# 6 Sturm-Liouville boundary value problems

## 6.1 Introduction

In Section 5.5 we described the remarkable result that an arbitrary piecewise differentiable function $f(x)$ could be expanded in either a pure sine series of the form

$$f(x) = \sum_{n=1}^{\infty} b_n \sin\frac{n\pi x}{l} \tag{1}$$

or a pure cosine series of the form

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi x}{l} \tag{2}$$

on the interval $0 < x < l$. We were led to the trigonometric functions appearing in the series (1) and (2) by considering the 2 point boundary value problems

$$y'' + \lambda y = 0, \qquad y(0) = 0, y(l) = 0, \tag{3}$$

and

$$y'' + \lambda y = 0, \qquad y'(0) = 0, y'(l) = 0. \tag{4}$$

Recall that Equations (3) and (4) have nontrivial solutions

$$y_n(x) = c\sin\frac{n\pi x}{l} \quad \text{and} \quad y_n(x) = c\cos\frac{n\pi x}{l},$$

respectively, only if $\lambda = \lambda_n = \dfrac{n^2\pi^2}{l^2}$. These special values of $\lambda$ were called eigenvalues, and the corresponding solutions were called eigenfunctions.

There was another instance in our study of differential equations where something interesting happened for special values of a parameter $\lambda$. To wit, in Section 2.8 we studied, either in the text or in the exercises, the four differential equations

$$y'' - 2xy + \lambda y = 0, \tag{5}$$

$$(1 - x^2)y'' - 2xy' + \lambda(\lambda + 1)y = 0, \tag{6}$$

$$(1 - x^2)y'' - xy' + \lambda^2 y = 0, \tag{7}$$

and

$$xy'' + (1 - x)y' + \lambda y = 0. \tag{8}$$

Equations (5)–(8) are the famous Hermite, Legendre, Tchebycheff, and Laguerre differential equations, respectively. The Legendre, Tchebycheff, and Laguerre equations each have a polynomial solution of degree $n$ if $\lambda = n$, while the Hermite equation has a polynomial solution of degree $n$ if $\lambda = 2n$. These polynomials, when properly normalized, i.e., when multiplied by a suitable constant, are known as the Hermite, Legendre, Tchebycheff, and Laguerre polynomials. It turns out, remarkably, that any piecewise differentiable function $f(x)$ can also be expanded in a series of Hermite, Legendre, Tchebycheff, and Laguerre polynomials, on an appropriate interval.

There is a very pretty theory that ties together not only the trigonometric functions and the Hermite, Legendre, Tchebycheff, and Laguerre polynomials, but also many of the other famous functions of mathematical physics, such as the various Bessel functions. This theory is commonly called Sturm-Liouville Theory; it has its roots, essentially, in an area of linear algebra known as inner product spaces, and it is to this area that we now turn our attention.

## 6.2 Inner product spaces

Up to this point, our study of linear algebra in general and linear vector spaces in particular was algebraic in nature. We were able to add two vectors together and multiply a vector by a constant. By means of these operations we can define the geometric concepts of dimension, line, plane, and even parallelism of lines. Recall that the dimension of a space $V$ is the number of elements in a basis, i.e., the fewest number of linearly independent vectors that span $V$. Once we have the concept of dimension, we can define a line in $V$ as a subspace of dimension 1, a plane as a subspace of dimension 2, etc. Finally, two vectors are parallel if one is a constant multiple of the other.

Many important geometric concepts of so-called Euclidean geometry, however, still cannot be defined for arbitrary linear vector spaces. Specifically, we have no way of formulating, as yet, the definition of *length* of a vector and the *angle* between two vectors. To accomplish this, we need to superimpose on $V$ some additional structure. This structure is known as an *inner product.*

Our definition of an inner product is modelled after the traditional *dot product* of two vectors in $R^2$ and $R^3$, which we studied in calculus. Recall that if $\mathbf{x}$ and $\mathbf{y}$ are vectors in $R^2$ or $R^3$, then

$$\mathbf{x}\cdot\mathbf{y} = x_1y_1 + x_2y_2 \quad \text{in } R^2 \tag{1}$$

and

$$\mathbf{x}\cdot\mathbf{y} = x_1y_1 + x_2y_2 + x_3y_3 \quad \text{in } R^3 \tag{2}$$

where $x_1, x_2, \ldots$, and $y_1, y_2, \ldots$ are the components of $\mathbf{x}$ and $\mathbf{y}$, respectively. Recall too, the famous identity, proven in most calculus courses, that

$$\mathbf{x}\cdot\mathbf{y} = |\mathbf{x}||\mathbf{y}|\cos\theta \tag{3}$$

where $\theta$ is the angle between $\mathbf{x}$ and $\mathbf{y}$ and

$$|\mathbf{x}| = (x_1^2 + x_2^2)^{1/2} \quad \text{or} \quad (x_1^2 + x_2^2 + x_3^2)^{1/2}$$

in $R^2$ and $R^3$, respectively, is the Euclidean length of $\mathbf{x}$. Rewriting Equation (3) in the form

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}||\mathbf{y}|} \tag{4}$$

enables us to compute the angle $\theta$ between two vectors. Finally, if we observe that

$$|\mathbf{x}| = (\mathbf{x}\cdot\mathbf{x})^{1/2} \tag{5}$$

then both the length of a vector and the angle between two vectors can be computed from the dot product alone.

We are now ready to generalize this concept of dot product. This generalization will be called an *inner product.* We will first define a *real* inner product, and then toward the end of this section, a *complex* inner product.

**Definition.** Let $V$ be a real vector space. A *real inner product* on $V$ is a real-valued function that associates with each pair of vectors $\mathbf{x}$, $\mathbf{y}$ a real number, denoted by $\langle\mathbf{x},\mathbf{y}\rangle$, that satisfies the following properties:

(i) $\langle\mathbf{x},\mathbf{y}\rangle = \langle\mathbf{y},\mathbf{x}\rangle$ for all $\mathbf{x}$ and $\mathbf{y}$ in $V$,
(ii) $\langle k\mathbf{x},\mathbf{y}\rangle = k\langle\mathbf{x},\mathbf{y}\rangle$ for all scalars $k$ and vectors $\mathbf{x}$, $\mathbf{y}$,
(iii) $\langle\mathbf{x}+\mathbf{y},\mathbf{z}\rangle = \langle\mathbf{x},\mathbf{z}\rangle + \langle\mathbf{y},\mathbf{z}\rangle$ for all $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ in $V$,
(iv) $\langle\mathbf{x},\mathbf{x}\rangle \geqslant 0$ and $\langle\mathbf{x},\mathbf{x}\rangle = 0$ only if $\mathbf{x} = \mathbf{0}$.

It is customary to refer to the vector space $V$, together with some inner product $\langle\ ,\ \rangle$ as a real inner product space, and a finite-dimensional real inner product space is often called a Euclidean space.

**Remark.** Observe from (i) that properties (ii) and (iii) hold equally well on opposite sides of the inner product bracket, i.e.,

$$\langle \mathbf{x}, k\mathbf{y}\rangle = k\langle \mathbf{x},\mathbf{y}\rangle \quad \text{and} \quad \langle \mathbf{x},\mathbf{y}+\mathbf{z}\rangle = \langle \mathbf{x},\mathbf{y}\rangle + \langle \mathbf{x},\mathbf{z}\rangle.$$

The classic example of an inner product on $R^n$, of course, is the dot product

$$\langle \mathbf{x},\mathbf{y}\rangle = \mathbf{x}\cdot\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

It is easily verified that the dot product satisfies properties (i)–(iv). Here are some additional examples, which have proven extremely useful in applications.

**Example 1.** Let $V$ be the space of all continuous functions on the interval $[a,b]$, and define the inner product of two functions $f(x)$ and $g(x)$ in $V$ as

$$\langle f,g\rangle = \int_a^b f(x)g(x)\,dx. \tag{6}$$

It is easily verified that the definition (6) satisfies properties (i)–(iv).

**Example 2.** Let $V$ be as in Example 1, and define

$$\langle f,g\rangle = \int_a^b r(x)f(x)g(x)\,dx \tag{7}$$

where $r(x)$ is positive on the open interval $(a,b)$. Again, it is trivial to verify that the inner product (7) satisfies properties (i)–(iv).

**Example 3.** Let $V = R^3$ and set

$$\langle \mathbf{x},\mathbf{y}\rangle = x_1y_1 + 2x_2y_2 + 2x_3y_3 - x_1y_2 - x_2y_1 - x_2y_3 - x_3y_2.$$

Properties (ii) and (iii) are trivial to check, and property (i) is simple to verify. To check property (iv), we write

$$\begin{aligned}\langle \mathbf{x},\mathbf{x}\rangle &= x_1^2 + 2x_2^2 + 2x_3^2 - x_1x_2 - x_2x_1 - x_2x_3 - x_3x_2\\ &= x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3\\ &= x_1^2 - 2x_1x_2 + x_2^2 + x_2^2 - 2x_2x_3 + x_3^2 + x_3^2\\ &= (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2,\end{aligned}$$

and it is now clear that $\langle \mathbf{x},\mathbf{x}\rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in an arbitrary inner product space $V$, and suppose, following Equation (5), that we define the lengths of $\mathbf{x}$ and $\mathbf{y}$, denoted by $\|\mathbf{x}\|$ and $\|\mathbf{y}\|$, respectively, as

$$\|\mathbf{x}\| = \langle \mathbf{x},\mathbf{x}\rangle^{1/2}, \qquad \|\mathbf{y}\| = \langle \mathbf{y},\mathbf{y}\rangle^{1/2}, \tag{8}$$

where the double vertical lines are used to denote length. It is certainly very plausible to try to mimic Equation (4) and define the angle $\theta$ between $\mathbf{x}$ and y via the equation

$$\cos\theta = \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|\,\|\mathbf{y}\|} = \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{x},\mathbf{x}\rangle^{1/2}\langle \mathbf{y},\mathbf{y}\rangle^{1/2}}. \tag{9}$$

Clearly, the right-hand side of Equation (9) must be $\leqslant 1$ in order for this equation to have any meaning at all. A second and more subtle requirement is that the quantity $\|\mathbf{x}\|$, defined as the *length* of $\mathbf{x}$, really is a length; that is, it satisfies the geometric properties usually associated with length. Let us therefore take one last digression and discuss the concept of length of a vector, or as it is traditionally called in linear algebra, the *norm* of a vector.

**Definition.** A *norm* on a real or complex vector space $V$ is a real-valued function, usually denoted by $\|\ \ \|$, which satisfies

(i) $\|\mathbf{x}\| \geqslant 0$ for all $\mathbf{x}$ in $V$ and $\|\mathbf{x}\| = 0$ only if $\mathbf{x} = \mathbf{0}$,
(ii) $\|k\mathbf{x}\| = |k|\,\|\mathbf{x}\|$ for all scalars $k$,
(iii) $\|\mathbf{x}+\mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$.

Properties (i) and (ii) are fairly obvious properties of length, while property (iii) is simply the triangle law, which states that the length of one side of a triangle is always less than (or equal to) the sum of the lengths of the other two sides (see Figure 1).

The classical example of a norm is ordinary length in $R^n$; i.e., if

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

then

$$\|\mathbf{x}\| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$$

obviously satisfies (i)–(iii). Here are some additional examples, which have proven quite useful in applications.

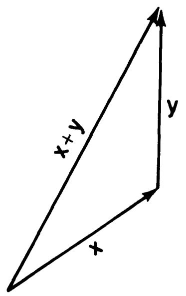


Figure 1

**Example 4.** Let $V = R^n$ and define

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|.$$

It is extremely easy to verify that $\|\ \|_1$ is a norm on $R^n$.

**Example 5.** Let $V = R^n$ and define

$$\|\mathbf{x}\|_\infty = \max[|x_1|, |x_2|, \ldots, |x_n|].$$

Again, it is extremely simple to verify that $\|\ \|_\infty$ defines a norm on $R^n$.

**Example 6.** Let $V = R^n$ and define

$$\|\mathbf{x}\|_p = [|x_1|^p + |x_2|^p + \cdots + |x_n|^p]^{1/p}. \tag{10}$$

It is extremely simple to verify that $\|\ \|_p$, referred to as the $p$ norm, satisfies properties (i) and (ii). Property (iii) is also true but quite difficult to verify, hence we will not do so here. We wish to point out (see Exercise 6), however, that

$$\|\mathbf{x}\|_\infty = \lim_{p\to\infty} \|\mathbf{x}\|_p,$$

and this is the motivation for the strange notation $\|\ \|_\infty$.

**Example 7.** Let $V$ be the space of all continuous functions $f(x)$ on the interval $[0, 1]$, and define

$$\|f\| = \int_0^1 |f(x)|\,dx. \tag{11}$$

It is easily verified that Equation (11) defines a norm on $V$.

Let us return now to the questions raised via the definitions (8) and (9). As mentioned previously, we must show that the right-hand side of Equation (9) is always between $-1$ and $+1$ in order for this equation to make sense. But this is equivalent to the inequality

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} \langle \mathbf{y}, \mathbf{y} \rangle^{1/2}$$

or, equivalently,

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \|\mathbf{x}\|\,\|\mathbf{y}\|. \tag{12}$$

Equation (12) is the precise statement of a very famous theorem, known as Schwarz's Inequality, which we now prove.

**Theorem 1** (Schwarz's Inequality). *Let $V$ be a real inner product space with inner product $\langle\ ,\ \rangle$. Then*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \|\mathbf{x}\|\,\|\mathbf{y}\| \tag{13}$$

*for all vectors* **x**, **y**, *where* $\|\mathbf{x}\|$ *is defined by Equation* (8). *Furthermore, equality holds in Equation* (13) *only if the vectors* **x** *and* **y** *are linearly dependent, i.e.,* $\mathbf{y} = k\mathbf{x}$ *for some scalar k.*

PROOF. There are several ways of proving Theorem 1, but we will choose here the proof that generalizes most easily to the complex case. Observe that

$$\left\langle \mathbf{x} - \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\mathbf{y},\, \mathbf{x} - \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\mathbf{y} \right\rangle \geqslant 0 \tag{14}$$

for all nonzero vectors **x** and **y** in $V$. This follows immediately from property (iv) of inner products. Next, using properties (i)–(iii) of inner products, we can rewrite Equation (14) in the form

$$\begin{aligned}\langle \mathbf{x},\mathbf{x}\rangle - \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\langle \mathbf{y},\mathbf{x}\rangle - \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\langle \mathbf{x},\mathbf{y}\rangle + \frac{\langle \mathbf{x},\mathbf{y}\rangle^2}{\langle \mathbf{y},\mathbf{y}\rangle^2}\langle \mathbf{y},\mathbf{y}\rangle \\ &= \langle \mathbf{x},\mathbf{x}\rangle - 2\frac{\langle \mathbf{x},\mathbf{y}\rangle^2}{\langle \mathbf{y},\mathbf{y}\rangle} + \frac{\langle \mathbf{x},\mathbf{y}\rangle^2}{\langle \mathbf{y},\mathbf{y}\rangle} \\ &= \langle \mathbf{x},\mathbf{x}\rangle - \frac{\langle \mathbf{x},\mathbf{y}\rangle^2}{\langle \mathbf{y},\mathbf{y}\rangle} \\ &= \|\mathbf{x}\|^2 - \frac{\langle \mathbf{x},\mathbf{y}\rangle^2}{\|\mathbf{y}\|^2}.\end{aligned}$$

Hence,

$$\frac{\langle \mathbf{x},\mathbf{y}\rangle^2}{\|\mathbf{y}\|^2} \leqslant \|\mathbf{x}\|^2$$

or, equivalently,

$$|\langle \mathbf{x},\mathbf{y}\rangle| \leqslant \|\mathbf{x}\|\,\|\mathbf{y}\|.$$

Finally, again from property (iv), equality holds in Equation (14) only if

$$\mathbf{x} - \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\mathbf{y} = \mathbf{0}$$

or

$$\mathbf{x} = k\mathbf{y}, \qquad k = \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}.$$

**Remark 1.** The proof we just gave assumes that $\mathbf{y} \neq \mathbf{0}$. We leave it to the reader to verify (see Exercise 13) Schwarz's Inequality in the trivial case that $\mathbf{y} = \mathbf{0}$.

**Remark 2.** The vector $(\langle \mathbf{x},\mathbf{y}\rangle/\langle \mathbf{y},\mathbf{y}\rangle)\mathbf{y}$ introduced in Equation (14) was not a lucky guess. This vector is actually the *projection* of the vector **x** onto the vector **y**, as shown in Figure 2.

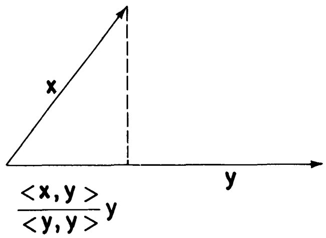


Figure 2

The second question raised was whether

$$\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} \tag{15}$$

satisfies properties (i)–(iii) of norms. Property (i) follows immediately from property (iv) of inner products. To verify property (ii), we compute

$$\|k\mathbf{x}\| = \langle k\mathbf{x}, k\mathbf{x} \rangle^{1/2} = [k^2 \langle \mathbf{x}, \mathbf{x} \rangle]^{1/2} = |k| \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} = |k| \, \|\mathbf{x}\|.$$

Property (iii) is a bit more difficult to verify. Observe first that

$$\begin{aligned}
\|\mathbf{x} + \mathbf{y}\|^2 &= \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle \\
&= \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \\
&= \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \\
&\leqslant \|\mathbf{x}\|^2 + 2\|\mathbf{x}\| \, \|\mathbf{y}\| + \|\mathbf{y}\|^2 \\
&= [\|\mathbf{x}\| + \|\mathbf{y}\|]^2,
\end{aligned}$$

where the inequality above follows directly from Schwarz's Inequality. Taking square roots gives the triangle inequality, and this completes our proof.

**Remark.** Schwarz's Inequality is an extremely powerful tool, as witnessed by the preceding proof. Here is another illustration of its strength. Let $V$ be the space of all continuous functions on the interval $[a, b]$, and define

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

Then, from Schwarz's Inequality,

$$|\langle f, g \rangle| = \left| \int_a^b f(x)g(x)\,dx \right| \leqslant \|f(x)\| \, \|g(x)\|.$$

Hence,

$$\left| \int_a^b f(x)g(x)\,dx \right| \leqslant \left[ \int_a^b f^2(x)\,dx \right]^{1/2} \left[ \int_a^b g^2(x)\,dx \right]^{1/2}. \tag{16}$$

We defy the reader to give a pure calculus proof of the inequality (16).

Once we have an inner product space $V$, we can define two vectors in $V$ to be *orthogonal*, or perpendicular to each other, if their inner product is zero; i.e., $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

**Example 8.** Let $V$ be the space of all continuous functions on the interval $[-\pi, \pi]$, and define an inner product on $V$ via the relation

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

The set of functions

$$f_k(x) = \sin kx, \qquad k = 1, 2, \ldots,$$

are all mutually orthogonal; that is,

$$\langle f_j(x), f_k(x) \rangle = 0, \qquad j \neq k,$$

since

$$\langle f_j(x), f_k(x) \rangle = \int_{-\pi}^{\pi} \sin jx \sin kx\, dx = 0, \qquad j \neq k,$$

following the discussion in Section 5.4.

**Example 9.** Let $V$ be the space of all continuous functions on the interval $[-1, 1]$, and define

$$\langle f, g \rangle = \int_{-1}^{1} \frac{f(x)g(x)}{\sqrt{1-x^2}}\,dx$$

for $f$ and $g$ in $V$. The set of functions

$$f_k(x) = \cos(k \cos^{-1} x), \qquad k = 0, 1, 2, \ldots,$$

are mutually orthogonal with respect to $\langle\ ,\ \rangle$, i.e.,

$$\int_{-1}^{1} \frac{\cos(j \cos^{-1} x)\cos(k \cos^{-1} x)}{\sqrt{1-x^2}}\,dx = 0, \qquad j \neq k. \tag{17}$$

To establish Equation (17) we make the substitution

$$u = \cos^{-1} x \qquad du = \frac{-1}{\sqrt{1-x^2}}\,dx.$$

Then the left-hand side of Equation (17) becomes

$$\begin{aligned} -\int_{\pi}^{0} \cos ju \cos ku\, du &= \int_{0}^{\pi} \cos ju \cos ku\, du \\ &= \frac{1}{2}\int_{0}^{\pi} [\cos(j+k)u + \cos(j-k)u]\,du = 0, \qquad j \neq k. \end{aligned}$$

**Remark.** It can be shown that the functions $f_k(x)$ defined in Example 9 are the Tchebycheff polynomials $T_k(x)$.

We conclude this section by extending the notion and definition of inner product to complex vector spaces. The main problem we have to overcome is the requirement that $\langle \mathbf{x}, \mathbf{x} \rangle$ be positive for all nonzero vectors $\mathbf{x}$. For example, if $V$ is the space of all complex-valued functions

$$h(t) = f(t) + ig(t), \qquad a \leqslant t \leqslant b,$$

then we cannot define

$$\langle h_1, h_2 \rangle = \int_a^b h_1(t) h_2(t)\, dt$$

as in Example 1, since

$$\langle h, h \rangle = \int_a^b [f(t) + ig(t)][f(t) + ig(t)]\, dt$$

is not even real [for $\langle f, g \rangle \neq 0$], let alone positive.

To motivate the proper extension, we return to the familiar dot product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \tag{18}$$

where $\mathbf{x}$ and $\mathbf{y}$ are *real* vectors, i.e., their components are real numbers. We can extend this definition to complex vector spaces by making the simple adjustment

$$\mathbf{x} \cdot \mathbf{y} = x_1 \bar{y}_1 + x_2 \bar{y}_2 + \cdots + x_n \bar{y}_n. \tag{19}$$

Then, following Equation (19),

$$\langle \mathbf{x}, \mathbf{x} \rangle = x_1 \bar{x}_1 + x_2 \bar{x}_2 + \cdots + x_n \bar{x}_n$$

is always real and nonnegative. The only thing we have to be extra careful about now is that

$$\mathbf{y} \cdot \mathbf{x} = y_1 \bar{x}_1 + \cdots + y_n \bar{x}_n = \overline{\langle \mathbf{x}, \mathbf{y} \rangle}$$

instead of $\langle \mathbf{x}, \mathbf{y} \rangle$ and

$$\mathbf{x} \cdot k\mathbf{y} = x_1 \bar{k} \bar{y}_1 + \cdots + x_n \bar{k} \bar{y}_n = \bar{k} \langle \mathbf{x}, \mathbf{y} \rangle$$

instead of $k \langle \mathbf{x}, \mathbf{y} \rangle$. This leads us to the following definition.

**Definition.** Let $V$ be a complex vector space. A *complex*, or Hermitian, *inner product* on $V$ is a complex-valued function that associates with each pair of vectors $\mathbf{x}$, $\mathbf{y}$ a complex number, denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$, that satisfies the following properties:

(i) $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$,
(ii) $\langle k\mathbf{x}, \mathbf{y} \rangle = k \langle \mathbf{x}, \mathbf{y} \rangle$,

(iii) $\langle \mathbf{x}+\mathbf{y}, \mathbf{z}\rangle = \langle \mathbf{x}, \mathbf{z}\rangle + \langle \mathbf{y}, \mathbf{z}\rangle$,
(iv) $\langle \mathbf{x}, \mathbf{x}\rangle \geqslant 0$ and $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ only if $\mathbf{x} = \mathbf{0}$

for all vectors $\mathbf{x}$ and $\mathbf{y}$ and complex numbers $k$.

Property (i) is known as *conjugate symmetry* in contrast to the ordinary symmetry of the real case. It implies that

$$\langle \mathbf{x}, k\mathbf{y}\rangle = \overline{\langle k\mathbf{y}, \mathbf{x}\rangle} = \overline{k\langle \mathbf{y}, \mathbf{x}\rangle} = \bar{k}\overline{\langle \mathbf{y}, \mathbf{x}\rangle} = \bar{k}\langle \mathbf{x}, \mathbf{y}\rangle.$$

Thus, scalars pulled out from the right-hand side of a complex inner product must be conjugated. Finally, we leave it to the reader to verify (see Exercise 14) that

$$\langle \mathbf{x}, \mathbf{y}+\mathbf{z}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{x}, \mathbf{z}\rangle.$$

**Example 10.** Let $V$ be the space of all complex-valued functions $h(t) = f(t) + ig(t)$ on the interval $[0, 1]$, and define

$$\begin{aligned} \langle h_1, h_2\rangle &= \int_0^1 h_1(t)\overline{h_2(t)}\,dt \\ &= \int_0^1 [f_1(t) + ig_1(t)][f_2(t) - ig_2(t)]\,dt \qquad (20) \\ &= \int_0^1 [f_1(t)f_2(t) + g_1(t)g_2(t)]\,dt + i\int_0^1 [f_2(t)g_1(t) - f_1(t)g_2(t)]\,dt \end{aligned}$$

It is easily verified that Equation (20) satisfies properties (i)–(iv) of our definition.

Complex inner product spaces are the same as real inner product spaces, essentially, except for the conjugate symmetry property, as noted above. For example, we define

$$\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x}\rangle^{1/2} \qquad (21)$$

to be the length of a vector $\mathbf{x}$ in a complex vector space, just as in a real vector space. The only thing that we must do, though, is extend the proof of Schwarz's Inequality to the complex case, since our proof made use of the symmetry property of real inner products.

**Theorem 2** (Schwarz's Inequality). *Let $V$ be a complex inner product space with inner product $\langle\ ,\ \rangle$. Then*

$$|\langle \mathbf{x}, \mathbf{y}\rangle| \leqslant \|\mathbf{x}\|\,\|\mathbf{y}\| \qquad (22)$$

*for all* $\mathbf{x}$, $\mathbf{y}$ *in* $V$, *where* $\|\mathbf{x}\|$ *is defined by Equation* (21). *Furthermore, equality holds in Equation* (22) *if and only if the vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *are linearly dependent.*

PROOF. Observe from property (iv) that

$$\left\langle \mathbf{x} - \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\mathbf{y}, \mathbf{x} - \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\mathbf{y} \right\rangle \geqslant 0 \tag{23}$$

for all vectors $\mathbf{x}$ and nonzero vectors $\mathbf{y}$ in $V$. Using properties (i)–(iii), we can rewrite the left-hand side of Equation (23) in the form

$$\begin{aligned}
&\langle \mathbf{x},\mathbf{x}\rangle - \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\langle \mathbf{y},\mathbf{x}\rangle - \frac{\overline{\langle \mathbf{x},\mathbf{y}\rangle}}{\langle \mathbf{y},\mathbf{y}\rangle}\langle \mathbf{x},\mathbf{y}\rangle + \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\frac{\overline{\langle \mathbf{x},\mathbf{y}\rangle}}{\langle \mathbf{y},\mathbf{y}\rangle}\langle \mathbf{y},\mathbf{y}\rangle \\
&\quad = \langle \mathbf{x},\mathbf{x}\rangle - \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\overline{\langle \mathbf{x},\mathbf{y}\rangle} - \frac{\overline{\langle \mathbf{x},\mathbf{y}\rangle}}{\langle \mathbf{y},\mathbf{y}\rangle}\langle \mathbf{x},\mathbf{y}\rangle + \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{y},\mathbf{y}\rangle}\overline{\langle \mathbf{x},\mathbf{y}\rangle} \\
&\quad = \|\mathbf{x}\|^2 - \frac{|\langle \mathbf{x},\mathbf{y}\rangle|^2}{\|\mathbf{y}\|^2} \\
&\quad = \frac{\|\mathbf{x}\|^2\|\mathbf{y}\|^2 - |\langle \mathbf{x},\mathbf{y}\rangle|^2}{\|\mathbf{y}\|^2}.
\end{aligned}$$

Hence,

$$|\langle \mathbf{x},\mathbf{y}\rangle|^2 \leqslant \|\mathbf{x}\|^2\|\mathbf{y}\|^2, \tag{24}$$

and Schwarz's Inequality follows immediately by taking square roots in Equation (24).

**Remark.** As in the real case, our proof assumes that $\mathbf{y} \neq \mathbf{0}$. We leave it to the reader (see Exercise 13) to complete the proof for $\mathbf{y} = \mathbf{0}$ and to verify that equality holds if and only if $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.

## EXERCISES

**1.** Show that Equation (6) defines an inner product on $C[a,b]$, the space of all continuous functions on the interval $[a,b]$.

**2.** Show that Equation (7) defines an inner product on $C[a,b]$.

**3.** Show that $\|\mathbf{x}\|_1$ defined in Example 4 is a norm.

**4.** Show that $\|\mathbf{x}\|_\infty$ defined in Example 5 is a norm.

**5.** Show that $\|\mathbf{x}\|_p$ defined in Example 6 is a norm.

**6.** Show that $\|\mathbf{x}\|_\infty = \lim\limits_{p\to\infty} \|\mathbf{x}\|_p$.

**7.** Show that $\|\mathbf{x}\| = |x_1| + 2|x_2| + 4|x_3|$ is a norm on $R^3$.

**8.** Show that $|x_1| + |x_2| + |x_3|^2$ does not define a norm on $R^3$.

**9.** Show that $e^{|x_1|+|x_2|}$ does not define a norm on $R^2$.

**10.** Show that Equation (11) defines a norm on $C[a,b]$.

**11.** Let $V$ be a real inner product space. Prove that

$$\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{4}\|\mathbf{x} + \mathbf{y}\|^2 - \frac{1}{4}\|\mathbf{x} - \mathbf{y}\|^2.$$

**12.** Let $\mathbf{x} \times \mathbf{y}$ be the vector with components

$$x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1,$$

respectively. Show that $\mathbf{x} \times \mathbf{y}$ is perpendicular to both $\mathbf{x}$ and $\mathbf{y}$ under the inner product $(\mathbf{x}, \mathbf{y})$.

**13.** Complete the proof of Schwarz's Inequality in both $R^n$ and $C^n$ for the case $\mathbf{y} = \mathbf{0}$.

**14.** Show that

$$\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$$

for all vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ in $C^n$.

## 6.3 Orthogonal bases, Hermitian operators

It is often the case that a crucial step in the solution of many problems in a linear vector space revolves around a judicious choice of basis. A type of basis that is especially useful is one that is *orthogonal*.

**Definition.** A set of vectors is said to be *orthogonal* if the inner product of any two distinct vectors in the set is zero.

One of the nice things about an orthogonal set of vectors is that they are automatically linearly independent. This is the content of Lemma 1.

**Lemma 1.** *Let* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ *be mutually orthogonal, that is,*

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0, \qquad i \neq j.$$

*Then,* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ *are linearly independent.*

PROOF. Suppose that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_N\mathbf{x}_N = \mathbf{0}. \tag{1}$$

Taking inner products of both sides of Equation (1) with $\mathbf{x}_j$ gives

$$c_1\langle \mathbf{x}_1, \mathbf{x}_j \rangle + c_2\langle \mathbf{x}_2, \mathbf{x}_j \rangle + \cdots + c_N\langle \mathbf{x}_N, \mathbf{x}_j \rangle = 0. \tag{2}$$

The left-hand side of Equation (2) reduces to $c_j\langle \mathbf{x}_j, \mathbf{x}_j \rangle$. Since $\langle \mathbf{x}_j, \mathbf{x}_j \rangle > 0$, we see that $c_j = 0$, $j = 1, 2, \ldots, N$, which proves Lemma 1.

Another benefit of working with orthogonal bases is that it is relatively easy to find the coordinates of a vector with respect to a given orthogonal basis. To wit, let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be a mutually orthogonal set of vectors in a real $n$-dimensional vector space $V$. By Lemma 1, this set of vectors is also a basis for $V$, and every vector $\mathbf{x}$ in $V$ can be expanded in the form

$$\mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n. \tag{3}$$

Taking inner products of both sides of Equation (3) with $\mathbf{u}_j$ gives $\langle \mathbf{x}, \mathbf{u}_j\rangle = c_j\langle \mathbf{u}_j, \mathbf{u}_j\rangle$ so that

$$c_j = \frac{\langle \mathbf{x}, \mathbf{u}_j\rangle}{\langle \mathbf{u}_j, \mathbf{u}_j\rangle}, \qquad j = 1, 2, \ldots, n. \tag{4}$$

**Example 1.** Let $V = R^2$ and define $\langle \mathbf{x}, \mathbf{y}\rangle = x_1y_1 + x_2y_2$. The vectors

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

are orthogonal and, thus, form a basis for $R^2$. Consequently, from Equations (3) and (4), any vector $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ can also be written in the form

$$\begin{aligned}
\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= c_1\begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2\begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
&= \frac{\langle \mathbf{x}, \mathbf{u}_1\rangle}{\langle \mathbf{u}_1, \mathbf{u}_1\rangle}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{\langle \mathbf{x}, \mathbf{u}_2\rangle}{\langle \mathbf{u}_2, \mathbf{u}_2\rangle}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
&= \frac{x_1 + x_2}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{x_1 - x_2}{2}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.
\end{aligned}$$

Given the benefit of working with orthogonal bases, the following question naturally arises: Does every $n$-dimensional Euclidean space have an orthogonal basis? The answer to this question is a resounding yes and is given by the following theorem.

**Theorem 3** (Gram-Schmidt). *Every n-dimensional Euclidean space V has an orthogonal basis.*

PROOF. Choose a basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ for $V$. We will inductively construct an orthogonal basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ by taking suitable combinations of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. To this end, let $\mathbf{v}_1 = \mathbf{u}_1$ and set

$$\mathbf{v}_2 = \mathbf{u}_2 + \lambda\mathbf{v}_1. \tag{5}$$

Taking the inner product of $\mathbf{v}_2$ with $\mathbf{v}_1$ gives

$$\langle \mathbf{v}_2, \mathbf{v}_1\rangle = \langle \mathbf{u}_2, \mathbf{v}_1\rangle + \lambda\langle \mathbf{v}_1, \mathbf{v}_1\rangle$$

so that $\mathbf{v}_2$ will be orthogonal to $\mathbf{v}_1$ if $\lambda = -\langle \mathbf{u}_2, \mathbf{v}_1\rangle/\langle \mathbf{v}_1, \mathbf{v}_1\rangle$. Note that $\mathbf{v}_2$ cannot equal $\mathbf{0}$ by virtue of the linear independence of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Proceeding inductively, let us assume that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are mutually orthogonal, and set

$$\mathbf{v}_{k+1} = \mathbf{u}_{k+1} + \lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k. \tag{6}$$

The requirement that $\mathbf{v}_{k+1}$ be orthogonal to $\mathbf{v}_1, \ldots, \mathbf{v}_k$ yields

$$0 = \langle \mathbf{v}_{k+1}, \mathbf{v}_j \rangle = \langle \mathbf{u}_{k+1}, \mathbf{v}_j \rangle + \lambda_j \langle \mathbf{v}_j, \mathbf{v}_j \rangle,$$

which gives the desired relation

$$\lambda_j = -\frac{\langle \mathbf{u}_{k+1}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}, \qquad j = 1, \ldots, k.$$

With this choice of $\lambda_1, \ldots, \lambda_k$, the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$, $\mathbf{v}_{k+1}$ are mutually orthogonal. Moreover, $\mathbf{v}_{k+1}$ cannot be $\mathbf{0}$ by virtue of the linear independence of $\mathbf{u}_1, \ldots, \mathbf{u}_{k+1}$. Proceeding inductively until $k = n$, we obtain $n$ mutually orthogonal nonzero vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

**Remark 1.** The procedure outlined in Theorem 3 is usually referred to as the Gram-Schmidt orthogonalization procedure.

**Remark 2.** If $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is an orthogonal set of vectors, and, in addition, each of the vectors $\mathbf{u}_j$ has length 1, i.e.,

$$\|\mathbf{u}_j\| = \langle \mathbf{u}_j, \mathbf{u}_j \rangle^{1/2} = 1, \qquad j = 1, \ldots, n,$$

then $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is called an *orthonormal* set of vectors. It is a simple matter to construct an orthonormal basis from an orthogonal one: if $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is an orthogonal basis, then

$$\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}, \qquad \mathbf{v}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|}, \qquad \ldots, \qquad \mathbf{v}_n = \frac{\mathbf{u}_n}{\|\mathbf{u}_n\|}$$

are orthonormal, since

$$\|\mathbf{v}_j\| = \left\| \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|} \right\| = \frac{1}{\|\mathbf{u}_j\|} \|\mathbf{u}_j\| = 1.$$

**Example 2.** Let $V$ be the space of all polynomials of degree $n - 1$, and define

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x)\, dx$$

for all functions $f$ and $g$ in $V$. It is easily verified that $V$ is an $n$-dimensional Euclidean space and that

$$f_0(x) = 1, \qquad f_1(x) = x, \qquad \ldots, \qquad f_{n-1}(x) = x^{n-1}$$

form a basis for $V$. Applying the Gram-Schmidt orthogonalization procedure to the functions $f_0(x), \ldots, f_{n-1}(x)$ gives

$$p_0(x) = 1;$$

$$p_1(x) = x + \lambda = x - \frac{\langle f_1, p_0\rangle}{\langle p_0, p_0\rangle} = x - \frac{\int_{-1}^{1} x\,dx}{\int_{-1}^{1} 1\,dx} = x;$$

$$p_2(x) = x^2 + \lambda_2 x + \lambda_1$$

where

$$\lambda_2 = -\frac{\int_{-1}^{1} x^2 \cdot x\,dx}{\int_{-1}^{1} x^2\,dx} = 0, \qquad \lambda_1 = -\frac{\int_{-1}^{1} x^2\,dx}{\int_{-1}^{1} 1\,dx} = -\frac{1}{3}$$

so that

$$p_2(x) = x^2 - \frac{1}{3},$$

$$p_3(x) = x^3 + \lambda_3\left(x^2 - \frac{1}{3}\right) + \lambda_2 x + \lambda_1$$

with

$$\lambda_3 = \frac{-\int_{-1}^{1} x^3(x^2 - \frac{1}{3})\,dx}{\int_{-1}^{1} (x^2 - \frac{1}{3})^2\,dx} = 0, \qquad \lambda_2 = \frac{-\int_{-1}^{1} x^3 \cdot x\,dx}{\int_{-1}^{1} x^2\,dx} = -\frac{3}{5},$$

and

$$\lambda_1 = \frac{-\int_{-1}^{1} x^3\,dx}{\int_{-1}^{1} 1\,dx} = 0$$

so that

$$p_3(x) = x^3 - \frac{3}{5}x.$$

These polynomials $p_0(x)$, $p_1(x)$, $p_2(x)$, $p_3(x), \ldots$, when properly normalized, i.e., when multiplied by a suitable constant, are the Legendre Polynomials $P_k(x)$ discussed in Chapter 2.

We now digress for a moment to discuss the following seemingly totally unrelated problem. Let $V = C^n$ with inner product

$$\langle \mathbf{x}, \mathbf{y}\rangle = (\mathbf{x}, \mathbf{y}) = x_1\bar{y}_1 + x_2\bar{y}_2 + \cdots + x_n\bar{y}_n$$

where $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are the coordinates of $\mathbf{x}$ and $\mathbf{y}$, respectively, and let $A$ be a given $n \times n$ matrix. Does there exist a matrix $B$ such that

$$(A\mathbf{x},\mathbf{y}) = (\mathbf{x}, B\mathbf{y}) \tag{7}$$

for all vectors $\mathbf{x}$ and $\mathbf{y}$?

To answer this question, we let $A = (a_{ij})$, $B = (b_{ij})$ and observe that

$$(A\mathbf{x},\mathbf{y}) = \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}x_i\right)\bar{y}_j = \sum_{i,j=1}^{n} a_{ji}x_i\bar{y}_j$$

and

$$(\mathbf{x}, B\mathbf{y}) = \sum_{i=1}^{n} x_i\overline{\left(\sum_{j=1}^{n} b_{ij}y_j\right)} = \sum_{i,j=1}^{n} \overline{b_{ij}}x_i\bar{y}_j.$$

Hence, Equation (7) is satisfied if and only if $\overline{b_{ij}} = a_{ji}$ or, equivalently,

$$b_{ij} = \overline{a_{ji}}. \tag{8}$$

**Definition.** The matrix $B$ whose elements are given by Equation (8) is called the *adjoint* of $A$ and is denoted by $A^*$. If $A^* = A$, then $A$ is said to be *selfadjoint* or *Hermitian*, and

$$(A\mathbf{x},\mathbf{y}) = (\mathbf{x}, A\mathbf{y}) \quad \text{for all } \mathbf{x} \text{ and } \mathbf{y}.$$

**Remark 1.** If $A$ is real, then $A^*$ is simply $A^{\mathrm{T}}$, the transpose of $A$. Thus, if $A$ is real and symmetric, i.e., $A = A^{\mathrm{T}}$, then $A$ is Hermitian.

**Remark 2.** It is a simple matter to show (see Exercise 6) that $(A^*)^* = A$ and that

$$(\mathbf{x}, A\mathbf{y}) = (A^*\mathbf{x},\mathbf{y}) \quad \text{for all } \mathbf{x} \text{ and } \mathbf{y}.$$

**Example 3.** Let

$$A = \begin{bmatrix} 1+i & 2 & i \\ 3 & -i & 1-i \\ i & 2i & 3i \end{bmatrix}.$$

Then,

$$A^* = \begin{bmatrix} 1-i & 3 & -i \\ 2 & i & -2i \\ -i & 1+i & -3i \end{bmatrix}.$$

**Example 4.** Let

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}.$$

Then $A^* = A^{\mathrm{T}} = A$.

Selfadjoint matrices have several very important properties, which we now wish to describe.

**Property 1.** Let $A$ be selfadjoint and let $\mathbf{x}$ be an eigenvector of $A$ with eigenvalue $\lambda$. Since $A$ is selfadjoint, $(A\mathbf{x},\mathbf{x}) = (\mathbf{x},A\mathbf{x})$, but

$$(A\mathbf{x},\mathbf{x}) = (\lambda\mathbf{x},\mathbf{x}) = \lambda(\mathbf{x},\mathbf{x}),$$

while

$$(\mathbf{x},A\mathbf{x}) = (\mathbf{x},\lambda\mathbf{x}) = \bar{\lambda}(\mathbf{x},\mathbf{x}).$$

Hence, $\lambda = \bar{\lambda}$.

**Corollary.** *The eigenvalues of a real symmetric matrix are real.*

**Property 2.** Let $A$ be selfadjoint, and let $\mathbf{u}_1$ and $\mathbf{u}_2$ be eigenvectors of $A$ with eigenvalues $\lambda_1$ and $\lambda_2$, respectively, with $\lambda_1 \neq \lambda_2$. Then, $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal, i.e., $(\mathbf{u}_1,\mathbf{u}_2) = 0$.

PROOF. Observe first that

$$(A\mathbf{u}_1,\mathbf{u}_2) = (\lambda_1\mathbf{u}_1,\mathbf{u}_2) = \lambda_1(\mathbf{u}_1,\mathbf{u}_2).$$

Since $A$ is selfadjoint,

$$(A\mathbf{u}_1,\mathbf{u}_2) = (\mathbf{u}_1,A\mathbf{u}_2) = (\mathbf{u}_1,\lambda_2\mathbf{u}_2) = \lambda_2(\mathbf{u}_1,\mathbf{u}_2)$$

since $\lambda_2$ is real. Hence,

$$\lambda_1(\mathbf{u}_1,\mathbf{u}_2) = \lambda_2(\mathbf{u}_1,\mathbf{u}_2), \tag{9}$$

and since $\lambda_1 \neq \lambda_2$, Equation (9) now forces $(\mathbf{u}_1,\mathbf{u}_2) = 0$, which proves Property 2.

In actuality, we can state a much stronger property than Property 2: Not only are the eigenvectors belonging to distinct eigenvalues of a selfadjoint matrix orthogonal, but every $n \times n$ selfadjoint matrix posesses a full complement of $n$ mutually orthogonal eigenvectors. This is the content of Property 3.

**Property 3.** Every $n \times n$ selfadjoint matrix $A$ has $n$ mutually orthogonal eigenvectors.

PROOF. Every matrix has at least one eigenvector. Let $\mathbf{v}_1$ be an eigenvector of $A$ with eigenvalue $\lambda_1$, and let $W$ be the subspace of $V$, which consists of all vectors in $V$ that are orthogonal to $\mathbf{v}_1$. We leave it to the reader to verify (see Exercise 7) that $W$ is a linear subspace of $V$ of dimension $n-1$. The crucial observation now is that $A$ takes $W$ into itself; i.e., if $\mathbf{x}$ is orthogonal to $\mathbf{v}_1$, then so is $A\mathbf{x}$. But this follows immediately from the observation that

$$(A\mathbf{x},\mathbf{v}_1) = (\mathbf{x},A^*\mathbf{v}_1) = (\mathbf{x},A\mathbf{v}_1) = (\mathbf{x},\lambda_1\mathbf{v}_1) = \lambda_1(\mathbf{x},\mathbf{v}_1).$$

Thus, $(\mathbf{x}, \mathbf{v}_1) = 0 \Rightarrow (A\mathbf{x}, \mathbf{v}_1) = 0$.

The next part of our proof is a little tricky. Recall from Section 3.7 that every $n \times n$ matrix $A$ induces a linear transformation $\mathcal{A}$ on $C^n(R^n)$ with

$$\mathcal{A}(\mathbf{x}) = A\mathbf{x}. \tag{10}$$

Conversely, every linear transformation $\mathcal{A}$ on $C^n(R^n)$ can be written in the form (10) for some appropriate matrix $A$. Let $\mathcal{A}$ be the linear transformation induced by $A^*$. From the remarks above, $\mathcal{A}$ takes $W$ into itself. Thus, assuming that (see Exercise 8) every linear transformation that takes a vector space into itself has at least one eigenvector, we see that $\mathcal{A}$ has at least one eigenvector $\mathbf{v}_2$ in $W$; i.e.,

$$\mathcal{A}(\mathbf{v}_2) = A\mathbf{v}_2 = \lambda_2\mathbf{v}_2$$

for some number $\lambda_2$. Proceeding inductively, that is, by considering the subspace of $W$ perpendicular to $\mathbf{v}_2$, we can produce $n$ eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of $A$ that are mutually orthogonal, and this completes the proof of Property 3.

**Corollary.** *Every symmetric $n \times n$ matrix $A$ has $n$ mutually orthogonal eigenvectors.*

At the beginning of this section we mentioned that orthogonal bases were often very useful in applications. In Chapter 3, we observed that a basis that consists of eigenvectors of a matrix $A$ is also very useful since in this basis the matrix is diagonal. The ideal situation in practice is a basis of eigenvectors of a selfadjoint matrix $A$, for then we have the best of both worlds, i.e., the eigenvectors form an orthogonal basis. This is the real secret behind the famous Sturm-Liouville Theory to which we now turn our attention.

## Exercises

**1.** Let $V$ be the space of all polynomials of degree $\leqslant 2$ on the interval $[-2, 2]$, with inner product

$$\langle f, g\rangle = \int_{-2}^{2} x^2 f(x)g(x)\,dx.$$

Starting with the basis 1, $x$, and $x^2$, use the Gram-Schmidt procedure to construct an orthonormal basis for $V$.

**2.** Let $V$ be the space of all polynomials of degree $< n$, on the interval $[a, b]$, and define

$$\langle f, g\rangle = \int_{a}^{b} r(x)f(x)g(x)\,dx.$$

Use the Gram-Schmidt procedure to obtain an orthogonal basis for $V$ in the case
(a) $(a, b) = (-1, 1)$, $r(x) = |x|^{1/2}$;
(b) $(a, b) = (-\infty, \infty)$, $r(x) = e^{-x}$.

3. Obtain the first four Hermite polynomials by applying the Gram-Schmidt procedure to the monomials 1, $x$, $x^2$, and $x^3$, with weight function $r(x) = e^{-x^2/2}$.

4. Let $V$ be the $n$-dimensional space spanned by the $n$ functions $e^{-kx}$, $k = 0, 1, \ldots, n-1$, and let

$$(f, g) = \int_0^\infty f(x)g(x)\,dx$$

for $f$, $g$ in $V$. Obtain an orthogonal basis for $V$.

5. Obtain the first four Laguerre polynomials by applying the Gram-Schmidt procedure to the monomials 1, $x$, $x^2$, and $x^3$, with weight function $r(x) = e^{-x}$.

6. Show that $(A^*)^* = A$.

7. Show that the subspace $W$ introduced in the proof of Property 3 has dimension $n - 1$.

8. Show that every linear transformation that takes a vector space into itself has at least one eigenvector.

## 6.4 Sturm-Liouville theory

In Chapter 5 we studied equations of the form

$$L[y](x) = y''(x) + \lambda y(x) = 0 \tag{1}$$

subject to the boundary conditions

$$a_1 y(0) + b_1 y'(0) = 0, \qquad a_2 y(l) + b_2 y'(l) = 0. \tag{1$'$}$$

Equation (1) together with the boundary conditions (1′) was referred to as a boundary value problem. We then stated a theorem to the effect that this boundary value problem has a nontrivial solution only for a special set of numbers

$$\lambda_1 < \lambda_2 < \lambda_3 < \cdots < \lambda_n < \cdots$$

called *eigenvalues* and every differentiable function $y(x)$ can be expanded in a series of these eigenfunctions, i.e.,

$$y(x) = \sum_{n=1}^{\infty} c_n y_n(x), \qquad 0 < x < l.$$

For example, the boundary value problem

$$y''(x) + \lambda y(x) = 0, \qquad y(0) = 0,\ y(l) = 0, \tag{2}$$

has eigenvalues $\lambda_n = n^2\pi^2/l^2$, with corresponding eigenfunctions

$$y_n(x) = \sin\frac{n\pi x}{l}, \qquad n = 1, 2, \ldots,$$

and every differentiable function $f(x)$ can be expanded in a Fourier series of the form

$$y(x) = \sum_{n=1}^{\infty} c_n \sin\frac{n\pi x}{l}, \qquad 0 < x < l.$$

The theorem we just stated is a special case of a much more general theorem, which applies to the more general equation

$$L[y](x) = a(x)y''(x) + b(x)y'(x) + c(x)y(x) = -\lambda r(x)y(x) \tag{3}$$

and the boundary conditions

$$a_1 y(\alpha) + b_1 y'(\alpha) = 0, \qquad a_2 y(\beta) + b_2 y'(\beta) = 0. \tag{3'}$$

We can gain a lot of insight into this problem by studying it from a linear algebra point of view. To wit, let $V$ be the space of all twice continuously differentiable functions. $V$ is a linear vector space but is infinite dimensional, rather than finite dimensional. Next, define an inner product on $V$ via the relation

$$(f, g) = \int_{\alpha}^{\beta} f(x)g(x)\,dx. \tag{4}$$

The *operator* $L$ defined by

$$L[y](x) = a(x)y''(x) + b(x)y'(x) + c(x)y(x) \tag{5}$$

is a *linear* operator defined on $V$ (see Section 2.1) and it is natural to ask whether $L$ is *selfadjoint* in the sense that

$$(Lu, v) = (u, Lv) \tag{6}$$

for all functions $u$, $v$ in $V$. The answer to this question is given in Theorem 4.

**Theorem 4.** *Let $V$ be the space of all twice continuously differentiable functions on the interval $[\alpha, \beta]$. Then, the operator $L$ defined by Equation* (5) *is selfadjoint on $V$, in the sense of Equation* (6), *if and only if*

$$b(x) = a'(x). \tag{7}$$

PROOF. Computing

$$(Lu, v) = \int_{\alpha}^{\beta} [a(x)u''(x) + b(x)u'(x) + c(x)u(x)]v(x)\,dx$$

and

$$(u, Lv) = \int_{\alpha}^{\beta} [a(x)v''(x) + b(x)v'(x) + c(x)v(x)]u(x)\,dx$$

we see that

$$(Lu,v)-(u,Lv)=\int_\alpha^\beta [a(x)v(x)u''(x)+b(x)v(x)u'(x)]\,dx$$
$$-\int_\alpha^\beta [a(x)u(x)v''(x)+b(x)u(x)v'(x)]\,dx.$$

Integrating the terms containing $u''(x)$ and $v''(x)$ by parts gives

$$\begin{aligned}(Lu,v)-(u,Lv)=&-\int_\alpha^\beta a(x)u'(x)v'(x)\,dx-\int_\alpha^\beta a'(x)v(x)u'(x)\,dx\\&+\int_\alpha^\beta b(x)v(x)u'(x)\,dx+\int_\alpha^\beta a(x)u'(x)v'(x)\,dx\\&+\int_\alpha^\beta a'(x)u(x)v'(x)\,dx-\int_\alpha^\beta b(x)u(x)v'(x)\,dx\\&+a(x)v(x)u'(x)|_\alpha^\beta-a(x)u(x)v'(x)|_\alpha^\beta\\=&\int_\alpha^\beta [b(x)-a'(x)][v(x)u'(x)-u(x)v'(x)]\,dx\\&+a(x)[v(x)u'(x)-u(x)v'(x)]_\alpha^\beta.\end{aligned}\tag{8}$$

The boundary terms in Equation (8) are zero. To see this, assume that the constants $b_1, b_2$ in Equation (3′) are not zero (the proof is trivial if they are—see Exercise 1), and set $c_1=a_1/b_1$, $c_2=a_2/b_2$. Then

$$\begin{aligned}&a(x)[v(x)u'(x)-u(x)v'(x)]_\alpha^\beta\\&\quad=a(\beta)[v(\beta)u'(\beta)-u(\beta)v'(\beta)]-a(\alpha)[v(\alpha)u'(\alpha)-u(\alpha)v'(\alpha)];\end{aligned}\tag{9}$$

but

$$u'(\beta)=c_2u(\beta),\quad v'(\beta)=c_2v(\beta),\qquad u'(\alpha)=c_1u(\alpha),\quad v'(\alpha)=c_1v(\alpha)$$

since $u(x)$ and $v(x)$ satisfy the boundary conditions (3′). Hence, the right-hand side of Equation (9) becomes

$$a(\beta)[v(\beta)c_2u(\beta)-u(\beta)c_2v(\beta)]-a(\alpha)[v(\alpha)c_1u(\alpha)-u(\alpha)c_1v(\alpha)]=0.$$

Thus, we have shown that

$$(Lu,v)-(u,Lv)=\int_\alpha^\beta [b(x)-a'(x)][v(x)u'(x)-u(x)v'(x)]\,dx\tag{10}$$

for all $u$ and $v$ in $V$. The right-hand side of Equation (10) is zero if $b(x)=a'(x)$. Thus, $L$ is certainly selfadjoint in the sense of Equation (6) if $b(x)=a'(x)$. In addition, it is intuitively clear that $b(x)=a'(x)$ is also a necessary condition for $L$ to be selfadjoint; the right-hand side of Equation (10) cannot vanish for all $u$ and $v$ in $V$ if $b(x)\neq a'(x)$. A proof of this assertion is outlined in the exercises.

In summary, the boundary value problem (3), (3′) is self-adjoint if and only if $b(x) = a'(x)$. Setting $a(x) = p(x)$ and $c(x) = -q(x)$, we can now write all selfadjoint boundary value problems in the canonical form

$$L[y](x) = [p(x)y'(x)]' - q(x)y(x) = -\lambda r(x)y(x) \tag{11}$$

$$a_1 y(\alpha) + b_1 y'(\alpha) = 0, \qquad a_2 y(\beta) + b_2 y'(\beta) = 0. \tag{11'}$$

Equation (11) together with the boundary conditions (11′) is often referred to as a Sturm-Liouville boundary value problem.

**Remark 1.** In the notation of Equation (11), Equation (8) becomes

$$(Lu, v) - (u, Lv) = -p(x)[u'(x)v(x) - u(x)v'(x)]_\alpha^\beta. \tag{12}$$

Equation (12) is known as Lagrange's identity, and if $u$ and $v$ satisfy the boundary conditions (11′), then Lagrange's identity reduces to

$$(Lu, v) - (u, Lv) = 0. \tag{13}$$

**Remark 2.** We have derived Equation (13) in the case that $u$ and $v$ are both real, because that is how Lagrange's identity is usually expressed. It is a simple matter, however, to verify (see Exercise 3) that Equation (13) is also true for $u$ and $v$ complex, provided we define

$$(f, g) = \int_\alpha^\beta f(x)\overline{g(x)}\,dx. \tag{14}$$

**Definition.** A Sturm-Liouville boundary value problem is said to be *regular* if each of the following conditions hold:

(i) $r(x) > 0$ and $p(x) > 0$ for $x \in [\alpha, \beta]$.
(ii) $p(x)$, $p'(x)$, $q(x)$, and $r(x)$ are continuous on $[\alpha, \beta]$.
(iii) $(\alpha, \beta)$ is finite.

We are now ready to state and prove the main theorem of this chapter.

**Theorem 5.** *For any regular Sturm-Liouville boundary value problem*:

(1) *All the eigenvalues (and consequently all the eigenfunctions) are real.*

(2) *Eigenfunctions belonging to different eigenvalues are orthogonal under the inner product*

$$\langle f, g\rangle = \int_\alpha^\beta r(x)f(x)g(x)\,dx. \tag{15}$$

(3) *To each eigenvalue, there corresponds one, and only one, eigenfunction.*

(4) *There are a countably infinite number of eigenvalues $\lambda_0, \lambda_1, \lambda_2, \ldots$ with corresponding eigenfunctions $y_0(x)$, $y_1(x)$, $y_2(x), \ldots$. The eigenvalues $\lambda_n$ can be ordered so that $n$ refers to the number of zeros of $y_n$ in the interval $[\alpha, \beta]$.*

*In addition, the zeros of $y_{n+1}$ interlace those of $y_n$. Finally, $\lambda_0 < \lambda_1 < \lambda_2 < \lambda_3 \cdots$ with $\lambda_n \to \infty$ as $n \to \infty$.*

(5) *Let $f(x)$ be any continuously differentiable function on the interval $[\alpha, \beta]$. Then, $f(x)$ can be expanded in a convergent series of the eigenfunctions of $L$; i.e.,*

$$f(x) = \sum_{n=0}^{\infty} a_n y_n(x)$$

*where*

$$a_n = \frac{\displaystyle\int_\alpha^\beta r(x) f(x) y_n(x)\,dx}{\displaystyle\int_\alpha^\beta r(x) y_n^2(x)\,dx}. \tag{16}$$

PROOF. (1) The proof of (1) follows immediately from the selfadjointness of $L$: if $Ly = -\lambda y$ then

$$(Ly, y) = (-\lambda r y, y) = -\lambda \int_\alpha^\beta r(x)|y^2(x)|\,dx \tag{17}$$

and

$$(y, Ly) = (y, -\lambda r y) = -\bar{\lambda} \int_\alpha^\beta r(x)|y^2(x)|\,dx \tag{18}$$

since $r(x)$ is real. Equating Equations (17) and (18), we see that $\lambda$ must be real.

(2) Let $y_n(x)$ and $y_m(x)$ be eigenfunctions of $L$ with eigenvalues $\lambda_n$ and $\lambda_m$, respectively, with $\lambda_n \neq \lambda_m$. Then

$$\begin{aligned}(Ly_n, y_m) &= (-\lambda_n r y_n, y_m)\\ &= -\lambda_n \int_\alpha^\beta r(x) y_n(x) y_m(x)\,dx;\end{aligned}$$

however, since $L$ is selfadjoint,

$$\begin{aligned}(Ly_n, y_m) &= (y_n, Ly_m)\\ &= (y_n, -\lambda_m r y_m)\\ &= -\lambda_m \int_\alpha^\beta r(x) y_n(x) y_m(x)\,dx,\end{aligned}$$

and since $\lambda_n \neq \lambda_m$, we see that

$$\int_\alpha^\beta r(x) y_n(x) y_m(x)\,dx = \langle y_n, y_m \rangle = 0.$$

(3) Suppose that $u_1(x)$ and $u_2(x)$ are two independent eigenfunctions with eigenvalue $\lambda$. Then both $u_1$ and $u_2$ satisfy the second-order differential

equation

$$(p(x)y')' - q(x)y = -\lambda r(x)y \tag{19}$$

in addition to the boundary conditions

$$a_1 y(\alpha) + b_1 y'(\alpha) = 0, \qquad a_2 y(\beta) + b_2 y'(\beta) = 0. \tag{20}$$

Since $u_1$ and $u_2$ are independent, they form a basis for all the solutions of Equation (19), that is, every solution $y(x)$ of Equation (19) can be written in the form

$$y(x) = c_1 u_1(x) + c_2 u_2(x) \tag{21}$$

for some choice of constants $c_1$, $c_2$. Finally, observe that if $u_1(x)$ and $u_2(x)$ satisfy the boundary conditions (20), then so does the linear combination (21). Hence, every solution of Equation (19) must satisfy the boundary conditions (20), which is clearly absurd. Thus, to each eigenvalue $\lambda$ of $L$, there corresponds only one eigenfunction.

(4) The proof of (4), though not very difficult, is quite long, so it will be omitted here.

(5) The convergence part of (5) is too difficult to present here. We wish to point out though, that if $f(x)$ can be expanded in a series of the eigenfunctions $y_n(x)$, i.e.,

$$f(x) = \sum_{n=0}^{\infty} a_n y_n(x) \tag{22}$$

then $a_n$ must, of necessity, be given by Equation (16). To see this, assume that the series (22) is convergent and that the inner product of $f$ with any function $y(x)$ can be distributed, that is,

$$\langle f, y \rangle = \sum_{n=0}^{\infty} a_n \langle y_n, y \rangle. \tag{23}$$

If $y(x) = y_j(x)$, then Equation (23) reduces to

$$\langle f, y_j \rangle = a_j \langle y_j, y_j \rangle \tag{24}$$

by virtue of the orthogonality of the eigenfunctions $y_0(x), y_1(x), y_2(x), \ldots$, and it now follows immediately from Equation (24) that

$$a_j = \frac{\langle f, y_j \rangle}{\langle y_j, y_j \rangle} = \frac{\int_\alpha^\beta r(x) f(x) y_j(x)\, dx}{\int_\alpha^\beta r(x) y_j^2(x)\, dx}. \tag{25}$$

Part (5) of Theorem 5 deals with *pointwise* convergence, i.e., the series (22) converges for each $x$ in $(\alpha, \beta)$ if $f(x)$ is continuously differentiable. A different type of convergence, which has proven quite useful in applications, is known as convergence in the mean.

**Definition.** A sequence of functions $f_n(x)$, $\alpha < x < \beta$, is said to converge in the mean to $f(x)$ if

$$\|f - f_n\| = \left[\int_\alpha^\beta r(x)(f(x) - f_n(x))^2\,dx\right]^{1/2}$$

approaches zero as $n \to 0$.

Part (5) of Theorem 5 now has the following analogue in terms of *mean convergence.*

**Theorem 6.** *Let*

$$f_n(x) = \sum_{j=0}^{n} a_j y_j(x) \tag{26}$$

*where $a_j$ is given by Equation* (25). *Then, $f_n(x)$ converges to $f(x)$ in the mean for all functions $f(x)$ with*

$$\int_\alpha^\beta r(x)f^2(x)\,dx < \infty,$$

*i.e., for all functions $f(x)$ with finite norm.*

**Example 1.** Consider the Sturm-Liouville boundary value problem

$$L[y] = y'' = -\lambda y, \qquad y'(0) = 0,\, y'(1) = 0. \tag{27}$$

Here $p(x) = r(x) = 1$ and $q(x) = 0$. It is easily verified that

$$y = y_n(x) = \cos n\pi x, \qquad n = 0, 1, 2, \ldots,$$

are the nontrivial solutions, i.e., eigenfunctions, of Equation (27), with corresponding eigenvalues

$$\lambda_n = n^2\pi^2, \qquad n = 0, 1, 2, \ldots.$$

Parts (1)–(3) of Theorem 5 are trivial to verify and, in fact, have been done already in Chapter 5. The zeros of $y_n(x)$ are located at the $n$ points

$$\frac{1}{2n}, \frac{3}{2n}, \frac{5}{2n}, \ldots, \frac{2n-1}{2n},$$

and it is easily verified that the zeros of $y_{n+1}$ interlace those of $y_n$. For example, the four zeros $\frac{1}{8}, \frac{3}{8}, \frac{5}{8}$, and $\frac{7}{8}$ of $y_4(x)$ interlace the three zeros $\frac{1}{6}, \frac{3}{6}, \frac{5}{6}$ of $y_3(x)$ since

$$\frac{1}{8} < \frac{1}{6} < \frac{3}{8} < \frac{3}{6} < \frac{5}{8} < \frac{5}{6} < \frac{7}{8}.$$

Finally, part (5) of Theorem 5 is simply Fourier's Theorem of Section 5.4.

Let us return now to the Hermite, Legendre, Tchebycheff, and Laguerre differential equations

$$y'' - xy' + \lambda y = 0, \qquad -\infty < x < \infty, \tag{28}$$

$$(1 - x^2)y'' - 2xy' + \lambda(\lambda + 1)y = 0, \qquad -1 < x < 1, \tag{29}$$

$$(1 - x^2)y'' - xy' + \lambda^2 y = 0, \qquad -1 < x < 1, \tag{30}$$

$$xy'' + (1 - x)y' + \lambda y = 0, \qquad 0 < x < \infty, \tag{31}$$

respectively. Except for the Legendre equation (29), the coefficient of $y'$ in each of these equations is *not* the derivative of the coefficient of $y''$. We remedy this situation by multiplying through by an appropriate integrating factor (see Section 1.2). Then we can rewrite Equations (28)–(31) in the equivalent form

$$L[y](x) = \frac{d}{dx} e^{-x^2/2} y'(x) = -\lambda e^{-x^2/2} y(x), \qquad -\infty < x < \infty, \tag{28'}$$

$$L[y](x) = \frac{d}{dx}(1 - x^2)y'(x) = -\lambda(\lambda + 1)y(x), \qquad -1 < x < 1, \tag{29'}$$

$$L[y](x) = \frac{d}{dx}(1 - x^2)^{1/2} y'(x) = \frac{-\lambda^2}{(1 - x^2)^{1/2}} y(x), \qquad -1 < x < 1, \tag{30'}$$

$$L[y](x) = \frac{d}{dx} x e^{-x} y'(x) = -\lambda e^{-x} y(x), \qquad 0 < x < \infty. \tag{31'}$$

Indeed, we can now begin to understand the presence of the factor $r(x)$ in the Sturm-Liouville problem

$$L[y](x) = -\lambda r(x) y(x).$$

Our next step is to determine what boundary conditions should be imposed on the solutions of Equations (28)–(31). To this end, observe that none of the Equations (28′)–(31′) are regular in the sense of the definition preceding Theorem 5. To wit, the functions

$$p(x) = 1 - x^2 \quad \text{and} \quad p(x) = (1 - x^2)^{1/2}$$

for the Legendre and Tchebycheff equations vanish at the end points $-1$ and $+1$ (which in a limiting sense is also true for the functions $e^{-x^2/2}$ and $xe^{-x}$, which occur in the Hermite and Laguerre equations), and the interval $(\alpha, \beta)$ for both the Hermite and Laguerre equations is not finite.

Let us return now to our proof of Theorem 4 where we derived the relation

$$(Lu, v) - (u, Lv) = -p(x)[u'(x)v(x) - u(x)v'(x)]_\alpha^\beta. \tag{32}$$

For the operators $L$ appearing in Equations (28)–(31), the integrals appearing in the left-hand sides of Equation (32) may now have to be considered as improper integrals. In the case of the Legendre and Tchebycheff equations, the function $p(x)$ vanishes at the end points. Thus, $L$ will certainly be self-adjoint if we impose the boundary conditions that $y'(x)$ [and hence $y(x)$, see Exercise 12] is bounded as $x \to \pm 1$. In the case of the Hermite equation, we require that

$$\lim_{x\to\pm\infty} e^{-x^2/2}[u(x)v'(x) - u'(x)v(x)] = 0. \tag{33}$$

In the case of the Laguerre equation, $p(x)$ vanishes at $x = 0$. Thus, one boundary condition is that $y'(x)$ [and hence $y(x)$] is bounded as $x \to 0$. The second boundary condition is determined from the requirement that

$$\lim_{x\to\infty} xe^{-x}[u(x)v'(x) - u'(x)v(x)] = 0. \tag{34}$$

Finally, we note that the conditions (33) and (34) translate into the boundary conditions

$$\lim_{x\to\pm\infty} e^{-x^2/2}y(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} xe^{-x}y(x) = 0$$

for the Hermite and Laguerre equations, respectively.

It is interesting to note (but difficult to prove) that the boundary conditions imposed on the solutions of Equations (28)–(31) are exactly sufficient to force the power series solutions of these equations to terminate at a finite point. In other words, the only solutions of Equations (28)–(31) that satisfy our boundary conditions are the Hermite, Legendre, Tchebycheff, and Laguerre polynomials, respectively; or, in the language of linear algebra, the only nontrivial solutions of the equation

$$L[y](x) = -\lambda r(x)y(x)$$

for the operators $L$ in Equations (28′)–(31′) and the boundary conditions discussed above are the eigenfunctions $y_n(x)$ with eigenvalues $\lambda_n$, and these eigenfunctions, when properly normalized, are the Hermite, Legendre, Tchebycheff, and Laguerre polynomials, respectively.

**Remark.** We have imposed boundary conditions on Equations (28)–(31) to ensure that the respective operators $L$ are selfadjoint. This will guarantee that parts (1) and (2) of Theorem 5 are true. More generally, parts (1) and (2) of Theorem 5 are true even for singular Sturm-Liouville problems, but parts (3)–(5) may sometimes fail to be true, even though they are true for the Hermite, Legendre, Tchebycheff, and Laguerre equations.

We conclude this section by proving that the only nontrivial solutions of the Hermite equation

$$y'' - 2xy' + \lambda y = 0 \quad \text{or} \quad (e^{-x^2/2}y')' = -\lambda e^{-x^2/2}y \tag{35}$$

satisfying the boundary conditions

$$\lim_{x\to\pm\infty} e^{-x^2/2}y(x) = 0 \tag{36}$$

are the Hermite polynomials $H_n(x)$, which are the polynomial solutions of the equation

$$y'' - 2xy' + 2ny = 0.$$

In other words, all nonpolynomial solutions of Equation (35) fail to satisfy the boundary conditions (36). To this end we set

$$y(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Computing

$$y'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1} \quad \text{and} \quad y''(x) = \sum_{n=0}^{\infty} n(n-1) a_n x^{n-2}$$

and plugging into Equation (35) yields

$$\sum_{n=0}^{\infty} (n+2)(n+1) a_{n+2} x^n - \sum_{n=0}^{\infty} 2n a_n x^n + \sum_{n=0}^{\infty} \lambda a_n x^n = 0.$$

This, in turn, implies that

$$a_{n+2} = \frac{(2n-\lambda)}{(n+2)(n+1)} a_n. \tag{37}$$

For the moment, let us set $a_1 = 0$ and $a_0 = 1$. Then Equation (37) forces all the odd coefficients $a_3, a_5, \ldots,$ to be zero. Next, observe that

$$a_{n+2} = \frac{2}{(n+2)} \frac{(n-\lambda/2)}{(n+1)} a_n. \tag{38}$$

For large $n$,

$$\frac{(n-\lambda/2)}{(n+1)} \approx 1.$$

Thus, for $n > N$ sufficiently large,

$$a_{n+2} \geqslant \frac{2}{n+2}(.9) a_n, \qquad n > N. \tag{39}$$

Next, look at

$$z(x) = \sum_{n=0}^{\infty} b_n x^n$$

where

$$b_{n+2} = \frac{2}{n+2}(.9) b_n \quad \text{and} \quad b_0 = 1. \tag{40}$$

Then,

$$b_2 = \frac{2}{2}(.9) = \frac{1}{1!}(.9)$$

$$b_4 = \frac{2^2}{2 \cdot 4}(.9)^2 = \frac{1}{2!}(.9)^2$$

$$b_6 = \frac{2^3}{2\cdot 4\cdot 6}(.9)^3 = \frac{1}{3!}(.9)^3$$

$$\vdots$$

and

$$\begin{aligned} z(x) &= 1 + \frac{(.9)}{1!}x^2 + \frac{(.9)^2}{2!}x^4 + \frac{(.9)^3}{3!}x^6 + \cdots \\ &= 1 + \frac{(.9x^2)}{1!} + \frac{(.9x^2)^2}{2!} + \frac{(.9x^2)^3}{3!} + \cdots \\ &= e^{.9x^2}. \end{aligned}$$

It is clear from Equations (39) and (40) that for $\lambda \neq 2n$ (so that $a_{n+2}$ is never 0)

$$y(x) \geqslant p_N(x) + e^{.9x^2}$$

for some appropriate polynomial $p_N(x)$ of degree $N$. But then

$$\lim_{x\to\pm\infty} e^{-x^2/2}y(x) = \lim_{x\to\pm\infty} [e^{-x^2/2}p_N(x) + e^{.4x^2}] = \infty.$$

Hence, the only way $y(x)$ can satisfy the boundary condition (36) is if $\lambda = 2n$, so that its series terminates.

The case $a_0 = 0$, $a_1 = 1$ is similar, and we leave it as an exercise for the reader (see Exercise 13).

## EXERCISES

**1.** Show that the boundary terms in Equation (8) are trivially zero if either $b_1$ or $b_2$ is zero.

**2.** Here is the outline of a proof that $L$ [see Equation (3)] is not selfadjoint if $b(x) \neq a'(x)$. Observe first that if $b(x) \neq a'(x)$ and we can choose $u$ and $v$ so that

$$v(x)u'(x) - u(x)v'(x) = b(x) - a'(x),$$

then the integral in Equation (10) reduces to

$$\int_\alpha^\beta [b(x) - a'(x)]^2\,dx \neq 0.$$

The only problem is that $u$ and $v$ have to satisfy the boundary conditions (3′). To modify the proof,

(a) Observe that

$$v(x)u'(x) - u(x)v'(x) = v^2(x)\left(\frac{u}{v}\right)'.$$

(b) Choose $u$, $v$, and constants $c_1, c_3, \ldots, c_{k+1}$ (with $k$ to be determined) so that

$$\left(\frac{u}{v}\right)' = c_1[b(x)-a'(x)] + c_3[b(x)-a'(x)]^3 + c_5[b(x)-a'(x)]^5 + \cdots,$$

$u$ and $v$ satisfy the boundary conditions (3′), and the integral in Equation (10) is $\neq 0$.

3. Show that Equation (13) is true even for $u$, $v$ complex.

4. Show that Liouville's transformation

$$y = \frac{z}{(p(x)r(x))^{1/4}}, \qquad t'(x) = \sqrt{\frac{r(x)}{p(x)}}$$

reduces Equation (11) to

$$z'' - f(t)z = -\lambda z.$$

What is $f(t)$? Note that under this transformation, the eigenvalues remain fixed and the weight function becomes unity.

5. Let $P_n(x)$ be the Legendre polynomial of degree $n$.
   (a) Show that $P_n'(x)$ satisfies a selfadjoint equation with $\lambda = n(n+1) - 2$.
   (b) Show that $\int_{-1}^{1} P_n'(x)P_m'(x)(1 - x^2)\,dx = 0$, $m \neq n$.

6. Find the eigenvalues and eigenfunctions of the boundary value problem

$$x^2y'' = -\lambda y, \qquad y(1) = 0, y(2) = 0.$$

(*Hint*: Try $y = x^p$.)

In each of Problems 7–9, find the eigenvalues and eigenfunctions of the given boundary value problem.

7. $y'' + \lambda y = 0; \quad y(0) = 0, y(\pi) - y'(\pi) = 0$

8. $y'' + \lambda y = 0; \quad y(0) - y'(0) = 0, y(1) = 0$

9. $y'' + \lambda y = 0; \quad y(0) - y'(0) = 0, y(\pi) - y'(\pi) = 0$

10. For which values of $\lambda$ does the boundary value problem

$$y'' - 2y' + (1 + \lambda)y = 0, \qquad y(0) = 0, y(1) = 0,$$

have a nontrivial solution?

11. Show that the singular boundary value problem

$$y'' + \lambda y = 0, \qquad y(0) = y(2\pi), y'(0) = y'(2\pi),$$

has a continuum of eigenvalues.

12. Suppose that $y'(0)$ is bounded as $x \to \pm 1$. Show that $y(x)$ must also be bounded as $x \to \pm 1$. [*Hint*: Write $y(x)$ as an integral involving $y'(x)$.]

13. Let $a_0 = 0$ and $a_1 = 1$. Show that the power series solution $\sum_{n=0}^{\infty} a_n x^n$ of Equation (35), with $a_n$ determined from Equation (37), will not satisfy the boundary conditions (36) unless the series terminates, i.e., $\lambda = 2n$ for some integer $n$.