \text{and} \qquad M[\,L[\,\mathbf{v}\,]\,] = \mathbf{v} \quad \text{for all} \quad \mathbf{v} \in V.$$

In Exercise 7.1.55, you are asked to prove that, when it exists, the inverse is unique. Of course, if $M = L^{-1}$ is the inverse of L, then $L = M^{-1}$ is the inverse of M since the conditions are symmetric, and, in such cases, $(L^{-1})^{-1} = L$.

Lemma 7.15. If it exists, the inverse of a linear function is also a linear function.

Proof: Let L, M satisfy the conditions of Definition 7.14. Given $\mathbf{w}, \widetilde{\mathbf{w}} \in W$, we note

$$\mathbf{w} = (L \circ M)[\,\mathbf{w}\,] = L[\,\mathbf{v}\,], \quad \widetilde{\mathbf{w}} = (L \circ M)[\,\widetilde{\mathbf{w}}\,] = L[\,\widetilde{\mathbf{v}}\,], \quad \text{where} \quad \mathbf{v} = M[\,\mathbf{w}\,], \quad \widetilde{\mathbf{v}} = M[\,\widetilde{\mathbf{w}}\,].$$

Therefore, given scalars c, d, and using only the linearity of L,

$$M[\,c\,\mathbf{w} + d\,\widetilde{\mathbf{w}}\,] = M[\,c\,L[\,\mathbf{v}\,] + d\,L[\,\widetilde{\mathbf{v}}\,]\,] = (M \circ L)[\,c\,\mathbf{v} + d\,\widetilde{\mathbf{v}}\,] = c\,\mathbf{v} + d\,\widetilde{\mathbf{v}} = c\,M[\,\mathbf{w}\,] + d\,M[\,\widetilde{\mathbf{w}}\,],$$

proving linearity of M. *Q.E.D.*

If $V = \mathbb{R}^n$, $W = \mathbb{R}^m$, so that L and M are given by matrix multiplication, by A and B respectively, then the conditions (7.19) reduce to the usual conditions

$$A\,B = \mathrm{I}, \qquad B\,A = \mathrm{I},$$

for matrix inversion, cf. (1.37). Therefore, $B = A^{-1}$ is the inverse matrix. In particular, for $L\colon \mathbb{R}^m \to \mathbb{R}^n$ to have an inverse, we must have $m = n$, and its coefficient matrix A must be nonsingular.

The invertibility of linear transformations on infinite-dimensional function spaces is more subtle. Here is a familiar example from calculus.

Example 7.16. The Fundamental Theorem of Calculus says, roughly, that differentiation $D[f] = f'$ and (indefinite) integration $J[f] = g$, where $g(x) = \displaystyle\int_a^x f(y)\,dy$, are "inverse" operations. More precisely, the derivative of the indefinite integral of f is equal to f, and hence

$$D[J[f]] = D[g] = g' = f, \qquad \text{since} \qquad g'(x) = \frac{d}{dx}\int_a^x f(y)\,dy = f(x).$$

In other words, the composition $D \circ J = \mathrm{I}_{\mathrm{C}^0[a,b]}$ defines the identity operator on the function space $\mathrm{C}^0[a,b]$. On the other hand, if we integrate the derivative of a continuously differentiable function $f \in \mathrm{C}^1[a,b]$, we obtain $J[D[f]] = J[f'] = h$, where

$$h(x) = \int_a^x f'(y)\,dy = f(x) - f(a) \neq f(x) \qquad \text{unless} \qquad f(a) = 0.$$

Therefore, the composition is *not* the identity operator: $J \circ D \neq \mathrm{I}_{\mathrm{C}^1[a,b]}$. In other words, the differentiation operator D is a left inverse for the integration operator J but not a right inverse!

If we restrict D to the subspace $V = \{ f \mid f(a) = 0 \} \subset \mathrm{C}^1[a,b]$ consisting of all continuously differentiable functions that vanish at the left-hand endpoint, then $J\colon \mathrm{C}^0[a,b] \to V$, and $D\colon V \to \mathrm{C}^0[a,b]$ are, by the preceding argument, inverse linear operators: $D \circ J = \mathrm{I}_{\mathrm{C}^0[a,b]}$, and $J \circ D = \mathrm{I}_V$. Note that $V \subsetneq \mathrm{C}^1[a,b] \subsetneq \mathrm{C}^0[a,b]$. Thus, we discover the curious and disconcerting infinite-dimensional phenomenon that J defines a one-to-one, invertible, linear map from a vector space $\mathrm{C}^0[a,b]$ to a proper subspace $V \subsetneq \mathrm{C}^0[a,b]$. This paradoxical situation *cannot* occur in finite dimensions. A linear map $L\colon \mathbb{R}^n \to \mathbb{R}^n$ can be invertible only when its image is the entire space — because it represents multiplication by a nonsingular square matrix.

Two vector spaces V, W are said to be *isomorphic*, written $V \simeq W$, if there exists an invertible linear function $L\colon V \to W$. For example, if V is finite-dimensional, then $V \simeq W$ if and only if W has the same dimension as V. In particular, if V has dimension n, then $V \simeq \mathbb{R}^n$. One way to construct the required invertible linear map is the choose a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V, and map it to the standard basis of $\mathbb{R}^n$, so $L[\mathbf{v}_k] = \mathbf{e}_k$ for $k = 1, \ldots, n$. In general, given $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$, then, by linearity,

$$\begin{aligned} L[v] = L[x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n] &= x_1 L[\mathbf{v}_1] + \cdots + x_n L[\mathbf{v}_n] \\ &= x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n = (x_1, x_2, \ldots, x_n)^T = \mathbf{x}, \end{aligned}$$

and hence L maps $\mathbf{v}$ to the column vector $\mathbf{x} = \mathbb{R}^n$ whose entries are its coordinates with respect to the chosen basis. The inverse $L^{-1}\colon \mathbb{R}^n \to V$ maps $\mathbf{x} \in \mathbb{R}^n$ to the element $L^{-1}[\mathbf{x}] = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n \in V$. As the above example makes clear, isomorphism of infinite-dimensional vector spaces is more subtle, and one often imposes additional restrictions on the allowable linear maps.

Exercises

7.1.51. Determine which of the following linear functions $L:\mathbb{R}^2 \to \mathbb{R}^2$ has an inverse, and, if so, describe it: (*a*) the scaling transformation that doubles the length of each vector; (*b*) clockwise rotation by 45°; (*c*) reflection through the y-axis; (*d*) orthogonal projection onto the line $y = x$; (*e*) the shearing transformation defined by the matrix $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$.

7.1.52. For each of the linear functions in Exercise 7.1.51, write down its matrix representative, the matrix representative of its inverse, and verify that the matrices are mutual inverses.

7.1.53. Let $L:\mathbb{R}^2 \to \mathbb{R}^2$ be the linear function such that $L[\mathbf{e}_1] = (1,-1)^T$, $L[\mathbf{e}_2] = (3,-2)^T$. Find $L^{-1}[\mathbf{e}_1]$ and $L^{-1}[\mathbf{e}_2]$.

7.1.54. Let $L:\mathbb{R}^3 \to \mathbb{R}^3$ be the linear function such that $L[\mathbf{e}_1] = (2,1,-1)^T$, $L[\mathbf{e}_2] = (1,2,1)^T$, $L[\mathbf{e}_3] = (-1,2,2)^T$. Find $L^{-1}[\mathbf{e}_1]$, $L^{-1}[\mathbf{e}_2]$, and $L^{-1}[\mathbf{e}_3]$.

♢ 7.1.55. Prove that the inverse of a linear transformation is unique; i.e., given L, there is at most one linear transformation M that can satisfy (7.19).

♢ 7.1.56. Let $L:V \to W$ be a linear function. Suppose $M, N: W \to V$ are linear functions that satisfy $L \circ M = \mathrm{I}_V = N \circ L$. Prove that $M = N = L^{-1}$. Thus, a linear function may have only a left or a right inverse, but if it has both, then they must be the same.

7.1.57. Give an example of a matrix with a left inverse, but not a right inverse. Is your left inverse unique?

♡ 7.1.58. Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis for V and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ a basis for W. (*a*) Prove that there is a unique linear function $L:V \to W$ such that $L[\mathbf{v}_i] = \mathbf{w}_i$ for $i = 1, \ldots, n$. (*b*) Prove that L is invertible. (*c*) If $V = W = \mathbb{R}^n$, find a formula for the matrix representative of the linear functions L and L^{-1}. (*d*) Apply your construction to produce a linear function that takes: (*i*) $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to $\mathbf{w}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$, $\mathbf{w}_2 = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$,
(*ii*) $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ to $\mathbf{w}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\mathbf{w}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$,
(*iii*) $\mathbf{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $\mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ to $\mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{w}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $\mathbf{w}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

7.1.59. Suppose $V, W \subset \mathbb{R}^n$ are subspaces of the same dimension. Prove that there is an invertible linear function $L:\mathbb{R}^n \to \mathbb{R}^n$ that takes V to W. *Hint*: Use Exercise 7.1.58.

♢ 7.1.60. Let W, Z be complementary subspaces of a vector space V, as in Exercise 2.2.24. Let V/W denote the quotient vector space, as defined in Exercise 2.2.29. Show that the map $L: Z \to V/W$ that maps $L[\mathbf{z}] = [\mathbf{z}]_W$ defines an invertible linear map, and hence $Z \simeq V/W$ are isomorphic vector spaces.

♢ 7.1.61. Let $L:V \to W$ be a linear map. (*a*) Suppose V, W are finite-dimensional vector spaces, and let A be a matrix representative of L. Explain why we can identify $\operatorname{coker} A \simeq W/\operatorname{img} A$ and $\operatorname{coimg} A = V/\ker A$ as quotient vector spaces, cf. Exercise 2.2.29.
Remark. These characterizations are used to give intrinsic definitions of the cokernel and coimage of a general linear function $L:V \to W$ without any reference to a transpose (or, as defined below, adjoint) operation. Namely, set $\operatorname{coker} L \simeq W/\operatorname{img} L$ and $\operatorname{coimg} L = V/\ker L$.
(*b*) The *index* of the linear map is defined as $\operatorname{index} L = \dim\ker L - \dim\operatorname{coker} L$, using the above intrinsic definitions. Prove that, when V, W are finite-dimensional, $\operatorname{index} L = \dim V - \dim W$.

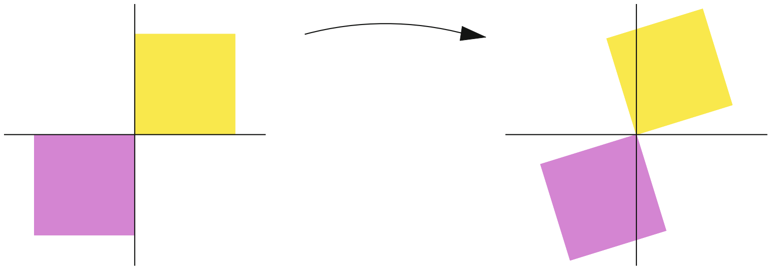


Figure 7.4. Rotation.

♢ 7.1.62. Let V be a finite-dimensional real inner product space and let V^* be its dual. Using Theorem 7.10, prove that the map $J\colon V^* \to V$ that takes the linear function $\ell \in V^*$ to the vector $J[\ell] = \mathbf{a} \in V$ satisfying $\ell[\mathbf{v}] = \langle \mathbf{a}, \mathbf{v} \rangle$ defines a linear isomorphism between the inner product space and its dual: $V^* \simeq V$.

7.1.63. (a) Prove that $L[p] = p' + p$ defines an invertible linear map on the space $\mathcal{P}^{(2)}$ of quadratic polynomials. Find a formula for its inverse.
(b) Does the derivative $D[p] = p'$ have either a left or a right inverse on $\mathcal{P}^{(2)}$?

♡ 7.1.64. (a) Show that the set of all functions of the form $f(x) = (a x^2 + b x + c)\, e^x$ for $a, b, c, \in \mathbb{R}$ is a vector space. What is its dimension? (b) Show that the derivative $D[f(x)] = f'(x)$ defines an invertible linear transformation on this vector space, and determine its inverse. (c) Generalize your result in part (b) to the infinite-dimensional vector space consisting of all functions of the form $p(x)\, e^x$, where $p(x)$ is an arbitrary polynomial.

7.2 Linear Transformations

Consider a linear function $L\colon \mathbb{R}^n \to \mathbb{R}^n$ that maps n-dimensional Euclidean space to itself. The function L maps a point $\mathbf{x} \in \mathbb{R}^n$ to its image point $L[\mathbf{x}] = A\mathbf{x}$, where A is its $n \times n$ matrix representative. As such, it can be assigned a geometrical interpretation that leads to further insight into the nature and scope of linear functions on Euclidean space. The geometrically inspired term *linear transformation* is often used to refer to such linear functions. The two-, three-, and four-dimensional (viewing time as the fourth dimension of space-time) cases have particular relevance to our physical universe. Many of the notable maps that appear in geometry, computer graphics, elasticity, symmetry, crystallography, and Einstein's special relativity, to name a few, are defined by linear transformations.

Most of the important classes of linear transformations already appear in the two-dimensional case. Every linear function $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ has the form

$$L\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a x + b y \\ c x + d y \end{pmatrix}, \qquad \text{where} \qquad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{7.20}$$

is an arbitrary 2×2 matrix. We have already encountered the *rotation matrices*

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{7.21}$$

whose effect is to rotate every vector in $\mathbb{R}^2$ through an angle θ; in Figure 7.4 we illustrate the effect on a couple of square regions in the plane. Planar rotations coincide with 2×2

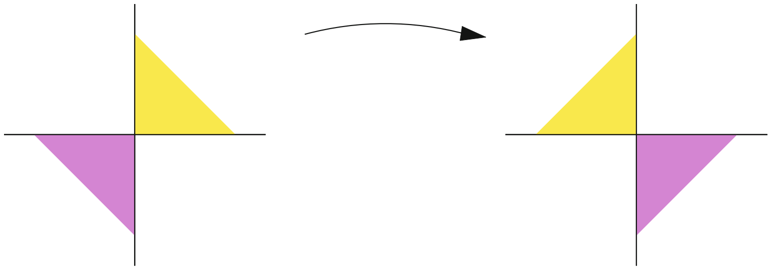


Figure 7.5. Reflection through the y-axis.

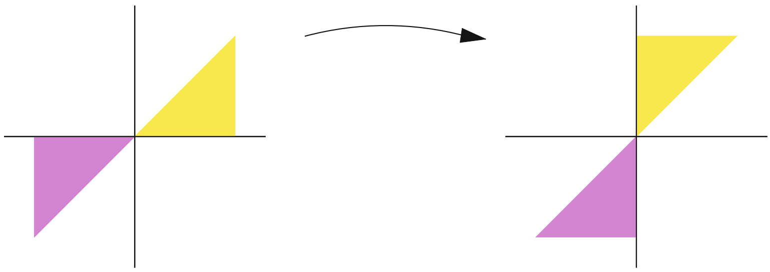


Figure 7.6. Reflection through the Diagonal.

proper orthogonal matrices, meaning matrices Q that satisfy

$$Q^T Q = \mathrm{I}, \qquad \det Q = +1. \tag{7.22}$$

The improper orthogonal matrices, i.e., those with determinant -1, define *reflections*. For example, the matrix

$$A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{corresponds to the linear transformation} \quad L\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix}, \tag{7.23}$$

which reflects the plane through the y-axis. It can be visualized by thinking of the y-axis as a mirror, as illustrated in Figure 7.5. Another simple example is the improper orthogonal matrix

$$R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad \text{The corresponding linear transformation} \quad L\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix} \tag{7.24}$$

is a reflection through the diagonal line $y = x$, as illustrated in Figure 7.6.

A similar classification of orthogonal matrices carries over to three-dimensional (and even higher-dimensional) space. The proper orthogonal matrices correspond to rotations and the improper orthogonal matrices to reflections, or, more generally, reflections combined with rotations. For example, the proper orthogonal matrix

$$Z_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7.25}$$

corresponds to a counterclockwise rotation through an angle θ around the z-axis, while

$$Y_\varphi = \begin{pmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{pmatrix} \tag{7.26}$$

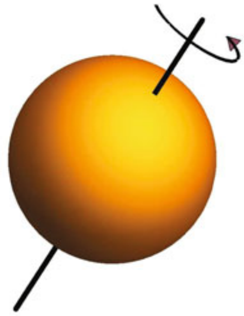


Figure 7.7. A Three–Dimensional Rotation.

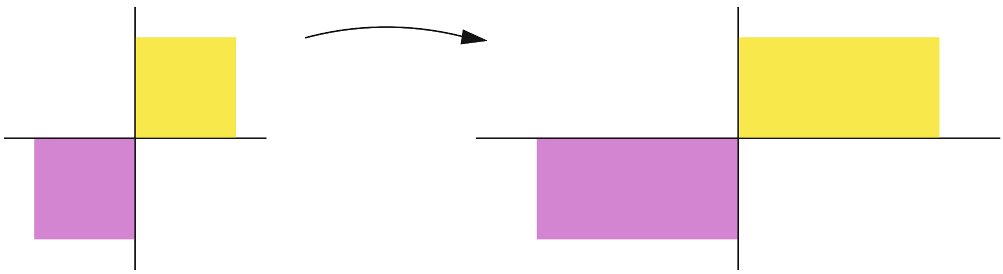


Figure 7.8. Stretch along the x-axis.

corresponds to a clockwise rotation through an angle φ around the y-axis. In general, a proper orthogonal matrix $Q = (\,\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3\,)$ with columns $\mathbf{u}_i = Q\,\mathbf{e}_i$ corresponds to the rotation in which the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are rotated to new positions given by the orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$. It can be shown — see Exercise 8.2.44 — that *every* 3×3 orthogonal matrix corresponds to a rotation around a line through the origin in $\mathbb{R}^3$ — the axis of the rotation, as sketched in Figure 7.7.

Since the product of two (proper) orthogonal matrices is also (proper) orthogonal, the composition of two rotations is also a rotation. Unlike the planar case, the order in which the rotations are performed is important! Multiplication of $n \times n$ orthogonal matrices is *not* commutative when $n \geq 3$. For example, rotating first around the z-axis and then rotating around the y-axis does *not* have the same effect as first rotating around the y-axis and then around the z-axis. If you don't believe this, try it out with a solid object such as this book. Rotate through 90°, say, around each axis; the final configuration of the book will depend upon the order in which you do the rotations. Then prove this mathematically by showing that the two rotation matrices (7.25, 26) do not commute.

Other important linear transformations arise from elementary matrices. First, the elementary matrices corresponding to the third type of row operations — multiplying a row by a scalar — correspond to simple stretching transformations. For example, if

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \text{then the linear transformation} \qquad L\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2\,x \\ y \end{pmatrix}$$

has the effect of stretching along the x-axis by a factor of 2; see Figure 7.8. A negative diagonal entry corresponds to a reflection followed by a stretch. For example, the elementary

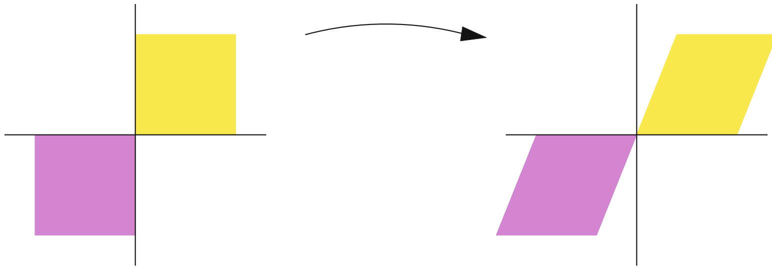


Figure 7.9. Shear in the x Direction.

matrix

$$\begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

corresponds to a reflection through the y-axis followed by a stretch along the x-axis. In this case, the order of these operations is immaterial, since the matrices commute.

In the 2×2 case, there is only one type of elementary row interchange, namely the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, which corresponds to a reflection through the diagonal $y = x$, as in (7.24).

The elementary matrices of type #1 correspond to *shearing transformations* of the plane. For example, the matrix

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \qquad \text{represents the linear transformation} \qquad L\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + 2y \\ y \end{pmatrix},$$

which has the effect of shearing the plane along the x-axis. The constant 2 will be called the *shear factor*, and can be either positive or negative. Under the shearing transformation, each point moves parallel to the x-axis by an amount proportional to its (signed) distance from the axis. Similarly, the elementary matrix

$$\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix} \qquad \text{represents the linear transformation} \qquad L\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y - 3x \end{pmatrix},$$

which is a shear along the y-axis of magnitude -3. As illustrated in Figure 7.9, shears map rectangles to parallelograms; distances are altered, but areas are unchanged.

All of the preceding linear maps are invertible, and so are represented by nonsingular matrices. Besides the zero map/matrix, which sends every point $\mathbf{x} \in \mathbb{R}^2$ to the origin, the simplest singular map is

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \text{corresponding to the linear transformation} \qquad L\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix},$$

which defines the orthogonal projection of the vector $(x, y)^T$ onto the x-axis. Other rank one matrices represent various kinds of projections from the plane to a line through the origin; see Exercise 7.2.16 for details.

A similar classification of linear maps can be established in higher dimensions. The linear transformations constructed from elementary matrices can be built up from the following four basic types:

(*i*) a stretch in a single coordinate direction;
(*ii*) a reflection through a coordinate plane;†
(*iii*) a reflection through a diagonal plane;
(*iv*) a shear along a coordinate axis.

Moreover, we already proved — see (1.47) — that every nonsingular matrix can be written as a product of elementary matrices. This has the remarkable consequence that *every* invertible linear transformation can be constructed from a sequence of elementary stretches, reflections, and shears. In addition, there is one further, non-invertible, type of basic linear transformation:

(*v*) an orthogonal projection onto a lower-dimensional subspace.

All linear transformations of $\mathbb{R}^n$ can be built up, albeit non-uniquely, as a composition of these five basic types.

Example 7.17. Consider the matrix $A = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$ corresponding to a plane rotation through $\theta = 30^\circ$, cf. (7.21). Rotations are not elementary linear transformations. To express this particular rotation as a product of elementary matrices, we need to perform the Gauss-Jordan Elimination procedure to reduce it to the identity matrix. Let us indicate the basic steps:

$$E_1 = \begin{pmatrix} 1 & 0 \\ -\frac{1}{\sqrt{3}} & 1 \end{pmatrix}, \qquad E_1 A = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & \frac{2}{\sqrt{3}} \end{pmatrix},$$

$$E_2 = \begin{pmatrix} 1 & 0 \\ 0 & \frac{\sqrt{3}}{2} \end{pmatrix}, \qquad E_2 E_1 A = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & 1 \end{pmatrix},$$

$$E_3 = \begin{pmatrix} \frac{2}{\sqrt{3}} & 0 \\ 0 & 1 \end{pmatrix}, \qquad E_3 E_2 E_1 A = \begin{pmatrix} 1 & -\frac{1}{\sqrt{3}} \\ 0 & 1 \end{pmatrix},$$

$$E_4 = \begin{pmatrix} 1 & \frac{1}{\sqrt{3}} \\ 0 & 1 \end{pmatrix}, \qquad E_4 E_3 E_2 E_1 A = \mathrm{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We conclude that

$$\begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} = A = E_1^{-1} E_2^{-1} E_3^{-1} E_4^{-1} = \begin{pmatrix} 1 & 0 \\ \frac{1}{\sqrt{3}} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{2}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{2} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{\sqrt{3}} \\ 0 & 1 \end{pmatrix}.$$

As a result, a 30° rotation can be effected by composing the following elementary transformations in the prescribed order, bearing in mind that the last matrix in the product will act first on the vector $\mathbf{x}$:

(1) First, a shear in the x direction with shear factor $-\frac{1}{\sqrt{3}}$.
(2) Then a stretch (or, rather, a contraction) in the direction of the x-axis by a factor of $\frac{\sqrt{3}}{2}$.
(3) Then a stretch in the y direction by the reciprocal factor $\frac{2}{\sqrt{3}}$.
(4) Finally, a shear in the direction of the y-axis with shear factor $\frac{1}{\sqrt{3}}$.

† In n-dimensional space, this should read "hyperplane", i.e., a subspace of dimension $n - 1$.

The fact that this combination of elementary transformations results in a pure rotation is surprising and non-obvious.

Exercises

7.2.1. For each of the following linear transformations $L: \mathbb{R}^2 \to \mathbb{R}^2$, find a matrix representative, and then describe its effect on (i) the x-axis; (ii) the unit square $S = \{0 \le x, y \le 1\}$; (iii) the unit disk $D = \{x^2 + y^2 \le 1\}$: (a) counterclockwise rotation by 45°; (b) rotation by 180°; (c) reflection in the line $y = 2x$; (d) shear along the y-axis of magnitude 2; (e) shear along the line $x = y$ of magnitude 3; (f) orthogonal projection on the line $y = 2x$.

7.2.2. Let L be the linear transformation represented by the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Show that $L^2 = L \circ L$ is rotation by 180°. Is L itself a rotation or a reflection?

7.2.3. Let L be the linear transformation determined by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Show $L^2 = \mathrm{I}$, and interpret geometrically.

7.2.4. What is the geometric interpretation of the linear transformation with matrix $A = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}$? Use this to explain why $A^2 = \mathrm{I}$.

7.2.5. Describe the image of the line ℓ that goes through the points $\begin{pmatrix} -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ under the linear transformation $\begin{pmatrix} 2 & 3 \\ -1 & 0 \end{pmatrix}$.

7.2.6. Draw the parallelogram spanned by the vectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$. Then draw its image under the linear transformations defined by the following matrices:

(a) $\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$, (b) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$, (d) $\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$,

(e) $\begin{pmatrix} -1 & -2 \\ 2 & 1 \end{pmatrix}$, (f) $\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$, (g) $\begin{pmatrix} 2 & -1 \\ -4 & 2 \end{pmatrix}$.

7.2.7. Find a linear transformation that maps the unit circle $x^2 + y^2 = 1$ to the ellipse $\frac{1}{4}x^2 + \frac{1}{9}y^2 = 1$. Is your answer unique?

7.2.8. Find a linear transformation that maps the unit sphere $x^2 + y^2 + z^2 = 1$ to the ellipsoid $x^2 + \frac{1}{4}y^2 + \frac{1}{16}z^2 = 1$.

7.2.9. *True or false*: A linear transformation $L: \mathbb{R}^2 \to \mathbb{R}^2$ maps (a) straight lines to straight lines; (b) triangles to triangles; (c) squares to squares; (d) circles to circles; (e) ellipses to ellipses.

♢ 7.2.10. (a) Prove that the linear transformation associated with the improper orthogonal matrix $\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$ is a reflection through the line that makes an angle $\frac{1}{2}\theta$ with the x-axis.
(b) Show that the composition of two such reflections, with angles θ, φ, is a rotation. What is the angle of the rotation? Does the composition depend upon the order of the two reflections?

7.2.11. (a) Find the matrix in $\mathbb{R}^3$ that corresponds to a counterclockwise rotation around the x-axis through an angle 60°. (b) Write it as a product of elementary matrices, and interpret each of the factors.

♢ 7.2.12. Let $L \subset \mathbb{R}^2$ be the line through the origin in the direction of a unit vector $\mathbf{u}$. (a) Prove that the matrix representative of reflection through L is $R = 2\mathbf{u}\mathbf{u}^T - \mathrm{I}$. (b) Find the corresponding formula for reflection through the line in the direction of a general nonzero vector $\mathbf{v} \neq \mathbf{0}$. (c) Determine the matrix representative for reflection through the line in the direction (i) $(1,0)^T$, (ii) $\left(\frac{3}{5}, -\frac{4}{5}\right)^T$, (iii) $(1,1)^T$, (iv) $(2,-3)^T$.

7.2.13. Decompose the following matrices into a product of elementary matrices. Then interpret each of the factors as a linear transformation.
(a) $\begin{pmatrix} 0 & 2 \\ -3 & 1 \end{pmatrix}$, (b) $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, (c) $\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$, (d) $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, (e) $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 2 & 1 & 1 \end{pmatrix}$.

7.2.14. (a) Prove that $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$, where $a = -\tan\frac{1}{2}\theta$ and $b = \sin\theta$. (b) Is the factorization valid for all values of θ? (c) Interpret the factorization geometrically. **Remark.** The factored version is less prone to numerical errors due to round-off, and so can be used when extremely accurate numerical computations involving rotations are required.

7.2.15. Determine the matrix representative for orthogonal projection $P\colon \mathbb{R}^2 \to \mathbb{R}^2$ on the line through the origin in the direction (a) $(1,0)^T$, (b) $(1,1)^T$, (c) $(2,-3)^T$.

♢ 7.2.16. (a) Prove that every 2×2 matrix of rank 1 can be written in the form $A = \mathbf{u}\mathbf{v}^T$ where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ are non-zero column vectors. (b) Which rank one matrices correspond to orthogonal projection onto a one-dimensional subspace of $\mathbb{R}^2$?

7.2.17. Give a geometrical interpretation of the linear transformations on $\mathbb{R}^3$ defined by each of the six 3×3 permutation matrices (1.30).

7.2.18. Write down the 3×3 matrix X_ψ representing a clockwise rotation in $\mathbb{R}^3$ around the x-axis by angle ψ.

7.2.19. Explain why the linear map defined by $-\mathrm{I}$ defines a rotation in two-dimensional space, but a reflection in three-dimensional space.

♢ 7.2.20. Let $\mathbf{u} = (u_1, u_2, u_3)^T \in \mathbb{R}^3$ be a unit vector. Show that $Q_\pi = 2\mathbf{u}\mathbf{u}^T - \mathrm{I}$ represents rotation around the axis $\mathbf{u}$ through an angle π.

♢ 7.2.21. Let $\mathbf{u} \in \mathbb{R}^3$ be a unit vector. (a) Explain why the *elementary reflection matrix* $R = \mathrm{I} - 2\mathbf{u}\mathbf{u}^T$ represents a reflection through the plane orthogonal to $\mathbf{u}$. (b) Prove that R is an orthogonal matrix. Is it proper or improper? (c) Write out R when $\mathbf{u} =$ (i) $\left(\frac{3}{5}, 0, -\frac{4}{5}\right)^T$, (ii) $\left(\frac{3}{13}, \frac{4}{13}, -\frac{12}{13}\right)^T$, (iii) $\left(\frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)^T$. (d) Give a geometrical explanation why $Q_\pi = -R$ represents the rotation of Exercise 7.2.20.

♢ 7.2.22. Let $\mathbf{a} \in \mathbb{R}^3$, and let Q be any 3×3 rotation matrix such that $Q\mathbf{a} = \mathbf{e}_3$. (a) Show, using the notation of (7.25), that $R_\theta = Q^T Z_\theta Q$ represents rotation around $\mathbf{a}$ by angle θ. (b) Verify this formula in the case $\mathbf{a} = \mathbf{e}_2$ by comparing with (7.26).

♡ 7.2.23. *Quaternions*: The *skew field* $\mathbb{H}$ of quaternions can be identified with the vector space $\mathbb{R}^4$ equipped with a *noncommutative* multiplication operation. The standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$ are traditionally denoted by the letters $1, \mathrm{i}, \mathrm{j}, \mathrm{k}$; the vector $(a,b,c,d)^T \in \mathbb{R}^4$ corresponds to the quaternion $q = a + b\,\mathrm{i} + c\,\mathrm{j} + d\,\mathrm{k}$. Quaternion addition coincides with vector addition. Quaternion multiplication is defined so that

$$1\,q = q = q\,1, \quad \mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1, \quad \mathrm{i}\,\mathrm{j} = \mathrm{k} = -\mathrm{j}\,\mathrm{i}, \quad \mathrm{i}\,\mathrm{k} = -\mathrm{j} = -\mathrm{k}\,\mathrm{i}, \quad \mathrm{j}\,\mathrm{k} = \mathrm{i} = -\mathrm{k}\,\mathrm{j},$$

along with the distributive laws

$$(q+r)s = qs + rs, \qquad q(r+s) = qr + qs, \qquad \text{for all} \qquad q, r, s \in \mathbb{H}.$$

(a) Compute the following quaternion products: (*i*) $\mathrm{j}(2-3\mathrm{j}+\mathrm{k})$, (*ii*) $(1+\mathrm{i})(1-2\mathrm{i}+\mathrm{j})$, (*iii*) $(1+\mathrm{i}-\mathrm{j}-3\mathrm{k})^2$, (*iv*) $(2+2\mathrm{i}+3\mathrm{j}-\mathrm{k})(2-2\mathrm{i}-3\mathrm{j}+\mathrm{k})$. (*b*) The *conjugate* of the quaternion $q=a+b\mathrm{i}+c\mathrm{j}+d\mathrm{k}$ is defined to be $\overline{q}=a-b\mathrm{i}-c\mathrm{j}-d\mathrm{k}$. Prove that $q\,\overline{q}=\|q\|^2=\overline{q}\,q$, where $\|\cdot\|$ is the usual Euclidean norm on $\mathbb{R}^4$. (*c*) Prove that quaternion multiplication is associative. (*d*) Let $q=a+b\mathrm{i}+c\mathrm{j}+d\mathrm{k}\in\mathbb{H}$. Show that $L_q[r]=q\,r$ and $R_q[r]=r\,q$ define linear transformations on the vector space $\mathbb{H}\simeq\mathbb{R}^4$. Write down their 4×4 matrix representatives, and observe that they are not the same, since quaternion multiplication is not commutative. (*e*) Show that L_q and R_q are orthogonal matrices if $\|q\|^2=a^2+b^2+c^2+d^2=1$. (*f*) We can identify a quaternion $q=b\mathrm{i}+c\mathrm{j}+d\mathrm{k}$ with zero real part, $a=0$, with a vector $\mathbf{q}=(b,c,d)^T\in\mathbb{R}^3$. Show that, in this case, the quaternion product $q\,r=\mathbf{q}\times\mathbf{r}-\mathbf{q}\cdot\mathbf{r}$ can be identified with the difference between the cross and dot product of the two vectors. Which vector identities result from the associativity of quaternion multiplication? **Remark.** The *quaternions* were discovered by the Irish mathematician William Rowan Hamilton in 1843. Much of our modern vector calculus notation is of quaternionic origin, [**17**].

Change of Basis

Sometimes a linear transformation represents an elementary geometrical transformation, but this is not evident because the matrix happens to be written in the "wrong" coordinates. The characterization of linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ as multiplication by $m\times n$ matrices in Theorem 7.5 relies on using the standard bases for both the domain and codomain. In many cases, these bases are not particularly well adapted to the linear transformation in question, and one can often gain additional insight by adopting more suitable bases. To this end, we first need to understand how to rewrite a linear transformation in terms of a new basis.

The following result says that, in *any* basis, a linear function on finite-dimensional vector spaces can always be realized by matrix multiplication of the coordinates. But bear in mind that the particular matrix representative will depend upon the choice of bases.

Theorem 7.18. Let $L\colon V\to W$ be a linear function. Suppose V has basis $\mathbf{v}_1,\ldots,\mathbf{v}_n$ and W has basis $\mathbf{w}_1,\ldots,\mathbf{w}_m$. We can write

$$\mathbf{v}=x_1\mathbf{v}_1+\cdots+x_n\mathbf{v}_n\in V,\qquad \mathbf{w}=y_1\mathbf{w}_1+\cdots+y_m\mathbf{w}_m\in W,$$

where $\mathbf{x}=(x_1,x_2,\ldots,x_n)^T$ are the coordinates of $\mathbf{v}$ relative to the basis of V and $\mathbf{y}=(y_1,y_2,\ldots,y_m)^T$ are those of $\mathbf{w}$ relative to the basis of W. Then, in these coordinates, the linear function $\mathbf{w}=L[\mathbf{v}]$ is given by multiplication by an $m\times n$ matrix B, so $\mathbf{y}=B\,\mathbf{x}$.

Proof: We mimic the proof of Theorem 7.5, replacing the standard basis vectors by more general basis vectors. In other words, we will apply L to the basis vectors of V and express the result as a linear combination of the basis vectors in W. Specifically, we write

$$L[\mathbf{v}_j]=\sum_{i=1}^{m} b_{ij}\,\mathbf{w}_i.$$

The coefficients b_{ij} form the entries of the desired coefficient matrix. Indeed, by linearity,

$$L[\mathbf{v}]=L[x_1\mathbf{v}_1+\cdots+x_n\mathbf{v}_n]=x_1L[\mathbf{v}_1]+\cdots+x_nL[\mathbf{v}_n]=\sum_{i=1}^{m}\left(\sum_{j=1}^{n} b_{ij}\,x_j\right)\mathbf{w}_i,$$

and so $y_i = \sum_{j=1}^{n} b_{ij}\, x_j$, as claimed. *Q.E.D.*

Suppose that the linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is represented by a certain $m \times n$ matrix A relative to the standard bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\widehat{\mathbf{e}}_1, \ldots, \widehat{\mathbf{e}}_m$ of the domain and codomain. If we introduce alternative bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, then the *same* linear transformation may have a completely different matrix representation. Therefore, different matrices may represent the same underlying linear transformation, but relative to different bases of its domain and codomain.

Example 7.19. Consider the linear transformation

$$L[\mathbf{x}] = L\left[\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right] = \begin{pmatrix} x_1 - x_2 \\ 2\,x_1 + 4\,x_2 \end{pmatrix}, \tag{7.27}$$

which we write in the standard, Cartesian coordinates on $\mathbb{R}^2$. The corresponding coefficient matrix

$$A = \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \tag{7.28}$$

is the matrix representation of L, relative to the standard basis $\mathbf{e}_1, \mathbf{e}_2$ of $\mathbb{R}^2$, and can be read directly off the explicit formula (7.27):

$$L[\mathbf{e}_1] = L\left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right] = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \mathbf{e}_1 + 2\,\mathbf{e}_2, \qquad L[\mathbf{e}_2] = L\left[\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right] = \begin{pmatrix} -1 \\ 4 \end{pmatrix} = -\,\mathbf{e}_1 + 4\,\mathbf{e}_2.$$

Let us see what happens if we replace the standard basis by the alternative basis

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

What is the corresponding matrix formulation of the same linear transformation? According to the recipe of Theorem 7.18, we must compute

$$L[\mathbf{v}_1] = \begin{pmatrix} 2 \\ -2 \end{pmatrix} = 2\,\mathbf{v}_1, \qquad L[\mathbf{v}_2] = \begin{pmatrix} 3 \\ -6 \end{pmatrix} = 3\,\mathbf{v}_2.$$

The linear transformation acts by stretching in the direction $\mathbf{v}_1$ by a factor of 2 and simultaneously stretching in the direction $\mathbf{v}_2$ by a factor of 3. Therefore, the matrix form of L with respect to this new basis is the diagonal matrix

$$D = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}. \tag{7.29}$$

In general,

$$L[a\,\mathbf{v}_1 + b\,\mathbf{v}_2] = 2\,a\,\mathbf{v}_1 + 3\,b\,\mathbf{v}_2,$$

and the effect is to multiply the new basis coordinates $\mathbf{a} = (\,a, b\,)^T$ by the diagonal matrix D. Both (7.28) and (7.29) represent the *same* linear transformation — the former in the standard basis and the latter in the new basis. The hidden geometry of this linear transformation is thereby exposed through an inspired choice of basis. The secret behind such well-adapted bases will be revealed in Chapter 8.

How does one effect a change of basis in general? According to formula (2.23), if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis of $\mathbb{R}^n$, then the coordinates $\mathbf{y} = (\,y_1, y_2, \ldots, y_n\,)^T$ of a vector

$$(\,x_1, x_2, \ldots, x_n\,)^T = \mathbf{x} = y_1\,\mathbf{v}_1 + y_2\,\mathbf{v}_2 + \;\cdots\; + y_n\,\mathbf{v}_n$$

are found by solving the linear system

$$S\,\mathbf{y} = \mathbf{x}, \qquad \text{where} \qquad S = (\,\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_n\,) \tag{7.30}$$

is the nonsingular $n \times n$ matrix whose columns are the basis vectors.

Consider first a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^n$ from $\mathbb{R}^n$ to itself. When written in terms of the standard basis, $L[\mathbf{x}] = A\,\mathbf{x}$ has a certain $n \times n$ coefficient matrix A. To change to the new basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$, we use (7.30) to rewrite the standard $\mathbf{x}$ coordinates in terms of the new $\mathbf{y}$ coordinates. We also need to find the coordinates $\mathbf{g}$ of an image vector $\mathbf{f} = A\,\mathbf{x}$ with respect to the new basis. By the same reasoning that led to (7.30), its new coordinates are found by solving the linear system $\mathbf{f} = S\,\mathbf{g}$. Therefore, the new codomain coordinates are expressed in terms of the new domain coordinates via

$$\mathbf{g} = S^{-1}\mathbf{f} = S^{-1}A\,\mathbf{x} = S^{-1}A\,S\,\mathbf{y} = B\,\mathbf{y}.$$

We conclude that, in the new basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$, the matrix form of our linear transformation is

$$B = S^{-1}A\,S, \qquad \text{where} \qquad S = (\,\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_n\,). \tag{7.31}$$

Two matrices A and B that are related by such an equation for some nonsingular matrix S are called *similar*. Similar matrices represent the *same* linear transformation, but relative to *different* bases of the underlying vector space $\mathbb{R}^n$, the matrix S serving to encode the *change of basis*.

Example 7.19 (continued). Returning to the preceding example, we assemble the new basis vectors to form the change of basis matrix $S = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix}$, and verify that

$$S^{-1}A\,S = \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} = D,$$

reconfirming our earlier computation.

More generally, a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is represented by an $m \times n$ matrix A with respect to the standard bases on both the domain and codomain. What happens if we introduce a new basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ on the domain space $\mathbb{R}^n$ *and* a new basis $\mathbf{w}_1, \ldots, \mathbf{w}_m$ on the codomain $\mathbb{R}^m$? Arguing as above, we conclude that the matrix representative of L with respect to these new bases is given by

$$B = T^{-1}A\,S, \tag{7.32}$$

where $S = (\,\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_n\,)$ is the domain basis matrix, while $T = (\,\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_m\,)$ is the image basis matrix.

In particular, suppose that L has rank $r = \dim \operatorname{img} A = \dim \operatorname{coimg} A$. Let us choose a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbb{R}^n$ such that $\mathbf{v}_1, \ldots, \mathbf{v}_r$ form a basis of $\operatorname{coimg} A$, while $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ form a basis for $\ker A = (\operatorname{coimg} A)^\perp$. According to Theorem 4.49, the image vectors $\mathbf{w}_1 = L[\mathbf{v}_1], \ldots, \mathbf{w}_r = L[\mathbf{v}_r]$ form a basis for $\operatorname{img} A$, while $L[\mathbf{v}_{r+1}] = \cdots = L[\mathbf{v}_n] = \mathbf{0}$. We further choose a basis $\mathbf{w}_{r+1}, \ldots, \mathbf{w}_m$ for $\operatorname{coker} A = (\operatorname{img} A)^\perp$, and note that the combination $\mathbf{w}_1, \ldots, \mathbf{w}_m$ is a basis for $\mathbb{R}^m$. The matrix form of L relative to these two adapted bases is simply

$$B = T^{-1} A\, S = \begin{pmatrix} \mathrm{I}_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{7.33}$$

In this matrix, the first r columns have a single 1 in the diagonal slot, indicating that the first r basis vectors of the domain space are mapped to the first r basis vectors of the codomain, while the last $n - r$ columns are all zero, indicating that the last $n - r$ basis vectors in the domain are all mapped to $\mathbf{0}$. Thus, by a suitable choice of bases on both the domain and codomain, every linear transformation has an extremely simple *canonical form* (7.33) that depends *only* on its rank.

Example 7.20. According to the example following Theorem 2.49, the matrix

$$A = \begin{pmatrix} 2 & -1 & 1 & 2 \\ -8 & 4 & -6 & -4 \\ 4 & -2 & 3 & 2 \end{pmatrix}$$

has rank 2. Based on those calculations, we choose the domain space basis

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ -1 \\ 1 \\ 2 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 0 \\ -2 \\ 4 \end{pmatrix}, \qquad \mathbf{v}_3 = \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{v}_4 = \begin{pmatrix} -2 \\ 0 \\ 2 \\ 1 \end{pmatrix},$$

noting that $\mathbf{v}_1, \mathbf{v}_2$ are a basis for $\operatorname{coimg} A$, while $\mathbf{v}_3, \mathbf{v}_4$ are a basis for $\ker A$. For our basis of the codomain, we first compute $\mathbf{w}_1 = A\mathbf{v}_1$ and $\mathbf{w}_2 = A\mathbf{v}_2$, which form a basis for $\operatorname{img} A$. We supplement these by the single basis vector $\mathbf{w}_3$ for $\operatorname{coker} A$, where

$$\mathbf{w}_1 = \begin{pmatrix} 10 \\ -34 \\ 17 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} 6 \\ -4 \\ 2 \end{pmatrix}, \qquad \mathbf{w}_3 = \begin{pmatrix} 0 \\ \frac{1}{2} \\ 1 \end{pmatrix}.$$

By construction, $B[\mathbf{v}_1] = \mathbf{w}_1$, $B[\mathbf{v}_2] = \mathbf{w}_2$, $B[\mathbf{v}_3] = B[\mathbf{v}_4] = \mathbf{0}$, and thus the canonical matrix form of this particular linear function is given in terms of these two bases as

$$B = T^{-1} A\, S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where the bases are assembled to form the matrices

$$S = \begin{pmatrix} 2 & 0 & \frac{1}{2} & -2 \\ -1 & 0 & 1 & 0 \\ 1 & -2 & 0 & 2 \\ 2 & 4 & 0 & 1 \end{pmatrix}, \qquad T = \begin{pmatrix} 10 & 6 & 0 \\ -34 & -4 & \frac{1}{2} \\ 17 & 2 & 1 \end{pmatrix}.$$

Exercises

7.2.24. Find the matrix form of the linear transformation $L(x,y) = \begin{pmatrix} x-4y \\ -2x+3y \end{pmatrix}$ with respect to the following bases of $\mathbb{R}^2$:

(a) $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, (b) $\begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix}$, (c) $\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, (d) $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, (e) $\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

7.2.25. Find the matrix form of $L[\mathbf{x}] = \begin{pmatrix} -3 & 2 & 2 \\ -3 & 1 & 3 \\ -1 & 2 & 0 \end{pmatrix} \mathbf{x}$ with respect to the following bases of $\mathbb{R}^3$:

(a) $\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -2 \end{pmatrix}$, (b) $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, (c) $\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$.

7.2.26. Find bases of the domain and codomain that place the following matrices in the canonical form (7.33). Use (7.32) to check your answer. (a) $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$,

(b) $\begin{pmatrix} 1 & -3 & 4 \\ -2 & 6 & -8 \end{pmatrix}$, (c) $\begin{pmatrix} 2 & 3 \\ 0 & 4 \\ -1 & 1 \end{pmatrix}$, (d) $\begin{pmatrix} 1 & 2 & 1 \\ 1 & -1 & -1 \\ 2 & 1 & 0 \end{pmatrix}$, (e) $\begin{pmatrix} 1 & 3 & 0 & 1 \\ 2 & 6 & 1 & -2 \\ -1 & -3 & -1 & 3 \\ 0 & 0 & -1 & 4 \end{pmatrix}$.

7.2.27. (a) Show that every invertible linear function $L\colon \mathbb{R}^n \to \mathbb{R}^n$ can be represented by the identity matrix by choosing appropriate (and not necessarily the same) bases on the domain and codomain. (b) Which linear transformations are represented by the identity matrix when the domain and codomain are required to have the same basis? (c) Find bases of $\mathbb{R}^2$ so that the following linear transformations are represented by the identity matrix: (i) the scaling map $S[\mathbf{x}] = 2\mathbf{x}$; (ii) counterclockwise rotation by 45°; (iii) the shear $\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$.

♢ 7.2.28. Suppose a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^n$ is represented by a symmetric matrix with respect to the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. (a) Prove that its matrix representative with respect to any orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is symmetric. (b) Is it symmetric when expressed in terms of a non-orthonormal basis?

♢ 7.2.29. In this exercise, we show that every inner product $\langle \cdot\,,\cdot \rangle$ on $\mathbb{R}^n$ can be reduced to the dot product when expressed in a suitably adapted basis. (a) Specifically, prove that there exists a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbb{R}^n$ such that $\langle \mathbf{x}\,,\mathbf{y} \rangle = \sum_{i=1}^{n} c_i\, d_i = \mathbf{c} \cdot \mathbf{d}$, where $\mathbf{c} = (c_1, c_2, \ldots, c_n)^T$ are the coordinates of $\mathbf{x}$ and $\mathbf{d} = (d_1, d_2, \ldots, d_n)^T$ those of $\mathbf{y}$ with respect to the basis. Is the basis uniquely determined? (b) Find bases that reduce the following inner products to the dot product on $\mathbb{R}^2$:

(i) $\langle \mathbf{v}\,,\mathbf{w} \rangle = 2v_1w_1 + 3v_2w_2$, (ii) $\langle \mathbf{v}\,,\mathbf{w} \rangle = v_1w_1 - v_1w_2 - v_2w_1 + 3v_2w_2$.

♡ 7.2.30. *Dual functions*: Let $L\colon V \to W$ be a linear function between vector spaces. The *dual linear function*, denoted by $L^*\colon W^* \to V^*$ (note the change in direction) is defined so that $L^*(m) = m \circ L$ for all linear functions $m \in W^*$. (a) Prove that L^* is a linear function. (b) If $M\colon W \to Z$ is linear, prove that $(M \circ L)^* = L^* \circ M^*$. (c) Suppose $\dim V = n$ and $\dim W = m$. Prove that if L is represented by the $m \times n$ matrix A with respect to bases of V, W, then L^* is represented by the $n \times m$ transposed matrix A^T with respect to the dual bases, as defined in Exercise 7.1.32.

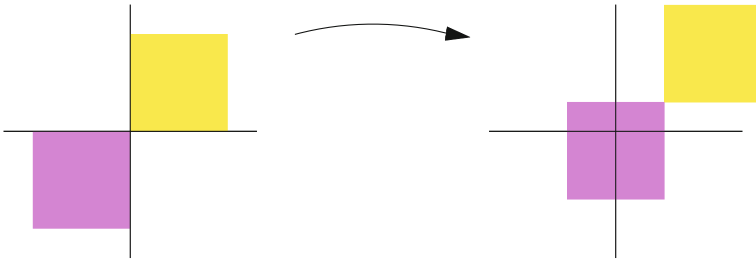


Figure 7.10. Translation.

◇ 7.2.31. Suppose A is an $m\times n$ matrix. (a) Let $\mathbf{v}_1,\ldots,\mathbf{v}_n$ be a basis of $\mathbb{R}^n$, and $A\mathbf{v}_i=\mathbf{w}_i\in\mathbb{R}^m$, for $i=1,\ldots,n$. Prove that the vectors $\mathbf{v}_1,\ldots,\mathbf{v}_n$, $\mathbf{w}_1,\ldots,\mathbf{w}_n$, serve to uniquely specify A. (b) Write down a formula for A.

7.3 Affine Transformations and Isometries

Not every transformation of importance in geometrical applications arises as a linear function. A simple example is a *translation*, whereby all the points in $\mathbb{R}^n$ are moved in the same direction by a common distance. The function that accomplishes this is

$$T[\mathbf{x}]=\mathbf{x}+\mathbf{b}, \qquad \mathbf{x}\in\mathbb{R}^n, \tag{7.34}$$

where $\mathbf{b}\in\mathbb{R}^n$ determines the direction and the distance that the points are translated. Except in the trivial case $\mathbf{b}=\mathbf{0}$, the translation T is *not* a linear function because

$$T[\mathbf{x}+\mathbf{y}]=\mathbf{x}+\mathbf{y}+\mathbf{b} \neq T[\mathbf{x}]+T[\mathbf{y}]=\mathbf{x}+\mathbf{y}+2\mathbf{b}.$$

Or, even more simply, we note that $T[\mathbf{0}]=\mathbf{b}$, which must be $\mathbf{0}$ if T is to be linear.

Combining translations and linear functions leads us to an important class of geometrical transformations.

Definition 7.21. A function $F:\mathbb{R}^n\to\mathbb{R}^n$ of the form

$$F[\mathbf{x}]=A\,\mathbf{x}+\mathbf{b}, \tag{7.35}$$

where A is an $n\times n$ matrix and $\mathbf{b}\in\mathbb{R}^n$, is called an *affine transformation*.

In general, $F[\mathbf{x}]$ is an affine transformation if and only if $L[\mathbf{x}]=F[\mathbf{x}]-F[\mathbf{0}]$ is a linear function. In the particular case (7.35), $F[\mathbf{0}]=\mathbf{b}$, and so $L[\mathbf{x}]=A\,\mathbf{x}$. The word "affine" comes from the Latin "affinus", meaning "related", because such transformations preserve the relation of parallelism between lines; see Exercise 7.3.2.

For example, every affine transformation from $\mathbb{R}$ to $\mathbb{R}$ has the form

$$f(x)=\alpha\,x+\beta. \tag{7.36}$$

As mentioned earlier, even though the graph of $f(x)$ is a straight line, f is *not* a linear function — unless $\beta=0$, and the line goes through the origin. Thus, to be mathematically accurate, we should refer to (7.36) as a *one-dimensional affine transformation*.

Example 7.22. The affine transformation

$$F(x,y) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} -y+1 \\ x-2 \end{pmatrix}$$

has the effect of first rotating the plane $\mathbb{R}^2$ by 90° about the origin, and then translating by the vector $(1,-2)^T$. The reader may enjoy proving that this combination has the same effect as just rotating the plane through an angle of 90° centered at the point $\left(\frac{3}{4},-\frac{1}{2}\right)$. For details, see Exercise 7.3.14.

Note that the affine transformation (7.35) can be obtained by composing a linear function $L[\mathbf{x}] = A\mathbf{x}$ with a translation $T[\mathbf{x}] = \mathbf{x} + \mathbf{b}$, so

$$F[\mathbf{x}] = T \circ L[\mathbf{x}] = T[L[\mathbf{x}]] = T[A\mathbf{x}] = A\mathbf{x} + \mathbf{b}.$$

The order of composition is important, since $G = L \circ T$ defines the slightly different affine transformation

$$G[\mathbf{x}] = L \circ T[\mathbf{x}] = L[T[\mathbf{x}]] = L[\mathbf{x}+\mathbf{b}] = A(\mathbf{x}+\mathbf{b}) = A\mathbf{x} + \mathbf{c}, \qquad \text{where} \qquad \mathbf{c} = A\mathbf{b}.$$

More generally, the composition of any two affine transformations is again an affine transformation. Specifically, given

$$F[\mathbf{x}] = A\mathbf{x} + \mathbf{a}, \qquad G[\mathbf{y}] = B\mathbf{y} + \mathbf{b},$$

then

$$\begin{aligned}(G \circ F)[\mathbf{x}] = G[F[\mathbf{x}]] = G[A\mathbf{x}+\mathbf{a}] = B(A\mathbf{x}+\mathbf{a}) + \mathbf{b} = C\mathbf{x} + \mathbf{c}, \\ \text{where} \quad C = BA, \quad \mathbf{c} = B\mathbf{a} + \mathbf{b}.\end{aligned} \tag{7.37}$$

Note that the coefficient matrix of the composition is the product of the coefficient matrices, but the resulting vector of translation is *not* the sum of the two translation vectors.

Exercises

7.3.1. *True or false*: An affine transformation takes (*a*) straight lines to straight lines; (*b*) triangles to triangles; (*c*) squares to squares; (*d*) circles to circles; (*e*) ellipses to ellipses.

♢ 7.3.2. (*a*) Let $F: \mathbb{R}^n \to \mathbb{R}^n$ be an affine transformation. Let $L_1, L_2 \subset \mathbb{R}^n$ be two parallel lines. Prove that $F[L_1]$ and $F[L_2]$ are also parallel lines.
(*b*) Is the converse valid: if $F: \mathbb{R}^n \to \mathbb{R}^n$ maps parallel lines to parallel lines, then F is necessarily an affine transformation?

7.3.3. Describe the image of (*i*) the x-axis, (*ii*) the unit disk $x^2+y^2 \le 1$, (*iii*) the unit square $0 \le x, y \le 1$, under the following affine transformations:

(*a*) $T_1\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 2 \\ -1 \end{pmatrix}$, (*b*) $T_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \end{pmatrix}$,

(*c*) $T_3\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, (*d*) $T_4\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$,

(*e*) $T_5\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} .6 & .8 \\ -.8 & .6 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -3 \\ 2 \end{pmatrix}$, (*f*) $T_6\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$,

(*g*) $T_7\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 2 \\ -3 \end{pmatrix}$, (*h*) $T_8\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

7.3.4. Using the affine transformations in Exercise 7.3.3, write out the following compositions and verify that they satisfy (7.37):
(*a*) $T_3 \circ T_4$, (*b*) $T_4 \circ T_3$, (*c*) $T_3 \circ T_6$, (*d*) $T_6 \circ T_3$, (*e*) $T_7 \circ T_8$, (*f*) $T_8 \circ T_7$.

7.3.5. Describe the image of the triangle with vertices $(-1,0),(1,0),(0,2)$ under the affine transformation $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 & -1 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 3 \\ -4 \end{pmatrix}$.

7.3.6. Under what conditions is the composition of two affine transformations
(*a*) a translation? (*b*) a linear function?

7.3.7. (*a*) Under what conditions does an affine transformation have an inverse? (*b*) Is the inverse an affine transformation? If so, find a formula for its matrix and vector constituents. (*c*) Find the inverse, when it exists, of each of the the affine transformations in Exercise 7.3.3.

♢ 7.3.8. Let $\mathbf{v}_1,\ldots,\mathbf{v}_n$ be a basis for $\mathbb{R}^n$. (*a*) Show that every affine transformation $F[\mathbf{x}] = A\mathbf{x} + \mathbf{b}$ on $\mathbb{R}^n$ is uniquely determined by the $n+1$ vectors $\mathbf{w}_0 = F[\mathbf{0}]$, $\mathbf{w}_1 = F[\mathbf{v}_1]$, $\ldots$,
$\mathbf{w}_n = F[\mathbf{v}_n]$. (*b*) Find the formula for A and $\mathbf{b}$ when $\mathbf{v}_1 = \mathbf{e}_1$, $\ldots$, $\mathbf{v}_n = \mathbf{e}_n$ are the standard basis vectors. (*c*) Find the formula for $A, \mathbf{b}$ for a general basis $\mathbf{v}_1,\ldots,\mathbf{v}_n$.

7.3.9. Show that the space of all affine transformations on $\mathbb{R}^n$ is a vector space. What is its dimension?

♢ 7.3.10. In this exercise, we establish a useful matrix representation for affine transformations. We identify $\mathbb{R}^n$ with the n-dimensional affine subspace (as in Exercise 2.2.28)
$$V_n = \{\, (\mathbf{x},1)^T = (x_1,\ldots,x_n,1)^T \,\} \subset \mathbb{R}^{n+1}$$
consisting of vectors whose last coordinate is fixed at $x_{n+1} = 1$. (*a*) Show that multiplication of vectors $\begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} \in V_n$ by the $(n+1)\times(n+1)$ *affine matrix* $\begin{pmatrix} A & \mathbf{b} \\ \mathbf{0} & 1 \end{pmatrix}$ coincides with the action (7.35) of an affine transformation on $\mathbf{x}\in\mathbb{R}^n$. (*b*) Prove that the composition law (7.37) for affine transformations corresponds to multiplication of their affine matrices. (*c*) Define the inverse of an affine transformation in the evident manner, and show that it corresponds to the inverse affine matrix.

Isometry

A transformation that preserves distance is known as an *isometry*. (The mathematical term *metric* refers to the underlying norm or distance on the space; thus, "isometric" translates as "distance-preserving".) In Euclidean geometry, the isometries coincide with the *rigid motions* — translations, rotations, reflections, and the affine maps they generate through composition.

Definition 7.23. Let V be a normed vector space. A function $F: V \to V$ is called an *isometry* if it preserves distance, meaning
$$d(F[\mathbf{v}], F[\mathbf{w}]) = d(\mathbf{v},\mathbf{w}) \qquad \text{for all} \qquad \mathbf{v},\mathbf{w}\in V. \tag{7.38}$$

Since the distance between points is just the norm of the vector connecting them, $d(\mathbf{v},\mathbf{w}) = \|\,\mathbf{v}-\mathbf{w}\,\|$, cf. (3.33), the isometry condition (7.38) can be restated as
$$\big\|\, F[\mathbf{v}] - F[\mathbf{w}]\,\big\| = \|\,\mathbf{v}-\mathbf{w}\,\| \qquad \text{for all} \qquad \mathbf{v},\mathbf{w}\in V. \tag{7.39}$$
Clearly, any translation
$$T[\mathbf{v}] = \mathbf{v} + \mathbf{a}, \qquad \text{where} \qquad \mathbf{a}\in V,$$

defines an isometry, since $T[\mathbf{v}]-T[\mathbf{w}] = \mathbf{v}-\mathbf{w}$. A linear transformation $L: V \to V$ defines an isometry if and only if

$$\| L[\mathbf{v}] \| = \| \mathbf{v} \| \qquad \text{for all} \qquad \mathbf{v} \in V, \tag{7.40}$$

because, by linearity,

$$\| L[\mathbf{v}] - L[\mathbf{w}] \| = \| L[\mathbf{v}-\mathbf{w}] \| = \| \mathbf{v}-\mathbf{w} \|.$$

A similar computation proves that an affine transformation $F[\mathbf{v}] = L[\mathbf{v}]+\mathbf{a}$ is an isometry if and only if its linear part $L[\mathbf{v}]$ is.

As noted above, the simplest class of isometries comprises the translations

$$T[\mathbf{x}] = \mathbf{x} + \mathbf{b} \tag{7.41}$$

in the direction $\mathbf{b}$. For the standard Euclidean norm on $V = \mathbb{R}^n$, the linear isometries consist of rotations and reflections. As we shall prove, both are characterized by orthogonal matrices:

$$L[\mathbf{x}] = Q\,\mathbf{x}, \qquad \text{where} \qquad Q^T Q = \mathrm{I}\,. \tag{7.42}$$

The *proper isometries* correspond to the rotations, with $\det Q = +1$, and can be realized as physical motions; *improper isometries*, with $\det Q = -1$, are then obtained by reflection in a mirror.

Proposition 7.24. A linear transformation $L[\mathbf{x}] = Q\,\mathbf{x}$ defines a Euclidean isometry of $\mathbb{R}^n$ if and only if Q is an orthogonal matrix.

Proof: The linear isometry condition (7.40) requires that

$$\| Q\,\mathbf{x} \|^2 = (Q\,\mathbf{x})^T Q\,\mathbf{x} = \mathbf{x}^T Q^T Q\,\mathbf{x} = \mathbf{x}^T \mathbf{x} = \| \mathbf{x} \|^2 \qquad \text{for all} \qquad \mathbf{x} \in \mathbb{R}^n.$$

According to Exercise 4.3.16, this holds if and only if $Q^T Q = \mathrm{I}$, which is precisely the condition (4.29) that Q be an orthogonal matrix. *Q.E.D.*

It can be proved, [**93**], that the most general Euclidean isometry of $\mathbb{R}^n$ is an affine transformation, and hence of the form $F[\mathbf{x}] = Q\,\mathbf{x} + \mathbf{b}$, where Q is an orthogonal matrix and $\mathbf{b}$ is a vector. Therefore, every Euclidean isometry or rigid motion is a combination of translations, rotations, and reflections.

In the two-dimensional case, the proper linear isometries $R[\mathbf{x}] = Q\,\mathbf{x}$ with $\det Q = 1$ represent rotations around the origin. More generally, a rotation of the plane around a center at $\mathbf{c}$ is represented by the affine isometry

$$R[\mathbf{x}] = Q(\mathbf{x}-\mathbf{c}) + \mathbf{c} = Q\,\mathbf{x} + \mathbf{b}, \qquad \text{where} \qquad \mathbf{b} = (\,\mathrm{I} - Q\,)\,\mathbf{c}, \tag{7.43}$$

and where Q is a rotation matrix. In Exercise 7.3.14, we ask you to prove that every plane isometry is either a translation or a rotation around a center.

In three-dimensional space, both translations (7.41) and rotations around a center (7.43) continue to define proper isometries. There is one additional type, representing the motion of a point on the head of a screw. A *screw motion* is an affine transformation of the form

$$S[\mathbf{x}] = Q\,\mathbf{x} + \mathbf{a}, \tag{7.44}$$

where the 3×3 orthogonal matrix Q represents a rotation through an angle θ around a fixed axis in the direction of the vector $\mathbf{a}$, which is also the direction of the translation

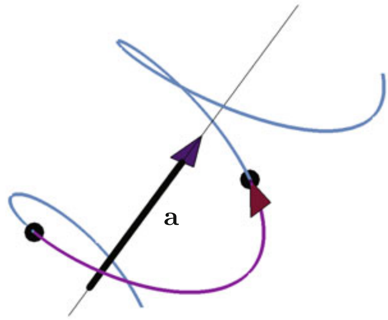


Figure 7.11. A Screw Motion.

term. The result is indicated in Figure 7.11; the trajectory followed by a point not on the axis is a circular helix centered on the axis. For example,

$$S_\theta \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ a \end{pmatrix}$$

represents a vertical screw along the z-axis through an angle θ by a distance a. In Exercise 8.2.45 you are asked to prove that every proper isometry of $\mathbb{R}^3$ is either a translation, a rotation, or a screw motion.

The isometries of $\mathbb{R}^2$ and $\mathbb{R}^3$ are indispensable for understanding of how physical objects move in three-dimensional space. Basic computer graphics and animation require efficient implementation of rigid isometries in three-dimensional space and their compositions — coupled with appropriate (nonlinear) perspective maps prescribing the projection of three-dimensional objects onto a two-dimensional viewing screen, [**12**, **72**].

Exercises

Note: All exercises are based on the Euclidean norm unless otherwise noted.

7.3.11. Which of the indicated maps $\mathbf{F}(x,y)$ define isometries of the Euclidean plane? (a) $\begin{pmatrix} y \\ -x \end{pmatrix}$, (b) $\begin{pmatrix} x-2 \\ y-1 \end{pmatrix}$, (c) $\begin{pmatrix} x-y+1 \\ x+2 \end{pmatrix}$, (d) $\dfrac{1}{\sqrt{2}}\begin{pmatrix} x+y-3 \\ x+y-2 \end{pmatrix}$, (e) $\dfrac{1}{5}\begin{pmatrix} 3x+4y \\ -4x+3y+1 \end{pmatrix}$.

7.3.12. Prove that the planar affine isometry $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y+1 \\ x-2 \end{pmatrix}$ represents a rotation through an angle of 90° around the center $\left(\frac{3}{2}, -\frac{1}{2}\right)^T$.

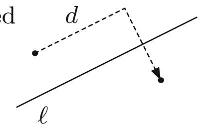
7.3.13. *True or false*: The map $L[\mathbf{x}] = -\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$ defines (*a*) an isometry; (*b*) a rotation.

♢ 7.3.14. Prove that every *proper* affine plane isometry $F[\mathbf{x}] = Q\,\mathbf{x} + \mathbf{b}$ of $\mathbb{R}^2$, where $\det Q = 1$, is either (*i*) a translation, or (*ii*) a rotation (7.43) centered at some point $\mathbf{c} \in \mathbb{R}^2$.
Hint: Use Exercise 1.5.7.

7.3.15. Compute both compositions $F \circ G$ and $G \circ F$ of the following affine transformations on $\mathbb{R}^2$. Which pairs commute? (*a*) $F =$ counterclockwise rotation around the origin by 45°; $G =$ translation in the y direction by 3 units. (*b*) $F =$ counterclockwise rotation around the point $(1,1)^T$ by 30°; $G =$ counterclockwise rotation around the point $(-2,1)^T$ by 90°. (*c*) $F =$ reflection through the line $y = x+1$; $G =$ rotation around $(1,1)^T$ by 180°.

♡ 7.3.16. In $\mathbb{R}^2$, show the following: (*a*) The composition of two affine isometries is another affine isometry. (*b*) The composition of two translations is another translation. (*c*) The composition of a translation and a rotation (not necessarily centered at the origin) in either order is a rotation. (*d*) The composition of two plane rotations is either another rotation or a translation. What is the condition for the latter possibility? (*e*) Every plane translation can be written as the composition of two rotations.

♢ 7.3.17. Let ℓ be a line in $\mathbb{R}^2$. A *glide reflection* is an affine map on $\mathbb{R}^2$ composed of a translation in the direction of ℓ by a distance d followed by a reflection through ℓ. Find the formula for a glide reflection along (*a*) the x-axis by a distance 2; (*b*) the line $y = x$ by a distance 3 in the direction of increasing x; (*c*) the line $x + y = 1$ by a distance 2 in the direction of increasing x.



♢ 7.3.18. Let ℓ be the line in the direction of the unit vector $\mathbf{u}$ through the point $\mathbf{a}$. (*a*) Write down the formula for the affine map defining the reflection through the line ℓ. *Hint*: Use Exercise 7.2.12. (*b*) Write down the formula for the glide reflection, as defined in Exercise 7.3.17, along ℓ by a distance d in the direction of $\mathbf{u}$. (*c*) Prove that every improper affine plane isometry is either a reflection or a glide reflection. *Hint*: Use Exercise 7.2.10.

♡ 7.3.19. A set of $n+1$ points $\mathbf{a}_0, \ldots, \mathbf{a}_n \in \mathbb{R}^n$ is said to be *in general position* if the differences $\mathbf{a}_i - \mathbf{a}_j$ span $\mathbb{R}^n$. (*a*) Show that the points are in general position if and only if they do not all lie in a proper affine subspace $A \subsetneq \mathbb{R}^n$, cf. Exercise 2.2.28. (*b*) Let $\mathbf{a}_0, \ldots, \mathbf{a}_n$ and $\mathbf{b}_0, \ldots, \mathbf{b}_n$ be two sets in general position. Show that there is an isometry $F\colon \mathbb{R}^n \to \mathbb{R}^n$ such that $F[\mathbf{a}_i] = \mathbf{b}_i$ for all $i = 0, \ldots, n$, if and only if their interpoint distances agree: $\|\mathbf{a}_i - \mathbf{a}_j\| = \|\mathbf{b}_i - \mathbf{b}_j\|$ for all $0 \le i < j \le n$. *Hint*: Use Exercise 4.3.19.

♢ 7.3.20. Suppose that V is an inner product space and $L\colon V \to V$ is an isometry, so $\|L[\mathbf{v}]\| = \|\mathbf{v}\|$ for all $\mathbf{v} \in V$. Prove that L also preserves the inner product: $\langle L[\mathbf{v}], L[\mathbf{w}] \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$. *Hint*: Look at $\|L[\mathbf{v}+\mathbf{w}]\|^2$.

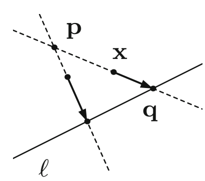
♢ 7.3.21. Let V be a normed vector space. Prove that a linear map $L\colon V \to V$ defines an isometry of V for the given norm if and only if it maps the unit sphere $S_1 = \{\|\mathbf{u}\| = 1\}$ to itself: $L[S_1] = \{L[\mathbf{u}] \mid \mathbf{u} \in S_1\} = S_1$.

7.3.22. (*a*) List all linear and affine isometries of $\mathbb{R}^2$ with respect to the ∞ norm. *Hint*: Use Exercise 7.3.21. (*b*) Can you generalize your results to $\mathbb{R}^3$?

7.3.23. Answer Exercise 7.3.22 for the 1 norm.

♡ 7.3.24. A matrix of the form $H = \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}$ for $\alpha \in \mathbb{R}$ defines a *hyperbolic rotation* of $\mathbb{R}^2$. (*a*) Prove that all hyperbolic rotations preserve the indefinite quadratic form $q(\mathbf{x}) = x^2 - y^2$ in the sense that $q(H\mathbf{x}) = q(\mathbf{x})$ for all $\mathbf{x} = (x, y)^T \in \mathbb{R}^2$. Observe that ordinary rotations preserve circles $x^2 + y^2 = a$, while hyperbolic rotations preserve hyperbolas $x^2 - y^2 = a$. (*b*) Are there any other affine transformations of $\mathbb{R}^2$ that preserve the quadratic form $q(\mathbf{x})$? **Remark.** The four-dimensional version of this construction, i.e., affine maps preserving the indefinite *Minkowski form* $t^2 - x^2 - y^2 - z^2$, forms the geometrical foundation for Einstein's theory of special relativity, [**55**].

♡ 7.3.25. Let $\ell \subset \mathbb{R}^2$ be a line, and $\mathbf{p} \notin \ell$ a point. A *perspective map* takes a point $\mathbf{x} \in \mathbb{R}^2$ to the point $\mathbf{q} \in \ell$ that is the intersection of ℓ with the line going through $\mathbf{p}$ and $\mathbf{x}$. If the line is parallel to ℓ, then the map is not defined. Find the formula for the perspective map when (*a*) ℓ is the x-axis and $\mathbf{p} = (0,1)^T$, (*b*) ℓ is the line $y = x$ and $\mathbf{p} = (1,0)^T$. Is either map affine? An isometry? **Remark.** Mapping three-dimensional objects onto a two-dimensional screen (or your retina) is based on perspective maps, which are thus of fundamental importance in art, optics, computer vision, computer graphics and animation, and computer games.



7.4 Linear Systems

The abstract notion of a linear system serves to unify, in a common conceptual framework, linear systems of algebraic equations, linear differential equations, both ordinary and partial, linear boundary value problems, linear integral equations, linear control systems, and a huge variety of other linear systems that appear in all aspects of mathematics and its applications. The idea is simply to replace matrix multiplication by a general linear function. Many of the structural results we learned in the matrix context have, when suitably formulated, direct counterparts in these more general frameworks. The result is a unified understanding of the basic properties and nature of solutions to all such linear systems.

Definition 7.25. A *linear system* is an equation of the form

$$L[\mathbf{u}] = \mathbf{f}, \tag{7.45}$$

in which $L: U \to V$ is a linear function between vector spaces, the right-hand side is an element of the codomain, $\mathbf{f} \in V$, while the desired solution belongs to the domain, $\mathbf{u} \in U$. The system is *homogeneous* if $\mathbf{f} = \mathbf{0}$; otherwise, it is called *inhomogeneous*.

Example 7.26. If $U = \mathbb{R}^n$ and $V = \mathbb{R}^m$, then, according to Theorem 7.5, every linear function $L: \mathbb{R}^n \to \mathbb{R}^m$ is given by matrix multiplication: $L[\mathbf{u}] = A\mathbf{u}$. Therefore, in this particular case, every linear system is a matrix system, namely $A\mathbf{u} = \mathbf{f}$.

Example 7.27. A *linear ordinary differential equation* takes the form $L[u] = f$, where L is an n^{th} order linear differential operator of the form (7.15), and the right-hand side is, say, a continuous function. Written out, the differential equation takes the familiar form

$$L[u] = a_n(x)\,\frac{d^n u}{dx^n} + a_{n-1}(x)\,\frac{d^{n-1}u}{dx^{n-1}} + \ \cdots\ + a_1(x)\,\frac{du}{dx} + a_0(x)u = f(x). \tag{7.46}$$

You should have already gained some familiarity with solving the constant coefficient case as covered, for instance, in [**7**, **22**].

Example 7.28. Let $K(x,y)$ be a function of two variables that is continuous for all $a \le x, y \le b$. Then the integral

$$I_K[u] = \int_a^b K(x,y)\,u(y)\,dy$$

defines a linear operator $I_K: \mathrm{C}^0[a,b] \to \mathrm{C}^0[a,b]$, known as an *integral transform*. Important examples include the Fourier and Laplace transforms, [**61**, **79**]. Finding the inverse transform requires solving a *linear integral equation* $I_K[u] = f$, which has the explicit form

$$\int_a^b K(x,y)\,u(y)\,dy = f(x).$$

Example 7.29. We can combine linear maps to form more complicated, "mixed" types of linear systems. For example, consider a typical initial value problem

$$u'' + u' - 2\,u = x, \qquad u(0) = 1, \qquad u'(0) = -1, \tag{7.47}$$

for an unknown scalar function $u(x)$. The differential equation can be written as a linear system

$$L[u] = x, \qquad \text{where} \qquad L[u] = (D^2 + D - 2)[u] = u'' + u' - 2\,u$$

is a linear, constant coefficient differential operator. Further,

$$M[u] = \begin{pmatrix} L[u] \\ u(0) \\ u'(0) \end{pmatrix} = \begin{pmatrix} u''(x) + u'(x) - 2\,u(x) \\ u(0) \\ u'(0) \end{pmatrix}$$

defines a linear map whose domain is the space $U = \mathrm{C}^2$ of twice continuously differentiable functions $u(x)$, and whose image is the vector space V consisting of all triples[†] $\mathbf{v} = \begin{pmatrix} f(x) \\ a \\ b \end{pmatrix}$, where $f \in \mathrm{C}^0$ is a continuous function and $a, b \in \mathbb{R}$ are real constants. You should convince yourself that V is indeed a vector space under the evident addition and scalar multiplication operations. In this way, we can write the initial value problem (7.47) in linear systems form as $M[u] = \mathbf{f}$, where $\mathbf{f} = (\,x, 1, -1\,)^T$.

A similar construction applies to linear boundary value problems. For example, the boundary value problem

$$u'' + u = e^x, \qquad u(0) = 1, \qquad u(1) = 2,$$

is in the form of a linear system

$$B[u] = \mathbf{f}, \qquad \text{where} \qquad B[u] = \begin{pmatrix} u''(x) + u(x) \\ u(0) \\ u(1) \end{pmatrix}, \qquad \mathbf{f} = \begin{pmatrix} e^x \\ 1 \\ 2 \end{pmatrix}.$$

Note that $B\colon \mathrm{C}^2 \to V$ defines a linear map having the same domain and codomain as the initial value problem map M.

Exercises

7.4.1. *True or false*: If $F[\mathbf{x}]$ is an affine transformation on $\mathbb{R}^n$, then the equation $F[\mathbf{x}] = \mathbf{c}$ defines a linear system.

7.4.2. Place each of the following linear systems in the form (7.45). Carefully describe the linear function, its domain, its codomain, and the right-hand side of the system. Which systems are homogeneous? (a) $3x + 5 = 0$, (b) $x = y + z$, (c) $a = 2b - 3$, $b = c - 1$, (d) $3(p-2) = 2(q-3)$, $p+q = 0$, (e) $u' + 3xu = 0$, (f) $u' + 3x = 0$, (g) $u' = u$, $u(0) = 1$, (h) $u'' - u = e^x$, $u(0) = 3u(1)$, (i) $u'' + x^2 u = 3x$, $u(0) = 1$, $u'(0) = 0$, (j) $u' = v$, $v' = 2u$, (k) $u'' - v'' = 2u - v$, $u(0) = v(0)$, $u(1) = v(1)$, (l) $u(x) = 1 - 3\int_0^x u(y)\,dy$, (m) $\int_0^\infty u(t)\,e^{-s\,t}\,dt = 1 + s^2$, (n) $\int_0^1 u(x)\,dx = u\left(\tfrac{1}{2}\right)$, (o) $\int_0^1 u(y)\,dy = \int_0^1 y\,v(y)\,dy$, (p) $\dfrac{\partial u}{\partial t} + 2\,\dfrac{\partial u}{\partial x} = 1$, (q) $\dfrac{\partial u}{\partial x} = \dfrac{\partial v}{\partial y}$, $\dfrac{\partial u}{\partial y} = -\dfrac{\partial v}{\partial x}$, (r) $-\dfrac{\partial^2 u}{\partial x^2} - \dfrac{\partial^2 u}{\partial y^2} = x^2 + y^2 - 1$.

7.4.3. The Fredholm Alternative of Theorem 4.46 first appeared in the study of what are now known as *Fredholm integral equations*: $u(x) + \int_a^b K(x,y)\,u(y)\,dy = f(x)$, in which $K(x,y)$ and $f(x)$ are prescribed continuous functions. Explain how the integral equation is a linear system; i.e., describe the linear map L, its domain and codomain, and prove linearity.

[†] This is a particular case of the general Cartesian product construction between vector spaces; here $V = \mathrm{C}^0 \times \mathbb{R}^2$. See Exercise 2.1.13 for details.

7.4.4. Answer Exercise 7.4.3 for the *Volterra integral equation* $u(t) + \int_a^t K(t,s)\,u(s)\,ds = f(t)$, where $a \le t \le b$.

7.4.5. (a) Prove that the solution to the linear integral equation $u(t) = a + \int_0^t k(s)\,u(s)\,ds$ solves the linear initial value problem $\dfrac{du}{dt} = k(t)\,u(t), \;\; u(0) = a$.

(*b*) Use part (a) to solve the following integral equations

(*i*) $u(t) = 2 - \int_0^t u(s)\,ds$, (*ii*) $u(t) = 1 + 2\int_1^t s\,u(s)\,ds$, (*iii*) $u(t) = 3 + \int_0^t e^s\,u(s)\,ds$.

The Superposition Principle

Before attempting to tackle general inhomogeneous linear systems, we should look first at the homogeneous version. The most important fact is that homogeneous linear systems admit a superposition principle that allows one to construct new solutions from known solutions. Recall that the word "superposition" refers to taking linear combinations of solutions.

Consider a general homogeneous linear system

$$L[\mathbf{z}] = \mathbf{0}, \tag{7.48}$$

where $L: U \to V$ is a linear function. If we are given two solutions, say $\mathbf{z}_1$ and $\mathbf{z}_2$, meaning that

$$L[\mathbf{z}_1] = \mathbf{0}, \qquad L[\mathbf{z}_2] = \mathbf{0},$$

then their sum $\mathbf{z}_1 + \mathbf{z}_2$ is automatically a solution, since, in view of the linearity of L,

$$L[\mathbf{z}_1 + \mathbf{z}_2] = L[\mathbf{z}_1] + L[\mathbf{z}_2] = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

Similarly, given a solution $\mathbf{z}$ and any scalar c, the scalar multiple $c\,\mathbf{z}$ is automatically a solution, since

$$L[c\,\mathbf{z}] = c\,L[\mathbf{z}] = c\,\mathbf{0} = \mathbf{0}.$$

Combining these two elementary observations, we can now state the general *superposition principle*. The proof is an immediate consequence of formula (7.4).

Theorem 7.30. If $\mathbf{z}_1, \ldots, \mathbf{z}_k$ are all solutions to the same homogeneous linear system $L[\mathbf{z}] = \mathbf{0}$, then every linear combination $c_1\mathbf{z}_1 + \cdots + c_k\mathbf{z}_k$ is also a solution.

As with matrices, we call the solution space to the homogeneous linear system (7.48) the *kernel* of the linear function L. The superposition principle implies that the kernel always forms a subspace.

Proposition 7.31. If $L: U \to V$ is a linear function, then its *kernel*

$$\ker L = \{\, \mathbf{z} \in U \mid L[\mathbf{z}] = 0 \,\} \subset U \tag{7.49}$$

is a subspace of the domain space U.

As we know, in the case of linear matrix systems, the kernel can be explicitly determined by applying the usual Gaussian Elimination algorithm. To solve more general homogeneous linear systems, e.g., linear differential equations, one must develop appropriate analytical solution techniques.

Example 7.32. Consider the second order linear differential operator

$$L = D^2 - 2\,D - 3, \tag{7.50}$$

which maps the function $u(x)$ to the function

$$L[\,u\,] = (D^2 - 2\,D - 3)[\,u\,] = u'' - 2\,u' - 3\,u.$$

The associated homogeneous system takes the form of a homogeneous, linear, constant coefficient second order ordinary differential equation

$$L[\,u\,] = u'' - 2\,u' - 3\,u = 0. \tag{7.51}$$

In accordance with the standard solution method, we plug the exponential ansatz†

$$u = e^{\lambda\,x}$$

into the equation. The result is

$$L[\,e^{\lambda\,x}\,] = D^2[\,e^{\lambda\,x}\,] - 2\,D[\,e^{\lambda\,x}\,] - 3\,e^{\lambda\,x} = \lambda^2\,e^{\lambda\,x} - 2\,\lambda\,e^{\lambda\,x} - 3\,e^{\lambda\,x} = (\lambda^2 - 2\,\lambda - 3)e^{\lambda\,x}.$$

Therefore, $u = e^{\lambda\,x}$ is a solution if and only if λ satisfies the *characteristic equation*

$$0 = \lambda^2 - 2\,\lambda - 3 = (\lambda - 3)(\lambda + 1).$$

The two roots are $\lambda_1 = 3,\ \ \lambda_2 = -1$, and hence

$$u_1(x) = e^{3\,x}, \qquad\qquad u_2(x) = e^{-\,x}, \tag{7.52}$$

are two linearly independent solutions of (7.51). According to the general superposition principle, every linear combination

$$u(x) = c_1\,u_1(x) + c_2\,u_2(x) = c_1\,e^{3\,x} + c_2\,e^{-\,x} \tag{7.53}$$

of these two basic solutions is also a solution, for any choice of constants c_1, c_2. In fact, this two-parameter family (7.53) constitutes the most general solution to the ordinary differential equation (7.51); indeed, this is a consequence of Theorem 7.34 below. Thus, the kernel of the second order differential operator (7.50) is two-dimensional, with basis given by the independent exponential solutions (7.52).

In general, the solution space to an n^{th} order homogeneous linear ordinary differential equation

$$L[\,u\,] = a_n(x)\,\frac{d^n u}{dx^n} + a_{n-1}(x)\,\frac{d^{n-1} u}{dx^{n-1}} + \ \cdots\ + a_1(x)\,\frac{du}{dx} + a_0(x)u = 0 \tag{7.54}$$

is a subspace of the vector space $\mathrm{C}^n(\,a,b\,)$ of n times continuously differentiable functions defined on an open interval‡ $a < x < b$, since it is just the kernel of a linear differential

---

† The German word *Ansatz* refers to the method of finding a solution to a complicated equation by guessing the solution's form in advance. Typically, one is not clever enough to guess the precise solution, and so the ansatz will have one or more free parameters — in this case the constant exponent λ — that, with some luck, can be rigged up to fulfill the requirements imposed by the equation. Thus, a reasonable English translation of "ansatz" is "inspired guess".

‡ We allow a and/or b to be infinite.

operator $L: \mathrm{C}^n(a,b) \to \mathrm{C}^0(a,b)$. This implies that linear combinations of solutions are also solutions. To determine the number of solutions, or, more precisely, the dimension of the solution space, we need to impose some mild restrictions on the differential operator.

Definition 7.33. A differential operator L given by (7.54) is called *nonsingular* on an open interval (a,b) if all its coefficients are continuous functions, so $a_n(x), \ldots, a_0(x) \in \mathrm{C}^0(a,b)$, and its *leading coefficient* does not vanish: $a_n(x) \neq 0$ for all $a < x < b$.

The basic existence and uniqueness theorems governing nonsingular homogeneous linear ordinary differential equations can be reformulated as a characterization of the dimension of the solution space.

Theorem 7.34. The kernel of a nonsingular n^{th} order ordinary differential operator is an n-dimensional subspace $\ker L \subset \mathrm{C}^n(a,b)$.

A proof of this theorem relies on the fundamental existence and uniqueness theorems for ordinary differential equations, and can be found in [**7**, **36**]. The fact that the kernel has dimension n means that it has a basis consisting of n linearly independent solutions $u_1(x), \ldots, u_n(x) \in \mathrm{C}^n(a,b)$ with the property that every solution to the homogeneous differential equation (7.54) is given by a linear combination

$$u(x) = c_1\, u_1(x) + \;\cdots\; + c_n\, u_n(x),$$

where $c_1, \ldots, c_n$ are arbitrary constants. Therefore, once we find n linearly independent solutions of an n^{th} order homogeneous linear ordinary differential equation, we can immediately write down its most general solution.

The condition that the leading coefficient $a_n(x) \neq 0$ is essential. Points where $a_n(x) = 0$ are known as *singular points*. Singular points show up in many applications, and must be treated separately and with care, [**7**, **22**, **61**]. Of course, if the coefficients are constant, then there is nothing to worry about — either the leading coefficient is nonzero, $a_n \neq 0$, or the differential equation is, in fact, of lower order than advertised. Here is the prototypical example of an ordinary differential equation with a singular point.

Example 7.35. A second order *Euler differential equation* takes the form

$$E[u] = a\,x^2\,u'' + b\,x\,u' + c\,u = 0, \tag{7.55}$$

where $a \neq 0$ and b, c are constants. Here $E = a\,x^2 D^2 + b\,x\,D + c$ is a second order variable coefficient linear differential operator. Instead of the exponential solution ansatz used in the constant coefficient case, Euler equations are solved by using a power ansatz

$$u(x) = x^r \tag{7.56}$$

with unknown exponent r. Substituting into the differential equation, we find

$$\begin{aligned} E[x^r] &= a\,x^2 D^2[x^r] + b\,x\,D[x^r] + c\,x^r \\ &= a\,r\,(r-1)\,x^r + b\,r\,x^r + c\,x^r = [\,a\,r\,(r-1) + b\,r + c\,]\,x^r. \end{aligned}$$

Thus, x^r is a solution if and only if r satisfies the *characteristic equation*

$$a\,r\,(r-1) + b\,r + c = a\,r^2 + (b-a)\,r + c = 0. \tag{7.57}$$

If the quadratic characteristic equation has two distinct real roots, $r_1 \neq r_2$, then we obtain two linearly independent solutions $u_1(x) = x^{r_1}$ and $u_2(x) = x^{r_2}$, and so the general (real) solution to (7.55) has the form

$$u(x) = c_1 \,|\, x \,|^{r_1} + c_2 \,|\, x \,|^{r_2}. \tag{7.58}$$

(The absolute values are usually needed to ensure that the solutions remain real when $x < 0$.) The other cases — repeated roots and complex roots — will be discussed below.

The Euler equation has a singular point at $x = 0$, where its leading coefficient vanishes. Theorem 7.34 assures us that the differential equation has a two-dimensional solution space on every interval not containing the singular point. However, predicting the number of solutions that remain continuously differentiable at $x = 0$ is not so easy, since it depends on the values of the exponents r_1 and r_2. For instance, the case

$$x^2 u'' - 3\,x\,u' + 3\,u = 0 \qquad \text{has general solution} \qquad u = c_1 x + c_2 x^3,$$

which forms a two-dimensional subspace of $\mathrm{C}^0(\mathbb{R})$. However,

$$x^2 u'' + x\,u' - u = 0 \qquad \text{has general solution} \qquad u = c_1 x + \frac{c_2}{x}\,,$$

and only the multiples of the first solution x are continuous at $x = 0$. Therefore, the solutions that are continuous everywhere form only a one-dimensional subspace of $\mathrm{C}^0(\mathbb{R})$. Finally,

$$x^2 u'' + 5\,x\,u' + 3\,u = 0 \qquad \text{has general solution} \qquad u = \frac{c_1}{x} + \frac{c_2}{x^3}\,.$$

In this case, there are no nontrivial solutions $u(x) \not\equiv 0$ that are continuous at $x = 0$, and so the space of solutions defined on all of $\mathbb{R}$ is zero-dimensional.

The superposition principle is equally valid in the study of homogeneous linear partial differential equations. Here is a particularly noteworthy example.

Example 7.36. Consider the *Laplace equation*

$$\Delta[\,u\,] = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{7.59}$$

for a function $u(x, y)$ defined on a domain $\Omega \subset \mathbb{R}^2$. The Laplace equation is named after the renowned eighteenth-century French mathematician Pierre-Simon Laplace, and is *the* most important partial differential equation. Its applications range over almost all fields of mathematics, physics, and engineering, including complex analysis, differential geometry, fluid mechanics, electromagnetism, elasticity, thermodynamics, and quantum mechanics, [**61**]. The Laplace equation is a homogeneous linear partial differential equation corresponding to the partial differential operator $\Delta = \partial_x^2 + \partial_y^2$ known as the *Laplacian*. Linearity can either be proved directly, or by noting that Δ is built up from the basic linear partial derivative operators ∂_x, ∂_y by the processes of composition and addition, as detailed in Exercise 7.1.46. Solutions to the Laplace equation are known as *harmonic functions*.

Unlike homogeneous linear ordinary differential equations, there is an infinite number of linearly independent solutions to the Laplace equation. Examples include the trigonometric/exponential solutions

$$e^{\omega\,x} \cos \omega\,y, \qquad e^{\omega\,x} \sin \omega\,y, \qquad e^{\omega\,y} \cos \omega\,x, \qquad e^{\omega\,y} \sin \omega\,y,$$

where ω is *any* real constant. There are also infinitely many independent *harmonic polynomial* solutions, the first few of which are

$$1, \qquad x, \qquad y, \qquad x^2 - y^2, \qquad x\,y, \qquad x^3 - 3\,x\,y^2, \qquad \ldots\,.$$

The reader might enjoy finding some more polynomial solutions and trying to spot the pattern. (The answer will appear shortly.) As usual, we can build up more complicated solutions by taking general linear combinations of these particular ones; for instance, $u(x,y) = 1 - 4xy + 2e^{3x}\cos 3y$ is automatically a solution. See [**61**] for further developments.

Exercises

7.4.6. Solve the following homogeneous linear ordinary differential equations. What is the dimension of the solution space? (a) $u'' - 4u = 0$, (b) $u'' - 6u' + 8u = 0$, (c) $u''' - 9u' = 0$, (d) $u'''' + 4u''' - u'' - 16u' - 12u = 0$.

7.4.7. Define $L[y] = y'' + y$. (a) Prove directly from the definition that $L\colon \mathrm{C}^2[a,b] \to \mathrm{C}^0[a,b]$ is a linear transformation. (b) Determine $\ker L$.

7.4.8. Answer Exercise 7.4.7 when $L = 3D^2 - 2D - 5$.

7.4.9. Consider the linear differential equation $y''' + 5y'' + 3y' - 9y = 0$. (a) Write the equation in the form $L[y] = 0$ for a differential operator $L = p(D)$. (b) Find a basis for $\ker L$, and then write out the general solution to the differential equation.

7.4.10. The following functions are solutions to a constant coefficient homogeneous scalar ordinary differential equation. (i) Determine the least possible order of the differential equation, and (ii) write down an appropriate differential equation.

(a) $e^{2x} + e^{-3x}$, (b) $1 + e^{-x}$, (c) $x e^{x}$, (d) $e^{x} + 2e^{2x} + 3e^{3x}$.

7.4.11. Solve the following Euler differential equations: (a) $x^2u'' + 5xu' - 5u = 0$, (b) $2x^2u'' - xu' - 2u = 0$, (c) $x^2u'' - u = 0$, (d) $x^2u'' + xu' - 3u = 0$, (e) $3x^2u'' - 5xu' - 3u = 0$, (f) $\dfrac{d^2u}{dx^2} + \dfrac{2}{x}\dfrac{du}{dx} = 0$.

7.4.12. Solve the third order Euler differential equation $x^3u''' + 2x^2u'' - 3xu' + 3u = 0$ by using the power ansatz (7.56). What is the dimension of the solution space for $x > 0$? For all x?

7.4.13. (i) Show that if $u(x)$ solves the Euler equation $ax^2\dfrac{d^2u}{dx^2} + bx\dfrac{du}{dx} + cu = 0$, then $v(t) = u(e^t)$ solves a linear, constant coefficient differential equation. (ii) Use this alternative technique to solve the Euler differential equations in Exercise 7.4.11.

♢ 7.4.14. (a) Use the method in Exercise 7.4.13 to solve an Euler equation whose characteristic equation has a double root $r_1 = r_2 = r$. (b) Solve the specific equations

(i) $x^2u'' - xu' + u = 0$, (ii) $\dfrac{d^2u}{dx^2} + \dfrac{1}{x}\dfrac{du}{dx} = 0$.

7.4.15. Show that if $u(x)$ solves $xu'' + 2u' - 4xu = 0$, then $v(x) = x\,u(x)$ solves a linear, constant coefficient equation. Use this to find the general solution to the given differential equation. Which of your solutions are continuous at the singular point $x = 0$? Differentiable?

7.4.16. Let $S \subset \mathbb{R}$ be an open subset (i.e., a union of open intervals), and let $D\colon \mathrm{C}^1(S) \to \mathrm{C}^0(S)$ be the derivative operator $D[f] = f'$. *True or false*: $\ker D$ is a one-dimensional subspace of $\mathrm{C}^1(S)$.

7.4.17. Show that $\log(x^2 + y^2)$ and $\dfrac{x}{x^2 + y^2}$ are harmonic functions, that is, solutions of the two-dimensional Laplace equation.

7.4.18. Find all solutions $u = f(r)$ of the two-dimensional Laplace equation that depend only on the radial coordinate $r = \sqrt{x^2 + y^2}$. Do these solutions form a vector space? If so, what is its dimension?

7.4.19. Find all (real) solutions to the two-dimensional Laplace equation of the form $u = \log p(x, y)$, where $p(x, y)$ is a quadratic polynomial. Do these solutions form a vector space? If so, what is its dimension?

♡ 7.4.20. (a) Show that the function $e^x \cos y$ is a solution to the two-dimensional Laplace equation. (b) Show that its quadratic Taylor polynomial at $x = y = 0$ is harmonic. (c) What about its degree 3 Taylor polynomial? (d) Can you state a general theorem? (e) Test your result by looking at the Taylor polynomials of the harmonic function
$$\log\left[\,(x-1)^2 + y^2\,\right].$$

7.4.21. (a) Find a basis for, and the dimension of, the vector space consisting of all quadratic polynomial solutions of the three-dimensional Laplace equation $\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} + \dfrac{\partial^2 u}{\partial z^2} = 0$. (b) Do the same for the homogeneous cubic polynomial solutions.

7.4.22. Find all solutions $u = f(r)$ of the three-dimensional Laplace equation $\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} + \dfrac{\partial^2 u}{\partial z^2} = 0$ that depend only on the radial coordinate $r = \sqrt{x^2 + y^2 + z^2}$. Do these solutions form a vector space? If so, what is its dimension?

7.4.23. Let L, M be linear functions. (a) Prove that $\ker(L \circ M) \supseteq \ker M$. (b) Find an example in which $\ker(L \circ M) \neq \ker M$.

Inhomogeneous Systems

Now we turn our attention to inhomogeneous linear systems
$$L[\mathbf{u}] = \mathbf{f}, \tag{7.60}$$
where $L\colon U \to V$ is a linear function, $\mathbf{f} \in V$, and the desired solution $\mathbf{u} \in U$. Unless $\mathbf{f} = \mathbf{0}$, the set of solutions to (7.60) is *not* a subspace of U, but, rather, forms an affine subspace, as defined in Exercise 2.2.28. Here, the crucial question is existence — is there a solution to the system? In contrast, for the homogeneous system $L[\mathbf{z}] = \mathbf{0}$, existence is not an issue, since $\mathbf{0}$ is always a solution. The key question for homogeneous systems is uniqueness: either $\ker L = \{\mathbf{0}\}$, in which case $\mathbf{0}$ is the only solution, or $\ker L \neq \{\mathbf{0}\}$, in which case there are infinitely many nontrivial solutions $\mathbf{0} \neq \mathbf{z} \in \ker L$.

In the matrix case, the compatibility of an inhomogeneous system $A\mathbf{x} = \mathbf{b}$ — which was required for the existence of a solution — led to the general definition of the image of a matrix, which we copy verbatim for linear functions.

Definition 7.37. The *image* of a linear function $L\colon U \to V$ is the subspace
$$\operatorname{img} L = \{\, L[\mathbf{u}] \mid \mathbf{u} \in U \,\} \subset V.$$

The proof that $\operatorname{img} L$ is a subspace of the codomain is straightforward: If $\mathbf{f} = L[\mathbf{u}]$ and $\mathbf{g} = L[\mathbf{v}]$ are any two elements of the image, so is any linear combination, since, by linearity
$$c\,\mathbf{f} + d\,\mathbf{g} = c\,L[\mathbf{u}] + d\,L[\mathbf{v}] = L[c\,\mathbf{u} + d\,\mathbf{v}] \in \operatorname{img} L.$$
For example, if $L[\mathbf{u}] = A\mathbf{u}$ is given by multiplication by an $m \times n$ matrix, then its image is the subspace $\operatorname{img} L = \operatorname{img} A \subset \mathbb{R}^m$ spanned by the columns of A — the *column space*

of the coefficient matrix. When L is a linear differential operator, or more general linear operator, characterizing its image can be a much more challenging problem.

The fundamental theorem regarding solutions to inhomogeneous linear equations exactly mimics our earlier result, Theorem 2.39, for matrix systems.

Theorem 7.38. Let $L: U \to V$ be a linear function. Let $\mathbf{f} \in V$. Then the inhomogeneous linear system

$$L[\mathbf{u}] = \mathbf{f} \tag{7.61}$$

has a solution if and only if $\mathbf{f} \in \operatorname{img} L$. In this case, the general solution to the system has the form

$$\mathbf{u} = \mathbf{u}^\star + \mathbf{z}, \tag{7.62}$$

where $\mathbf{u}^\star$ is a particular solution, so $L[\mathbf{u}^\star] = \mathbf{f}$, and $\mathbf{z}$ is any element of $\ker L$, i.e., a solution to the corresponding homogeneous system

$$L[\mathbf{z}] = \mathbf{0}. \tag{7.63}$$

Proof: We merely repeat the proof of Theorem 2.39. The existence condition $\mathbf{f} \in \operatorname{img} L$ is an immediate consequence of the definition of the image. Suppose $\mathbf{u}^\star$ is a particular solution to (7.61). If $\mathbf{z}$ is a solution to (7.63), then, by linearity,

$$L[\mathbf{u}^\star + \mathbf{z}] = L[\mathbf{u}^\star] + L[\mathbf{z}] = \mathbf{f} + \mathbf{0} = \mathbf{f},$$

and hence $\mathbf{u}^\star + \mathbf{z}$ is also a solution to (7.61). To show that every solution has this form, let $\mathbf{u}$ be a second solution, so that $L[\mathbf{u}] = \mathbf{f}$. Setting $\mathbf{z} = \mathbf{u} - \mathbf{u}^\star$, we find that

$$L[\mathbf{z}] = L[\mathbf{u} - \mathbf{u}^\star] = L[\mathbf{u}] - L[\mathbf{u}^\star] = \mathbf{f} - \mathbf{f} = \mathbf{0}.$$

Therefore $\mathbf{z} \in \ker L$, and so $\mathbf{u}$ has the proper form (7.62). *Q.E.D.*

Corollary 7.39. The inhomogeneous linear system (7.61) has a *unique* solution if and only if $\mathbf{f} \in \operatorname{img} L$ and $\ker L = \{\mathbf{0}\}$.

Therefore, to prove that a linear system has a unique solution, we first need to prove an *existence result* that there is at least one solution, which requires the right-hand side $\mathbf{f}$ to lie in the image of the operator L, and then a *uniqueness result*, that the only solution to the homogeneous system $L[\mathbf{z}] = \mathbf{0}$ is the trivial zero solution $\mathbf{z} = \mathbf{0}$. Observe that whenever an inhomogeneous system $L[\mathbf{u}] = \mathbf{f}$ has a unique solution, then every other inhomogeneous system $L[\mathbf{u}] = \mathbf{g}$ that is defined by the *same* linear function also has a unique solution, provided $\mathbf{g} \in \operatorname{img} L$. In other words, uniqueness does not depend upon the external forcing — although existence might.

Remark. In physical systems, the inhomogeneity $\mathbf{f}$ typically corresponds to an external force. The decomposition formula (7.62) states that its effect on the linear system can be viewed as a combination of one specific response $\mathbf{u}^\star$ to the forcing and the system's internal, unencumbered motion, as represented by the homogeneous solution $\mathbf{z}$. Keep in mind that the particular solution is *not* uniquely defined (unless $\ker L = \{\mathbf{0}\}$), and any one solution can serve in this role.

Example 7.40. Consider the inhomogeneous linear second order differential equation

$$u'' + u' - 2u = x. \tag{7.64}$$

Note that this can be written in the linear system form

$$L[\,u\,] = x, \qquad \text{where} \qquad L = D^2 + D - 2$$

is a linear second order differential operator. The kernel of the differential operator L is found by solving the associated homogeneous linear equation

$$L[\,z\,] = z'' + z' - 2\,z = 0. \tag{7.65}$$

Applying the usual solution method, we find that the homogeneous differential equation (7.65) has a two-dimensional solution space, with basis functions

$$z_1(x) = e^{-2x}, \qquad\qquad z_2(x) = e^{x}.$$

Therefore, the general element of $\ker L$ is a linear combination

$$z(x) = c_1\, z_1(x) + c_2\, z_2(x) = c_1\, e^{-2x} + c_2\, e^{x}.$$

To find a particular solution to the inhomogeneous differential equation (7.64), we rely on the method of *undetermined coefficients*†. We introduce the solution ansatz $u = a\,x + b$, and compute

$$L[\,u\,] = L[\,a\,x + b\,] = a - 2\,(a\,x + b) = -\,2\,a\,x + (a - 2\,b) = x.$$

Equating the coefficients of x and 1, and then solving for $a = -\frac{1}{2}$, $b = -\frac{1}{4}$, we deduce that

$$u^\star(x) = -\,\tfrac{1}{2}\,x - \tfrac{1}{4}$$

is a particular solution to the inhomogeneous differential equation. Theorem 7.38 then says that the general solution is

$$u(x) = u^\star(x) + z(x) = -\,\tfrac{1}{2}\,x - \tfrac{1}{4} + c_1\, e^{-2x} + c_2\, e^{x}.$$

Example 7.41. By inspection, we see that

$$u(x,y) = -\,\tfrac{1}{2}\,\sin(x+y)$$

is a solution to the particular *Poisson equation*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \sin(x+y). \tag{7.66}$$

Theorem 7.38 implies that *every* solution to this inhomogeneous version of the Laplace equation (7.59) takes the form

$$u(x,y) = -\,\tfrac{1}{2}\,\sin(x+y) + z(x,y),$$

where $z(x,y)$ is an arbitrary harmonic function, i.e., a solution to the homogeneous Laplace equation.

† One could also employ the method of *variation of parameters*, although usually the undetermined coefficient method, when applicable, is the more straightforward of the two. Details can be found in most ordinary differential equations texts, including [**7**, **22**].

Example 7.42. Let us solve the second order linear boundary value problem

$$u'' + u = x, \qquad u(0) = 0, \qquad u(\pi) = 0. \tag{7.67}$$

As with initial value problems, the first step is to solve the differential equation. To this end, we first solve the corresponding homogeneous differential equation $z'' + z = 0$. The usual method — see [**7**] or Example 7.50 below — shows that $\cos x$ and $\sin x$ form a basis for its solution space. The method of undetermined coefficients then produces the particular solution $u^\star(x) = x$ to the inhomogeneous differential equation, and so its general solution is

$$u(x) = x + c_1 \cos x + c_2 \sin x. \tag{7.68}$$

The next step is to see whether any solutions also satisfy the boundary conditions. Plugging formula (7.68) into the boundary conditions yields

$$u(0) = c_1 = 0, \qquad u(\pi) = \pi - c_1 = 0.$$

However, these two conditions are incompatible, and so there is *no* solution to the linear system (7.67). The function $f(x) = x$ does not lie in the image of the differential operator $L[u] = u'' + u$ when u is subjected to the boundary conditions. Or, to state it another way, $(x, 0, 0)^T$ does not belong to the image of the linear operator $M[u] = (u'' + u, u(0), u(\pi))^T$ defining the boundary value problem.

On the other hand, if we slightly modify the inhomogeneity, the boundary value problem

$$u'' + u = x - \tfrac{1}{2}\pi, \qquad u(0) = 0, \qquad u(\pi) = 0, \tag{7.69}$$

does admit a solution, but it fails to be unique. Applying the preceding solution techniques, we find that

$$u(x) = x - \tfrac{1}{2}\pi + \tfrac{1}{2}\pi \cos x + c \sin x$$

solves the system for *any* choice of constant c, and so the boundary value problem (7.69) admits infinitely many solutions. Observe that $z(x) = \sin x$ is a basis for the kernel or solution space of the corresponding homogeneous boundary value problem

$$z'' + z = 0, \qquad z(0) = 0, \qquad z(\pi) = 0,$$

while $u^\star(x) = x - \frac{1}{2}\pi + \frac{1}{2}\pi \cos x$ represents a particular solution to the inhomogeneous system. Thus, $u(x) = u^\star(x) + z(x)$, in conformity with the general formula (7.62).

Incidentally, if we modify the interval of definition, considering

$$u'' + u = f(x), \qquad u(0) = 0, \qquad u\left(\tfrac{1}{2}\pi\right) = 0, \tag{7.70}$$

then the homogeneous boundary value problem, with $f(x) \equiv 0$, has only the trivial solution, and so the inhomogeneous system admits a unique solution for *any* inhomogeneity $f(x)$. For example, if $f(x) = x$, then

$$u(x) = x - \tfrac{1}{2}\pi \sin x \tag{7.71}$$

is the unique solution to the resulting boundary value problem.

This example highlights some crucial differences between boundary value problems and initial value problems for ordinary differential equations. Nonsingular initial value problems have a unique solution for every suitable set of initial conditions. Boundary value problems have more of the flavor of linear algebraic systems, either possessing a unique solution for all possible inhomogeneities, or admitting either no solution or infinitely many solutions, depending on the right-hand side. An interesting question is how to characterize the inhomogeneities $f(x)$ that admit a solution, i.e., that lie in the image of the associated linear operator. These issues are explored in depth in [**61**].

Exercises

7.4.24. For each of the following inhomogeneous systems, determine whether the right-hand side lies in the image of the coefficient matrix, and, if so, write out the general solution, clearly identifying the particular solution and the kernel element.

(a) $\begin{pmatrix} 1 & -1 \\ 3 & -3 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, (b) $\begin{pmatrix} 2 & 1 & 4 \\ -1 & 2 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 2 & -1 \\ 2 & 0 & 1 \\ 1 & -2 & 2 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix}$,

(d) $\begin{pmatrix} -2 & 1 \\ -2 & 3 \\ 3 & -5 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, (e) $\begin{pmatrix} -1 & 3 & 0 & 2 \\ 2 & -6 & 1 & -1 \\ -3 & 9 & -2 & 0 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 2 \\ -2 \\ 2 \end{pmatrix}$.

7.4.25. Which of the following systems have a unique solution?

(a) $\begin{pmatrix} 3 & 1 \\ -1 & -1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$, (b) $\begin{pmatrix} 1 & 2 & -1 \\ -2 & 3 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$,

(c) $\begin{pmatrix} 2 & 1 & -1 \\ 0 & -3 & -3 \\ 2 & 0 & -2 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 5 \end{pmatrix}$, (d) $\begin{pmatrix} 1 & 4 & -1 \\ 1 & 3 & -3 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \\ 1 \end{pmatrix}$.

7.4.26. Solve the following inhomogeneous linear ordinary differential equations:
(a) $u' - 4u = x - 3$, (b) $5u'' - 4u' + 4u = e^x \cos x$, (c) $u'' - 3u' = e^{3x}$.

7.4.27. Solve the following initial value problems: (a) $u' + 3u = e^x$, $u(1) = 0$, (b) $u'' + 4u = 1$, $u(\pi) = u'(\pi) = 0$, (c) $u'' - u' - 2u = e^x + e^{-x}$, $u(0) = u'(0) = 0$, (d) $u'' + 2u' + 5u = \sin x$, $u(0) = 1$, $u'(0) = 0$, (e) $u''' - u'' + u' - u = x$, $u(0) = 0$, $u'(0) = 1$, $u''(0) = 0$.

7.4.28. Solve the following inhomogeneous Euler equations using either variation of parameters or the change of variables method discussed in Exercise 7.4.13:
(a) $x^2 u'' + x u' - u = x$, (b) $x^2 u'' - 2x u' + 2u = \log x$, (c) $x^2 u'' - 3x u' - 5u = 3x - 5$.

7.4.29. Write down all solutions to the following boundary value problems. Label your answer as (*i*) unique solution, (*ii*) no solution, (*iii*) infinitely many solutions.
(a) $u'' + 2u = 2x$, $u(0) = 0$, $u(\pi) = 0$, (b) $u'' + 4u = \cos x$, $u(-\pi) = 0$, $u(\pi) = 1$,
(c) $u'' - 2u' + u = x - 2$, $u(0) = -1$, $u(1) = 1$,
(d) $u'' + 2u' + 2u = 1$, $u(0) = \frac{1}{2}$, $u(\pi) = \frac{1}{2}$, (e) $u'' - 3u' + 2u = 4x$, $u(0) = 0$, $u(1) = 0$,
(f) $x^2 u'' + x u' - u = 0$, $u(0) = 1$, $u(1) = 0$, (g) $x^2 u'' - 6u = 0$, $u(1) = 1$, $u(2) = -1$,
(h) $x^2 u'' - 2x u' + 2u = 0$, $u(0) = 0$, $u(1) = 1$.

♢ 7.4.30. Let $L: U \to V$ be a linear function, and let $W \subset U$ be a subspace of the domain space. (a) Prove that $Y = \{ L[\mathbf{w}] \mid \mathbf{w} \in W \} \subset \operatorname{img} L \subset V$ is a subspace of the image.
(b) Prove that $\dim Y \leq \dim W$. Conclude that a linear transformation can never increase the dimension of a subspace.

♢ 7.4.31. (a) Show that if $L: V \to V$ is linear and $\ker L \neq \{\mathbf{0}\}$, then L is not invertible.
(b) Show that if $\operatorname{img} L \neq V$, then L is not invertible.
(c) Give an example of a linear map with $\ker L = \{\mathbf{0}\}$ that is not invertible. *Hint*: First explain why your example must be on an infinite-dimensional vector space.

Superposition Principles for Inhomogeneous Systems

The *superposition principle* for inhomogeneous linear systems allows us to combine different inhomogeneities — provided that we do not change the underlying linear operator. The result is a straightforward generalization of the matrix version described in Theorem 2.44.

Theorem 7.43. Let $L: U \to V$ be a linear function. Suppose that, for each $i = 1, \ldots, k$, we know a particular solution $\mathbf{u}_i^\star$ to the inhomogeneous linear system $L[\mathbf{u}] = \mathbf{f}_i$ for some $\mathbf{f}_i \in \operatorname{img} L$. Then, given scalars $c_1, \ldots, c_k$, a particular solution to the combined inhomogeneous system

$$L[\mathbf{u}] = c_1 \mathbf{f}_1 + \cdots + c_k \mathbf{f}_k \tag{7.72}$$

is the corresponding linear combination

$$\mathbf{u}^\star = c_1 \mathbf{u}_1^\star + \cdots + c_k \mathbf{u}_k^\star \tag{7.73}$$

of particular solutions. The general solution to the inhomogeneous system (7.72) is

$$\mathbf{u} = \mathbf{u}^\star + \mathbf{z} = c_1 \mathbf{u}_1^\star + \cdots + c_k \mathbf{u}_k^\star + \mathbf{z}, \tag{7.74}$$

where $\mathbf{z} \in \ker L$ is an arbitrary solution to the associated homogeneous system $L[\mathbf{z}] = \mathbf{0}$.

The proof is an easy consequence of linearity, and left to the reader. In physical terms, the superposition principle can be interpreted as follows. If we know the response of a linear physical system to several different external forces, represented by $\mathbf{f}_1, \ldots, \mathbf{f}_k$, then the response of the system to a linear combination of these forces is just the self-same linear combination of the individual responses. The homogeneous solution $\mathbf{z}$ represents an internal motion that the system acquires independent of any external forcing. Superposition relies on the linearity of the system, and so is always applicable in quantum mechanics, which is an inherently linear theory. On the other hand, in classical and relativistic mechanics, superposition is valid only in the linear approximation regime governing small motions/displacements/etc. Large-scale motions of a fully nonlinear physical system are more subtle, and combinations of external forces may lead to unexpected results.

Example 7.44. In Example 7.42, we found that a particular solution to the linear differential equation

$$u'' + u = x \qquad \text{is} \qquad u_1^\star = x.$$

The method of undetermined coefficients can be used to solve the inhomogeneous equation

$$u'' + u = \cos x.$$

Since $\cos x$ and $\sin x$ are already solutions to the homogeneous equation, we must use the solution ansatz

$$u = a\,x \cos x + b\,x \sin x,$$

which, when substituted into the differential equation, produces the particular solution

$$u_2^\star = -\tfrac{1}{2}\,x \sin x.$$

Therefore, by the superposition principle, the combined inhomogeneous system

$$u'' + u = 3\,x - 2\cos x$$

has a particular solution

$$u^\star = 3\,u_1^\star - 2\,u_2^\star = 3\,x + x \sin x.$$

The general solution is obtained by appending an arbitrary solution to the homogeneous equation:

$$u = 3x + x\sin x + c_1 \cos x + c_2 \sin x.$$

Example 7.45. Consider the boundary value problem

$$u'' + u = x, \qquad u(0) = 2, \qquad u\left(\tfrac{1}{2}\pi\right) = -1, \tag{7.75}$$

which is a modification of (7.70) with inhomogeneous boundary conditions. The superposition principle applies here, and allows us to decouple the inhomogeneity due to the forcing from the inhomogeneity due to the boundary conditions. We decompose the right-hand side, written in vectorial form, into simpler constituents[†]

$$\begin{pmatrix} x \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The first vector on the right-hand side corresponds to the preceding boundary value problem (7.70), whose solution was found in (7.71). The second and third vectors correspond to the unforced boundary value problems

$$u'' + u = 0, \quad u(0) = 1, \quad u\left(\tfrac{1}{2}\pi\right) = 0, \qquad \text{and} \qquad u'' + u = 0, \quad u(0) = 0, \quad u\left(\tfrac{1}{2}\pi\right) = 1,$$

with respective solutions $u(x) = \cos x$ and $u(x) = \sin x$. Therefore, the solution to the combined boundary value problem (7.75) is the same linear combination of these individual solutions:

$$u(x) = \left(x - \tfrac{1}{2}\pi \sin x\right) + 2\cos x - \sin x = x + 2\cos x - \left(1 + \tfrac{1}{2}\pi\right)\sin x.$$

The solution is unique because the corresponding homogeneous boundary value problem

$$z'' + z = 0, \qquad z(0) = 0, \qquad z\left(\tfrac{1}{2}\pi\right) = 0,$$

has only the trivial solution $z(x) \equiv 0$, as you can verify.

Exercises

7.4.32. Use superposition to solve the following inhomogeneous ordinary differential equations: (a) $u' + 2u = 1 + \cos x$, (b) $u'' - 9u = x + \sin x$, (c) $9u'' - 18u' + 10u = 1 + e^x \cos x$, (d) $u'' + u' - 2u = \sinh x$, where $\sinh x = \frac{1}{2}(e^x - e^{-x})$, (e) $u''' + 9u' = 1 + e^{3x}$.

7.4.33. Consider the differential equation $u'' + x u = 2$. Suppose you know solutions to the two boundary value problems $u(0) = 1$, $u(1) = 0$ and $u(0) = 0$, $u(1) = 1$. List all possible boundary value problems you can solve using superposition.

[†] **Warning.** When writing out a linear combination, make sure the scalars are *constants*! Writing the first summand as $x\,(1,0,0)^T$ will lead to an *incorrect* application of the superposition principle.

7.4.34. Consider the differential equation $x u'' - (x+1)u' + u = 0$. Suppose we know the solution to the initial value problem $u(1) = 2,\ u'(1) = 1$ is $u(x) = x+1$, while the solution to the initial value problem $u(1) = 1,\ u'(1) = 1$ is $u(x) = e^{x-1}$. (*a*) What is the solution to the initial value problem $u(1) = 3,\ u'(1) = -2$? (*b*) What is the general solution to the differential equation?

7.4.35. Consider the differential equation $4x u'' + 2u' + u = 0$. Given that $\cos\sqrt{x}$ solves the boundary value problem $u\left(\frac14\pi^2\right) = 0,\ u(\pi^2) = -1$, and $\sin\sqrt{x}$ solves the boundary value problem $u\left(\frac14\pi^2\right) = 1,\ u(\pi^2) = 0$, write down the solution to the boundary value problem $u\left(\frac14\pi^2\right) = -3,\ u(\pi^2) = 7$.

7.4.36. Solve the following boundary value problems by using superposition: (*a*) $u'' + 9u = x$, $u(0) = 1,\ u'(\pi) = 0$, (*b*) $u'' - 8u' + 16u = e^{4x},\ u(0) = 1,\ u(1) = 0$, (*c*) $u'' + 4u = \sin 3x$, $u'(0) = 0,\ u(2\pi) = 3$, (*d*) $u'' - 2u' + u = 1 + e^x,\ u'(0) = -1,\ u'(1) = 1$.

7.4.37. Given that $x^2 + y^2$ solves the Poisson equation $\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} = 4$, while $x^4 + y^4$ solves $\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} = 12(x^2 + y^2)$, write down a solution to $\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} = 1 + x^2 + y^2$.

♡ 7.4.38. *Reduction of order*: Suppose you know one solution $u_1(x)$ to the second order homogeneous differential equation $u'' + a(x)u' + b(x)u = 0$. (*a*) Show that if $u(x) = v(x)u_1(x)$ is any other solution, then $w(x) = v'(x)$ satisfies a first order differential equation. (*b*) Use reduction of order to find the general solution to the following equations, based on the indicated solution:
(*i*) $u'' - 2u' + u = 0,\ u_1(x) = e^x$, (*ii*) $x u'' + (x-1)u' - u = 0,\ u_1(x) = x - 1$, (*iii*) $u'' + 4x u' + (4x^2 + 2)u = 0,\ u_1(x) = e^{-x^2}$, (*iv*) $u'' - (x^2+1)u = 0,\ u_1(x) = e^{x^2/2}$.

♢ 7.4.39. Write out the details of the proof of Theorem 7.43.

Complex Solutions to Real Systems

As we know, solutions to a linear, homogeneous, constant coefficient ordinary differential equation are found by substituting an exponential ansatz, which effectively reduces the differential equation to the polynomial characteristic equation. Complex roots of the characteristic equation yield complex exponential solutions. But, if the equation is real, then the real and imaginary parts of the complex solutions are automatically real solutions. This solution technique is a particular case of a general principle for producing real solutions to real linear systems from, typically, simpler complex solutions. To work, the method requires us to impose some additional structure on the complex vector spaces involved.

Definition 7.46. A complex vector space V is called *conjugated* if it admits an operation of *complex conjugation* taking $\mathbf{u} \in V$ to $\overline{\mathbf{u}} \in V$ with the following properties:

(*a*) conjugating twice returns one to the original vector: $\overline{\overline{\mathbf{u}}} = \mathbf{u}$;

(*b*) compatibility with vector addition: $\overline{\mathbf{u} + \mathbf{v}} = \overline{\mathbf{u}} + \overline{\mathbf{v}}$ for all $\mathbf{u}, \mathbf{v} \in V$;

(*c*) compatibility with scalar multiplication, $\overline{\lambda\,\mathbf{u}} = \overline{\lambda}\,\overline{\mathbf{u}}$, for all $\lambda \in \mathbb{C}$ and $\mathbf{u} \in V$.

The simplest example of a conjugated vector space is $\mathbb{C}^n$. The complex conjugate of a vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)^T$ is obtained by conjugating all its entries, whereby $\overline{\mathbf{u}} = (\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_n)^T$. Thus,

$$\begin{aligned} \mathbf{u} &= \mathbf{v} + \mathrm{i}\,\mathbf{w}, \\ \overline{\mathbf{u}} &= \mathbf{v} - \mathrm{i}\,\mathbf{w}, \end{aligned} \qquad \text{where} \qquad \mathbf{v} = \operatorname{Re}\mathbf{u} = \frac{\mathbf{u}+\overline{\mathbf{u}}}{2}, \qquad \mathbf{w} = \operatorname{Im}\mathbf{u} = \frac{\mathbf{u}-\overline{\mathbf{u}}}{2\,\mathrm{i}}, \tag{7.76}$$

are the real and imaginary parts of $\mathbf{u} \in \mathbb{C}^n$. For example, if

$$\mathbf{u} = \begin{pmatrix} 1-2\,\mathrm{i} \\ 3\,\mathrm{i} \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix} + \mathrm{i} \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix}, \qquad \text{then} \qquad \overline{\mathbf{u}} = \begin{pmatrix} 1+2\,\mathrm{i} \\ -3\,\mathrm{i} \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix} - \mathrm{i} \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix},$$

whence

$$\operatorname{Re}\mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix}, \qquad \operatorname{Im}\mathbf{u} = \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix}.$$

The other prototypical example of a conjugated vector space is the space of complex-valued functions $f(x) = r(x) + \mathrm{i}\,s(x)$ defined on the interval $a \le x \le b$. The complex conjugate function is $\overline{f}(x) = \overline{f(x)} = r(x) - \mathrm{i}\,s(x)$. Thus, the complex conjugate of $e^{(1+3\,\mathrm{i})x} = e^x \cos 3x + \mathrm{i}\,e^x \sin 3x$ is $\overline{e^{(1+3\,\mathrm{i})x}} = e^{(1-3\,\mathrm{i})x} = e^x \cos 3x - \mathrm{i}\,e^x \sin 3x$, with $\operatorname{Re} e^{(1+3\,\mathrm{i})x} = e^x \cos 3x$, $\operatorname{Im} e^{(1+3\,\mathrm{i})x} = \mathrm{i}\,e^x \sin 3x$.

An element $\mathbf{v} \in V$ of a conjugated vector space is called *real* if $\overline{\mathbf{v}} = \mathbf{v}$. One easily checks that the real and imaginary parts of a general element, as defined by (7.76), are both real elements.

Warning. Not all subspaces of a conjugated vector space are conjugated. For example, the one-dimensional subspace of $\mathbb{C}^2$ spanned by $\mathbf{v}_1 = (1,2)^T$ is conjugated. Indeed, the complex conjugate of a general element $c\,\mathbf{v}_1 = (c, 2c)^T$ is $(\overline{c}, 2\overline{c})^T = \overline{c}\,\mathbf{v}_1$ which also belongs to the subspace. On the other hand, the subspace spanned by $(1, \mathrm{i})^T$ is *not* conjugated, because the complex conjugate of the element $(c, \mathrm{i}\,c)^T$ is $(\overline{c}, -\mathrm{i}\,\overline{c})^T$, which does not belong to the subspace unless $c = 0$. In Exercise 7.4.50 you are asked to prove that a subspace $V \subset \mathbb{C}^n$ is conjugated if and only if it has a basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ consisting entirely of real vectors. While conjugated subspaces play a role in certain applications, in practice we will deal only with $\mathbb{C}^n$ and the entire space of complex-valued functions, and so can suppress most of these somewhat technical details.

Definition 7.47. A linear function $L : U \to V$ between conjugated vector spaces is called *real* if it commutes with complex conjugation:

$$L[\,\overline{\mathbf{u}}\,] = \overline{L[\,\mathbf{u}\,]}. \tag{7.77}$$

For example, any linear function $L : \mathbb{C}^n \to \mathbb{C}^m$ is given by multiplication by an $m \times n$ matrix: $L[\mathbf{u}] = A\mathbf{u}$. The function is real if and only if A is a real matrix. Similarly, a differential operator (7.15) is real if and only if its coefficients are real-valued functions.

The solutions to a homogeneous system defined by a real linear function satisfy the following general *Reality Principle*.

Theorem 7.48. If $L[\mathbf{u}] = \mathbf{0}$ is a real homogeneous linear system and $\mathbf{u} = \mathbf{v} + \mathrm{i}\,\mathbf{w}$ is a complex solution, then its complex conjugate $\overline{\mathbf{u}} = \mathbf{v} - \mathrm{i}\,\mathbf{w}$ is also a solution. Moreover, both the real and imaginary parts, $\mathbf{v}$ and $\mathbf{w}$, of a complex solution are real solutions.

Proof: First note that, by reality, $L[\,\overline{\mathbf{u}}\,] = \overline{L[\,\mathbf{u}\,]} = \mathbf{0}$ whenever $L[\mathbf{u}] = \mathbf{0}$, and hence the complex conjugate $\overline{\mathbf{u}}$ of any solution is also a solution. Therefore, by linear superposition, $\mathbf{v} = \operatorname{Re}\mathbf{u} = \frac{1}{2}(\mathbf{u} + \overline{\mathbf{u}})$ and $\mathbf{w} = \operatorname{Im}\mathbf{u} = \frac{1}{2\,\mathrm{i}}(\mathbf{u} - \overline{\mathbf{u}})$ are also solutions. *Q.E.D.*

Example 7.49. The real linear matrix system

$$\begin{pmatrix} 2 & -1 & 3 & 0 \\ -2 & 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

has a complex solution

$$\mathbf{u} = \begin{pmatrix} -1-3\,\mathrm{i} \\ 1 \\ 1+2\,\mathrm{i} \\ -2-4\,\mathrm{i} \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 1 \\ -2 \end{pmatrix} + \mathrm{i} \begin{pmatrix} -3 \\ 0 \\ 2 \\ -4 \end{pmatrix}.$$

Since the coefficient matrix is real, the real and imaginary parts,

$$\mathbf{v} = (\,-1, 1, 1, -2\,)^T\,, \qquad \mathbf{w} = (\,-3, 0, 2, -4\,)^T\,,$$

are both solutions of the system, as can easily be checked.

On the other hand, the complex linear system

$$\begin{pmatrix} 2 & -2\,\mathrm{i} & \mathrm{i} & 0 \\ 1+\mathrm{i} & 0 & -2-\mathrm{i} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

has the complex solution

$$\mathbf{u} = \begin{pmatrix} 1-\mathrm{i} \\ -\mathrm{i} \\ 2 \\ 2+2\,\mathrm{i} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \end{pmatrix} + \mathrm{i} \begin{pmatrix} -1 \\ -1 \\ 0 \\ 2 \end{pmatrix}.$$

However, neither its real nor its imaginary part is a solution to the system.

Example 7.50. Consider the real homogeneous ordinary differential equation

$$u'' + 2\,u' + 5\,u = 0.$$

To solve it, we use the usual exponential ansatz $u = e^{\lambda x}$, leading to the characteristic equation

$$\lambda^2 + 2\,\lambda + 5 = 0.$$

There are two roots,

$$\lambda_1 = -1 + 2\,\mathrm{i}\,, \qquad \lambda_2 = -1 - 2\,\mathrm{i}\,,$$

leading, via Euler's formula (3.92), to the complex solutions

$$u_1(x) = e^{(-1+2\,\mathrm{i})\,x} = e^{-x}\cos 2\,x + \mathrm{i}\,e^{-x}\sin 2\,x,$$
$$u_2(x) = e^{(-1-2\,\mathrm{i})\,x} = e^{-x}\cos 2\,x - \mathrm{i}\,e^{-x}\sin 2\,x.$$

The complex conjugate of the first solution is the second, in accordance with Theorem 7.48. Moreover, the real and imaginary parts of the two solutions

$$v(x) = e^{-x}\cos 2\,x, \qquad w(x) = e^{-x}\sin 2\,x,$$

are individual real solutions. Since the solution space is two-dimensional, the general solution is a linear combination

$$u(x) = c_1\,e^{-x}\cos 2\,x + c_2\,e^{-x}\sin 2\,x,$$

of the two linearly independent real solutions.

Example 7.51. Consider the real second order Euler differential equation

$$L[u] = x^2 u'' + 7\,x\,u' + 13\,u = 0.$$

The roots of the associated characteristic equation

$$r\,(r-1) + 7\,r + 13 = r^2 + 6\,r + 13 = 0$$

are complex: $r = -3 \pm 2\,\mathrm{i}$, and the resulting solutions $x^{-3+2\,\mathrm{i}}, x^{-3-2\,\mathrm{i}}$ are complex conjugate powers. We use Euler's formula (3.92), to obtain their real and imaginary parts:

$$x^{-3+2\,\mathrm{i}} = x^{-3}\, e^{2\,\mathrm{i}\,\log x} = x^{-3}\cos(2\log x) + \mathrm{i}\, x^{-3}\sin(2\log x),$$

valid for $x > 0$. (For $x < 0$ just replace x by $-x$ in the above formula.) Again, by Theorem 7.48, the real and imaginary parts of the complex solution are by themselves real solutions to the equation. Therefore, the general real solution to this differential equation for $x > 0$ is

$$u(x) = c_1\, x^{-3}\cos(2\,\log x) + c_2\, x^{-3}\sin(2\,\log x).$$

Example 7.52. The complex monomial

$$u(x,y) = (x + \mathrm{i}\,y)^n$$

is a solution to the Laplace equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

because, by the chain rule,

$$\frac{\partial^2 u}{\partial x^2} = n(n-1)(x+\mathrm{i}\,y)^{n-2}, \qquad \frac{\partial^2 u}{\partial y^2} = n(n-1)\,\mathrm{i}^2\,(x+\mathrm{i}\,y)^{n-2} = -\,n(n-1)(x+\mathrm{i}\,y)^{n-2}.$$

Since the Laplacian operator is real, Theorem 7.48 implies that the real and imaginary parts of this complex solution are real solutions. The resulting real solutions are the harmonic polynomials introduced in Example 7.36.

Knowing this, it is relatively easy to find the explicit formulas for the harmonic polynomials. We appeal to the binomial formula

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{k} x^k y^{n-k}, \qquad \text{where} \qquad \binom{n}{k} = \frac{n!}{k!\,(n-k)!} \tag{7.78}$$

is the standard notation for the *binomial coefficients*. Since $\mathrm{i}^2 = -1,\ \ \mathrm{i}^3 = -\,\mathrm{i},\ \ \mathrm{i}^4 = 1,$ etc., we have

$$\begin{aligned}(x+\mathrm{i}\,y)^n &= x^n + n\,x^{n-1}(\mathrm{i}\,y) + \binom{n}{2} x^{n-2}(\mathrm{i}\,y)^2 + \binom{n}{3} x^{n-3}(\mathrm{i}\,y)^3 + \ \cdots\ + (\mathrm{i}\,y)^n \\ &= x^n + \mathrm{i}\,n\,x^{n-1}\,y - \binom{n}{2} x^{n-2}\,y^2 - \mathrm{i}\binom{n}{3} x^{n-3}\,y^3 + \ \cdots\ .\end{aligned}$$

Separating the real and imaginary terms, we obtain the explicit formulas

$$\begin{aligned}\operatorname{Re}\,(x+\mathrm{i}\,y)^n &= x^n - \binom{n}{2} x^{n-2}\,y^2 + \binom{n}{4} x^{n-4}\,y^4 + \ \cdots\ , \\ \operatorname{Im}\,(x+\mathrm{i}\,y)^n &= n\,x^{n-1}\,y - \binom{n}{3} x^{n-3}\,y^3 + \binom{n}{5} x^{n-5}\,y^5 + \ \cdots\ ,\end{aligned} \tag{7.79}$$

for the two independent harmonic polynomials of degree n. The first few of these polynomials were described in Example 7.36. In fact, it can be proved that the most general solution to the Laplace equation can be written as a convergent infinite series in the basic harmonic polynomials, cf. [**61**].

Exercises

7.4.40. Can you find a complex matrix A such that $\ker A \neq \{\mathbf{0}\}$ and the real and imaginary parts of every complex solution to $A\mathbf{u} = \mathbf{0}$ are also solutions?

7.4.41. Find the general real solution to the following homogeneous differential equations: (a) $u'' + 4u = 0$, (b) $u'' + 6u' + 10u = 0$, (c) $2u''' + 3u' - 5u = 0$, (d) $u'''' + u = 0$, (e) $u'''' + 13u'' + 36u = 0$, (f) $x^2 u'' - x u' + 3u = 0$, (g) $x^3 u''' + x^2 u'' + 3x u' - 8u = 0$.

7.4.42. The following functions are solutions to a real constant coefficient homogeneous scalar ordinary differential equation. (i) Determine the least possible order of the differential equation, and (ii) write down an appropriate differential equation. (a) $e^{-x}\sin 3x$, (b) $x\sin x$, (c) $1 + x e^{-x}\cos 2x$, (d) $\sin x + \cos 2x$, (e) $\sin x + x^2\cos x$.

7.4.43. Find the general solution to the following complex ordinary differential equations. Verify that, in these cases, the real and imaginary parts of a complex solution are *not* real solutions. (a) $u' + \mathrm{i}\,u = 0$, (b) $u'' - \mathrm{i}\,u' + (\mathrm{i} - 1)u = 0$, (c) $u'' - \mathrm{i}\,u = 0$.

7.4.44. (a) Write down the explicit formulas for the harmonic polynomials of degree 4 and check that they are indeed solutions to the Laplace equation. (b) Prove that every homogeneous polynomial solution of degree 4 is a linear combination of the two basic harmonic polynomials.

7.4.45. Find all complex exponential solutions $u(t,x) = e^{\omega t + k x}$ of the *beam equation* $\dfrac{\partial^2 u}{\partial t^2} = \dfrac{\partial^4 u}{\partial x^4}$. How many different real solutions can you produce?

♡ 7.4.46. (a) Show that, if $k \in \mathbb{R}$, then $u(t,x) = e^{-k^2 t + \mathrm{i}\,k x}$ is a complex solution to the *heat equation* $\dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2}$. (b) Use complex conjugation to write down another complex solution. (c) Find two independent real solutions to the heat equation. (d) Can k be complex? If so, what real solutions are produced? (e) Which of your solutions decay to zero as $t \to \infty$? (f) Can you solve the exercise assuming $k \in \mathbb{C}\setminus\mathbb{R}$ is not real?

7.4.47. Show that the free space *Schrödinger equation* $\mathrm{i}\,\dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2}$ is not a real linear system by constructing a complex quadratic polynomial solution and verifying that its real and imaginary parts are not solutions.

7.4.48. Which of the following sets of vectors span conjugated subspaces of $\mathbb{C}^3$?

(a) $\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$; (b) $\begin{pmatrix} 1 \\ -\mathrm{i} \\ 2\,\mathrm{i} \end{pmatrix}$; (c) $\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$; (d) $\begin{pmatrix} 1 \\ 0 \\ \mathrm{i} \end{pmatrix}, \begin{pmatrix} \mathrm{i} \\ 1 \\ 0 \end{pmatrix}$; (e) $\begin{pmatrix} \mathrm{i} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -\mathrm{i} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \mathrm{i} \end{pmatrix}$.

◇ 7.4.49. Prove that the real and imaginary parts of a general element of a conjugated vector space, as defined by (7.76), are both real elements.

◇ 7.4.50. Prove that a subspace $V \subset \mathbb{C}^n$ is conjugated if and only if it admits a basis all of whose elements are real.

◇ 7.4.51. Prove that if $L[\mathbf{u}] = \mathbf{f}$ is a real inhomogeneous linear system with real right-hand side $\mathbf{f}$, and $\mathbf{u} = \mathbf{v} + \mathrm{i}\,\mathbf{w}$ is a complex solution, then its real part $\mathbf{v}$ is a solution to the system, $L[\mathbf{v}] = \mathbf{f}$, while its imaginary part $\mathbf{w}$ solves the homogeneous system $L[\mathbf{w}] = \mathbf{0}$.

♢ 7.4.52. Prove that a linear function $L: \mathbb{C}^n \to \mathbb{C}^m$ is real if and only if $L[\mathbf{u}] = A\mathbf{u}$, where A is a real $m \times n$ matrix.

♢ 7.4.53. Let $\mathbf{u} = \mathbf{x} + \mathrm{i}\,\mathbf{y}$ be a complex solution to a real linear system. Under what conditions are its real and imaginary parts $\mathbf{x}, \mathbf{y}$ linearly independent real solutions?

7.5 Adjoints, Positive Definite Operators, and Minimization Principles

Sections 2.5 and 4.4 revealed the importance of the adjoint system $A^T\mathbf{y} = \mathbf{f}$ in the analysis of systems of linear algebraic equations $A\mathbf{x} = \mathbf{b}$. Two of the four fundamental matrix subspaces are based on the transposed matrix. While the $m \times n$ matrix A defines a linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$, its transpose, A^T, has size $n \times m$ and hence characterizes a linear function in the *reverse* direction, from $\mathbb{R}^m$ to $\mathbb{R}^n$.

As with most basic concepts for linear algebraic systems, the adjoint system and transpose operation on the coefficient matrix are the prototypes of a more general construction that is valid for general linear functions. However, it is not immediately obvious how to "transpose" a more general linear operator $L[u]$, e.g., a differential operator acting on function space. In this section, we shall introduce the abstract concept of the *adjoint* of a linear function that generalizes the transpose operation on matrices. This will be followed by a general characterization of positive definite linear operators and the characterization of the solutions to the associated linear systems via minimization principles. Unfortunately, we will not have sufficient analytical tools to develop most of the interesting examples, and instead refer the interested reader to [**61**, **79**].

The adjoint (and transpose) rely on an inner product structure on both the domain and codomain spaces. For simplicity, we restrict our attention to real inner product spaces, leaving the complex version to the interested reader. Thus, we begin with a linear function $L: U \to V$ that maps an inner product space U to a second inner product space V. We distinguish the inner products on U and V (which may be different even when U and V are the same vector space) by using a single angle bracket

$$\langle\, \mathbf{u} \,,\widetilde{\mathbf{u}}\,\rangle \qquad \text{to denote the inner product between} \qquad \mathbf{u}, \widetilde{\mathbf{u}} \in U,$$

and a double angle bracket

$$\langle\!\langle\, \mathbf{v} \,,\widetilde{\mathbf{v}}\,\rangle\!\rangle \qquad \text{to denote the inner product between} \qquad \mathbf{v}, \widetilde{\mathbf{v}} \in V.$$

Once inner products on both the domain and codomain are prescribed, the abstract definition of the adjoint of a linear function can be formulated.

Definition 7.53. Let U, V be inner product spaces, and let $L: U \to V$ be a linear function. The *adjoint* of L is the function† $L^*: V \to U$ that satisfies

$$\langle\!\langle\, L[\mathbf{u}] \,, \mathbf{v}\,\rangle\!\rangle = \langle\, \mathbf{u} \,, L^*[\mathbf{v}]\,\rangle \qquad \text{for all} \qquad \mathbf{u} \in U, \quad \mathbf{v} \in V. \tag{7.80}$$

† The notation L^* unfortunately coincides with that of the dual linear function introduced in Exercise 7.2.30. These clashing notations are both well established in the literature, although occasionally a prime, as in V', L', is used for dual spaces, maps, etc. However, it is possible to reconcile the two notations in a natural manner; see Exercise 7.5.10. In this book, the dual notation appears only in these few exercises.

Note that the adjoint function goes in the *opposite* direction to L, just like the transposed matrix. Also, the left-hand side of equation (7.80) indicates the inner product on V, while the right-hand side is the inner product on U — which is where the respective vectors live. In infinite-dimensional situations, the adjoint may not exist. But if it does, then it is uniquely determined by (7.80); see Exercise 7.5.7.

Remark. Technically, (7.80) serves to define the "formal adjoint" of the linear operator L. For the infinite-dimensional function spaces arising in analysis, a true adjoint must satisfy certain additional analytical requirements, [**50**, **67**]. However, for pedagogical reasons, it is better to suppress such advanced analytical complications in this introductory treatment.

Lemma 7.54. The adjoint of a linear function is a linear function.

Proof: Given $\mathbf{u} \in U$, $\mathbf{v}, \mathbf{w} \in V$, and scalars $c, d \in \mathbb{R}$, using the defining property of the adjoint and the bilinearity of the two inner products produces

$$\begin{aligned}\langle\, \mathbf{u}\,, L^*[\,c\,\mathbf{v} + d\,\mathbf{w}\,]\,\rangle &= \langle\!\langle\, L[\mathbf{u}]\,, c\,\mathbf{v} + d\,\mathbf{w}\,\rangle\!\rangle = c\,\langle\!\langle\, L[\mathbf{u}]\,, \mathbf{v}\,\rangle\!\rangle + d\,\langle\!\langle\, L[\mathbf{u}]\,, \mathbf{w}\,\rangle\!\rangle \\ &= c\,\langle\, \mathbf{u}\,, L^*[\,\mathbf{v}\,]\,\rangle + d\,\langle\, \mathbf{u}\,, L^*[\,\mathbf{w}\,]\,\rangle = \langle\, \mathbf{u}\,, c\,L^*[\,\mathbf{v}\,] + d\,L^*[\,\mathbf{w}\,]\,\rangle.\end{aligned}$$

Since this holds for all $\mathbf{u} \in U$, we must have

$$L^*[\,c\,\mathbf{v} + d\,\mathbf{w}\,] = c\,L^*[\,\mathbf{v}\,] + d\,L^*[\,\mathbf{w}\,], \qquad \text{thereby proving linearity.} \qquad Q.E.D.$$

Example 7.55. Let us first show how the defining equation (7.80) for the adjoint leads directly to the transpose of a matrix. Let $L\colon \mathbb{R}^n \to \mathbb{R}^m$ be the linear function $L[\mathbf{v}] = A\,\mathbf{v}$ defined by multiplication by the $m \times n$ matrix A. Then $L^*\colon \mathbb{R}^m \to \mathbb{R}^n$ is linear, and so is represented by matrix multiplication, $L^*[\mathbf{v}] = A^*\,\mathbf{v}$, by an $n \times m$ matrix A^*. We impose the ordinary Euclidean dot products

$$\langle\, \mathbf{u}\,, \widetilde{\mathbf{u}}\,\rangle = \mathbf{u} \cdot \widetilde{\mathbf{u}} = \mathbf{u}^T\,\widetilde{\mathbf{u}}, \qquad \mathbf{u}, \widetilde{\mathbf{u}} \in \mathbb{R}^n, \qquad \langle\!\langle\, \mathbf{v}\,, \widetilde{\mathbf{v}}\,\rangle\!\rangle = \mathbf{v} \cdot \widetilde{\mathbf{v}} = \mathbf{v}^T\,\widetilde{\mathbf{v}}, \qquad \mathbf{v}, \widetilde{\mathbf{v}} \in \mathbb{R}^m,$$

as our inner products on both $\mathbb{R}^n$ and $\mathbb{R}^m$. Evaluation of both sides of the adjoint identity (7.80) yields

$$\begin{aligned}\langle\!\langle\, L[\mathbf{u}]\,, \mathbf{v}\,\rangle\!\rangle &= \langle\!\langle\, A\,\mathbf{u}\,, \mathbf{v}\,\rangle\!\rangle = (A\,\mathbf{u})^T\mathbf{v} = \mathbf{u}^T A^T\,\mathbf{v}, \\ \langle\, \mathbf{u}\,, L^*[\mathbf{v}]\,\rangle &= \langle\, \mathbf{u}\,, A^*\,\mathbf{v}\,\rangle = \mathbf{u}^T A^*\,\mathbf{v}.\end{aligned} \tag{7.81}$$

Since these expressions must agree for all $\mathbf{u}, \mathbf{v}$, the matrix A^* representing L^* is equal to the transposed matrix A^T, as justified in Exercise 1.6.13. We conclude that *the adjoint of a matrix with respect to the Euclidean dot product is its transpose*: $A^* = A^T$.

Remark. See Exercise 7.2.30 for another interpretation of the transpose in terms of dual vector spaces. Again, keep in mind that the * notation has a different meaning there.

Example 7.56. Let us now adopt different, weighted inner products on the domain and codomain for the linear map $L\colon \mathbb{R}^n \to \mathbb{R}^m$ given by $L[\mathbf{u}] = A\,\mathbf{v}$. Suppose that

(i) the inner product on the domain space $\mathbb{R}^n$ is given by $\langle\, \mathbf{u}\,, \widetilde{\mathbf{u}}\,\rangle = \mathbf{u}^T M\,\widetilde{\mathbf{u}}$, while

(ii) the inner product on the codomain $\mathbb{R}^m$ is given by $\langle\!\langle\, \mathbf{v}\,, \widetilde{\mathbf{v}}\,\rangle\!\rangle = \mathbf{v}^T C\,\widetilde{\mathbf{v}}$.

Here M and C are positive definite matrices of respective sizes $n \times n$ and $m \times m$. Then, in place of (7.81), we have

$$\langle\!\langle\, A\,\mathbf{u}\,, \mathbf{v}\,\rangle\!\rangle = (A\,\mathbf{u})^T C\,\mathbf{v} = \mathbf{u}^T A^T C\,\mathbf{v}, \qquad \langle\, \mathbf{u}\,, A^*\,\mathbf{v}\,\rangle = \mathbf{u}^T M A^*\,\mathbf{v}.$$

Equating these expressions, we deduce that $A^T C = M A^*$. Therefore, the *weighted adjoint* of the matrix A is given by the more complicated formula

$$A^* = M^{-1} A^T C. \tag{7.82}$$

In mechanical applications, M plays the role of the mass matrix, and explicitly appears in the dynamical systems to be studied in Chapter 10. In particular, suppose A is square, defining a linear transformation $L: \mathbb{R}^n \to \mathbb{R}^n$. If we adopt the same inner product $\langle \mathbf{v}, \widetilde{\mathbf{v}} \rangle = \mathbf{v}^T C \widetilde{\mathbf{v}}$ on both the domain and codomain spaces $\mathbb{R}^n$, then its adjoint matrix $A^* = C^{-1} A^T C$ is similar to its transpose.

Everything that we learned about transposes can be reinterpreted in the more general language of adjoints. First, applying the adjoint operation twice returns you to where you began; this is an immediate consequence of the defining equation (7.80).

Proposition 7.57. The adjoint of the adjoint of L is just $L = (L^*)^*$.

The next result generalizes the fact, (1.55), that the transpose of the product of two matrices is the product of the transposes, in the reverse order.

Proposition 7.58. Let U, V, W be inner product spaces. If $L: U \to V$ and $M: V \to W$ have respective adjoints $L^*: V \to U$ and $M^*: W \to V$, then the composite linear function $M \circ L: U \to W$ has adjoint $(M \circ L)^* = L^* \circ M^*$, which maps W to U.

Proof: Let $\langle \mathbf{u}, \widetilde{\mathbf{u}} \rangle$, $\langle\!\langle \mathbf{v}, \widetilde{\mathbf{v}} \rangle\!\rangle$, $\langle\!\langle\!\langle \mathbf{w}, \widetilde{\mathbf{w}} \rangle\!\rangle\!\rangle$, denote, respectively, the inner products on U, V, W. For $\mathbf{u} \in U$, $\mathbf{w} \in W$, we compute using the definition (7.80) repeatedly:

$$\begin{aligned} \langle \mathbf{u}, (M \circ L)^*[\mathbf{w}] \rangle &= \langle\!\langle\!\langle M \circ L[\mathbf{u}], \mathbf{w} \rangle\!\rangle\!\rangle = \langle\!\langle\!\langle M[L[\mathbf{u}]], \mathbf{w} \rangle\!\rangle\!\rangle \\ &= \langle\!\langle L[\mathbf{u}], M^*[\mathbf{w}] \rangle\!\rangle = \langle \mathbf{u}, L^*[M^*[\mathbf{w}]] \rangle = \langle \mathbf{u}, L^* \circ M^*[\mathbf{w}] \rangle. \end{aligned}$$

Since this holds for all $\mathbf{u}$ and $\mathbf{w}$, the identification follows. *Q.E.D.*

In this chapter, we have been able to actually compute adjoints in just the finite-dimensional situation, when the linear functions are given by matrix multiplication. For the more challenging case of adjoints of linear operators on function spaces, e.g., differential operators appearing in boundary value problems, the reader should consult [**61**].

Exercises

7.5.1. Choose one from the following list of inner products on $\mathbb{R}^2$. Then find the adjoint of $A = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}$ when your inner product is used on both its domain and codomain. (*a*) the Euclidean dot product; (*b*) the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = 2 v_1 w_1 + 3 v_2 w_2$; (*c*) the inner product $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T K \mathbf{w}$ defined by the positive definite matrix $K = \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix}$.

7.5.2. From the list in Exercise 7.5.1, choose different inner products on the domain and codomain, and then determine the adjoint of the matrix A.

7.5.3. Choose one from the following list of inner products on $\mathbb{R}^3$ for both the domain and codomain, and find the adjoint of $A = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 2 \end{pmatrix}$: (a) the Euclidean dot product;

(*b*) the weighted inner product $\langle\, \mathbf{v}\,,\mathbf{w}\,\rangle = v_1\, w_1 + 2\, v_2\, w_2 + 3\, v_3\, w_3$; (*c*) the inner product $\langle\, \mathbf{v}\,,\mathbf{w}\,\rangle = \mathbf{v}^T K\, \mathbf{w}$ defined by the positive definite matrix $K = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$.

7.5.4. From the list in Exercise 7.5.3, choose different inner products on the domain and codomain, and then compute the adjoint of the matrix A.

7.5.5. Choose an inner product on $\mathbb{R}^2$ from the list in Exercise 7.5.1, and an inner product on $\mathbb{R}^3$ from the list in Exercise 7.5.3, and then compute the adjoint of $A = \begin{pmatrix} 1 & 3 \\ 0 & 2 \\ -1 & 1 \end{pmatrix}$.

7.5.6. Let $\mathcal{P}^{(2)}$ be the space of quadratic polynomials equipped with the inner product $\langle\, p\,,q\,\rangle = \displaystyle\int_0^1 p(x)\, q(x)\, dx$. Find the adjoint of the derivative operator $D[\,p\,] = p'$ on $\mathcal{P}^{(2)}$.

♢ 7.5.7. Prove that, if it exists, the adjoint of a linear function is uniquely determined by (7.80).

♢ 7.5.8. Prove that (*a*) $(L + M)^* = L^* + M^*$, (*b*) $(c\,L)^* = c\,L^*$ for $c \in \mathbb{R}$, (*c*) $(L^*)^* = L$, (*d*) $(L^{-1})^* = (L^*)^{-1}$.

♢ 7.5.9. Let $L: U \to V$ be a linear function between inner product spaces. Prove that $\mathbf{u} \in \mathbb{R}^n$ solves the inhomogeneous linear system $L[\,\mathbf{u}\,] = \mathbf{f}$ if and only if

$$\langle\, \mathbf{u}\,, L^*[\,\mathbf{v}\,]\,\rangle = \langle\, \mathbf{f}\,,\mathbf{v}\,\rangle \qquad \text{for all} \qquad \mathbf{v} \in V. \tag{7.83}$$

Explain why Exercise 3.1.11 is a special case of this result. **Remark.** Equation (7.83) is known as the *weak formulation* of the linear system. It plays an essential role in the analysis of differential equations and their numerical approximations, [**61**].

♢ 7.5.10. Suppose V, W are finite-dimensional inner product spaces with dual space V^*, W^*. Let $L: V \to W$ be a linear function, and let $\widetilde{L}^*: W^* \to V^*$ denote the dual linear function, as in Exercise 7.2.30 (without the tilde), while $L^*: W \to V$ denotes its adjoint. (As noted above, the same notation denotes two mathematically different objects.) Prove that if we identify $V^* \simeq V$ and $W^* \simeq W$ using the linear isomorphism in Exercise 7.1.62, then the dual and adjoint functions are identified $\widetilde{L}^* \simeq L^*$, thus reconciling the unfortunate clash in notation. In particular, this includes the two possible interpretations of the transpose of a matrix.

Self-Adjoint and Positive Definite Linear Functions

Throughout this section U will be an inner product space. We will show how to generalize the notions of symmetric and positive definite matrices to linear operators on U in a natural fashion. First, we define the analogue of a symmetric matrix.

Definition 7.59. A linear function $J: U \to U$ is called *self-adjoint* if $J^* = J$. A self-adjoint linear function is *positive definite*, written $J > 0$, if

$$\langle\, \mathbf{u}\,, J[\,\mathbf{u}\,]\,\rangle > 0 \qquad \text{for all} \qquad \mathbf{0} \neq \mathbf{u} \in U. \tag{7.84}$$

In particular, if $J > 0$ then $\ker J = \{\mathbf{0}\}$. (Why?) Thus, a positive definite linear system $J[\,\mathbf{u}\,] = \mathbf{f}$ with $\mathbf{f} \in \operatorname{img} J$ must have a unique solution. The next result generalizes our basic observation that the Gram matrices $A^T A$ and $A^T C\, A$, cf. (3.62, 64), are symmetric and positive (semi-)definite.

Theorem 7.60. Let $L: U \to V$ be a linear map between inner product spaces with adjoint $L^*: V \to U$. Then the composite map $J = L^* \circ L: U \to U$ is self-adjoint. Moreover, J is positive definite if and only if $\ker L = \{\mathbf{0}\}$.

Proof: First, by Propositions 7.58 and 7.57,

$$J^* = (L^* \circ L)^* = L^* \circ (L^*)^* = L^* \circ L = J,$$

proving self-adjointness. Furthermore, for $\mathbf{u} \in U$, the inner product

$$\langle\, \mathbf{u}\, ,J[\,\mathbf{u}\,]\,\rangle = \langle\, \mathbf{u}\, ,L^*[\,L[\,\mathbf{u}\,]\,]\,\rangle = \langle\, L[\,\mathbf{u}\,]\, ,L[\,\mathbf{u}\,]\,\rangle = \|\, L[\,\mathbf{u}\,]\,\|^2 > 0$$

is strictly positive, provided that $L[\,\mathbf{u}\,] \neq \mathbf{0}$. Thus, if $\ker L = \{\mathbf{0}\}$, then the positivity condition (7.84) holds, and conversely. *Q.E.D.*

Let us specialize to the case of a linear function $L: \mathbb{R}^n \to \mathbb{R}^m$ that is represented by the $m \times n$ matrix A, so that $L[\,\mathbf{u}\,] = A\,\mathbf{u}$. When the Euclidean dot product is used on the two spaces, the adjoint L^* is represented by the transpose A^T, and hence the map $J = L^* \circ L$ has matrix representation $J[\,\mathbf{u}\,] = K\,\mathbf{u}$, where $K = A^T A$. Therefore, in this case Theorem 7.60 reduces to our earlier Proposition 3.36, governing the positive definiteness of the Gram matrix product $A^T A$. If we change the inner product on the codomain to $\langle\!\langle\, \mathbf{w}\, ,\widetilde{\mathbf{w}}\,\rangle\!\rangle = \mathbf{w}^T C\, \widetilde{\mathbf{w}}$ for some $C > 0$, then L^* is represented by $A^T C$, and hence $J = L^* \circ L$ has matrix form $K = A^T C\, A$, which is the general symmetric, positive definite Gram matrix constructed in (3.64) that underlay our development of the equations of equilibrium in Chapter 6.

Finally, if we further replace the dot product on the domain space $\mathbb{R}^n$ by the alternative inner product $\langle\, \mathbf{v}\, ,\widetilde{\mathbf{v}}\,\rangle = \mathbf{v}^T M\, \widetilde{\mathbf{v}}$ for $M > 0$, then, according to formula (7.82), the adjoint of L has matrix form

$$A^* = M^{-1} A^T C, \qquad \text{and therefore} \qquad K = A^* A = M^{-1} A^T C\, A \tag{7.85}$$

is a self-adjoint, positive (semi-)definite matrix with respect to the weighted inner product on $\mathbb{R}^n$ prescribed by the positive definite matrix M. In this case, the positive definite, self-adjoint operator J is *no longer represented by a symmetric matrix*. So, we did not quite tell the truth when we said we would allow only symmetric matrices to be positive definite — we really meant only self-adjoint matrices.

General self-adjoint matrices will be important in our discussion of the vibrations of mass–spring chains that have unequal masses. Extensions of these constructions to differential operators underlies the analysis of the boundary value problems of continuum mechanics, to be studied in [**61**].

Exercises

7.5.11. Show that the following linear transformations of $\mathbb{R}^2$ are self-adjoint with respect to the Euclidean dot product: (*a*) rotation through the angle $\theta = \pi$; (*b*) reflection about the line $y = x$. (*c*) The scaling map $S[\,\mathbf{x}\,] = 3\,\mathbf{x}$; (*d*) orthogonal projection onto the line $y = x$.

♢ 7.5.12. Let M be a positive definite matrix. Show that $A: \mathbb{R}^n \to \mathbb{R}^n$ is self-adjoint with respect to the inner product $\langle\, \mathbf{v}\, ,\mathbf{w}\,\rangle = \mathbf{v}^T M\, \mathbf{w}$ if and only if $M\, A$ is a symmetric matrix.

7.5.13. Prove that $A = \begin{pmatrix} 6 & 3 \\ 2 & 4 \end{pmatrix}$ is self-adjoint with respect to the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = 2 v_1 w_1 + 3 v_2 w_2$. *Hint*: Use the criterion in Exercise 7.5.12.

7.5.14. Consider the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + \frac{1}{2} v_2 w_2 + \frac{1}{3} v_3 w_3$ on $\mathbb{R}^3$. (*a*) What are the conditions on the entries of a 3×3 matrix A in order that it be self-adjoint? *Hint*: Use the criterion in Exercise 7.5.12. (*b*) Write down an example of a non-diagonal self-adjoint matrix.

7.5.15. Answer Exercise 7.5.14 for the inner product based on $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$.

7.5.16. *True or false*: The identity transformation is self-adjoint for an arbitrary inner product on the underlying vector space.

7.5.17. *True or false*: A diagonal matrix is self-adjoint for an arbitrary inner product on $\mathbb{R}^n$.

7.5.18. Suppose $L: U \to U$ has an adjoint $L^*: U \to U$. (*a*) Show that $L + L^*$ is self-adjoint. (*b*) Show that $L \circ L^*$ is self-adjoint.

♢ 7.5.19. Suppose $J, M: U \to U$ are self-adjoint linear functions on an inner product space U. (*a*) Prove that $\langle J[\mathbf{u}], \mathbf{u} \rangle = \langle M[\mathbf{u}], \mathbf{u} \rangle$ for all $\mathbf{u} \in U$ if and only if $J = M$. (*b*) Explain why this result is false if the self-adjointness hypothesis is dropped.

7.5.20. Prove that if $L: U \to U$ is an invertible linear transformation on an inner product space U, then the following three statements are equivalent: (*a*) $\langle L[\mathbf{u}], L[\mathbf{v}] \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$ for all $\mathbf{u}, \mathbf{v} \in U$. (*b*) $\| L[\mathbf{u}] \| = \| \mathbf{u} \|$ for all $\mathbf{u} \in U$. (*c*) $L^* = L^{-1}$. *Hint*: Use Exercise 7.5.19.

7.5.21. (*a*) Prove that the operation $M_a[u(x)] = a(x)\, u(x)$ of multiplication by a continuous function $a(x)$ defines a self-adjoint linear operator on the function space $\mathrm{C}^0[a, b]$ with respect to the L^2 inner product. (*b*) Is M_a also self-adjoint with respect to the weighted inner product $\langle\!\langle f, g \rangle\!\rangle = \int_a^b f(x)\, g(x)\, w(x)\, dx$?

♡ 7.5.22. A linear function $S: U \to U$ is called *skew-adjoint* if $S^* = -S$. (*a*) Prove that a skew-symmetric matrix is skew-adjoint with respect to the standard dot product on $\mathbb{R}^n$. (*b*) Under what conditions is $S[\mathbf{x}] = A\mathbf{x}$ skew-adjoint with respect to the inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T M \mathbf{y}$ on $\mathbb{R}^n$? (*c*) Let $L: U \to U$ have an adjoint L^*. Prove that $L - L^*$ is skew-adjoint. (*d*) Explain why every linear operator $L: U \to U$ that has an adjoint L^* can be written as the sum of a self-adjoint and a skew-adjoint operator.

♢ 7.5.23. (*a*) Let $L_1: U \to V_1$ and $L_2: U \to V_2$ be linear maps between inner product spaces, with V_1, V_2 not necessarily the same. Let $J_1 = L_1^* \circ L_1$, $J_2 = L_2^* \circ L_2$. Show that the sum $J = J_1 + J_2$ can be written as a self-adjoint combination $J = L^* \circ L$ for some linear operator L. *Hint*: See Exercise 3.4.35 for the matrix case.

Minimization

In Chapter 5, we learned how the solution to a linear algebraic system $K\mathbf{u} = \mathbf{f}$ with positive definite coefficient matrix K can be characterized as the unique minimizer for the quadratic function $p(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T K \mathbf{u} - \mathbf{u}^T \mathbf{f}$. There is an analogous minimization principle that characterizes the solutions to linear systems defined by positive definite linear operators. This general result is of tremendous importance in analysis of boundary value problems for differential equations, for both physical and mathematical reasons, and also inspires the finite element numerical solution algorithm, [**61**].

We restrict our attention to real linear functions on real vector spaces in this section.

Theorem 7.61. Let $J: U \to U$ be a positive definite linear function on a real inner product space U. If $\mathbf{f} \in \operatorname{img} J$, then the quadratic function

$$p(\mathbf{u}) = \tfrac{1}{2} \langle \mathbf{u}, J[\mathbf{u}] \rangle - \langle \mathbf{u}, \mathbf{f} \rangle \tag{7.86}$$

has a unique minimizer $\mathbf{u} = \mathbf{u}^\star$, which is the solution to the linear system $J[\mathbf{u}] = \mathbf{f}$.

Proof: The proof mimics that of its matrix counterpart in Theorem 5.2. Our assumption that $\mathbf{f} \in \operatorname{img} J$ implies that there is a $\mathbf{u}^\star \in U$ such that $J[\mathbf{u}^\star] = \mathbf{f}$. Thus, we can write

$$p(\mathbf{u}) = \tfrac{1}{2} \langle \mathbf{u}, J[\mathbf{u}] \rangle - \langle \mathbf{u}, J[\mathbf{u}^\star] \rangle = \tfrac{1}{2} \langle \mathbf{u} - \mathbf{u}^\star, J[\mathbf{u} - \mathbf{u}^\star] \rangle - \tfrac{1}{2} \langle \mathbf{u}^\star, J[\mathbf{u}^\star] \rangle, \tag{7.87}$$

where we used linearity, along with the fact that J is self-adjoint to identify the terms $\langle \mathbf{u}, J[\mathbf{u}^\star] \rangle = \langle \mathbf{u}^\star, J[\mathbf{u}] \rangle$. Since $J > 0$, the first term on the right-hand side of (7.87) is always ≥ 0; moreover, it equals its minimal value 0 if and only if $\mathbf{u} = \mathbf{u}^\star$. On the other hand, the second term does not depend upon $\mathbf{u}$ at all, and hence is unaffected by variations in $\mathbf{u}$. Therefore, to minimize $p(\mathbf{u})$, we must make the first term as small as possible, which is accomplished by setting $\mathbf{u} = \mathbf{u}^\star$. *Q.E.D.*

Remark. For linear functions given by matrix multiplication, positive definiteness automatically implies invertibility, and so the linear system $K\mathbf{u} = \mathbf{f}$ has a solution for every right-hand side. This is not so immediate when J is a positive definite operator on an infinite-dimensional function space. Therefore, the existence of a solution or minimizer is a significant issue. And, in fact, many modern analytical existence results rely on the determination of suitable minimization principles. On the other hand, once existence is assured, uniqueness follows immediately from the positive definiteness of the operator J.

Theorem 7.62. Suppose $L: U \to V$ is a linear function between inner product spaces with $\ker L = \{\mathbf{0}\}$ and adjoint function $L^*: V \to U$. Let $J = L^* \circ L: U \to U$ be the associated positive definite linear function. If $\mathbf{f} \in \operatorname{img} J$, then the quadratic function

$$p(\mathbf{u}) = \tfrac{1}{2} \| L[\mathbf{u}] \|^2 - \langle \mathbf{u}, \mathbf{f} \rangle \tag{7.88}$$

has a unique minimizer $\mathbf{u}^\star$, which is the solution to the linear system $J[\mathbf{u}^\star] = \mathbf{f}$.

Proof: It suffices to note that the quadratic term in (7.88) can be written in the alternative form

$$\| L[\mathbf{u}] \|^2 = \langle\!\langle L[\mathbf{u}], L[\mathbf{u}] \rangle\!\rangle = \langle \mathbf{u}, L^*[L[\mathbf{u}]] \rangle = \langle \mathbf{u}, J[\mathbf{u}] \rangle.$$

Thus, (7.88) reduces to the quadratic function of the form (7.86) with $J = L^* \circ L$, and so Theorem 7.62 follows directly from Theorem 7.61. *Q.E.D.*

Warning. In (7.88), the first term $\| L[\mathbf{u}] \|^2$ is computed using the norm based on the inner product on V, while the second term $\langle \mathbf{u}, \mathbf{f} \rangle$ employs the inner product on U.

Example 7.63. For a general positive definite matrix (7.85), the quadratic function (7.88) is computed with respect to the alternative inner product $\langle \mathbf{u}, \widetilde{\mathbf{u}} \rangle = \mathbf{u}^T M \widetilde{\mathbf{u}}$, so

$$p(\mathbf{u}) = \tfrac{1}{2} \| A\mathbf{u} \|^2 - \langle \mathbf{u}, \mathbf{f} \rangle = \tfrac{1}{2} (A\mathbf{u})^T C A \mathbf{u} - \mathbf{u}^T M \mathbf{f} = \tfrac{1}{2} \mathbf{u}^T (A^T C A) \mathbf{u} - \mathbf{u}^T (M\mathbf{f}).$$

Theorem 7.62 tells us that the minimizer of the quadratic function is the solution to

$$A^T C A \mathbf{u} = M\mathbf{f}, \qquad \text{which we rewrite as} \qquad K\mathbf{u} = M^{-1} A^T C A \mathbf{u} = \mathbf{f}.$$

This conclusion also follows from our earlier finite-dimensional Minimization Theorem 5.2.

In [**61**, **79**], it is shown that the most important minimization principles that characterize solutions to the linear boundary value problems of physics and engineering all arise through this remarkably general mathematical construction.

Exercises

7.5.24. Find the minimum value of $p(\mathbf{u}) = \dfrac{1}{2}\,\mathbf{u}^T \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix} \mathbf{u} - \mathbf{u}^T \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ for $\mathbf{u} \in \mathbb{R}^2$.

7.5.25. Minimize the function $p(\mathbf{u}) = \dfrac{1}{2}\,\mathbf{u}^T \begin{pmatrix} 2 & -1 & 0 \\ -1 & 4 & -2 \\ 0 & -2 & 3 \end{pmatrix} \mathbf{u} - \mathbf{u}^T \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$ for $\mathbf{u} \in \mathbb{R}^3$.

7.5.26. Minimize $\|\,(\,2x - y, x + y\,)^T\,\|^2 - 6x$ over all x, y, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^2$.

7.5.27. Answer Exercise 7.5.26 for (*a*) the weighted norm $\|\,(\,x, y\,)^T\,\| = \sqrt{2x^2 + 3y^2}$; (*b*) the norm based on $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$; (*c*) the norm based on $\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$.

7.5.28. Let $L(x, y) = \begin{pmatrix} x - 2y \\ x + y \\ -x + 3y \end{pmatrix}$ and $\mathbf{f} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Minimize $p(\mathbf{x}) = \frac{1}{2}\,\|\,L[\mathbf{x}]\,\|^2 - \langle\,\mathbf{x}\,,\mathbf{f}\,\rangle$ using (*a*) the Euclidean inner products and norms on both $\mathbb{R}^2$ and $\mathbb{R}^3$; (*b*) the Euclidean inner product on $\mathbb{R}^2$ and the weighted norm $\|\,\mathbf{w}\,\| = \sqrt{w_1^2 + 2w_2^2 + 3w_3^2}$ on $\mathbb{R}^3$; (*c*) the inner product given by $\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ on $\mathbb{R}^2$ and the Euclidean norm on $\mathbb{R}^3$; (*d*) the inner product given by $\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ on $\mathbb{R}^2$ and the weighted norm $\|\,\mathbf{w}\,\| = \sqrt{w_1^2 + 2w_2^2 + 3w_3^2}$ on $\mathbb{R}^3$.

7.5.29. Find the minimum distance between the point $(\,1, 0, 0\,)^T$ and the plane $x + y - z = 0$ when distance is measured in (*a*) the Euclidean norm; (*b*) the weighted norm $\|\,\mathbf{w}\,\| = \sqrt{w_1^2 + 2w_2^2 + 3w_3^2}$; (*c*) the norm based on the positive definite matrix $\begin{pmatrix} 3 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 3 \end{pmatrix}$.

♢ 7.5.30. How would you modify the statement of Theorem 7.62 if $\ker L \neq \{\mathbf{0}\}$?