M.A. Al-Gwaiz

# Sturm-Liouville Theory and its Applications

Springer

Springer Undergraduate Mathematics Series

Springer Undergraduate Mathematics Series

## Advisory Board



## Other books in this series

M.A. Al-Gwaiz

# Sturm-Liouville Theory and its Applications

Springer

M.A. Al-Gwaiz
Department of Mathematics
King Saud University
Riyadh, Saudi Arabia
malgwaiz@ksu.edu.sa

*Cover illustration elements reproduced by kind permission of:*
Aptech Systems, Inc., Publishers of the GAUSS Mathematical and Statistical System, 23804 S.E. Kent-Kangley Road, Maple Valley, WA 98038, USA. Tel: (206) 432 - 7855 Fax (206) 432 - 7832 email: info@aptech.com URL: www.aptech.com.
American Statistical Association: Chance Vol 8 No 1, 1995 article by KS and KW Heiner 'Tree Rings of the Northern Shawangunks' page 32 fig 2.
Springer-Verlag: Mathematica in Education and Research Vol 4 Issue 3 1995 article by Roman E Maeder, Beatrice Amrhein and Oliver Gloor 'Illustrated Mathematics: Visualization of Mathematical Objects' page 9 fig 11, originally published as a CD ROM 'Illustrated Mathematics' by TELOS: ISBN 0-387-14222-3, German edition by Birkhauser: ISBN 3-7643-5100-4.
Mathematica in Education and Research Vol 4 Issue 3 1995 article by Richard J Gaylord and Kazume Nishidate 'Traffic Engineering with Cellular Automata' page 35 fig 2. Mathematica in Education and Research Vol 5 Issue 2 1996 article by Michael Trott 'The Implicitization of a Trefoil Knot' page 14.
Mathematica in Education and Research Vol 5 Issue 2 1996 article by Lee de Cola 'Coins, Trees, Bars and Bells: Simulation of the Binomial Process' page 19 fig 3. Mathematica in Education and Research Vol 5 Issue 2 1996 article by Richard Gaylord and Kazume Nishidate 'Contagious Spreading' page 33 fig 1. Mathematica in Education and Research Vol 5 Issue 2 1996 article by Joe Buhler and Stan Wagon 'Secrets of the Madelung Constant' page 50 fig 1.

Mathematics Subject Classification (2000): 34B24, 34L10

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2007938910

Springer Undergraduate Mathematics Series ISSN 1615-2085
ISBN 978-1-84628-971-2 e-ISBN 978-1-84628-972-9



Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

# Preface

This book is based on lecture notes which I have used over a number of years to teach a course on mathematical methods to senior undergraduate students of mathematics at King Saud University. The course is offered here as a prerequisite for taking partial differential equations in the final (fourth) year of the undergraduate program. It was initially designed to cover three main topics: special functions, Fourier series and integrals, and a brief sketch of the Sturm–Liouville problem and its solutions. Using separation of variables to solve a boundary-value problem for a second-order partial differential equation often leads to a Sturm–Liouville eigenvalue problem, and the solution set is likely to be a sequence of special functions, hence the relevance of these topics. Typically, the solution of the partial differential equation can then be represented (pointwise) by a Fourier series or a Fourier integral, depending on whether the domain is finite or infinite.

But it soon became clear that these "mathematical methods" could be developed into a more coherent and substantial course by presenting them within the more general Sturm–Liouville theory in $\mathcal{L}^2$. According to this theory, a linear second-order differential operator which is self-adjoint has an orthogonal sequence of eigenfunctions that spans $\mathcal{L}^2$. This immediately leads to the fundamental theorem of Fourier series in $\mathcal{L}^2$ as a special case in which the operator is simply $d^2/dx^2$. The other orthogonal functions of mathematical physics, such as the Legendre and Hermite polynomials or the Bessel functions, are similarly generated as eigenfunctions of particular differential operators. The result is a generalized version of the classical theory of Fourier series, which ties up the topics of the course mentioned above and provides a common theme for the book.

In Chapter 1 the stage is set by defining the inner product space of square integrable functions $\mathcal{L}^2$, and the basic analytical tools needed in the chapters to follow. These include the convergence properties of sequences and series of functions and the important notion of completeness of $\mathcal{L}^2$, which is defined through Cauchy sequences.

The difficulty with building Fourier analysis on the Sturm–Liouville theory is that the latter is deeply rooted in functional analysis, in particular the spectral theory of compact operators, which is beyond the scope of an undergraduate treatment such as this. We need a simpler proof of the existence and completeness of the eigenfunctions. In the case of the regular Sturm–Liouville problem, this is achieved in Chapter 2 by invoking the existence theorem for linear differential equations to construct Green's function for the Sturm–Liouville operator, and then using the Ascoli–Arzela theorem to arrive at the desired conclusions. This is covered in Sections 2.4.1 and 2.4.2 along the lines of Coddington and Levinson in [6].

Chapters 3 through 5 present special applications of the Sturm–Liouville theory. Chapter 3, which is on Fourier series, provides the prime example of a regular Sturm–Liouville problem. In this chapter the pointwise theory of Fourier series is also covered, and the classical theorem (Theorem 3.9) in this context is proved. The advantage of the $\mathcal{L}^2$ theory is already evident from the simple statement of Theorem 3.2, that a function can be represented by a Fourier series if and only if it lies in $\mathcal{L}^2$, as compared to the statement of Theorem 3.9.

In Chapters 4 and 5 we discuss some of the more important examples of a singular Sturm–Liouville problem. These lead to the orthogonal polynomials and Bessel functions which are familiar to students of science and engineering. Each chapter concludes with applications to some well-known equations of mathematical physics, including Laplace's equation, the heat equation, and the wave equation.

Chapters 6 and 7 on the Fourier and Laplace transformations are not really part of the Sturm–Liouville theory, but are included here as extensions of the Fourier series method for representing functions. These have important applications in heat transfer and signal transmission. They also allow us to solve nonhomogeneous differential equations, a subject which is not discussed in the previous chapters where the emphasis is mainly on the eigenfunctions.

The reader is assumed to be familiar with the convergence properties of sequences and series of functions, which are usually presented in advanced calculus, and with elementary ordinary differential equations. In addition, we have used some standard results of real analysis, such as the density of continuous functions in $\mathcal{L}^2$ and the Ascoli–Arzela theorem. These are used to prove the existence of eigenfunctions for the Sturm–Liouville operator in Chapter 2, and they have the advantage of avoiding any need for Lebesgue measure and integration. It is for that reason that smoothness conditions are imposed on the coefficients of the Sturm–Liouville operator, for otherwise integrability conditions would have sufficed. The only exception is the dominated convergence theorem, which is invoked in Chapter 6 to establish the continuity of the Fourier transform. This is a marginal result which lies outside the context of the Sturm–Liouville theory and could have been handled differently, but the temptation to use that powerful theorem as a shortcut was irresistible.

This book follows a strict mathematical style of presentation, but the subject is important for students of science and engineering. In these disciplines, Fourier analysis and special functions are used quite extensively for solving linear differential equations, but it is only through the Sturm–Liouville theory in $\mathcal{L}^2$ that one discovers the underlying principles which clarify why the procedure works. The theoretical treatment in Chapter 2 need not hinder students outside mathematics who may have some difficulty with the analysis. Proof of the existence and completeness of the eigenfunctions (Sections 2.4.1 and 2.4.2) may be skipped by those who are mainly interested in the results of the theory. But the operator-theoretic approach to differential equations in Hilbert space has proved extremely convenient and fruitful in quantum mechanics, where it is introduced at the undergraduate level, and it should not be avoided where it seems to brings clarity and coherence in other disciplines.

I have occasionally used the symbols $\Rightarrow$ (for "implies") and $\Leftrightarrow$ (for "if and only if") to connect mathematical statements. This is done mainly for the sake of typographical convenience and economy of expression, especially where displayed relations are involved.

A first draft of this book was written in the summer of 2005 while I was on vacation in Lebanon. I should like to thank the librarian of the American University of Beirut for allowing me to use the facilities of their library during my stay there. A number of colleagues in our department were kind enough to check the manuscript for errors and misprints, and to comment on parts of it. I am grateful to them all. Professor Saleh Elsanousi prepared the figures for the book, and my former student Mohammed Balfageh helped me to set up the software used in the SUMS Springer series. I would not have been able to complete these tasks without their help. Finally, I wish to express my deep appreciation to Karen Borthwick at Springer-Verlag for her gracious handling of all the communications leading to publication.

M.A. Al-Gwaiz
Riyadh, March 2007

# Contents

# 1
# Inner Product Space

An inner product space is the natural generalization of the Euclidean space $\mathbb{R}^n$, with its well-known topological and geometric properties. It constitutes the framework, or setting, for much of our work in this book, as it provides the appropriate mathematical structure that we need.

## 1.1 Vector Space

We use the symbol $\mathbb{F}$ to denote either the real number field $\mathbb{R}$ or the complex number field $\mathbb{C}$.

### Definition 1.1

A *linear vector space*, or simply a *vector space*, over $\mathbb{F}$ is a set $X$ on which two operations, *addition*

$$+ : X \times X \to X,$$

and *scalar multiplication*

$$\cdot : \mathbb{F} \times X \to X,$$

are defined such that:

1. $X$ is a commutative group under addition; that is,

   (a) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for all $\mathbf{x}, \mathbf{y} \in X$.

   (b) $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X$.

(c) There is a *zero*, or *null*, *element* $\mathbf{0} \in X$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x} \in X$.

(d) For each $\mathbf{x} \in X$ there is an *additive inverse* $-\mathbf{x} \in X$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.

2. Scalar multiplication between the elements of $\mathbb{F}$ and $X$ satisfies

(a) $a \cdot (b \cdot \mathbf{x}) = (ab) \cdot \mathbf{x}$ for all $a, b \in \mathbb{F}$ and all $\mathbf{x} \in X$,

(b) $1 \cdot \mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in X$.

3. The two distributive properties

(a) $a \cdot (\mathbf{x} + \mathbf{y}) = a \cdot \mathbf{x} + a \cdot \mathbf{y}$

(b) $(a + b) \cdot \mathbf{x} = a \cdot \mathbf{x} + b \cdot \mathbf{x}$

hold for any $a, b \in \mathbb{F}$ and $\mathbf{x}, \mathbf{y} \in X$.

$X$ is called a *real vector space* or a *complex vector space* depending on whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. The elements of $X$ are called *vectors* and those of $\mathbb{F}$ *scalars.*

From these properties it can be shown that the zero vector $\mathbf{0}$ is unique, and that every $\mathbf{x} \in X$ has a unique inverse $-\mathbf{x}$. Furthermore, it follows that $0 \cdot \mathbf{x} = \mathbf{0}$ and $(-1) \cdot \mathbf{x} = -\mathbf{x}$ for every $\mathbf{x} \in X$, and that $a \cdot \mathbf{0} = \mathbf{0}$ for every $a \in \mathbb{F}$. As usual, we often drop the multiplication dot in $a \cdot \mathbf{x}$ and write $a\mathbf{x}$.

## Example 1.2

(i) The set of $n$-tuples of real numbers

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_i \in \mathbb{R}\},$$

under addition, defined by

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$

and scalar multiplication, defined by

$$a \cdot (x_1, \ldots, x_n) = (ax_1, \ldots, ax_n),$$

where $a \in \mathbb{R}$, is a real vector space.

(ii) The set of $n$-tuples of complex numbers

$$\mathbb{C}^n = \{(z_1, \ldots, z_n) : z_i \in \mathbb{C}\},$$

on the other hand, under the operations

$$(z_1, \dots, z_n) + (w_1, \dots, w_n) = (z_1 + w_1, \dots, z_n + w_n),$$
$$a \cdot (z_1, \dots, z_n) = (az_1, \dots, az_n), \quad a \in \mathbb{C},$$

is a complex vector space.

(iii) The set $\mathbb{C}^n$ over the field $\mathbb{R}$ is a real vector space.

(iv) Let $I$ be a real interval which may be closed, open, half-open, finite, or infinite. $\mathcal{P}(I)$ denotes the set of polynomials on $I$ with real (complex) coefficients. This becomes a real (complex) vector space under the usual operation of addition of polynomials, and scalar multiplication

$$b \cdot (a_n x^n + \cdots + a_1 x + a_0) = ba_n x^n + \cdots + ba_1 x + ba_0,$$

where $b$ is a real (complex) number. We also abbreviate $\mathcal{P}(\mathbb{R})$ as $\mathcal{P}$.

(v) The set of real (complex) continuous functions on the real interval $I$, which is denoted $C(I)$, is a real (complex) vector space under the usual operations of addition of functions and multiplication of a function by a real (complex) number.

Let $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be any finite set of vectors in a vector space $X$. The sum

$$a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n = \sum_{i=1}^n a_i\mathbf{x}_i, \quad a_i \in \mathbb{F},$$

is called a *linear combination* of the vectors in the set, and the scalars $a_i$ are the *coefficients* in the linear combination.

## Definition 1.3

(i) A finite set of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is said to be *linearly independent* if

$$\sum_{i=1}^n a_i\mathbf{x}_i = \mathbf{0} \Rightarrow a_i = 0 \quad \text{for all } i \in \{1, \dots, n\},$$

that is, if every linear combination of the vectors is not equal to zero except when all the coefficients are zeros. The set $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is *linearly dependent* if it is not linearly independent, that is, if there is a collection of coefficients $a_1, \dots, a_n$, not all zeros, such that $\sum_{i=1}^n a_i\mathbf{x}_i = \mathbf{0}$.

(ii) An infinite set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots\}$ is *linearly independent* if every finite subset of the set is linearly independent. It is *linearly dependent* if it is not linearly independent, that is, if there is a finite subset of $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots\}$ which is linearly dependent.

It should be noted at this point that a finite set of vectors is linearly dependent if, and only if, one of the vectors can be represented as a linear combination of the others (see Exercise 1.3).

## Definition 1.4

Let $X$ be a vector space.

(i) A set $\mathcal{A}$ of vectors in $X$ is said to *span* $X$ if every vector in $X$ can be expressed as a linear combination of elements of $\mathcal{A}$. If, in addition, the vectors in $\mathcal{A}$ are linearly independent, then $\mathcal{A}$ is called a *basis* of $X$.

(ii) A subset $Y$ of $X$ is called a *subspace* of $X$ if every linear combination of vectors in $Y$ lies in $Y$. This is equivalent to saying that $Y$ is a vector space in its own right (over the same scalar field as $X$).

If $X$ has a finite basis then any other basis of $X$ is also finite, and both bases have the same number of elements (Exercise 1.4). This number is called the *dimension* of $X$ and is denoted $\dim X$. If the basis is infinite, we take $\dim X = \infty$.

In Example 1.2, the vectors

$$\begin{aligned}\mathbf{e}_1 &= (1, 0, \ldots, 0),\\ \mathbf{e}_2 &= (0, 1, 0, \ldots, 0),\\ &\vdots\\ \mathbf{e}_n &= (0, \ldots, 0, 1)\end{aligned}$$

form a basis for $\mathbb{R}^n$ over $\mathbb{R}$ and $\mathbb{C}^n$ over $\mathbb{C}$. The vectors

$$\begin{aligned}\mathbf{d}_1 &= (i, 0, \ldots, 0),\\ \mathbf{d}_2 &= (0, i, 0, \ldots, 0),\\ &\vdots\\ \mathbf{d}_n &= (0, \ldots, 0, i),\end{aligned}$$

together with $\mathbf{e}_1, \ldots, \mathbf{e}_n$, form a basis of $\mathbb{C}^n$ over $\mathbb{R}$. On the other hand, the powers of $x \in \mathbb{R}$,

$$1,\ x,\ x^2,\ x^3, \ldots,$$

span $\mathcal{P}$ and, being linearly independent (Exercise 1.5), they form a basis for the space of real (complex) polynomials over $\mathbb{R}$ ($\mathbb{C}$). Thus both real $\mathbb{R}^n$ and complex $\mathbb{C}^n$ have dimension $n$, whereas real $\mathbb{C}^n$ has dimension $2n$. The space of polynomials, on the other hand, has infinite dimension. So does the space of continuous functions $C(I)$, as it includes all the polynomials on $I$ (Exercise 1.6).

Let $\mathcal{P}_n(I)$ be the vector space of polynomials on the interval $I$ of degree $\leq n$. This is clearly a subspace of $\mathcal{P}(I)$ of dimension $n+1$. Similarly, if we denote the set of (real or complex) functions on $I$ whose first derivatives are continuous by $C^1(I)$, then, under the usual operations of addition of functions and multiplication by scalars, $C^1(I)$ is a vector subspace of $C(I)$ over the same (real or complex) field. As usual, when $I$ is closed at one (or both) of its endpoints, the derivative at that endpoint is the one-sided derivative. More generally, by defining

$$C^n(I) = \{f \in C(I) : f^{(n)} \in C(I),\ n \in \mathbb{N}\},$$
$$C^\infty(I) = \bigcap_{n=1}^{\infty} C^n(I),$$

we obtain a sequence of vector spaces

$$C(I) \supset C^1(I) \supset C^2(I) \supset \cdots \supset C^\infty(I)$$

such that $C^k(I)$ is a (proper) vector subspace of $C^m(I)$ whenever $k > m$. Here $\mathbb{N}$ is the set of natural numbers $\{1, 2, 3, \ldots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. The set of integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$ is denoted $\mathbb{Z}$. If we identify $C^0(I)$ with $C(I)$, all the spaces $C^n(I)$, $n \in \mathbb{N}_0$, have infinite dimensions as each includes the polynomials $\mathcal{P}(I)$. When $I = \mathbb{R}$, or when $I$ is not relevant, we simply write $C^n$.

## *EXERCISES*

1.1 Use the properties of the vector space $X$ over $\mathbb{F}$ to prove the following.

(a) $0 \cdot \mathbf{x} = \mathbf{0}$ for all $\mathbf{x} \in X$.

(b) $a \cdot \mathbf{0} = \mathbf{0}$ for all $a \in \mathbb{F}$.

(c) $(-1) \cdot \mathbf{x} = -\mathbf{x}$ for all $\mathbf{x} \in X$.

(d) If $a \cdot \mathbf{x} = \mathbf{0}$ then either $a = 0$ or $\mathbf{x} = \mathbf{0}$.

1.2 Determine which of the following sets is a vector space under the usual operations of addition and scalar multiplication, and whether it is a real or a complex vector space.

(a) $\mathcal{P}_n(I)$ with complex coefficients over $\mathbb{C}$

(b) $\mathcal{P}(I)$ with imaginary coefficients over $\mathbb{R}$

(c) The set of real numbers over $\mathbb{C}$

(d) The set of complex functions of class $C^n(I)$ over $\mathbb{R}$

1.3 Prove that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly dependent if, and only if, there is an integer $k \in \{1, \ldots, n\}$ such that

$$\mathbf{x}_k = \sum_{i \neq k} a_i \mathbf{x}_i, \qquad a_i \in \mathbb{F}.$$

Conclude from this that any set of vectors, whether finite or infinite, is linearly dependent if, and only if, one of its vectors is a finite linear combination of the other vectors.

1.4 Let $X$ be a vector space. Prove that, if $\mathcal{A}$ and $\mathcal{B}$ are bases of $X$ and one of them is finite, then so is the other and they have the same number of elements.

1.5 Show that any finite set of powers of $x$, $\{1, x, x^2, \ldots, x^n : x \in I\}$, is linearly independent. Hence conclude that the infinite set $\{1, x, x^2, \ldots : x \in I\}$ is linearly independent.

1.6 If $Y$ is a subspace of the vector space $X$, prove that $\dim Y \leq \dim X$.

1.7 Prove that the vectors

$$\begin{aligned} \mathbf{x}_1 &= (x_{11}, \ldots, x_{1n}), \\ &\vdots \\ \mathbf{x}_n &= (x_{n1}, \ldots, x_{nn}), \end{aligned}$$

where $x_{ij} \in \mathbb{R}$, are linearly dependent if, and only if, $\det(x_{ij}) = 0$, where $\det(x_{ij})$ is the determinant of the matrix $(x_{ij})$.

## 1.2 Inner Product Space

### Definition 1.5

Let $X$ be a vector space over $\mathbb{F}$. A function from $X \times X$ to $\mathbb{F}$ is called an *inner product* in $X$ if, for any pair of vectors $\mathbf{x}, \mathbf{y} \in X$, the inner product $(\mathbf{x}, \mathbf{y}) \mapsto \langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{F}$ satisfies the following conditions.

(i) $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ for all $\mathbf{x}, \mathbf{y} \in X$.

(ii) $\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a \langle \mathbf{x}, \mathbf{z} \rangle + b \langle \mathbf{y}, \mathbf{z} \rangle$ for all $a, b \in \mathbb{F}$, $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X$.

(iii) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ for all $\mathbf{x} \in X$.

(iv) $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$.

A vector space on which an inner product is defined is called an *inner product space*.

The symbol $\overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ in (i) denotes the complex conjugate of $\langle \mathbf{y}, \mathbf{x} \rangle$, so that $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ if $X$ is a real vector space. Note also that (i) and (ii) imply

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

which means that the linearity property which holds in the first component of the inner product, as expressed by (ii), does not apply to the second component unless $\mathbb{F} = \mathbb{R}$.

## Theorem 1.6 (Cauchy–Bunyakowsky–Schwarz Inequality)

If $X$ is an inner product space, then

$$|\langle \mathbf{x}, \mathbf{y} \rangle|^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle \quad \text{for all } \mathbf{x}, \mathbf{y} \in X.$$

## Proof

If either $\mathbf{x} = \mathbf{0}$ or $\mathbf{y} = \mathbf{0}$ this inequality clearly holds, so we need only consider the case where $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$. Furthermore, neither side of the inequality is affected if we replace $\mathbf{x}$ by $a\mathbf{x}$ where $|a| = 1$. Choose $a$ so that $\langle a\mathbf{x}, \mathbf{y} \rangle$ is a real number; that is, if $\langle \mathbf{x}, \mathbf{y} \rangle = |\langle \mathbf{x}, \mathbf{y} \rangle| e^{i\theta}$, let $a = e^{-i\theta}$. Therefore we may assume, without loss of generality, that $\langle \mathbf{x}, \mathbf{y} \rangle$ is a real number. Using the above properties of the inner product, we have, for any real number $t$,

$$0 \leq \langle \mathbf{x}+t\mathbf{y}, \mathbf{x}+t\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + 2 \langle \mathbf{x}, \mathbf{y} \rangle t + \langle \mathbf{y}, \mathbf{y} \rangle t^2. \tag{1.1}$$

This is a real quadratic expression in $t$ which achieves its minimum at $t = -\langle \mathbf{x}, \mathbf{y} \rangle / \langle \mathbf{y}, \mathbf{y} \rangle$. Substituting this value for $t$ into (1.1) gives

$$0 \leq \langle \mathbf{x}, \mathbf{x} \rangle - \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\langle \mathbf{y}, \mathbf{y} \rangle},$$

and hence the desired inequality. □

We now define the *norm* of the vector $\mathbf{x}$ as

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

Hence, in view of (iii) and (iv), $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in X$, and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$. The Cauchy–Bunyakowsky–Schwarz inequality, which we henceforth refer to as the CBS inequality, then takes the form

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\| \quad \text{for all } \mathbf{x}, \mathbf{y} \in X. \tag{1.2}$$

## Corollary 1.7

If $X$ is an inner product space, then

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \quad \text{for all } \mathbf{x}, \mathbf{y} \in X. \tag{1.3}$$

## Proof

By definition of the norm,

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

But $\mathrm{Re}\,\langle \mathbf{x}, \mathbf{y} \rangle \leq |\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|\,\|\mathbf{y}\|$ by the CBS inequality, hence

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 &\leq \|\mathbf{x}\|^2 + 2\,\|\mathbf{x}\|\,\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2. \end{aligned}$$

Inequality (1.3) now follows by taking the square roots of both sides. □

By defining the *distance* between the vectors $\mathbf{x}$ and $\mathbf{y}$ to be $\|\mathbf{x}-\mathbf{y}\|$, we see that for any three vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X$,

$$\begin{aligned} \|\mathbf{x}-\mathbf{y}\| &= \|\mathbf{x}-\mathbf{z}+\mathbf{z}-\mathbf{y}\| \\ &\leq \|\mathbf{x}-\mathbf{z}\| + \|\mathbf{z}-\mathbf{y}\| . \end{aligned}$$

This inequality, and by extension (1.3), is called the *triangle inequality,* as it generalizes a well known inequality between the sides of a triangle in the plane whose vertices are the points $\mathbf{x}, \mathbf{y}, \mathbf{z}$. The inner product space $X$ is now a *topological space*, in which the topology is defined by the norm $\|\cdot\|$, which is derived from the inner product $\langle \cdot, \cdot \rangle$.

## Example 1.8

(a) In $\mathbb{R}^n$ we define the inner product of the vectors

$$\mathbf{x} = (x_1, \ldots, x_n), \quad \mathbf{y} = (y_1, \ldots, y_n)$$

by

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \ldots + x_n y_n, \tag{1.4}$$

which implies

$$\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

In this topology the vector space $\mathbb{R}^n$ is the familiar $n$-dimensional Euclidean space. Note that there are other choices for defining the inner product $\langle \mathbf{x}, \mathbf{y} \rangle$, such as $c(x_1y_1 + \cdots + x_ny_n)$ where $c$ is any positive number, or $c_1x_1y_1 + \cdots + c_nx_ny_n$ where $c_i > 0$ for every $i$. In either case the provisions of Definition 1.5 are all satisfied, but the resulting inner product space would not in general be Euclidean.

(b) In $\mathbb{C}^n$ we define

$$\langle \mathbf{z}, \mathbf{w} \rangle = z_1\bar{w}_1 + \cdots + z_n\bar{w}_n \tag{1.5}$$

for any pair $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$. Consequently,

$$\|\mathbf{z}\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

(c) A natural choice for the definition of an inner product on $C([a,b])$, by analogy with (1.5), is

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx, \qquad f, g \in C([a,b]), \tag{1.6}$$

so that

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

It is a simple matter to verify that the properties (i) through (iv) of the inner product are satisfied in each case, provided of course that $\mathbb{F}= \mathbb{C}$ when the vector space is $\mathbb{C}^n$ or complex $C([a,b])$. To check (iv) in Example 1.8(c), we have to show that

$$\left[\int_a^b |f(x)|^2 \, dx\right]^{1/2} = 0 \quad \Leftrightarrow \quad f(x) = 0 \quad \text{for all } x \in [a,b].$$

We need only verify the forward implication ($\Rightarrow$), as the backward implication ($\Leftarrow$) is trivial. But this follows from a well-known property of continuous, nonnegative functions: If $\varphi$ is continuous on $[a,b]$, $\varphi \geq 0$, and $\int_a^b \varphi(x)dx = 0$, then $\varphi = 0$ (see [1], for example). Because $|f|^2$ is continuous and nonnegative on $[a,b]$ for any $f \in C([a,b])$,

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

In this study, we are mainly concerned with function spaces on which an inner product of the type (1.6) is defined. In addition to the topological structure which derives from the norm $\|\cdot\|$, this inner product endows the space with a geometrical structure that extends some desirable notions, such as orthogonality, from Euclidean space to infinite-dimensional spaces. This is taken up in Section 1.3. Here we examine the Euclidean space $\mathbb{R}^n$ more closely.

Although we proved the CBS and the triangle inequalities for any inner product in Theorem 1.6 and its corollary, we can also derive these inequalities directly in $\mathbb{R}^n$. Consider the inequality

$$(a-b)^2 = a^2 - 2ab + b^2 \geq 0 \tag{1.7}$$

which holds for any pair of real numbers $a$ and $b$. Let

$$a = \frac{x_i}{\sqrt{x_1^2 + \cdots + x_n^2}}, \quad b = \frac{y_i}{\sqrt{y_1^2 + \cdots + y_n^2}}, \quad x_i, y_i \in \mathbb{R}.$$

If $\sum_{j=1}^n x_j^2 \neq 0$ and $\sum_{j=1}^n y_j^2 \neq 0$, then (1.7) implies

$$\frac{x_i y_i}{\sqrt{\sum x_j^2}\sqrt{\sum y_j^2}} \leq \frac{1}{2}\frac{x_i^2}{\sum x_j^2} + \frac{1}{2}\frac{y_i^2}{\sum y_j^2},$$

where the summation over the index $j$ is from 1 to $n$. After summing on $i$ from 1 to $n$, the right-hand side of this inequality reduces to 1, and we obtain

$$\sum x_i y_i \leq \sqrt{\sum x_i^2}\sqrt{\sum y_i^2}.$$

This inequality remains valid regardless of the signs of $x_i$ and $y_i$, therefore we can write

$$\left|\sum x_i y_i\right| \leq \sqrt{\sum x_i^2}\sqrt{\sum y_i^2}$$

for all $\mathbf{x} = (x_1, \ldots, x_n) \neq \mathbf{0}$ and $\mathbf{y} = (y_1, \ldots, y_n) \neq \mathbf{0}$ in $\mathbb{R}^n$. But because the inequality becomes an equality if either $\|\mathbf{x}\|$ or $\|\mathbf{y}\|$ is 0, this proves the CBS inequality

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \, \|\mathbf{y}\| \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

From this the triangle inequality $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ immediately follows.

Now we define the angle $\theta \in [0, \pi]$ between any pair of nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ by the equation

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \, \|\mathbf{y}\| \cos\theta.$$

Because the function $\cos : [0, \pi] \to [-1, 1]$ is injective, this defines the angle $\theta$ uniquely and agrees with the usual definition of the angle between $\mathbf{x}$ and $\mathbf{y}$ in both $\mathbb{R}^2$ and $\mathbb{R}^3$. With $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$,

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0 \quad \Leftrightarrow \quad \cos\theta = 0,$$

which is the condition for the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ to be orthogonal. Consequently, we adopt the following definition.

## Definition 1.9

(i) A pair of nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ in the inner product space $X$ is said to be *orthogonal* if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, symbolically expressed by writing $\mathbf{x} \perp \mathbf{y}$. A set of nonzero vectors $\mathcal{V}$ in $X$ is *orthogonal* if every pair in $\mathcal{V}$ is orthogonal.

(ii) An orthogonal set $\mathcal{V} \subseteq X$ is said to be *orthonormal* if $\|\mathbf{x}\| = 1$ for every $\mathbf{x} \in \mathcal{V}$.

A typical example of an orthonormal set in the Euclidean space $\mathbb{R}^n$ is given by

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, \ldots, 0), \\ \mathbf{e}_2 &= (0, 1, \ldots, 0), \\ &\vdots \\ \mathbf{e}_n &= (0, \ldots, 0, 1), \end{aligned}$$

which, as we have already seen, forms a basis of $\mathbb{R}^n$.

In general, if the vectors

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \tag{1.8}$$

in the inner product space $X$ are orthogonal, then they are necessarily linearly independent. To see that, let

$$a_1 \mathbf{x}_1 + \cdots + a_n \mathbf{x}_n = \mathbf{0}, \quad a_i \in \mathbb{F},$$

and take the inner product of each side of this equation with $\mathbf{x}_k$, $1 \leq k \leq n$. In as much as $\langle \mathbf{x}_i, \mathbf{x}_k \rangle = 0$ whenever $i \neq k$, we obtain

$$\begin{aligned} a_k \langle \mathbf{x}_k, \mathbf{x}_k \rangle = a_k \|\mathbf{x}_k\|^2 = 0, \quad k \in \{1, \cdots, n\} \\ \Rightarrow \ a_k = 0 \quad \text{for all } k. \end{aligned}$$

By dividing each vector in (1.8) by its norm, we obtain the orthonormal set $\{\mathbf{x}_i / \|\mathbf{x}_i\| : 1 \leq i \leq n\}$.

Let us go back to the Euclidean space $\mathbb{R}^n$ and assume that $\mathbf{x}$ is any vector in $\mathbb{R}^n$. We can therefore represent it in the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ by

$$\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{e}_i. \tag{1.9}$$

Taking the inner product of Equation (1.9) with $\mathbf{e}_k$, and using the orthonormal property of $\{\mathbf{e}_i\}$,

$$\langle \mathbf{x}, \mathbf{e}_k \rangle = a_k, \quad k \in \{1, \ldots, n\}.$$

This determines the coefficients $a_i$ in (1.9), and means that any vector $\mathbf{x}$ in $\mathbb{R}^n$ is represented by the formula

$$\mathbf{x} = \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{e}_i \rangle \mathbf{e}_i.$$

The number $\langle \mathbf{x}, \mathbf{e}_i \rangle$ is called the *projection* of $\mathbf{x}$ on $\mathbf{e}_i$, and $\langle \mathbf{x}, \mathbf{e}_i \rangle \mathbf{e}_i$ is the *projection vector* in the direction of $\mathbf{e}_i$. More generally, if $\mathbf{x}$ and $\mathbf{y} \neq \mathbf{0}$ are any vectors in the inner product space $X$, then $\langle \mathbf{x}, \mathbf{y}/\left\|\mathbf{y}\right\| \rangle$ is the projection of $\mathbf{x}$ on $\mathbf{y}$, and the vector

$$\left\langle \mathbf{x}, \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\rangle \frac{\mathbf{y}}{\|\mathbf{y}\|} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{y}\|^2} \mathbf{y}$$

is its projection vector along $\mathbf{y}$.

Suppose now that we have a linearly independent set of vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in the inner product space $X$. Can we form an orthogonal set out of this set? In what follows we present the so-called Gram–Schmidt method for constructing an orthogonal set $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ out of $\{\mathbf{x}_i\}$ having the same number of vectors: We first choose

$$\mathbf{y}_1 = \mathbf{x}_1.$$

The second vector is obtained from $\mathbf{x}_2$ after extracting the projection vector of $\mathbf{x}_2$ in the direction of $\mathbf{y}_1$,

$$\mathbf{y}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{y}_1 \rangle}{\|\mathbf{y}_1\|^2} \mathbf{y}_1.$$

The third vector is $\mathbf{x}_3$ minus the projections of $\mathbf{x}_3$ in the directions of $\mathbf{y}_1$ and $\mathbf{y}_2$,

$$\mathbf{y}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{y}_1 \rangle}{\|\mathbf{y}_1\|^2} \mathbf{y}_1 - \frac{\langle \mathbf{x}_3, \mathbf{y}_2 \rangle}{\|\mathbf{y}_2\|^2} \mathbf{y}_2.$$

We continue in this fashion until the last vector

$$\mathbf{y}_n = \mathbf{x}_n - \frac{\langle \mathbf{x}_n, \mathbf{y}_1 \rangle}{\|\mathbf{y}_1\|^2} \mathbf{y}_1 - \cdots - \frac{\langle \mathbf{x}_n, \mathbf{y}_{n-1} \rangle}{\|\mathbf{y}_{n-1}\|^2} \mathbf{y}_{n-1},$$

and the reader can verify that the set $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ is orthogonal.

## EXERCISES

1.8 Given two vectors $\mathbf{x}$ and $\mathbf{y}$ in an inner product space, under what conditions does the equality $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ hold? Can this equation hold even if the vectors are not orthogonal?

1.9 Let $\mathbf{x}, \mathbf{y} \in X$, where $X$ is an inner product space.

(a) If the vectors $\mathbf{x}$ and $\mathbf{y}$ are linearly independent, prove that $\mathbf{x}+\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ are also linearly independent.

(b) If $\mathbf{x}$ and $\mathbf{y}$ are orthogonal and nonzero, when are $\mathbf{x}+\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ orthogonal?

1.10 Let $\varphi_1(x) = 1,\ \varphi_2(x) = x,\ \varphi_3(x) = x^2,\ -1 \le x \le 1$. Use (1.6) to calculate

(a) $\langle \varphi_1, \varphi_2 \rangle$

(b) $\langle \varphi_1, \varphi_3 \rangle$

(c) $\|\varphi_1 - \varphi_2\|^2$

(d) $\|2\varphi_1 + 3\varphi_2\|$.

1.11 Determine all orthogonal pairs on $[0,1]$ among the functions $\varphi_1(x) = 1,\ \varphi_2(x) = x,\ \varphi_3(x) = \sin 2\pi x,\ \varphi_4(x) = \cos 2\pi x$. What is the largest orthogonal subset of $\{\varphi_1, \varphi_2, \varphi_3, \varphi_4\}$?

1.12 Determine the projection of $f(x) = \cos^2 x$ on each of the functions $f_1(x) = 1,\ f_2(x) = \cos x,\ f_3(x) = \cos 2x,\ -\pi \le x \le \pi$.

1.13 Verify that the functions $\varphi_1, \varphi_2, \varphi_3$ in Exercise 1.10 are linearly independent, and use the Gram–Schmidt method to construct a corresponding orthogonal set.

1.14 Prove that the set of functions $\{1, x, |x|\}$ is linearly independent on $[-1,1]$, and construct a corresponding orthonormal set. Is the given set linearly independent on $[0,1]$?

1.15 Use the result of Exercise 1.3 and the properties of determinants to prove that any set of functions $\{f_1, \ldots, f_n\}$ in $C^{n-1}(I)$, $I$ being a real interval, is linearly dependent if, and only if, $\det(f_i^{(j)}) = 0$ on $I$, where $1 \le i \le n, 0 \le j \le n-1$.

1.16 Verify that the following functions are orthogonal on $[-1,1]$.

$$\varphi_1(x) = 1, \quad \varphi_2(x) = x^2 - \frac{1}{3}, \quad \varphi_3(x) = \begin{cases} x/|x|, & x \ne 0 \\ 0, & x = 0. \end{cases}$$

Determine the corresponding orthonormal set.

1.17 Determine the values of the coefficients $a$ and $b$ which make the function $x^2 + ax + b$ orthogonal to both $x+1$ and $x-1$ on $[0,1]$.

1.18 Using the definition of the inner product as expressed by Equation (1.6), show that $\|f\| = 0$ does not necessarily imply that $f = 0$ unless $f$ is continuous.

## 1.3 The Space $\mathcal{L}^2$

For any two functions $f$ and $g$ in the vector space $C([a,b])$ of complex continuous functions on a real interval $[a,b]$, we defined the inner product

$$\langle f,g\rangle=\int_a^b f(x)\overline{g(x)}dx, \tag{1.10}$$

from which followed the definition of the norm

$$\|f\|=\sqrt{\langle f,f\rangle}=\sqrt{\int_a^b |f(x)|^2\,dx}. \tag{1.11}$$

As in $\mathbb{R}^n$, we can also show directly that the CBS inequality holds in $C([a,b])$. For any $f,g\in C([a,b])$, we have

$$\left\|\frac{|f|}{\|f\|}-\frac{|g|}{\|g\|}\right\|^2=\int_a^b\left[\frac{|f(x)|}{\|f\|}-\frac{|g(x)|}{\|g\|}\right]^2dx\geq 0,$$

where we assume that $\|f\|\neq 0$ and $\|g\|\neq 0$. Hence

$$\int_a^b \frac{|f(x)|}{\|f\|}\frac{|g(x)|}{\|g\|}dx\leq\frac{1}{2\|f\|^2}\int_a^b|f(x)|^2\,dx+\frac{1}{2\|g\|^2}\int_a^b|g(x)|^2\,dx=1$$
$$\Rightarrow\langle|f|,|g|\rangle\leq\|f\|\,\|g\|.$$

Using the monotonicity property of the integral

$$\left|\int_a^b\varphi(x)dx\right|\leq\int_a^b|\varphi(x)|\,dx,$$

we therefore conclude that

$$|\langle f,g\rangle|\leq\langle|f|,|g|\rangle\leq\|f\|\,\|g\|.$$

If either $\|f\|=0$ or $\|g\|=0$ the inequality remains valid, as it becomes an equality. The triangle inequality

$$\|f+g\|\leq\|f\|+\|g\|$$

then easily follows from the relation $f\bar{g}+\bar{f}g=\ 2\operatorname{Re}f\bar{g}\leq 2\,|fg|$.

As we have already observed, the nonnegative number $\|f-g\|$ may be regarded as a measure of the "distance" between the functions $f,g\in C([a,b])$. In this case we clearly have $\|f-g\|=0$ if, and only if, $f=g$ on $[a,b]$. This is the advantage of dealing with continuous functions, for if we admit discontinuous functions, such as

$$h(x)=\begin{cases}1, & x=1\\ 0, & x\in(1,2],\end{cases} \tag{1.12}$$

then $\|h\|=0$ whereas $h\neq 0$.

Nevertheless, $C([a,b])$ is not a suitable inner product space for pursuing this study, for it is not closed under limit operations as we show in the next section. That is to say, if a sequence of functions in $C([a,b])$ "converges" (in a sense to be defined in Section 1.4) its "limit" may not be in $C([a,b])$. So we need to enlarge the space of continuous functions over $[a,b]$ in order to avoid this difficulty. But in this larger space, call it $X([a,b])$, we can only admit functions for which the inner product

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

is defined for every pair $f, g \in X([a,b])$. Now the CBS inequality $|\langle f,g\rangle| \leq \|f\| \, \|g\|$ ensures that the inner product of $f$ and $g$ is well defined if $\|f\|$ and $\|g\|$ exist (i.e., if $|f|^2$ and $|g|^2$ are integrable). Strictly speaking, this is only true if the integrals are interpreted as Lebesgue integrals, for the Riemann integrability of $f^2$ and $g^2$ does not guarantee the Riemann integrability of $fg$ (see Exercise 1.21); but in this study we shall have no occasion to deal with functions which are integrable in the sense of Lebesgue but not in the sense of Riemann. For our purposes, Riemann integration, and its extension to improper integrals, is adequate. The space $X([a,b])$ which we seek should therefore be made up of functions $f$ such that $|f|^2$ is integrable on $[a,b]$.

We use the symbol $\mathcal{L}^2(a,b)$ to denote the set of functions $f : [a,b] \to \mathbb{C}$ such that

$$\int_a^b |f(x)|^2 \, dx < \infty.$$

By defining the inner product (1.10) and the norm (1.11) on $\mathcal{L}^2(a,b)$, we can use the triangle inequality to obtain

$$\begin{aligned} \|\alpha f + \beta g\| &\leq \|\alpha f\| + \|\beta g\| \\ &= |\alpha| \, \|f\| + |\beta| \, \|g\| \quad \text{for all } f, g \in \mathcal{L}^2(a,b), \quad \alpha, \beta \in \mathbb{C}, \end{aligned}$$

hence $\alpha f + \beta g \in \mathcal{L}^2(a,b)$ whenever $f, g \in \mathcal{L}^2(a,b)$. Thus $\mathcal{L}^2(a,b)$ is a linear vector space which, under the inner product (1.10), becomes an inner product space and includes $C([a,b])$ as a proper subspace.

In $\mathcal{L}^2(a,b)$ the equality $\|f\| = 0$ does not necessarily mean $f(x) = 0$ at every point $x \in [a,b]$. For example, in the case where $f(x) = 0$ on all but a finite number of points in $[a,b]$ we clearly have $\|f\| = 0$. We say that $f = 0$ *pointwise* on a real interval $I$ if $f(x) = 0$ at every $x \in I$. If $\|f\| = 0$ we say that $f = 0$ *in* $\mathcal{L}^2(I)$. Thus the function $h$ defined in (1.12) equals 0 in $\mathcal{L}^2(I)$, but not pointwise. The function 0 in $\mathcal{L}^2(I)$ really denotes an *equivalence class of functions*, each of which has norm 0. The function which is pointwise equal to 0 is only one member, indeed the only continuous member, of that class. Similarly, we say that two functions $f$ and $g$ in $\mathcal{L}^2(I)$ are *equal in* $\mathcal{L}^2(I)$ if $\|f-g\|=0$, although $f$ and $g$ may not be equal pointwise on $I$. In the terminology of measure theory, $f$ and $g$ are said to be "equal almost everywhere." Hence the space $\mathcal{L}^2(a,b)$ is, in fact, made up of equivalence classes of functions defined by equality in $\mathcal{L}^2(a,b)$, that is, functions which are equal almost everywhere.

Thus far we have used the symbol $\mathcal{L}^2(a,b)$ to denote the linear space of functions $f:[a,b]\to\mathbb{C}$ such that $\int_a^b |f(x)|^2\,dx<\infty$. But because this integral is not affected by replacing the closed interval $[a,b]$ by $[a,b)$, $(a,b]$, or $(a,b)$, $\mathcal{L}^2(a,b)$ coincides with $\mathcal{L}^2([a,b))$, $\mathcal{L}^2((a,b])$ and $\mathcal{L}^2((a,b))$. The interval $(a,b)$ need not be bounded at either or both ends, and so we have $\mathcal{L}^2(a,\infty)$, $\mathcal{L}^2(-\infty,b)$ and $\mathcal{L}^2(-\infty,\infty)=\mathcal{L}^2(\mathbb{R})$. In such cases, as in the case when the function is unbounded, we interpret the integral of $|f|^2$ on $(a,b)$ as an improper Riemann integral. Sometimes we simply write $\mathcal{L}^2$ when the underlying interval is not specified or irrelevant to the discussion.

## Example 1.10

Determine each function which belongs to $\mathcal{L}^2$ and calculate its norm.

(i) $f(x)=\begin{cases}1, & 0\le x<1/2\\ 0, & 1/2\le x\le 1.\end{cases}$

(ii) $f(x)=1/\sqrt{x},\ 0<x<1.$

(iii) $f(x)=1/\sqrt[3]{x},\ 0<x<1.$

(iv) $f(x)=1/x,\ 1<x<\infty.$

## Solution

(i)

$$\|f\|^2=\int_0^1 f^2(x)dx=\int_0^{1/2}dx=\frac{1}{2}.$$

Therefore $f\in\mathcal{L}^2(0,1)$ and $\|f\|=1/\sqrt{2}$.

(ii)

$$\|f\|^2=\int_0^1\frac{1}{x}dx=\lim_{\varepsilon\to 0^+}\int_\varepsilon^1\frac{1}{x}dx=-\lim_{\varepsilon\to 0^+}\log\varepsilon=\infty$$
$$\Rightarrow f\notin\mathcal{L}^2(0,1).$$

(iii)

$$\|f\|^2=\int_0^1\frac{1}{x^{2/3}}dx=\lim_{\varepsilon\to 0^+}3(1-\varepsilon^{1/3})=3$$
$$\Rightarrow f\in\mathcal{L}^2(0,1),\ \|f\|=\sqrt{3}.$$

(iv)

$$\|f\|^2 = \int_1^\infty \frac{1}{x^2} dx = \lim_{b\to\infty} -\left(\frac{1}{b} - 1\right) = 1$$
$$\Rightarrow f \in \mathcal{L}^2(1,\infty),\ \|f\| = 1.$$

## Example 1.11

The infinite set of functions $\{1, \cos x, \sin x, \cos 2x, \ \sin 2x, \ldots\}$ is orthogonal in the real inner product space $\mathcal{L}^2(-\pi,\pi)$. This can be seen by calculating the inner product of each pair in the set:

$$\langle 1, \cos nx\rangle = \int_{-\pi}^{\pi} \cos nx \ dx = 0, \quad n \in \mathbb{N}.$$
$$\langle 1, \sin nx\rangle = \int_{-\pi}^{\pi} \sin nx \ dx = 0, \quad n \in \mathbb{N}.$$

$$\begin{aligned}
\langle \cos nx, \cos mx\rangle &= \int_{-\pi}^{\pi} \cos nx \cos mx \ dx \\
&= \frac{1}{2}\int_{-\pi}^{\pi} [\cos(n-m)x + \cos(n+m)x] dx \\
&= \frac{1}{2}\left[\frac{1}{n-m}\sin(n-m)x + \frac{1}{n+m}\sin(n+m)x\right]\Big|_{-\pi}^{\pi} \\
&= 0, \quad n \neq m.
\end{aligned}$$

$$\begin{aligned}
\langle \sin nx, \sin mx\rangle &= \int_{-\pi}^{\pi} \sin nx \sin mx \ dx \\
&= \frac{1}{2}\int_{-\pi}^{\pi} [\cos(n-m)x \ - \cos(n+m)x] dx \\
&= 0, \quad n \neq m.
\end{aligned}$$

$$\langle \cos nx, \sin mx\rangle = \int_{-\pi}^{\pi} \cos nx \sin mx \ dx = 0, \quad n, m \in \mathbb{N},$$

because $\cos nx \sin mx$ is an odd function. Furthermore,

$$\begin{aligned}
\|1\| &= \sqrt{2\pi}, \\
\|\cos nx\| &= \left[\int_{-\pi}^{\pi} \cos^2 nx \ dx\right]^{1/2} = \sqrt{\pi}, \\
\|\sin nx\| &= \left[\int_{-\pi}^{\pi} \sin^2 nx \ dx\right]^{1/2} = \sqrt{\pi}, \quad n \in \mathbb{N}.
\end{aligned}$$

Thus the set

$$\left\{\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \cdots\right\},$$

which is obtained by dividing each function in the orthogonal set by its norm, is orthonormal in $\mathcal{L}^2(-\pi, \pi)$.

## Example 1.12

The set of functions

$$\{e^{inx} : n \in \mathbb{Z}\} = \{\ldots, e^{-i2x}, e^{-ix}, 1, e^{ix}, e^{i2x}, \ldots\}$$

is orthogonal in the complex space $\mathcal{L}^2(-\pi, \pi)$, because, for any $n \neq m$,

$$\begin{aligned} \langle e^{inx}, e^{imx} \rangle &= \int_{-\pi}^{\pi} e^{inx}\overline{e^{imx}}dx \\ &= \int_{-\pi}^{\pi} e^{inx}e^{-imx}dx \\ &= \frac{1}{i(n-m)}e^{i(n-m)x}\Big|_{-\pi}^{\pi} = 0. \end{aligned}$$

By dividing the functions in this set by

$$\left\|e^{inx}\right\| = \left[\int_{-\pi}^{\pi} e^{inx}\overline{e^{inx}}dx\right]^{1/2} = \sqrt{2\pi}, \quad n \in \mathbb{Z},$$

we obtain the corresponding orthonormal set

$$\left\{\frac{1}{\sqrt{2\pi}}e^{inx} : n \in \mathbb{Z}\right\}.$$

If $\rho$ is a positive continuous function on $(a, b)$, we define the inner product of two functions $f, g \in C(a, b)$ with respect to the weight function $\rho$ by

$$\langle f, g\rangle_\rho = \int_a^b f(x)\bar{g}(x)\rho(x)dx, \tag{1.13}$$

and we leave it to the reader to verify that all the properties of the inner product, as given in Definition 1.5, are satisfied. $f$ is then said to be orthogonal to $g$ with respect to the weight function $\rho$ if $\langle f, g\rangle_\rho = 0$. The induced norm

$$\|f\|_\rho = \left[\int_a^b |f(x)|^2 \rho(x)dx\right]^{1/2}$$

satisfies all the properties of the norm (1.11), including the CBS inequality and the triangle inequality. We use $\mathcal{L}^2_\rho(a, b)$ to denote the set of functions $f$ :

$(a, b) \to \mathbb{C}$, where $(a, b)$ may be finite or infinite, such that $\|f\|_\rho < \infty$. This is clearly an inner product space, and $\mathcal{L}^2(a, b)$ is then the special case in which $\rho \equiv 1$.

## *EXERCISES*

1.19 Prove the triangle inequality $\|f + g\| \leq \|f\| + \|g\|$ for any $f, g \in \mathcal{L}^2(a, b)$.

1.20 Verify the CBS inequality for the functions $f(x) = 1$ and $g(x) = x$ on $[0, 1]$.

1.21 Let the functions $f$ and $g$ be defined on $[0, 1]$ by

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q} \cap [0,1] \\ -1, & x \in \mathbb{Q}^c \cap [0,1] \end{cases}, \quad g(x) = 1 \quad \text{for all } x \in [0,1],$$

where $\mathbb{Q}$ is the set of rational numbers. Show that both $f^2$ and $g^2$ are Riemann integrable on $[0, 1]$ but that $fg$ is not.

1.22 Determine which of the following functions belongs to $\mathcal{L}^2(0, \infty)$ and calculate its norm.

$$(i)\ e^{-x}, \quad (ii)\ \sin x, \quad (iii)\ \frac{1}{1+x}, \quad (iv)\ \frac{1}{\sqrt[3]{x}}.$$

1.23 If $f$ and $g$ are positive, continuous functions in $\mathcal{L}^2(a, b)$, prove that $\langle f, g\rangle = \|f\| \|g\|$ if, and only if, $f$ and $g$ are linearly dependent.

1.24 Discuss the conditions under which the equality $\|f + g\| = \|f\| + \|g\|$ holds in $\mathcal{L}^2(a, b)$.

1.25 Determine the real values of $\alpha$ for which $x^\alpha$ lies in $\mathcal{L}^2(0, 1)$.

1.26 Determine the real values of $\alpha$ for which $x^\alpha$ lies in $\mathcal{L}^2(1, \infty)$.

1.27 If $f \in \mathcal{L}^2(0, \infty)$ and $\lim_{x\to\infty} f(x)$ exists, prove that $\lim_{x\to\infty} f(x) = 0$.

1.28 Assuming that the interval $(a, b)$ is finite, prove that if $f \in \mathcal{L}^2(a, b)$ then the integral $\int_a^b |f(x)|\, dx$ exists. Show that the converse is false by giving an example of a function $f$ such that $|f|$ is integrable on $(a, b)$, but $f \notin \mathcal{L}^2(a, b)$.

1.29 If the function $f : [0, \infty) \to \mathbb{R}$ is bounded and $|f|$ is integrable, prove that $f \in \mathcal{L}^2(0, \infty)$. Show that the converse is false by giving an example of a bounded function in $\mathcal{L}^2(0, \infty)$ which is not integrable on $[0, \infty)$.

1.30 In $\mathcal{L}^2(-\pi,\pi)$, express the function $\sin^3 x$ as a linear combination of the orthogonal functions $\{1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots\}$.

1.31 Define a function $f \in \mathcal{L}^2(-1,1)$ such that $\langle f, x^2+1 \rangle = 0$ and $\|f\| = 2$.

1.32 Given $\rho(x) = e^{-x}$, prove that any polynomial in $x$ belongs to $\mathcal{L}^2_\rho(0,\infty)$.

1.33 Show that if $\rho$ and $\sigma$ are two weight functions such that $\rho \geq \sigma \geq 0$ on $(a,b)$, then $\mathcal{L}^2_\rho(a,b) \subseteq \mathcal{L}^2_\sigma(a,b)$.

## 1.4 Sequences of Functions

Much of the subject of this book deals with sequences and series of functions, and this section presents the background that we need on their convergence properties. We assume that the reader is familiar with the basic theory of numerical sequences and series which is usually covered in advanced calculus.

Suppose that for each $n \in \mathbb{N}$ we have a (real or complex) function $f_n : I \to \mathbb{F}$ defined on a real interval $I$. We then say that we have a *sequence of functions* $(f_n : n \in \mathbb{N})$ defined on $I$. Suppose, furthermore, that, for every fixed $x \in I$, the *sequence of numbers* $(f_n(x) : n \in \mathbb{N})$ converges as $n \to \infty$ to some limit in $\mathbb{F}$. Now we define the function $f : I \to \mathbb{F}$, for each $x \in I$, by

$$f(x) = \lim_{n\to\infty} f_n(x). \tag{1.14}$$

That means, given any positive number $\varepsilon$, there is a positive integer $N$ such that

$$n \geq N \;\Rightarrow\; |f_n(x) - f(x)| < \varepsilon. \tag{1.15}$$

Note that the number $N$ depends on the point $x$ as much as it depends on $\varepsilon$, hence $N = N(\varepsilon, x)$. The function $f$ defined in Equation (1.14) is called the *pointwise limit* of the sequence $(f_n)$.

### Definition 1.13

A sequence of functions $f_n : I \to \mathbb{F}$ is said to *converge pointwise* to the function $f : I \to \mathbb{F}$, expressed symbolically by

$$\lim_{n\to\infty} f_n = f, \quad \lim f_n = f, \quad \text{or } f_n \to f,$$

if, for every $x \in I$, $\lim_{n\to\infty} f_n(x) = f(x)$.

### Example 1.14

(i) Let $f_n(x) = \frac{1}{n}\sin nx,\ x \in \mathbb{R}$. In as much as

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{1}{n}\sin nx = 0 \quad \text{for every } x \in \mathbb{R},$$

the pointwise limit of this sequence is the function $f(x) = 0,\ x \in \mathbb{R}$.

(ii) For all $x \in [0,1]$,

$$f_n(x) = x^n \to \begin{cases} 0, & 0 \le x < 1 \\ 1, & x = 1, \end{cases}$$

hence the limit function is

$$f(x) = \begin{cases} 0, & 0 \le x < 1 \\ 1, & x = 1, \end{cases} \tag{1.16}$$

as shown in Figure 1.1.

(iii) For all $x \in [0,\infty)$,

$$f_n(x) = \frac{nx}{1+nx} \to f(x) = \begin{cases} 0, & x = 0 \\ 1, & x > 0. \end{cases}$$

### Example 1.15

For each $n \in \mathbb{N}$, define the sequence $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 0, & x = 0 \\ n, & 0 < x \le 1/n \\ 0, & 1/n < x \le 1. \end{cases}$$

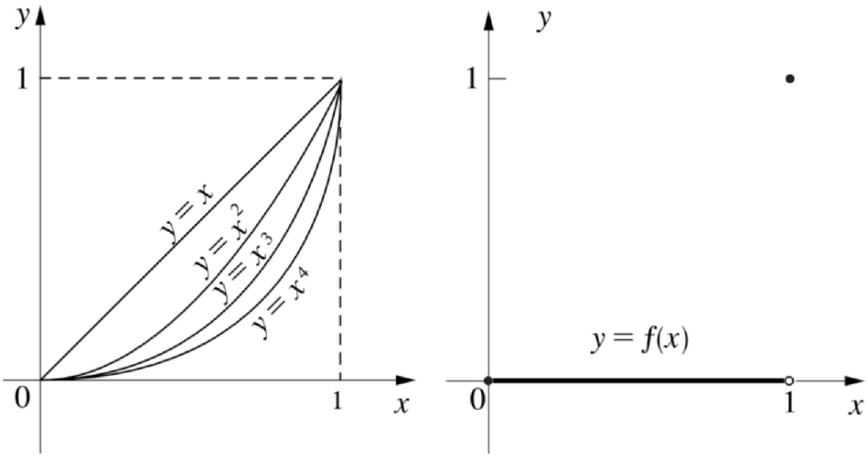


**Figure 1.1** The sequence $f_n(x) = x^n$.

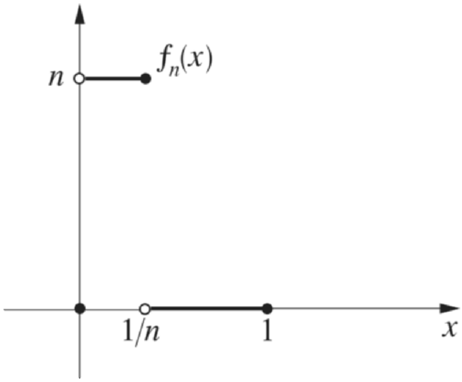


**Figure 1.2**

To determine the limit $f$, we first note that $f_n(0) = 0$ for all $n$. If $x > 0$, then there is an integer $N$ such that $1/N < x$, in which case

$$n \geq N \;\Rightarrow\; \frac{1}{n} \leq \frac{1}{N} < x \;\Rightarrow\; f_n(x) = 0.$$

Therefore $f_n \to 0$ (see Figure 1.2).

If the number $N$ in the implication (1.15) does not depend on $x$, that is, if for every $\varepsilon > 0$ there is an integer $N = N(\varepsilon)$ such that

$$n \geq N \;\Rightarrow\; |f_n(x) - f(x)| < \varepsilon \quad \text{for all } x \in I, \tag{1.17}$$

then the convergence $f_n \to f$ is called *uniform*, and we distinguish this from pointwise convergence by writing

$$f_n \xrightarrow{u} f.$$

Going back to Example 1.14, we note the following.

(i) Since

$$|f_n(x) - 0| = \left|\frac{1}{n}\sin nx\right| \leq \frac{1}{n} \quad \text{for all } x \in \mathbb{R},$$

we see that any choice of $N$ greater than $1/\varepsilon$ will satisfy the implication (1.17), hence

$$\frac{1}{n}\sin nx \xrightarrow{u} 0 \;\text{ on } \mathbb{R}.$$

(ii) The convergence $x^n \to 0$ is not uniform on $[0, 1)$ because the implication

$$n \geq N \;\Rightarrow\; |x^n - 0| = x^n < \varepsilon$$

cannot be satisfied on the whole interval $[0,1)$ if $0<\varepsilon<1$, but only on $[0,\sqrt[n]{\varepsilon})$, because $x^n>\varepsilon$ for all $x\in(\sqrt[n]{\varepsilon},1)$. Hence the convergence $f_n\to f$, where $f$ is given in (1.16), is not uniform.

(iii) The convergence

$$\frac{nx}{1+nx}\to 1,\quad x\in(0,\infty)$$

is also not uniform in as much as the inequality

$$\left|\frac{nx}{1+nx}-1\right|=\frac{1}{1+nx}<\varepsilon$$

cannot be satisfied for values of $x$ in $(0,(1-\varepsilon)/n\varepsilon]$ if $0<\varepsilon<1$.

## Remark 1.16

1. The uniform convergence $f_n\xrightarrow{u}f$ clearly implies the pointwise convergence $f_n\to f$ (but not vice versa). Hence, when we wish to test for the uniform convergence of a sequence $f_n$, the candidate function $f$ for the uniform limit of $f_n$ should always be the pointwise limit.

2. In the inequalities (1.15) and (1.17) we can replace the relation $<$ by $\leq$ and the positive number $\varepsilon$ by $c\varepsilon$, where $c$ is a positive constant (which does not depend on $n$).

3. Because the statement $|f_n(x)-f(x)|\leq\varepsilon$ for all $x\in I$ is equivalent to

$$\sup_{x\in I}|f_n(x)-f(x)|\leq\varepsilon,$$

we see that $f_n\xrightarrow{u}f$ on $I$ if, and only if, for every $\varepsilon>0$ there is an integer $N$ such that

$$n\geq N\ \Rightarrow\ \sup_{x\in I}|f_n(x)-f(x)|\leq\varepsilon,$$

which is equivalent to the statement

$$\sup_{x\in I}|f_n(x)-f(x)|\to 0\quad\text{as } n\to\infty. \tag{1.18}$$

Using the criterion (1.18) for uniform convergence on the sequences of Example 1.14, we see that, in (i),

$$\sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sin nx\right|\leq\frac{1}{n}\to 0,$$

thus confirming the uniform convergence of $\sin nx/n$ to 0. In (ii) and (iii), we have

$$\sup_{x\in[0,1]}|x^n-f(x)|=\sup_{x\in[0,1)}x^n=1\nrightarrow 0,$$

$$\sup_{x\in[0,\infty)}\left|\frac{nx}{1+nx}-f(x)\right| = \sup_{x\in(0,\infty)}\left(1-\frac{nx}{1+nx}\right) = 1 \nrightarrow 0,$$

hence neither sequence converges uniformly.

Although all three sequences discussed in Example 1.14 are continuous, only the first one, $(\sin nx/n)$, converges to a continuous limit. This would seem to indicate that uniform convergence preserves the property of continuity as the sequence passes to the limit. We should also be interested to know under what conditions we can interchange the operations of integration or differentiation with the process of passage to the limit. In other words, when can we write

$$\int_I \lim f_n(x)dx = \lim \int_I f_n(x)dx, \quad \text{or} \quad (\lim f_n)' = \lim\ f_n' \text{ on } I?$$

The answer is contained in the following theorem, which gives sufficient conditions for the validity of these equalities. This is a standard result in classical real analysis whose proof may be found in a number of references, such as [1] or [14].

## Theorem 1.17

Let $(f_n)$ be a sequence of functions defined on the interval $I$ which converges pointwise to $f$ on $I$.

(i) If $f_n$ is continuous for every $n$, and $f_n \xrightarrow{u} f$, then $f$ is continuous on $I$.

(ii) If $f_n$ is integrable for every $n$, $I$ is bounded, and $f_n \xrightarrow{u} f$, then $f$ is integrable on $I$ and

$$\int_I f(x)dx = \lim \int_I f_n(x)dx.$$

(iii) If $f_n$ is differentiable on $I$ for every $n$, $I$ is bounded, and $f_n'$ converges uniformly on $I$, then $f_n$ converges uniformly to $f$, $f$ is differentiable on $I$, and

$$f_n' \xrightarrow{u} f' \text{ on } I.$$

## Remark 1.18

Part (iii) of Theorem 1.17 remains valid if pointwise convergence of $f_n$ on $I$ is replaced by the weaker condition that $f_n$ converges at any single point in $I$, for such a condition is only needed to ensure the convergence of the constants of integration in going from $f_n'$ to $f_n$.

Going back to Example 1.14, we observe that the uniform convergence of $\sin nx/n$ to 0 satisfies part (i) of Theorem 1.17. It also satisfies (ii) over any bounded interval in $\mathbb{R}$. But (iii) is not satisfied, in as much as the sequence

$$\frac{d}{dx}\left(\frac{1}{n}\sin nx\right) = \cos nx$$

is not convergent. The sequence $(x^n)$ is continuous on $[0, 1]$ for every $n$, but its limit is not. This is consistent with (i), because the convergence is not uniform. The same observation applies to the sequence $nx/(1 + nx)$.

In Example 1.15 we have

$$\int_0^1 f_n(x)dx = \int_0^{1/n} ndx = 1 \quad \text{for all } n \in \mathbb{N}$$
$$\Rightarrow \lim \int_0^1 f_n(x)dx = 1,$$

whereas

$$\int_0^1 \lim f_n(x)dx = 0.$$

This implies that the convergence $f_n \to 0$ is not uniform, which is confirmed by the fact that

$$\sup_{0\leq x\leq 1} f_n(x) = n.$$

On the other hand,

$$\lim \int_0^1 x^n dx = 0 = \int_0^1 \lim x^n dx,$$

although the convergence $x^n \to 0$ is not uniform, which indicates that not all the conditions of Theorem 1.17 are necessary.

Given a sequence of (real or complex) functions $(f_n)$ defined on a real interval $I$, we define its $n$th partial sum by

$$S_n(x) = f_1(x) + \cdots + f_n(x) = \sum_{k=1}^{n} f_k(x), \quad x \in I.$$

The sequence of functions $(S_n)$, defined on $I$, is called an *infinite series* (of functions) and is denoted $\sum f_k$. The series is said to *converge pointwise* on $I$ if the sequence $(S_n)$ converges pointwise on $I$, in which case $\sum f_k$ is called *convergent.* Its limit is the sum of the series

$$\lim_{n\to\infty} S_n(x) = \sum_{k=1}^{\infty} f_k(x), \ x \in I.$$

Sometimes we shall find it convenient to identify a convergent series with its sum, just as we occasionally identify a function $f$ with its value $f(x)$. A series which does not converge at a point is said to *diverge* at that point. The series $\sum f_k$ is *absolutely convergent* on $I$ if the positive series $\sum |f_k|$ is pointwise convergent on $I$, and *uniformly convergent* on $I$ if the sequence $(S_n)$ is uniformly convergent on $I$. In investigating the convergence properties of series of functions we naturally rely on the corresponding convergence properties of sequences of functions, as discussed earlier, because a series is ultimately a sequence. But we shall often resort to the convergence properties of series of numbers, which we assume that the reader is familiar with, such as the various tests of convergence (comparison test, ratio test, root test, alternating series test), and the behaviour of such series as the geometric series and the $p$-series (see [1] or [3]).

Applying Theorem 1.17 to series, we arrive at the following result.

## Corollary 1.19

Suppose the series $\sum f_n$ converges pointwise on the interval $I$.

(i) If $f_n$ is continuous on $I$ for every $n$ and $\sum f_n$ converges uniformly on $I$, then its sum $\sum_{n=1}^{\infty} f_n$ is continuous.

(ii) If $f_n$ is integrable on $I$ for every $n$, $I$ is bounded, and $\sum f_n$ converges uniformly, then $\sum_{n=1}^{\infty} f_n$ is integrable on $I$ and

$$\int_I \sum_{n=1}^{\infty} f_n(x)dx = \sum_{n=1}^{\infty} \int_I f_n(x)dx.$$

(iii) If $f_n$ is differentiable on $I$ for every $n$, $I$ is bounded, and $\sum f_n'$ converges uniformly on $I$, then $\sum f_n$ converges uniformly and its limit is differentiable on $I$ and satisfies

$$\left(\sum_{n=1}^{\infty} f_n\right)' = \sum_{n=1}^{\infty} f_n'.$$

This corollary points out the relevance of uniform convergence to manipulating series, and it would be helpful if we had a simpler and more practical test for the uniform convergence of a series than applying the definition. This is provided by the following theorem, which gives a sufficient condition for the uniform convergence of a series of functions.

## Theorem 1.20 (Weierstrass M-Test)

Let $(f_n)$ be a sequence of functions on $I$, and suppose that there is a sequence of (nonnegative) numbers $M_n$ such that

$$|f_n(x)| \leq M_n \quad \text{for all } x \in I, \quad n \in \mathbb{N}.$$

If $\sum M_n$ converges, then $\sum f_n$ converges uniformly and absolutely on $I$.

## Proof

Let $\varepsilon > 0$. We have

$$\begin{aligned}\left|\sum_{k=1}^{\infty} f_k(x) - \sum_{k=1}^{n} f_k(x)\right| &\leq \sum_{k=n+1}^{\infty} |f_k(x)| \\ &\leq \sum_{k=n+1}^{\infty} M_k \quad \text{for all } x \in I, \quad n \in \mathbb{N}.\end{aligned}$$

Because the series $\sum M_k$ is convergent, there is an integer $N$ such that

$$\begin{aligned}n \geq N \Rightarrow & \sum_{k=n+1}^{\infty} M_k < \varepsilon \\ & \Rightarrow \left|\sum_{k=1}^{\infty} f_k(x) - \sum_{k=1}^{n} f_k(x)\right| < \varepsilon \quad \text{for all } x \in I.\end{aligned}$$

By definition, this means $\sum f_k$ is uniformly convergent on $I$. Absolute convergence follows by comparison with $M_n$. □

## Example 1.21

(i) The trigonometric series

$$\sum \frac{1}{n^2} \sin nx$$

is uniformly convergent on $\mathbb{R}$ because

$$\left|\frac{1}{n^2} \sin nx\right| \leq \frac{1}{n^2}$$

and the series $\sum 1/n^2$ is convergent. Because $\sin nx/n^2$ is continuous on $\mathbb{R}$ for every $n$, the function $\sum_{n=1}^{\infty} \sin nx/n^2$ is also continuous on $\mathbb{R}$. Furthermore, by Corollary 1.19, the integral of the series on any finite interval $[a, b]$ is

$$\begin{aligned}\int_a^b \left(\sum_{n=1}^{\infty} \frac{1}{n^2} \sin nx\right) dx &= \sum_{n=1}^{\infty} \frac{1}{n^2} \int_a^b \sin nx \; dx \\ &= \sum_{n=1}^{\infty} \frac{1}{n^3} (\cos na - \cos nb) \\ &\leq 2 \sum_{n=1}^{\infty} \frac{1}{n^3},\end{aligned}$$

which is convergent. On the other hand, the series of derivatives

$$\sum_{n=1}^{\infty} \frac{d}{dx} \left(\frac{1}{n^2} \sin nx\right) = \sum_{n=1}^{\infty} \frac{1}{n} \cos nx$$

is not uniformly convergent. In fact, it is not even convergent at some values of $x$, such as the integral multiples of $2\pi$. Hence we cannot write

$$\frac{d}{dx}\sum_{n=1}^{\infty}\frac{1}{n^2}\sin nx=\sum_{n=1}^{\infty}\frac{1}{n}\cos nx \quad \text{for all } x\in\mathbb{R}.$$

(ii) By the M-test, both the series

$$\sum\frac{1}{n^3}\sin nx$$

and

$$\sum_{n=1}^{\infty}\frac{d}{dx}\left(\frac{1}{n^3}\sin nx\right)=\sum_{n=1}^{\infty}\frac{1}{n^2}\cos nx$$

are uniformly convergent on $\mathbb{R}$. Hence the equality

$$\frac{d}{dx}\sum_{n=1}^{\infty}\frac{1}{n^3}\sin nx=\sum_{n=1}^{\infty}\frac{1}{n^2}\cos nx$$

is valid for all $x$ in $\mathbb{R}$.

## *EXERCISES*

1.34 Calculate the pointwise limit where it exists.

(a) $\dfrac{x^n}{1+x^n}$, $x\in\mathbb{R}$.

(b) $\sqrt[n]{x}$, $0\le x<\infty$.

(c) $\sin nx$, $x\in\mathbb{R}$.

1.35 Determine the type of convergence (pointwise or uniform) for each of the following sequences.

(a) $\dfrac{x^n}{1+x^n}$, $0\le x\le 2$.

(b) $\sqrt[n]{x}$, $1/2\le x\le 1$.

(c) $\sqrt[n]{x}$, $0\le x\le 1$.

1.36 Determine the type of convergence for the sequence

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

and decide whether the equality

$$\lim\int_0^1 f_n(x)dx=\int_0^1 \lim f_n(x)dx$$

is valid.

1.37 Evaluate the limit of the sequence

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

and determine the type of convergence.

1.38 Determine the limit and the type of convergence for the sequence $f_n(x) = nx(1-x^2)^n$ on $[0,1]$.

1.39 Prove that the convergence

$$\frac{x}{n+x} \to 0$$

is uniform on $[0,a]$ for any $a > 0$, but not on $[0,\infty)$.

1.40 Given

$$f_n(x) = \begin{cases} 1/n, & |x| < n \\ 0, & |x| > n, \end{cases}$$

prove that $f_n \xrightarrow{u} 0$. Evaluate $\lim \int_{-\infty}^{\infty} f_n(x)dx$ and explain why it is not 0.

1.41 If the sequence $(f_n)$ converges uniformly to $f$ on $[a,b]$, prove that $|f_n - f|$, and hence $|f_n - f|^2$, converges uniformly to 0 on $[a,b]$.

1.42 Determine the domain of convergence of the series $\sum f_n$, where

(a) $f_n(x) = \dfrac{1}{n^2 + x^2}$.

(b) $f_n(x) = \dfrac{x^n}{1 + x^n}$.

1.43 If the series $\sum a_n$ is absolutely convergent, prove that $\sum a_n \sin nx$ is uniformly convergent on $\mathbb{R}$.

1.44 Prove that

$$\lim_{n\to\infty} \int_{n\pi}^{(n+1)\pi} \frac{|\sin x|}{x} dx = 0.$$

Use this to conclude that the improper integral

$$\int_0^{\infty} \frac{\sin x}{x} dx$$

exists. Show that the integral $\int_0^{\infty} (|\sin x|/x)dx = \infty$. Hint: Use the alternating series test and the divergence of the harmonic series $\sum 1/n$.

1.45 The series

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots$$

is called a *power series* about the point 0. It is known (see [1]) that this series converges in $(-R, R)$ and diverges outside $[-R, R]$, where

$$R = \left[\lim_{n\to\infty} \sqrt[n]{|a_n|}\right]^{-1} = \lim_{n\to\infty} \left|\frac{a_n}{a_{n+1}}\right| \geq 0.$$

If $R > 0$, use the Weierstrass M-test to prove that the power series converges uniformly on $[-R+\varepsilon, R-\varepsilon]$, where $\varepsilon$ is any positive number less than $R$.

1.46 Use the result of Exercise 1.45 to show that the function

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

is continuous on $(-R, R)$; then show that $f$ is also differentiable on $(-R, R)$ with

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

1.47 From Exercise 1.46 conclude that the power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is differentiable any number of times on $(-R, R)$, and that $a_n = f^{(n)}(0)/n!$ for all $n \in \mathbb{N}$.

1.48 Use the result of Exercise 1.47 to obtain the following power series (Taylor series) representations of the exponential and trigonometric functions on $\mathbb{R}$.

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!},$$

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

1.49 Use the result of Exercise 1.48 to prove Euler's formula $e^{ix} = \cos x + i \sin x$ for all $x \in \mathbb{R}$, where $i = \sqrt{-1}$.

## 1.5 Convergence in $\mathcal{L}^2$

Having discussed pointwise and uniform convergence for a sequence of functions, we now consider a third type: convergence in $\mathcal{L}^2$.

### Definition 1.22

A sequence of functions $(f_n)$ in $\mathcal{L}^2(a,b)$ is said to *converge in* $\mathcal{L}^2$ if there is a function $f \in \mathcal{L}^2(a,b)$ such that

$$\lim_{n\to\infty} \|f_n - f\| = 0, \tag{1.19}$$

that is, if for every $\varepsilon > 0$ there is an integer $N$ such that

$$n \geq N \Rightarrow \|f_n - f\| < \varepsilon.$$

Equation (1.19) is equivalent to writing

$$f_n \overset{\mathcal{L}^2}{\to} f,$$

and $f$ is called the *limit in* $\mathcal{L}^2$ of the sequence $(f_n)$.

### Example 1.23

(i) In Example 1.14(ii) we saw that, pointwise,

$$x^n \to \begin{cases} 0, & 0 \leq x < 1 \\ 1, & x = 1. \end{cases}.$$

Because $\mathcal{L}^2([0,1]) = \mathcal{L}^2([0,1))$, we have

$$\|x^n - 0\| = \left[\int_0^1 x^{2n} dx\right]^{1/2} = \left[\frac{1}{2n+1}\right]^{1/2} \to 0.$$

Therefore $x^n \overset{\mathcal{L}^2}{\to} 0$.

(ii) The sequence of functions $(f_n)$ defined in Example 1.15 by

$$f_n(x) = \begin{cases} 0, & x = 0 \\ n, & 0 < x \leq 1/n \\ 0, & 1/n < x \leq 1 \end{cases}$$

also converges pointwise to 0 on $[0,1]$. But in this case,

$$\begin{aligned}\|f_n - 0\|^2 &= \int_0^1 f_n^2(x)dx \\ &= \int_0^{1/n} n^2 dx \\ &= n \quad \text{for all } n \in \mathbb{N}.\end{aligned}$$

Thus $\|f_n - 0\| = \sqrt{n} \nrightarrow 0$, which means the sequence $f_n$ does not converge to 0 in $\mathcal{L}^2$.

This last example shows that pointwise convergence does not imply convergence in $\mathcal{L}^2$. Conversely, convergence in $\mathcal{L}^2$ cannot imply pointwise convergence, because the limit in this case is a class of functions (which are equal in $\mathcal{L}^2$ but not pointwise). It is legitimate to ask, however, whether a sequence that converges pointwise to some limit $f$ can converge to a different limit in $\mathcal{L}^2$. For example, can the sequence $(f_n)$ in Example 1.23(ii) converge in $\mathcal{L}^2$ to some function other than 0? The answer is no. In other words, if a sequence converges both pointwise and in $\mathcal{L}^2$, then its limit is the same in both cases. More precisely, we should say that the two limits are not distinguishable in $\mathcal{L}^2$ as they belong to the same equivalence class.

On the other hand, uniform convergence $f_n \xrightarrow{u} f$ over $I$ implies pointwise convergence, as we have already observed, and we now show that it also implies $f_n \xrightarrow{\mathcal{L}^2} f$ provided the sequence $(f_n)$ and $f$ lie in $\mathcal{L}^2(I)$ and $I$ is bounded: Because $f_n - f \xrightarrow{u} 0$, it is a simple matter to show that $|f_n - f|^2 \xrightarrow{u} 0$ (Exercise 1.41). By Theorem 1.17(ii), we therefore have

$$\begin{aligned}\lim_{n\to\infty} \|f_n - f\|^2 &= \lim_{n\to\infty} \int_I |f_n(x) - f(x)|^2 \, dx \\ &= \int_I \lim_{n\to\infty} |f_n(x) - f(x)|^2 \, dx = 0.\end{aligned}$$

The condition that $f$ belong to $\mathcal{L}^2(I)$ is actually not needed, as we shall discover in Theorem 1.26.

## Example 1.24

We saw in Example 1.21 that

$$S_n(x) = \sum_{k=1}^{n} \frac{1}{k^2} \sin kx \xrightarrow{u} S(x) = \sum_{k=1}^{\infty} \frac{1}{k^2} \sin kx, \quad x \in \mathbb{R},$$

hence the function $S(x)$ is continuous on $[-\pi, \pi]$. Moreover, both $S_n$ and $S$ lie in $\mathcal{L}^2(-\pi, \pi)$ because each is uniformly bounded above by the convergent series

$\sum 1/k^2$. Therefore $S_n$ converges to $S$ in $\mathcal{L}^2(-\pi,\pi)$. Equivalently, we say that the series $\sum \sin kx/k^2$ converges to $\sum_{k=1}^{\infty} \sin kx/k^2$ in $\mathcal{L}^2(-\pi,\pi)$ and write

$$\lim \sum_{k=1}^{n} \frac{1}{k^2} \sin kx = \sum_{k=1}^{\infty} \frac{1}{k^2} \sin kx \text{ in } \mathcal{L}^2(-\pi,\pi).$$

The series $\sum \sin kx/k$, on the other hand, cannot be tested for convergence in $\mathcal{L}^2$ with the tools available, and we have to develop the theory a little further. First we define a Cauchy sequence in $\mathcal{L}^2$ along the lines of the corresponding notion in $\mathbb{R}$. This allows us to test a sequence for convergence without having to guess its limit beforehand.

## Definition 1.25

A sequence in $\mathcal{L}^2$ is called a *Cauchy sequence* if, for every $\varepsilon > 0$, there is an integer $N$ such that

$$m, n \geq N \Rightarrow \|f_n - f_m\| < \varepsilon.$$

Clearly, every convergent sequence $(f_n)$ in $\mathcal{L}^2$ is a Cauchy sequence; for if $f_n \xrightarrow{\mathcal{L}^2} f$, then, by the triangle inequality,

$$\|f_n - f_m\| \leq \|f_n - f\| + \|f_m - f\|,$$

and we can make the right-hand side of this inequality arbitrarily small by taking $m$ and $n$ large enough. The converse of this statement (i.e., that every Cauchy sequence in $\mathcal{L}^2$ converges to some function in $\mathcal{L}^2$) is also true and expresses the *completeness property* of $\mathcal{L}^2$.

## Theorem 1.26 (Completeness of $\mathcal{L}^2$)

For every Cauchy sequence $(f_n)$ in $\mathcal{L}^2$ there is a function $f \in \mathcal{L}^2$ such that $f_n \xrightarrow{\mathcal{L}^2} f$.

There is another theorem which states that, for every function $f \in \mathcal{L}^2(a,b)$, there is a sequence of continuous functions $(f_n)$ on $[a,b]$ such that $f_n \xrightarrow{\mathcal{L}^2} f$. In other words, the set of functions $C([a,b])$ is dense in $\mathcal{L}^2(a,b)$ in much the same way that the rationals $\mathbb{Q}$ are dense in $\mathbb{R}$, keeping in mind of course the different topologies of $\mathbb{R}$ and $\mathcal{L}^2$, the first being defined by the absolute value $|\cdot|$ and the second by the norm $\|\cdot\|$. For example, the $\mathcal{L}^2(-1,1)$ function

$$f(x) = \begin{cases} 0, & -1 \leq x < 0 \\ 1, & 0 \leq x \leq 1, \end{cases}$$

which is discontinuous at $x = 0$, can be approached in the $\mathcal{L}^2$ norm by the sequence of continuous functions

$$f_n = \begin{cases} 0, & -1 \le x \le -1/n \\ nx + 1, & -1/n < x < 0 \\ 1, & 0 \le x \le 1. \end{cases}$$

This is clear from

$$\begin{aligned} \lim_{n\to\infty} \|f_n - f\| &= \lim_{n\to\infty} \left[ \int_{-1}^{1} |f_n(x) - f(x)|^2 \, dx \right]^{1/2} \\ &= \lim_{n\to\infty} \left[ \int_{-1/n}^{0} (nx+1)^2 dx \right]^{1/2} \\ &= \lim_{n\to\infty} 1/\sqrt{3n} = 0. \end{aligned}$$

Needless to say, there are many other sequences in $C([-1,1])$ which converge to $f$ in $\mathcal{L}^2(-1,1)$, just as there are many sequences in $\mathbb{Q}$ which converge to the irrational number $\sqrt{2}$.

As we shall have occasion to refer to this result in the following chapter, we give here its precise statement.

## Theorem 1.27 (Density of $C$ in $\mathcal{L}^2$)

For any $f \in \mathcal{L}^2(a,b)$ and any $\varepsilon > 0$, there is a continuous function $g$ on $[a,b]$ such that $\|f - g\| < \varepsilon$.

The proofs of Theorems 1.26 and 1.27 may be found in [14]. The space $\mathcal{L}^2$ is one of the most important examples of a *Hilbert space,* which is an inner product space that is complete under the norm defined by the inner product. It is named after David Hilbert (1862–1943), the German mathematician whose work and inspiration did much to develop the ideas of Hilbert space (see [7], vol. I). Many of the ideas that we work with are articulated within the context of $\mathcal{L}^2$.

## Example 1.28

Using Theorem 1.26, we can now look into the question of convergence of the sequence $S_n(x) = \sum_{k=1}^{n} \sin kx/k$ in $\mathcal{L}^2(-\pi,\pi)$. Noting that

$$\|S_n(x) - S_m(x)\|^2 = \left\| \sum_{k=m+1}^{n} \frac{1}{k} \sin kx \right\|^2, \quad m < n,$$

we can use the orthogonality of $\{\sin kx : k \in \mathbb{N}\}$ in $\mathcal{L}^2(-\pi,\pi)$ (Example 1.11) to obtain

$$\left\| \sum_{k=m+1}^{n} \frac{1}{k} \sin kx \right\|^2 = \sum_{k=m+1}^{n} \frac{1}{k^2} \|\sin kx\|^2 = \pi \sum_{k=m+1}^{n} \frac{1}{k^2}.$$

Suppose $\varepsilon > 0$. Since $\sum 1/k^2$ is convergent, we can choose $N$ so that

$$\begin{aligned} n > m \geq N &\Rightarrow \sum_{k=m+1}^{n} \frac{1}{k^2} < \frac{\varepsilon^2}{\pi} \\ &\Rightarrow \|S_n(x) - S_m(x)\| < \varepsilon. \end{aligned}$$

Thus $\sum_{k=1}^{n} \sin kx/k$ is a Cauchy sequence and hence converges in $\mathcal{L}^2(-\pi,\pi)$, although we cannot as yet tell to what limit.

Similarly, the series $\sum \cos kx/k$ converges in $\mathcal{L}^2(-\pi,\pi)$, although this series diverges pointwise at certain values of $x$, such as all integral multiples of $2\pi$.

This section was devoted to convergence in $\mathcal{L}^2$ because of its importance to the theory of Fourier series, but we could just as easily have been discussing convergence in the weighted space $\mathcal{L}^2_\rho$. Definitions 1.22 and 1.25 and Theorems 1.26 and 1.27 would remain unchanged, with the norm $\|\cdot\|$ replaced by $\|\cdot\|_\rho$ and convergence in $\mathcal{L}^2$ by convergence in $\mathcal{L}^2_\rho$.

## *EXERCISES*

1.50 Determine the limit in $\mathcal{L}^2$ of each of the following sequences where it exists.

(a) $f_n(x) = \sqrt[n]{x},\ 0 \leq x \leq 1.$

(b) $f_n(x) = \begin{cases} nx, & 0 \leq x < 1/n \\ 1, & 1/n \leq x \leq 1. \end{cases}$

(c) $f_n(x) = nx(1-x)^n,\ 0 \leq x \leq 1.$

1.51 Test the following series for convergence in $\mathcal{L}^2$.

(a) $\sum \dfrac{1}{k^{2/3}} \sin kx,\ -\pi \leq x \leq \pi.$

(b) $\sum \dfrac{1}{k} e^{ikx},\ -\pi \leq x \leq \pi.$

(c) $\sum \dfrac{1}{\sqrt{k+1}} \cos kx,\ -\pi \leq x \leq \pi.$

1.52 If $(f_n)$ is a sequence in $\mathcal{L}^2(a,b)$ which converges to $f$ in $\mathcal{L}^2$, show that $\langle f_n, g\rangle \xrightarrow{\mathcal{L}^2} \langle f, g\rangle$ for any $g \in \mathcal{L}^2(a,b)$.

1.53 Prove that $|\|f\| - \|g\|| \leq \|f-g\|$, and hence conclude that if $f_n \xrightarrow{\mathcal{L}^2} f$ then $\|f_n\| \to \|f\|$.

1.54 If the numerical series $\sum |a_n|$ is convergent, prove that $\sum |a_n|^2$ is also convergent, and that the series $\sum a_n \sin nx$ and $\sum a_n \cos nx$ are both continuous on $[-\pi, \pi]$.

1.55 Prove that if the weight functions $\rho$ and $\sigma$ are related by $\rho \geq \sigma$ on $(a,b)$, then a sequence which converges in $\mathcal{L}^2_\rho(a,b)$ also converges in $\mathcal{L}^2_\sigma(a,b)$.

## 1.6 Orthogonal Functions

Let

$$\{\varphi_1, \varphi_2, \varphi_3, \ldots\}$$

be an orthogonal set of (nonzero) functions in the complex space $\mathcal{L}^2$, which may be finite or infinite, and suppose that the function $f \in \mathcal{L}^2$ is a finite linear combination of elements in the set $\{\varphi_i\}$,

$$f = \sum_{i=1}^n \alpha_i \varphi_i, \quad \alpha_i \in \mathbb{C}. \tag{1.20}$$

Taking the inner product of $f$ with $\varphi_k$,

$$\langle f, \varphi_k \rangle = \alpha_k \|\varphi_k\|^2 \quad \text{for all } k = 1, \ldots, n,$$

we conclude that

$$\alpha_k = \frac{\langle f, \varphi_k\rangle}{\|\varphi_k\|^2},$$

and the representation (1.20) takes the form

$$f = \sum_{k=1}^n \frac{\langle f, \varphi_k\rangle}{\|\varphi_k\|^2} \varphi_k.$$

In other words, the coefficients $\alpha_k$ in the linear combination (1.20) are determined by the projections of $f$ on $\varphi_k$. In terms of the corresponding orthonormal set $\{\psi_k = \varphi_k / \|\varphi_k\|\}$,

$$f = \sum_{k=1}^n \langle f, \psi_k\rangle \psi_k,$$

and the coefficients coincide with the projections of $f$ on $\psi_k$.

Suppose, on the other hand, that $f$ is an arbitrary function in $\mathcal{L}^2$ and that we want to obtain the best approximation of $f$ in $\mathcal{L}^2$, that is, in the norm $\|\cdot\|$, by a finite linear combination of the elements of $\{\varphi_k\}$. We should then look for the coefficients $\alpha_k$ which minimize the nonnegative number

$$\left\| f - \sum_{k=1}^{n} \alpha_k \varphi_k \right\|.$$

We have

$$\begin{aligned}
\left\| f - \sum_{k=1}^{n} \alpha_k \varphi_k \right\|^2 &= \left\langle f - \sum_{k=1}^{n} \alpha_k \varphi_k, f - \sum_{k=1}^{n} \alpha_k \varphi_k \right\rangle \\
&= \|f\|^2 - 2 \sum_{k=1}^{n} \operatorname{Re} \bar{\alpha}_k \langle f, \varphi_k \rangle + \sum_{k=1}^{n} |\alpha_k|^2 \|\varphi_k\|^2 \\
&= \|f\|^2 - \sum_{k=1}^{n} \frac{|\langle f, \varphi_k \rangle|^2}{\|\varphi_k\|^2} \\
&\quad + \sum_{k=1}^{n} \|\varphi_k\|^2 \left[ |\alpha_k^2| - 2 \operatorname{Re} \bar{\alpha}_k \frac{\langle f, \varphi_k \rangle}{\|\varphi_k\|^2} + \frac{|\langle f, \varphi_k \rangle|^2}{\|\varphi_k\|^4} \right] \\
&= \|f\|^2 - \sum_{k=1}^{n} \frac{|\langle f, \varphi_k \rangle|^2}{\|\varphi_k\|^2} + \sum_{k=1}^{n} \|\varphi_k\|^2 \left| \alpha_k - \frac{\langle f, \varphi_k \rangle}{\|\varphi_k\|^2} \right|^2.
\end{aligned}$$

Since the coefficients $\alpha_k$ appear only in the last term

$$\sum_{k=1}^{n} \|\varphi_k\|^2 \left| \alpha_k - \frac{\langle f, \varphi_k \rangle}{\|\varphi_k\|^2} \right|^2 \geq 0,$$

we obviously achieve the minimum of $\left\| f - \sum_{k=1}^{n} \alpha_k \varphi_k \right\|^2$, and hence of $\left\| f - \sum_{k=1}^{n} \alpha_k \varphi_k \right\|$, by choosing

$$\alpha_k = \frac{\langle f, \varphi_k \rangle}{\|\varphi_k\|^2}.$$

This minimum is given by

$$\left\| f - \sum_{k=1}^{n} \frac{\langle f, \varphi_k \rangle}{\|\varphi_k\|^2} \varphi_k \right\|^2 = \|f\|^2 - \sum_{k=1}^{n} \frac{|\langle f, \varphi_k \rangle|^2}{\|\varphi_k\|^2} \geq 0. \tag{1.21}$$

This yields the relation,

$$\sum_{k=1}^{n} \frac{|\langle f, \varphi_k \rangle|^2}{\|\varphi_k\|^2} \leq \|f\|^2.$$

Since this relation is true for any $n$, it is also true in the limit as $n \to \infty$. The resulting inequality

$$\sum_{k=1}^{\infty} \frac{|\langle f, \varphi_k \rangle|^2}{\|\varphi_k\|^2} \leq \|f\|^2, \tag{1.22}$$

known as *Bessel's inequality*, holds for any orthogonal set $\{\varphi_k : k \in \mathbb{N}\}$ and any $f \in \mathcal{L}^2$.

In view of (1.21), Bessel's inequality becomes an equality if, and only if,

$$\left\| f - \sum_{k=1}^{\infty} \frac{\langle f, \varphi_k \rangle}{\|\varphi_k\|^2} \varphi_k \right\| = 0,$$

or, equivalently,

$$f = \sum_{k=1}^{\infty} \frac{\langle f, \varphi_k \rangle}{\|\varphi_k\|^2} \varphi_k \text{ in } \mathcal{L}^2,$$

which means that $f$ is represented in $\mathcal{L}^2$ by the sum $\sum_{k=1}^{\infty} \alpha_k \varphi_k$, where $\alpha_k = \langle f, \varphi_k \rangle / \|\varphi_k\|^2$.

## Definition 1.29

An orthogonal set $\{\varphi_n : n \in \mathbb{N}\}$ in $\mathcal{L}^2$ is said to be *complete* if, for any $f \in \mathcal{L}^2$,

$$\sum_{k=1}^{n} \frac{\langle f, \varphi_k \rangle}{\|\varphi_k\|^2} \varphi_k \overset{\mathcal{L}^2}{\to} f.$$

Thus a complete orthogonal set in $\mathcal{L}^2$ becomes a basis for the space, and because $\mathcal{L}^2$ is infinite-dimensional the basis has to be an infinite set. When Bessel's inequality becomes an equality, the resulting relation

$$\|f\|^2 = \sum_{n=1}^{\infty} \frac{|\langle f, \varphi_n \rangle|^2}{\|\varphi_n\|^2} \tag{1.23}$$

is called *Parseval's relation* or the *completeness relation.* The second term is justified by the following theorem, which is really a restatement of Definition 1.29.

## Theorem 1.30

An orthogonal set $\{\varphi_n : n \in \mathbb{N}\}$ is complete if, and only if, it satisfies Parseval's relation (1.23) for any $f \in \mathcal{L}^2$.

## Remark 1.31

1. Given any orthogonal set $\{\varphi_n : n \in \mathbb{N}\}$ in $\mathcal{L}^2$, we have shown that we obtain the best $\mathcal{L}^2$-approximation

$$\sum_{k=1}^{n} \alpha_k \varphi_k$$

of the function $f \in \mathcal{L}^2$ by choosing $\alpha_k = \langle f, \varphi_k \rangle / \|\varphi_k\|^2$, and this choice is independent of $n$. If $\{\varphi_n\}$ is complete then the equality $f = \sum_{n=1}^{\infty} \alpha_n \varphi_n$ holds in $\mathcal{L}^2$.

2. When the orthogonal set $\{\varphi_n\}$ is normalized to $\{\psi_k = \varphi_k / \|\varphi_k\|\}$, Bessel's inequality takes the form

$$\sum_{k=1}^{\infty} |\langle f, \psi_k \rangle|^2 \leq \|f\|^2 ,$$

and Parseval's relation becomes

$$\|f\|^2 = \sum_{n=1}^{\infty} |\langle f, \psi_n \rangle|^2 .$$

3. For any $f \in \mathcal{L}^2$, because $\|f\| < \infty$, we conclude from Bessel's inequality that $\langle f, \psi_n \rangle \to 0$ whether the orthonormal set $\{\psi_n\}$ is complete or not.

Parseval's relation may be regarded as a generalization of the theorem of Pythagoras from $\mathbb{R}^n$ to $\mathcal{L}^2$, where $\|f\|^2$ replaces the square of the length of the vector, and $\sum_{n=1}^{\infty} |\langle f, \psi_n \rangle|^2$ represents the sum of the squares of its projections on the orthonormal basis. That is one reason why $\mathcal{L}^2$ is considered the natural generalization of the finite-dimensional Euclidean space to infinite dimensions. It preserves some of the basic geometric structure of $\mathbb{R}^n$, and the completeness property (Theorem 1.26) guarantees its closure under limiting operations on Cauchy sequences.

## *EXERCISES*

1.56 If $l$ is any positive number, show that $\{\sin(n\pi x/l) : n \in \mathbb{N}\}$ and $\{\cos(n\pi x/l) : n \in \mathbb{N}_0\}$ are orthogonal sets in $\mathcal{L}^2(0, l)$. Determine the corresponding orthonormal sets.

1.57 Determine the coefficients $c_i$ in the linear combination

$$c_1 + c_2 \sin \pi x + c_3 \sin 2\pi x$$

which give the best approximation in $\mathcal{L}^2(0, 2)$ of the function $f(x) = x,\ 0 < x < 2$.

1.58 Determine the coefficients $a_i$ and $b_i$ in the linear combination

$$a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x$$

which give the best approximation in $\mathcal{L}^2(-\pi, \pi)$ of $f(x) = |x|$, $-\pi \leq x \leq \pi$.

1.59 Let $p_1, p_2$, and $p_3$ be the three orthogonal polynomials formed from the set $\{1, x, x^2\}$ by the Gram–Schmidt method, where $-1 \leq x \leq 1$. Determine the constant coefficients in the second-degree polynomial $a_1 p_1(x) + a_2 p_2(x) + a_3 p_3(x)$ which give the best approximation in $\mathcal{L}^2(-1, 1)$ of $e^x$. Can you think of another polynomial $p$ of degree 2 which approximates $e^x$ on $(-1, 1)$ in a different sense?

1.60 Assuming that

$$1 - x = \frac{8}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{(2n-1)^2} \cos \frac{(2n-1)\pi}{2} x, \quad 0 \leq x \leq 2,$$

use Parseval's identity to prove that

$$\pi^4 = 96 \sum_{n=1}^{\infty} \frac{1}{(2n-1)^4}.$$

1.61 Define a real sequence $(a_k)$ such that $\sum a_k^2$ converges and $\sum a_k$ diverges. What type of convergence can the series $\sum a_n \cos nx$, $-\pi \leq x \leq \pi$ have?

1.62 Suppose $\{f_n : n \in \mathbb{N}\}$ is an orthogonal set in $\mathcal{L}^2(0, l)$, and let

$$\varphi_n(x) = \frac{1}{2}[f_n(x) + f_n(-x)],$$
$$\psi_n(x) = \frac{1}{2}[f_n(x) - f_n(-x)], \quad -l \leq x \leq l,$$

be the even and odd extensions, respectively, of $f_n$ from $[0, l]$ to $[-l, l]$. Show that the set $\{\varphi_n\} \cup \{\psi_n\}$ is orthogonal in $\mathcal{L}^2(-l, l)$. If $\{f_n\}$ is orthonormal in $\mathcal{L}^2(0, l)$, what is the corresponding orthonormal set in $\mathcal{L}^2(-l, l)$?