Chapter 3

Inner Products and Norms

The geometry of Euclidean space is founded on the familiar properties of length and angle. The abstract concept of a norm on a vector space formalizes the geometrical notion of the length of a vector. In Euclidean geometry, the angle between two vectors is specified by their dot product, which is itself formalized by the abstract concept of an inner product. Inner products and norms lie at the heart of linear (and nonlinear) analysis, in both finite-dimensional vector spaces and infinite-dimensional function spaces. A vector space equipped with an inner product and its associated norm is known as an inner product space. It is impossible to overemphasize their importance for theoretical developments, practical applications, and the design of numerical solution algorithms.

Mathematical analysis relies on the exploitation of inequalities. The most fundamental is the Cauchy–Schwarz inequality, which is valid in every inner product space. The more familiar triangle inequality for the associated norm is then derived as a simple consequence. Not every norm comes from an inner product, and, in such cases, the triangle inequality becomes part of the general definition. Both inequalities retain their validity in both finite-dimensional and infinite-dimensional vector spaces. Indeed, their abstract formulation exposes the key ideas behind the proof, avoiding all distracting particularities appearing in the explicit formulas.

The characterization of general inner products on Euclidean space will lead us to the noteworthy class of positive definite matrices. Positive definite matrices appear in a wide variety of applications, including minimization, least squares, data analysis and statistics, as well as, for example, mechanical systems, electrical circuits, and the differential equations describing both static and dynamical processes. The test for positive definiteness relies on Gaussian Elimination, and we can reinterpret the resulting matrix factorization as the algebraic process of completing the square for the associated quadratic form. In applications, positive definite matrices most often arise as Gram matrices, whose entries are formed by taking inner products between selected elements of an inner product space.

So far, we have focussed our attention on real vector spaces. Complex numbers, vectors, and functions also arise in numerous applications, and so, in the final section, we take the opportunity to formally introduce complex vector spaces. Most of the theory proceeds in direct analogy with the real version, but the notions of inner product and norm on complex vector spaces require some thought. Applications of complex vector spaces and their inner products are of particular significance in Fourier analysis, signal processing, and partial differential equations, [**61**], and they play an absolutely essential role in modern quantum mechanics, [**54**].

3.1 Inner Products

The most basic example of an inner product is the familiar *dot product*

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = \sum_{i=1}^{n} v_i w_i, \tag{3.1}$$

P. J. Olver, C. Shakiban, *Applied Linear Algebra*, Undergraduate Texts in Mathematics,
https://doi.org/10.1007/978-3-319-91041-3_3

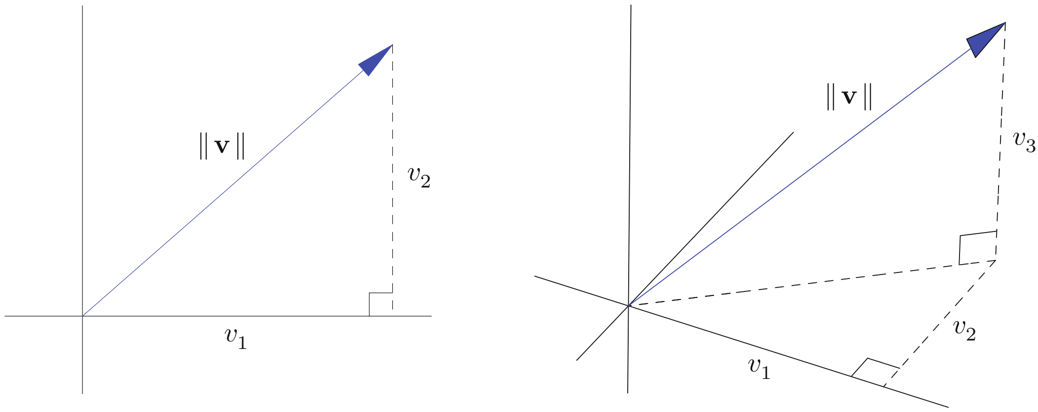

Figure 3.1. The Euclidean Norm in $\mathbb{R}^2$ and $\mathbb{R}^3$.

between (column) vectors $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T$, $\mathbf{w} = (w_1, w_2, \ldots, w_n)^T$, both lying in the Euclidean space $\mathbb{R}^n$. A key observation is that the dot product (3.1) is equal to the matrix product

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = (\, v_1 \quad v_2 \quad \ldots \quad v_n \,) \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \tag{3.2}$$

between the row vector $\mathbf{v}^T$ and the column vector $\mathbf{w}$.

The dot product is the cornerstone of Euclidean geometry. The key fact is that the dot product of a vector with itself,

$$\mathbf{v} \cdot \mathbf{v} = v_1^2 + v_2^2 + \cdots + v_n^2,$$

is the sum of the squares of its entries, and hence, by the classical Pythagorean Theorem, equals the square of its length; see Figure 3.1. Consequently, the *Euclidean norm* or *length* of a vector is found by taking the square root:

$$\| \mathbf{v} \| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}\,. \tag{3.3}$$

Note that every nonzero vector, $\mathbf{v} \neq \mathbf{0}$, has positive Euclidean norm, $\| \mathbf{v} \| > 0$, while only the zero vector has zero norm: $\| \mathbf{v} \| = 0$ if and only if $\mathbf{v} = \mathbf{0}$. The elementary properties of dot product and Euclidean norm serve to inspire the abstract definition of more general inner products.

Definition 3.1. An *inner product* on the real vector space V is a pairing that takes two vectors $\mathbf{v}, \mathbf{w} \in V$ and produces a real number $\langle \mathbf{v}, \mathbf{w} \rangle \in \mathbb{R}$. The inner product is required to satisfy the following three axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, and scalars $c, d \in \mathbb{R}$.

(*i*) *Bilinearity*:
$$\begin{aligned} \langle c\,\mathbf{u} + d\,\mathbf{v}, \mathbf{w} \rangle &= c \langle \mathbf{u}, \mathbf{w} \rangle + d \langle \mathbf{v}, \mathbf{w} \rangle, \\ \langle \mathbf{u}, c\,\mathbf{v} + d\,\mathbf{w} \rangle &= c \langle \mathbf{u}, \mathbf{v} \rangle + d \langle \mathbf{u}, \mathbf{w} \rangle. \end{aligned} \tag{3.4}$$

(*ii*) *Symmetry*:
$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle. \tag{3.5}$$

(*iii*) *Positivity*:
$$\langle \mathbf{v}, \mathbf{v} \rangle > 0 \quad \text{whenever} \quad \mathbf{v} \neq \mathbf{0}, \quad \text{while} \quad \langle \mathbf{0}, \mathbf{0} \rangle = 0. \tag{3.6}$$

A vector space equipped with an inner product is called an *inner product space*. As we shall see, a vector space can admit many different inner products. Verification of the inner product axioms for the Euclidean dot product is straightforward, and left as an exercise for the reader.

Given an inner product, the associated *norm* of a vector $\mathbf{v} \in V$ is defined as the positive square root of the inner product of the vector with itself:

$$\| \mathbf{v} \| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}\,. \tag{3.7}$$

The positivity axiom implies that $\| \mathbf{v} \| \geq 0$ is real and non-negative, and equals 0 if and only if $\mathbf{v} = \mathbf{0}$ is the zero vector.

Example 3.2. While certainly the most common inner product on $\mathbb{R}^2$, the dot product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2$$

is by no means the only possibility. A simple example is provided by the *weighted inner product*

$$\langle \mathbf{v}, \mathbf{w} \rangle = 2 v_1 w_1 + 5 v_2 w_2, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \qquad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{3.8}$$

Let us verify that this formula does indeed define an inner product. The symmetry axiom (3.5) is immediate. Moreover,

$$\begin{aligned} \langle c\,\mathbf{u} + d\,\mathbf{v}, \mathbf{w} \rangle &= 2(c u_1 + d v_1) w_1 + 5(c u_2 + d v_2) w_2 \\ &= c(2 u_1 w_1 + 5 u_2 w_2) + d(2 v_1 w_1 + 5 v_2 w_2) = c \langle \mathbf{u}, \mathbf{w} \rangle + d \langle \mathbf{v}, \mathbf{w} \rangle, \end{aligned}$$

which verifies the first bilinearity condition; the second follows by a very similar computation. (Or, one can use the symmetry axiom to deduce the second bilinearity identity from the first; see Exercise 3.1.9.) Moreover, $\langle \mathbf{0}, \mathbf{0} \rangle = 0$, while

$$\langle \mathbf{v}, \mathbf{v} \rangle = 2 v_1^2 + 5 v_2^2 > 0 \qquad \text{whenever} \qquad \mathbf{v} \neq \mathbf{0},$$

since at least one of the summands is strictly positive. This establishes (3.8) as a legitimate inner product on $\mathbb{R}^2$. The associated *weighted norm* $\| \mathbf{v} \| = \sqrt{2 v_1^2 + 5 v_2^2}$ defines an alternative, "non-Pythagorean" notion of length of vectors and distance between points in the plane.

A less evident example of an inner product on $\mathbb{R}^2$ is provided by the expression

$$\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 - v_1 w_2 - v_2 w_1 + 4 v_2 w_2. \tag{3.9}$$

Bilinearity is verified in the same manner as before, and symmetry is immediate. Positivity is ensured by noticing that the expression

$$\langle \mathbf{v}, \mathbf{v} \rangle = v_1^2 - 2 v_1 v_2 + 4 v_2^2 = (v_1 - v_2)^2 + 3 v_2^2 \geq 0$$

is always non-negative, and, moreover, is equal to zero if and only if $v_1 - v_2 = 0$, $v_2 = 0$, i.e., only when $v_1 = v_2 = 0$ and so $\mathbf{v} = \mathbf{0}$. We conclude that (3.9) defines yet another inner product on $\mathbb{R}^2$, with associated norm

$$\| \mathbf{v} \| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{v_1^2 - 2 v_1 v_2 + 4 v_2^2}\,.$$

The second example (3.8) is a particular case of a general class of inner products.

Example 3.3. Let $c_1, \ldots, c_n > 0$ be a set of *positive* numbers. The corresponding *weighted inner product* and *weighted norm* on $\mathbb{R}^n$ are defined by

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} c_i v_i w_i, \qquad \| \mathbf{v} \| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\sum_{i=1}^{n} c_i v_i^2}\,. \tag{3.10}$$

The numbers $c_i > 0$ are the *weights*. Observe that the larger the weight c_i, the more the i^{th} coordinate of $\mathbf{v}$ contributes to the norm. Weighted norms are particularly relevant in statistics and data fitting, [**43**, **87**], when one wants to emphasize the importance of certain measurements and de-emphasize others; this is done by assigning appropriate weights to the different components of the data vector $\mathbf{v}$. Section 5.4, on least squares approximation methods, will contain further details.

Exercises

3.1.1. Prove that the formula $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 - v_1 w_2 - v_2 w_1 + b v_2 w_2$ defines an inner product on $\mathbb{R}^2$ if and only if $b > 1$.

3.1.2. Which of the following formulas for $\langle \mathbf{v}, \mathbf{w} \rangle$ define inner products on $\mathbb{R}^2$?
(*a*) $2 v_1 w_1 + 3 v_2 w_2$, (*b*) $v_1 w_2 + v_2 w_1$, (*c*) $(v_1 + v_2)(w_1 + w_2)$, (*d*) $v_1^2 w_1^2 + v_2^2 w_2^2$,
(*e*) $\sqrt{v_1^2 + v_2^2}\,\sqrt{w_1^2 + w_2^2}$, (*f*) $2 v_1 w_1 + (v_1 - v_2)(w_1 - w_2)$,
(*g*) $4 v_1 w_1 - 2 v_1 w_2 - 2 v_2 w_1 + 4 v_2 w_2$.

3.1.3. Show that $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + v_1 w_2 + v_2 w_1 + v_2 w_2$ does *not* define an inner product on $\mathbb{R}^2$.

3.1.4. Prove that each of the following formulas for $\langle \mathbf{v}, \mathbf{w} \rangle$ defines an inner product on $\mathbb{R}^3$. Verify all the inner product axioms in careful detail:
(*a*) $v_1 w_1 + 2 v_2 w_2 + 3 v_3 w_3$, (*b*) $4 v_1 w_1 + 2 v_1 w_2 + 2 v_2 w_1 + 4 v_2 w_2 + v_3 w_3$,
(*c*) $2 v_1 w_1 - 2 v_1 w_2 - 2 v_2 w_1 + 3 v_2 w_2 - v_2 w_3 - v_3 w_2 + 2 v_3 w_3$.

3.1.5. The *unit circle* for an inner product on $\mathbb{R}^2$ is defined as the set of all vectors of unit length: $\| \mathbf{v} \| = 1$. Graph the unit circles for (*a*) the Euclidean inner product, (*b*) the weighted inner product (3.8), (*c*) the non-standard inner product (3.9). (*d*) Prove that cases (*b*) and (*c*) are, in fact, both ellipses.

♢ 3.1.6. (*a*) Explain why the formula for the Euclidean norm in $\mathbb{R}^2$ follows from the Pythagorean Theorem. (*b*) How do you use the Pythagorean Theorem to justify the formula for the Euclidean norm in $\mathbb{R}^3$? *Hint*: Look at Figure 3.1.

♢ 3.1.7. Prove that the norm on an inner product space satisfies $\| c \mathbf{v} \| = | c | \, \| \mathbf{v} \|$ for every scalar c and vector $\mathbf{v}$.

3.1.8. Prove that $\langle a \mathbf{v} + b \mathbf{w}, c \mathbf{v} + d \mathbf{w} \rangle = a c \| \mathbf{v} \|^2 + (a d + b c) \langle \mathbf{v}, \mathbf{w} \rangle + b d \| \mathbf{w} \|^2$.

♢ 3.1.9. Prove that the second bilinearity formula (3.4) is a consequence of the first and the other two inner product axioms.

♢ 3.1.10. Let V be an inner product space. (*a*) Prove that $\langle \mathbf{x}, \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in V$ if and only if $\mathbf{x} = \mathbf{0}$. (*b*) Prove that $\langle \mathbf{x}, \mathbf{v} \rangle = \langle \mathbf{y}, \mathbf{v} \rangle$ for all $\mathbf{v} \in V$ if and only if $\mathbf{x} = \mathbf{y}$. (*c*) Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis for V. Prove that $\langle \mathbf{x}, \mathbf{v}_i \rangle = \langle \mathbf{y}, \mathbf{v}_i \rangle$, $i = 1, \ldots, n$, if and only if $\mathbf{x} = \mathbf{y}$.

♢ 3.1.11. Prove that $\mathbf{x} \in \mathbb{R}^n$ solves the linear system $A \mathbf{x} = \mathbf{b}$ if and only if

$$\mathbf{x}^T A^T \mathbf{v} = \mathbf{b}^T \mathbf{v} \qquad \text{for all} \qquad \mathbf{v} \in \mathbb{R}^m.$$

The latter is known as the *weak formulation* of the linear system, and its generalizations are of great importance in the study of differential equations and numerical analysis, [**61**].

♢ 3.1.12. (*a*) Prove the identity

$$\langle \mathbf{u}, \mathbf{v} \rangle = \tfrac{1}{4} \left(\| \mathbf{u} + \mathbf{v} \|^2 - \| \mathbf{u} - \mathbf{v} \|^2 \right), \tag{3.11}$$

which allows one to reconstruct an inner product from its norm. (*b*) Use (3.11) to find the inner product on $\mathbb{R}^2$ corresponding to the norm $\| \mathbf{v} \| = \sqrt{v_1^2 - 3 v_1 v_2 + 5 v_2^2}$.

3.1.13. (a) Show that, for all vectors $\mathbf{x}$ and $\mathbf{y}$ in an inner product space,

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\right).$$

(b) Interpret this result pictorially for vectors in $\mathbb{R}^2$ under the Euclidean norm.

3.1.14. Suppose $\mathbf{u}, \mathbf{v}$ satisfy $\|\mathbf{u}\| = 3$, $\|\mathbf{u}+\mathbf{v}\| = 4$, and $\|\mathbf{u}-\mathbf{v}\| = 6$. What must $\|\mathbf{v}\|$ equal? Does your answer depend upon which norm is being used?

3.1.15. Let A be any $n \times n$ matrix. Prove that the dot product identity $\mathbf{v}\cdot(A\mathbf{w}) = (A^T\mathbf{v})\cdot\mathbf{w}$ is valid for all vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.

♢ 3.1.16. Prove that $A = A^T$ is a symmetric $n \times n$ matrix if and only if $(A\mathbf{v})\cdot\mathbf{w} = \mathbf{v}\cdot(A\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.

3.1.17. Prove that $\langle A, B\rangle = \operatorname{tr}(A^T B)$ defines an inner product on the vector space $\mathcal{M}_{n\times n}$ of real $n \times n$ matrices.

3.1.18. Suppose $\langle \mathbf{v}, \mathbf{w}\rangle$ defines an inner product on a vector space V. Explain why it also defines an inner product on every subspace $W \subset V$.

3.1.19. Prove that if $\langle \mathbf{v}, \mathbf{w}\rangle$ and $\langle\!\langle \mathbf{v}, \mathbf{w}\rangle\!\rangle$ are two inner products on the same vector space V, then their sum $\langle\!\langle\!\langle \mathbf{v}, \mathbf{w}\rangle\!\rangle\!\rangle = \langle \mathbf{v}, \mathbf{w}\rangle + \langle\!\langle \mathbf{v}, \mathbf{w}\rangle\!\rangle$ defines an inner product on V.

♢ 3.1.20. Let V and W be inner product spaces with respective inner products $\langle \mathbf{v}, \widetilde{\mathbf{v}}\rangle$ and $\langle\!\langle \mathbf{w}, \widetilde{\mathbf{w}}\rangle\!\rangle$. Show that $\langle\!\langle\!\langle (\mathbf{v},\mathbf{w}), (\widetilde{\mathbf{v}},\widetilde{\mathbf{w}})\rangle\!\rangle\!\rangle = \langle \mathbf{v}, \widetilde{\mathbf{v}}\rangle + \langle\!\langle \mathbf{w}, \widetilde{\mathbf{w}}\rangle\!\rangle$ for $\mathbf{v}, \widetilde{\mathbf{v}} \in V$, $\mathbf{w}, \widetilde{\mathbf{w}} \in W$, defines an inner product on their Cartesian product $V \times W$.

Inner Products on Function Spaces

Inner products and norms on function spaces lie at the foundation of modern analysis and its applications, particularly Fourier analysis, boundary value problems, ordinary and partial differential equations, and numerical analysis. Let us introduce the most important examples.

Example 3.4. Let $[a,b] \subset \mathbb{R}$ be a bounded closed interval. Consider the vector space $\mathrm{C}^0[a,b]$ consisting of all continuous scalar functions f defined on the interval $[a,b]$. The integral of the product of two continuous functions,

$$\langle f, g\rangle = \int_a^b f(x)\,g(x)\,dx, \tag{3.12}$$

defines an inner product on the vector space $\mathrm{C}^0[a,b]$, as we shall prove below. The associated norm is, according to the basic definition (3.7),

$$\| f \| = \sqrt{\int_a^b f(x)^2\,dx}\ , \tag{3.13}$$

and is known as the L^2 *norm* of the function f over the interval $[a,b]$. The L^2 inner product and norm of functions can be viewed as the infinite-dimensional function space versions of the dot product and Euclidean norm of vectors in $\mathbb{R}^n$. The reason for the name L^2 will become clearer later on.

For example, if we take $[a,b] = \left[0, \frac{1}{2}\pi\right]$, then the L^2 inner product between $f(x) = \sin x$ and $g(x) = \cos x$ is equal to

$$\langle \sin x, \cos x\rangle = \int_0^{\pi/2} \sin x\,\cos x\,dx = \frac{1}{2}\sin^2 x\,\Big|_{x=0}^{\pi/2} = \frac{1}{2}.$$

Similarly, the norm of the function $\sin x$ is

$$\| \sin x \| = \sqrt{\int_0^{\pi/2} (\sin x)^2 \, dx} = \sqrt{\frac{\pi}{4}} \; .$$

One must always be careful when evaluating function norms. For example, the constant function $c(x) \equiv 1$ has norm

$$\| 1 \| = \sqrt{\int_0^{\pi/2} 1^2 \, dx} = \sqrt{\frac{\pi}{2}} \; ,$$

not 1 as you might have expected. We also note that the value of the norm depends upon which interval the integral is taken over. For instance, on the longer interval $[0, \pi]$,

$$\| 1 \| = \sqrt{\int_0^{\pi} 1^2 \, dx} = \sqrt{\pi} \; .$$

Thus, when dealing with the L^2 inner product or norm, one must always be careful to specify the function space, or, equivalently, the interval on which it is being evaluated.

Let us prove that formula (3.12) does, indeed, define an inner product. First, we need to check that $\langle f , g \rangle$ is well defined. This follows because the product $f(x)\,g(x)$ of two continuous functions is also continuous, and hence its integral over a bounded interval is defined and finite. The symmetry requirement is immediate:

$$\langle f , g \rangle = \int_a^b f(x)\, g(x) \, dx = \int_a^b g(x)\, f(x)\, dx = \langle g , f \rangle,$$

because multiplication of functions is commutative. The first bilinearity axiom

$$\langle c\, f + d\, g , h \rangle = c \langle f , h \rangle + d \langle g , h \rangle$$

amounts to the following elementary integral identity

$$\int_a^b \big[\, c\, f(x) + d\, g(x) \,\big]\, h(x)\, dx = c \int_a^b f(x)\, h(x)\, dx + d \int_a^b g(x)\, h(x)\, dx,$$

valid for arbitrary continuous functions f, g, h and scalars (constants) c, d. The second bilinearity axiom is proved similarly; alternatively, one can use symmetry to deduce it from the first as in Exercise 3.1.9. Finally, positivity requires that

$$\| f \|^2 = \langle f , f \rangle = \int_a^b f(x)^2 \, dx \geq 0.$$

This is clear because $f(x)^2 \geq 0$, and the integral of a nonnegative function is nonnegative. Moreover, since the function $f(x)^2$ is continuous and nonnegative, its integral will vanish, $\displaystyle\int_a^b f(x)^2 \, dx = 0$, if and only if $f(x) \equiv 0$ is the zero function, cf. Exercise 3.1.29. This completes the proof that (3.12) defines a bona fide inner product on the space $\mathrm{C}^0[a, b]$.

Remark. The L^2 inner product formula can also be applied to more general functions, but we have restricted our attention to continuous functions in order to avoid certain technical complications. The most general function space admitting this inner product is known as *Hilbert space*, which forms the basis of much of modern analysis, function theory, and Fourier analysis, as well as providing the theoretical setting for all of quantum mechanics, [**54**]. Unfortunately, we cannot provide the mathematical details of the Hilbert space construction, since it requires that you be familiar with measure theory and the Lebesgue integral. See [**61**] for a basic introduction and [**19**, **68**, **77**] for the fully rigorous theory.

Warning. One needs to be extremely careful when trying to extend the L^2 inner product to other spaces of functions. Indeed, there are *nonzero* discontinuous functions with *zero* "L^2 norm". For example, the function

$$f(x) = \begin{cases} 1, & x = 0, \\ 0, & \text{otherwise}, \end{cases} \qquad \text{satisfies} \qquad \| f \|^2 = \int_{-1}^{1} f(x)^2 \, dx = 0, \tag{3.14}$$

because every function that is zero except at finitely many (or even countably many) points has zero integral.

The L^2 inner product is but one of a vast number of possible inner products on function spaces. For example, one can also define weighted inner products on the space $C^0[a,b]$. The weighting along the interval is specified by a (continuous) positive scalar function $w(x) > 0$. The corresponding *weighted inner product* and *norm* are

$$\langle f, g \rangle = \int_a^b f(x)\, g(x)\, w(x)\, dx, \qquad \| f \| = \sqrt{\int_a^b f(x)^2\, w(x)\, dx}\ . \tag{3.15}$$

The verification of the inner product axioms in this case is left as an exercise for the reader. As in the finite-dimensional version, weighted inner products are often used in statistics and data analysis, [**20**, **43**, **87**].

Exercises

3.1.21. For each of the given pairs of functions in $C^0[0,1]$, find their L^2 inner product $\langle f, g \rangle$ and their L^2 norms $\| f \|, \| g \|$: (a) $f(x) = 1$, $g(x) = x$; (b) $f(x) = \cos 2\pi x$, $g(x) = \sin 2\pi x$; (c) $f(x) = x$, $g(x) = e^x$; (d) $f(x) = (x+1)^2$, $g(x) = \dfrac{1}{x+1}$.

3.1.22. Let $f(x) = x$, $g(x) = 1 + x^2$. Compute $\langle f, g \rangle$, $\| f \|$, and $\| g \|$ for (a) the L^2 inner product $\langle f, g \rangle = \int_0^1 f(x)\, g(x)\, dx$; (b) the L^2 inner product $\langle f, g \rangle = \int_{-1}^1 f(x)\, g(x)\, dx$; (c) the weighted inner product $\langle f, g \rangle = \int_0^1 f(x)\, g(x)\, x\, dx$.

3.1.23. Which of the following formulas for $\langle f, g \rangle$ define inner products on the space $C^0[-1,1]$? (a) $\int_{-1}^1 f(x)\, g(x)\, e^{-x}\, dx$, (b) $\int_{-1}^1 f(x)\, g(x)\, x\, dx$, (c) $\int_{-1}^1 f(x)\, g(x)\, (x+2)\, dx$, (d) $\int_{-1}^1 f(x)\, g(x)\, x^2\, dx$.

3.1.24. Prove that $\langle f, g \rangle = \int_0^1 f(x)\, g(x)\, dx$ does *not* define an inner product on the vector space $C^0[-1,1]$. Explain why this does not contradict the fact that it defines an inner product on the vector space $C^0[0,1]$. Does it define an inner product on the subspace $\mathcal{P}^{(n)} \subset C^0[-1,1]$ consisting of all polynomial functions?

3.1.25. Does either of the following define an inner product on $\mathrm{C}^0[0,1]$?
(a) $\langle f , g \rangle = f(0)\,g(0) + f(1)\,g(1)$, (b) $\langle f , g \rangle = f(0)\,g(0) + f(1)\,g(1) + \int_0^1 f(x)\,g(x)\,dx$.

3.1.26. Let $f(x)$ be a function, and $\| f \|$ its L^2 norm on $[a,b]$. Is $\| f^2 \| = \| f \|^2$? If yes, prove the statement. If no, give a counterexample.

♢ 3.1.27. Prove that $\langle f , g \rangle = \int_a^b \big[\, f(x)\,g(x) + f'(x)\,g'(x) \,\big]\, dx$ defines an inner product on the space $\mathrm{C}^1[a,b]$ of continuously differentiable functions on the interval $[a,b]$. Write out the corresponding norm, known as the *Sobolev* H^1 *norm*; it and its generalizations play an extremely important role in advanced mathematical analysis, [**49**].

3.1.28. Let $V = \mathrm{C}^1[-1,1]$ denote the vector space of continuously differentiable functions for $-1 \le x \le 1$. (a) Does the expression $\langle f , g \rangle = \int_{-1}^1 f'(x)\,g'(x)\,dx$ define an inner product on V? (b) Answer the same question for the subspace $W = \{\, f \in V \mid f(0) = 0 \,\}$ consisting of all continuously differentiable functions that vanish at 0.

♢ 3.1.29. (a) Let $h(x) \ge 0$ be a continuous, non-negative function defined on an interval $[a,b]$. Prove that $\int_a^b h(x)\,dx = 0$ if and only if $h(x) \equiv 0$. *Hint*: Use the fact that $\int_c^d h(x)\,dx > 0$ if $h(x) > 0$ for $c \le x \le d$. (b) Give an example that shows that this result is not valid if h is allowed to be discontinuous.

♢ 3.1.30. (a) Prove the inner product axioms for the weighted inner product (3.15), assuming $w(x) > 0$ for all $a \le x \le b$. (b) Explain why it does *not* define an inner product if w is continuous and $w(x_0) < 0$ for some $x_0 \in [a,b]$. (c) If $w(x) \ge 0$ for $a \le x \le b$, does (3.15) define an inner product? *Hint*: Your answer may depend upon $w(x)$.

♡ 3.1.31. Let $\Omega \subset \mathbb{R}^2$ be a closed bounded subset. Let $\mathrm{C}^0(\Omega)$ denote the vector space consisting of all continuous, bounded real-valued functions $f(x,y)$ defined for $(x,y) \in \Omega$. (a) Prove that if $f(x,y) \ge 0$ is continuous and $\iint_\Omega f(x,y)\,dx\,dy = 0$, then $f(x,y) \equiv 0$. *Hint*: Mimic Exercise 3.1.29. (b) Use this result to prove that

$$\langle f , g \rangle = \iint_\Omega f(x,y)\,g(x,y)\,dx\,dy \tag{3.16}$$

defines an inner product on $\mathrm{C}^0(\Omega)$, called the L^2 inner product on the domain Ω. What is the corresponding norm?

3.1.32. Compute the L^2 inner product (3.16) and norms of the functions $f(x,y) \equiv 1$ and $g(x,y) = x^2 + y^2$, when (a) $\Omega = \{\, 0 \le x \le 1, 0 \le y \le 1 \,\}$ is the unit square;
(b) $\Omega = \{\, x^2 + y^2 \le 1 \,\}$ is the unit disk. *Hint*: Use polar coordinates.

♡ 3.1.33. Let V be the vector space consisting of all continuous, vector-valued functions $\mathbf{f}(x) = (\, f_1(x), f_2(x) \,)^T$ defined on the interval $0 \le x \le 1$.
(a) Prove that $\langle\!\langle\, \mathbf{f} , \mathbf{g} \,\rangle\!\rangle = \int_0^1 \big[\, f_1(x)\,g_1(x) + f_2(x)\,g_2(x) \,\big]\,dx$ defines an inner product on V.
(b) Prove, more generally, that if $\langle\, \mathbf{v} , \mathbf{w} \,\rangle$ is any inner product on $\mathbb{R}^2$, then $\langle\!\langle\, \mathbf{f} , \mathbf{g} \,\rangle\!\rangle = \int_a^b \langle\, \mathbf{f}(x) , \mathbf{g}(x) \,\rangle\,dx$ defines an inner product on V. (Part (a) corresponds to the dot product.) (c) Use part (b) to prove that

$$\langle\!\langle\, \mathbf{f} , \mathbf{g} \,\rangle\!\rangle = \int_a^b \big[\, f_1(x)\,g_1(x) - f_1(x)\,g_2(x) - f_2(x)\,g_1(x) + 3\,f_2(x)\,g_2(x) \,\big]\,dx$$

defines an inner product on V.

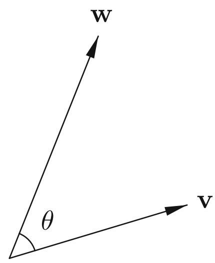

Figure 3.2. Angle Between Two Vectors.

3.2 Inequalities

There are two absolutely basic inequalities that are valid for *any* inner product space. The first is inspired by the geometric interpretation of the dot product on Euclidean space in terms of the angle between vectors. It is named after two of the founders of modern analysis, the nineteenth-century mathematicians Augustin Cauchy, of France, and Herman Schwarz, of Germany, who established it in the case of the L^2 inner product on function space.† The more familiar triangle inequality, that the length of any side of a triangle is bounded by the sum of the lengths of the other two sides, is, in fact, an immediate consequence of the Cauchy–Schwarz inequality, and hence also valid for any norm based on an inner product.

We will present these two inequalities in their most general, abstract form, since this brings their essence into the limelight. Specializing to different inner products and norms on both finite-dimensional and infinite-dimensional vector spaces leads to a wide variety of striking and useful inequalities.

The Cauchy–Schwarz Inequality

In Euclidean geometry, the dot product between two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ can be geometrically characterized by the equation

$$\mathbf{v} \cdot \mathbf{w} = \| \mathbf{v} \| \, \| \mathbf{w} \| \, \cos\theta, \tag{3.17}$$

where $\theta = \measuredangle\,(\mathbf{v}, \mathbf{w})$ measures the angle between the two vectors, as illustrated in Figure 3.2. Since $|\cos\theta\,| \leq 1$, the absolute value of the dot product is bounded by the product of the lengths of the vectors:

$$|\,\mathbf{v} \cdot \mathbf{w}\,| \;\leq\; \| \mathbf{v} \| \, \| \mathbf{w} \|.$$

This is the simplest form of the general *Cauchy–Schwarz inequality*. We present a direct algebraic proof that does not rely on the geometrical notions of length and angle and thus demonstrates its universal validity for *any* inner product.

Theorem 3.5. Every inner product satisfies the Cauchy–Schwarz inequality

$$|\,\langle\, \mathbf{v} \,, \mathbf{w} \,\rangle\,| \;\leq\; \| \mathbf{v} \| \, \| \mathbf{w} \|, \qquad \text{for all} \qquad \mathbf{v}, \mathbf{w} \in V. \tag{3.18}$$

Here, $\| \mathbf{v} \|$ is the associated norm, while $|\cdot|$ denotes the absolute value of a real number. Equality holds in (3.18) if and only if $\mathbf{v}$ and $\mathbf{w}$ are parallel vectors.

† Russians also give credit for its discovery to their compatriot Viktor Bunyakovsky, and, indeed, some authors append his name to the inequality.

Proof: The case when $\mathbf{w} = \mathbf{0}$ is trivial, since both sides of (3.18) are equal to 0. Thus, we concentrate on the case when $\mathbf{w} \neq \mathbf{0}$. Let $t \in \mathbb{R}$ be an arbitrary scalar. Using the three inner product axioms, we have

$$\begin{aligned} 0 \le \| \mathbf{v} + t\,\mathbf{w} \|^2 = \langle \mathbf{v} + t\,\mathbf{w}\,, \mathbf{v} + t\,\mathbf{w} \rangle &= \langle \mathbf{v}\,, \mathbf{v} \rangle + 2\,t\,\langle \mathbf{v}\,, \mathbf{w} \rangle + t^2\,\langle \mathbf{w}\,, \mathbf{w} \rangle \\ &= \| \mathbf{v} \|^2 + 2\,t\,\langle \mathbf{v}\,, \mathbf{w} \rangle + t^2\,\| \mathbf{w} \|^2, \end{aligned} \tag{3.19}$$

with equality holding if and only if $\mathbf{v} = -\,t\,\mathbf{w}$ — which requires $\mathbf{v}$ and $\mathbf{w}$ to be parallel vectors. We fix $\mathbf{v}$ and $\mathbf{w}$, and consider the right-hand side of (3.19) as a quadratic function of the scalar variable t:

$$0 \le p(t) = a\,t^2 + 2\,b\,t + c, \qquad \text{where} \qquad a = \| \mathbf{w} \|^2, \qquad b = \langle \mathbf{v}\,, \mathbf{w} \rangle, \qquad c = \| \mathbf{v} \|^2.$$

To get the maximum mileage out of the fact that $p(t) \ge 0$, let us look at where it assumes its minimum, which occurs when its derivative is zero:

$$p'(t) = 2\,a\,t + 2\,b = 0, \qquad \text{and so} \qquad t \;=\; -\,\frac{b}{a} \;=\; -\,\frac{\langle \mathbf{v}\,, \mathbf{w} \rangle}{\| \mathbf{w} \|^2}\,.$$

Substituting this particular value of t into (3.19), we obtain

$$0 \;\le\; \| \mathbf{v} \|^2 - 2\,\frac{\langle \mathbf{v}\,, \mathbf{w} \rangle^2}{\| \mathbf{w} \|^2} + \frac{\langle \mathbf{v}\,, \mathbf{w} \rangle^2}{\| \mathbf{w} \|^2} \;=\; \| \mathbf{v} \|^2 - \frac{\langle \mathbf{v}\,, \mathbf{w} \rangle^2}{\| \mathbf{w} \|^2}\,.$$

Rearranging this last inequality, we conclude that

$$\frac{\langle \mathbf{v}\,, \mathbf{w} \rangle^2}{\| \mathbf{w} \|^2} \;\le\; \| \mathbf{v} \|^2, \qquad \text{or} \qquad \langle \mathbf{v}\,, \mathbf{w} \rangle^2 \;\le\; \| \mathbf{v} \|^2\,\| \mathbf{w} \|^2.$$

Also, as noted above, equality holds if and only if $\mathbf{v}$ and $\mathbf{w}$ are parallel. Equality also holds when $\mathbf{w} = \mathbf{0}$, which is of course parallel to every vector $\mathbf{v}$. Taking the (positive) square root of both sides of the final inequality completes the proof of (3.18). *Q.E.D.*

Given any inner product, we can use the quotient

$$\cos\theta = \frac{\langle \mathbf{v}\,, \mathbf{w} \rangle}{\| \mathbf{v} \|\;\| \mathbf{w} \|} \tag{3.20}$$

to define the "angle" $\theta = \sphericalangle\,(\mathbf{v}, \mathbf{w})$ between the vector space elements $\mathbf{v}, \mathbf{w} \in V$. The Cauchy–Schwarz inequality tells us that the ratio lies between -1 and $+1$, and hence the angle θ is well defined modulo 2π, and, in fact, unique if we restrict it to lie in the range $0 \le \theta \le \pi$.

For example, the vectors $\mathbf{v} = (\,1, 0, 1\,)^T$, $\mathbf{w} = (\,0, 1, 1\,)^T$ have dot product $\mathbf{v} \cdot \mathbf{w} = 1$ and norms $\| \mathbf{v} \| = \| \mathbf{w} \| = \sqrt{2}$. Hence the Euclidean angle between them is given by

$$\cos\theta = \frac{1}{\sqrt{2}\cdot\sqrt{2}} = \frac{1}{2}\,, \qquad \text{and so} \qquad \theta = \sphericalangle\,(\mathbf{v}, \mathbf{w}) = \tfrac{1}{3}\,\pi = 1.0471\,\ldots\,.$$

On the other hand, if we adopt the weighted inner product $\langle \mathbf{v}\,, \mathbf{w} \rangle = v_1\,w_1 + 2\,v_2\,w_2 + 3\,v_3\,w_3$, then $\mathbf{v} \cdot \mathbf{w} = 3$, $\| \mathbf{v} \| = 2$, $\| \mathbf{w} \| = \sqrt{5}$, and hence their "weighted" angle becomes

$$\cos\theta = \frac{3}{2\,\sqrt{5}} = .67082\,\ldots\,, \qquad \text{with} \qquad \theta = \sphericalangle\,(\mathbf{v}, \mathbf{w}) = .83548\,\ldots\,.$$

Thus, the measurement of angle (and length) depends on the choice of an underlying inner product.

Similarly, under the L^2 inner product on the interval $[0,1]$, the "angle" θ between the polynomials $p(x) = x$ and $q(x) = x^2$ is given by

$$\cos\theta = \frac{\langle x, x^2 \rangle}{\|x\| \, \|x^2\|} = \frac{\displaystyle\int_0^1 x^3\,dx}{\sqrt{\displaystyle\int_0^1 x^2\,dx}\ \sqrt{\displaystyle\int_0^1 x^4\,dx}} = \frac{\frac14}{\sqrt{\frac13}\ \sqrt{\frac15}} = \sqrt{\frac{15}{16}}\,,$$

so that $\theta = \sphericalangle(p,q) = .25268\ldots$ radians.

Warning. You should not try to give this notion of angle between functions more significance than the formal definition warrants — it does not correspond to any "angular" properties of their graphs. Also, the value depends on the choice of inner product and the interval upon which it is being computed. For example, if we change to the L^2 inner product on the interval $[-1,1]$, then $\langle x, x^2 \rangle = \displaystyle\int_{-1}^1 x^3\,dx = 0$. Hence, (3.20) becomes $\cos\theta = 0$, so the "angle" between x and x^2 is now $\theta = \sphericalangle(p,q) = \frac12\pi$.

Exercises

3.2.1. Verify the Cauchy–Schwarz inequality for each of the following pairs of vectors $\mathbf{v}, \mathbf{w}$, using the standard dot product, and then determine the angle between them:
(*a*) $(1,2)^T, (-1,2)^T$, (*b*) $(1,-1,0)^T, (-1,0,1)^T$, (*c*) $(1,-1,0)^T, (2,2,2)^T$,
(*d*) $(1,-1,1,0)^T, (-2,0,-1,1)^T$, (*e*) $(2,1,-2,-1)^T, (0,-1,2,-1)^T$.

3.2.2. (*a*) Find the Euclidean angle between the vectors $(1,1,1,1)^T$ and $(1,1,1,-1)^T$ in $\mathbb{R}^4$.
(*b*) List the possible angles between $(1,1,1,1)^T$ and $(a_1,a_2,a_3,a_4)^T$, where each a_i is either 1 or -1.

3.2.3. Prove that the points $(0,0,0), (1,1,0), (1,0,1), (0,1,1)$ form the vertices of a regular tetrahedron, meaning that all sides have the same length. What is the common Euclidean angle between the edges? What is the angle between any two rays going from the center $\left(\frac12,\frac12,\frac12\right)$ to the vertices? **Remark.** Methane molecules assume this geometric configuration, and the angle influences their chemistry.

3.2.4. Verify the Cauchy–Schwarz inequality for the vectors $\mathbf{v} = (1,2)^T$, $\mathbf{w} = (1,-3)^T$, using (*a*) the dot product; (*b*) the weighted inner product $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + 2 v_2 w_2$;
(*c*) the inner product (3.9).

3.2.5. Verify the Cauchy–Schwarz inequality for the vectors $\mathbf{v} = (3,-1,2)^T$, $\mathbf{w} = (1,-1,1)^T$, using (*a*) the 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 $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \mathbf{w}$.

♢ 3.2.6. Show that one can determine the angle θ between $\mathbf{v}$ and $\mathbf{w}$ via the formula
$\cos\theta = \dfrac{\|\mathbf{v}+\mathbf{w}\|^2 - \|\mathbf{v}-\mathbf{w}\|^2}{4\,\|\mathbf{v}\|\,\|\mathbf{w}\|}$. Draw a picture illustrating what is being measured.

♢ 3.2.7. *The Law of Cosines*: Prove that the formula

$$\|\mathbf{v}-\mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2\,\|\mathbf{v}\|\,\|\mathbf{w}\|\,\cos\theta, \qquad (3.21)$$

where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$, is valid in every inner product space.

3.2.8. Use the Cauchy–Schwarz inequality to prove $(a\cos\theta + b\sin\theta)^2 \le a^2 + b^2$ for any θ, a, b.

3.2.9. Prove that $(a_1 + a_2 + \cdots + a_n)^2 \le n\,(a_1^2 + a_2^2 + \cdots + a_n^2)$ for any real numbers $a_1, \ldots, a_n$. When does equality hold?

♡ 3.2.10. The *cross product* of two vectors in $\mathbb{R}^2$ is defined as the scalar

$$\mathbf{v} \times \mathbf{w} = v_1 w_2 - v_2 w_1 \qquad \text{for} \qquad \mathbf{v} = (\,v_1, v_2\,)^T, \qquad \mathbf{w} = (\,w_1, w_2\,)^T. \tag{3.22}$$

(*a*) Does the cross product define an inner product on $\mathbb{R}^2$? Carefully explain which axioms are valid and which are not. (*b*) Prove that $\mathbf{v} \times \mathbf{w} = \|\mathbf{v}\|\,\|\mathbf{w}\|\,\sin\theta$, where θ denotes the angle from $\mathbf{v}$ to $\mathbf{w}$ as in Figure 3.2. (*c*) Prove that $\mathbf{v} \times \mathbf{w} = 0$ if and only if $\mathbf{v}$ and $\mathbf{w}$ are parallel vectors. (*d*) Show that $|\,\mathbf{v} \times \mathbf{w}\,|$ equals the area of the parallelogram defined by $\mathbf{v}$ and $\mathbf{w}$.

♢ 3.2.11. Explain why the inequality $\langle\, \mathbf{v}\,,\mathbf{w}\,\rangle \le \|\mathbf{v}\|\,\|\mathbf{w}\|$, obtained by omitting the absolute value sign on the left-hand side of Cauchy–Schwarz, is valid.

3.2.12. Verify the Cauchy–Schwarz inequality for the functions $f(x) = x$ and $g(x) = e^x$ with respect to (*a*) the L^2 inner product on the interval $[0,1]$, (*b*) the L^2 inner product on $[-1,1]$, (*c*) the weighted inner product $\langle\, f\,,g\,\rangle = \displaystyle\int_0^1 f(x)\,g(x)\,e^{-x}\,dx$.

3.2.13. Using the L^2 inner product on the interval $[\,0\,,\pi\,]$, find the angle between the functions (*a*) 1 and $\cos x$; (*b*) 1 and $\sin x$; (*c*) $\cos x$ and $\sin x$.

3.2.14. Verify the Cauchy–Schwarz inequality for the two particular functions appearing in Exercise 3.1.32 using the L^2 inner product on (*a*) the unit square; (*b*) the unit disk.

Orthogonal Vectors

In Euclidean geometry, a particularly noteworthy configuration occurs when two vectors are *perpendicular*. Perpendicular vectors meet at a right angle, $\theta = \frac{1}{2}\pi$ or $\frac{3}{2}\pi$, with $\cos\theta = 0$. The angle formula (3.17) implies that the vectors $\mathbf{v}, \mathbf{w}$ are perpendicular if and only if their dot product vanishes: $\mathbf{v} \cdot \mathbf{w} = 0$. Perpendicularity is of interest in general inner product spaces, but, for historical reasons, has been given a more suggestive name.

Definition 3.6. Two elements $\mathbf{v}, \mathbf{w} \in V$ of an inner product space V are called *orthogonal* if their inner product vanishes: $\langle\, \mathbf{v}\,,\mathbf{w}\,\rangle = 0$.

In particular, the zero element is orthogonal to all other vectors: $\langle\, \mathbf{0}\,,\mathbf{v}\,\rangle = 0$ for all $\mathbf{v} \in V$. Orthogonality is a remarkably powerful tool that appears throughout the manifold applications of linear algebra, and often serves to dramatically simplify many computations. We will devote all of Chapter 4 to a detailed exploration of its manifold implications.

Example 3.7. The vectors $\mathbf{v} = (\,1,2\,)^T$ and $\mathbf{w} = (\,6,-3\,)^T$ are orthogonal with respect to the Euclidean dot product in $\mathbb{R}^2$, since $\mathbf{v} \cdot \mathbf{w} = 1 \cdot 6 + 2 \cdot (-3) = 0$. We deduce that they meet at a right angle. However, these vectors are *not* orthogonal with respect to the weighted inner product (3.8):

$$\langle\, \mathbf{v}\,,\mathbf{w}\,\rangle = \left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 6 \\ -3 \end{pmatrix} \right\rangle = 2 \cdot 1 \cdot 6 + 5 \cdot 2 \cdot (-3) = -18 \ne 0.$$

Thus, the property of orthogonality, like angles in general, depends upon which inner product is being used.

Example 3.8. The polynomials $p(x) = x$ and $q(x) = x^2 - \frac{1}{2}$ are orthogonal with respect to the inner product $\langle p, q \rangle = \int_0^1 p(x)\, q(x)\, dx$ on the interval $[0,1]$, since

$$\left\langle x\,,\, x^2 - \tfrac{1}{2} \right\rangle = \int_0^1 x\left(x^2 - \tfrac{1}{2} \right) dx = \int_0^1 \left(x^3 - \tfrac{1}{2}\,x \right) dx = 0.$$

They fail to be orthogonal on most other intervals. For example, on the interval $[0,2]$,

$$\left\langle x\,,\, x^2 - \tfrac{1}{2} \right\rangle = \int_0^2 x\left(x^2 - \tfrac{1}{2} \right) dx = \int_0^2 \left(x^3 - \tfrac{1}{2}\,x \right) dx = 3.$$

Warning. There is no obvious connection between the orthogonality of two functions and the geometry of their graphs.

Exercises

Note: Unless stated otherwise, the inner product is the standard dot product on $\mathbb{R}^n$.

3.2.15. (*a*) Find a such that $(2, a, -3)^T$ is orthogonal to $(-1, 3, -2)^T$. (*b*) Is there any value of a for which $(2, a, -3)^T$ is parallel to $(-1, 3, -2)^T$?

3.2.16. Find all vectors in $\mathbb{R}^3$ that are orthogonal to both $(1, 2, 3)^T$ and $(-2, 0, 1)^T$.

3.2.17. Answer Exercises 3.2.15 and 3.2.16 using the weighted inner product
$$\langle \mathbf{v}, \mathbf{w} \rangle = 3 v_1 w_1 + 2 v_2 w_2 + v_3 w_3.$$

3.2.18. Find all vectors in $\mathbb{R}^4$ that are orthogonal to both $(1, 2, 3, 4)^T$ and $(5, 6, 7, 8)^T$.

3.2.19. Determine a basis for the subspace $W \subset \mathbb{R}^4$ consisting of all vectors which are orthogonal to the vector $(1, 2, -1, 3)^T$.

3.2.20. Find three vectors $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$ such that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, $\mathbf{u}$ and $\mathbf{w}$ are orthogonal, but $\mathbf{v}$ and $\mathbf{w}$ are *not* orthogonal. Are your vectors linearly independent or linearly dependent? Can you find vectors of the opposite dependency satisfying the same conditions? Why or why not?

3.2.21. For what values of a, b are the vectors $(1, 1, a)^T$ and $(b, -1, 1)^T$ orthogonal
(*a*) with respect to the dot product?
(*b*) with respect to the weighted inner product of Exercise 3.2.17?

3.2.22. When is a vector orthogonal to itself?

◇ 3.2.23. Prove that the only element $\mathbf{w}$ in an inner product space V that is orthogonal to every vector, so $\langle \mathbf{w}, \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in V$, is the zero vector: $\mathbf{w} = \mathbf{0}$.

3.2.24. A vector with $\| \mathbf{v} \| = 1$ is known as a *unit vector*. Prove that if $\mathbf{v}, \mathbf{w}$ are both unit vectors, then $\mathbf{v} + \mathbf{w}$ and $\mathbf{v} - \mathbf{w}$ are orthogonal. Are they also unit vectors?

◇ 3.2.25. Let V be an inner product space and $\mathbf{v} \in V$. Prove that the set of all vectors $\mathbf{w} \in V$ that are orthogonal to $\mathbf{v}$ is a subspace of V.

3.2.26. (*a*) Show that the polynomials $p_1(x) = 1,\ p_2(x) = x - \frac{1}{2},\ p_3(x) = x^2 - x + \frac{1}{6}$ are mutually orthogonal with respect to the L^2 inner product on the interval $[0,1]$.
(*b*) Show that the functions $\sin n\pi x$, $n = 1, 2, 3, \ldots$, are mutually orthogonal with respect to the same inner product.

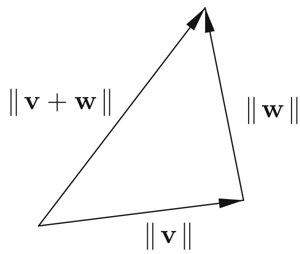

Figure 3.3. Triangle Inequality.

3.2.27. Find a non-zero quadratic polynomial that is orthogonal to both $p_1(x) = 1$ and $p_2(x) = x$ under the L^2 inner product on the interval $[-1, 1]$.

3.2.28. Find all quadratic polynomials that are orthogonal to the function e^x with respect to the L^2 inner product on the interval $[0, 1]$.

3.2.29. Determine all pairs among the functions 1, x, $\cos \pi x$, $\sin \pi x$, e^x, that are orthogonal with respect to the L^2 inner product on $[-1, 1]$.

3.2.30. Find two non-zero functions that are orthogonal with respect to the weighted inner product $\langle f , g \rangle = \displaystyle\int_0^1 f(x)\, g(x)\, x \, dx$.

The Triangle Inequality

The familiar triangle inequality states that the length of one side of a triangle is at most equal to the sum of the lengths of the other two sides. Referring to Figure 3.3, if the first two sides are represented by vectors $\mathbf{v}$ and $\mathbf{w}$, then the third corresponds to their sum $\mathbf{v} + \mathbf{w}$. The triangle inequality turns out to be an elementary consequence of the Cauchy–Schwarz inequality (3.18), and hence is valid in *every* inner product space.

Theorem 3.9. The norm associated with an inner product satisfies the *triangle inequality*

$$\| \mathbf{v} + \mathbf{w} \| \le \| \mathbf{v} \| + \| \mathbf{w} \| \qquad \text{for all} \qquad \mathbf{v}, \mathbf{w} \in V. \tag{3.23}$$

Equality holds if and only if $\mathbf{v}$ and $\mathbf{w}$ are parallel vectors.

Proof: We compute

$$\begin{aligned} \| \mathbf{v} + \mathbf{w} \|^2 = \langle \mathbf{v} + \mathbf{w} , \mathbf{v} + \mathbf{w} \rangle &= \| \mathbf{v} \|^2 + 2 \langle \mathbf{v} , \mathbf{w} \rangle + \| \mathbf{w} \|^2 \\ &\le \| \mathbf{v} \|^2 + 2 \| \mathbf{v} \| \, \| \mathbf{w} \| + \| \mathbf{w} \|^2 = \big(\| \mathbf{v} \| + \| \mathbf{w} \| \big)^2, \end{aligned}$$

where the middle inequality follows from Cauchy–Schwarz, cf. Exercise 3.2.11. Taking square roots of both sides and using the fact that the resulting expressions are both positive completes the proof. *Q.E.D.*

Example 3.10. The vectors $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix}$ sum to $\mathbf{v} + \mathbf{w} = \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix}$. Their Euclidean norms are $\| \mathbf{v} \| = \sqrt{6}$ and $\| \mathbf{w} \| = \sqrt{13}$, while $\| \mathbf{v} + \mathbf{w} \| = \sqrt{17}$. The triangle inequality (3.23) in this case says $\sqrt{17} \le \sqrt{6} + \sqrt{13}$, which is true.

Example 3.11. Consider the functions $f(x) = x - 1$ and $g(x) = x^2 + 1$. Using the L^2 norm on the interval $[0,1]$, we find that

$$\|f\| = \sqrt{\int_0^1 (x-1)^2\,dx} = \sqrt{\frac{1}{3}}, \qquad \|g\| = \sqrt{\int_0^1 (x^2+1)^2\,dx} = \sqrt{\frac{28}{15}},$$
$$\|f+g\| = \sqrt{\int_0^1 (x^2+x)^2\,dx} = \sqrt{\frac{31}{30}}.$$

The triangle inequality requires $\sqrt{\frac{31}{30}} \le \sqrt{\frac{1}{3}} + \sqrt{\frac{28}{15}}$, which is valid.

The Cauchy–Schwarz and triangle inequalities look much more impressive when written out in full detail. For the Euclidean dot product (3.1), they are

$$\left|\sum_{i=1}^n v_i w_i\right| \le \sqrt{\sum_{i=1}^n v_i^2}\ \sqrt{\sum_{i=1}^n w_i^2},$$
$$\sqrt{\sum_{i=1}^n (v_i+w_i)^2} \le \sqrt{\sum_{i=1}^n v_i^2} + \sqrt{\sum_{i=1}^n w_i^2}. \tag{3.24}$$

Theorems 3.5 and 3.9 imply that these inequalities are valid for arbitrary real numbers $v_1, \ldots, v_n, w_1, \ldots, w_n$. For the L^2 inner product (3.13) on function space, they produce the following splendid integral inequalities:

$$\left|\int_a^b f(x)\,g(x)\,dx\right| \le \sqrt{\int_a^b f(x)^2\,dx}\ \sqrt{\int_a^b g(x)^2\,dx},$$
$$\sqrt{\int_a^b \big[f(x)+g(x)\big]^2\,dx} \le \sqrt{\int_a^b f(x)^2\,dx} + \sqrt{\int_a^b g(x)^2\,dx}, \tag{3.25}$$

which hold for arbitrary continuous (and, in fact, rather general) functions. The first of these is the original Cauchy–Schwarz inequality, whose proof appeared to be quite deep when it first appeared. Only after the abstract notion of an inner product space was properly formalized did its innate simplicity and generality become evident.

Exercises

3.2.31. Use the dot product on $\mathbb{R}^3$ to answer the following: (*a*) Find the angle between the vectors $(1,2,3)^T$ and $(1,-1,2)^T$. (*b*) Verify the Cauchy–Schwarz and triangle inequalities for these two particular vectors. (*c*) Find all vectors that are orthogonal to both of these vectors.

3.2.32. Verify the triangle inequality for each pair of vectors in Exercise 3.2.1.

3.2.33. Verify the triangle inequality for the vectors and inner products in Exercise 3.2.4.

3.2.34. Verify the triangle inequality for the functions in Exercise 3.2.12 for the indicated inner products.

3.2.35. Verify the triangle inequality for the two particular functions appearing in Exercise 3.1.32 with respect to the L^2 inner product on (*a*) the unit square; (*b*) the unit disk.

3.2.36. Use the L^2 inner product $\langle f , g \rangle = \int_{-1}^{1} f(x)\, g(x)\, dx$ to answer the following: (*a*) Find the "angle" between the functions 1 and x. Are they orthogonal? (*b*) Verify the Cauchy–Schwarz and triangle inequalities for these two functions. (*c*) Find all quadratic polynomials $p(x) = a + b\,x + c\,x^2$ that are orthogonal to both of these functions.

3.2.37. (*a*) Write down the explicit formulae for the Cauchy–Schwarz and triangle inequalities based on the weighted inner product $\langle f , g \rangle = \int_0^1 f(x)\, g(x)\, e^x\, dx$. (*b*) Verify that the inequalities hold when $f(x) = 1$, $g(x) = e^x$ by direct computation. (*c*) What is the "angle" between these two functions in this inner product?

3.2.38. Answer Exercise 3.2.37 for the Sobolev H^1 inner product

$$\langle f , g \rangle = \int_0^1 \left[\, f(x)\, g(x) + f'(x)\, g'(x) \,\right] dx, \qquad \text{cf. Exercise 3.1.27.}$$

3.2.39. Prove that $\|\, \mathbf{v} - \mathbf{w} \,\| \;\geq\; \Big|\, \|\, \mathbf{v} \,\| - \|\, \mathbf{w} \,\| \,\Big|$. Interpret this result pictorially.

3.2.40. *True or false*: $\|\, \mathbf{w} \,\| \leq \|\, \mathbf{v} \,\| + \|\, \mathbf{v} + \mathbf{w} \,\|$ for all $\mathbf{v}, \mathbf{w} \in V$.

♡ 3.2.41. (*a*) Prove that the space $\mathbb{R}^\infty$ consisting of all infinite sequences $\mathbf{x} = (x_1, x_2, x_3, \ \ldots\ )$ of real numbers $x_i \in \mathbb{R}$ is a vector space. (*b*) Prove that the set of all sequences $\mathbf{x}$ such that $\sum_{k=1}^{\infty} x_k^2 < \infty$ is a subspace, commonly denoted by $\ell^2 \subset \mathbb{R}^\infty$. (*c*) Write down two examples of sequences $\mathbf{x}$ belonging to ℓ^2 and two that do not belong to ℓ^2. (*d*) *True or false*: If $\mathbf{x} \in \ell^2$, then $x_k \to 0$ and $k \to \infty$. (*e*) *True or false*: If $x_k \to 0$ as $k \to \infty$, then $\mathbf{x} \in \ell^2$. (*f*) Given $\alpha \in \mathbb{R}$, let $\mathbf{x}$ be the sequence with $x_k = \alpha^k$. For which values of α is $\mathbf{x} \in \ell^2$? (*g*) Answer part (*f*) when $x_k = k^\alpha$. (*h*) Prove that $\langle \mathbf{x} , \mathbf{y} \rangle = \sum_{k=1}^{\infty} x_k\, y_k$ defines an inner product on the vector space ℓ^2. What is the corresponding norm? (*i*) Write out the Cauchy–Schwarz and triangle inequalities for the inner product space ℓ^2.

3.3 Norms

Every inner product gives rise to a norm that can be used to measure the magnitude or length of the elements of the underlying vector space. However, not every norm that is used in analysis and applications arises from an inner product. To define a general norm on a vector space, we will extract those properties that do not directly rely on the inner product structure.

Definition 3.12. A *norm* on a vector space V assigns a non-negative real number $\|\, \mathbf{v} \,\|$ to each vector $\mathbf{v} \in V$, subject to the following axioms, valid for every $\mathbf{v}, \mathbf{w} \in V$ and $c \in \mathbb{R}$:

(*i*) *Positivity*: $\|\, \mathbf{v} \,\| \geq 0$, with $\|\, \mathbf{v} \,\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

(*ii*) *Homogeneity*: $\|\, c\, \mathbf{v} \,\| = |\, c \,|\, \|\, \mathbf{v} \,\|$.

(*iii*) *Triangle inequality*: $\|\, \mathbf{v} + \mathbf{w} \,\| \leq \|\, \mathbf{v} \,\| + \|\, \mathbf{w} \,\|$.

As we now know, every inner product gives rise to a norm. Indeed, positivity of the norm is one of the inner product axioms. The homogeneity property follows since

$$\|\, c\, \mathbf{v} \,\| \;=\; \sqrt{\langle\, c\, \mathbf{v} , c\, \mathbf{v} \,\rangle} \;=\; \sqrt{c^2 \,\langle\, \mathbf{v} , \mathbf{v} \,\rangle} \;=\; |\, c \,|\, \sqrt{\langle\, \mathbf{v} , \mathbf{v} \,\rangle} \;=\; |\, c \,|\, \|\, \mathbf{v} \,\|.$$

Finally, the triangle inequality for an inner product norm was established in Theorem 3.9. Let us introduce some of the principal examples of norms that do not come from inner products.

First, let $V = \mathbb{R}^n$. The 1 *norm* of a vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T$ is defined as the sum of the absolute values of its entries:

$$\| \mathbf{v} \|_1 = | v_1 | + | v_2 | + \cdots + | v_n |. \tag{3.26}$$

The *max* or ∞ *norm* is equal to its maximal entry (in absolute value):

$$\| \mathbf{v} \|_\infty = \max \{ | v_1 |, | v_2 |, \ldots, | v_n | \}. \tag{3.27}$$

Verification of the positivity and homogeneity properties for these two norms is straightforward; the triangle inequality is a direct consequence of the elementary inequality

$$| a + b | \le | a | + | b |, \qquad a, b \in \mathbb{R},$$

for absolute values.

The Euclidean norm, 1 norm, and ∞ norm on $\mathbb{R}^n$ are just three representatives of the general p *norm*

$$\| \mathbf{v} \|_p = \sqrt[p]{\sum_{i=1}^{n} | v_i |^p}\ . \tag{3.28}$$

This quantity defines a norm for all $1 \le p < \infty$. The ∞ norm is a limiting case of (3.28) as $p \to \infty$. Note that the Euclidean norm (3.3) is the 2 norm, and is often designated as such; it is the only p norm which comes from an inner product. The positivity and homogeneity properties of the p norm are not hard to establish. The triangle inequality, however, is not trivial; in detail, it reads

$$\sqrt[p]{\sum_{i=1}^{n} | v_i + w_i |^p} \ \le\ \sqrt[p]{\sum_{i=1}^{n} | v_i |^p} \ +\ \sqrt[p]{\sum_{i=1}^{n} | w_i |^p}\ , \tag{3.29}$$

and is known as *Minkowski's inequality*. A complete proof can be found in [**50**].

There are analogous norms on the space $\mathrm{C}^0[a, b]$ of continuous functions on an interval $[a, b]$. Basically, one replaces the previous sums by integrals. Thus, the L^p norm is defined as

$$\| f \|_p = \sqrt[p]{\int_a^b | f(x) |^p \, dx}\ . \tag{3.30}$$

In particular, the L^1 norm is given by integrating the absolute value of the function:

$$\| f \|_1 = \int_a^b | f(x) | \, dx. \tag{3.31}$$

The L^2 norm (3.13) appears as a special case, $p = 2$, and, again, is the only one arising from an inner product. The limiting L^∞ norm is defined by the maximum

$$\| f \|_\infty = \max \{ | f(x) | \ : \ a \le x \le b \}. \tag{3.32}$$

Positivity of the L^p norms again relies on the fact that the only continuous non-negative function with zero integral is the zero function. Homogeneity is easily established. On the

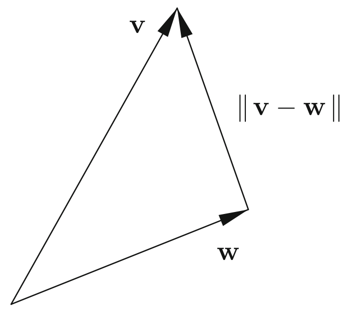

Figure 3.4. Distance Between Vectors.

other hand, the proof of the general triangle, or Minkowski, inequality for $p \neq 1, 2, \infty$ is again not trivial, [**19**, **68**].

Example 3.13. Consider the polynomial $p(x) = 3x^2 - 2$ on the interval $-1 \le x \le 1$. Its L^2 norm is

$$\| p \|_2 = \sqrt{\int_{-1}^{1} (3x^2 - 2)^2 \, dx} = \sqrt{\frac{18}{5}} = 1.8973\ldots .$$

Its L^∞ norm is

$$\| p \|_\infty = \max \left\{ \, | \, 3x^2 - 2 \, | \; : \; -1 \le x \le 1 \, \right\} = 2,$$

with the maximum occurring at $x = 0$. Finally, its L^1 norm is

$$\begin{aligned} \| p \|_1 &= \int_{-1}^{1} | \, 3x^2 - 2 \, | \, dx \\ &= \int_{-1}^{-\sqrt{2/3}} (3x^2 - 2) \, dx + \int_{-\sqrt{2/3}}^{\sqrt{2/3}} (2 - 3x^2) \, dx + \int_{\sqrt{2/3}}^{1} (3x^2 - 2) \, dx \\ &= \left(\tfrac{4}{3} \sqrt{\tfrac{2}{3}} - 1 \right) + \tfrac{8}{3} \sqrt{\tfrac{2}{3}} + \left(\tfrac{4}{3} \sqrt{\tfrac{2}{3}} - 1 \right) = \tfrac{16}{3} \sqrt{\tfrac{2}{3}} - 2 = 2.3546\ldots . \end{aligned}$$

Every norm defines a *distance* between vector space elements, namely

$$d(\mathbf{v}, \mathbf{w}) = \| \mathbf{v} - \mathbf{w} \|. \tag{3.33}$$

For the standard dot product norm, we recover the usual notion of distance between points in Euclidean space. Other types of norms produce alternative (and sometimes quite useful) notions of distance that are, nevertheless, subject to all the familiar properties:

(a) *Symmetry*: $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v})$;
(b) *Positivity*: $d(\mathbf{v}, \mathbf{w}) = 0$ if and only if $\mathbf{v} = \mathbf{w}$;
(c) *Triangle inequality*: $d(\mathbf{v}, \mathbf{w}) \le d(\mathbf{v}, \mathbf{z}) + d(\mathbf{z}, \mathbf{w})$.

Just as the distance between vectors measures how close they are to each other — keeping in mind that this measure of proximity depends on the underlying choice of norm — so the distance between functions in a normed function space tells something about how close they are to each other, which is related, albeit subtly, to how close their graphs are. Thus, the norm serves to define the *topology* of the underlying vector space, which determines notions of open and closed sets, convergence, and so on, [**19**, **68**].

Exercises

3.3.1. Compute the 1, 2, 3, and ∞ norms of the vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Verify the triangle inequality in each case.

3.3.2. Answer Exercise 3.3.1 for (*a*) $\begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix}$, (*b*) $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$, (*c*) $\begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ -3 \end{pmatrix}$.

3.3.3. Which two of the vectors $\mathbf{u} = (-2,2,1)^T$, $\mathbf{v} = (1,4,1)^T$, $\mathbf{w} = (0,0,-1)^T$ are closest to each other in distance for (*a*) the Euclidean norm? (*b*) the ∞ norm? (*c*) the 1 norm?

3.3.4. (a) Compute the L^∞ norm on $[0,1]$ of the functions $f(x) = \frac{1}{3} - x$ and $g(x) = x - x^2$. (*b*) Verify the triangle inequality for these two particular functions.

3.3.5. Answer Exercise 3.3.4 using the L^1 norm.

3.3.6. Which two of the functions $f(x) = 1$, $g(x) = x$, $h(x) = \sin \pi x$ are closest to each other on the interval $[0,1]$ under (*a*) the L^1 norm? (*b*) the L^2 norm? (*c*) the L^∞ norm?

3.3.7. Consider the functions $f(x) = 1$ and $g(x) = x - \frac{3}{4}$ as elements of the vector space $C^0[0,1]$. For each of the following norms, compute $\| f \|$, $\| g \|$, $\| f + g \|$, and verify the triangle inequality: (*a*) the L^1 norm; (*b*) the L^2 norm; (*c*) the L^3 norm; (*d*) the L^∞ norm.

3.3.8. Answer Exercise 3.3.7 when $f(x) = e^x$ and $g(x) = e^{-x}$.

3.3.9. Carefully prove that $\| (x,y)^T \| = | x | + 2 | x - y |$ defines a norm on $\mathbb{R}^2$.

3.3.10. Prove that the following formulas define norms on $\mathbb{R}^2$: (a) $\| \mathbf{v} \| = \sqrt{2 v_1^2 + 3 v_2^2}$, (*b*) $\| \mathbf{v} \| = \sqrt{2 v_1^2 - v_1 v_2 + 2 v_2^2}$, (*c*) $\| \mathbf{v} \| = 2 | v_1 | + | v_2 |$, (*d*) $\| \mathbf{v} \| = \max\big\{ 2 | v_1 |, | v_2 | \big\}$, (*e*) $\| \mathbf{v} \| = \max\big\{ | v_1 - v_2 |, | v_1 + v_2 | \big\}$, (*f*) $\| \mathbf{v} \| = | v_1 - v_2 | + | v_1 + v_2 |$.

3.3.11. Which of the following formulas define norms on $\mathbb{R}^3$? (*a*) $\| \mathbf{v} \| = \sqrt{2 v_1^2 + v_2^2 + 3 v_3^2}$, (*b*) $\| \mathbf{v} \| = \sqrt{v_1^2 + 2 v_1 v_2 + v_2^2 + v_3^2}$, (*c*) $\| \mathbf{v} \| = \max\big\{ | v_1 |, | v_2 |, | v_3 | \big\}$, (*d*) $\| \mathbf{v} \| = | v_1 - v_2 | + | v_2 - v_3 | + | v_3 - v_1 |$, (*e*) $\| \mathbf{v} \| = | v_1 | + \max\big\{ | v_2 |, | v_3 | \big\}$.

3.3.12. Prove that two parallel vectors $\mathbf{v}$ and $\mathbf{w}$ have the same norm if and only if $\mathbf{v} = \pm \mathbf{w}$.

3.3.13. *True or false*: If $\| \mathbf{v} + \mathbf{w} \| = \| \mathbf{v} \| + \| \mathbf{w} \|$, then $\mathbf{v}, \mathbf{w}$ are parallel vectors.

3.3.14. Prove that the ∞ norm on $\mathbb{R}^2$ does not come from an inner product. *Hint*: Look at Exercise 3.1.13.

3.3.15. Can formula (3.11) be used to define an inner product for (*a*) the 1 norm $\| \mathbf{v} \|_1$ on $\mathbb{R}^2$? (*b*) the ∞ norm $\| \mathbf{v} \|_\infty$ on $\mathbb{R}^2$?

♢ 3.3.16. Prove that $\lim_{p \to \infty} \| \mathbf{v} \|_p = \| \mathbf{v} \|_\infty$ for all $\mathbf{v} \in \mathbb{R}^2$.

♢ 3.3.17. Justify the triangle inequality for (a) the L^1 norm (3.31); (*b*) the L^∞ norm (3.32).

♢ 3.3.18. Let $w(x) > 0$ for $a \le x \le b$ be a weight function. (*a*) Prove that $\| f \|_{1,w} = \int_a^b | f(x) | \, w(x) \, dx$ defines a norm on $C^0[a,b]$, called the *weighted* L^1 *norm*. (*b*) Do the same for the *weighted* L^∞ *norm* $\| f \|_{\infty,w} = \max\big\{ | f(x) | \, w(x) \; : \; a \le x \le b \big\}$.

3.3.19. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two different norms on a vector space V. (*a*) Prove that $\|\mathbf{v}\| = \max\big\{\|\mathbf{v}\|_1, \|\mathbf{v}\|_2\big\}$ defines a norm on V. (*b*) Does $\|\mathbf{v}\| = \min\big\{\|\mathbf{v}\|_1, \|\mathbf{v}\|_2\big\}$ define a norm? (*c*) Does the arithmetic mean $\|\mathbf{v}\| = \frac{1}{2}\big(\|\mathbf{v}\|_1 + \|\mathbf{v}\|_2\big)$ define a norm?
(*d*) Does the geometric mean $\|\mathbf{v}\| = \sqrt{\|\mathbf{v}\|_1\,\|\mathbf{v}\|_2}$ define a norm?

Unit Vectors

Let V be a normed vector space. The elements $\mathbf{u} \in V$ that have unit norm, $\|\mathbf{u}\| = 1$, play a special role, and are known as *unit vectors* (or functions or elements). The following easy lemma shows how to construct a unit vector pointing in the same direction as any given nonzero vector.

Lemma 3.14. If $\mathbf{v} \neq \mathbf{0}$ is any nonzero vector, then the vector $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$ obtained by dividing $\mathbf{v}$ by its norm is a unit vector parallel to $\mathbf{v}$.

Proof: We compute, making use of the homogeneity property of the norm and the fact that $\|\mathbf{v}\|$ is a scalar,

$$\|\mathbf{u}\| = \left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\| = \frac{\|\mathbf{v}\|}{\|\mathbf{v}\|} = 1. \qquad \textit{Q.E.D.}$$

Example 3.15. The vector $\mathbf{v} = (1, -2)^T$ has length $\|\mathbf{v}\|_2 = \sqrt{5}$ with respect to the standard Euclidean norm. Therefore, the unit vector pointing in the same direction is

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|_2} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix}.$$

On the other hand, for the 1 norm, $\|\mathbf{v}\|_1 = 3$, and so

$$\widetilde{\mathbf{u}} = \frac{\mathbf{v}}{\|\mathbf{v}\|_1} = \frac{1}{3} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}$$

is the unit vector parallel to $\mathbf{v}$ in the 1 norm. Finally, $\|\mathbf{v}\|_\infty = 2$, and hence the corresponding unit vector for the ∞ norm is

$$\widehat{\mathbf{u}} = \frac{\mathbf{v}}{\|\mathbf{v}\|_\infty} = \frac{1}{2} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -1 \end{pmatrix}.$$

Thus, the notion of unit vector will depend upon which norm is being used.

Example 3.16. Similarly, on the interval $[0, 1]$, the quadratic polynomial $p(x) = x^2 - \frac{1}{2}$ has L^2 norm

$$\|p\|_2 = \sqrt{\int_0^1 \left(x^2 - \tfrac{1}{2}\right)^2 dx} = \sqrt{\int_0^1 \left(x^4 - x^2 + \tfrac{1}{4}\right) dx} = \sqrt{\frac{7}{60}}.$$

Therefore, $u(x) = \dfrac{p(x)}{\|p\|} = \sqrt{\frac{60}{7}}\, x^2 - \sqrt{\frac{15}{7}}$ is a "unit polynomial", $\|u\|_2 = 1$, which is "parallel" to (or, more precisely, a scalar multiple of) the polynomial p. On the other hand, for the L^∞ norm,

$$\|p\|_\infty = \max\left\{ \left| x^2 - \tfrac{1}{2} \right| \;\middle|\; 0 \le x \le 1 \right\} = \tfrac{1}{2},$$

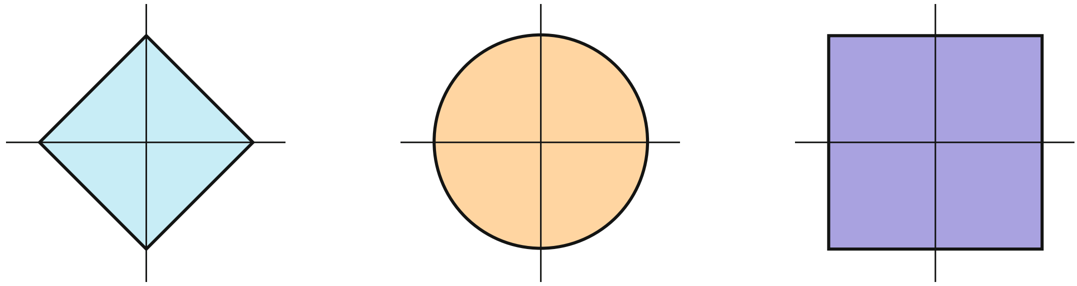

Figure 3.5. Unit Balls and Spheres for 1, 2, and ∞ Norms in $\mathbb{R}^2$.

and hence, in this case, $\widetilde{u}(x) = 2\,p(x) = 2\,x^2 - 1$ is the corresponding unit polynomial.

The *unit sphere* for the given norm is defined as the set of all unit vectors

$$S_1 = \{\, \|\mathbf{u}\| = 1 \,\}, \qquad \text{while} \qquad S_r = \{\, \|\mathbf{u}\| = r \,\} \tag{3.34}$$

is the sphere of radius $r \geq 0$. Thus, the unit sphere for the Euclidean norm on $\mathbb{R}^n$ is the usual round sphere

$$S_1 = \{\, \|\mathbf{x}\|^2 = x_1^2 + x_2^2 + \cdots + x_n^2 = 1 \,\}.$$

The unit sphere for the ∞ norm is the surface of a unit cube:

$$S_1 = \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{c} |\,x_i\,| \leq 1, \ i = 1, \ldots, n, \quad \text{and either} \\ x_1 = \pm 1 \text{ or } x_2 = \pm 1 \text{ or } \ \ldots \ \text{ or } x_n = \pm 1 \end{array} \right\}.$$

For the 1 norm,

$$S_1 = \{\, \mathbf{x} \in \mathbb{R}^n \mid |\,x_1\,| + |\,x_2\,| + \cdots + |\,x_n\,| = 1 \,\}$$

is the unit diamond in two dimensions, unit octahedron in three dimensions, and unit *cross polytope* in general. See Figure 3.5 for the two-dimensional pictures.

In all cases, the *unit ball* $B_1 = \{\, \|\mathbf{u}\| \leq 1 \,\}$ consists of all vectors of norm less than or equal to 1, and has the unit sphere as its boundary. If V is a finite-dimensional normed vector space, then the unit ball B_1 is a *compact* subset, meaning that it is closed and bounded. This basic topological fact, which is *not* true in infinite-dimensional normed spaces, underscores the distinction between finite-dimensional vector analysis and the vastly more complicated infinite-dimensional realm.

Exercises

3.3.20. Find a unit vector in the same direction as $\mathbf{v} = (1, 2, -3)^T$ for (*a*) the Euclidean norm, (*b*) the weighted norm $\|\mathbf{v}\|^2 = 2v_1^2 + v_2^2 + \frac{1}{3}v_3^2$, (*c*) the 1 norm, (*d*) the ∞ norm, (*e*) the norm based on the inner product $2v_1w_1 - v_1w_2 - v_2w_1 + 2v_2w_2 - v_2w_3 - v_3w_2 + 2v_3w_3$.

3.3.21. Show that, for every choice of given angles θ, ϕ, and ψ, the following are unit vectors in the Euclidean norm: (*a*) $(\cos\theta\cos\phi, \cos\theta\sin\phi, \sin\theta)^T$. (*b*) $\frac{1}{\sqrt{2}}(\cos\theta, \sin\theta, \cos\phi, \sin\phi)^T$. (*c*) $(\cos\theta\cos\phi\cos\psi, \cos\theta\cos\phi\sin\psi, \cos\theta\sin\phi, \sin\theta)^T$.

3.3.22. How many unit vectors are parallel to a given vector $\mathbf{v} \neq \mathbf{0}$? (*a*) 1, (*b*) 2, (*c*) 3, (*d*) ∞, (*e*) depends on the norm. Explain your answer.

3.3.23. Plot the unit circle (sphere) for (*a*) the weighted norm $\|\mathbf{v}\| = \sqrt{v_1^2 + 4v_2^2}$; (*b*) the norm based on the inner product (3.9); (*c*) the norm of Exercise 3.3.9.

3.3.24. Draw the unit circle for each norm in Exercise 3.3.10.

3.3.25. Sketch the unit sphere $S_1 \subset \mathbb{R}^3$ for (*a*) the L^1 norm, (*b*) the L^∞ norm, (*c*) the weighted norm $\|\mathbf{v}\|^2 = 2v_1^2 + v_2^2 + 3v_3^2$, (*d*) $\|\mathbf{v}\| = \max\bigl\{ |v_1+v_2|, |v_1+v_3|, |v_2+v_3| \bigr\}$.

3.3.26. Let $\mathbf{v} \neq \mathbf{0}$ be any nonzero vector in a normed vector space V. Show how to construct a new norm on V that changes $\mathbf{v}$ into a unit vector.

3.3.27. *True or false*: Two norms on a vector space have the same unit sphere if and only if they are the same norm.

3.3.28. Find the unit function that is a constant multiple of the function $f(x) = x - \frac{1}{3}$ with respect to the (*a*) L^1 norm on $[0,1]$; (*b*) L^2 norm on $[0,1]$; (*c*) L^∞ norm on $[0,1]$; (*d*) L^1 norm on $[-1,1]$; (*e*) L^2 norm on $[-1,1]$; (*f*) L^∞ norm on $[-1,1]$.

3.3.29. For which norms is the constant function $f(x) \equiv 1$ a unit function?
(*a*) L^1 norm on $[0,1]$; (*b*) L^2 norm on $[0,1]$; (*c*) L^∞ norm on $[0,1]$;
(*d*) L^1 norm on $[-1,1]$; (*e*) L^2 norm on $[-1,1]$; (*f*) L^∞ norm on $[-1,1]$;
(*g*) L^1 norm on $\mathbb{R}$; (*h*) L^2 norm on $\mathbb{R}$; (*i*) L^∞ norm on $\mathbb{R}$.

♢ 3.3.30. A subset $S \subset \mathbb{R}^n$ is called *convex* if, for all $\mathbf{x}, \mathbf{y} \in S$, the line segment joining $\mathbf{x}$ to $\mathbf{y}$ is also in S, i.e., $t\mathbf{x} + (1-t)\mathbf{y} \in S$ for all $0 \leq t \leq 1$. Prove that the unit ball is a convex subset of a normed vector space. Is the unit sphere convex?

Equivalence of Norms

While there are many different types of norms, in a finite-dimensional vector space they are all more or less equivalent. "Equivalence" does not mean that they assume the same values, but rather that they are, in a certain sense, always close to one another, and so, for many analytical purposes, may be used interchangeably. As a consequence, we may be able to simplify the analysis of a problem by choosing a suitably adapted norm; examples can be found in Chapter 9.

Theorem 3.17. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be any two norms on $\mathbb{R}^n$. Then there exist positive constants $0 < c^\star \leq C^\star$ such that

$$c^\star \|\mathbf{v}\|_1 \leq \|\mathbf{v}\|_2 \leq C^\star \|\mathbf{v}\|_1 \qquad \text{for every} \qquad \mathbf{v} \in \mathbb{R}^n. \tag{3.35}$$

Proof: We just sketch the basic idea, leaving the details to a more rigorous real analysis course, cf. [**19**; §7.6]. We begin by noting that a norm defines a continuous real-valued function $f(\mathbf{v}) = \|\mathbf{v}\|$ on $\mathbb{R}^n$. (Continuity is, in fact, a consequence of the triangle inequality.) Let $S_1 = \bigl\{ \|\mathbf{u}\|_1 = 1 \bigr\}$ denote the unit sphere of the first norm. Every continuous function defined on a compact set achieves both a maximum and a minimum value. Thus, restricting the second norm function to the unit sphere S_1 of the first norm, we can set

$$c^\star = \min\{ \|\mathbf{u}\|_2 \mid \mathbf{u} \in S_1 \}, \qquad C^\star = \max\{ \|\mathbf{u}\|_2 \mid \mathbf{u} \in S_1 \}. \tag{3.36}$$

Moreover, $0 < c^\star \leq C^\star < \infty$, with equality holding if and only if the norms are the same. The minimum and maximum (3.36) will serve as the constants in the desired inequalities (3.35). Indeed, by definition,

$$c^\star \leq \|\mathbf{u}\|_2 \leq C^\star \qquad \text{when} \qquad \|\mathbf{u}\|_1 = 1, \tag{3.37}$$

which proves that (3.35) is valid for all unit vectors $\mathbf{v} = \mathbf{u} \in S_1$. To prove the inequalities in general, assume $\mathbf{v} \neq \mathbf{0}$. (The case $\mathbf{v} = \mathbf{0}$ is trivial.) Lemma 3.14 says that $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|_1 \in S_1$

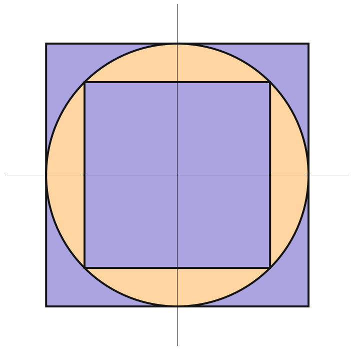

Figure 3.6. Equivalence of the ∞ and 2 Norms.

is a unit vector in the first norm: $\|\mathbf{u}\|_1 = 1$. Moreover, by the homogeneity property of the norm, $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 / \|\mathbf{v}\|_1$. Substituting into (3.37) and clearing denominators completes the proof of (3.35). *Q.E.D.*

Example 3.18. Consider the Euclidean norm $\|\cdot\|_2$ and the max norm $\|\cdot\|_\infty$ on $\mathbb{R}^n$. According to (3.36), the bounding constants are found by minimizing and maximizing $\|\mathbf{u}\|_\infty = \max\{\,|u_1|,\ldots,|u_n|\,\}$ over all unit vectors $\|\mathbf{u}\|_2 = 1$ on the (round) unit sphere. The maximal value is achieved at the poles $\pm\mathbf{e}_k$, with $\|\pm\mathbf{e}_k\|_\infty = C^\star = 1$. The minimal value is attained at the points $\left(\pm\frac{1}{\sqrt{n}},\ldots,\pm\frac{1}{\sqrt{n}}\right)$, whereby $c^\star = \frac{1}{\sqrt{n}}$. Therefore,

$$\frac{1}{\sqrt{n}}\,\|\mathbf{v}\|_2 \le \|\mathbf{v}\|_\infty \le \|\mathbf{v}\|_2. \tag{3.38}$$

We can interpret these inequalities as follows. Suppose $\mathbf{v}$ is a vector lying on the unit sphere in the Euclidean norm, so $\|\mathbf{v}\|_2 = 1$. Then (3.38) tells us that its ∞ norm is bounded from above and below by $\frac{1}{\sqrt{n}} \le \|\mathbf{v}\|_\infty \le 1$. Therefore, the Euclidean unit sphere sits inside the ∞ norm unit sphere and outside the ∞ norm sphere of radius $\frac{1}{\sqrt{n}}$. Figure 3.6 illustrates the two-dimensional situation: the unit circle is inside the unit square, and contains the square of size $\frac{1}{\sqrt{2}}$.

One significant consequence of the equivalence of norms is that, in $\mathbb{R}^n$, convergence is independent of the norm. The following are all equivalent to the standard notion of convergence of a sequence $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \mathbf{u}^{(3)}, \ \ldots$ of vectors in $\mathbb{R}^n$:

(*a*) the vectors converge: $\mathbf{u}^{(k)} \longrightarrow \mathbf{u}^\star$:

(*b*) the individual coordinates all converge: $u_i^{(k)} \longrightarrow u_i^\star$ for $i = 1,\ldots,n$.

(*c*) the difference in norms goes to zero: $\|\mathbf{u}^{(k)} - \mathbf{u}^\star\| \longrightarrow 0$.

The last version, known as *convergence in norm*, does not depend on which norm is chosen. Indeed, the inequality (3.35) implies that if one norm goes to zero, so does any other norm. A consequence is that all norms on $\mathbb{R}^n$ induce the same topology — convergence of sequences, notions of open and closed sets, and so on. None of this is true in infinite-dimensional function space! A rigorous development of the underlying topological and analytical properties of compactness, continuity, and convergence is beyond the scope of this course. The motivated student is encouraged to consult a text in real analysis, e.g., [**19**, **68**], to find the relevant definitions, theorems, and proofs.

Example 3.19. Consider the infinite-dimensional vector space $C^0[0,1]$ consisting of all continuous functions on the interval $[0,1]$. The functions

$$f_n(x) = \begin{cases} 1 - nx, & 0 \le x \le \frac{1}{n}, \\ 0, & \frac{1}{n} \le x \le 1, \end{cases}$$

have identical L^∞ norms

$$\| f_n \|_\infty = \sup \{ \, | f_n(x) | \, | \; 0 \le x \le 1 \} = 1.$$

On the other hand, their L^2 norm

$$\| f_n \|_2 = \sqrt{\int_0^1 f_n(x)^2 \, dx} = \sqrt{\int_0^{1/n} (1 - nx)^2 \, dx} = \frac{1}{\sqrt{3n}}$$

goes to zero as $n \to \infty$. This example shows that there *is no* constant $C^\star$ such that

$$\| f \|_\infty \le C^\star \| f \|_2$$

for all $f \in C^0[0,1]$. Thus, the L^∞ and L^2 norms on $C^0[0,1]$ are not equivalent — there exist functions that have unit L^∞ norm, but arbitrarily small L^2 norm. Similar comparative results can be established for the other function space norms. Analysis and topology on function space is intimately linked to the underlying choice of norm.

Exercises

3.3.31. Check the validity of the inequalities (3.38) for the particular vectors
(*a*) $(1,-1)^T$, (*b*) $(1,2,3)^T$, (*c*) $(1,1,1,1)^T$, (*d*) $(1,-1,-2,-1,1)^T$.

3.3.32. Find all $\mathbf{v} \in \mathbb{R}^2$ such that
(*a*) $\| \mathbf{v} \|_1 = \| \mathbf{v} \|_\infty$, (*b*) $\| \mathbf{v} \|_1 = \| \mathbf{v} \|_2$, (*c*) $\| \mathbf{v} \|_2 = \| \mathbf{v} \|_\infty$, (*d*) $\| \mathbf{v} \|_\infty = \frac{1}{\sqrt{2}} \| \mathbf{v} \|_2$.

3.3.33. How would you quantify the following statement: The norm of a vector is small if and only if all its entries are small.

3.3.34. Can you find an elementary proof of the inequalities $\| \mathbf{v} \|_\infty \le \| \mathbf{v} \|_2 \le \sqrt{n} \, \| \mathbf{v} \|_\infty$ for $\mathbf{v} \in \mathbb{R}^n$ directly from the formulas for the norms?

3.3.35. (*i*) Show the equivalence of the Euclidean norm and the 1 norm on $\mathbb{R}^n$ by proving $\| \mathbf{v} \|_2 \le \| \mathbf{v} \|_1 \le \sqrt{n} \, \| \mathbf{v} \|_2$. (*ii*) Verify that the vectors in Exercise 3.3.31 satisfy both inequalities. (*iii*) For which vectors $\mathbf{v} \in \mathbb{R}^n$ is (*a*) $\| \mathbf{v} \|_2 = \| \mathbf{v} \|_1$? (*b*) $\| \mathbf{v} \|_1 = \sqrt{n} \, \| \mathbf{v} \|_2$?

3.3.36. (*i*) Establish the equivalence inequalities (3.35) between the 1 and ∞ norms.
(*ii*) Verify them for the vectors in Exercise 3.3.31.
(*iii*) For which vectors $\mathbf{v} \in \mathbb{R}^n$ are your inequalities equality?

3.3.37. Let $\| \cdot \|_2$ denote the usual Euclidean norm on $\mathbb{R}^n$. Determine the constants in the norm equivalence inequalities $c^\star \, \| \mathbf{v} \| \le \| \mathbf{v} \|_2 \le C^\star \, \| \mathbf{v} \|$ for the following norms: (*a*) the weighted norm $\| \mathbf{v} \| = \sqrt{2 v_1^2 + 3 v_2^2}$, (*b*) the norm $\| \mathbf{v} \| = \max\big\{ \, | v_1 + v_2 |, | v_1 - v_2 | \, \big\}$.

3.3.38. Let $\| \cdot \|$ be a norm on $\mathbb{R}^n$. Prove that there is a constant $C > 0$ such that the entries of every $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T \in \mathbb{R}^n$ are all bounded, in absolute value, by $| v_i | \le C \, \| \mathbf{v} \|$.

3.3.39. Prove that if $[a,b]$ is a bounded interval and $f \in C^0[a,b]$, then $\| f \|_2 \le \sqrt{b-a} \, \| f \|_\infty$.

♡ 3.3.40. In this exercise, the indicated function norms are taken over all of $\mathbb{R}$.

(a) Let $f_n(x) = \begin{cases} 1, & -n \le x \le n, \\ 0, & \text{otherwise.} \end{cases}$ Prove that $\| f_n \|_\infty = 1$, but $\| f_n \|_2 \to \infty$ as $n \to \infty$.

(b) Explain why there is no constant C such that $\| f \|_2 \le C \| f \|_\infty$ for all functions f.

(c) Let $f_n(x) = \begin{cases} \sqrt{\dfrac{n}{2}}, & -\dfrac{1}{n} \le x \le \dfrac{1}{n}, \\ 0, & \text{otherwise.} \end{cases}$ Prove that $\| f_n \|_2 = 1$, but $\| f_n \|_\infty \to \infty$ as $n \to \infty$. Conclude that there is no constant C such that $\| f \|_\infty \le C \| f \|_2$.

(d) Construct similar examples that disprove the related inequalities

(i) $\| f \|_\infty \le C \| f \|_1$, ($ii$) $\| f \|_1 \le C \| f \|_2$, ($iii$) $\| f \|_2 \le C \| f \|_1$.

♡ 3.3.41. (a) Prove that the L^∞ and L^2 norms on the vector space $\mathrm{C}^0[-1,1]$ are not equivalent. *Hint*: Look at Exercise 3.3.40 for ideas. (b) Can you establish a bound in either direction, i.e., $\| f \|_\infty \le C \| f \|_2$ or $\| f \|_2 \le \widetilde{C} \| f \|_\infty$ for all $f \in \mathrm{C}^0[-1,1]$ for some positive constants $C, \widetilde{C}$? (c) Are the L^1 and L^∞ norms equivalent?

♢ 3.3.42. What does it mean if the constants defined in (3.36) are equal: $c^\star = C^\star$?

3.3.43. Suppose $\langle \mathbf{v}, \mathbf{w} \rangle_1$ and $\langle \mathbf{v}, \mathbf{w} \rangle_2$ are two inner products on the same vector space V. For which $\alpha, \beta \in \mathbb{R}$ is the linear combination $\langle \mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{w} \rangle_1 + \beta \langle \mathbf{v}, \mathbf{w} \rangle_2$ a legitimate inner product? *Hint*: The case $\alpha, \beta \ge 0$ is easy. However, some negative values are also permitted, and your task is to decide which.

♢ 3.3.44. Suppose $\| \cdot \|_1, \| \cdot \|_2$ are two norms on $\mathbb{R}^n$. Prove that the corresponding matrix norms satisfy $\widehat{c}^\star \| A \|_1 \le \| A \|_2 \le \widehat{C}^\star \| A \|_1$ for any $n \times n$ matrix A for some positive constants $0 < \widehat{c}^\star \le \widehat{C}^\star$.

Matrix Norms

Each norm on $\mathbb{R}^n$ will naturally induce a norm on the vector space $\mathcal{M}_{n \times n}$ of all $n \times n$ matrices. Roughly speaking, the matrix norm tells us how much a linear transformation stretches vectors relative to the given norm. Matrix norms will play an important role in Chapters 8 and 9, particularly in our analysis of linear iterative systems and iterative numerical methods for solving both linear and nonlinear systems.

We work exclusively with real $n \times n$ matrices in this section, although the results straightforwardly extend to complex matrices. We begin by fixing a norm $\| \cdot \|$ on $\mathbb{R}^n$. The norm may or may not come from an inner product — this is irrelevant as far as the construction goes.

Theorem 3.20. If $\| \cdot \|$ is any norm on $\mathbb{R}^n$, then the quantity

$$\| A \| = \max \{ \| A \mathbf{u} \| \mid \| \mathbf{u} \| = 1 \} \tag{3.39}$$

defines the norm of an $n \times n$ matrix $A \in \mathcal{M}_{n \times n}$, called the associated *natural matrix norm*.

Proof: First note that $\| A \| < \infty$, since the maximum is taken on a closed and bounded subset, namely the unit sphere $S_1 = \{ \| \mathbf{u} \| = 1 \}$ for the given norm. To show that (3.39) defines a norm, we need to verify the three basic axioms of Definition 3.12.

Non-negativity, $\| A \| \ge 0$, is immediate. Suppose $\| A \| = 0$. This means that, for every unit vector, $\| A \mathbf{u} \| = 0$, and hence $A \mathbf{u} = \mathbf{0}$ whenever $\| \mathbf{u} \| = 1$. If $\mathbf{0} \ne \mathbf{v} \in \mathbb{R}^n$ is any nonzero vector, then $\mathbf{u} = \mathbf{v}/r$, where $r = \| \mathbf{v} \|$, is a unit vector, so

$$A \mathbf{v} = A(r \mathbf{u}) = r A \mathbf{u} = \mathbf{0}. \tag{3.40}$$

Therefore, $A\mathbf{v} = \mathbf{0}$ for every $\mathbf{v} \in \mathbb{R}^n$, which implies that $A = \mathrm{O}$ is the zero matrix. This serves to prove the positivity property: $\| A \| = 0$ if and only if $A = \mathrm{O}$.

As for homogeneity, if $c \in \mathbb{R}$ is any scalar, then

$$\| c A \| = \max \{ \| c A \mathbf{u} \| \} = \max \{ | c | \, \| A \mathbf{u} \| \} = | c | \max \{ \| A \mathbf{u} \| \} = | c | \, \| A \|.$$

Finally, to prove the triangle inequality, we use the fact that the maximum of the sum of quantities is bounded by the sum of their individual maxima. Therefore, since the norm on $\mathbb{R}^n$ satisfies the triangle inequality,

$$\begin{aligned} \| A + B \| &= \max \{ \| A \mathbf{u} + B \mathbf{u} \| \} \leq \max \{ \| A \mathbf{u} \| + \| B \mathbf{u} \| \} \\ &\leq \max \{ \| A \mathbf{u} \| \} + \max \{ \| B \mathbf{u} \| \} = \| A \| + \| B \|. \end{aligned}$$

Q.E.D.

The property that distinguishes a matrix norm from a generic norm on the space of matrices is the fact that it also obeys a very useful *product inequality*.

Theorem 3.21. A natural matrix norm satisfies

$$\| A \mathbf{v} \| \leq \| A \| \, \| \mathbf{v} \|, \qquad \text{for all} \qquad A \in \mathcal{M}_{n \times n}, \quad \mathbf{v} \in \mathbb{R}^n. \tag{3.41}$$

Furthermore,

$$\| A B \| \leq \| A \| \, \| B \|, \qquad \text{for all} \qquad A, B \in \mathcal{M}_{n \times n}. \tag{3.42}$$

Proof: Note first that, by definition $\| A \mathbf{u} \| \leq \| A \|$ for all unit vectors $\| \mathbf{u} \| = 1$. Then, letting $\mathbf{v} = r\,\mathbf{u}$ where $\mathbf{u}$ is a unit vector and $r = \| \mathbf{v} \|$, we have

$$\| A \mathbf{v} \| = \| A (r\,\mathbf{u}) \| = r \, \| A \mathbf{u} \| \leq r \, \| A \| = \| \mathbf{v} \| \, \| A \|,$$

proving the first inequality. To prove the second, we apply the first, replacing $\mathbf{v}$ by $B\,\mathbf{u}$:

$$\begin{aligned} \| A B \| &= \max \{ \| A B \mathbf{u} \| \} = \max \{ \| A (B \mathbf{u}) \| \} \\ &\leq \max \{ \| A \| \, \| B \mathbf{u} \| \} = \| A \| \max \{ \| B \mathbf{u} \| \} = \| A \| \, \| B \|. \end{aligned}$$

Q.E.D.

Remark. In general, a norm on the vector space of $n \times n$ matrices is called a *matrix norm* if it also satisfies the multiplicative inequality (3.42). Most, but not all, matrix norms used in applications come from norms on the underlying vector space.

The multiplicative inequality (3.42) implies, in particular, that $\| A^2 \| \leq \| A \|^2$; equality is not necessarily valid. More generally:

Proposition 3.22. If A is a square matrix, then $\| A^k \| \leq \| A \|^k$.

Let us determine the explicit formula for the matrix norm induced by the ∞ norm

$$\| \mathbf{v} \|_\infty = \max \{ | v_1 |, \ \ldots \ , | v_n | \}.$$

The corresponding formula for the 1 norm is left as Exercise 3.3.48. The formula for the Euclidean matrix norm (2 norm) will be deferred until Theorem 8.71.

Definition 3.23. The i^{th} *absolute row sum* of a matrix A is the sum of the absolute values of the entries in the i^{th} row:

$$s_i = |\,a_{i1}\,| + \cdots + |\,a_{in}\,| = \sum_{j=1}^{n} |\,a_{ij}\,|. \tag{3.43}$$

Proposition 3.24. The ∞ *matrix norm* of a matrix A is equal to its maximal absolute row sum:

$$\|\,A\,\|_\infty = \max\{s_1,\ldots,s_n\} = \max\left\{ \sum_{j=1}^{n} |\,a_{ij}\,| \;\middle|\; 1 \le i \le n \right\}. \tag{3.44}$$

Proof: Let $s = \max\{s_1,\ldots,s_n\}$ denote the right-hand side of (3.44). Given any $\mathbf{v} \in \mathbb{R}^n$, we compute the ∞ norm of the image vector $A\mathbf{v}$:

$$\begin{aligned}
\|\,A\mathbf{v}\,\|_\infty &= \max\left\{ \left| \sum_{j=1}^{n} a_{ij} v_j \right| \right\} \le \max\left\{ \sum_{j=1}^{n} |\,a_{ij} v_j\,| \right\} \\
&\le \max\left\{ \sum_{j=1}^{n} |\,a_{ij}\,| \right\} \max\left\{\,|\,v_j\,|\,\right\} = s\,\|\,\mathbf{v}\,\|_\infty.
\end{aligned}$$

In particular, by specializing to a unit vector, $\|\,\mathbf{v}\,\|_\infty = 1$, we deduce that $\|\,A\,\|_\infty \le s$.

On the other hand, suppose the maximal absolute row sum occurs at row i, so

$$s_i = \sum_{j=1}^{n} |\,a_{ij}\,| = s. \tag{3.45}$$

Let $\mathbf{u} \in \mathbb{R}^n$ be the specific vector that has the following entries: $u_j = +1$ if $a_{ij} \ge 0$, while $u_j = -1$ if $a_{ij} < 0$. Then $\|\,\mathbf{u}\,\|_\infty = 1$. Moreover, since $a_{ij}\,u_j = |\,a_{ij}\,|$, the i^{th} entry of $A\mathbf{u}$ is equal to the i^{th} absolute row sum (3.45). This implies that $\|\,A\,\|_\infty \ge \|\,A\mathbf{u}\,\|_\infty \ge s$. *Q.E.D.*

Example 3.25. Consider the symmetric matrix $A = \begin{pmatrix} \frac12 & -\frac13 \\ -\frac13 & \frac14 \end{pmatrix}$. Its two absolute row sums are $\left|\,\frac12\,\right| + \left|\,-\frac13\,\right| = \frac56$, $\left|\,-\frac13\,\right| + \left|\,\frac14\,\right| = \frac{7}{12}$, so $\|\,A\,\|_\infty = \max\left\{\,\frac56, \frac{7}{12}\,\right\} = \frac56$.

Exercises

3.3.45. Compute the ∞ matrix norm of the following matrices.

(a) $\begin{pmatrix} \frac12 & \frac14 \\ \frac13 & \frac16 \end{pmatrix}$, (b) $\begin{pmatrix} \frac53 & \frac43 \\ -\frac76 & -\frac56 \end{pmatrix}$, (c) $\begin{pmatrix} 0 & .1 & .8 \\ -.1 & 0 & .1 \\ -.8 & -.1 & 0 \end{pmatrix}$, (d) $\begin{pmatrix} \frac13 & 0 & 0 \\ -\frac13 & 0 & \frac13 \\ 0 & \frac23 & \frac13 \end{pmatrix}$.

3.3.46. Find a matrix A such that $\|\,A^2\,\|_\infty \ne \|\,A\,\|_\infty^2$.

3.3.47. *True or false*: If $B = S^{-1} A S$ are similar matrices, then $\|\,B\,\|_\infty = \|\,A\,\|_\infty$.

♢ 3.3.48. (*i*) Find an explicit formula for the 1 matrix norm $\|\,A\,\|_1$.
(*ii*) Compute the 1 matrix norm of the matrices in Exercise 3.3.45.

3.3.49. Prove directly from the axioms of Definition 3.12 that (3.44) defines a norm on the space of $n \times n$ matrices.

3.3.50. Let $A = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$. Compute the natural matrix norm $\| A \|$ for (*a*) the weighted ∞ norm $\| \mathbf{v} \| = \max\{ 2 | v_1 |, 3 | v_2 | \}$; (*b*) the weighted 1 norm $\| \mathbf{v} \| = 2 | v_1 | + 3 | v_2 |$.

♡ 3.3.51. The *Frobenius norm* of an $n \times n$ matrix A is defined as $\| A \|_F = \sqrt{\sum_{i,j=1}^{n} a_{ij}^2}\,.$ Prove that this defines a matrix norm by checking the three norm axioms plus the multiplicative inequality (3.42).

3.3.52. Explain why $\| A \| = \max | a_{ij} |$ defines a norm on the space of $n \times n$ matrices. Show by example that this is *not* a matrix norm, i.e., (3.42) is not necessarily valid.

3.4 Positive Definite Matrices

Let us now return to the study of inner products and fix our attention on the finite-dimensional situation. Our immediate goal is to determine the most general inner product that can be placed on the finite-dimensional vector space $\mathbb{R}^n$. The answer will lead us to the important class of positive definite matrices, which appear in a wide range of applications, including minimization problems, mechanics, electrical circuits, differential equations, statistics, and numerical methods. Moreover, their infinite-dimensional counterparts, positive definite linear operators, govern most boundary value problems arising in continuum physics and engineering.

Suppose we are given an inner product $\langle \mathbf{x} , \mathbf{y} \rangle$ between vectors $\mathbf{x} = (x_1 \; x_2 \; \ldots \; x_n)^T$ and $\mathbf{y} = (y_1 \; y_2 \; \ldots \; y_n)^T$ in $\mathbb{R}^n$. Our goal is to determine its explicit formula. We begin by writing the vectors in terms of the standard basis vectors (2.17):

$$\mathbf{x} = x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n = \sum_{i=1}^{n} x_i \mathbf{e}_i, \qquad \mathbf{y} = y_1 \mathbf{e}_1 + \cdots + y_n \mathbf{e}_n = \sum_{j=1}^{n} y_j \mathbf{e}_j. \tag{3.46}$$

To evaluate their inner product, we will appeal to the three basic axioms. We first employ bilinearity to expand

$$\langle \mathbf{x} , \mathbf{y} \rangle = \left\langle \sum_{i=1}^{n} x_i \mathbf{e}_i \, , \, \sum_{j=1}^{n} y_j \mathbf{e}_j \right\rangle = \sum_{i,j=1}^{n} x_i \, y_j \langle \mathbf{e}_i , \mathbf{e}_j \rangle.$$

Therefore,

$$\langle \mathbf{x} , \mathbf{y} \rangle = \sum_{i,j=1}^{n} k_{ij} \, x_i \, y_j = \mathbf{x}^T K \, \mathbf{y}, \tag{3.47}$$

where K denotes the $n \times n$ matrix of inner products of the basis vectors, with entries

$$k_{ij} = \langle \mathbf{e}_i , \mathbf{e}_j \rangle, \qquad i, j = 1, \ldots, n. \tag{3.48}$$

We conclude that any inner product must be expressed in the general *bilinear form* (3.47).

The two remaining inner product axioms will impose certain constraints on the inner product matrix K. Symmetry implies that

$$k_{ij} = \langle \mathbf{e}_i , \mathbf{e}_j \rangle = \langle \mathbf{e}_j , \mathbf{e}_i \rangle = k_{ji}, \qquad i, j = 1, \ldots, n.$$

Consequently, the inner product matrix K must be symmetric:

$$K = K^T.$$

Conversely, symmetry of K ensures symmetry of the bilinear form:

$$\langle\, \mathbf{x}\,,\mathbf{y}\,\rangle = \mathbf{x}^T K\,\mathbf{y} = (\mathbf{x}^T K\,\mathbf{y})^T = \mathbf{y}^T K^T \mathbf{x} = \mathbf{y}^T K\,\mathbf{x} = \langle\, \mathbf{y}\,,\mathbf{x}\,\rangle,$$

where the second equality follows from the fact that the quantity $\mathbf{x}^T K\,\mathbf{y}$ is a scalar, and hence equals its transpose.

The final condition for an inner product is positivity, which requires that

$$\|\,\mathbf{x}\,\|^2 = \langle\, \mathbf{x}\,,\mathbf{x}\,\rangle = \mathbf{x}^T K\,\mathbf{x} = \sum_{i,j=1}^{n} k_{ij}\, x_i\, x_j \geq 0 \qquad \text{for all} \qquad \mathbf{x} \in \mathbb{R}^n, \tag{3.49}$$

with equality if and only if $\mathbf{x} = \mathbf{0}$. The precise meaning of this positivity condition on the matrix K is not so immediately evident, and so will be encapsulated in a definition.

Definition 3.26. An $n \times n$ matrix K is called *positive definite* if it is symmetric, $K^T = K$, and satisfies the positivity condition

$$\mathbf{x}^T K\,\mathbf{x} > 0 \qquad \text{for all} \qquad \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n. \tag{3.50}$$

We will sometimes write $K > 0$ to mean that K is a positive definite matrix.

Warning. The condition $K > 0$ does *not* mean that all the entries of K are positive. There are many positive definite matrices that have some negative entries; see Example 3.28 below. Conversely, many symmetric matrices with all positive entries are not positive definite!

Remark. Although some authors allow non-symmetric matrices to be designated as positive definite, we will say that a matrix is positive definite *only* when it is symmetric. But, to underscore our convention and remind the casual reader, we will often include the superfluous adjective "symmetric" when speaking of positive definite matrices.

Our preliminary analysis has resulted in the following general characterization of inner products on a finite-dimensional vector space.

Theorem 3.27. Every inner product on $\mathbb{R}^n$ is given by

$$\langle\, \mathbf{x}\,,\mathbf{y}\,\rangle = \mathbf{x}^T K\,\mathbf{y} \qquad \text{for} \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \tag{3.51}$$

where K is a symmetric, positive definite $n \times n$ matrix.

Given a symmetric matrix K, the homogeneous quadratic polynomial

$$q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x} = \sum_{i,j=1}^{n} k_{ij}\, x_i\, x_j, \tag{3.52}$$

is known as a *quadratic form*† on $\mathbb{R}^n$. The quadratic form is called *positive definite* if

$$q(\mathbf{x}) > 0 \qquad \text{for all} \qquad 0 \neq \mathbf{x} \in \mathbb{R}^n. \tag{3.53}$$

So the quadratic form (3.52) is positive definite if and only if its coefficient matrix K is.

---

† Exercise 3.4.15 shows that the coefficient matrix K in any quadratic form can be taken to be symmetric without any loss of generality.

Example 3.28. Even though the symmetric matrix $K = \begin{pmatrix} 4 & -2 \\ -2 & 3 \end{pmatrix}$ has two negative entries, it is, nevertheless, a positive definite matrix. Indeed, the corresponding quadratic form

$$q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x} = 4\,x_1^2 - 4\,x_1\,x_2 + 3\,x_2^2 = (2\,x_1 - x_2)^2 + 2\,x_2^2 \geq 0$$

is a sum of two non-negative quantities. Moreover, $q(\mathbf{x}) = 0$ if and only if both $2\,x_1 - x_2 = 0$ and $x_2 = 0$, which implies $x_1 = 0$ also. This proves $q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$, and hence K is indeed a positive definite matrix. The corresponding inner product on $\mathbb{R}^2$ is

$$\langle\, \mathbf{x}\,, \mathbf{y}\,\rangle = (\,x_1\;\, x_2\,) \begin{pmatrix} 4 & -2 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 4\,x_1\,y_1 - 2\,x_1\,y_2 - 2\,x_2\,y_1 + 3\,x_2\,y_2.$$

On the other hand, despite the fact that $K = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ has all positive entries, it is *not* a positive definite matrix. Indeed, writing out

$$q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x} = x_1^2 + 4\,x_1\,x_2 + x_2^2,$$

we find, for instance, that $q(1,-1) = -2 < 0$, violating positivity. These two simple examples should be enough to convince the reader that the problem of determining whether a given symmetric matrix is positive definite is not completely elementary.

Example 3.29. By definition, a general symmetric 2×2 matrix $K = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is positive definite if and only if the associated quadratic form satisfies

$$q(\mathbf{x}) = a\,x_1^2 + 2\,b\,x_1\,x_2 + c\,x_2^2 > 0 \qquad \text{for all} \qquad \mathbf{x} \neq \mathbf{0}. \tag{3.54}$$

Analytic geometry tells us that this is the case if and only if

$$a > 0, \qquad\qquad a\,c - b^2 > 0, \tag{3.55}$$

i.e., the quadratic form has positive leading coefficient and positive determinant (or negative discriminant). A direct proof of this well-known fact will appear shortly.

With a little practice, it is not difficult to read off the coefficient matrix K from the explicit formula for the quadratic form (3.52).

Example 3.30. Consider the quadratic form

$$q(x,y,z) = x^2 + 4\,x\,y + 6\,y^2 - 2\,x\,z + 9\,z^2$$

depending upon three variables. The corresponding coefficient matrix is

$$K = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 6 & 0 \\ -1 & 0 & 9 \end{pmatrix} \qquad \text{whereby} \qquad q(x,y,z) = (\,x \quad y \quad z\,) \begin{pmatrix} 1 & 2 & -1 \\ 2 & 6 & 0 \\ -1 & 0 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Note that the squared terms in q contribute directly to the diagonal entries of K, while the mixed terms are split in half to give the symmetric off-diagonal entries. As a challenge, the reader might wish to try proving that this particular matrix is positive definite by establishing positivity of the quadratic form: $q(x,y,z) > 0$ for all nonzero $(\,x, y, z\,)^T \in \mathbb{R}^3$. Later, we will devise a simple, systematic test for positive definiteness.

Slightly more generally, a quadratic form and its associated symmetric coefficient matrix are called *positive semi-definite* if

$$q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x} \geq 0 \qquad \text{for all} \qquad \mathbf{x} \in \mathbb{R}^n, \tag{3.56}$$

in which case we write $K \geq 0$. A positive semi-definite matrix may have *null directions*, meaning non-zero vectors $\mathbf{z}$ such that $q(\mathbf{z}) = \mathbf{z}^T K \mathbf{z} = \mathbf{0}$. Clearly, every nonzero vector $\mathbf{z} \in \ker K$ that lies in the coefficient matrix's kernel defines a null direction, but there may be others. A positive definite matrix is not allowed to have null directions, and so $\ker K = \{\mathbf{0}\}$. Recalling Proposition 2.42, we deduce that all positive definite matrices are nonsingular. The converse, however, is *not* valid; many symmetric, nonsingular matrices fail to be positive definite.

Proposition 3.31. If a matrix is positive definite, then it is nonsingular.

Example 3.32. The matrix $K = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ is positive semi-definite, but not positive definite. Indeed, the associated quadratic form

$$q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x} = x_1^2 - 2x_1 x_2 + x_2^2 = (x_1 - x_2)^2 \geq 0$$

is a perfect square, and so clearly non-negative. However, the elements of $\ker K$, namely the scalar multiples of the vector $(1,1)^T$, define null directions: $q(c,c) = 0$.

In a similar fashion, a quadratic form $q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x}$ and its associated symmetric matrix K are called *negative semi-definite* if $q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ and *negative definite* if $q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$. A quadratic form is called *indefinite* if it is neither positive nor negative semi-definite, equivalently, if there exist points $\mathbf{x}_+$ where $q(\mathbf{x}_+) > 0$ and points $\mathbf{x}_-$ where $q(\mathbf{x}_-) < 0$. Details can be found in the exercises.

Only positive definite matrices define inner products. However, indefinite matrices play a fundamental role in Einstein's theory of special relativity, [**55**]. In particular, the quadratic form associated with the matrix

$$K = \begin{pmatrix} c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{namely} \quad q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x} = c^2 t^2 - x^2 - y^2 - z^2 \quad \text{where} \quad \mathbf{x} = \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix}, \tag{3.57}$$

with c representing the speed of light, is the so-called Minkowski "metric" on relativistic space-time $\mathbb{R}^4$. The null directions form the light cone; see Exercise 3.4.20.

Exercises

3.4.1. Which of the following 2×2 matrices are positive definite?
(a) $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, (b) $\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, (d) $\begin{pmatrix} 5 & 3 \\ 3 & -2 \end{pmatrix}$, (e) $\begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix}$, (f) $\begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}$.
In the positive definite cases, write down the formula for the associated inner product.

3.4.2. Let $K = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$. Prove that the associated quadratic form $q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x}$ is indefinite by finding a point $\mathbf{x}^+$ where $q(\mathbf{x}^+) > 0$ and a point $\mathbf{x}^-$ where $q(\mathbf{x}^-) < 0$.

♢ 3.4.3. (a) Prove that a diagonal matrix $D = \operatorname{diag}(c_1, c_2, \ldots, c_n)$ is positive definite if and only if all its diagonal entries are positive: $c_i > 0$. (b) Write down and identify the associated inner product.

3.4.4. Write out the Cauchy–Schwarz and triangle inequalities for the inner product defined in Example 3.28.

♢ 3.4.5. (a) Show that every diagonal entry of a positive definite matrix must be positive. (b) Write down a symmetric matrix with all positive diagonal entries that is not positive definite. (c) Find a nonzero matrix with one or more zero diagonal entries that is positive semi-definite.

3.4.6. Prove that if K is any positive definite matrix, then every positive scalar multiple cK, $c > 0$, is also positive definite.

♢ 3.4.7. (a) Show that if K and L are positive definite matrices, so is $K + L$. (b) Give an example of two matrices that are not positive definite whose sum is positive definite.

3.4.8. Find two positive definite matrices K and L whose product KL is not positive definite.

3.4.9. Write down a nonsingular symmetric matrix that is not positive or negative definite.

♢ 3.4.10. Let K be a nonsingular symmetric matrix. (a) Show that $\mathbf{x}^T K^{-1}\mathbf{x} = \mathbf{y}^T K\,\mathbf{y}$, where $K\mathbf{y} = \mathbf{x}$. (b) Prove that if K is positive definite, then so is K^{-1}.

♢ 3.4.11. Prove that an $n \times n$ symmetric matrix K is positive definite if and only if, for every $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^n$, the vectors $\mathbf{v}$ and $K\mathbf{v}$ meet at an acute Euclidean angle: $|\,\theta\,| < \frac{1}{2}\pi$.

♢ 3.4.12. Prove that the inner product associated with a positive definite quadratic form $q(\mathbf{x})$ is given by the *polarization formula* $\langle\, \mathbf{x}\,,\mathbf{y}\,\rangle = \frac{1}{2}\big[\,q(\mathbf{x}+\mathbf{y}) - q(\mathbf{x}) - q(\mathbf{y})\,\big]$.

3.4.13. (a) Is it possible for a quadratic form to be positive, $q(\mathbf{x}_+) > 0$, at only one point $\mathbf{x}_+ \in \mathbb{R}^n$? (b) Under what conditions is $q(\mathbf{x}_0) = 0$ at only one point?

♢ 3.4.14. (a) Let K and L be symmetric $n \times n$ matrices. Prove that $\mathbf{x}^T K\,\mathbf{x} = \mathbf{x}^T L\,\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$ if and only if $K = L$. (b) Find an example of two non-symmetric matrices $K \neq L$ such that $\mathbf{x}^T K\,\mathbf{x} = \mathbf{x}^T L\,\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

♢ 3.4.15. Suppose $q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} = \displaystyle\sum_{i,j=1}^{n} a_{ij}\,x_i\,x_j$ is a general quadratic form on $\mathbb{R}^n$, whose coefficient matrix A is not necessarily symmetric. Prove that $q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x}$, where $K = \frac{1}{2}(A + A^T)$ is a symmetric matrix. Therefore, we do not lose any generality by restricting our discussion to quadratic forms that are constructed from symmetric matrices.

3.4.16. (a) Show that a symmetric matrix N is negative definite if and only if $K = -N$ is positive definite. (b) Write down two explicit criteria that tell whether a 2×2 matrix $N = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is negative definite. (c) Use your criteria to check whether (i) $\begin{pmatrix} -1 & 1 \\ 1 & -2 \end{pmatrix}$, (ii) $\begin{pmatrix} -4 & -5 \\ -5 & -6 \end{pmatrix}$, (iii) $\begin{pmatrix} -3 & -1 \\ -1 & 2 \end{pmatrix}$ are negative definite.

3.4.17. Show that $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is a null direction for $K = \begin{pmatrix} 1 & -2 \\ -2 & 3 \end{pmatrix}$, but $\mathbf{x} \notin \ker K$.

3.4.18. Explain why an indefinite quadratic form necessarily has a non-zero null direction.

3.4.19. Let $K = K^T$. *True or false*: (a) If K admits a null direction, then $\ker K \neq \{\mathbf{0}\}$. (b) If K has no null directions, then K is either positive or negative definite.

♢ 3.4.20. In special relativity, light rays in Minkowski space-time $\mathbb{R}^n$ travel along the *light cone* which, by definition, consists of all null directions associated with an indefinite quadratic form $q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x}$. Find and sketch a picture of the light cone when the coefficient matrix K is (a) $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, (b) $\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$. **Remark.** In the physical universe, space-time is $n = 4$-dimensional, and K is given in (3.57), [**55**].

♢ 3.4.21. A function $f(\mathbf{x})$ on $\mathbb{R}^n$ is called *homogeneous* of *degree* k if $f(c\,\mathbf{x}) = c^k f(\mathbf{x})$ for all scalars c. (*a*) Given $\mathbf{a} \in \mathbb{R}^n$, show that the *linear form* $\ell(\mathbf{x}) = \mathbf{a}\cdot\mathbf{x} = a_1 x_1 + \cdots + a_n x_n$ is homogeneous of degree 1. (*b*) Show that the quadratic form $q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x} = \sum_{i,j=1}^{n} k_{ij}\, x_i\, x_j$ is homogeneous of degree 2.
(*c*) Find a homogeneous function of degree 2 on $\mathbb{R}^2$ that is not a quadratic form.

Gram Matrices

Symmetric matrices whose entries are given by inner products of elements of an inner product space will appear throughout this text. They are named after the nineteenth-century Danish mathematician Jørgen Gram — not the metric mass unit!

Definition 3.33. Let V be an inner product space, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. The associated *Gram matrix*

$$K = \begin{pmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle & \ldots & \langle \mathbf{v}_1, \mathbf{v}_n \rangle \\ \langle \mathbf{v}_2, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle & \ldots & \langle \mathbf{v}_2, \mathbf{v}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{v}_n, \mathbf{v}_1 \rangle & \langle \mathbf{v}_n, \mathbf{v}_2 \rangle & \ldots & \langle \mathbf{v}_n, \mathbf{v}_n \rangle \end{pmatrix} \tag{3.58}$$

is the $n \times n$ matrix whose entries are the inner products between the selected vector space elements.

Symmetry of the inner product implies symmetry of the Gram matrix:

$$k_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \langle \mathbf{v}_j, \mathbf{v}_i \rangle = k_{ji}, \qquad \text{and hence} \qquad K^T = K. \tag{3.59}$$

In fact, the most direct method for producing positive definite and semi-definite matrices is through the Gram matrix construction.

Theorem 3.34. All Gram matrices are positive semi-definite. The Gram matrix (3.58) is positive definite if and only if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent.

Proof: To prove positive (semi-)definiteness of K, we need to examine the associated quadratic form

$$q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x} = \sum_{i,j=1}^{n} k_{ij}\, x_i\, x_j .$$

Substituting the values (3.59) for the matrix entries, we obtain

$$q(\mathbf{x}) = \sum_{i,j=1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle\, x_i\, x_j .$$

Bilinearity of the inner product on V implies that we can assemble this summation into a single inner product

$$q(\mathbf{x}) = \left\langle \sum_{i=1}^{n} x_i \mathbf{v}_i\,,\ \sum_{j=1}^{n} x_j \mathbf{v}_j \right\rangle = \langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2 \geq 0, \quad \text{where} \quad \mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i$$

lies in the subspace of V spanned by the given vectors. This immediately proves that K is positive semi-definite.

Moreover, $q(\mathbf{x}) = \|\mathbf{v}\|^2 > 0$ as long as $\mathbf{v} \neq \mathbf{0}$. If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent, then

$$\mathbf{v} = x_1 \mathbf{v}_1 + \cdots + x_n \mathbf{v}_n = \mathbf{0} \qquad \text{if and only if} \qquad x_1 = \cdots = x_n = 0,$$

and hence $q(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$. This implies that $q(\mathbf{x})$ and hence K are positive definite. *Q.E.D.*

Example 3.35. Consider the vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 3 \\ 0 \\ 6 \end{pmatrix}$. For the standard Euclidean dot product on $\mathbb{R}^3$, the Gram matrix is

$$K = \begin{pmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 \end{pmatrix} = \begin{pmatrix} 6 & -3 \\ -3 & 45 \end{pmatrix}.$$

Since $\mathbf{v}_1, \mathbf{v}_2$ are linearly independent, $K > 0$. Positive definiteness implies that

$$q(x_1, x_2) = 6x_1^2 - 6x_1x_2 + 45x_2^2 > 0 \qquad \text{for all} \qquad (x_1, x_2) \neq \mathbf{0}.$$

Indeed, this can be checked directly, by using the criteria in (3.55).

On the other hand, for the weighted inner product

$$\langle \mathbf{v}, \mathbf{w} \rangle = 3v_1w_1 + 2v_2w_2 + 5v_3w_3, \tag{3.60}$$

the corresponding Gram matrix is

$$\widetilde{K} = \begin{pmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle \\ \langle \mathbf{v}_2, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle \end{pmatrix} = \begin{pmatrix} 16 & -21 \\ -21 & 207 \end{pmatrix}. \tag{3.61}$$

Since $\mathbf{v}_1, \mathbf{v}_2$ are still linearly independent (which, of course, does not depend upon which inner product is used), the matrix $\widetilde{K}$ is also positive definite.

In the case of the Euclidean dot product, the construction of the Gram matrix K can be directly implemented as follows. Given column vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$, let us form the $m \times n$ matrix $A = (\mathbf{v}_1 \; \mathbf{v}_2 \; \ldots \; \mathbf{v}_n)$. In view of the identification (3.2) between the dot product and multiplication of row and column vectors, the (i, j) entry of K is given as the product

$$k_{ij} = \mathbf{v}_i \cdot \mathbf{v}_j = \mathbf{v}_i^T \mathbf{v}_j$$

of the i^{th} row of the transpose A^T and the j^{th} column of A. In other words, the Gram matrix can be evaluated as a matrix product:

$$K = A^T A. \tag{3.62}$$

For the preceding Example 3.35,

$$A = \begin{pmatrix} 1 & 3 \\ 2 & 0 \\ -1 & 6 \end{pmatrix}, \qquad \text{and so} \qquad K = A^T A = \begin{pmatrix} 1 & 2 & -1 \\ 3 & 0 & 6 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 0 \\ -1 & 6 \end{pmatrix} = \begin{pmatrix} 6 & -3 \\ -3 & 45 \end{pmatrix}.$$

Theorem 3.34 implies that the Gram matrix (3.62) is positive definite if and only if the columns of A are linearly independent vectors. This implies the following result.

Proposition 3.36. Given an $m \times n$ matrix A, the following are equivalent:

(*a*) The $n \times n$ Gram matrix $K = A^T A$ is positive definite.

(*b*) A has linearly independent columns.

(*c*) $\operatorname{rank} A = n \leq m$.

(*d*) $\ker A = \{0\}$.

Changing the underlying inner product will, of course, change the Gram matrix. As noted in Theorem 3.27, every inner product on $\mathbb{R}^m$ has the form

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T C \mathbf{w} \qquad \text{for} \qquad \mathbf{v}, \mathbf{w} \in \mathbb{R}^m, \tag{3.63}$$

where $C > 0$ is a symmetric, positive definite $m \times m$ matrix. Therefore, given n vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$, the entries of the Gram matrix with respect to this inner product are

$$k_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \mathbf{v}_i^T C \mathbf{v}_j.$$

If, as above, we assemble the column vectors into an $m \times n$ matrix $A = (\mathbf{v}_1 \; \mathbf{v}_2 \; \ldots \; \mathbf{v}_n)$, then the Gram matrix entry k_{ij} is obtained by multiplying the i^{th} row of A^T by the j^{th} column of the product matrix $C A$. Therefore, the Gram matrix based on the alternative inner product (3.63) is given by

$$K = A^T C A. \tag{3.64}$$

Theorem 3.34 immediately implies that K is positive definite — provided that the matrix A has rank n.

Theorem 3.37. Suppose A is an $m \times n$ matrix with linearly independent columns. Suppose C is any positive definite $m \times m$ matrix. Then the Gram matrix $K = A^T C A$ is a positive definite $n \times n$ matrix.

The Gram matrices constructed in (3.64) arise in a wide variety of applications, including least squares approximation theory (cf. Chapter 5), and mechanical structures and electrical circuits (cf. Chapters 6 and 10). In the majority of applications, $C = \operatorname{diag}(c_1, \ldots, c_m)$ is a diagonal positive definite matrix, which requires it to have strictly positive diagonal entries $c_i > 0$. This choice corresponds to a weighted inner product (3.10) on $\mathbb{R}^m$.

Example 3.38. Returning to the situation of Example 3.35, the weighted inner product (3.60) corresponds to the diagonal positive definite matrix $C = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix}$. Therefore, the weighted Gram matrix (3.64) based on the vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 3 \\ 0 \\ 6 \end{pmatrix}$, is

$$\widetilde{K} = A^T C A = \begin{pmatrix} 1 & 2 & -1 \\ 3 & 0 & 6 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 0 \\ -1 & 6 \end{pmatrix} = \begin{pmatrix} 16 & -21 \\ -21 & 207 \end{pmatrix},$$

reproducing (3.61).

The Gram matrix construction is not restricted to finite-dimensional vector spaces, but also applies to inner products on function space. Here is a particularly important example.

Example 3.39. Consider the vector space $\mathrm{C}^0[0,1]$ consisting of continuous functions on the interval $0 \le x \le 1$, equipped with the L^2 inner product $\langle f, g \rangle = \int_0^1 f(x)\, g(x)\, dx$. Let us construct the Gram matrix corresponding to the simple monomial functions $1, x, x^2$.

We compute the required inner products

$$\langle 1,1\rangle = \|1\|^2 = \int_0^1 dx = 1, \qquad \langle 1,x\rangle = \int_0^1 x\,dx = \frac{1}{2},$$
$$\langle x,x\rangle = \|x\|^2 = \int_0^1 x^2\,dx = \frac{1}{3}, \qquad \langle 1,x^2\rangle = \int_0^1 x^2\,dx = \frac{1}{3},$$
$$\langle x^2,x^2\rangle = \|x^2\|^2 = \int_0^1 x^4\,dx = \frac{1}{5}, \qquad \langle x,x^2\rangle = \int_0^1 x^3\,dx = \frac{1}{4}.$$

Therefore, the Gram matrix is

$$K = \begin{pmatrix} \langle 1,1\rangle & \langle 1,x\rangle & \langle 1,x^2\rangle \\ \langle x,1\rangle & \langle x,x\rangle & \langle x,x^2\rangle \\ \langle x^2,1\rangle & \langle x^2,x\rangle & \langle x^2,x^2\rangle \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}.$$

As we know, the monomial functions $1, x, x^2$ are linearly independent, and so Theorem 3.34 immediately implies that the matrix K is positive definite.

The alert reader may recognize this particular Gram matrix as the 3×3 Hilbert matrix that we encountered in (1.72). More generally, the Gram matrix corresponding to the monomials $1, x, x^2, \ldots, x^n$ has entries

$$k_{ij} = \langle x^{i-1}, x^{j-1}\rangle = \int_0^1 x^{i+j-2}\,dx = \frac{1}{i+j-1}, \qquad i,j = 1,\ldots,n+1,$$

and is thus the $(n+1)\times(n+1)$ Hilbert matrix (1.72): $K = H_{n+1}$. As a consequence of Theorem 3.34 and Proposition 3.31 (and also Exercise 2.3.36), we have proved the following non-trivial result.

Proposition 3.40. The $n\times n$ Hilbert matrix H_n is positive definite. Consequently, H_n is a nonsingular matrix.

Example 3.41. Let us construct the Gram matrix corresponding to the trigonometric functions 1, $\cos x$, $\sin x$, with respect to the inner product $\langle f,g\rangle = \int_{-\pi}^{\pi} f(x)\,g(x)\,dx$ on the interval $[-\pi,\pi]$. We compute the inner products

$$\langle 1,1\rangle = \|1\|^2 = \int_{-\pi}^{\pi} dx = 2\pi, \qquad \langle 1,\cos x\rangle = \int_{-\pi}^{\pi} \cos x\,dx = 0,$$
$$\langle \cos x,\cos x\rangle = \|\cos x\|^2 = \int_{-\pi}^{\pi} \cos^2 x\,dx = \pi, \qquad \langle 1,\sin x\rangle = \int_{-\pi}^{\pi} \sin x\,dx = 0,$$
$$\langle \sin x,\sin x\rangle = \|\sin x\|^2 = \int_{-\pi}^{\pi} \sin^2 x\,dx = \pi, \qquad \langle \cos x,\sin x\rangle = \int_{-\pi}^{\pi} \cos x\sin x\,dx = 0.$$

Therefore, the Gram matrix is a simple diagonal matrix: $K = \begin{pmatrix} 2\pi & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & \pi \end{pmatrix}$. Positive definiteness of K is immediately evident.

If the columns of A are linearly dependent, then the associated Gram matrix is only positive semi-definite. In this case, the Gram matrix will have nontrivial null directions $\mathbf{v}$, so that $\mathbf{0} \neq \mathbf{v} \in \ker K = \ker A$.

Proposition 3.42. Let $K = A^T C A$ be the $n \times n$ Gram matrix constructed from an $m \times n$ matrix A and a positive definite $m \times m$ matrix $C > 0$. Then $\ker K = \ker A$, and hence $\operatorname{rank} K = \operatorname{rank} A$.

Proof: Clearly, if $A\mathbf{x} = \mathbf{0}$, then $K\mathbf{x} = A^T C A\mathbf{x} = \mathbf{0}$, and so $\ker A \subset \ker K$. Conversely, if $K\mathbf{x} = \mathbf{0}$, then

$$0 = \mathbf{x}^T K \mathbf{x} = \mathbf{x}^T A^T C A \mathbf{x} = \mathbf{y}^T C \mathbf{y}, \qquad \text{where} \qquad \mathbf{y} = A\mathbf{x}.$$

Since $C > 0$, this implies $\mathbf{y} = \mathbf{0}$, and hence $\mathbf{x} \in \ker A$. Finally, by Theorem 2.49, $\operatorname{rank} K = n - \dim \ker K = n - \dim \ker A = \operatorname{rank} A$. *Q.E.D.*

Exercises

3.4.22. (*a*) Find the Gram matrix corresponding to each of the following sets of vectors using the Euclidean dot product on $\mathbb{R}^n$. (*i*) $\begin{pmatrix} -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}$, (*ii*) $\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}$, (*iii*) $\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 2 \end{pmatrix}$, (*iv*) $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, (*v*) $\begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$, (*vi*) $\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix}$, (*vii*) $\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ -4 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \\ -1 \\ -2 \end{pmatrix}$, (*viii*) $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ -3 \\ 0 \end{pmatrix}$.
(*b*) Which are positive definite? (*c*) If the matrix is positive semi-definite, find all its null directions.

3.4.23. Recompute the Gram matrices for cases (*iii*)–(*v*) in the previous exercise using the weighted inner product $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + 2 x_2 y_2 + 3 x_3 y_3$. Does this change its positive definiteness?

3.4.24. Recompute the Gram matrices for cases (*vi*)–(*viii*) in Exercise 3.4.22 for the weighted inner product $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \frac{1}{2} x_2 y_2 + \frac{1}{3} x_3 y_3 + \frac{1}{4} x_4 y_4$.

3.4.25. Find the Gram matrix K for the functions $1, e^x, e^{2x}$ using the L^2 inner product on $[0,1]$. Is K positive definite?

3.4.26. Answer Exercise 3.4.25 using the weighted inner product $\langle f, g \rangle = \int_0^1 f(x)\, g(x)\, e^{-x}\, dx$.

3.4.27. Find the Gram matrix K for the monomials $1, x, x^2, x^3$ using the L^2 inner product on $[-1,1]$. Is K positive definite?

3.4.28. Answer Exercise 3.4.27 using the weighted inner product $\langle f, g \rangle = \int_{-1}^1 f(x)\, g(x)\, (1+x)\, dx$.

3.4.29. Let K be a 2×2 Gram matrix. Explain why the positive definiteness criterion (3.55) is equivalent to the Cauchy–Schwarz inequality.

♢ 3.4.30. (*a*) Prove that if K is a positive definite matrix, then K^2 is also positive definite. (*b*) More generally, if $S = S^T$ is symmetric and nonsingular, then S^2 is positive definite.

3.4.31. Let A be an $m \times n$ matrix. (*a*) Explain why the product $L = A A^T$ is a Gram matrix. (*b*) Show that, even though they may be of different sizes, both Gram matrices $K = A^T A$ and $L = A A^T$ have the same rank. (*c*) Under what conditions are both K and L positive definite?

♢ 3.4.32. Let $K = A^T C A$, where $C > 0$. Prove that
(a) $\ker K = \operatorname{coker} K = \ker A$; (b) $\operatorname{img} K = \operatorname{coimg} K = \operatorname{coimg} A$.

♢ 3.4.33. Prove that every positive definite matrix K can be written as a Gram matrix.

3.4.34. Suppose K is the Gram matrix computed from $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ relative to a given inner product. Let $\widetilde{K}$ be the Gram matrix for the same elements, but computed relative to a *different* inner product. Show that $K > 0$ if and only if $\widetilde{K} > 0$.

♢ 3.4.35. Let $K_1 = A_1^T C_1 A_1$ and $K_2 = A_2^T C_2 A_2$ be any two $n \times n$ Gram matrices. Let $K = K_1 + K_2$. (a) Show that if $K_1, K_2 > 0$ then $K > 0$. (b) Give an example in which K_1 and K_2 are not positive definite, but $K > 0$. (c) Show that K is also a Gram matrix, by finding a matrix A such that $K = A^T C A$. *Hint*: A will have size $(m_1 + m_2) \times n$, where m_1 and m_2 are the numbers of rows in A_1, A_2, respectively.

3.4.36. Show that $\mathbf{0} \neq \mathbf{z}$ is a null direction for the quadratic form $q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x}$ based on the Gram matrix $K = A^T C A$ if and only if $\mathbf{z} \in \ker K$.

3.5 Completing the Square

Gram matrices furnish us with an almost inexhaustible supply of positive definite matrices. However, we still do not know how to test whether a given symmetric matrix is positive definite. As we shall soon see, the secret already appears in the particular computations in Examples 3.2 and 3.28.

You may recall the algebraic technique known as "completing the square", first arising in the derivation of the formula for the solution to the quadratic equation

$$q(x) = a x^2 + 2 b x + c = 0, \tag{3.65}$$

and, later, helping to facilitate the integration of various types of rational and algebraic functions. The idea is to combine the first two terms in (3.65) as a perfect square, and thereby rewrite the quadratic function in the form

$$q(x) = a \left(x + \frac{b}{a} \right)^2 + \frac{a c - b^2}{a} = 0. \tag{3.66}$$

As a consequence,

$$\left(x + \frac{b}{a} \right)^2 = \frac{b^2 - a c}{a^2}.$$

The familiar *quadratic formula*

$$x = \frac{-b \pm \sqrt{b^2 - a c}}{a}$$

follows by taking the square root of both sides and then solving for x. The intermediate step (3.66), where we eliminate the linear term, is known as *completing the square*.

We can perform the same kind of manipulation on a homogeneous quadratic form

$$q(x_1, x_2) = a x_1^2 + 2 b x_1 x_2 + c x_2^2. \tag{3.67}$$

In this case, provided $a \neq 0$, completing the square amounts to writing

$$q(x_1, x_2) = a x_1^2 + 2 b x_1 x_2 + c x_2^2 = a \left(x_1 + \frac{b}{a} x_2 \right)^2 + \frac{a c - b^2}{a} x_2^2 = a y_1^2 + \frac{a c - b^2}{a} y_2^2. \tag{3.68}$$

The net result is to re-express $q(x_1, x_2)$ as a simpler sum of squares of the new variables

$$y_1 = x_1 + \frac{b}{a}\, x_2, \qquad y_2 = x_2. \tag{3.69}$$

It is not hard to see that the final expression in (3.68) is positive definite, as a function of y_1, y_2, if and only if both coefficients are positive:

$$a > 0, \qquad \frac{a\,c - b^2}{a} > 0. \tag{3.70}$$

Therefore, $q(x_1, x_2) \geq 0$, with equality if and only if $y_1 = y_2 = 0$, or, equivalently, $x_1 = x_2 = 0$. This conclusively proves that conditions (3.70) are necessary and sufficient for the quadratic form (3.67) to be positive definite.

Our goal is to adapt this simple idea to analyze the positivity of quadratic forms depending on more than two variables. To this end, let us rewrite the quadratic form identity (3.68) in matrix form. The original quadratic form (3.67) is

$$q(\mathbf{x}) = \mathbf{x}^T K\, \mathbf{x}, \qquad \text{where} \qquad K = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{3.71}$$

Similarly, the right-hand side of (3.68) can be written as

$$\widehat{q}\,(\mathbf{y}) = \mathbf{y}^T D\, \mathbf{y}, \qquad \text{where} \qquad D = \begin{pmatrix} a & 0 \\ 0 & \dfrac{a\,c - b^2}{a} \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}. \tag{3.72}$$

Anticipating the final result, the equations (3.69) connecting $\mathbf{x}$ and $\mathbf{y}$ can themselves be written in matrix form as

$$\mathbf{y} = L^T \mathbf{x} \qquad \text{or} \qquad \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + \frac{b}{a}\, x_2 \\ x_2 \end{pmatrix}, \qquad \text{where} \qquad L^T = \begin{pmatrix} 1 & \frac{b}{a} \\ 0 & 1 \end{pmatrix}.$$

Substituting into (3.72), we obtain

$$\mathbf{y}^T D\, \mathbf{y} = (L^T \mathbf{x})^T D\, (L^T \mathbf{x}) = \mathbf{x}^T L\, D\, L^T \mathbf{x} = \mathbf{x}^T K\, \mathbf{x}, \qquad \text{where} \qquad K = L\, D\, L^T. \tag{3.73}$$

The result is the *same* factorization (1.61) of the coefficient matrix that we previously obtained via Gaussian Elimination. We are thus led to the realization that *completing the square is the same as the* $L\,D\,L^T$ *factorization of a symmetric matrix*!

Recall the definition of a regular matrix as one that can be reduced to upper triangular form without any row interchanges. Theorem 1.34 says that the regular symmetric matrices are precisely those that admit an $L\,D\,L^T$ factorization. The identity (3.73) is therefore valid for all regular $n \times n$ symmetric matrices, and shows how to write the associated quadratic form as a sum of squares:

$$q(\mathbf{x}) = \mathbf{x}^T K\, \mathbf{x} = \mathbf{y}^T D\, \mathbf{y} = d_1\, y_1^2 + \ \cdots\ + d_n\, y_n^2, \qquad \text{where} \qquad \mathbf{y} = L^T \mathbf{x}. \tag{3.74}$$

The coefficients d_i are the diagonal entries of D, which are the pivots of K. Furthermore, the diagonal quadratic form is positive definite, $\mathbf{y}^T D\, \mathbf{y} > 0$ for all $\mathbf{y} \neq \mathbf{0}$, if and only if all the pivots are positive, $d_i > 0$. Invertibility of L^T tells us that $\mathbf{y} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$, and hence, positivity of the pivots is equivalent to positive definiteness of the original quadratic form: $q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$. We have thus almost proved the main result that completely characterizes positive definite matrices.

Theorem 3.43. A symmetric matrix is positive definite if and only if it is regular and has all positive pivots.

Equivalently, a square matrix K is positive definite if and only if it can be factored $K = L D L^T$, where L is lower unitriangular and D is diagonal with all positive diagonal entries.

Example 3.44. Consider the symmetric matrix $K = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 6 & 0 \\ -1 & 0 & 9 \end{pmatrix}$. Gaussian Elimination produces the factors

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix}, \qquad L^T = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

in its factorization $K = L D L^T$. Since the pivots — the diagonal entries $1, 2$, and 6 in D — are all positive, Theorem 3.43 implies that K is positive definite, which means that the associated quadratic form satisfies

$$q(\mathbf{x}) = x_1^2 + 4 x_1 x_2 - 2 x_1 x_3 + 6 x_2^2 + 9 x_3^2 > 0, \qquad \text{for all} \qquad \mathbf{x} = (\, x_1, x_2, x_3 \,)^T \neq \mathbf{0}.$$

Indeed, the $L D L^T$ factorization implies that $q(\mathbf{x})$ can be explicitly written as a sum of squares:

$$q(\mathbf{x}) = x_1^2 + 4 x_1 x_2 - 2 x_1 x_3 + 6 x_2^2 + 9 x_3^2 = y_1^2 + 2 y_2^2 + 6 y_3^2, \tag{3.75}$$

where

$$y_1 = x_1 + 2 x_2 - x_3, \qquad y_2 = x_2 + x_3, \qquad y_3 = x_3,$$

are the entries of $\mathbf{y} = L^T \mathbf{x}$. Positivity of the coefficients of the y_i^2 (which are the pivots) implies that $q(\mathbf{x})$ is positive definite.

Example 3.45. Let's test whether the matrix $K = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 7 \\ 3 & 7 & 8 \end{pmatrix}$ is positive definite. When we perform Gaussian Elimination, the second pivot turns out to be -1, which immediately implies that K is not positive definite — even though all its entries are positive. (The third pivot is 3, but this does not affect the conclusion; all it takes is one non-positive pivot to disqualify a matrix from being positive definite. Also, row interchanges aren't of any help, since we are not allowed to perform them when checking for positive definiteness.) This means that the associated quadratic form

$$q(\mathbf{x}) = x_1^2 + 4 x_1 x_2 + 6 x_1 x_3 + 3 x_2^2 + 14 x_2 x_3 + 8 x_3^2$$

assumes negative values at some points. For instance, $q(-2, 1, 0) = -1$.

A direct method for completing the square in a quadratic form goes as follows: The first step is to put all the terms involving x_1 in a suitable square, at the expense of introducing extra terms involving only the other variables. For instance, in the case of the quadratic form in (3.75), the terms involving x_1 can be written as

$$x_1^2 + 4 x_1 x_2 - 2 x_1 x_3 = (x_1 + 2 x_2 - x_3)^2 - 4 x_2^2 + 4 x_2 x_3 - x_3^2.$$

Therefore,

$$q(\mathbf{x}) = (x_1 + 2 x_2 - x_3)^2 + 2 x_2^2 + 4 x_2 x_3 + 8 x_3^2 = (x_1 + 2 x_2 - x_3)^2 + \widetilde{q}(x_2, x_3),$$

where

$$\widetilde{q}(x_2, x_3) = 2 x_2^2 + 4 x_2 x_3 + 8 x_3^2$$

is a quadratic form that involves only x_2, x_3. We then repeat the process, combining all the terms involving x_2 in the remaining quadratic form into a square, writing

$$\widetilde{q}(x_2, x_3) = 2\,(x_2 + x_3)^2 + 6\,x_3^2.$$

This gives the final form

$$q(\mathbf{x}) = (x_1 + 2\,x_2 - x_3)^2 + 2(x_2 + x_3)^2 + 6\,x_3^2,$$

which reproduces (3.75).

In general, as long as $k_{11} \neq 0$, we can write

$$\begin{aligned} q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x} &= k_{11}\,x_1^2 + 2\,k_{12}\,x_1\,x_2 + \;\cdots\; + 2\,k_{1n}\,x_1\,x_n + k_{22}\,x_2^2 + \;\cdots\; + k_{nn}\,x_n^2 \\ &= k_{11}\left(x_1 + \frac{k_{12}}{k_{11}}\,x_2 + \;\cdots\; + \frac{k_{1n}}{k_{11}}\,x_n\right)^2 + \widetilde{q}(x_2, \ldots, x_n) \\ &= k_{11}\,(x_1 + l_{21}\,x_2 + \;\cdots\; + l_{n1}\,x_n)^2 + \widetilde{q}(x_2, \ldots, x_n), \end{aligned} \tag{3.76}$$

where

$$l_{21} = \frac{k_{21}}{k_{11}} = \frac{k_{12}}{k_{11}}, \qquad \ldots \qquad l_{n1} = \frac{k_{n1}}{k_{11}} = \frac{k_{1n}}{k_{11}}$$

are precisely the multiples appearing in the matrix L obtained from applying Gaussian Elimination to K, while

$$\widetilde{q}(x_2, \ldots, x_n) = \sum_{i,j=2}^{n} \widetilde{k}_{ij}\,x_i\,x_j$$

is a quadratic form involving one less variable. The entries of its symmetric coefficient matrix $\widetilde{K}$ are

$$\widetilde{k}_{ij} = \widetilde{k}_{ji} = k_{ij} - l_{j1}\,k_{1i} = k_{ij} - \frac{k_{1j}\,k_{1i}}{k_{11}}, \qquad\qquad i, j = 2, \ldots n,$$

which are exactly the same as the entries appearing below and to the right of the first pivot after applying the the first phase of the Gaussian Elimination process to K. In particular, the second pivot of K is the diagonal entry $\widetilde{k}_{22}$. We continue by applying the same procedure to the reduced quadratic form $\widetilde{q}(x_2, \ldots, x_n)$ and repeating until only the final variable remains. Completing the square at each stage reproduces the corresponding phase of the Gaussian Elimination process. The final result is our formula (3.74) rewriting the original quadratic form as a sum of squares whose coefficients are the pivots.

With this in hand, we can now complete the proof of Theorem 3.43. First, if the upper left entry k_{11}, namely the first pivot, is not strictly positive, then K cannot be positive definite, because $q(\mathbf{e}_1) = \mathbf{e}_1^T K\,\mathbf{e}_1 = k_{11} \leq 0$. Otherwise, suppose $k_{11} > 0$, and so we can write $q(\mathbf{x})$ in the form (3.76). We claim that $q(\mathbf{x})$ is positive definite if and only if the reduced quadratic form $\widetilde{q}(x_2, \ldots, x_n)$ is positive definite. Indeed, if $\widetilde{q}$ is positive definite and $k_{11} > 0$, then $q(\mathbf{x})$ is the sum of two positive quantities, which simultaneously vanish if and only if $x_1 = x_2 = \cdots = x_n = 0$. On the other hand, suppose $\widetilde{q}(x_2^\star, \ldots, x_n^\star) \leq 0$ for some $x_2^\star, \ldots, x_n^\star$, not all zero. Setting $x_1^\star = -\,l_{21}\,x_2^\star - \;\cdots\; - l_{n1}\,x_n^\star$ makes the initial square term in (3.76) equal to 0, so

$$q(x_1^\star, x_2^\star, \ldots, x_n^\star) = \widetilde{q}(x_2^\star, \ldots, x_n^\star) \leq 0,$$

proving the claim. In particular, positive definiteness of $\widetilde{q}$ requires that the second pivot satisfy $\widetilde{k}_{22} > 0$. We then continue the reduction procedure outlined in the preceding paragraph; if a non-positive entry appears in the diagonal pivot position at any stage, the original quadratic form and matrix cannot be positive definite. On the other hand, finding all positive pivots (without using any row interchanges) will, in the absence of numerical errors, ensure positive definiteness. *Q.E.D.*

Exercises

3.5.1. Are the following matrices are positive definite? (*a*) $\begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix}$, (*b*) $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, (*c*) $\begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{pmatrix}$, (*d*) $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & -2 \\ 1 & -2 & 4 \end{pmatrix}$, (*e*) $\begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}$, (*f*) $\begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$.

3.5.2. Find an $L D L^T$ factorization of the following symmetric matrices. Which are positive definite? (*a*) $\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$, (*b*) $\begin{pmatrix} 5 & -1 \\ -1 & 3 \end{pmatrix}$, (*c*) $\begin{pmatrix} 3 & -1 & 3 \\ -1 & 5 & 1 \\ 3 & 1 & 5 \end{pmatrix}$, (*d*) $\begin{pmatrix} -2 & 1 & -1 \\ 1 & -2 & 1 \\ -1 & 1 & -2 \end{pmatrix}$, (*e*) $\begin{pmatrix} 2 & 1 & -2 \\ 1 & 1 & -3 \\ -2 & -3 & 11 \end{pmatrix}$, (*f*) $\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix}$, (*g*) $\begin{pmatrix} 3 & 2 & 1 & 0 \\ 2 & 3 & 0 & 1 \\ 1 & 0 & 3 & 2 \\ 0 & 1 & 2 & 3 \end{pmatrix}$, (*h*) $\begin{pmatrix} 2 & 1 & -2 & 0 \\ 1 & 1 & -3 & 2 \\ -2 & -3 & 10 & -1 \\ 0 & 2 & -1 & 7 \end{pmatrix}$.

3.5.3. (*a*) For which values of c is the matrix $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & c & 1 \\ 0 & 1 & 1 \end{pmatrix}$ positive definite? (*b*) For the particular value $c = 3$, carry out elimination to find the factorization $A = L D L^T$. (*c*) Use your result from part (*b*) to rewrite the quadratic form $q(x, y, z) = x^2 + 2xy + 3y^2 + 2yz + z^2$ as a sum of squares. (*d*) Explain how your result is related to the positive definiteness of A.

3.5.4. Write the quadratic form $q(\mathbf{x}) = x_1^2 + x_1 x_2 + 2x_2^2 - x_1 x_3 + 3x_3^2$ in the form $q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x}$ for some symmetric matrix K. Is $q(\mathbf{x})$ positive definite?

3.5.5. Write the following quadratic forms on $\mathbb{R}^2$ as a sum of squares. Which are positive definite? (*a*) $x^2 + 8xy + y^2$, (*b*) $x^2 - 4xy + 7y^2$, (*c*) $x^2 - 2xy - y^2$, (*d*) $x^2 + 6xy$.

3.5.6. Prove that the following quadratic forms on $\mathbb{R}^3$ are positive definite by writing each as a sum of squares: (*a*) $x^2 + 4xz + 3y^2 + 5z^2$, (*b*) $x^2 + 3xy + 3y^2 - 2xz + 8z^2$, (*c*) $2x_1^2 + x_1 x_2 - 2x_1 x_3 + 2x_2^2 - 2x_2 x_3 + 2x_3^2$.

3.5.7. Write the following quadratic forms in matrix notation and determine if they are positive definite: (*a*) $x^2 + 4xz + 2y^2 + 8yz + 12z^2$, (*b*) $3x^2 - 2y^2 - 8xy + xz + z^2$, (*c*) $x^2 + 2xy + 2y^2 - 4xz - 6yz + 6z^2$, (*d*) $3x_1^2 - x_2^2 + 5x_3^2 + 4x_1x_2 - 7x_1x_3 + 9x_2x_3$, (*e*) $x_1^2 + 4x_1 x_2 - 2x_1 x_3 + 5x_2^2 - 2x_2 x_4 + 6x_3^2 - x_3 x_4 + 4x_4^2$.

3.5.8. For what values of a, b, and c is the quadratic form $x^2 + axy + y^2 + bxz + cyz + z^2$ positive definite?

3.5.9. *True or false*: Every planar quadratic form $q(x, y) = ax^2 + 2bxy + cy^2$ can be written as a sum of squares.

3.5.10. (*a*) Prove that a positive definite matrix has positive determinant: $\det K > 0$. (*b*) Show that a positive definite matrix has positive trace: $\operatorname{tr} K > 0$. (*c*) Show that every 2×2 symmetric matrix with positive determinant and positive trace is positive definite. (*d*) Find a symmetric 3×3 matrix with positive determinant and positive trace that is not positive definite.

3.5.11. (a) Prove that if K_1, K_2 are positive definite $n \times n$ matrices, then $K = \begin{pmatrix} K_1 & \mathrm{O} \\ \mathrm{O} & K_2 \end{pmatrix}$ is a positive definite $2n \times 2n$ matrix. (b) Is the converse true?

3.5.12. Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. (a) Show that $q(\mathbf{x})$ is a positive definite quadratic form if and only if $q(\mathbf{u}) > 0$ for all unit vectors, $\|\mathbf{u}\| = 1$. (b) Prove that if $S = S^T$ is any symmetric matrix, then $K = S + c\,\mathrm{I} > 0$ is positive definite if c is sufficiently large.

3.5.13. Prove that every symmetric matrix $S = K + N$ can be written as the sum of a positive definite matrix K and a negative definite matrix N. *Hint*: Use Exercise 3.5.12(b).

♢ 3.5.14. (a) Prove that every regular symmetric matrix can be decomposed as a linear combination

$$K = d_1 \mathbf{l}_1 \mathbf{l}_1^T + d_2 \mathbf{l}_2 \mathbf{l}_2^T + \cdots + d_n \mathbf{l}_n \mathbf{l}_n^T \tag{3.77}$$

of symmetric rank 1 matrices, as in Exercise 1.8.15, where $\mathbf{l}_1, \ldots, \mathbf{l}_n$ are the columns of the lower unitriangular matrix L and $d_1, \ldots, d_n$ are the pivots, i.e., the diagonal entries of D. *Hint*: See Exercise 1.2.34. (b) Decompose $\begin{pmatrix} 4 & -1 \\ -1 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 & 1 \\ 2 & 6 & 1 \\ 1 & 1 & 4 \end{pmatrix}$ in this manner.

♡ 3.5.15. There is an alternative criterion for positive definiteness based on subdeterminants of the matrix. The 2×2 version already appears in (3.70). (a) Prove that a 3×3 matrix $K = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$ is positive definite if and only if $a > 0$, $a\,d - b^2 > 0$, and $\det K > 0$. (b) Prove the general version: an $n \times n$ matrix $K > 0$ is positive definite if and only if its upper left square $k \times k$ submatrices have positive determinant for all $k = 1, \ldots, n$. *Hint*: See Exercise 1.9.17.

♢ 3.5.16. Let K be a symmetric matrix. Prove that if a non-positive diagonal entry appears anywhere (not necessarily in the pivot position) in the matrix during Regular Gaussian Elimination, then K is not positive definite.

♢ 3.5.17. Formulate a determinantal criterion similar to that in Exercise 3.5.15 for negative definite matrices. Write out the 2×2 and 3×3 cases explicitly.

3.5.18. *True or false*: A negative definite matrix must have negative trace and negative determinant.

The Cholesky Factorization

The identity (3.73) shows us how to write an arbitrary regular quadratic form $q(\mathbf{x})$ as a linear combination of squares. We can push this result slightly further in the positive definite case. Since each pivot d_i is positive, we can write the diagonal quadratic form (3.74) as a sum of pure squares:

$$d_1 y_1^2 + \cdots + d_n y_n^2 = \left(\sqrt{d_1}\, y_1\right)^2 + \cdots + \left(\sqrt{d_n}\, y_n\right)^2 = z_1^2 + \cdots + z_n^2,$$

where $z_i = \sqrt{d_i}\, y_i$. In matrix form, we are writing

$$\widehat{q}\,(\mathbf{y}) = \mathbf{y}^T D\,\mathbf{y} = \mathbf{z}^T\mathbf{z} = \|\mathbf{z}\|^2, \quad \text{where} \quad \mathbf{z} = S\,\mathbf{y}, \quad \text{with} \quad S = \operatorname{diag}\left(\sqrt{d_1}, \ldots, \sqrt{d_n}\right).$$

Since $D = S^2$, the matrix S can be thought of as a "square root" of the diagonal matrix D. Substituting back into (1.58), we deduce the *Cholesky factorization*

$$K = L\,D\,L^T = L\,S\,S^T L^T = M\,M^T, \qquad \text{where} \qquad M = L\,S, \tag{3.78}$$

of a positive definite matrix, first proposed by the early twentieth-century French geographer André-Louis Cholesky for solving problems in geodetic surveying. Note that M is a lower triangular matrix with all positive diagonal entries, namely the square roots of the pivots: $m_{ii} = \sqrt{d_i}$. Applying the Cholesky factorization to the corresponding quadratic form produces

$$q(\mathbf{x}) = \mathbf{x}^T K\,\mathbf{x} = \mathbf{x}^T M\,M^T \mathbf{x} = \mathbf{z}^T \mathbf{z} = \|\,\mathbf{z}\,\|^2, \qquad \text{where} \qquad \mathbf{z} = M^T \mathbf{x}. \tag{3.79}$$

We can interpret (3.79) as a change of variables from $\mathbf{x}$ to $\mathbf{z}$ that converts an arbitrary inner product norm, as defined by the square root of the positive definite quadratic form $q(\mathbf{x})$, into the standard Euclidean norm $\|\,\mathbf{z}\,\|$.

Example 3.46. For the matrix $K = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 6 & 0 \\ -1 & 0 & 9 \end{pmatrix}$ considered in Example 3.44, the Cholesky formula (3.78) gives $K = M\,M^T$, where

$$M = L\,S = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{6} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & \sqrt{2} & 0 \\ -1 & \sqrt{2} & \sqrt{6} \end{pmatrix}.$$

The associated quadratic function can then be written as a sum of pure squares:

$$q(\mathbf{x}) = x_1^2 + 4\,x_1 x_2 - 2\,x_1 x_3 + 6\,x_2^2 + 9\,x_3^2 = z_1^2 + z_2^2 + z_3^2,$$

where

$$\mathbf{z} = M^T \mathbf{x}, \quad \text{or, explicitly,} \quad z_1 = x_1 + 2\,x_2 - x_3, \quad z_2 = \sqrt{2}\,x_2 + \sqrt{2}\,x_3, \quad z_3 = \sqrt{6}\,x_3.$$

Exercises

3.5.19. Find the Cholesky factorizations of the following matrices: (a) $\begin{pmatrix} 3 & -2 \\ -2 & 2 \end{pmatrix}$, (b) $\begin{pmatrix} 4 & -12 \\ -12 & 45 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & -2 \\ 1 & -2 & 14 \end{pmatrix}$, (d) $\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$, (e) $\begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$.

3.5.20. Which of the matrices in Exercise 3.5.1 have a Cholesky factorization? For those that do, write out the factorization.

3.5.21. Write the following positive definite quadratic forms as a sum of pure squares, as in (3.79): (a) $16x_1^2 + 25x_2^2$, (b) $x_1^2 - 2x_1x_2 + 4x_2^2$, (c) $5x_1^2 + 4x_1x_2 + 3x_2^2$, (d) $3x_1^2 - 2x_1x_2 - 2x_1x_3 + 2x_2^2 + 6x_3^2$, (e) $x_1^2 + x_1x_2 + x_2^2 + x_2x_3 + x_3^2$, (f) $4x_1^2 - 2x_1x_2 - 4x_1x_3 + \frac{1}{2}x_2^2 - x_2x_3 + 6x_3^2$, (g) $3x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_2x_3 + 3x_3^2 + 2x_3x_4 + 3x_4^2$.

3.6 Complex Vector Spaces

Although physical applications ultimately require real answers, complex numbers and complex vector spaces play an extremely useful, if not essential, role in the intervening analysis. Particularly in the description of periodic phenomena, complex numbers and complex exponentials help to simplify complicated trigonometric formulas. Complex variable methods

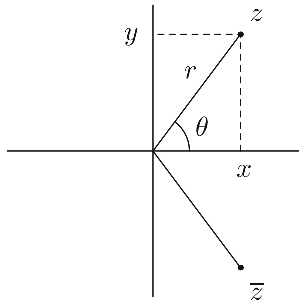

Figure 3.7. Complex Numbers.

are ubiquitous in electrical engineering, Fourier analysis, potential theory, fluid mechanics, electromagnetism, and many other applied fields, [**49**, **79**]. In quantum mechanics, the basic physical quantities are complex-valued wave functions, [**54**]. Moreover, the Schrödinger equation, which governs quantum dynamics, is an inherently complex partial differential equation.

In this section, we survey the principal properties of complex numbers and complex vector spaces. Most of the constructions are straightforward adaptations of their real counterparts, and so will not be dwelled on at length. The one exception is the complex version of an inner product, which does introduce some novelties not found in its simpler real sibling.

Complex Numbers

Recall that a *complex number* is an expression of the form $z = x + \mathrm{i}\, y$, where $x, y \in \mathbb{R}$ are real and[†] $\mathrm{i} = \sqrt{-1}$. The set of all complex numbers (scalars) is denoted by $\mathbb{C}$. We call $x = \operatorname{Re} z$ the *real part* of z and $y = \operatorname{Im} z$ the *imaginary part* of $z = x + \mathrm{i}\, y$. (*Note*: The imaginary part is the real number y, *not* $\mathrm{i}\, y$.) A real number x is merely a complex number with zero imaginary part, $\operatorname{Im} z = 0$, and so we may regard $\mathbb{R} \subset \mathbb{C}$. Complex addition and multiplication are based on simple adaptations of the rules of real arithmetic to include the identity $\mathrm{i}^2 = -1$, and so

$$\begin{aligned} (x + \mathrm{i}\, y) + (u + \mathrm{i}\, v) &= (x + u) + \mathrm{i}\,(y + v), \\ (x + \mathrm{i}\, y)\,(u + \mathrm{i}\, v) &= (x\,u - y\,v) + \mathrm{i}\,(x\,v + y\,u). \end{aligned} \tag{3.80}$$

Complex numbers enjoy all the usual laws of real addition and multiplication, *including commutativity*: $z\,w = w\,z$.

We can identify a complex number $x + \mathrm{i}\, y$ with a vector $(x, y)^T \in \mathbb{R}^2$ in the real plane. For this reason, $\mathbb{C}$ is sometimes referred to as the *complex plane*. Complex addition (3.80) corresponds to vector addition, but complex multiplication does not have a readily identifiable vector counterpart.

Another useful operation on complex numbers is that of complex conjugation.

Definition 3.47. The *complex conjugate* of $z = x + \mathrm{i}\, y$ is $\overline{z} = x - \mathrm{i}\, y$, whereby $\operatorname{Re} \overline{z} = \operatorname{Re} z$, while $\operatorname{Im} \overline{z} = -\operatorname{Im} z$.

[†] To avoid confusion with the symbol for current, electrical engineers prefer to use j to indicate the imaginary unit.

Geometrically, the complex conjugate of z is obtained by reflecting the corresponding vector through the real axis, as illustrated in Figure 3.7. In particular $\overline{z} = z$ if and only if z is real. Note that

$$\operatorname{Re} z = \frac{z + \overline{z}}{2}, \qquad \operatorname{Im} z = \frac{z - \overline{z}}{2\,\mathrm{i}}. \tag{3.81}$$

Complex conjugation is compatible with complex arithmetic:

$$\overline{z + w} = \overline{z} + \overline{w}, \qquad \overline{z\,w} = \overline{z}\,\overline{w}.$$

In particular, the product of a complex number and its conjugate,

$$z\,\overline{z} = (x + \mathrm{i}\,y)\,(x - \mathrm{i}\,y) = x^2 + y^2, \tag{3.82}$$

is real and non-negative. Its square root is known as the *modulus* or *norm* of the complex number $z = x + \mathrm{i}\,y$, and written

$$|\,z\,| = \sqrt{x^2 + y^2}\,. \tag{3.83}$$

Note that $|\,z\,| \geq 0$, with $|\,z\,| = 0$ if and only if $z = 0$. The modulus $|\,z\,|$ generalizes the absolute value of a real number, and coincides with the standard Euclidean norm in the $x\,y$–plane, which implies the validity of the triangle inequality

$$|\,z + w\,| \leq |\,z\,| + |\,w\,|. \tag{3.84}$$

Equation (3.82) can be rewritten in terms of the modulus as

$$z\,\overline{z} = |\,z\,|^2. \tag{3.85}$$

Rearranging the factors, we deduce the formula for the reciprocal of a nonzero complex number:

$$\frac{1}{z} = \frac{\overline{z}}{|\,z\,|^2}, \qquad z \neq 0, \qquad \text{or, equivalently,} \qquad \frac{1}{x + \mathrm{i}\,y} = \frac{x - \mathrm{i}\,y}{x^2 + y^2}. \tag{3.86}$$

The general formula for complex division,

$$\frac{w}{z} = \frac{w\,\overline{z}}{|\,z\,|^2} \qquad \text{or} \qquad \frac{u + \mathrm{i}\,v}{x + \mathrm{i}\,y} = \frac{(x\,u + y\,v) + \mathrm{i}\,(x\,v - y\,u)}{x^2 + y^2}, \tag{3.87}$$

is an immediate consequence.

The modulus of a complex number,

$$r = |\,z\,| = \sqrt{x^2 + y^2}\,,$$

is one component of its polar coordinate representation

$$x = r\cos\theta, \qquad y = r\sin\theta \qquad \text{or} \qquad z = r(\cos\theta + \mathrm{i}\,\sin\theta). \tag{3.88}$$

The polar angle, which measures the angle that the line connecting z to the origin makes with the horizontal axis, is known as the *phase*, and written

$$\theta = \operatorname{ph} z. \tag{3.89}$$

As such, the phase is defined only up to an integer multiple of $2\,\pi$. The more common term for the angle is the *argument*, written $\arg z = \operatorname{ph} z$. However, we prefer to use "phase" throughout this text, in part to avoid confusion with the argument z of a function $f(z)$.

We note that the modulus and phase of a product of complex numbers can be readily computed:

$$|\,z\,w\,| = |\,z\,|\,|\,w\,|, \qquad \operatorname{ph}(z\,w) = \operatorname{ph} z + \operatorname{ph} w. \tag{3.90}$$

Complex conjugation preserves the modulus, but reverses the sign of the phase:

$$|\,\overline{z}\,| = |\,z\,|, \qquad \operatorname{ph}\overline{z} = -\operatorname{ph} z. \tag{3.91}$$

One of the most profound formulas in all of mathematics is *Euler's formula*

$$e^{\,\mathrm{i}\,\theta} = \cos\theta + \mathrm{i}\,\sin\theta, \tag{3.92}$$

relating the complex exponential with the real sine and cosine functions. It has a variety of mathematical justifications; see Exercise 3.6.23 for one that is based on comparing power series. Euler's formula can be used to compactly rewrite the polar form (3.88) of a complex number as

$$z = r\,e^{\,\mathrm{i}\,\theta} \qquad \text{where} \qquad r = |\,z\,|, \qquad \theta = \operatorname{ph} z. \tag{3.93}$$

The complex conjugation identity

$$e^{-\,\mathrm{i}\,\theta} = \cos(-\theta) + \mathrm{i}\,\sin(-\theta) = \cos\theta - \mathrm{i}\,\sin\theta = \overline{e^{\,\mathrm{i}\,\theta}}$$

permits us to express the basic trigonometric functions in terms of complex exponentials:

$$\cos\theta = \frac{e^{\,\mathrm{i}\,\theta} + e^{-\,\mathrm{i}\,\theta}}{2}\,, \qquad \sin\theta = \frac{e^{\,\mathrm{i}\,\theta} - e^{-\,\mathrm{i}\,\theta}}{2\,\mathrm{i}}\,. \tag{3.94}$$

These formulas are very useful when working with trigonometric identities and integrals.

The exponential of a general complex number is easily derived from the Euler formula and the standard properties of the exponential function — which carry over unaltered to the complex domain; thus,

$$e^{z} = e^{x+\mathrm{i}\,y} = e^{x}\,e^{\,\mathrm{i}\,y} = e^{x}\cos y + \mathrm{i}\,e^{x}\sin y. \tag{3.95}$$

Note that $e^{2\,\pi\,\mathrm{i}} = 1$, and hence the exponential function is periodic,

$$e^{z+2\,\pi\,\mathrm{i}} = e^{z}, \tag{3.96}$$

with imaginary period $2\,\pi\,\mathrm{i}$ — indicative of the periodicity of the trigonometric functions in Euler's formula.

Exercises

3.6.1. Write down a single equation that relates the five most important numbers in mathematics, which are $0, 1, e, \pi$, and i.

3.6.2. For any integer k, prove that $e^{k\,\pi\,\mathrm{i}} = (-1)^k$.

3.6.3. Is the formula $1^z = 1$ valid for all complex values of z?

3.6.4. What is wrong with the calculation $e^{2\,a\,\pi\,\mathrm{i}} = (e^{2\,\pi\,\mathrm{i}})^a = 1^a = 1$?

3.6.5. (*a*) Write i in phase–modulus form. (*b*) Use this expression to find $\sqrt{\mathrm{i}}$, i.e., a complex number z such that $z^2 = \mathrm{i}$. Can you find a second square root? (*c*) Find explicit formulas for the three third roots and four fourth roots of i.

3.6.6. In Figure 3.7, where would you place the point $1/z$?

3.6.7. (a) If z moves counterclockwise around a circle of radius r in the complex plane, around which circle and in which direction does $w = 1/z$ move? (b) What about $w = \overline{z}$? (c) What if the circle is not centered at the origin?

♢ 3.6.8. Show that $-|\,z\,| \le \operatorname{Re} z \le |\,z\,|$ and $-|\,z\,| \le \operatorname{Im} z \le |\,z\,|$.

♢ 3.6.9. Prove that if φ is real, then $\operatorname{Re}(e^{\mathrm{i}\,\varphi} z) \le |\,z\,|$, with equality if and only if $\varphi = -\operatorname{ph} z$.

3.6.10. Prove the identities in (3.90) and (3.91).

3.6.11. Prove $\operatorname{ph}(z/w) = \operatorname{ph} z - \operatorname{ph} w = \operatorname{ph}(z\,\overline{w})$ is equal to the angle between the vectors representing z and w.

3.6.12. The phase of a complex number $z = x + \mathrm{i}\,y$ is often written as $\operatorname{ph} z = \tan^{-1}(y/x)$. Explain why this formula is ambiguous, and does not uniquely define $\operatorname{ph} z$.

3.6.13. Show that if we identify the complex numbers z, w with vectors in the plane, then their Euclidean dot product is equal to $\operatorname{Re}(z\,\overline{w})$.

3.6.14. (a) Prove that the complex numbers z and w correspond to orthogonal vectors in $\mathbb{R}^2$ if and only if $\operatorname{Re} z\,\overline{w} = 0$. (b) Prove that z and $\mathrm{i}\,z$ are always orthogonal.

3.6.15. Prove that $e^{z+w} = e^z\, e^w$. Conclude that $e^{m\,z} = (e^z)^m$ whenever m is an integer.

3.6.16. (a) Use the formula $e^{2\,\mathrm{i}\,\theta} = (e^{\mathrm{i}\,\theta})^2$ to deduce the well-known trigonometric identities for $\cos 2\theta$ and $\sin 2\theta$. (b) Derive the corresponding identities for $\cos 3\theta$ and $\sin 3\theta$. (c) Write down the explicit identities for $\cos m\theta$ and $\sin m\theta$ as polynomials in $\cos\theta$ and $\sin\theta$. *Hint*: Apply the Binomial Formula to $(e^{\mathrm{i}\,\theta})^m$.

♢ 3.6.17. Use complex exponentials to prove the identity $\cos\theta - \cos\varphi = 2\cos\dfrac{\theta-\varphi}{2}\,\cos\dfrac{\theta+\varphi}{2}$.

3.6.18. Prove that if $z = x + \mathrm{i}\,y$, then $|\,e^z\,| = e^x$, $\quad \operatorname{ph} e^z = y$.

3.6.19. The formulas $\cos z = \dfrac{e^{\mathrm{i}\,z} + e^{-\mathrm{i}\,z}}{2}$ and $\sin z = \dfrac{e^{\mathrm{i}\,z} - e^{-\mathrm{i}\,z}}{2\,\mathrm{i}}$ serve to define the basic complex trigonometric functions. Write out the formulas for their real and imaginary parts in terms of $z = x + \mathrm{i}\,y$, and show that $\cos z$ and $\sin z$ reduce to their usual real forms when $z = x$ is real. What do they become when $z = \mathrm{i}\,y$ is purely imaginary?

3.6.20. The complex *hyperbolic functions* are defined as $\cosh z = \dfrac{e^z + e^{-z}}{2}$, $\sinh z = \dfrac{e^z - e^{-z}}{2}$.
(a) Write out the formulas for their real and imaginary parts in terms of $z = x + \mathrm{i}\,y$.
(b) Prove that $\cos \mathrm{i}\,z = \cosh z$ and $\sin \mathrm{i}\,z = \mathrm{i}\sinh z$.

♡ 3.6.21. Generalizing Example 2.17c, by a *trigonometric polynomial* of *degree* $\le n$, we mean a function $T(x) = \sum_{0\le j+k\le n} c_{jk}\,(\cos\theta)^j\,(\sin\theta)^k$ in the powers of the sine and cosine functions up to degree n. (a) Use formula (3.94) to prove that every trigonometric polynomial of degree $\le n$ can be written as a complex linear combination of the $2n+1$ complex exponentials $e^{-n\,\mathrm{i}\,\theta}, \ \ldots\ e^{-\mathrm{i}\,\theta},\ e^{0\,\mathrm{i}\,\theta} = 1,\ e^{\mathrm{i}\,\theta},\ e^{2\,\mathrm{i}\,\theta}, \ \ldots\ e^{n\,\mathrm{i}\,\theta}$. (b) Prove that every trigonometric polynomial of degree $\le n$ can be written as a real linear combination of the trigonometric functions $1,\ \cos\theta,\ \sin\theta,\ \cos 2\theta,\ \sin 2\theta, \ \ldots\ \cos n\theta,\ \sin n\theta$.
(c) Write out the following trigonometric polynomials in both of the preceding forms:
$(i)\ \cos^2\theta,\ (ii)\ \cos\theta\sin\theta,\ (iii)\ \cos^3\theta,\ (iv)\ \sin^4\theta,\ (v)\ \cos^2\theta\sin^2\theta$.

♢ 3.6.22. Write out the real and imaginary parts of the power function x^c with complex exponent $c = a + \mathrm{i}\,b \in \mathbb{C}$.

♢ 3.6.23. Write the power series expansions for $e^{\mathrm{i}\,x}$. Prove that the real terms give the power series for $\cos x$, while the imaginary terms give that of $\sin x$. Use this identification to justify Euler's formula (3.92).

♢ 3.6.24. The derivative of a complex-valued function $f(x) = u(x) + \mathrm{i}\, v(x)$, depending on a real variable x, is given by $f'(x) = u'(x) + \mathrm{i}\, v'(x)$. (*a*) Prove that if $\lambda = \mu + \mathrm{i}\,\nu$ is any complex scalar, then $\dfrac{d}{dx}\, e^{\lambda x} = \lambda\, e^{\lambda x}$. (*b*) Prove, conversely, $\displaystyle\int_a^b e^{\lambda x}\, dx = \frac{1}{\lambda}\left(e^{\lambda b} - e^{\lambda a}\right)$ provided $\lambda \neq 0$.

3.6.25. Use the complex trigonometric formulas (3.94) and Exercise 3.6.24 to evaluate the following trigonometric integrals: (*a*) $\displaystyle\int \cos^2 x\, dx$, (*b*) $\displaystyle\int \sin^2 x\, dx$, (*c*) $\displaystyle\int \cos x \sin x\, dx$, (*d*) $\displaystyle\int \cos 3x \sin 5x\, dx$. How did you calculate them in first-year calculus? If you're not convinced this method is easier, try the more complicated integrals
(*e*) $\displaystyle\int \cos^4 x\, dx$, (*f*) $\displaystyle\int \sin^4 x\, dx$, (*g*) $\displaystyle\int \cos^2 x \sin^2 x\, dx$, (*h*) $\displaystyle\int \cos 3x \sin 5x \cos 7x\, dx$.

Complex Vector Spaces and Inner Products

A *complex vector space* is defined in exactly the same manner as its real counterpart, as in Definition 2.1, the only difference being that we replace real scalars by complex scalars. The most basic example is the n-dimensional complex vector space $\mathbb{C}^n$ consisting of all column vectors $\mathbf{z} = (\, z_1, z_2, \ldots, z_n\,)^T$ that have n complex entries $z_1, \ldots, z_n \in \mathbb{C}$. Vector addition and scalar multiplication are defined in the obvious manner, and verification of each of the vector space axioms is immediate.

We can write any complex vector $\mathbf{z} = \mathbf{x} + \mathrm{i}\,\mathbf{y} \in \mathbb{C}^n$ as a linear combination of two real vectors $\mathbf{x} = \operatorname{Re} \mathbf{z}$ and $\mathbf{y} = \operatorname{Im} \mathbf{z} \in \mathbb{R}^n$ called its *real* and *imaginary parts*. Its complex conjugate $\overline{\mathbf{z}} = \mathbf{x} - \mathrm{i}\,\mathbf{y}$ is obtained by taking the complex conjugates of its individual entries. Thus, for example, if

$$\mathbf{z} = \begin{pmatrix} 1+2\,\mathrm{i} \\ -3 \\ 5\,\mathrm{i} \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \\ 0 \end{pmatrix} + \mathrm{i} \begin{pmatrix} 2 \\ 0 \\ 5 \end{pmatrix}, \qquad \text{then} \qquad \operatorname{Re} \mathbf{z} = \begin{pmatrix} 1 \\ -3 \\ 0 \end{pmatrix}, \quad \operatorname{Im} \mathbf{z} = \begin{pmatrix} 2 \\ 0 \\ 5 \end{pmatrix},$$

$$\text{and so its complex conjugate is} \qquad \overline{\mathbf{z}} = \begin{pmatrix} 1-2\,\mathrm{i} \\ -3 \\ -5\,\mathrm{i} \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \\ 0 \end{pmatrix} - \mathrm{i} \begin{pmatrix} 2 \\ 0 \\ 5 \end{pmatrix}.$$

In particular, $\mathbf{z} \in \mathbb{R}^n \subset \mathbb{C}^n$ is a real vector if and only if $\mathbf{z} = \overline{\mathbf{z}}$.

Most of the vector space concepts we developed in the real domain, including span, linear independence, basis, and dimension, can be straightforwardly extended to the complex regime. The one exception is the concept of an inner product, which requires a little thought. In analysis, the primary applications of inner products and norms rely on the associated inequalities: Cauchy–Schwarz and triangle. But there is no natural ordering of the complex numbers, and so one *cannot* assign a meaning to a complex inequality like $z < w$. Inequalities make sense only in the real domain, and so the norm of a complex vector should still be a positive and real. With this in mind, the naïve idea of simply summing the squares of the entries of a complex vector will *not* define a norm on $\mathbb{C}^n$, since the result will typically be complex. Moreover, some nonzero complex vectors, e.g., $(\, 1, \mathrm{i}\,)^T$, would then have zero "norm".

The correct definition is modeled on the formula

$$|\, z\,| = \sqrt{z\,\overline{z}}\,,$$

which defines the modulus of a complex scalar $z \in \mathbb{C}$. If, in analogy with the real definition (3.7), the quantity inside the square root should represent the inner product of z with itself, then we should define the "dot product" between two complex numbers to be

$$z \cdot w = z\,\overline{w}, \qquad \text{so that} \qquad z \cdot z = z\,\overline{z} = |\,z\,|^2.$$

Writing out the formula when $z = x + \mathrm{i}\,y$ and $w = u + \mathrm{i}\,v$, we obtain

$$z \cdot w = z\,\overline{w} = (x + \mathrm{i}\,y)\,(u - \mathrm{i}\,v) = (x\,u + y\,v) + \mathrm{i}\,(y\,u - x\,v). \tag{3.97}$$

Thus, the dot product of two complex numbers is, in general, complex. The real part of $z \cdot w$ is, in fact, the Euclidean dot product between the corresponding vectors in $\mathbb{R}^2$, while its imaginary part is, interestingly, their scalar cross product, cf. (3.22).

The vector version of this construction is named after the nineteenth-century French mathematician Charles Hermite, and called the *Hermitian dot product* on $\mathbb{C}^n$. It has the explicit formula

$$\mathbf{z} \cdot \mathbf{w} = \mathbf{z}^T\,\overline{\mathbf{w}} = z_1\,\overline{w}_1 + z_2\,\overline{w}_2 + \;\cdots\; + z_n\,\overline{w}_n, \quad \text{for} \quad \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}. \tag{3.98}$$

Pay attention to the fact that we must apply complex conjugation to all the entries of the second vector. For example, if

$$\mathbf{z} = \begin{pmatrix} 1 + \mathrm{i} \\ 3 + 2\,\mathrm{i} \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} 1 + 2\,\mathrm{i} \\ \mathrm{i} \end{pmatrix}, \quad \text{then} \quad \mathbf{z} \cdot \mathbf{w} = (1 + \mathrm{i}\,)(1 - 2\,\mathrm{i}\,) + (3 + 2\,\mathrm{i}\,)(-\,\mathrm{i}\,) = 5 - 4\,\mathrm{i}\,.$$

On the other hand,

$$\mathbf{w} \cdot \mathbf{z} = (1 + 2\,\mathrm{i}\,)(1 - \mathrm{i}\,) + \mathrm{i}\,(3 - 2\,\mathrm{i}\,) = 5 + 4\,\mathrm{i}\,,$$

and we conclude that the Hermitian dot product is *not* symmetric. Indeed, reversing the order of the vectors conjugates their dot product:

$$\mathbf{w} \cdot \mathbf{z} \;=\; \overline{\mathbf{z} \cdot \mathbf{w}}.$$

This is an unexpected complication, but it does have the desired effect that the induced norm, namely

$$0 \le \|\,\mathbf{z}\,\| = \sqrt{\mathbf{z} \cdot \mathbf{z}} = \sqrt{\mathbf{z}^T\,\overline{\mathbf{z}}} = \sqrt{|\,z_1\,|^2 + \;\cdots\; + |\,z_n\,|^2}\,, \tag{3.99}$$

is strictly positive for all $\mathbf{0} \ne \mathbf{z} \in \mathbb{C}^n$. For example, if

$$\mathbf{z} = \begin{pmatrix} 1 + 3\,\mathrm{i} \\ -\,2\,\mathrm{i} \\ -\,5 \end{pmatrix}, \qquad \text{then} \qquad \|\,\mathbf{z}\,\| = \sqrt{\,|\,1 + 3\,\mathrm{i}\,|^2 + |-2\,\mathrm{i}\,|^2 + |-5\,|^2} = \sqrt{39}\,.$$

The Hermitian dot product is well behaved under complex vector addition:

$$(\mathbf{z} + \widetilde{\mathbf{z}}\,) \cdot \mathbf{w} = \mathbf{z} \cdot \mathbf{w} + \widetilde{\mathbf{z}} \cdot \mathbf{w}, \qquad \mathbf{z} \cdot (\mathbf{w} + \widetilde{\mathbf{w}}\,) = \mathbf{z} \cdot \mathbf{w} + \mathbf{z} \cdot \widetilde{\mathbf{w}}.$$

However, while complex scalar multiples can be extracted from the first vector without alteration, when they multiply the second vector, they emerge as complex conjugates:

$$(c\,\mathbf{z}) \cdot \mathbf{w} = c\,(\mathbf{z} \cdot \mathbf{w}), \qquad \mathbf{z} \cdot (c\,\mathbf{w}) = \overline{c}\,(\mathbf{z} \cdot \mathbf{w}), \qquad c \in \mathbb{C}.$$

Thus, the Hermitian dot product is not bilinear in the strict sense, but satisfies something that, for lack of a better name, is known as *sesquilinearity*.

The general definition of an inner product on a complex vector space is modeled on the preceding properties of the Hermitian dot product.

Definition 3.48. An *inner product* on the complex vector space V is a pairing that takes two vectors $\mathbf{v}, \mathbf{w} \in V$ and produces a complex number $\langle \mathbf{v}, \mathbf{w} \rangle \in \mathbb{C}$, subject to the following requirements, for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, and $c, d \in \mathbb{C}$:

(*i*) *Sesquilinearity*:

$$\begin{aligned} \langle c\,\mathbf{u} + d\,\mathbf{v}, \mathbf{w} \rangle &= c\,\langle \mathbf{u}, \mathbf{w} \rangle + d\,\langle \mathbf{v}, \mathbf{w} \rangle, \\ \langle \mathbf{u}, c\,\mathbf{v} + d\,\mathbf{w} \rangle &= \overline{c}\,\langle \mathbf{u}, \mathbf{v} \rangle + \overline{d}\,\langle \mathbf{u}, \mathbf{w} \rangle. \end{aligned} \tag{3.100}$$

(*ii*) *Conjugate Symmetry*:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}. \tag{3.101}$$

(*iii*) *Positivity*:

$$\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle \geq 0, \qquad \text{and} \qquad \langle \mathbf{v}, \mathbf{v} \rangle = 0 \quad \text{if and only if} \quad \mathbf{v} = \mathbf{0}. \tag{3.102}$$

Thus, when dealing with a complex inner product space, one must pay careful attention to the complex conjugate that appears when the second argument in the inner product is multiplied by a complex scalar, as well as the complex conjugate that appears when the order of the two arguments is reversed. But, once this initial complication has been properly taken into account, the further properties of the inner product carry over directly from the real domain. Exercise 3.6.45 contains the formula for a general inner product on the complex vector space $\mathbb{C}^n$.

Theorem 3.49. The Cauchy–Schwarz inequality,

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \;\leq\; \|\mathbf{v}\|\,\|\mathbf{w}\|, \tag{3.103}$$

with $|\cdot|$ now denoting the complex modulus, and the triangle inequality

$$\|\mathbf{v} + \mathbf{w}\| \;\leq\; \|\mathbf{v}\| + \|\mathbf{w}\| \tag{3.104}$$

are both valid on an arbitrary complex inner product space.

The proof of (3.103–104) is modeled on the real case, and the details are left to the reader.

Example 3.50. The vectors $\mathbf{v} = (\,1 + \mathrm{i}\,, 2\,\mathrm{i}\,, -3\,)^T$, $\mathbf{w} = (\,2 - \mathrm{i}\,, 1, 2 + 2\,\mathrm{i}\,)^T$, satisfy

$$\|\mathbf{v}\| = \sqrt{2+4+9} = \sqrt{15}, \qquad \|\mathbf{w}\| = \sqrt{5+1+8} = \sqrt{14},$$
$$\mathbf{v} \cdot \mathbf{w} = (1+\mathrm{i}\,)(2+\mathrm{i}\,) + 2\,\mathrm{i} + (-3)(2-2\,\mathrm{i}\,) = -5 + 11\,\mathrm{i}\,.$$

Thus, the Cauchy–Schwarz inequality reads

$$|\langle \mathbf{v}, \mathbf{w} \rangle| = |-5 + 11\,\mathrm{i}\,| = \sqrt{146} \;\leq\; \sqrt{210} = \sqrt{15}\,\sqrt{14} = \|\mathbf{v}\|\,\|\mathbf{w}\|.$$

Similarly, the triangle inequality tells us that

$$\|\mathbf{v} + \mathbf{w}\| = \|(\,3, 1 + 2\,\mathrm{i}\,, -1 + 2\,\mathrm{i}\,)^T\| = \sqrt{9+5+5} = \sqrt{19} \leq \sqrt{15} + \sqrt{14} = \|\mathbf{v}\| + \|\mathbf{w}\|.$$

Example 3.51. Let $\mathrm{C}^0[-\pi, \pi]$ denote the complex vector space consisting of all complex-valued continuous functions $f(x) = u(x) + \mathrm{i}\,v(x)$ depending upon the *real* variable

$-\pi \le x \le \pi$. The Hermitian L^2 inner product on $\mathrm{C}^0[-\pi,\pi]$ is defined as

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)\, \overline{g(x)}\, dx\,, \tag{3.105}$$

i.e., the integral of f times the complex conjugate of g, with corresponding norm

$$\| f \| = \sqrt{\int_{-\pi}^{\pi} |\, f(x) \,|^2 \, dx} = \sqrt{\int_{-\pi}^{\pi} \big[\, u(x)^2 + v(x)^2 \,\big]\, dx}\,. \tag{3.106}$$

The reader can verify that (3.105) satisfies the Hermitian inner product axioms.

In particular, if k, l are integers, then the inner product of the complex exponential functions $e^{\,\mathrm{i}\,kx}$ and $e^{\,\mathrm{i}\,lx}$ is

$$\langle e^{\,\mathrm{i}\,kx}, e^{\,\mathrm{i}\,lx} \rangle = \int_{-\pi}^{\pi} e^{\,\mathrm{i}\,kx} e^{-\,\mathrm{i}\,lx}\, dx = \int_{-\pi}^{\pi} e^{\,\mathrm{i}\,(k-l)x}\, dx = \begin{cases} 2\pi, & k = l, \\ \left.\dfrac{e^{\,\mathrm{i}\,(k-l)x}}{\mathrm{i}\,(k-l)}\right|_{x=-\pi}^{\pi} = 0, & k \ne l. \end{cases}$$

We conclude that when $k \ne l$, the complex exponentials $e^{\,\mathrm{i}\,kx}$ and $e^{\,\mathrm{i}\,lx}$ are orthogonal, since their inner product is zero. The complex formulation of Fourier analysis, [**61**, **77**], is founded on this key example.

Exercises

3.6.26. Determine whether the indicated sets of complex vectors are linearly independent or dependent. (*a*) $\begin{pmatrix} \mathrm{i} \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix}$, (*b*) $\begin{pmatrix} 1+\mathrm{i} \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1-\mathrm{i} \end{pmatrix}$, (*c*) $\begin{pmatrix} 1+3\,\mathrm{i} \\ 2-\mathrm{i} \end{pmatrix}, \begin{pmatrix} 2-3\,\mathrm{i} \\ 1-\mathrm{i} \end{pmatrix}$, (*d*) $\begin{pmatrix} -2+\mathrm{i} \\ \mathrm{i} \end{pmatrix}, \begin{pmatrix} 4-3\,\mathrm{i} \\ 1 \end{pmatrix}, \begin{pmatrix} 2\,\mathrm{i} \\ 1-5\,\mathrm{i} \end{pmatrix}$, (*e*) $\begin{pmatrix} 1+2\,\mathrm{i} \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1-\mathrm{i} \end{pmatrix}$, (*f*) $\begin{pmatrix} 1 \\ 3\,\mathrm{i} \\ 2-\mathrm{i} \end{pmatrix}, \begin{pmatrix} 1+2\,\mathrm{i} \\ -3 \\ 0 \end{pmatrix}, \begin{pmatrix} 1-\mathrm{i} \\ -\mathrm{i} \\ 1 \end{pmatrix}$, (*g*) $\begin{pmatrix} 1+\mathrm{i} \\ 2-\mathrm{i} \\ 1 \end{pmatrix}, \begin{pmatrix} 1-\mathrm{i} \\ -3\,\mathrm{i} \\ 1-2\,\mathrm{i} \end{pmatrix}, \begin{pmatrix} -1+\mathrm{i} \\ 2+3\,\mathrm{i} \\ 1+2\,\mathrm{i} \end{pmatrix}$.

3.6.27. *True or false*: The set of complex vectors of the form $\begin{pmatrix} z \\ \overline{z} \end{pmatrix}$ for $z \in \mathbb{C}$ is a subspace of $\mathbb{C}^2$.

3.6.28. (a) Determine whether the vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ \mathrm{i} \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1+\mathrm{i} \\ 2 \end{pmatrix}$, $\mathbf{v}_3 = \begin{pmatrix} -1+\mathrm{i} \\ 1+\mathrm{i} \\ -1 \end{pmatrix}$, are linearly independent or linearly dependent. (*b*) Do they form a basis of $\mathbb{C}^3$? (*c*) Compute the Hermitian norm of each vector. (*d*) Compute the Hermitian dot products between all different pairs. Which vectors are orthogonal?

3.6.29. Find the dimension of and a basis for the following subspaces of $\mathbb{C}^3$: (*a*) The set of all complex multiples of $(\,1, \mathrm{i}\,, 1-\mathrm{i}\,)^T$. (*b*) The plane $z_1 + \mathrm{i}\,z_2 + (1-\mathrm{i}\,)z_3 = 0$. (*c*) The image of the matrix $A = \begin{pmatrix} 1 & \mathrm{i} & 2-\mathrm{i} \\ 2+\mathrm{i} & 1+3\,\mathrm{i} & -1-\mathrm{i} \end{pmatrix}$. (*d*) The kernel of the same matrix. (*e*) The set of vectors that are orthogonal to $(\,1-\mathrm{i}\,, 2\,\mathrm{i}\,, 1+\mathrm{i}\,)^T$.

3.6.30. Find bases for the four fundamental subspaces associated with the complex matrices

(a) $\begin{pmatrix} \mathrm{i} & 2 \\ -1 & 2\,\mathrm{i} \end{pmatrix}$, (b) $\begin{pmatrix} 2 & -1+\mathrm{i} & 1-2\,\mathrm{i} \\ -4 & 3-\mathrm{i} & 1+\mathrm{i} \end{pmatrix}$, (c) $\begin{pmatrix} \mathrm{i} & -1 & 2-\mathrm{i} \\ -1+2\,\mathrm{i} & -2-\mathrm{i} & 3 \\ \mathrm{i} & -1 & 1+\mathrm{i} \end{pmatrix}$.

3.6.31. Prove that $\mathbf{v} = \mathbf{x} + \mathrm{i}\,\mathbf{y}$ and $\overline{\mathbf{v}} = \mathbf{x} - \mathrm{i}\,\mathbf{y}$ are linearly independent complex vectors if and only if their real and imaginary parts $\mathbf{x}$ and $\mathbf{y}$ are linearly independent real vectors.

3.6.32. Prove that the space of complex $m \times n$ matrices is a complex vector space. What is its dimension?

3.6.33. Determine which of the following are subspaces of the vector space consisting of all complex 2×2 matrices. (*a*) All matrices with real diagonals. (*b*) All matrices for which the sum of the diagonal entries is zero. (*c*) All singular complex matrices. (*d*) All matrices whose determinant is real. (*e*) All matrices of the form $\begin{pmatrix} a & b \\ \overline{a} & \overline{b} \end{pmatrix}$, where $a, b \in \mathbb{C}$.

3.6.34. *True or false*: The set of all complex-valued functions $u(x) = v(x) + \mathrm{i}\,w(x)$ with $u(0) = \mathrm{i}$ is a subspace of the vector space of complex-valued functions.

3.6.35. Let V denote the complex vector space spanned by the functions 1, $e^{\mathrm{i}\,x}$ and $e^{-\mathrm{i}\,x}$, where x is a real variable. Which of the following functions belong to V?
(*a*) $\sin x$, (*b*) $\cos x - 2\,\mathrm{i} \sin x$, (*c*) $\cosh x$, (*d*) $\sin^2 \frac{1}{2} x$, (*e*) $\cos^2 x$?

3.6.36. Prove that the following define Hermitian inner products on $\mathbb{C}^2$:
(*a*) $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 \overline{w}_1 + 2 v_2 \overline{w}_2$, (*b*) $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 \overline{w}_1 + \mathrm{i}\, v_1 \overline{w}_2 - \mathrm{i}\, v_2 \overline{w}_1 + 2 v_2 \overline{w}_2$.

3.6.37. Which of the following define inner products on $\mathbb{C}^2$? (*a*) $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 \overline{w}_1 + 2\,\mathrm{i}\, v_2 \overline{w}_2$, (*b*) $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + 2 v_2 w_2$, (*c*) $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 \overline{w}_2 + v_2 \overline{w}_1$, (*d*) $\langle \mathbf{v}, \mathbf{w} \rangle = 2 v_1 \overline{w}_1 + v_1 \overline{w}_2 + v_2 \overline{w}_1 + 2 v_2 \overline{w}_2$, (*e*) $\langle \mathbf{v}, \mathbf{w} \rangle = 2 v_1 \overline{w}_1 + (1 + \mathrm{i})\, v_1 \overline{w}_2 + (1 - \mathrm{i})\, v_2 \overline{w}_1 + 3 v_2 \overline{w}_2$.

♢ 3.6.38. Let $A = A^T$ be a real symmetric $n \times n$ matrix. Show that $(A\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (A\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$.

3.6.39. Let $\mathbf{z} = \mathbf{x} + \mathrm{i}\,\mathbf{y} \in \mathbb{C}^n$.
(*a*) Prove that, for the Hermitian dot product, $\| \mathbf{z} \|^2 = \| \mathbf{x} \|^2 + \| \mathbf{y} \|^2$.
(*b*) Does this formula remain valid under a more general Hermitian inner product on $\mathbb{C}^n$?

♢ 3.6.40. Let V be a complex inner product space. Prove that, for all $\mathbf{z}, \mathbf{w} \in V$,
(*a*) $\| \mathbf{z} + \mathbf{w} \|^2 = \| \mathbf{z} \|^2 + 2 \operatorname{Re} \langle \mathbf{z}, \mathbf{w} \rangle + \| \mathbf{w} \|^2$;
(*b*) $\langle \mathbf{z}, \mathbf{w} \rangle = \frac{1}{4} \left(\| \mathbf{z} + \mathbf{w} \|^2 - \| \mathbf{z} - \mathbf{w} \|^2 + \mathrm{i} \| \mathbf{z} + \mathrm{i}\,\mathbf{w} \|^2 - \mathrm{i} \| \mathbf{z} - \mathrm{i}\,\mathbf{w} \|^2 \right)$.

♢ 3.6.41. (*a*) How would you define the angle between two elements of a complex inner product space? (*b*) What is the angle between $(-1, 2 - \mathrm{i}, -1 + 2\,\mathrm{i})^T$ and $(-2 - \mathrm{i}, -\mathrm{i}, 1 - \mathrm{i})^T$ relative to the Hermitian dot product?

3.6.42. Let $\mathbf{0} \neq \mathbf{v} \in \mathbb{C}^n$. Which scalar multiples $c\,\mathbf{v}$ have the same Hermitian norm as $\mathbf{v}$?

♢ 3.6.43. Prove the Cauchy–Schwarz inequality (3.103) and the triangle inequality (3.104) for a general complex inner product. *Hint*: Use Exercises 3.6.8, 3.6.40(*a*).

♢ 3.6.44. The *Hermitian adjoint* of a complex $m \times n$ matrix A is the complex conjugate of its transpose, written $A^\dagger = \overline{A^T} = \overline{A}^T$.
For example, if $A = \begin{pmatrix} 1 + \mathrm{i} & 2\,\mathrm{i} \\ -3 & 2 - 5\,\mathrm{i} \end{pmatrix}$, then $A^\dagger = \begin{pmatrix} 1 - \mathrm{i} & -3 \\ -2\,\mathrm{i} & 2 + 5\,\mathrm{i} \end{pmatrix}$. Prove that
(*a*) $(A^\dagger)^\dagger = A$, (*b*) $(zA + wB)^\dagger = \overline{z} A^\dagger + \overline{w} B^\dagger$ for $z, w \in \mathbb{C}$, (*c*) $(AB)^\dagger = B^\dagger A^\dagger$.

♢ 3.6.45. A complex matrix H is called *Hermitian* if it equals its Hermitian adjoint, $H^\dagger = H$, as defined in the preceding exercise. (*a*) Prove that the diagonal entries of a Hermitian matrix are real. (*b*) Prove that $(H\mathbf{z}) \cdot \mathbf{w} = \mathbf{z} \cdot (H\mathbf{w})$ for $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$. (*c*) Prove that every Hermitian inner product on $\mathbb{C}^n$ has the form $\langle \mathbf{z}, \mathbf{w} \rangle = \mathbf{z}^T H \overline{\mathbf{w}}$, where H is an $n \times n$ positive definite Hermitian matrix. (*d*) How would you verify positive definiteness of a complex matrix?

3.6.46. *Multiple choice*: Let V be a complex normed vector space. How many unit vectors are parallel to a given vector $\mathbf{0} \neq \mathbf{v} \in V$? (a) none; (b) 1; (c) 2; (d) 3; (e) ∞; (f) depends upon the vector; (g) depends on the norm. Explain your answer.

♢ 3.6.47. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be elements of a complex inner product space. Let K denote the corresponding $n \times n$ *Gram matrix*, defined in the usual manner.

(a) Prove that K is a Hermitian matrix, as defined in Exercise 3.6.45.

(b) Prove that K is positive semi-definite, meaning $\mathbf{z}^T K \mathbf{z} \geq 0$ for all $\mathbf{z} \in \mathbb{C}^n$.

(c) Prove that K is positive definite if and only if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent.

3.6.48. For each of the following pairs of complex-valued functions,
(i) compute their L^2 norm and Hermitian inner product on the interval $[0,1]$, and then
(ii) check the validity of the Cauchy–Schwarz and triangle inequalities.

(a) $1,\ e^{\mathrm{i}\,\pi x}$; (b) $x + \mathrm{i},\ x - \mathrm{i}$; (c) $\mathrm{i}\,x^2,\ (1 - 2\,\mathrm{i})x + 3\,\mathrm{i}$.

3.6.49. Formulate conditions on a weight function $w(x)$ that guarantee that the weighted integral $\langle f, g \rangle = \int_a^b f(x)\,\overline{g(x)}\,w(x)\,dx$ defines an inner product on the space of continuous complex-valued functions on $[a,b]$.

3.6.50. (a) Formulate a general definition of a norm on a complex vector space.

(b) How would you define analogues of the $\mathrm{L}^1, \mathrm{L}^2$ and L^∞ norms on $\mathbb{C}^n$?