Chapter 3

Inner Products and Norms

The geometry of Euclidean space relies 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 governed 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 analysis, both linear and nonlinear, in both finite-dimensional vector spaces and infinite-dimensional function spaces. It is impossible to overemphasize their importance for both theoretical developments, practical applications, and in the design of numerical solution algorithms. We begin this chapter with a discussion of the basic properties of inner products, illustrated by some of the most important examples.

Mathematical analysis is founded on inequalities. The most basic is the Cauchy–Schwarz inequality, which is valid in any inner product space. The more familiar triangle inequality for the associated norm is then derived as a simple consequence. Not every norm arises from an inner product, and, in more general norms, the triangle inequality becomes part of the definition. Both inequalities retain their validity in both finite-dimensional and infinite-dimensional vector spaces. Indeed, their abstract formulation helps us focus on the key ideas in the proof, avoiding all distracting complications resulting from the explicit formulas.

In Euclidean space $\mathbb{R}^n$, the characterization of general inner products will lead us to an extremely important class of matrices. Positive definite matrices play a key role in a variety of applications, including minimization problems, least squares, mechanical systems, electrical circuits, and the differential equations describing dynamical processes. Later, we will generalize the notion of positive definiteness to more general linear operators, governing the ordinary and partial differential equations arising in continuum mechanics and dynamics. Positive definite matrices most commonly appear in so-called Gram matrix form, consisting of the inner products between selected elements of an inner product space. The test for positive definiteness is based on Gaussian elimination. Indeed, the associated matrix factorization can be reinterpreted as the process of completing the square for the associated quadratic form.

So far, we have confined our attention to real vector spaces. Complex numbers, vectors and functions also play an important role in applications, and so, in the final section, we formally introduce complex vector spaces. Most of the formulation proceeds in direct analogy with the real version, but the notions of inner product and norm on complex vector spaces requires some thought. Applications of complex vector spaces and their inner products are of particular importance in Fourier analysis and signal processing, and absolutely essential in modern quantum mechanics.
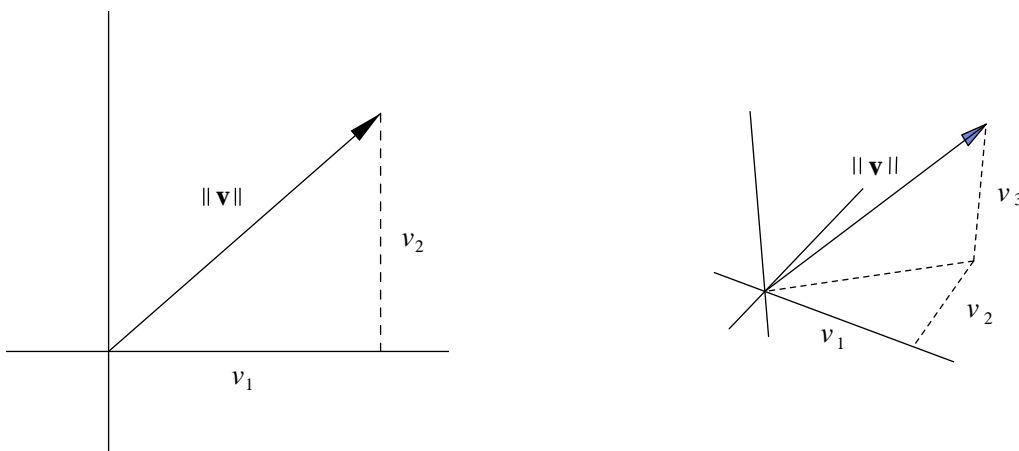
Figure 3.1. The Euclidean Norm in $\mathbb{R}^2$ and $\mathbb{R}^3$.

3.1. Inner Products.

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

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

between (column) vectors $\mathbf{v} = (\, v_1, v_2, \ldots, v_n \,)^T, \mathbf{w} = (\, w_1, w_2, \ldots, w_n \,)^T$ lying in the Euclidean space $\mathbb{R}^n$. An important observation is that the dot product (3.1) can be identified with 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 a row vector $\mathbf{v}^T$ and a 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, as a consequence of the classical Pythagorean Theorem, equal to 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 length, $\|\, \mathbf{v} \,\| \geq 0$, while only the zero vector has length $\|\, \mathbf{0} \,\| = 0$. The dot product and Euclidean norm satisfy certain evident properties, and these 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 $c, d \in \mathbb{R}$.

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

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

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

A vector space equipped with an inner product is called an *inner product space*. As we shall see, a given vector space can admit many different inner products. Verification of the inner product axioms for the Euclidean dot product is straightforward, and left to 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 basic inner product on $\mathbb{R}^2$, the dot product $\mathbf{v}\cdot\mathbf{w} = v_1\,w_1 + v_2\,w_2$ 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,

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

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 ■.) 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, verifying the positivity requirement (3.6). This serves to establish (3.8) as an 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 obvious. Positivity is ensured by noticing that

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

and is strictly positive for all $\mathbf{v} \ne \mathbf{0}$. Therefore, (3.9) defines another inner product on $\mathbb{R}^2$, with associated norm $\|\,\mathbf{v}\,\| \;=\; \sqrt{v_1^2 - 2\,v_1\,v_2 + 4\,v_2^2}\,$.

Example 3.3. Let $c_1, \ldots, c_n$ 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*. The larger the weight c_i, the more the ith coordinate of $\mathbf{v}$ contributes to the norm. Weighted norms are particularly important in statistics and data fitting, where one wants to emphasize certain quantities and de-emphasize others; this is done by assigning suitable weights to the different components of the data vector $\mathbf{v}$. Section 4.3 on least squares approximation methods will contain further details.

Inner Products on Function Space

Inner products and norms on function spaces play an absolutely essential role in 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 $f(x)$ defined for $a \le x \le b$. The integral of the product of two continuous functions

$$\langle\, f\,;g\,\rangle = \int_a^b f(x)\,g(x)\,dx \tag{3.11}$$

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),

$$\|\,f\,\| = \sqrt{\int_a^b f(x)^2\,dx}\,, \tag{3.12}$$

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 $\mathbb{R}^n$.

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

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

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 (3.11) 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:

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

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 ∎. 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, $\int_a^b f(x)^2 \, dx = 0$ if and only if $f(x) \equiv 0$ is the zero function, cf. Exercise ∎. This completes the demonstration that (3.11) defines a bona fide inner product on the function space $\mathrm{C}^0[a, b]$.
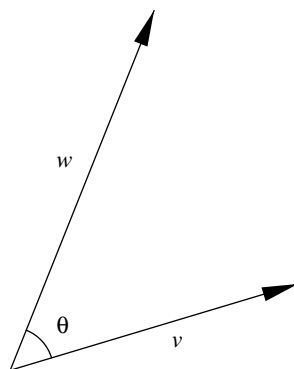
Figure 3.2. Angle Between Two Vectors.

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 important inner product is known as *Hilbert space*, which forms the foundation for modern analysis, [**126**], including Fourier series, [**51**], and also lies at the heart of modern quantum mechanics, [**100**, **104**, **122**]. One does need to be extremely careful when trying to extend the inner product to more general functions. Indeed, there are nonzero, discontinuous functions with zero "L^2 norm". An example is

$$f(x) = \begin{cases} 1, & x = 0, \\ 0, & \text{otherwise,} \end{cases} \qquad \text{which satisfies} \qquad \| f \|^2 = \int_{-1}^{1} f(x)^2 \, dx = 0 \qquad (3.13)$$

because any function which is zero except at finitely many (or even countably many) points has zero integral. We will discuss some of the details of the Hilbert space construction in Chapters 12 and 13.

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

$$\langle f \, ; g \, \rangle = \int_a^b f(x) \, g(x) \, w(x) \, dx, \qquad \| f \| = \sqrt{\int_a^b f(x)^2 \, w(x) \, dx} \, . \qquad (3.14)$$

The verification of the inner product axioms in this case is left as an exercise for the reader. As in the finite-dimensional versions, weighted inner products play a key role in statistics and data analysis.

3.2. Inequalities.

There are two absolutely fundamental inequalities that are valid for *any* inner product on any vector 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, Augustin Cauchy and Herman Schwarz, 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 spotlight. Specilizing to different inner products and norms on both finite-dimensional and infinite-dimensional vector spaces leads to a wide variety of striking, and useful particular cases.

The Cauchy–Schwarz Inequality

In two and three-dimensional Euclidean geometry, the dot product between two vectors can be geometrically characterized by the equation

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \theta, \tag{3.15}$$

where θ measures the angle between the vectors $\mathbf{v}$ and $\mathbf{w}$, as drawn 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 simple, 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*

$$|\langle \mathbf{v} \, ; \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\|, \qquad \mathbf{v}, \mathbf{w} \in V. \tag{3.16}$$

Here, $\|\mathbf{v}\|$ is the associated norm, while $|\cdot|$ denotes absolute value of real numbers. Equality holds if and only if $\mathbf{v}$ and $\mathbf{w}$ are parallel[†] vectors.

Proof: The case when $\mathbf{w} = \mathbf{0}$ is trivial, since both sides of (3.16) are equal to 0. Thus, we may suppose $\mathbf{w} \neq \mathbf{0}$. Let $t \in \mathbb{R}$ be an arbitrary scalar. Using the three basic inner product axioms, we have

$$0 \leq \|\mathbf{v} + t\,\mathbf{w}\|^2 = \langle \mathbf{v} + t\,\mathbf{w} \, ; \mathbf{v} + t\,\mathbf{w} \rangle = \|\mathbf{v}\|^2 + 2t\,\langle \mathbf{v} \, ; \mathbf{w} \rangle + t^2\,\|\mathbf{w}\|^2, \tag{3.17}$$

[†] Russians also give credit for its discovery to their compatriot Viktor Bunyakovskii, and, indeed, many authors append his name to the inequality.

[†] Recall that two vectors are parallel if and only if one is a scalar multiple of the other. The zero vector is parallel to every other vector, by convention.

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 (3.17) as a quadratic function,

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

of the scalar variable t. To get the maximum mileage out of the fact that $p(t) \geq 0$, let us look at where it assumes its minimum. This occurs when its derivative vanishes:

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

Substituting this particular minimizing value into (3.17), we find

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

Rearranging this last inequality, we conclude that

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

Taking the (positive) square root of both sides of the final inequality completes the proof of the Cauchy–Schwarz inequality (3.16). $\hfill$ *Q.E.D.*

Given any inner product on a vector space, we can use the quotient

$$\cos\theta = \frac{\langle\,\mathbf{v}\,;\mathbf{w}\,\rangle}{\|\,\mathbf{v}\,\|\,\|\,\mathbf{w}\,\|} \tag{3.18}$$

to *define* the "angle" between the 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, and, in fact, unique if we restrict it to lie in the range $0 \leq \theta \leq \pi$.

For example, using the standard dot product on $\mathbb{R}^3$, the angle between the vectors $\mathbf{v} = (\,1, 0, 1\,)^T$ and $\mathbf{w} = (\,0, 1, 1\,)^T$ is given by

$$\cos\theta = \frac{1}{\sqrt{2}\cdot\sqrt{2}} = \frac{1}{2}, \qquad \text{and so} \qquad \theta = \tfrac{1}{3}\,\pi = 1.0472\ldots, \quad \text{i.e.,} \quad 60°.$$

On the other hand, if we use 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

$$\cos\theta = \frac{3}{2\,\sqrt{5}} = .67082\ldots, \qquad \text{whereby} \qquad \theta = .835482\ldots.$$

Thus, the measurement of angle (and length) is dependent upon 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

$$\cos\theta \;=\; \frac{\langle\,x\,;x^2\,\rangle}{\|\,x\,\|\,\|\,x^2\,\|} \;=\; \frac{\displaystyle\int_0^1 x^3\,dx}{\sqrt{\displaystyle\int_0^1 x^2\,dx}\,\sqrt{\displaystyle\int_0^1 x^4\,dx}} \;=\; \frac{\frac{1}{4}}{\sqrt{\frac{1}{3}}\,\sqrt{\frac{1}{5}}} \;=\; \sqrt{\frac{15}{16}}\,,$$

so that $\theta = 0.25268$ radians. *Warning*: One 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 graph. 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 $\langle\, x \,;\, x^2 \,\rangle = \displaystyle\int_{-1}^{1} x^3 \, dx = 0$, and hence (3.18) becomes $\cos\theta = 0$, so the "angle" between x and x^2 is now $\theta = \frac{1}{2}\pi$.

Orthogonal Vectors

In Euclidean geometry, a particularly noteworthy configuration occurs when two vectors are *perpendicular*, which means that they meet at a right angle: $\theta = \frac{1}{2}\pi$ or $\frac{3}{2}\pi$, and so $\cos\theta = 0$. The angle formula (3.15) 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 also plays a key role 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: $\langle\, \mathbf{v} \,;\, \mathbf{w} \,\rangle = 0$.

Orthogonality is a remarkably powerful tool in all applications of linear algebra, and often serves to dramatically simplify many computations. We will devote all of Chapter 5 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 $90°$ angle. However, these vectors are *not* orthogonal with respect to the weighted inner product (3.8):

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

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 $\langle\, p \,;\, q \,\rangle = \displaystyle\int_0^1 p(x)\, q(x)\, dx$ 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 $[0, 2]$,

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

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
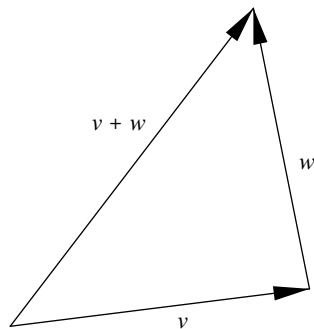
Figure 3.3.   Triangle Inequality.

first two side are represented by vectors $\mathbf{v}$ and $\mathbf{w}$, then the third corresponds to their sum $\mathbf{v} + \mathbf{w}$, and so $\|\,\mathbf{v} + \mathbf{w}\,\| \leq \|\,\mathbf{v}\,\| + \|\,\mathbf{w}\,\|$. The triangle inequality is a direct consequence of the Cauchy–Schwarz inequality, and hence holds for *any* inner product space.

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

$$\|\,\mathbf{v} + \mathbf{w}\,\| \ \leq \ \|\,\mathbf{v}\,\| + \|\,\mathbf{w}\,\| \tag{3.19}$$

for every $\mathbf{v}, \mathbf{w} \in V$. *Equality holds if and only if* $\mathbf{v}$ *and* $\mathbf{w}$ *are parallel vectors.*

Proof: We compute

$$\|\,\mathbf{v} + \mathbf{w}\,\|^2 = \langle\, \mathbf{v} + \mathbf{w} \,;\, \mathbf{v} + \mathbf{w} \,\rangle = \|\,\mathbf{v}\,\|^2 + 2\,\langle\, \mathbf{v}\,;\mathbf{w} \,\rangle + \|\,\mathbf{w}\,\|^2$$
$$\leq \|\,\mathbf{v}\,\|^2 + 2\,\|\,\mathbf{v}\,\|\,\|\,\mathbf{w}\,\| + \|\,\mathbf{w}\,\|^2 = \big(\,\|\,\mathbf{v}\,\| + \|\,\mathbf{w}\,\|\,\big)^2,$$

where the inequality follows from Cauchy–Schwarz. Taking square roots of both sides and using positivity 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 (3.19) in this case says $\sqrt{17} \leq \sqrt{6} + \sqrt{13}$, which is valid.

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

$$\|\,f\,\| \ = \ \sqrt{\int_0^1 (x-1)^2\, dx} \ = \ \sqrt{\frac{1}{3}}, \qquad\qquad \|\,g\,\| \ = \ \sqrt{\int_0^1 (x^2+1)^2\, dx} \ = \ \sqrt{\frac{23}{15}},$$

$$\|\,f+g\,\| \ = \ \sqrt{\int_0^1 (x^2+x)^2\, dx} \ = \ \sqrt{\frac{77}{60}}.$$

The triangle inequality requires $\sqrt{\frac{77}{60}} \leq \sqrt{\frac{1}{3}} + \sqrt{\frac{23}{15}}$, which is true.

2/25/04   97   ⓒ 2004   Peter J. Olver

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

$$\left| \sum_{i=1}^{n} v_i w_i \right| \leq \sqrt{\sum_{i=1}^{n} v_i^2} \; \sqrt{\sum_{i=1}^{n} w_i^2} \,,$$

$$\sqrt{\sum_{i=1}^{n} (v_i + w_i)^2} \leq \sqrt{\sum_{i=1}^{n} v_i^2} + \sqrt{\sum_{i=1}^{n} w_i^2} \,. \tag{3.20}$$

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 (3.12) on function space, they produce the following splendid integral inequalities:

$$\left| \int_a^b f(x)\, g(x)\, dx \right| \leq \sqrt{\int_a^b f(x)^2\, dx} \; \sqrt{\int_a^b g(x)^2\, dx} \,,$$

$$\sqrt{\int_a^b \left[f(x) + g(x) \right]^2 dx} \leq \sqrt{\int_a^b f(x)^2\, dx} + \sqrt{\int_a^b g(x)^2\, dx} \,, \tag{3.21}$$

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.

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 the vector space V assigns a real number $\| \mathbf{v} \|$ to each vector $\mathbf{v} \in V$, subject to the following axioms for all $\mathbf{v}, \mathbf{w} \in V$, 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. Here are some important examples of norms that do not come from inner products.

Example 3.13. Let $V = \mathbb{R}^n$. The 1–*norm* of a vector $\mathbf{v} = (\,\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_n\,)^T$ is defined as the sum of the absolute values of its entries:

$$\|\,\mathbf{v}\,\|_1 = |\,v_1\,| + |\,v_2\,| + \cdots + |\,v_n\,|. \tag{3.22}$$

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

$$\|\,\mathbf{v}\,\|_\infty = \sup\{\,|\,v_1\,|, |\,v_2\,|, \ldots, |\,v_n\,|\,\}. \tag{3.23}$$

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

$$|\,a + b\,| \le |\,a\,| + |\,b\,|$$

for absolute values.

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

$$\|\,\mathbf{v}\,\|_p = \sqrt[p]{\sum_{i=1}^{n} |\,v_i\,|^p\,}. \tag{3.24}$$

This quantity defines a norm for any $1 \le p < \infty$. The ∞–norm is a limiting case of 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 straightforward. The triangle inequality, however, is not trivial; in detail, it reads

$$\sqrt[p]{\sum_{i=1}^{n} |\,v_i + w_i\,|^p\,} \le \sqrt[p]{\sum_{i=1}^{n} |\,v_i\,|^p\,} + \sqrt[p]{\sum_{i=1}^{n} |\,w_i\,|^p\,}, \tag{3.25}$$

and is known as *Minkowski's inequality*. A proof can be found in [**97**].

Example 3.14. 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

$$\|\,f\,\|_p = \sqrt[p]{\int_a^b |\,f(x)\,|^p\,dx\,}. \tag{3.26}$$

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

$$\|\,f\,\|_1 = \int_a^b |\,f(x)\,|\,dx. \tag{3.27}$$

The L^2 norm (3.12) appears as a special case, $p = 2$, and, again, is the only one arising from an inner product. The proof of the general triangle or Minkowski inequality for $p \ne 1, 2$ is again not trivial, [**97**]. The limiting L^∞ norm is defined by the maximum

$$\|\,f\,\|_\infty = \max\{\,|\,f(x)\,| \,:\, a \le x \le b\,\}. \tag{3.28}$$

Example 3.15. Consider the polynomial $p(x) = 3x^2 - 2$ on the interval $-1 \leq x \leq 1$. Its L^2 norm is

$$\| p \|_2 = \sqrt{\int_{-1}^{1} (3x^2 - 2)^2 \, dx} \;\; = \;\; \sqrt{\frac{18}{5}} = 1.8974 \dots \, .$$

Its L^∞ norm is

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

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

$$\| p \|_1 = \int_{-1}^{1} | 3x^2 - 2 | \, dx$$

$$= \int_{-1}^{-\sqrt{2/3}} (3x^2 - 2) \, dx + \int_{-\sqrt{2/3}}^{\sqrt{2/3}} (2 - 3x^2) \, dx + \int_{\sqrt{2/3}}^{1} (3x^2 - 2) \, dx$$

$$= \left(\tfrac{4}{3} \sqrt{\tfrac{2}{3}} - 1 \right) + \tfrac{8}{3} \sqrt{\tfrac{2}{3}} + \left(\tfrac{4}{3} \sqrt{\tfrac{2}{3}} - 1 \right) = \tfrac{16}{3} \sqrt{\tfrac{2}{3}} - 2 = 2.3546 \dots \, .$$

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

$$d(\mathbf{v}, \mathbf{w}) = \| \mathbf{v} - \mathbf{w} \|. \tag{3.29}$$

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, nevertheless, satisfy all the familiar properties:

(a) Symmetry: $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v})$;

(b) $d(\mathbf{v}, \mathbf{w}) = 0$ if and only if $\mathbf{v} = \mathbf{w}$;

(c) The triangle inequality: $d(\mathbf{v}, \mathbf{w}) \leq d(\mathbf{v}, \mathbf{z}) + d(\mathbf{z}, \mathbf{w})$.

Unit Vectors

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

Lemma 3.16. *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:

$$\| \mathbf{u} \| = \left\| \frac{\mathbf{v}}{\| \mathbf{v} \|} \right\| = \frac{\| \mathbf{v} \|}{\| \mathbf{v} \|} = 1. \hspace{2cm} Q.E.D.$$

Example 3.17. 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 3.18. 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 correctly, a scalar multiple of) the polynomial p. On the other hand, for the L^∞ norm,

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

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

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

$$S_1 = \left\{ \| \mathbf{u} \| = 1 \right\} \subset V. \tag{3.30}$$

Thus, the unit sphere for the Euclidean norm on $\mathbb{R}^n$ is the usual round sphere

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

For the ∞ norm, it is the unit cube

$$S_1 = \left\{ \mathbf{x} \in \mathbb{R}^n \mid x_1 = \pm 1 \text{ or } x_2 = \pm 1 \text{ or } \ldots \text{ or } x_n = \pm 1 \right\}.$$

For the 1 norm, it is the unit diamond or "octahedron"

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

See Figure 3.4 for the two-dimensional pictures.

In all cases, the *closed unit ball* $B_1 = \left\{ \| \mathbf{u} \| \le 1 \right\}$ 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 forms a *compact* subset, meaning that it is closed and bounded. This basic topological fact, which is *not* true in infinite-dimensional
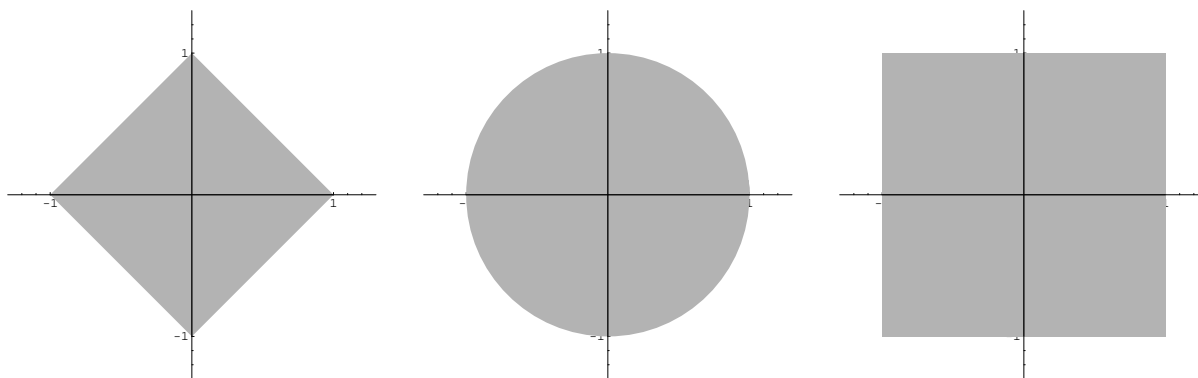
101

Figure 3.4. Unit Balls and Spheres for 1, 2 and ∞ Norms in $\mathbb{R}^2$.

spaces, underscores the fundamental distinction between finite-dimensional vector analysis and the vastly more complicated infinite-dimensional realm.

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 value, but rather that they are, in a certain sense, always close to one another, and so for most analytical purposes can be used interchangeably. As a consequence, we may be able to simplify the analysis of a problem by choosing a suitably adapted norm.

Theorem 3.19. *Let* $\| \cdot \|_1$ *and* $\| \cdot \|_2$ *be any two norms on* $\mathbb{R}^n$. *Then there exist positive constants* $c^\star, C^\star > 0$ *such that*

$$c^\star \, \| \, \mathbf{v} \, \|_1 \leq \| \, \mathbf{v} \, \|_2 \leq C^\star \, \| \, \mathbf{v} \, \|_1 \qquad \textit{for every} \qquad \mathbf{v} \in \mathbb{R}^n. \tag{3.31}$$

Proof: We just sketch the basic idea, leaving the details to a more rigorous real analysis course, cf. [**125**, **126**]. We begin by noting that a norm defines a continuous function $f(\mathbf{v}) = \| \, \mathbf{v} \, \|$ on $\mathbb{R}^n$. (Continuity is, in fact, a consequence of the triangle inequality.) Let $S_1 = \{ \, \| \, \mathbf{u} \, \|_1 = 1 \, \}$ denote the unit sphere of the first norm. Any 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

$$c^\star = \| \, \mathbf{u}^\star \, \|_2 = \min \{ \, \| \, \mathbf{u} \, \|_2 \mid \mathbf{u} \in S_1 \, \}, \quad C^\star = \| \, \mathbf{U}^\star \, \|_2 = \max \{ \, \| \, \mathbf{u} \, \|_2 \mid \mathbf{u} \in S_1 \, \}, \tag{3.32}$$

for certain vectors $\mathbf{u}^\star, \mathbf{U}^\star \in S_1$. Note that $0 < c^\star \leq C^\star < \infty$, with equality holding if and only if the the norms are the same. The minimum and maximum (3.32) will serve as the constants in the desired inequalities (3.31). Indeed, by definition,

$$c^\star \leq \| \, \mathbf{u} \, \|_2 \leq C^\star \qquad \text{when} \qquad \| \, \mathbf{u} \, \|_1 = 1, \tag{3.33}$$

and so (3.31) is valid for all $\mathbf{u} \in S_1$. To prove the inequalities in general, assume $\mathbf{v} \neq \mathbf{0}$. (The case $\mathbf{v} = \mathbf{0}$ is trivial.) Lemma 3.16 says that $\mathbf{u} = \mathbf{v}/\| \, \mathbf{v} \, \|_1 \in S_1$ 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 (3.33) and clearing denominators completes the proof of (3.31). *Q.E.D.*

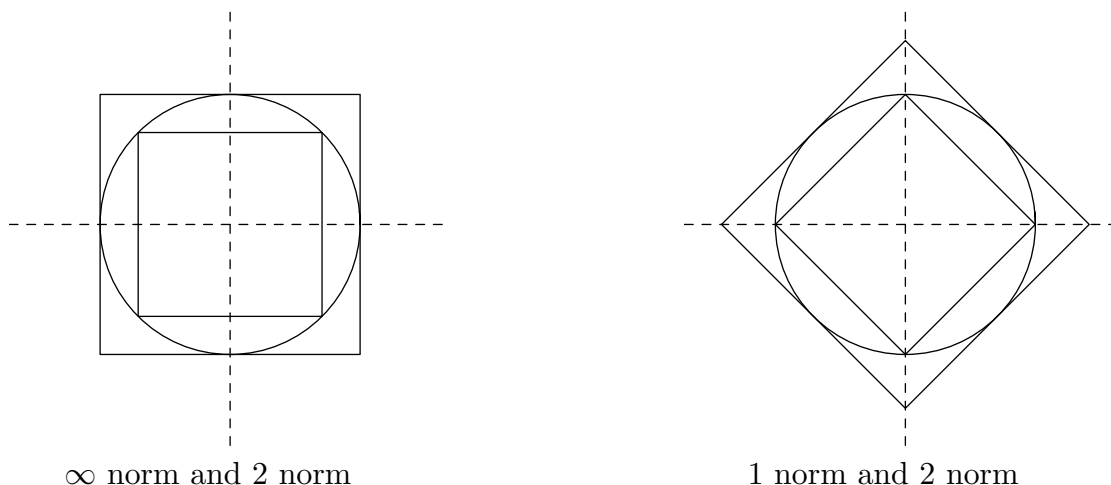∞ norm and 2 norm   1 norm and 2 norm

Figure 3.5.   Equivalence of Norms.

Example 3.20. For example, consider the Euclidean norm $\| \cdot \|_2$ and the max norm $\| \cdot \|_\infty$ on $\mathbb{R}^n$. According to (3.32), 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. Its maximal value is obtained at the poles, when $\mathbf{U}^\star = \pm\, \mathbf{e}_k$, with $\| \mathbf{e}_k \|_\infty = 1$. Thus, $C^\star = 1$. The minimal value is obtained when $\mathbf{u}^\star = \left(\dfrac{1}{\sqrt{n}}, \ldots, \dfrac{1}{\sqrt{n}} \right)$ has all equal components, whereby $c^\star = \| \mathbf{u} \|_\infty = 1/\sqrt{n}$. Therefore,

$$\frac{1}{\sqrt{n}}\, \| \mathbf{v} \|_2 \le \| \mathbf{v} \|_\infty \le \| \mathbf{v} \|_2. \tag{3.34}$$

One 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 (3.34) tells us that its ∞ norm is bounded from above and below by $1/\sqrt{n} \le \| \mathbf{v} \|_\infty \le 1$. Therefore, the unit Euclidean sphere sits inside the unit sphere in the ∞ norm, and outside the sphere of radius $1/\sqrt{n}$. Figure 3.5 illustrates the two-dimensional situation.

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 ε–δ 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 components 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 case, called *convergence in norm*, does not depend on which norm is chosen. Indeed, the basic inequality (3.31) implies that if one norm goes to zero, so does any other norm. An important 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., [**125**, **126**], to find the relevant definitions, theorems and proofs.

Example 3.21. 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 which have unit L^2 norm but arbitrarily small L^∞ norm. Similar comparative results can be established for the other function space norms. As a result, analysis and topology on function space is intimately related to the underlying choice of norm.

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 which can be placed on the finite-dimensional vector space $\mathbb{R}^n$. The resulting analysis will lead us to the extremely important class of positive definite matrices. Such matrices play a fundamental role in a wide variety of applications, including minimization problems, mechanics, electrical circuits, and differential equations. Moreover, their infinite-dimensional generalization to positive definite linear operators underlie all of the most important examples of boundary value problems for ordinary and partial differential equations.

Let $\langle \mathbf{x} \, ; \mathbf{y} \rangle$ denote an inner product between vectors $\mathbf{x} = (x_1 \ x_2 \ \ldots \ x_n)^T$ and $\mathbf{y} = (y_1 \ y_2 \ \ldots \ y_n)^T$, in $\mathbb{R}^n$. We begin by writing the vectors in terms of the standard basis vectors:

$$\mathbf{x} = x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n = \sum_{i=1}^n x_i \mathbf{e}_i, \qquad \mathbf{y} = y_1 \mathbf{e}_1 + \cdots + y_n \mathbf{e}_n = \sum_{j=1}^n y_j \mathbf{e}_j. \quad (3.35)$$

To evaluate their inner product, we will apply the three basic axioms. We first employ the bilinearity of the inner product 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 we can write

$$\langle \mathbf{x} \, ; \mathbf{y} \, \rangle = \sum_{i,j=1}^{n} k_{ij} \, x_i \, y_j = \mathbf{x}^T K \, \mathbf{y}, \tag{3.36}$$

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

$$k_{ij} = \langle \, \mathbf{e}_i \, ; \mathbf{e}_j \, \rangle, \qquad i, j = 1, \dots, n. \tag{3.37}$$

We conclude that any inner product must be expressed in the general *bilinear form* (3.36).

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

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

Consequently, the inner product matrix K is 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 is a scalar, and hence equals its transpose.

The final condition for an inner product is positivity. This requires that

$$\| \, \mathbf{x} \, \|^2 = \langle \, \mathbf{x} \, ; \mathbf{x} \, \rangle = \mathbf{x}^T K \mathbf{x} = \sum_{i,j=1}^{n} k_{ij} \, x_i \, x_j \geq 0 \qquad \text{for all} \qquad \mathbf{x} \in \mathbb{R}^n, \tag{3.38}$$

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

Definition 3.22. An $n \times n$ matrix K is called *positive definite* if it is symmetric, $K^T = K$, and satisfies the positivity condition

$$\mathbf{x}^T K \mathbf{x} > 0 \qquad \text{for all} \qquad \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n. \tag{3.39}$$

We will sometimes write $K > 0$ to mean that K is a symmetric, 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 which have some negative entries — see Example 3.24 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 *only* say that a matrix is positive definite 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 characterization of inner products on a finite-dimensional vector space.

Theorem 3.23. *Every inner product on $\mathbb{R}^n$ is given by*

$$\langle \mathbf{x} ; \mathbf{y} \rangle = \mathbf{x}^T K \mathbf{y}, \qquad \text{for} \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \tag{3.40}$$

where K is a symmetric, positive definite matrix.

Given any symmetric[†] matrix K, the homogeneous quadratic polynomial

$$q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x} = \sum_{i,j=1}^{n} k_{ij}\, x_i\, x_j, \tag{3.41}$$

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

$$q(\mathbf{x}) > 0 \qquad \text{for all} \qquad 0 \neq \mathbf{x} \in \mathbb{R}^n. \tag{3.42}$$

Thus, a quadratic form is positive definite if and only if its coefficient matrix is.

Example 3.24. 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

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

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

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

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

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

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 or is not positive definite is not completely elementary.

With a little practice, it is not difficult to read off the coefficient matrix K from the explicit formula for the quadratic form (3.41).

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

Example 3.25. Consider the quadratic form

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

depending upon three variables. The corresponding coefficient matrix is

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

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. 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

$$q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x} \geq 0 \qquad \text{for all} \qquad \mathbf{x} \in \mathbb{R}^n. \tag{3.43}$$

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 any nonzero vector $\mathbf{z} \in \ker K$ that lies in the 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 $\ker K = \{\mathbf{0}\}$. As a consequence of Proposition 2.39, we deduce that all positive definite matrices are invertible. (The converse, however, is *not* valid.)

Theorem 3.26. *If K is positive definite, then K is nonsingular.*

Example 3.27. 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, since $q(c, c) = 0$.

Example 3.28. 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

$$q(\mathbf{x}) = a x_1^2 + 2b x_1 x_2 + c x_2^2 > 0 \tag{3.44}$$

for all $\mathbf{x} \neq \mathbf{0}$. Analytic geometry tells us that this is the case if and only if

$$a > 0, \qquad\qquad a c - b^2 > 0, \tag{3.45}$$

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

Furthermore, 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, there exist one or more points $\mathbf{x}_+$ where $q(\mathbf{x}_+) > 0$ and one or more points $\mathbf{x}_-$ where $q(\mathbf{x}_-) < 0$. Details can be found in the exercises.

Gram Matrices

Symmetric matrices whose entries are given by inner products of elements of an inner product space play an important role. They are named after the nineteenth century Danish mathematician Jorgen Gram — not the metric mass unit!

Definition 3.29. Let V be an inner product space, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. The associated *Gram matrix*

$$
K = \begin{pmatrix}
\langle \mathbf{v}_1 ; \mathbf{v}_1 \rangle & \langle \mathbf{v}_1 ; \mathbf{v}_2 \rangle & \cdots & \langle \mathbf{v}_1 ; \mathbf{v}_n \rangle \\
\langle \mathbf{v}_2 ; \mathbf{v}_1 \rangle & \langle \mathbf{v}_2 ; \mathbf{v}_2 \rangle & \cdots & \langle \mathbf{v}_2 ; \mathbf{v}_n \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle \mathbf{v}_n ; \mathbf{v}_1 \rangle & \langle \mathbf{v}_n ; \mathbf{v}_2 \rangle & \cdots & \langle \mathbf{v}_n ; \mathbf{v}_n \rangle
\end{pmatrix} .
\tag{3.46}
$$

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

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

$$
k_{ij} = \langle \mathbf{v}_i ; \mathbf{v}_j \rangle = \langle \mathbf{v}_j ; \mathbf{v}_i \rangle = k_{ji}, \qquad \text{and hence} \qquad K^T = K.
\tag{3.47}
$$

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

Theorem 3.30. *All Gram matrices are positive semi-definite. The Gram matrix* (3.46) *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 (3.47) for the matrix entries, we find

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

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

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

where $\mathbf{v} = x_1 \mathbf{v}_1 + \cdots + x_n \mathbf{v}_n$ 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 $\mathbf{v} = \mathbf{0}$ if and only if $x_1 = \cdots = x_n = 0$, and hence $q(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$. Thus, in this case, $q(\mathbf{x})$ and K are positive definite. $\hfill Q.E.D.$

Example 3.31. 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 the associated quadratic form

$$q(x_1, x_2) = 6 x_1^2 - 6 x_1 x_2 + 45 x_2^2$$

is strictly positive for all $(x_1, x_2) \neq \mathbf{0}$. Indeed, this can be checked directly using the criteria in (3.45).

On the other hand, for the weighted inner product

$$\langle \mathbf{x} ; \mathbf{y} \rangle = 3 x_1 y_1 + 2 x_2 y_2 + 5 x_3 y_3, \tag{3.48}$$

the corresponding Gram matrix is

$$\widetilde{K} = \begin{pmatrix} \langle \mathbf{v}_1 ; \mathbf{v}_1 \rangle & \langle \mathbf{v}_1 ; \mathbf{v}_2 \rangle \\ \langle \mathbf{v}_2 ; \mathbf{v}_1 \rangle & \langle \mathbf{v}_2 ; \mathbf{v}_2 \rangle \end{pmatrix} = \begin{pmatrix} 16 & -21 \\ -21 & 207 \end{pmatrix}. \tag{3.49}$$

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 ith row of the transpose A^T with the jth column of A. In other words, the Gram matrix can be evaluated as a matrix product:

$$K = A^T A. \tag{3.50}$$

For the preceding Example 3.31,

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

Theorem 3.30 implies that the Gram matrix (3.50) is positive definite if and only if the columns of A are linearly independent vectors. This implies the following result.

Proposition 3.32. *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 3.23, every inner product on $\mathbb{R}^m$ has the form

$$\langle\, \mathbf{v}\,;\mathbf{w}\,\rangle = \mathbf{v}^T C\,\mathbf{w} \qquad \text{for} \qquad \mathbf{v}, \mathbf{w} \in \mathbb{R}^m, \tag{3.51}$$

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 ith row of A^T by the jth column of the product matrix $C\,A$. Therefore, the Gram matrix based on the alternative inner product (3.51) is given by

$$K = A^T C\,A. \tag{3.52}$$

Theorem 3.30 immediately implies that K is positive definite — provided A has rank n.

Theorem 3.33. *Suppose A is an $m \times n$ matrix with linearly independent columns. Suppose $C > 0$ is any positive definite $m \times m$ matrix. Then the matrix $K = A^T C\,A$ is a positive definite $n \times n$ matrix.*

The Gram matrices constructed in (3.52) arise in a wide variety of applications, including least squares approximation theory, cf. Chapter 4, and mechanical and electrical systems, cf. Chapters 6 and 9. 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 3.34. Returning to the situation of Example 3.31, the weighted inner product (3.48) 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 (3.52) based on the vectors $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 6 \end{pmatrix}$ 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 (3.49).

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 3.35. Consider vector space $C^0[0,1]$ consisting of continuous functions on the interval $0 \leq x \leq 1$, equipped with the L^2 inner product $\langle\, f\,;g\,\rangle = \displaystyle\int_0^1 f(x)\,g(x)\,dx$. 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 3.30 implies that this particular matrix is positive definite.

The alert reader may recognize this particular Gram matrix as the 3×3 *Hilbert matrix* that we encountered in (1.67). More generally, the Gram matrix corresponding to the monomials $1, x, x^2, \ldots, x^n$ has entries

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

Therefore, the monomial Gram matrix is the $(n+1) \times (n+1)$ Hilbert matrix (1.67): $K = H_{n+1}$. As a consequence of Theorems 3.26 and 3.33, we have proved the following non-trivial result.

Proposition 3.36. *The $n \times n$ Hilbert matrix H_n is positive definite. In particular, H_n is a nonsingular matrix.*

Example 3.37. Let us construct the Gram matrix corresponding to the functions $1, \cos x, \sin x$ with respect to the inner product $\langle\, f\,;g\,\rangle = \displaystyle\int_{-\pi}^{\pi} f(x)\,g(x)\,dx$ on the interval

$[-\pi, \pi]$. We compute the inner products

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

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

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

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.

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 3.24.

You may recall the importance of the method known as "completing the square", first in the derivation of the quadratic formula for the solution to

$$q(x) = a\,x^2 + 2\,b\,x + c = 0, \tag{3.53}$$

and, later, in facilitating the integration of various types of rational and algebraic functions. The idea is to combine the first two terms in (3.53) as a perfect square, and so rewrite the quadratic function in the form

$$q(x) = a\left(x + \frac{b}{a}\right)^2 + \frac{ac - b^2}{a} = 0. \tag{3.54}$$

As a consequence,

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

and the well-known quadratic formula

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

follows by taking the square root of both sides and then solving for x. The intermediate step (3.54), 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

$$q(x_1, x_2) = a\,x_1^2 + 2\,b\,x_1\,x_2 + c\,x_2^2. \tag{3.55}$$

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

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

(3.56)

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

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

(3.57)

It is not hard to see that the final expression in (3.56) is positive definite, as a function of y_1, y_2, if and only if both coefficients are positive:

$$a > 0, \qquad \frac{ac - b^2}{a} > 0.$$

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 (3.45) are necessary and sufficient for the quadratic form (3.55) 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 (3.56) in matrix form. The original quadratic form (3.55) is

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

(3.58)

Similarly, the right hand side of (3.56) can be written as

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

(3.59)

Anticipating the final result, the equations (3.57) connecting $\mathbf{x}$ and $\mathbf{y}$ can themselves be written in matrix form as

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

Substituting into (3.59), we find

$$\mathbf{y}^T D \mathbf{y} = (L^T \mathbf{x})^T D (L^T \mathbf{x}) = \mathbf{x}^T L D L^T \mathbf{x} = \mathbf{x}^T K \mathbf{x}, \qquad \text{where} \qquad K = L D L^T \quad (3.60)$$

is the *same* factorization (1.56) of the coefficient matrix, obtained earlier 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 1.32 says that the regular symmetric matrices are precisely those that admit an $L D L^T$ factorization. The identity (3.60) 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:

$$q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x} = \mathbf{y}^T D \mathbf{y} = d_1 y_1^2 + \cdots + d_n y_n^2, \qquad \text{where} \qquad \mathbf{y} = L^T \mathbf{x}. \qquad (3.61)$$

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 3.38. *A symmetric matrix K is positive definite if and only if it is regular and has all positive pivots.*

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

Example 3.39. 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 $1, 2, 6$ in D — are all positive, Theorem 3.38 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:

$$q(\mathbf{x}) = x_1^2 + 4 x_1 x_2 - 2 x_1 x_3 + 6 x_2^2 + 9 x_3^2 = y_1^2 + 2 y_2^2 + 6 y_3^2, \qquad (3.62)$$

where $y_1 = x_1 + 2 x_2 - x_3$, $y_2 = x_2 + x_3$, $y_3 = x_3$, 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 3.40. Lets test whether the matrix $K = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 8 \end{pmatrix}$ 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 doesn't help; all it takes is one non-positive pivot to disqualify a matrix from being positive definite.) This means that the associated quadratic form $q(\mathbf{x}) = x_1^2 + 4 x_1 x_2 + 6 x_1 x_3 + 3 x_2^2 + 8 x_2 x_3 + 8 x_3^2$ 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 (3.62), the terms involving x_1 are

$$x_1^2 + 4\,x_1\,x_2 - 2\,x_1\,x_3$$

which we write as

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

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 only involves 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 (3.62).

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

$$
\begin{aligned}
q(\mathbf{x}) &= k_{11}\,x_1^2 + 2\,k_{12}\,x_1\,x_2 + \;\cdots\; + 2\,k_{1n}\,x_1\,x_n + k_{22}\,x_2^2 + \;\cdots\; + k_{nn}\,x_n^2 \\
&= k_{11}\left(x_1 + \frac{k_{12}}{k_{11}}\,x_2 + \;\cdots\; + \frac{k_{1n}}{k_{11}}\,x_n\right)^2 + \widetilde{q}(x_2, \ldots, x_n) \\
&= k_{11}\,(x_1 + l_{21}\,x_2 + \;\cdots\; + l_{n1}\,x_n)^2 + \widetilde{q}(x_2, \ldots, x_n),
\end{aligned}
\tag{3.63}
$$

where

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

are precisely the multiples appearing in the matrix L obtained from Gaussian Elimination applied 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

$$\widetilde{k}_{ij} = \widetilde{k}_{ji} = k_{ij} - l_{j1}\,k_{1i}, \qquad \text{for} \qquad i \geq j.$$

Thus, the entries of $\widetilde{K}$ that lie on or below the diagonal are exactly the *same* as the entries appearing on or below the diagonal of K after the the first phase of the elimination process.

In particular, the second pivot of K is the diagonal entry $\widetilde{k}_{22}$. Continuing in this fashion, the steps involved in completing the square essentially reproduce the steps of Gaussian elimination, with the pivots appearing in the appropriate diagonal positions.

With this in hand, we can now complete the proof of Theorem 3.38. 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 (3.63). 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 (3.63) 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 $k_{22} > 0$. We then continue the reduction procedure outlined in the preceding paragraph; if a non-positive pivot appears an any stage, the original quadratic form and matrix cannot be positive definite, while having all positive pivots will ensure positive definiteness, thereby proving Theorem 3.38.

The Cholesky Factorization

The identity (3.60) shows us how to write any regular quadratic form $q(\mathbf{x})$ as a linear combination of squares. One can push this result slightly further in the positive definite case. Since each pivot $d_i > 0$, we can write the diagonal quadratic form (3.61) 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 (1.52), we deduce the *Cholesky factorization*

$$K = L D L^T = L S S^T L^T = M M^T, \qquad \text{where} \qquad M = L S \qquad (3.64)$$

of a positive definite matrix. Note that M is a lower triangular matrix with all positive entries, namely the square roots of the pivots $m_{ii} = \sqrt{d_i}$ on its diagonal. Applying the Cholesky factorization to the corresponding quadratic form produces

$$q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x} = \mathbf{x}^T M M^T \mathbf{x} = \mathbf{z}^T \mathbf{z} = \| \mathbf{z} \|^2, \qquad \text{where} \qquad \mathbf{z} = M^T \mathbf{x}. \qquad (3.65)$$

One can interpret (3.65) 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 3.41. For the matrix $K = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 6 & 0 \\ -1 & 0 & 9 \end{pmatrix}$ considered in Example 3.39, the Cholesky formula (3.64) 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 $\mathbf{z} = M^T\mathbf{x}$, or, explicitly, $z_1 = x_1 + 2\,x_2 - x_3$, $z_2 = \sqrt{2}\,x_2 + \sqrt{2}\,x_3$, $z_3 = \sqrt{6}\,x_3$..

3.6. Complex Vector Spaces.

Although physical applications ultimately require real answers, complex numbers and complex vector spaces assume 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 formulae. Complex variable methods are ubiquitous in electrical engineering, Fourier analysis, potential theory, fluid mechanics, electromagnetism, and so on. In quantum mechanics, the basic physical quantities are complex-valued wave functions. Moreover, the Schrödinger equation, which governs quantum dynamics, is an inherently complex partial differential equation.

In this section, we survey the basic facts about complex numbers and complex vector spaces. Most of the constructions are entirely analogous to 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 counterpart. Complex analysis (integration and differentiation of complex functions) and its applications to fluid flows, potential theory, waves and other areas of mathematics, physics and engineering, will be the subject of Chapter 16.

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 $x = \mathrm{Re}\,z$ the *real part* of z and $y = \mathrm{Im}\,z$ 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, $\mathrm{Im}\,z = 0$, 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

$$\begin{aligned} (x + \mathrm{i}\,y) + (u + \mathrm{i}\,v) &= (x + u) + \mathrm{i}\,(y + v), \\ (x + \mathrm{i}\,y) \cdot (u + \mathrm{i}\,v) &= (x\,u - y\,v) + \mathrm{i}\,(x\,v + y\,u). \end{aligned} \tag{3.66}$$

[†] Electrical engineers prefer to use j to indicate the imaginary unit.

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

We can identity a complex number $x + iy$ 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 (3.66) corresponds to vector addition, but complex multiplication does *not* have a readily identifiable vector counterpart.

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

Definition 3.42. The *complex conjugate* of $z = x + iy$ is $\overline{z} = x - iy$, whereby $\operatorname{Re} \overline{z} = \operatorname{Re} z$, while $\operatorname{Im} \overline{z} = -\operatorname{Im} z$.

Geometrically, the operation of complex conjugation coincides with reflection of the corresponding vector through the real axis, as illustrated in Figure 3.6. In particular $\overline{z} = z$ if and only if z is real. Note that

$$\operatorname{Re} z = \frac{z + \overline{z}}{2}, \qquad \operatorname{Im} z = \frac{z - \overline{z}}{2i}. \tag{3.67}$$

Complex conjugation is compatible with complex arithmetic:

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

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

$$z\,\overline{z} = (x + iy)(x - iy) = x^2 + y^2 \tag{3.68}$$

is real and non-negative. Its square root is known as the *modulus* of the complex number $z = x + iy$, and written

$$|z| = \sqrt{x^2 + y^2}. \tag{3.69}$$

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 xy–plane, which implies the validity of the triangle inequality

$$|z + w| \leq |z| + |w|. \tag{3.70}$$

Equation (3.68) can be rewritten in terms of the modulus as

$$z\,\overline{z} = |z|^2. \tag{3.71}$$

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

$$\frac{1}{z} = \frac{\overline{z}}{|z|^2}, \qquad z \neq 0, \qquad \text{or, equivalently,} \qquad \frac{1}{x + iy} = \frac{x - iy}{x^2 + y^2}. \tag{3.72}$$

The general formula for complex division

$$\frac{w}{z} = \frac{w\,\overline{z}}{|z|^2} \qquad \text{or} \qquad \frac{u + iv}{x + iy} = \frac{(xu + yv) + i(xv - yu)}{x^2 + y^2}, \tag{3.73}$$
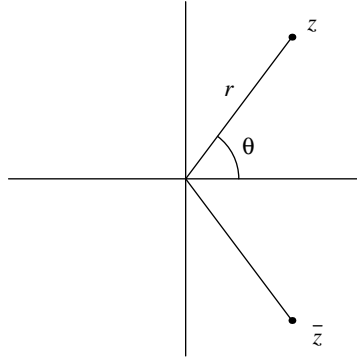
Figure 3.6. Complex Numbers.

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

$$x = r\cos\theta, \qquad y = r\sin\theta \qquad \text{or} \qquad z = r(\cos\theta + i\,\sin\theta). \tag{3.74}$$

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

$$\theta = \text{ph}\,z. \tag{3.75}$$

As such, the phase is only defined up to an integer multiple of 2π. The more common term for the angle is the *argument*, written arg $z = \text{ph}\,z$. However, we prefer to use "phase" throughout this text, in part to avoid confusion with the argument of a function. We note that the modulus and phase of a product of complex numbers can be readily computed:

$$|z\,w| = |z|\,|w|, \qquad \text{ph}\,(z\,w) = \text{ph}\,z + \text{ph}\,w. \tag{3.76}$$

Complex conjugation preserves the modulus, but negates the phase:

$$|\overline{z}| = |z|, \qquad \text{ph}\,\overline{z} = -\,\text{ph}\,z. \tag{3.77}$$

One of the most important equations in all of mathematics is *Euler's formula*

$$e^{i\theta} = \cos\theta + i\,\sin\theta, \tag{3.78}$$

relating the complex exponential with the real sine and cosine functions. This fundamental identity has a variety of mathematical justifications; see Exercise ■ for one that is based on comparing power series. Euler's formula (3.78) can be used to compactly rewrite the polar form (3.74) of a complex number as

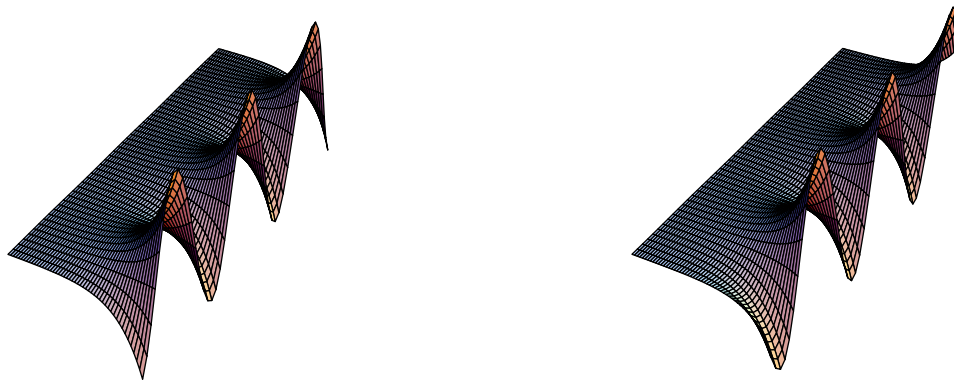$$z = r\,e^{i\theta} \qquad \text{where} \qquad r = |z|, \qquad \theta = \text{ph}\,z. \tag{3.79}$$

Figure 3.7. Real and Imaginary Parts of e^z.

The complex conjugate 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:

$$\cos\theta = \frac{e^{\mathrm{i}\,\theta} + e^{-\mathrm{i}\,\theta}}{2}, \qquad \sin\theta = \frac{e^{\mathrm{i}\,\theta} - e^{-\mathrm{i}\,\theta}}{2\,\mathrm{i}}. \qquad (3.80)$$

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

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

$$e^z = e^{x+\mathrm{i}\,y} = e^x\,e^{\mathrm{i}\,y} = e^x\cos y + \mathrm{i}\,e^x\sin y. \qquad (3.81)$$

Graphs of the real and imaginary parts of the complex exponential function appear in Figure 3.7. Note that $e^{2\pi\mathrm{i}} = 1$, and hence the exponential function is periodic

$$e^{z+2\pi\mathrm{i}} = e^z \qquad (3.82)$$

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

Complex Vector Spaces and Inner Products

A *complex vector space* is defined in exactly the same manner as its real cousin, cf. 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: $z_1, \ldots, z_n \in \mathbb{C}$. 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 $\mathbf{x} = \operatorname{Re}\mathbf{z}, \mathbf{y} = \operatorname{Im}\mathbf{z} \in \mathbb{R}^n$. 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

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

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 most important applications of inner products and norms are based on the associated inequalities: Cauchy–Schwarz and triangle. But there is no natural ordering of the complex numbers, and so one *cannot* make any sense of a complex inequality like $z < w$. Inequalities only make sense 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, this would give some nonzero complex vectors, e.g., $(\,1 \quad \mathrm{i}\,)^T$, a zero "norm", violating positivity.

The correct definition is modeled on the formula

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

that 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 is to 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 find

$$z \cdot w = z\,\overline{w} = (x + \mathrm{i}\,y)\,(u - \mathrm{i}\,v) = (x\,u + y\,v) + \mathrm{i}\,(y\,u - x\,v). \tag{3.83}$$

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. (cross2■).

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

$$\mathbf{z} \cdot \mathbf{w} = \mathbf{z}^T\,\overline{\mathbf{w}} = z_1\,\overline{w}_1 + z_2\,\overline{w}_2 + \cdots + z_n\,\overline{w}_n, \quad \text{for} \quad \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}. \tag{3.84}$$

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,

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

Therefore, the Hermitian dot product is *not* symmetric. Reversing the order of the vectors results in complex conjugation of the dot product:

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

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

$$0 \le \|\,\mathbf{z}\,\| = \sqrt{\mathbf{z} \cdot \mathbf{z}} = \sqrt{\mathbf{z}^T\,\overline{\mathbf{z}}} = \sqrt{\,|\,z_1\,|^2 + \;\cdots\; + |\,z_n\,|^2}\,, \tag{3.85}$$

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:

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

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:

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

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 *sesqui-linearity*.

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

Definition 3.43. 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*) *Sesqui-linearity*:

$$\begin{aligned} \langle\,c\,\mathbf{u} + d\,\mathbf{v}\,;\mathbf{w}\,\rangle &= c\,\langle\,\mathbf{u}\,;\mathbf{w}\,\rangle + d\,\langle\,\mathbf{v}\,;\mathbf{w}\,\rangle, \\ \langle\,\mathbf{u}\,;c\,\mathbf{v} + d\,\mathbf{w}\,\rangle &= \overline{c}\,\langle\,\mathbf{u}\,;\mathbf{v}\,\rangle + \overline{d}\,\langle\,\mathbf{u}\,;\mathbf{w}\,\rangle. \end{aligned} \tag{3.86}$$

(*ii*) *Conjugate Symmetry*:

$$\langle\,\mathbf{v}\,;\mathbf{w}\,\rangle = \overline{\langle\,\mathbf{w}\,;\mathbf{v}\,\rangle}. \tag{3.87}$$

(*iii*) *Positivity*:

$$\|\,\mathbf{v}\,\|^2 = \langle\,\mathbf{v}\,;\mathbf{v}\,\rangle \ge 0, \quad \text{and} \quad \langle\,\mathbf{v}\,;\mathbf{v}\,\rangle = 0 \quad \text{if and only if} \quad \mathbf{v} = \mathbf{0}. \tag{3.88}$$

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 reversing the order of the two arguments. But, once this initial complication has been properly dealt with, the further properties of the inner product carry over directly from the real domain.

Theorem 3.44. *The Cauchy–Schwarz inequality,*

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

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

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

are both valid on any complex inner product space.

The proof of this result is practically the same as in the real case, and the details are left to the reader.

Example 3.45. The vectors $\mathbf{v} = (1 + i , 2 i , -3)^T$, $\mathbf{w} = (2 - i , 1 , 2 + 2 i)^T$, satisfy

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

Thus, the Cauchy–Schwarz inequality reads

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

Similarly, the triangle inequality tells us that

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

Example 3.46. Let $C^0[-\pi, \pi]$ denote the complex vector space consisting of all complex valued continuous functions $f(x) = u(x) + i v(x)$ depending upon the *real* variable $-\pi \leq x \leq \pi$. The Hermitian L^2 inner product on $C^0[-\pi, \pi]$ is defined as

$$\langle f ; g \rangle = \int_{-\pi}^{\pi} f(x) \, \overline{g(x)} \, dx \, , \tag{3.89}$$

with corresponding norm

$$\| f \| \; = \; \sqrt{ \int_{-\pi}^{\pi} | f(x) |^2 \, dx } \; = \; \sqrt{ \int_{-\pi}^{\pi} \left[u(x)^2 + v(x)^2 \right] dx } \, . \tag{3.90}$$

The reader should verify that (3.89) satisfies the basic Hermitian inner product axioms.

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

$$\langle e^{i k x} ; e^{i l x} \rangle = \int_{-\pi}^{\pi} e^{i k x} e^{- i l x} \, dx = \int_{-\pi}^{\pi} e^{i (k-l) x} \, dx = \begin{cases} 2 \pi, & k = l, \\[2mm] \left. \dfrac{e^{i (k-l) x}}{i (k - l)} \right|_{x = -\pi}^{\pi} = 0, & k \neq l. \end{cases}$$

We conclude that when $k \neq l$, the complex exponentials e^{ikx} and e^{ilx} are orthogonal, since their inner product is zero. This example will be of fundamental significance in the complex formulation of Fourier analysis.