BIRKHÄUSER

Applied and Numerical Harmonic Analysis

Functions, Spaces, and Expansions

Mathematical Tools in Physics and Engineering

Ole Christensen

BIRKHÄUSER

Applied and Numerical Harmonic Analysis

Ole Christensen

Functions, Spaces, and Expansions

Mathematical Tools in Physics and Engineering

Birkhäuser
Boston • Basel • Berlin

Ole Christensen
Technical University of Denmark
Department of Mathematics
2800 Lyngby
Denmark
Ole.Christensen@mat.dtu.dk

ISBN 978-0-8176-4979-1 e-ISBN 978-0-8176-4980-7
DOI 10.1007/978-0-8176-4980-7
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010928840

Mathematics Subject Classification (2010): 40-01, 41-01, 42-01, 46-01

Printed on acid-free paper

Birkhäuser is part of Springer Science+Business Media (www.birkhauser.com)

ANHA Series Preface

The *Applied and Numerical Harmonic Analysis (ANHA)* book series aims to provide the engineering, mathematical, and scientific communities with significant developments in harmonic analysis, ranging from abstract harmonic analysis to basic applications. The title of the series reflects the importance of applications and numerical implementation, but richness and relevance of applications and implementation depend fundamentally on the structure and depth of theoretical underpinnings. Thus, from our point of view, the interleaving of theory and applications and their creative symbiotic evolution is axiomatic.

Harmonic analysis is a wellspring of ideas and applicability that has flourished, developed, and deepened over time within many disciplines and by means of creative cross-fertilization with diverse areas. The intricate and fundamental relationship between harmonic analysis and fields such as signal processing, partial differential equations (PDEs), and image processing is reflected in our state-of-the-art *ANHA* series.

Our vision of modern harmonic analysis includes mathematical areas such as wavelet theory, Banach algebras, classical Fourier analysis, time-frequency analysis, and fractal geometry, as well as the diverse topics that impinge on them.

For example, wavelet theory can be considered an appropriate tool to deal with some basic problems in digital signal processing, speech and image processing, geophysics, pattern recognition, biomedical engineering, and turbulence. These areas implement the latest technology from sampling methods on surfaces to fast algorithms and computer vision methods.

The underlying mathematics of wavelet theory depends not only on classical Fourier analysis, but also on ideas from abstract harmonic analysis, including von Neumann algebras and the affine group. This leads to a study of the Heisenberg group and its relationship to Gabor systems, and of the metaplectic group for a meaningful interaction of signal decomposition methods. The unifying influence of wavelet theory in the aforementioned topics illustrates the justification for providing a means for centralizing and disseminating information from the broader, but still focused, area of harmonic analysis. This will be a key role of *ANHA*. We intend to publish with the scope and interaction that such a host of issues demands.

Along with our commitment to publish mathematically significant works at the frontiers of harmonic analysis, we have a comparably strong commitment to publish major advances in the following applicable topics in which harmonic analysis plays a substantial role:

Antenna theory	*Prediction theory*
Biomedical signal processing	*Radar applications*
Digital signal processing	*Sampling theory*
Fast algorithms	*Spectral estimation*
Gabor theory and applications	*Speech processing*
Image processing	*Time-frequency and time-scale analysis*
Numerical partial differential equations	*Wavelet theory*

The above point of view for the *ANHA* book series is inspired by the history of Fourier analysis itself, whose tentacles reach into so many fields.

In the last two centuries Fourier analysis has had a major impact on the development of mathematics, on the understanding of many engineering and scientific phenomena, and on the solution of some of the most important problems in mathematics and the sciences. Historically, Fourier series were developed in the analysis of some of the classical PDEs of mathematical physics; these series were used to solve such equations. In order to understand Fourier series and the kinds of solutions they could represent, some of the most basic notions of analysis were defined, e.g., the concept of "function." Since the coefficients of Fourier series are integrals, it is no surprise that Riemann integrals were conceived to deal with uniqueness properties of trigonometric series. Cantor's set theory was also developed because of such uniqueness questions.

A basic problem in Fourier analysis is to show how complicated phenomena, such as sound waves, can be described in terms of elementary harmonics. There are two aspects of this problem: first, to find, or even define properly, the harmonics or spectrum of a given phenomenon, e.g., the spectroscopy problem in optics; second, to determine which phenomena can be constructed from given classes of harmonics, as done, for example, by the mechanical synthesizers in tidal analysis.

Fourier analysis is also the natural setting for many other problems in engineering, mathematics, and the sciences. For example, Wiener's Tauberian theorem in Fourier analysis not only characterizes the behavior of the prime numbers, but also provides the proper notion of spectrum for phenomena such as white light; this latter process leads to the Fourier analysis associated with correlation functions in filtering and prediction problems, and these problems, in turn, deal naturally with Hardy spaces in the theory of complex variables.

Nowadays, some of the theory of PDEs has given way to the study of Fourier integral operators. Problems in antenna theory are studied in terms of unimodular trigonometric polynomials. Applications of Fourier analysis abound in signal processing, whether with the fast Fourier transform (FFT), or filter design, or the adaptive modeling inherent in time-frequency-scale methods such as wavelet theory. The coherent states of mathematical physics are translated and modulated Fourier transforms, and these are used, in conjunction with the uncertainty principle, for dealing with signal reconstruction in communications theory. We are back to the raison d'être of the *ANHA* series!

John J. Benedetto
Series Editor
University of Maryland
College Park

Contents

Preface

The purpose of this book is to present some mathematical tools that play key roles in mathematics as well as in applied mathematics, physics, and engineering. The treatment is mathematical in nature, and we do not go into concrete applications; but it is important to stress that all the considered topics are selected because they actually play a role outside pure mathematics. The hope is that the book will be useful for students in many fields of science and engineering, and professionals who want a deeper insight in some of the topics appearing in the scientific literature.

A central theme throughout the work is the structure of various vector spaces (most importantly, normed vector spaces and Hilbert spaces) and expansions of elements in these spaces in terms of bases. Particular attention is given to the space of square-integrable functions, $L^2(\mathbb{R})$.

The goal is twofold. Besides the interest in these subjects by themselves, the book will also contribute to a deeper understanding of several themes from calculus and linear algebra, because these themes appear here again and are tied together. For example, we discuss Fourier series in the correct setting of an expansion in a Hilbert space, similar to the one that is obtained via an orthonormal basis in $\mathbb{C}^n$.

Before we go into detail about the content of the book, let us spend a few lines on the prerequisites. We expect the reader to

- Have a profound understanding of linear algebra, as well in $\mathbb{R}^n$ and $\mathbb{C}^n$ as in general vector spaces;

- Be familiar with the basic concepts of calculus and real analysis, including (Riemann) integration and infinite series of real or complex numbers.

The core of the book is formed by Chapters 2–7. Chapter 1 is a survey on topics from elementary mathematics courses, and Chapters 8–11 describe concrete functions and settings where the key concepts treated in Chapters 2–7 play a central role. Each chapter ends with a collection of exercises.

Let us describe the content in more detail. Chapter 1 collects some basic results from linear algebra and calculus, e.g., concerning topology in $\mathbb{C}^n$ and continuity of functions. We expect the reader to be familiar with most of the topics in this chapter. All results are stated without proofs, but in many cases a guide to a proof can be found in the exercises.

Chapters 2 and 3 deal with particular types of vector spaces on an abstract level. The aim is a detailed mathematical description. All results are presented either with a proof or the proof left as an exercise. In the first of these chapters, Chapter 2, the key concept of a normed vector space is presented. We discuss linear operators on such spaces, and infinite series consisting of vectors in normed spaces. Chapter 3 deals with Banach spaces, in particular the sequence spaces $\ell^p(\mathbb{N})$, and operators hereon. Chapter 4 specializes in the important case of a norm arising from an inner product. This leads to the concept of a Hilbert space. We continue the analysis of linear operators initiated in Chapter 3, now with focus on results that are particular for Hilbert spaces. The key concept of an orthonormal basis is introduced.

While Chapters 2–4 are abstract in nature, Chapter 5 marks the beginning of a more concrete part of the book. There is still emphasis on the mathematical formalism. However, we do not insist on a complete treatment: we skip discussions of certain technical issues, and some results are presented without a proof.

In Chapter 5 we consider an important class of Banach spaces consisting of functions, the so-called L^p-spaces. Special emphasis is given to the space $L^1(\mathbb{R})$ and integration techniques on that space. Chapter 6 specializes in the case $p = 2$, which leads to a Hilbert space. We consider various operators on $L^2(\mathbb{R})$. We also consider L^2-spaces on an interval, and relate these spaces to Fourier analysis. Chapter 7 deals with the Fourier transform, convolution, and the sampling problem.

The final part of the book discusses special classes of functions that appear in many areas of applied mathematics and are related to the themes presented in the book. All the considered functions naturally lead to bases for $L^2(\mathbb{R})$ or subspaces hereof. Chapter 8 provides a short description of wavelet theory in $L^2(\mathbb{R})$, based on Fourier analysis. A more detailed analysis of the key tool in wavelet theory, multiresolution analysis, is given in Chapter 9. Chapter 10 introduces the important B-splines and their main properties. In Chapter 11 we consider certain functions (typically polynomials) that arise as solutions to various differential equations appearing, for example, in physics. It turns out that for special differential equations, the collection of some particular solutions form an orthonormal basis for a certain L^2-space; that brings us back to the main theme of the book.

Appendix A collects certain proofs that are particularly long and technical. Appendix B contains a list of the vector spaces considered in the book and their main properties, as well as a list of some of the special functions considered in Chapter 11.

For use in a course at the master's level, the natural starting point is Chapter 2; to the extent that the results in Chapter 1 are unknown, they can be presented section by section during the course whenever relevant. Depending on the anticipated content of the course, one can proceed in various ways. For a course focusing on Hilbert spaces, one can skip most of Chapter 3 (except the definition of a Banach space) and move directly to Chapters 4–7; on the other hand, a profound understanding of general Banach spaces requires inclusion of Chapter 3. Concrete manifestations of the abstract concept of an orthonormal basis appear in Chapters 8–10 (wavelets, in particular, based on B-splines) and Chapter 11 (orthonormal bases consisting of solutions to special differential equations).

The list of references contains articles and books at several levels. In order to be more informative, we have introduced the following ranking system to the references: **(A)** elementary; **(B)** undergraduate level; **(C)** graduate level; **(D)** research paper; **(H)** historical paper.

I would like to thank Robert Burckel and Christopher Heil for many constructive comments to an earlier version of the manuscript. Their help greatly improved the presentation. I also thank the many students at the Technical University of Denmark who helped me by finding print errors and spotting unclear formulations in the preliminary manuscripts I used in the spring semesters of 2008 and 2009. Finally, I thank the staff at Birkhäuser, especially Tom Grasso and Patrick Keene, for their help and careful copyediting of the manuscript.

Ole Christensen
Kgs. Lyngby, Denmark
January 2010

Prologue: Spaces and Expansions

In brief, the content of this book is captured by the two themes *spaces* and *expansions.* The purpose of this prologue is to relate the key topics to physics and engineering.

In engineering and signal processing, a *signal* means a *function* f, typically with the time as variable. For example, the signal might be the current running in the loudspeaker cable when a certain recording is played. Such a signal is shown in Figure 1. In order to extract relevant features in the signal, the signal is often considered in a *transformed domain:* in the example with the recording, if we want to extract information about the frequencies appearing in the signal, one would consider the *Fourier transform* of the signal. The Fourier transform of f is formally defined as the function

$$\hat{f}(\gamma) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\gamma}\, dx, \ \gamma \in \mathbb{R}. \qquad (1)$$

For the function in Figure 1, the absolute value of the Fourier transform is depicted in Figure 2. It shows that the signal has a large content of frequencies around 500 Hz, which is close to the frequency 440 Hz that is used to tune the instruments in an orchestra.

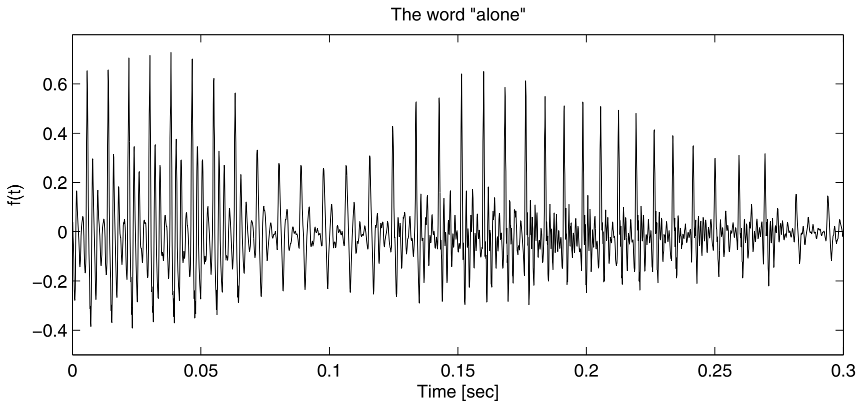

Figure 1. A speech signal. One can regard such a signal as the current in the cable to the loudspeaker when a recording of the speech is played. The actual signal is a recording of the word "alone".

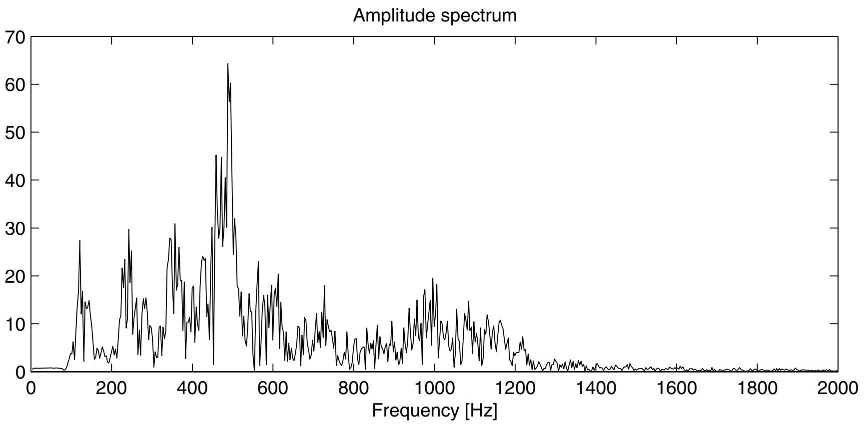

Figure 2. The absolute value of the Fourier transform of the signal in Figure 1.

However, the expression (1) only makes sense under certain restrictions on the function f. In other words: we have to specify the applicable signals. Mathematically, this is done by requiring all considered signals f to belong to certain *vector spaces:* in the concrete case discussed here, the relevant spaces are the (Banach) spaces $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$. These spaces play a central role in the book.

The concept *an expansion* is known from elementary linear algebra: it is merely another name for a representation of a vector $\mathbf{v}$ in a vector space in terms of a basis $\{\mathbf{e}_k\}$,

$$\mathbf{v} = \sum c_k \mathbf{e}_k.$$

We know that the choice of a convenient basis is crucial: for example, the representation of a linear operator might be very complicated with respect to some unfortunate bases, but very easy with respect to a well-chosen basis. One of the key concepts of the book is to discuss bases and expansions in infinite-dimensional vector spaces. Readers having knowledge about quantum mechanics and coherent states already know about series expansions in terms of eigenfunctions for certain differential equations; such expansions appear as special cases of the general theory presented in this book.

Already in Example 1.1.3 we will introduce the discrete Fourier transform basis for $\mathbb{C}^n$. Paying close attention to that example will help the reader to see the motivation behind much of the material to be presented later in the book.

1

Mathematical Background

This introductory chapter collects basic results from linear algebra and real analysis that will be needed throughout the book. The chapter has the status of a *tool box,* and the presentation is more compressed than in the rest of the book. None of the stated results are proved, but most of the definitions and key results are illustrated by examples. In many cases a guide to a proof is stated as an exercise.

Some basic concepts from linear algebra in $\mathbb{R}^n$ and $\mathbb{C}^n$ are reviewed in Section 1.1. The discrete Fourier transform is introduced as an example of an orthonormal basis in $\mathbb{C}^n$. The setting of abstract vector spaces is discussed in Section 1.2, and Section 1.3 deals with finite-dimensional vector spaces. Section 1.4 gives a short introduction to topology in $\mathbb{R}^n$, a subject that later is discussed in general normed spaces. Section 1.5 presents the concepts of supremum and infimum; they play a key role already in the subsequent Section 1.6 about continuity of functions. Section 1.7 states several useful inequalities related to integration and summation. Section 1.8 defines various special types of functions. Finally, Section 1.9 states the principles for proof by induction.

1.1 $\mathbb{R}^n$ and $\mathbb{C}^n$

Formally, the vector space $\mathbb{R}^n$ is defined as the set consisting of all sequences of n real numbers:

$$\mathbb{R}^n = \left\{(x_1, x_2, \ldots, x_n) \mid x_k \in \mathbb{R},\ k = 1, \ldots, n\right\}.$$

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4980-7_1,

Similarly,

$$\mathbb{C}^n = \{(x_1, x_2, \ldots, x_n) \mid x_k \in \mathbb{C},\ k = 1, \ldots, n\}.$$

Usually a vector belonging to $\mathbb{R}^n$ or $\mathbb{C}^n$ is written as

$$\mathbf{x} = (x_1, x_2, \ldots, x_n).$$

Whenever convenient we will also write the vector $\mathbf{x}$ as a column,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}.$$

We expect the reader to know the definition of inner products on these spaces (if not, look at the definition in general vector spaces in Definition 4.1.1). The canonical inner product in $\mathbb{R}^n$ is

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^{n} x_k y_k,\ \ \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \tag{1.1}$$

and the canonical inner product in $\mathbb{C}^n$ is

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^{n} x_k \overline{y_k},\ \ \mathbf{x}, \mathbf{y} \in \mathbb{C}^n. \tag{1.2}$$

The spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ can also be equipped with other inner products, but we will always use the inner products in (1.1) and (1.2). In both spaces, the *length* (later to be called the *norm*) of a vector $\mathbf{x}$ is

$$||\mathbf{x}|| = \left(\sum_{k=1}^{n} |x_k|^2 \right)^{1/2}. \tag{1.3}$$

We will focus on the vector space $\mathbb{C}^n$; the theory for $\mathbb{R}^n$ is parallel, except that we do not have the complex conjugation in the inner product.

In the theory for $\mathbb{C}^n$, the concept of a basis plays a crucial role. We state the formal definition:

Definition 1.1.1 (Basis in $\mathbb{C}^n$) *Consider a collection of vectors $\{\mathbf{e}_k\}_{k=1}^m$ in $\mathbb{C}^n$.*

(i) $\{\mathbf{e}_k\}_{k=1}^m$ is a basis for $\mathbb{C}^n$ if span$\{\mathbf{e}_k\}_{k=1}^m = \mathbb{C}^n$ and the vectors $\{\mathbf{e}_k\}_{k=1}^m$ are linearly independent.

(ii) $\{\mathbf{e}_k\}_{k=1}^m$ is an orthonormal basis for $\mathbb{C}^n$ if $\{\mathbf{e}_k\}_{k=1}^m$ is a basis and

$$\langle \mathbf{e}_k, \mathbf{e}_\ell \rangle := \delta_{k,\ell} = \begin{cases} 1 & \text{if } k = \ell, \\ 0 & \text{if } k \neq \ell. \end{cases}$$

It is well-known that any basis for $\mathbb{C}^n$ contains exactly $m = n$ elements. If $\{\mathbf{e}_k\}_{k=1}^n$ is a basis for $\mathbb{C}^n$, any vector $\mathbf{v} \in \mathbb{C}^n$ has a unique representation

$$\mathbf{v} = \sum_{k=1}^{n} c_k \mathbf{e}_k \tag{1.4}$$

for some scalars $c_k, \; k = 1, \ldots, n$; in case $\{\mathbf{e}_k\}_{k=1}^n$ is an orthonormal basis this representation takes the form (Exercise 1.1)

$$\mathbf{v} = \sum_{k=1}^{n} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k. \tag{1.5}$$

Example 1.1.2 (Canonical orthonormal basis for $\mathbb{C}^n$) Let the vectors $\{\mathbf{e}_k\}_{k=1}^n$ in $\mathbb{C}^n$ be defined by

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \; \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \\ 1 \end{pmatrix}.$$

Then $\{\mathbf{e}_k\}_{k=1}^n$ forms an orthonormal basis for $\mathbb{C}^n$ with respect to the canonical inner product. This basis is usually called the *canonical orthonormal basis for* $\mathbb{C}^n$. □

Let us present another orthonormal basis for $\mathbb{C}^n$, the *discrete Fourier transform basis*. An understanding of this basis will motivate our later study of Fourier series and the Fourier transform:

Example 1.1.3 (Discrete Fourier transform basis) For $k = 1, \ldots, n$, define the vectors $\mathbf{e}_k \in \mathbb{C}^n$ by

$$\mathbf{e}_k = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ e^{2\pi i(k-1)/n} \\ e^{4\pi i(k-1)/n} \\ \cdot \\ \cdot \\ e^{2\pi i(n-1)(k-1)/n} \end{pmatrix}. \tag{1.6}$$

That is, the ℓth coordinate of $\mathbf{e}_k$ is

$$(\mathbf{e}_k)_\ell = \frac{1}{\sqrt{n}} e^{2\pi i(\ell-1)(k-1)/n}, \;\; \ell = 1, \ldots, n.$$

We will prove that the vectors $\{\mathbf{e}_k\}_{k=1}^n$ defined by (1.6) constitute an orthonormal basis for $\mathbb{C}^n$. Since $\{\mathbf{e}_k\}_{k=1}^n$ are n vectors in an n-dimensional vector space, it is enough to prove that they constitute an orthonormal system; in fact, this implies that they are linearly independent and therefore span $\mathbb{C}^n$. Direct calculation reveals that $||\mathbf{e}_k|| = 1$ for all k. Now, given $k \neq j$, the definition of $\mathbf{e}_k$ and a change of summation index show that

$$\begin{aligned} \langle \mathbf{e}_k, \mathbf{e}_j \rangle &= \frac{1}{n} \sum_{\ell=1}^{n} e^{2\pi i(\ell-1)(k-1)/n} e^{-2\pi i(\ell-1)(j-1)/n} \\ &= \frac{1}{n} \sum_{\ell=0}^{n} e^{2\pi i\ell(k-1)/n} e^{-2\pi i\ell(j-1)/n} \\ &= \frac{1}{n} \sum_{\ell=0}^{n-1} e^{2\pi i\ell(k-j)/n}. \end{aligned}$$

Using the formula $(1-x)(1+x+\cdots+x^{n-1}) = 1-x^n$ with $x = e^{2\pi i(k-j)/n}$, we get

$$\langle \mathbf{e}_k, \mathbf{e}_j \rangle = \frac{1}{n} \frac{1-(e^{2\pi i(k-j)/n})^n}{1-e^{2\pi i(k-j)/n}} = 0,$$

as desired.

The basis $\{\mathbf{e}_k\}_{k=1}^n$ is called the *discrete Fourier transform basis.* Using this basis, every sequence $\mathbf{v} \in \mathbb{C}^n$,

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_n \end{pmatrix},$$

has the representation

$$\begin{aligned} \mathbf{v} &= \sum_{k=1}^{n} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k && (1.7) \\ &= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{\ell=1}^{n} v_\ell \, e^{-2\pi i(\ell-1)(k-1)/n} \mathbf{e}_k. && (1.8) \end{aligned}$$

Written out in coordinates, this means that the jth coordinate is

$$\begin{aligned} v_j &= \frac{1}{n} \sum_{k=1}^{n} \sum_{\ell=1}^{n} v_\ell \, e^{-2\pi i(\ell-1)(k-1)/n} e^{2\pi i(j-1)(k-1)/n} \\ &= \frac{1}{n} \sum_{k=1}^{n} \sum_{\ell=1}^{n} v_\ell \, e^{2\pi i(j-\ell)(k-1)/n}, \; j = 1, \ldots, n. \end{aligned}$$

The vector $\widehat{\mathbf{v}} := \{\langle \mathbf{v}, \mathbf{e}_k \rangle\}_{k=1}^n$ consisting of the coefficients in (1.7) is called the *discrete Fourier transform* (DFT) of $\mathbf{v}$. We return to the discrete Fourier transform in Section 7.5. □

From linear algebra we know many equivalent conditions for a set of vectors to constitute a basis for $\mathbb{C}^n$. Let us list the most important characterizations:

Theorem 1.1.4 (Characterization of basis for $\mathbb{C}^n$) *Consider n vectors in $\mathbb{C}^n$,*

$$\mathbf{e}_1 = \begin{pmatrix} e_{11} \\ e_{21} \\ \cdot \\ \cdot \\ e_{n1} \end{pmatrix}, \ \mathbf{e}_2 = \begin{pmatrix} e_{12} \\ e_{22} \\ \cdot \\ \cdot \\ e_{n2} \end{pmatrix}, \dots, \mathbf{e}_n = \begin{pmatrix} e_{1n} \\ e_{2n} \\ \cdot \\ \cdot \\ e_{nn} \end{pmatrix},$$

and write them as columns in an $n \times n$ matrix,

$$E = \begin{pmatrix} e_{11} & e_{12} & \cdot & \cdot & e_{1n} \\ e_{21} & e_{22} & \cdot & \cdot & e_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ e_{n1} & e_{n2} & \cdot & \cdot & e_{nn} \end{pmatrix}.$$

Then the following are equivalent:

(i) The columns in E (i.e., the given vectors) constitute a basis for $\mathbb{C}^n$.

(ii) The rows in E constitute a basis for $\mathbb{C}^n$.

(iii) The determinant of E is nonzero.

(iv) E is invertible.

(v) E defines an injective mapping from $\mathbb{C}^n$ into $\mathbb{C}^n$.

(vi) E defines a surjective mapping from $\mathbb{C}^n$ onto $\mathbb{C}^n$.

(vii) The columns in E are linearly independent.

(viii) E has rank equal to n.

Example 1.1.5 (Basis for $\mathbb{C}^2$) Consider the matrix

$$E = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

The operator associated with E acts on $\mathbb{R}^2$ by matrix multiplication. That is, for $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$,

$$E\mathbf{x} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{pmatrix}.$$

Since $\det(E) = -2$, the equivalent conditions stated in Theorem 1.1.4 are satisfied. Thus, the columns in the matrix E form a basis for $\mathbb{C}^2$. □

1.2 Abstract vector spaces

All the central concepts of linear algebra can be extended to *abstract vector spaces*, from now on just called *vector spaces.* A vector space is a nonempty set V that has been equipped with two operations, called *addition* and *scalar multiplication*, satisfying certain rules. The operation"addition" associates to each pair of elements $\mathbf{v}, \mathbf{w} \in V$ an element in V that will be denoted by $\mathbf{v}+\mathbf{w}$. The operation "scalar multiplication" associates to each $\mathbf{v} \in V$ and each $\alpha \in \mathbb{C}$ an element in V that will be denoted by $\alpha\mathbf{v}$ or $\mathbf{v}\alpha$. These operations have to satisfy the following requirements:

Definition 1.2.1 (Vector space) *Consider a nonempty set V, equipped with operations of addition and scalar multiplication. Assume that the following rules are satisfied:*

(i) *For all $\mathbf{v}, \mathbf{w} \in V$, we have that $\mathbf{v}+\mathbf{w} = \mathbf{w}+\mathbf{v}$;*

(ii) *For all $\mathbf{v}, \mathbf{w}, \mathbf{u} \in V$, we have that $(\mathbf{v}+\mathbf{w})+\mathbf{u} = \mathbf{v}+(\mathbf{w}+\mathbf{u})$;*

(iii) *There exists an element, called $\mathbf{0}$, in V, such that for all $\mathbf{v} \in V$,*

$$\mathbf{v}+\mathbf{0} = \mathbf{v};$$

(iv) *For each $\mathbf{v} \in V$ there exists an element, called $-\mathbf{v}$, in V, with the property that*

$$\mathbf{v}+(-\mathbf{v}) = \mathbf{0};$$

(v) *For all $\alpha, \beta \in \mathbb{C}$ and all $\mathbf{v} \in V$,*

$$\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v};$$

(vi) *For all $\alpha, \beta \in \mathbb{C}$ and all $\mathbf{v} \in V$,*

$$(\alpha+\beta)\mathbf{v} = \alpha\mathbf{v}+\beta\mathbf{v};$$

(vii) *For all $\alpha \in \mathbb{C}$ and all $\mathbf{v}, \mathbf{w} \in V$,*

$$\alpha(\mathbf{v}+\mathbf{w}) = \alpha\mathbf{v}+\alpha\mathbf{w};$$

(viii) *For all $\mathbf{v} \in V$,*

$$1\mathbf{v} = \mathbf{v}.$$

In that case we say that V, equipped with the operations of addition and scalar multiplication, forms a vector space.

It is immediate to check that the sets $\mathbb{R}^n$ and $\mathbb{C}^n$, equipped with the usual operations of addition and scalar multiplication, satisfy the conditions in Definition 1.2.1.

Because the scalar multiplication in Definition 1.2.1 is allowed to be by complex numbers, one frequently speaks about a *complex vector space.* A *real vector space* is defined the same way, except that the scalars $\mathbb{C}$ are replaced by the real numbers $\mathbb{R}$. Except $\mathbb{R}^n$, all vector spaces appearing in this book are complex. For this reason, all definitions will be formulated for the complex case.

Example 1.2.2 (Functions on a set A) Let A denote an arbitrary nonempty set, and let V denote the collection of all functions $f : A \to \mathbb{C}$. Given functions $f, g \in V$, we define the function $f + g \in V$ by

$$(f + g)(x) := f(x) + g(x), \ x \in A.$$

Also, given $f \in V$ and $\alpha \in \mathbb{C}$, define the function $\alpha f \in V$ by

$$(\alpha f)(x) := \alpha f(x), \ x \in A.$$

Direct verification shows that V equipped with these definitions of addition and scalar multiplication satisfies all the conditions in Definition 1.2.1. Thus, the set V forms a (complex) vector space. □

Example 1.2.3 (Polynomials on $\mathbb{R}$) Let W denote the set of polynomials on $\mathbb{R}$ of degree at most N for some $N \in \mathbb{N}$. That is, the elements in W have the form

$$P(x) = a_N x^N + a_{N-1} x^{N-1} + \cdots + a_0, \ x \in \mathbb{R},$$

for some scalar coefficients $a_0, a_1, \ldots, a_N \in \mathbb{C}$. Given a polynomial P of that form and another polynomial Q of the same form,

$$Q(x) = b_N x^N + b_{N-1} x^{N-1} + \cdots + b_0,$$

we define the polynomial $P + Q$ by

$$(P + Q)(x) = (a_N + b_N) x^N + (a_{N-1} + b_{N-1}) x^{N-1} + \cdots + (a_0 + b_0).$$

Also, for $\alpha \in \mathbb{C}$, we define the polynomial αP by

$$(\alpha P)(x) = \alpha a_N x^N + \alpha a_{N-1} x^{N-1} + \cdots + \alpha a_0.$$

We see that $P + Q \in W$ and that $\alpha P \in W$; furthermore, direct verification shows that all the conditions in Definition 1.2.1 are satisfied. Thus, with our definitions of addition and scalar multiplication, the set W forms a (complex) vector space. □

Given a collection of vectors $\{\mathbf{v}_k\}_{k=1}^N$ in a (complex) vector space V, a *linear combination* is a vector of the form

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_N \mathbf{v}_N \tag{1.9}$$

for some $\alpha_1, \alpha_2, \ldots, \alpha_N \in \mathbb{C}$. If we put $\alpha_1 = \alpha_2 = \cdots = \alpha_N = 0$, then the linear combination (1.9) yields the zero vector. In case no other linear combination yields the zero vector, we say that the vectors $\{\mathbf{v}_k\}_{k=1}^N$ are *linearly independent:*

Definition 1.2.4 (Linear independence) *Let* $\{\mathbf{v}_k\}_{k=1}^N$ *be a collection of vectors in* V*. If*

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_N\mathbf{v}_N = \mathbf{0} \Rightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_N = 0,$$

then $\{\mathbf{v}_k\}_{k=1}^N$ *are linearly independent; if not, the vectors are linearly dependent.*

Example 1.2.5 (Linear independence of polynomials) Consider the vector space V in Example 1.2.3. Note that the polynomials

$$1, x, \ldots, x^N \tag{1.10}$$

belong to V. Now, assume that for some $\alpha_0, \alpha_1, \ldots, \alpha_N \in \mathbb{C}$, we have that

$$\alpha_0 + \alpha_1 x + \cdots + \alpha_N x^N = 0 \text{ for all } x \in \mathbb{R}.$$

Since a nontrivial polynomial of degree N at most can have N roots, this implies that

$$\alpha_0 = \alpha_1 = \cdots = \alpha_N = 0.$$

Thus, the polynomials in (1.10) are linearly independent. □

Often, we encounter subsets of vector spaces having themselves the structure of a vector space:

Definition 1.2.6 (Subspace) *Let* V *be a vector space. A subset* $W \subseteq V$ *which itself is a vector space (when equipped with the operations of addition and scalar multiplication in* V*), is called a subspace of* V*.*

The set just containing the element $\mathbf{0}$ is always a subspace of V. Furthermore, V is a subspace of itself. Any subspace W for which $W \neq \mathbf{0}$ and $W \neq V$, is called a *nontrivial* subspace.

In practice, one often verifies that a subset W of V is a subspace via the following lemma. The reader is asked to provide the proof in Exercise 1.4.

Lemma 1.2.7 (Characterization of subspace) *A nonempty subset* W *of a vector space* V *is a subspace of* V *if and only if*

$$\alpha\mathbf{v} + \beta\mathbf{w} \in W \textit{ for all } \mathbf{v}, \mathbf{w} \in W,\ \alpha, \beta \in \mathbb{C}. \tag{1.11}$$

Example 1.2.8 (Subspace of polynomials) Let V denote the vector space consisting of all functions $f : \mathbb{R} \to \mathbb{C}$, i.e., V is the space in Example 1.2.2 with $A := \mathbb{R}$. The space W considered in Example 1.2.3 is a subset of V. Furthermore, any linear combination of two polynomials of degree at most N is again a polynomial of degree at most N, i.e., (1.11) holds. Thus, W is a (nontrivial) subspace of V. □

1.3 Finite-dimensional vector spaces

Let V be a (complex) vector space. Given a collection of vectors $\{\mathbf{v}_k\}_{k=1}^N$ of vectors in V, we define the *span* of the vectors as the set of all linear combinations:

$$\text{span}\{\mathbf{v}_k\}_{k=1}^N := \{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_N\mathbf{v}_N \mid \alpha_1, \alpha_2, \ldots, \alpha_N \in \mathbb{C}\}.$$

In general, $\text{span}\{\mathbf{v}_k\}_{k=1}^N$ will be a nontrivial subspace of V. However, for special choices of the vectors $\{\mathbf{v}_k\}_{k=1}^N$ it might happen that the span of the vectors equals V. This leads to a definition:

Definition 1.3.1 (Dimension) *A vector space V has dimension N, $N \in \mathbb{N}$, if there exists a collection of linearly independent vectors $\{\mathbf{v}_k\}_{k=1}^N$ such that*

$$V = \textit{span}\{\mathbf{v}_k\}_{k=1}^N.$$

A vector space, for which the condition in Definition 1.3.1 is satisfied for some number $N \in \mathbb{N}$, is said to be *finite-dimensional;* otherwise, the vector space is *infinite-dimensional.*

We now define the concept of a *basis* in a finite-dimensional vector space. In Section 2.5 we treat the infinite-dimensional case.

Definition 1.3.2 (Basis in finite-dimensional vector space) *A collection of vectors $\{\mathbf{e}_k\}_{k=1}^n$ in V is a basis for V if $\textit{span}\{\mathbf{e}_k\}_{k=1}^n = V$ and the vectors $\{\mathbf{e}_k\}_{k=1}^n$ are linearly independent.*

Example 1.3.3 (Basis for vector space of polynomials) The vector space V in Example 1.2.3 is finite-dimensional. In fact, as we have seen in Example 1.2.5, the polynomials

$$1, x, x^2, \ldots, x^N \tag{1.12}$$

are linearly independent, and each polynomial P of degree at most N is a linear combination of these, i.e.,

$$V = \text{span}\{1, x, \ldots, x^N\}.$$

The argument shows that V has dimension $N+1$, and that the vectors in (1.12) form a basis for V. □

In this book we will mainly consider vector spaces consisting of functions — and most of them will actually be infinite-dimensional.

1.4 Topology in $\mathbb{R}^n$

As we have seen in (1.3), the *length* (later to be called the *norm*) of a vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is

$$||\mathbf{x}|| = \left(\sum_{k=1}^{n} |x_k|^2 \right)^{1/2}.$$

Given a point $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, the *ball* centered at $\mathbf{x}$ and with radius $r > 0$ is defined as the set

$$\begin{aligned} B(\mathbf{x}, r) &:= \{\mathbf{y} \in \mathbb{R}^n \mid ||\mathbf{y} - \mathbf{x}|| < r\} \\ &= \left\{\mathbf{y} \in \mathbb{R}^n \mid \left(\sum_{k=1}^{n} |y_k - x_k|^2 \right)^{1/2} < r \right\}. \end{aligned}$$

Definition 1.4.1 (Open and closed sets in $\mathbb{R}^n$) *Consider a subset U of $\mathbb{R}^n$.*

(i) The subset U is open if for each $\mathbf{x} \in U$ there exists a number $\delta > 0$ such that $B(\mathbf{x}, \delta) \subseteq U$.

(ii) The complement of the subset U is defined as the set

$$U^c := \mathbb{R}^n \setminus U.$$

(iii) The subset U is said to be closed if the complement U^c is open.

(iv) The closure of U, to be denoted $\overline{U}$, is the smallest closed set in $\mathbb{R}^n$ that contains U.

Note that it is easy to find sets in $\mathbb{R}^n$ that are neither open nor closed. Intuitively, the closure of a subset of $\mathbb{R}^n$ is obtained by adding the "boundary":

Example 1.4.2 (Subsets of $\mathbb{R}$ and $\mathbb{R}^2$) We consider some subsets of $\mathbb{R}$ and $\mathbb{R}^2$:

(i) The subset $]-1, 1[$ of $\mathbb{R}$ is open; the closure of the set is $[-1, 1]$.

(ii) The subset $[-1, 1[$ of $\mathbb{R}$ is neither open nor closed; its closure is the set $[-1, 1]$.

(iii) The set $]0, 1[\times]4, 7[$ is open in $\mathbb{R}^2$; its closure is $[0, 1] \times [4, 7]$.

(iv) The set $[0, 1[\times]4, 7[$ is neither open nor closed in $\mathbb{R}^2$; its closure is $[0, 1] \times [4, 7]$. □

1.5 Supremum and infimum

We now introduce the concepts supremum and infimum for subsets of $\mathbb{R}$.

Definition 1.5.1 (Supremum) *Consider a subset E of $\mathbb{R}$.*

(i) E is bounded above if there exists a number $\beta \in \mathbb{R}$ such that

$$x \leq \beta, \ \forall x \in E. \tag{1.13}$$

(ii) If E is bounded above, the smallest number β satisfying (1.13) is called the supremum of E, and is written using one of the following three equivalent notations:

$$\sup E = \sup_{x \in E} x = \sup\{ \ x \mid x \in E\}.$$

(iii) For a set E that is not bounded above, we put $\sup E = \infty$.

Definition 1.5.2 (Infimum) *Consider a subset E of $\mathbb{R}$.*

(i) E is bounded below if there exists a number $\alpha \in \mathbb{R}$ such that

$$\alpha \leq x, \ \forall x \in E. \tag{1.14}$$

(ii) If E is bounded below, the largest number α satisfying (1.14) is called the infimum of E, and is written using one of the following three equivalent notations:

$$\inf E = \inf_{x \in E} x = \inf\{ \ x \mid x \in E\}.$$

(iii) For a set E that is not bounded below, we put $\inf E = -\infty$.

Example 1.5.3 (Supremum and infimum) By inspection, we see that

$$\sup]0,4] = 4, \quad \sup[-2,5[= 5, \quad \inf]-2,1] = -2, \quad \inf(\mathbb{Q} \cap [\pi, 7]) = \pi. \quad \square$$

In the particular case of a function f defined on a set A and taking real values, $f : A \to \mathbb{R}$, we can consider the set

$$E := \{ \ f(x) \mid x \in A\}. \tag{1.15}$$

The set A is called the *domain* of the function f, and E is the *range* or *image.* Now,

$$\sup E = \sup_{x \in A} f(x) = \sup\{ \ f(x) \mid x \in A\}.$$

A warning is in order. In general, the number $\sup_{x \in A} f(x)$ does not need to be a *function value* for the function f; that is, there might not exist an $x_0 \in A$ such that

$$f(x_0) = \sup_{x \in A} f(x). \tag{1.16}$$

In case an $x_0 \in A$ satisfying (1.16) exists, we write

$$\max_{x\in A} f(x) = \sup_{x\in A} f(x).$$

The expression $\max_{x\in A} f(x)$ is only used if the supremum's value appears as a function value.

Similarly, still under the assumption that f is real valued and with the set E as in (1.15), we have that

$$\inf E = \inf_{x\in A} f(x) = \inf\{\, f(x) \mid x \in A\}.$$

In case there exists an $x_0 \in A$ such that

$$f(x_0) = \inf_{x\in A} f(x),$$

we write

$$\min_{x\in A} f(x) = \inf_{x\in A} f(x).$$

Example 1.5.4 (Supremum and infimum for functions)

(i) Put $f(x) = x^2,\ x \in [0, 2]$; then

$$\sup_{x\in[0,2]} f(x) = 4 = f(2).$$

The supremum value is attained, so

$$\sup_{x\in[0,2]} f(x) = \max_{x\in[0,2]} f(x).$$

(ii) Put $f(x) = x^2,\ x \in [0, 2[$; then

$$\sup_{x\in[0,2[} f(x) = 4.$$

The supremum value is not attained.

(iii) Put $f(x) = x - x^2 = x(1 - x),\ x \in]0, 1[$; then

$$\sup_{x\in]0,1[} f(x) = \frac{1}{4} = f(\frac{1}{2}).$$

The supremum value is attained, so

$$\sup_{x\in]0,1[} f(x) = \max_{x\in]0,1[} f(x).$$

□

The concept of supremum is also used for sequences, e.g., indexed by $\mathbb{N}$; in fact, they appear as special cases of the above by letting $A = \mathbb{N}$. Let us consider such a case:

Example 1.5.5 (Supremum for a sequence) Direct inspection shows that

$$\sup_{n\in\mathbb{N}} \frac{n^2+4}{n^2+121} = 1;$$

the supremum is not attained. □

For any sequence $\{c_k\}_{k=1}^{\infty}$ of real numbers, the sequence $\{d_k\}_{k=1}^{\infty}$ given by

$$d_k := \inf_{n\geq k} c_n = \inf\{c_k, c_{k+1}, \dots\}$$

is increasing. If the sequence $\{c_k\}_{k=1}^{\infty}$ is bounded above, this implies that $\lim_{k\to\infty} d_k$ exists. In case $\{c_k\}_{k=1}^{\infty}$ is unbounded, we put $\lim_{k\to\infty} d_k = \infty$. In any case, $\lim_{k\to\infty} d_k$ is called *limes inferior,* or *lim inf* of the given sequence:

Definition 1.5.6 (lim inf and lim sup) *Given a sequence* $\{c_k\}_{k=1}^{\infty}$ *of real numbers, define*

$$\liminf_{k\to\infty} c_k := \lim_{k\to\infty}\left(\inf_{n\geq k} c_n\right).$$

Similarly, define

$$\limsup_{k\to\infty} c_k := \lim_{k\to\infty}\left(\sup_{n\geq k} c_n\right).$$

In case $\limsup_{k\to\infty} c_k$ and $\liminf_{k\to\infty} c_k$ are real numbers, they have the property that infinitely many elements from the sequence $\{c_k\}_{k=1}^{\infty}$ are located in arbitrarily small neighborhoods around these numbers. In order to give an exact description of this, we need the concept *accumulation point.*

Definition 1.5.7 (Accumulation point) *Let* $\{c_k\}_{k=1}^{\infty}$ *be a sequence of real numbers. A number* $c \in \mathbb{R}$ *is an accumulation point for* $\{c_k\}_{k=1}^{\infty}$ *if the set*

$$\{k \in \mathbb{N} \;\big|\; c_k \in [c-\epsilon, c+\epsilon]\}$$

is infinite for all $\epsilon > 0$.

Definition 1.5.8 (Bounded set, compact set) *Consider a subset* E *of* $\mathbb{R}$ *or* $\mathbb{C}$.

(i) E *is bounded if* E *is bounded above and below, i.e., if there exists a number* $\alpha > 0$ *such that*

$$|x| \leq \alpha, \;\forall x \in E.$$

(ii) E *is compact if* E *is bounded and closed.*

In particular, a sequence $\{c_k\}_{k=1}^{\infty}$ consisting of real or complex numbers is bounded if there exists a constant $C > 0$ such that

$$|c_k| \leq C, \ \forall k \in \mathbb{N}.$$

Any bounded sequence of real numbers has at least one accumulation point:

Lemma 1.5.9 (lim inf and lim sup are accumulation points) *Let $\{c_k\}_{k=1}^{\infty}$ be a bounded sequence of real numbers. Then the following hold:*

(i) $\liminf_{k\to\infty} c_k$ *and* $\limsup_{k\to\infty} c_k$ *are accumulation points for* $\{c_k\}_{k=1}^{\infty}$.

(ii) If $\lim_{k\to\infty} c_k$ *exists, then*

$$\liminf_{k\to\infty} c_k = \lim_{k\to\infty} c_k = \limsup_{k\to\infty} c_k.$$

We ask the reader to prove Lemma 1.5.9 in Exercise 1.11.

Given a bounded sequence $\{c_k\}_{k=1}^{\infty}$ of real numbers, put

$$c := \limsup_{k\to\infty} c_k.$$

Given any $\epsilon > 0$, Lemma 1.5.9 implies that for all $N \in \mathbb{N}$ there exists a $k > N$ such that

$$|c - c_k| \leq \epsilon.$$

In particular, there exists a $k_1 \in \mathbb{N}$ such that

$$|c - c_{k_1}| \leq \frac{1}{2};$$

and there exists a $k_2 > k_1$ such that

$$|c - c_{k_2}| \leq \frac{1}{4}.$$

Continuing this way, we obtain a sequence

$$\cdots > k_n > k_{n-1} > \cdots > k_2 > k_1 \geq 1$$

such that for any $n \in \mathbb{N}$,

$$|c - c_{k_n}| \leq \frac{1}{2^n}.$$

The sequence $\{c_{k_n}\}_{n=1}^{\infty}$ is said to be a *subsequence* of $\{c_k\}_{k=1}^{\infty}$. By definition,

$$c_{k_n} \to c \text{ as } n \to \infty.$$

This proves the so-called *Bolzano–Weierstrass Lemma:*

Lemma 1.5.10 (Bolzano–Weierstrass lemma) *Every bounded sequence* $\{c_k\}_{k=1}^{\infty}$ *of real numbers has a convergent subsequence.*

Example 1.5.11 (Convergent subsequence) Let

$$c_k := (-1)^k \frac{k+1}{k+100}, \ k \in \mathbb{N}.$$

The sequence $\{c_k\}_{k=1}^{\infty}$ does not converge. But the subsequence

$$\{c_{k_n}\}_{n=1}^{\infty} := \{c_{2n}\}_{n=1}^{\infty} = \left\{ \frac{2n+1}{2n+100} \right\}_{n=1}^{\infty}$$

is convergent. □

The concepts lim inf and lim sup are also applied to functions: given a sequence of real-valued functions

$$f_k : A \to \mathbb{R}, \ k = 1, 2, \dots,$$

we define the functions $\liminf_{k\to\infty} f_k$ and $\limsup_{k\to\infty} f_k$ by

$$\liminf_{k\to\infty} f_k(x) = \lim_{k\to\infty} \left(\inf_{n\geq k} f_n(x) \right), \ x \in A,$$

and

$$\limsup_{k\to\infty} f_k(x) := \lim_{k\to\infty} \left(\sup_{n\geq k} f_n(x) \right), \ x \in A.$$

Note that with our definitions of supremum and infimum, the functions $\liminf_{k\to\infty} f_k$ and $\limsup_{k\to\infty} f_k$ might assume the values $\pm\infty$.

At a few instances we will use the following inequality, which is a discrete version of the so-called *Fatou's Lemma*:

Lemma 1.5.12 (Fatou's lemma) *Let* $f_k : \mathbb{N} \to [0, \infty[, \ k \in \mathbb{N}$, *be a sequence of functions. Then*

$$\sum_{j=1}^{\infty} \liminf_{k\to\infty} f_k(j) \leq \liminf_{k\to\infty} \sum_{j=1}^{\infty} f_k(j), \ \forall j \in \mathbb{N}.$$

We guide the reader through a proof of Lemma 1.5.12 in Exercise 1.12.

1.6 Continuity of functions on $\mathbb{R}$

In this section we consider functions f defined on $\mathbb{R}$ or subintervals hereof, and taking values in $\mathbb{R}$ or $\mathbb{C}$. First we state an exact definition of the concept of continuity, as well as the related definition of uniform continuity.

Definition 1.6.1 (Continuity and uniform continuity) *Let $I \subseteq \mathbb{R}$ be an interval, and consider a function $f : I \to \mathbb{C}$.*

(i) The function f is continuous at the point $x_0 \in I$ if for each $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - f(x_0)| < \epsilon \text{ for all } x \in I \text{ for which } |x - x_0| < \delta. \qquad (1.17)$$

(ii) The function f is continuous if f is continuous at every point in I.

(iii) The function f is uniformly continuous if for each $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - f(y)| < \epsilon \text{ for all } x, y \in I \text{ for which } |x - y| < \delta. \qquad (1.18)$$

Observe the difference between (1.17) and (1.18): in the first case x_0 is a *fixed* point in the interval I, while in the second case we allow x as well as y to vary. Therefore, uniform continuity is a more restrictive condition than continuity.

We state some of the important properties for continuous functions on bounded and closed intervals:

Theorem 1.6.2 (Uniform continuity) *A continuous function on a bounded and closed interval $[a, b]$ is uniformly continuous.*

A function f defined on a set A, $f : A \to \mathbb{C}$, is *bounded* if the range

$$E := \{f(x) \mid x \in A\}$$

is bounded, i.e., if there exists a constant $K > 0$ such that

$$|f(x)| \leq K, \ \forall x \in A.$$

If no such constant K exists, the function f is *unbounded.*

Under certain conditions on the domain, a continuous function is bounded. We ask the reader to provide the proof of the following result in Exercise 1.13:

Theorem 1.6.3 (Continuous functions on $[a, b]$) *Consider a continuous real-valued function f defined on a bounded and closed interval $[a, b]$. Then the following hold:*

(i) f is bounded.

(ii) f attains its supremum, i.e., there exists $x_0 \in [a, b]$ such that

$$f(x_0) = \sup \{f(x) \mid x \in [a, b]\}.$$

(iii) f attains its infimum.

In Theorems 1.6.2 and 1.6.3 it is assumed that the function f is defined on a bounded and closed interval. The conclusions might fail if this hypothesis is removed:

Example 1.6.4 We illustrate the necessity of the hypotheses in Theorems 1.6.2 and 1.6.3:

(i) Let $f(x) = x^{-1}, \ x \in]0,1[$. The interval $]0,1[$ is bounded, but not closed. The function f is continuous, but not uniformly continuous. The function is unbounded.

(ii) Let $f(x) = x^2, \ x \in [0,\infty[$. The interval $[0,\infty[$ is closed, but not bounded. The function f is continuous, but not uniformly continuous. The function is unbounded. □

We will now consider a sequence of functions $\{f_k\}_{k=1}^{\infty}$, all of them defined on an interval I. Related to such a sequence one can introduce various types of convergence:

Definition 1.6.5 (Pointwise convergence, uniform convergence) *Let $\{f_k\}_{k=1}^{\infty}$ be a sequence of functions defined on an interval I and $f: I \to \mathbb{C}$ a given function.*

(i) If for each $x \in I$ and each $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$|f(x) - f_k(x)| < \epsilon$$

for all $k \geq N$, then $\{f_k\}_{k=1}^{\infty}$ is said to converge pointwise to f.

(ii) If for each $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$\sup_{x \in I} |f(x) - f_k(x)| < \epsilon$$

for all $k \geq N$, then $\{f_k\}_{k=1}^{\infty}$ is said to converge uniformly to f.

Note that the difference between the two types of convergence is rather subtle:

- In order to check pointwise convergence, we *fix* x and ask for $f_k(x)$ being close to $f(x)$ for large values of k;
- In order to check uniform convergence, we ask for $f_k(x)$ being close to $f(x)$ for large values of k, *simultaneously for all* x.

Note also that if (i) in Definition 1.6.5 holds, then necessarily

$$f(x) = \lim_{k \to \infty} f_k(x), \ x \in I.$$

The limit of a sequence of continuous functions might not be continuous itself. On the other hand, the limit of a uniformly convergent sequence of continuous functions is continuous:

Theorem 1.6.6 (Continuity of uniform limit) *Assume that $\{f_k\}_{k=1}^{\infty}$ is a sequence of continuous functions defined on an interval I. If $\{f_k\}_{k=1}^{\infty}$ converges uniformly to a function $f: I \to \mathbb{C}$, then f is continuous.*

We guide the reader through a proof of Theorem 1.6.6 in Exercise 1.15.

For later use we state the definition of *piecewise continuous functions* formally:

Definition 1.6.7 (Piecewise continuous function) *Let $I \subseteq \mathbb{R}$ be an interval. A function $f : I \to \mathbb{C}$ is piecewise continuous if the interval I can be split into a finite collection of subintervals on which f is continuous.*

We note that the results for continuous functions considered in this section do not extend to piecewise continuous functions. The following example illustrates this.

Example 1.6.8 (Piecewise continuous function) The function

$$f : [-1, 3] \to \mathbb{R}, \quad f(x) := \begin{cases} 0 & \text{if } x \in [-1, 0], \\ x^{-1} & \text{if } x \in]0, 2], \\ x & \text{if } x \in]2, 3], \end{cases}$$

is piecewise continuous. However, f is unbounded and not uniformly continuous. We ask the reader to prove these claims in Exercise 1.14. □

1.7 Integration and summation

The (Riemann) integral allows us to integrate scalar-valued piecewise continuous functions over closed and bounded intervals. For certain functions f we can also perform integration over unbounded intervals. For example, assume that the function f is defined on $[0, \infty[$ and that the expression

$$\int_0^\alpha f(x)\,dx$$

has a limit as $\alpha \to \infty$. Then the *improper Riemann integral*

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

is defined by

$$\int_0^\infty f(x)\,dx := \lim_{\alpha \to \infty} \int_0^\alpha f(x)\,dx.$$

Similarly, assuming that the limits

$$\lim_{\alpha \to \infty} \int_0^\alpha f(x)\,dx \quad \text{and} \lim_{\beta \to -\infty} \int_\beta^0 f(x)\,dx$$

exist, we define

$$\int_{-\infty}^{\infty} f(x)\,dx := \lim_{\alpha\to\infty}\int_0^{\alpha} f(x)\,dx + \lim_{\beta\to-\infty}\int_{\beta}^{0} f(x)\,dx.$$

In this section we collect some basic inequalities concerning integrals. Three types of integrals will appear in this book:

(i) The (Riemann) integral, concerning integration of bounded and piecewise continuous functions over bounded intervals;

(ii) The improper *Riemann integral*, concerning integration of piecewise continuous functions over unbounded domains;

(iii) The Lebesgue integral, to be discussed in Section 5.2.

The following results hold in all the settings (i)–(iii). We do not specify these assumptions on the functions and the intervals in the statements of the results.

Theorem 1.7.1 (Hölder's inequality and Minkowski's inequality) *Let $I \subseteq \mathbb{R}$ be an interval and consider functions $f, g : I \to \mathbb{C}$. Then the following inequalities hold:*

(i) (Hölder's inequality) For any numbers $p, q \in]1,\infty[$ with $1/p + 1/q = 1$,

$$\int_I |f(x)g(x)|\,dx \le \left(\int_I |f(x)|^p\,dx\right)^{1/p}\left(\int_I |g(x)|^q\,dx\right)^{1/q}. \tag{1.19}$$

(ii) (Minkowski's inequality) For any $p \in [1,\infty[$,

$$\left(\int_I |f(x)+g(x)|^p\,dx\right)^{1/p} \le \left(\int_I |f(x)|^p\,dx\right)^{1/p} + \left(\int_I |g(x)|^p\,dx\right)^{1/p}.$$

Proofs of these inequalities are outlined in Exercises 1.16 and 1.17. We state another important inequality that is valid under the same type of assumptions on the function f and the interval I:

Lemma 1.7.2 (Absolute integrability implies integrability) *Let I be an interval, and consider a function $f : I \to \mathbb{C}$. Assume that*

$$\int_I |f(x)|\,dx < \infty.$$

Then f is integrable, and

$$\left|\int_I f(x)\,dx\right| \le \int_I |f(x)|\,dx.$$

The results in Theorem 1.7.1 and Lemma 1.7.2 have discrete versions, reading as follows (Exercises 1.18 and 1.19):

Theorem 1.7.3 (Hölder's inequality, Minkowski's inequality) *Consider any scalar sequences* $\{x_k\}_{k=1}^{\infty}, \{y_k\}_{k=1}^{\infty}$. *Then the following inequalities hold:*

(i) (Hölder's inequality) For any numbers $p, q \in]1, \infty[$ *with* $1/p + 1/q = 1$,

$$\sum_{k=1}^{\infty} |x_k y_k| \leq \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} \left(\sum_{k=1}^{\infty} |y_k|^q \right)^{1/q}.$$

(ii) (Minkowski's inequality) For any $p \in [1, \infty[$,

$$\left(\sum_{k=1}^{\infty} |x_k + y_k|^p \right)^{1/p} \leq \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} + \left(\sum_{k=1}^{\infty} |y_k|^p \right)^{1/p}.$$

Lemma 1.7.4 (Absolute convergence implies convergence) *Let* $\{x_k\}_{k=1}^{\infty}$ *be a scalar sequence. If* $\sum_{k=1}^{\infty} |x_k|$ *is convergent, then* $\sum_{k=1}^{\infty} x_k$ *is convergent, and*

$$\left| \sum_{k=1}^{\infty} x_k \right| \leq \sum_{k=1}^{\infty} |x_k|.$$

1.8 Some special functions

In this section we define some special functions that are used in the book.

Definition 1.8.1 (Characteristic function) *Given a subset* $E \subseteq \mathbb{R}$, *let*

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases} \tag{1.20}$$

The function χ_E *is called the characteristic function for the set* E.

Usually, the set E will be an interval. Characteristic functions for intervals play a role in the description of signals that only take place over a limited time interval, e.g., a physical experiment that has a well-defined starting time and end time. For example, an electrical current that has the shape of a sine-function but only runs over a time interval of length 2π, might be described by a function of the type

$$f(x) = \sin x \ \chi_{[0,2\pi]}(x).$$

Definition 1.8.2 (Trigonometric polynomial) *A trigonometric polynomial is a finite linear combination of complex exponential functions having period 1, i.e., an expression*

$$H(x) = \sum_{k=N_1}^{N_2} c_k e^{2\pi ikx} \tag{1.21}$$

for some $N_1 \leq N_2$ *and some* $c_k \in \mathbb{C}$.

Note that a trigonometric polynomial can be rewritten as

$$H(x) = a_0 + \sum_{k=1}^{N} (a_k \cos(2\pi kx) + b_k \sin(2\pi kx)) \tag{1.22}$$

for $N = \max(|N_1|, |N_2|)$ and some coefficients $a_k, b_k \in \mathbb{C}$, see Exercise 1.20.

Trigonometric polynomials correspond to partial sums of Fourier series. They can also be considered with other periods than 1; for example, the expression

$$H(x) = \sum_{k=N_1}^{N_2} c_k e^{ikx} = a_0 + \sum_{k=1}^{N} (a_k \cos(kx) + b_k \sin(kx))$$

defines a trigonometric polynomial with period 2π.

Example 1.8.3 (Trigonometric polynomials) The set of trigonometric polynomials with period 2π is a subspace of

$$L^2(-\pi, \pi) := \left\{ f :]-\pi, \pi[\to \mathbb{C} \;\middle|\; \int_{-\pi}^{\pi} |f(x)|^2 \, dx < \infty \right\}. \tag{1.23}$$

In order to see this, we note that the trigonometric polynomials are continuous, so the integral in (1.23) is finite for such functions; this shows that the trigonometric polynomials form a subset of $L^2(-\pi, \pi)$. Now, a sum of two trigonometric polynomials is again a trigonometric polynomial, and a scalar multiple of a trigonometric polynomial is a trigonometric polynomial. According to Lemma 1.2.7 this implies that the set of trigonometric polynomials is a subspace of $L^2(-\pi, \pi)$. □

1.9 A useful technique: proof by induction

Suppose that we want to prove that a certain statement, involving a number n, holds true for all $n \in \mathbb{N}$. This can be done by showing that

(i) The statement holds for $n = 1$,

and

(ii) For an arbitrary $n \in \mathbb{N}$ it holds that if the statement is true for n, then it is also true when n is replaced by $n + 1$.

Part (ii) is called the *induction step;* the assumption that the considered statement holds for the value n is called the *induction hypothesis.* One can think about a proof by induction as climbing an infinitely high ladder: we can climb as high as we want if we can take the first step, and make sure that we always can go from one level to the next.

Example 1.9.1 (Induction) We will prove that for all $n \in \mathbb{N}$,

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}. \tag{1.24}$$

The statement certainly holds for $n = 1$. Now, assume that (1.24) holds for a certain value of n, and let us consider the statement with n replaced by $n + 1$: we want to verify that

$$1 + 2 + \cdots + n + (n+1) = \frac{(n+1)(n+2)}{2}.$$

In order to do so, we use that the hypothesis holds for n to derive that

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) &= \frac{n(n+1)}{2} + n + 1 \\ &= \frac{n(n+1)}{2} + \frac{2(n+1)}{2} \\ &= \frac{(n+1)(n+2)}{2}. \end{aligned}$$

This completes the proof of the induction step. □

Let us complete this section with a result that will be used at several instances. For $\ell, k \in \mathbb{N}_0$ with $\ell \geq k$, define the *binomial coefficient*

$$\binom{\ell}{k} := \frac{\ell!}{k!(\ell - k)!}.$$

The *binomial formula* states the following:

Lemma 1.9.2 (The binomial formula) *For all $y, z \in \mathbb{R}$ and any $\ell \in \mathbb{N}$,*

$$(y+z)^\ell = \sum_{k=0}^{\ell} \binom{\ell}{k} z^k y^{\ell-k}.$$

A guide to a proof of Lemma 1.9.2 is given in Exercise 1.26.

1.10 Exercises

1.1 Show that if $\{\mathbf{e}_k\}_{k=1}^n$ is an orthonormal basis for $\mathbb{C}^n$, then the representation (1.4) takes the form (1.5).

1.2 Do the vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \ \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$$

form a basis for $\mathbb{C}^3$?

1.3 Consider a bounded interval $[a,b] \subset \mathbb{R}$, and let $C[a,b]$ denote the set of continuous functions $f : [a,b] \to \mathbb{C}$, i.e.,

$$C[a,b] = \{f : [a,b] \to \mathbb{C} \mid f \text{ is continuous}\}. \tag{1.25}$$

(i) Show how to define appropriate operations of addition and scalar multiplication such that $C[a,b]$ equipped with these operations become a vector space.

(ii) Is $C[a,b]$ finite-dimensional? (*Hint:* for any $n \in \mathbb{N}, \ x^n \in C[a,b]$.)

1.4 Prove Lemma 1.2.7.

1.5 Give a geometric description (e.g., via a figure) of the set in $\mathbb{R}^2$ given by

$$W = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \mid c_1, c_2 \in [0,1[\},$$

where

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \ \mathbf{v}_2 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Describe also the closure $\overline{W}$.

1.6 Let $m \in \mathbb{N}$ be given, and assume that we for each $\ell \in \{0, 1, \ldots, m\}$ have chosen a polynomial P_ℓ of degree ℓ. Show that the collection of polynomials $\{P_\ell\}_{\ell=0}^m$ is a linearly independent set, and that

$$\operatorname{span}\{P_\ell\}_{\ell=0}^m = \operatorname{span}\{x^\ell\}_{\ell=0}^m.$$

1.7 Consider the function

$$f(x) := -xe^{-x}, \; x \in \mathbb{R}.$$

Calculate the number

$$\inf\{f(x) \mid x \in [0, \infty[\}.$$

Can "infimum" be replaced by "minimum"?

1.8 Determine the following numbers, and decide in each case whether "supremum" can be replaced by "maximum":

(i) $\sup_{x\in\mathbb{R}} \sin x$;

(ii) $\sup_{x\in\mathbb{R}} e^{-|x|}$;

(iii) $\sup_{x\in]1,\infty[} \frac{1}{x}$;

(iv) $\sup_{n\in\mathbb{N}} (-1)^n \frac{n+1}{n+100}$.

1.9 Determine the following numbers, and decide in each case whether "infimum" can be replaced by "minimum":

(i) $\inf_{x\in\mathbb{R}} \sin x$;

(ii) $\inf_{x\in\mathbb{R}} e^{-|x|}$;

(iii) $\inf_{x\in]1,\infty[} \frac{1}{x}$;

(iv) $\inf_{n\in\mathbb{N}} (-1)^n \frac{n+1}{n+100}$.

1.10 Determine the following numbers, and decide in each case whether "lim inf" resp. "lim sup" can be replaced by "limes":

(i) $\liminf_{n\to\infty} (-1)^n \frac{n+1}{n+100}$;

(ii) $\limsup_{n\to\infty} (-1)^n \frac{n+1}{n+100}$;

(iii) $\liminf_{n\to\infty} (-1)^n \frac{n+1}{n^2+100}$;

(iv) $\limsup_{n\to\infty} (-1)^n \frac{n+1}{n^2+100}$.

1.11 This exercise is related to Lemma 1.5.9 and its hypotheses.

(i) Prove Lemma 1.5.9.

(ii) Find a sequence $\{c_k\}_{k=1}^{\infty}$ of real numbers that does not have an accumulation point.

1.12 The purpose of this exercise is to prove Lemma 1.5.12. Under the conditions in Lemma 1.5.12, prove the following:

(i) For each $K \in \mathbb{N}$ there exists a number $m(K)$ such that

$$\liminf_{n\to\infty} f_n(k) \le f_m(k) + \frac{1}{K^2}$$

for all $k = 1, \dots, K$ and all $m \ge m(K)$.

(ii) With $m(K)$ chosen as in (i), show that for all $k = 1, \dots, K$ and all $m \ge m(K)$,

$$\sum_{k=1}^{K} \liminf_{n\to\infty} f_n(k) \le \sum_{k=1}^{\infty} f_m(k) + \frac{1}{K}.$$

(iii) Conclude that for all $k = 1, \dots, K$,

$$\sum_{k=1}^{K} \liminf_{n\to\infty} f_n(k) \le \liminf_{m\to\infty} \sum_{k=1}^{\infty} f_m(k) + \frac{1}{K}.$$

(iv) Conclude the proof by letting $K \to \infty$.

1.13 Prove Theorem 1.6.3. *Hint:* take a sequence of numbers $\{x_k\}_{k=1}^{\infty} \subset [a, b]$ such that

$$f(x_k) \to \sup_{x\in[a,b]} f(x) \text{ as } k \to \infty.$$

Use Lemma 1.5.10 to select a convergent subsequence $\{x_{k_n}\}_{n=1}^{\infty}$ of $\{x_k\}_{k=1}^{\infty} \subset [a, b]$, and let

$$x_0 := \lim_{n\to\infty} x_{k_n}.$$

1.14 Make a draft of the function f in Example 1.6.8, and argue that it is piecewise continuous. Argue further that f is unbounded and not uniformly continuous.

1.15 The purpose of this exercise is to prove Theorem 1.6.6. Assume that the hypotheses are satisfied, and let $x_0 \in I$ and $\epsilon > 0$ be given.

(i) Argue that there exists an $N \in \mathbb{N}$ such that

$$|f(x) - f_N(x)| \leq \epsilon/3, \ \forall x \in I.$$

(ii) Argue that we can choose $\delta > 0$ such that

$$|f_N(x) - f_N(x_0)| \leq \epsilon/3 \text{ whenever } x \in I \text{ and } |x - x_0| \leq \delta.$$

(iii) Use the triangle inequality to show that for $x \in I$ with $|x - x_0| \leq \delta$,

$$|f(x) - f(x_0)| \leq \epsilon.$$

1.16 The purpose of this exercise is to prove Theorem 1.7.1(i).

(i) Prove *Young's inequality:* for any $a, b > 0$ and any $p, q > 1$ with $p^{-1} + q^{-1} = 1$,

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q.$$

Hint: consider the xy-plane, the graph of the function $y = x^{p-1}$, and the lines $x = a, \ y = b$.

(ii) Prove Theorem 1.7.1(i). *Hint:* put

$$||f||_p = \left(\int_{-\infty}^{\infty} |f(x)|^p \, dx \right)^{1/p}, \quad ||g||_q = \left(\int_{-\infty}^{\infty} |g(x)|^q \, dx \right)^{1/q},$$

use Young's inequality with

$$a = \frac{|f(x)|}{||f||_p}, \quad b = \frac{|g(x)|}{||g||_q},$$

and perform an integration.

1.17 The purpose of this exercise is to prove Theorem 1.7.1(ii).

(i) Prove Theorem 1.7.1(ii) for $p = 1$.

We now assume that $p > 1$. Choose $q > 1$ such that $p^{-1}+q^{-1} = 1$.

(ii) Show that $q(p-1) = p$.

(iii) Show that

$$|f(x) + g(x)|^p \le |f(x)|\,|f(x) + g(x)|^{p-1} + |g(x)|\,|f(x) + g(x)|^{p-1}.$$

(iv) Show that

$$\int_I |f(x) + g(x)|^p\,dx \le \left(\int_I |f(x) + g(x)|^p\,dx\right)^{1-1/p} \times \left[\left(\int_I |f(x)|^p\,dx\right)^{1/p} + \left(\int_I |g(x)|^p\,dx\right)^{1/p}\right],$$

using the result in (ii) and Hölder's inequality applied on each of the two terms.

(v) Complete the proof of Theorem 1.7.1(ii) via (iv).

1.18 Prove Theorem 1.7.3(i), e.g., by appropriate modifications of the proof of Theorem 1.7.1(i) outlined in Exercise 1.16.

1.19 Prove Theorem 1.7.3(ii), e.g., by appropriate modifications of the proofs of Theorem 1.7.1(ii) outlined in Exercise 1.17.

1.20 Show that a trigonometric polynomial on the form (1.21) can be rewritten on the form (1.22).

1.21 Let $x \in \mathbb{R} \setminus \{1\}$. Show by induction that for any $N \in \mathbb{N}$,

$$1 + x + \cdots + x^N = \frac{1 - x^{N+1}}{1 - x}.$$

1.22 Show that for any $n \in \mathbb{N}$,

$$1 + 2^3 + 3^3 + \cdots + n^3 = \frac{1}{4}n^2(n+1)^2.$$

1.23 Show that $7^n - 4^n$ is a multiple of 3 for all $n \in \mathbb{N}$.

1.24 Show that the equality

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2$$

holds for all $n \in \mathbb{N}$.

1.25 Show that if $1 + x > 0$, then the inequality

$$(1 + x)^n \geq 1 + nx$$

holds for all $n \in \mathbb{N}$.

1.26 The purpose of this exercise is to prove Lemma 1.9.2.

(i) Let $\ell \in \mathbb{N}$. Prove that for $j = 0, 1, \ldots, \ell$,

$$\binom{\ell}{j} + \binom{\ell}{j-1} = \binom{\ell+1}{j}.$$

(ii) Prove Lemma 1.9.2 by induction.

2

Normed Vector Spaces

For the analysis of vector spaces, it is important to impose more structure on the space than merely the algebraic conditions in Definition 1.2.1. The purpose of this chapter is to consider norms on vector spaces and some of their properties. The key concept of a norm is presented in Section 2.1. In Section 2.2 the topological concepts treated in Section 1.4 are extended to general normed spaces. In Section 2.3 these concepts are linked with dense subsets, exemplified by Weierstrass' theorem on approximation of continuous functions by polynomials. Section 2.4 gives a short introduction to operators on normed vector spaces, and Section 2.5 deals with expansions in normed spaces in terms of bases.

2.1 Normed vector spaces

Our purpose in this section is to introduce *norms* on complex vector spaces. Intuitively, the norm of a vector shall measure the "size" of the vector; thus, the norm is the analogue of the concept of length of a vector $\mathbf{x} \in \mathbb{R}^n$, considered in (1.3).

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis,
DOI 10.1007/978-0-8176-4980-7_2,

Definition 2.1.1 (Norm) *Let V be a complex vector space. A norm on V is a function*

$$||\cdot||: V \to \mathbb{R}$$

that satisfies the following three conditions:

(i) $||\mathbf{v}|| \geq 0,\ \forall\, \mathbf{v} \in V,$ *and* $||\mathbf{v}|| = 0 \Leftrightarrow \mathbf{v} = \mathbf{0}$;

(ii) $||\alpha\mathbf{v}|| = |\alpha|\ ||\mathbf{v}||,\ \forall\, \mathbf{v} \in V,\ \alpha \in \mathbb{C}$;

(iii) $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||,\ \forall\, \mathbf{v}, \mathbf{w} \in V.$

A vector space equipped with a norm is called a normed vector space.

In situations where more than one vector space appears, we will frequently denote the norm on V by $||\cdot||_V$.

Note that we have stated the definition of a norm for a complex vector space. For a real vector space, the definition is the same, except that the scalars α in (ii) are assumed to be real numbers.

The inequality in Definition 2.1.1(iii) is called the *triangle inequality.* It has another important inequality, called the *reverse triangle inequality,* as a consequence:

Lemma 2.1.2 (Reverse triangle inequality) *Let V be a normed vector space. Then*

$$||\mathbf{v} - \mathbf{w}|| \geq |\ ||\mathbf{v}|| - ||\mathbf{w}||\ |,\ \forall\, \mathbf{v}, \mathbf{w} \in V. \tag{2.1}$$

Proof. Let $\mathbf{v}, \mathbf{w} \in V$. We need to show that

$$||\mathbf{v} - \mathbf{w}|| \geq ||\mathbf{v}|| - ||\mathbf{w}|| \text{ and } ||\mathbf{v} - \mathbf{w}|| \geq ||\mathbf{w}|| - ||\mathbf{v}||.$$

The proofs of these two inequalities are similar, so we only prove the first. Using the condition (iii) in Definition 2.1.1,

$$||\mathbf{v}|| = ||(\mathbf{v} - \mathbf{w}) + \mathbf{w}|| \leq ||\mathbf{v} - \mathbf{w}|| + ||\mathbf{w}||,$$

as desired. □

Example 2.1.3 ($\mathbb{R}^n$ and $\mathbb{C}^n$ are normed spaces) Using Minkowski's inequality in Theorem 1.7.3 with $p = 2$, one can prove directly that the spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ can be equipped with the norm

$$||\mathbf{x}|| = \left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2},\ \mathbf{x} = (x_1, x_2, \ldots, x_n).$$

Alternatively, the result is a direct consequence of a result proved later, Lemma 4.1.3, and the fact that $\mathbb{R}^n$ and $\mathbb{C}^n$ are equipped with the inner products in (1.1) and (1.2), respectively. □

Let us consider an important normed vector space consisting of functions:

Example 2.1.4 (Continuous functions on a bounded interval) Consider a bounded interval $[a, b] \subset \mathbb{R}$, and let $C[a, b]$ denote the set of continuous functions $f : [a, b] \to \mathbb{C}$, i.e.,

$$C[a, b] := \{f : [a, b] \to \mathbb{C} \mid f \text{ is continuous}\}.$$

Equip $C[a, b]$ with the natural operations of addition and scalar multiplication, see Exercise 1.3. By Theorem 1.6.3 we know that each function $f \in C[a, b]$ is bounded and assumes a maximum value; let

$$||f||_\infty := \max_{x \in [a,b]} |f(x)|. \tag{2.2}$$

We will verify that $|| \cdot ||_\infty$ defines a norm on $C[a, b]$, i.e., that it satisfies the requirements in Definition 2.1.1. First, it is clear that $||f||_\infty \geq 0$ for all $f \in C[a, b]$. Also, the function $f = 0$ belongs to $C[a, b]$ and satisfies that $||f||_\infty = 0$. On the other hand, if $||f||_\infty = 0$ for some function $f \in C[a, b]$, then the definition of $||\cdot||_\infty$ shows that $f(x) = 0$ for all $x \in [a, b]$, i.e., $f = 0$; this verifies (i) in Definition 2.1.1. The property (ii) is clearly satisfied. Now, in order to verify the condition (iii), let $f, g \in C[a, b]$. Then, for each $x \in [a, b]$,

$$\begin{aligned} |f(x) + g(x)| &\leq |f(x)| + |g(x)| \\ &\leq ||f||_\infty + ||g||_\infty; \end{aligned}$$

because this holds for all $x \in [a, b]$, it follows that

$$\begin{aligned} ||f + g||_\infty &= \max_{x \in [a,b]} |f(x) + g(x)| \\ &\leq ||f||_\infty + ||g||_\infty. \end{aligned}$$

We have now verified that $||\cdot||_\infty$ defines a norm on $C[a, b]$. The norm $||\cdot||_\infty$ is called the *supremums-norm.*

We will use the space $C[a, b]$ to illustrate the concepts and results appearing in the entire chapter. □

Frequently, a vector space can be equipped with different norms. For the case of the vector space $C[a, b]$ an alternative norm is discussed in Exercise 6.1.

We are now ready to introduce the important concept of convergence of a sequence of elements in a normed vector space. We will use the notation

$$\{\mathbf{v}_k\}_{k=1}^\infty = \{\mathbf{v}_1, \mathbf{v}_2, \ldots\},$$

indicating that we have chosen an ordering of the vectors $\mathbf{v}_k$ in V.

Definition 2.1.5 (Convergence in normed spaces) *A sequence* $\{\mathbf{v}_k\}_{k=1}^{\infty}$ *in a normed vector space* V *converges to* $\mathbf{v} \in V$ *if*

$$||\mathbf{v} - \mathbf{v}_k|| \to 0 \ \text{as} \ k \to \infty. \tag{2.3}$$

This is written as

$$\mathbf{v}_k \to \mathbf{v} \ \text{as} \ k \to \infty,$$

or

$$\mathbf{v} = \lim_{k\to\infty} \mathbf{v}_k.$$

Note that the precise meaning of the condition (2.3) is that there for all $\epsilon > 0$ exists an $N \in \mathbb{N}$ such that

$$||\mathbf{v} - \mathbf{v}_k|| \le \epsilon \text{ for all } k \ge N. \tag{2.4}$$

Let us illustrate the concept of convergence in the setting of the vector space $C[0, \frac{1}{2}]$ equipped with the norm $|| \cdot ||_\infty$ in (2.2):

Example 2.1.6 (Convergence of functions in $C[a, b]$) We will consider functions f and $\{f_k\}_{k=1}^{\infty}$ defined in terms of an infinite series and its partial sums. For $k \in \mathbb{N}$, let

$$f_k(x) := \sum_{n=0}^{k} x^n, \ x \in]-1, 1[.$$

Using that

$$(1 - x)(1 + x + \cdots + x^k) = 1 - x^{k+1},$$

it follows that

$$f_k(x) = 1 + x + \cdots + x^k = \frac{1 - x^{k+1}}{1 - x}, \ x \in]-1, 1[.$$

Thus,

$$f_k(x) \to \frac{1}{1 - x} \ \text{ as } k \to \infty.$$

This shows that the functions f_k converge pointwise toward the function

$$f(x) := \sum_{n=0}^{\infty} x^n = \frac{1}{1 - x}, \ x \in]-1, 1[.$$

Let us now consider the interval $[0, \frac{1}{2}]$. It is clear that all the functions f and $f_k,\ k \in \mathbb{N}$, belong to $C[0, \frac{1}{2}]$. Now, for $x \in [0, \frac{1}{2}]$,

$$\begin{aligned} |f(x) - f_k(x)| &= \left| \frac{1}{1-x} - \frac{1-x^{k+1}}{1-x} \right| \\ &= \frac{x^{k+1}}{1-x} \\ &\leq \left(\frac{1}{2}\right)^k . \end{aligned}$$

Thus, with the norm $||\cdot||_\infty$ introduced in Example 2.1.4,

$$\begin{aligned} ||f - f_k||_\infty &= \sup_{x \in [0,\frac{1}{2}]} |f(x) - f_k(x)| \\ &\leq \left(\frac{1}{2}\right)^k \to 0 \text{ as } k \to \infty. \end{aligned}$$

This shows that $f_k \to f$ in $C[0, \frac{1}{2}]$ equipped with the norm $||\cdot||_\infty$. □

2.2 Topology in normed vector spaces

The concepts of open and closed subsets in $\mathbb{R}^n$ can be extended to arbitrary normed vector spaces:

Definition 2.2.1 (Balls and neighborhoods) *Let V be a normed vector space.*

(i) Given a point $\mathbf{v}_0 \in V$, the ball centered at $\mathbf{v}_0$ and with radius $r > 0$ is the set

$$B(\mathbf{v}_0, r) := \{\mathbf{v} \in V \mid ||\mathbf{v} - \mathbf{v}_0|| < r\}.$$

(ii) For $\mathbf{v}_0 \in V$, a neighborhood of $\mathbf{v}_0$ is a subset of V that contains a ball $B(\mathbf{v}_0, \delta)$ for some $\delta > 0$.

Definition 2.2.2 (Open and closed sets) *Let V be a normed vector space, and W a subset of V.*

(i) W is open if for each $\mathbf{v}_0 \in W$ there exists a $\delta > 0$ such that

$$B(\mathbf{v}_0, \delta) \subseteq W.$$

(ii) The complement of W is

$$W^c := V \setminus W.$$

(iii) W is closed if the complement W^c is open.

The definition explains how one can verify that a subset of a normed vector space is open. In order to check that a subset is closed, one can either check that the complement is open, or use the following lemma.

Lemma 2.2.3 (Closed sets) *For a subset W of a normed vector space V the following are equivalent:*

(i) W is closed.

(ii) For any convergent sequence $\{\mathbf{v}_k\}_{k=1}^{\infty}$ of elements in W, the limit $\mathbf{v} = \lim_{k\to\infty} \mathbf{v}_k$ also belongs to W.

Proof. For the proof of (i) $\Rightarrow$ (ii), assume that W is closed, i.e., that W^c is open. Let $\{\mathbf{v}_k\}_{k=1}^{\infty}$ be a convergent sequence of elements in W. Let $\mathbf{v} = \lim_{k\to\infty} \mathbf{v}_k$. We will show that $\mathbf{v} \in W$. Assume the opposite, i.e., that $\mathbf{v} \in W^c$. Then there is a $\delta > 0$ such that $B(\mathbf{v}, \delta) \subseteq W^c$; but this implies that $\mathbf{v}_k \in W^c$ for k sufficiently large, which is a contradiction. Thus, $\mathbf{v} \in W$, which proves (ii).

For the proof of (ii) $\Rightarrow$ (i), assume that (ii) holds. We will show that W is closed by showing that the complement W^c is open. Let $\mathbf{v} \in W^c$. We want to prove that for $k \in \mathbb{N}$ sufficiently large, $B(\mathbf{v}, 1/k) \subseteq W^c$. In fact, if this was not the case, we could for infinitely many $k \in \mathbb{N}$ pick $\mathbf{v}_k \in B(\mathbf{v}, 1/k) \cap W$. This would yield a sequence $\mathbf{v}_k \in W$ converging to $\mathbf{v} \in W^c$. But this contradicts the assumption in (ii)! This proves that for $\delta > 0$ sufficiently small, the ball $B(\mathbf{v}, \delta)$ is contained in W^c. Thus, W^c is open, i.e., W is closed. □

Example 2.2.4 (The set of polynomials is not closed in $C[a, b]$) Consider the vector space $C[0, \frac{1}{2}]$ equipped with the norm $||\cdot||_\infty$, see Example 2.1.4. The set W consisting of all polynomials on $[0, \frac{1}{2}]$ is a subspace of $C[0, \frac{1}{2}]$. We will prove that W does not form a closed set. Consider the functions $P_k \in W,\ k \in \mathbb{N}$, given by

$$P_k(x) = \sum_{n=0}^{k} \frac{1}{n+1} x^n = 1 + \frac{1}{2}\, x + \cdots + \frac{1}{k+1}\, x^k,\ x \in [0, \frac{1}{2}].$$

The reader can check that the infinite series $\sum_{n=0}^{\infty} \frac{1}{n+1} x^n$ is convergent for all $x \in [0, \frac{1}{2}]$. Let

$$P(x) := \sum_{n=0}^{\infty} \frac{1}{n+1} x^n,\ x \in [0, \frac{1}{2}].$$

Then, for any $x \in [0, \frac{1}{2}]$,

$$\begin{aligned} |P(x) - P_k(x)| &= \left| \sum_{n=k+1}^{\infty} \frac{1}{n+1} x^n \right| \\ &\leq \sum_{n=k+1}^{\infty} \frac{1}{n+1} \left(\frac{1}{2}\right)^n. \end{aligned}$$

Thus,

$$\begin{aligned} ||P - P_k||_\infty &= \sup_{x \in [0,1/2]} |P(x) - P_k(x)| \\ &\leq \sum_{n=k+1}^{\infty} \frac{1}{n+1} \left(\frac{1}{2}\right)^n. \end{aligned}$$

Since $\sum_{n=0}^{\infty} \frac{1}{n+1} \frac{1}{2^n}$ is convergent, we infer that

$$||P - P_k||_\infty \to 0 \text{ as } k \to \infty.$$

By Theorem 1.6.6 it follows that P is a continuous function. Thus, the sequence P_k converges in $C[0, \frac{1}{2}]$ with respect to the norm $|| \cdot ||_\infty$. But the limit P is not a polynomial! By Lemma 2.2.3 this proves that W does not form a closed subset of $C[0, \frac{1}{2}]$.

A slight modification of the above argument shows that W is not open either (Exercise 2.8). □

2.3 Approximation in normed vector spaces

In the technical sense, normed vector spaces can contain elements that are very complicated to deal with (concrete instances will occur in the context of the L^p-spaces discussed in Chapter 5). In such cases it is important to have concepts for approximating complicated elements by more convenient elements. In this context we need the following concept.

Definition 2.3.1 (Dense subset) *A subset W of a normed vector space V is said to be dense in V if for each $\mathbf{v} \in V$ and each $\epsilon > 0$ there exists an element $\mathbf{w} \in W$ such that*

$$||\mathbf{v} - \mathbf{w}|| \leq \epsilon.$$

If W is a dense subspace of V, then all elements in V can be approximated arbitrarily well by elements in W. In fact, let $\mathbf{v} \in V$ and take $\epsilon = 1/k$ for $k \in \mathbb{N}$. Then the condition in Definition 2.3.1 says that we can find an element $\mathbf{w}_k \in W$ such that $||\mathbf{v} - \mathbf{w}_k|| \leq 1/k$. By construction, the sequence $\{\mathbf{w}_k\}_{k=1}^{\infty}$ satisfies that

$$\mathbf{w}_k \to \mathbf{v} \text{ as } k \to \infty.$$

For a subset W of a normed vector space V, it is convenient to have a formal notation for the set of elements in V that can be approximated arbitrarily well in norm by elements in W. This leads to the concept of the *closure* of a set:

Definition 2.3.2 (Closure) *Let W be a subset of a normed vector space V. The closure of W, to be denoted $\overline{W}$, consists of all the elements in $\mathbf{v} \in V$ having the property that we for each $\epsilon > 0$ can find an element $\mathbf{w} \in W$ such that $||\mathbf{v} - \mathbf{w}|| \leq \epsilon$.*

Via Lemma 2.2.3, it follows that $\overline{W}$ is a closed set; in fact, it is the smallest closed set in V containing W, see Exercise 2.9. Combining Definition 2.3.1 and Definition 2.3.2 leads to the following fundamental observation:

Lemma 2.3.3 (Characterization of dense subset) *Let W be a subset of the normed vector space V. Then W is dense in V if and only if $\overline{W} = V$.*

We have already considered examples dealing with continuous functions and polynomials. Let us now formulate the famous *Weierstrass' Theorem,* stating that every continuous function on a closed and bounded interval can be approximated arbitrarily well with a polynomial:

Theorem 2.3.4 (Weierstrass' theorem) *Let $[a, b] \subset \mathbb{R}$ be a closed and bounded interval and f a continuous function defined on $[a, b]$. Then, for every $\epsilon > 0$ there exists a polynomial P such that*

$$|f(x) - P(x)| \leq \epsilon \text{ for all } x \in [a, b]. \tag{2.5}$$

A proof of Weierstrass' theorem can be found in Appendix A.1. Note that in terms of the norm $|| \cdot ||_\infty$ in (2.2), the inequality (2.5) means that

$$||f - P||_\infty \leq \epsilon.$$

Formulated in terms of Definition 2.3.1, Weierstrass' theorem says that the set of polynomials on any closed and bounded interval $[a, b]$ is dense in $C[a, b]$.

Example 2.3.5 (The closure of the set of polynomials in $C[a, b]$) Let W be the vector space of polynomials on $[0, \frac{1}{2}]$, considered in Example 2.2.4. Then $\overline{W}$ consists of all functions $P : [0, \frac{1}{2}] \to \mathbb{C}$ for which we can find polynomials $P_k, \ k \in \mathbb{N}$, such that

$$||P - P_k||_\infty \to 0 \text{ as } k \to \infty.$$

By Theorem 1.6.6 any such function P is continuous, i.e., $\overline{W} \subseteq C[0, \frac{1}{2}]$. On the other hand, Theorem 2.3.4 implies that each function $f \in C[0, \frac{1}{2}]$ belongs to $\overline{W}$, so we conclude that

$$\overline{W} = C[0, \frac{1}{2}].$$

□

2.4 Linear operators on normed spaces

Given arbitrary (complex) vector space V_1 and V_2, a mapping $T : V_1 \to V_2$ is *linear* if

$$T(\alpha\mathbf{v} + \beta\mathbf{w}) = \alpha T(\mathbf{v}) + \beta T(\mathbf{w}), \ \forall\, \alpha, \beta \in \mathbb{C}, \ \mathbf{v}, \mathbf{w} \in V_1. \tag{2.6}$$

In the context of normed vector spaces, it is custom to use the word *operator* instead of *map*; we will adopt that terminology here. Usually, the action of a linear operator T on a vector $\mathbf{v}$ is written $T\mathbf{v}$ rather than $T(\mathbf{v})$; we will also adopt that convention. Often we will need to consider the norms of an element $\mathbf{v} \in V_1$, as well as the norm of the image $T\mathbf{v} \in V_2$. In cases where $V_1 \neq V_2$, we will frequently denote these norms by $||\mathbf{v}||_{V_1}$, respectively $||T\mathbf{v}||_{V_2}$; in cases where no confusion can arise we will omit the subscript.

Many normed vector spaces that appear in practice are infinite-dimensional. It is more complicated to deal with linear operators on such spaces than linear operators on $\mathbb{R}^n$. The following definition presents a condition that allows us to work with linear operators on normed vector spaces almost like with linear operators on $\mathbb{R}^n$; for the case of a linear operator on a finite-dimensional vector space, the condition is automatically satisfied.

Definition 2.4.1 (Bounded linear operator) *Let V_1 and V_2 be normed spaces. A linear operator*

$$T : V_1 \to V_2$$

is bounded if there exists a constant $K \geq 0$ such that

$$||T\mathbf{v}||_{V_2} \leq K\, ||\mathbf{v}||_{V_1}, \ \forall\, \mathbf{v} \in V_1. \tag{2.7}$$

The smallest possible value of K that can be used in (2.7) is called the norm of the operator T, and is denoted by $||T||$.

Example 2.4.2 (The identity operator) In the case of $V_1 = V_2$, we can consider the *identity operator*

$$I : V_1 \to V_1, \ I\mathbf{v} := \mathbf{v}.$$

The identity operator is linear and bounded, and $||I|| = 1$. □

A more interesting example of a bounded linear operator on $C[a, b]$ is given next:

Example 2.4.3 (Integral operator) Let $[a, b]$ be a bounded closed interval, and

$$K(\cdot, \cdot) : [a, b] \times [a, b] \to \mathbb{C}$$

a continuous function of two variables. Consider the linear operator T given by

$$T : C[a,b] \to C[a,b], \quad (Tf)(x) = \int_a^b K(x,y)f(y)\,dy.$$

We equip the space $C[a,b]$ with the norm $||\cdot||_\infty$ considered in Example 2.1.4, and want to show that T is bounded. In order to do so, we will use the inequality

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

see Lemma 1.7.2. Now, let $f \in C[a,b]$. Then,

$$\begin{aligned}
|(Tf)(x)| &= \left|\int_a^b K(x,y)f(y)\,dy\right| \\
&\leq \int_a^b |K(x,y)f(y)|\,dy \\
&\leq \int_a^b \left(\sup_{(x,y)\in[a,b]\times[a,b]} |K(x,y)|\right)\left(\sup_{y\in[a,b]} |f(y)|\right) dy \\
&= \left(\sup_{(x,y)\in[a,b]\times[a,b]} |K(x,y)|\right)\left(\sup_{y\in[a,b]} |f(y)|\right)\int_a^b dy \\
&= (b-a)\left(\sup_{(x,y)\in[a,b]\times[a,b]} |K(x,y)|\right)\left(\sup_{y\in[a,b]} |f(y)|\right) \\
&= (b-a)\left(\sup_{(x,y)\in[a,b]\times[a,b]} |K(x,y)|\right) ||f||_\infty.
\end{aligned}$$

This implies that

$$\begin{aligned}
||Tf||_\infty &= \sup_{x\in[a,b]} |(Tf)(x)| \\
&\leq (b-a)\left(\sup_{(x,y)\in[a,b]\times[a,b]} |K(x,y)|\right) ||f||_\infty.
\end{aligned}$$

Thus, T is bounded and

$$||T|| \leq (b-a)\left(\sup_{(x,y)\in[a,b]\times[a,b]} |K(x,y)|\right).$$

An operator T of the type considered here is called an *integral operator.* In the mathematical literature the operator is often analyzed on other vector spaces than the space $C[a,b]$ considered here. □

Many other types of bounded operators will be considered later, e.g., in Sections 3.3, 4.5 and 6.2.

A linear operator that does not satisfy the requirement in Definition 2.4.1 is called an *unbounded operator*. We state an example of an unbounded operator, but ask the reader to prove the claims (Exercise 2.12).

Example 2.4.4 (Differentiation operator) Consider the vector space

$$C^1[0, \frac{1}{2}] := \left\{ f : [0, \frac{1}{2}] \to \mathbb{C} \;\middle|\; f \text{ is differentiable and } f' \text{ is continuous} \right\}.$$

Then $C^1[0, \frac{1}{2}]$ is a subspace of $C[0, \frac{1}{2}]$. We now equip the spaces $C[0, \frac{1}{2}]$ and $C^1[0, \frac{1}{2}]$ with the supremums-norm, and consider the mapping

$$\mathcal{D} : C^1[0, \frac{1}{2}] \to C[0, \frac{1}{2}], \quad (\mathcal{D}f)(x) := f'(x), \; x \in [0, \frac{1}{2}].$$

Then $\mathcal{D}$ is a linear unbounded operator. □

We will now define some of the central concepts related to linear operators. The reader will notice that they are similar to concepts that are studied in linear algebra.

Definition 2.4.5 (Injective and surjective operator, isometry) *Let V_1 and V_2 be normed spaces and $T : V_1 \to V_2$ a bounded linear operator.*

(i) The operator T is injective if $T\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} = \mathbf{0}$.

(ii) The operator T is surjective if for each $\mathbf{w} \in V_2$ there exists a $\mathbf{v} \in V_1$ such that $T\mathbf{v} = \mathbf{w}$.

(iii) The operator T is bijective if T is injective and surjective.

(iv) The operator T is an isometry if $||T\mathbf{v}|| = ||\mathbf{v}||$ for all $\mathbf{v} \in V_1$.

Note that the definition of injectivity given here for a linear operator corresponds to the classical definition of an injective function:

Example 2.4.6 (Injective operator) In classical terminology, a linear mapping (or any other mapping) $T : V_1 \to V_2$ is injective if

$$T\mathbf{v} = T\mathbf{w} \Rightarrow \mathbf{v} = \mathbf{w}. \tag{2.8}$$

Assuming that T is linear, the requirement in (2.8) amounts to the requirement

$$T(\mathbf{v} - \mathbf{w}) = \mathbf{0} \Rightarrow \mathbf{v} - \mathbf{w} = \mathbf{0}; \tag{2.9}$$

condition (2.9) is exactly the one appearing in Definition 2.4.5(i). □

Definition 2.4.7 (Invertible operator) *Let V be a vector space. A linear operator $T : V \to V$ is invertible if there exists a linear operator $S : V \to V$ such that*

$$ST = TS = I.$$

The operator S is called the inverse operator of T, and is usually denoted by T^{-1}.

If V is finite-dimensional, it is enough to check either that $TT^{-1} = I$ or that $T^{-1}T = I$ in order to show that T^{-1} is the inverse of T. If V is an infinite-dimensional vector space, both conditions must be verified (Exercise 3.14).

2.5 Series in normed vector spaces

Eventually, we want to obtain expansions in an infinite-dimensional normed vector space V of the type we have at hand in $\mathbb{C}^n$, see (1.4). That is, we want to consider a collection of vectors $\{\mathbf{v}_k\}_{k=1}^{\infty}$ in V with the property that each $\mathbf{v} \in V$ has a representation

$$\mathbf{v} = \sum_{k=1}^{\infty} c_k \mathbf{v}_k \tag{2.10}$$

for appropriately chosen coefficients c_k.

The first step is to clarify what is meant by convergence of an infinite series consisting of elements in a normed vector space. In order to avoid confusion with the expansion (2.10) which involves the coefficients c_k, we will consider an infinite sequence $\{\mathbf{w}_k\}_{k=1}^{\infty}$ of elements in V. As discussed before, the notation $\{\mathbf{w}_k\}_{k=1}^{\infty}$ indicates that we have chosen an ordering of the vectors $\mathbf{w}_k$,

$$\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k, \dots \quad .$$

Our first goal is to give an exact definition of the infinite series $\sum_{k=1}^{\infty} \mathbf{w}_k$. In order to do so, we introduce the *Nth partial sum* by

$$\mathbf{S}_N := \sum_{k=1}^{N} \mathbf{w}_k.$$

Definition 2.5.1 (Convergence of infinite series in normed space) *Let* $\{\mathbf{w}_k\}_{k=1}^{\infty}$ *be a sequence of elements in a normed vector space* V*. We say that the infinite series* $\sum_{k=1}^{\infty} \mathbf{w}_k$ *is convergent with sum* $\mathbf{w} \in V$ *if*

$$\left\| \mathbf{w} - \sum_{k=1}^{N} \mathbf{w}_k \right\| \to 0 \text{ as } N \to \infty.$$

If this condition is satisfied, we write

$$\mathbf{w} = \sum_{k=1}^{\infty} \mathbf{w}_k. \tag{2.11}$$

Thus, the definition of a convergent infinite series in a normed vector space is analogous to the definition of a convergent series of numbers.

We will now define the span of an *infinite* collection of vectors. In order to avoid convergence issues, the span is defined as the collection of all *finite* linear combinations of the vectors:

Definition 2.5.2 (Span) *Given a sequence* $\{\mathbf{v}_k\}_{k=1}^{\infty}$ *in a normed vector space* V*, let span*$\{\mathbf{v}_k\}_{k=1}^{\infty}$ *denote the vector space consisting of all finite linear combinations of vectors* $\mathbf{v}_k$*, i.e.,*

$$\textit{span}\,\{\mathbf{v}_k\}_{k=1}^{\infty} = \{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_N\mathbf{v}_N \,|\, N \in \mathbb{N}, \alpha_1, \alpha_2, \ldots, \alpha_N \in \mathbb{C}\}.$$

The definition of convergence implies (see Exercise 2.6) that if each $\mathbf{v} \in V$ has a representation of the type

$$\mathbf{v} = \sum_{k=1}^{\infty} c_k \mathbf{v}_k \tag{2.12}$$

for some scalars $c_k \in \mathbb{C}$, then

$$\overline{\text{span}\{\mathbf{v}_k\}_{k=1}^{\infty}} = V. \tag{2.13}$$

On the other hand, the property (2.13) does not imply that each $\mathbf{v} \in V$ has a representation of the type (2.12), see Exercise 2.7.

Definition 2.5.3 (Total sequence) *Let* V *be a normed vector space. A sequence* $\{\mathbf{v}_k\}_{k=1}^{\infty}$ *having the property* (2.13) *is said to be complete or total in* V*.*

We note that there exist normed spaces where no sequence $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is complete. A normed vector space, in which a complete sequence $\{\mathbf{v}_k\}_{k=1}^{\infty}$ exists, is said to be *separable.*

We will now define the crucial concept of a *basis* in a normed vector space V. We will not go further into that concept now, but return to it later.

Definition 2.5.4 (Basis in normed vector space) *Consider a sequence* $\{\mathbf{v}_k\}_{k=1}^{\infty}$ *of vectors in a normed vector space* V. *The sequence* $\{\mathbf{v}_k\}_{k=1}^{\infty}$ *is a (Schauder) basis for* V *if for each* $\mathbf{v} \in V$ *there exist unique scalar coefficients* $\{c_k\}_{k=1}^{\infty}$ *such that*

$$\mathbf{v} = \sum_{k=1}^{\infty} c_k \mathbf{v}_k. \tag{2.14}$$

Definition 2.5.4 is the natural extension of Definition 1.3.2 to the setting of infinite-dimensional normed vector spaces. The infinite-dimensional case is much more complicated than the finite-dimensional, due to the fact that we have to deal with infinite sums. We will discuss an important class of bases in the context of Hilbert spaces in Section 4.7. The concept of a basis is central in Section 7.4, Chapter 8, and Chapter 11.

2.6 Exercises

2.1 Let V be a normed vector space. Show that if $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a sequence in V and $\mathbf{v}_k \to \mathbf{v}$ as $k \to \infty$, then

$$\lim_{k\to\infty} ||\mathbf{v}_k|| = ||\mathbf{v}||.$$

2.2 Let V be a normed vector space and $\{\mathbf{v}_k\}_{k=1}^{N}$ a collection of vectors in V. Assume that there exists a constant $A > 0$ such that the inequality

$$A \sum_{k=1}^{N} |c_k|^2 \leq ||\sum_{k=1}^{N} c_k \mathbf{v}_k||^2$$

holds for all scalar coefficients $c_1, \dots, c_N$. Show that the vectors $\{\mathbf{v}_k\}_{k=1}^{N}$ are linearly independent.

2.3 Consider the vector space $\mathbb{R}$ with the norm

$$||x|| = |x|, \;\; x \in \mathbb{R}.$$

(i) Is the subset $\mathbb{N}$ closed in $\mathbb{R}$? Describe the set $\overline{\mathbb{N}}$.

(ii) Is the subset $\mathbb{Q}$ closed in $\mathbb{R}$? Describe the set $\overline{\mathbb{Q}}$.

2.4 Consider the set V of trigonometric polynomials as defined in Definition 1.8.2.

(i) Show that V is a subspace of $C[0,1]$.

(ii) Equip $C[0,1]$ with the norm $||\cdot||_\infty$, as in Example 2.1.4. Is V a closed subspace of $C[0,1]$?

2.5 Consider the vector space

$$W := \left\{ f:\mathbb{R}\to\mathbb{C} \;\middle|\; f \text{ is continuous, and } \sum_{k\in\mathbb{Z}} ||f\chi_{[k,k+1[}||_\infty < \infty \right\}.$$

(i) Show that the expression $||\cdot||_W$ given by

$$||f||_W := \sum_{k\in\mathbb{Z}} ||f\chi_{[k,k+1[}||_\infty$$

defines a norm on W.

(ii) Check whether the function $f(x) = e^x,\ x\in\mathbb{R}$, belongs to the vector space W or not.

2.6 Let V be a normed vector space. Show that if each $\mathbf{v}\in V$ has a representation of the type (2.12) for some $\mathbf{v}_k\in V$, then (2.13) holds.

2.7 Consider the set W of all polynomials on $[0,\frac{1}{2}]$, as in Example 2.2.4.

(i) Argue that

$$W = \text{span}\{1, x, x^2, \dots\}.$$

(ii) Argue that there exist functions $f\in C[0,\frac{1}{2}]$ that cannot be written on the form

$$f(x) = \sum_{k=0}^{\infty} c_k x^k,\ \ x\in]0,\frac{1}{2}[.$$

This proves that the property (2.13) does not imply that each $\mathbf{v}\in V$ has a representation of the type (2.12). In fact, by Example 2.3.5 we know that $\overline{\text{span}}\{1,x,x^2,\dots\} = C[0,\frac{1}{2}]$, but we just saw that not all $f\in C[0,\frac{1}{2}]$ has a representation as an infinite sum of functions $x^k,\ k=0,1,\dots$.

2.8 Let W be the subspace of $C[0, \frac{1}{2}]$ considered in Example 2.2.4. Given $\delta > 0$, put

$$g(x) := \delta \sum_{n=1}^{\infty} \frac{1}{n+1} x^n, \ x \in [0, \frac{1}{2}].$$

(i) Show that $g \in B(\mathbf{0}, \delta)$.

(ii) Use (i) to conclude that W cannot be an open subset of $C[0, \frac{1}{2}]$.

2.9 Let W be a subspace of a normed vector space V. Show that the closure $\overline{W}$ is the smallest closed subspace of V that contains W.

2.10 Consider the linear map

$$T : \mathbb{R}^2 \to \mathbb{R}^2, \ T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ x_1 + x_2 \end{pmatrix}.$$

Equip $\mathbb{R}^2$ with the canonical norm, and answer the following:

(i) Is T injective?

(ii) Is T surjective?

(iii) Is T an isometry?

2.11 Consider the linear map

$$T : \mathbb{R}^2 \to \mathbb{R}^2, \ T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1}{3}}x_1 + \sqrt{\frac{2}{3}}x_2 \\ -\sqrt{\frac{2}{3}}x_1 + \sqrt{\frac{1}{3}}x_2 \end{pmatrix}.$$

Equip $\mathbb{R}^2$ with the canonical inner product and check the following:

(i) Is T injective?

(ii) Is T surjective?

(iii) Show that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$,

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

(iv) Is T an isometry?

2.12 Consider the vector space $C^1[0, \frac{1}{2}]$ defined in Example 2.4.4.

(i) Show that $C^1[0, \frac{1}{2}]$ is a subspace of $C[0, \frac{1}{2}]$.

Consider the mapping

$$\mathcal{D} : C^1[0, \frac{1}{2}] \to C[0, \frac{1}{2}], \quad (\mathcal{D}f)(x) := f'(x), \; x \in [0, \frac{1}{2}].$$

(ii) Show that $\mathcal{D}$ is a linear operator.

(iii) Show that $\mathcal{D}$ is unbounded. (*Hint:* consider the functions $f(x) = x^n$ for $n \in \mathbb{N}$.)

2.13 Let V be a normed vector space and T a bounded linear operator on V. Let W be a subset of V, and denote the image of W by $T(W)$.

(i) Show that

$$T(\overline{W}) \subseteq \overline{T(W)}.$$

(ii) Assume additionally that T is invertible and that T^{-1} is bounded. Show that

$$T(\overline{W}) = \overline{T(W)}.$$

2.14 Let V be a normed vector space and T a bounded linear operator on V.

(i) Assume that $\mathbf{w}_k \to \mathbf{w}$ in V as $k \to \infty$. Show that

$$T\mathbf{w}_k \to T\mathbf{w} \text{ as } k \to \infty.$$

(ii) Assume that $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a sequence of elements in V and that $\sum_{k=1}^{\infty} c_k \mathbf{v}_k$ is convergent for some scalar sequence $\{c_k\}_{k=1}^{\infty}$. Show that

$$T \sum_{k=1}^{\infty} c_k \mathbf{v}_k = \sum_{k=1}^{\infty} c_k T \mathbf{v}_k.$$

2.15 Let V denote a normed vector space, and let $\{\mathbf{v}_k\}_{k=1}^n$ denote a collection of vectors in V. Equip $\mathbb{C}^n$ with the canonical norm, and consider the mapping

$$T : \mathbb{C}^n \to V,\ T\{c_k\}_{k=1}^n := \sum_{k=1}^n c_k \mathbf{v}_k.$$

Show that T is a bounded linear operator with

$$||T|| \leq \left(\sum_{k=1}^n ||\mathbf{v}_k||^2\right)^{1/2}.$$

Hint: use the triangle inequality on $||\sum_{k=1}^n c_k \mathbf{v}_k||$, followed by an application of Hölder's inequality.

3

Banach Spaces

Banach spaces form a special class of normed vector spaces. Compared with other normed spaces, Banach spaces have the advantage that it is easier to check that a sequence of vectors in the space is convergent. We give the formal definition of a Banach space in Section 3.1. As example of a Banach space we consider the set of continuous functions on a closed and bounded interval. Other important Banach spaces, the sequence spaces $\ell^p(\mathbb{N})$, are introduced in Section 3.2. In Section 3.3 we continue the analysis of bounded linear operators initiated in Section 2.4.

3.1 Banach spaces

Already in Definition 2.1.5 we saw the concept of convergence for a sequence of elements in a normed vector space. We now present a related definition that turns out to be the key to the definition of Banach spaces.

Definition 3.1.1 (Cauchy sequence) *Let V be a normed vector space. A sequence $\{\mathbf{v}_k\}_{k=1}^{\infty}$ of elements in V is a Cauchy sequence if for each $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that*

$$||\mathbf{v}_k - \mathbf{v}_\ell|| \leq \epsilon \text{ whenever } k, \ell \geq N.$$

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis,
DOI 10.1007/978-0-8176-4980-7_3,

Example 3.1.2 (Sequences in $\mathbb{R}$) Consider the normed vector space $\mathbb{R}$. Then the reader can verify the following:

(i) The sequence $\{x_k\}_{k=1}^{\infty}$ given by $x_k := k^2$ is not a Cauchy sequence.

(ii) The sequence $\{x_k\}_{k=1}^{\infty}$ given by $x_k := 3 + 1/k$ is a Cauchy sequence (you might verify this geometrically rather than analytically). □

In any normed vector space, a convergent sequence is automatically a Cauchy sequence. We state the result formally, and ask the reader to provide the proof (Exercise 3.1):

Lemma 3.1.3 (Convergent sequences are Cauchy sequences) *Assume that V is a normed vector space, and that $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a convergent sequence in V. Then $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a Cauchy sequence.*

The opposite does not hold in general: there are normed vector spaces, in which there exist nonconvergent Cauchy sequences. But in many important normed vector spaces, a sequence is convergent if and only if it is a Cauchy sequence. Such spaces are called *Banach spaces:*

Definition 3.1.4 (Banach space) *A normed vector space V with the property that each Cauchy sequence $\{\mathbf{v}_k\}_{k=1}^{\infty}$ in V converges toward some $\mathbf{v} \in V$, is called a Banach space.*

In cases where the relevant vector space V is a subspace of a larger space, it is important to notice that there are two requirements in Definition 3.1.4:

- Each Cauchy sequence of elements in V must be convergent;
- The limit of the Cauchy sequence must belong to V.

See Exercises 5.4 and 6.1 for natural examples of spaces that are not Banach spaces.

Example 3.1.5 ($\mathbb{R}^n$ and $\mathbb{C}^n$ are Banach spaces) When equipped with the norm considered in Example 2.1.3, the spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ have the property that all Cauchy sequences are convergent, so both spaces are Banach spaces. We give the proof for the space $\mathbb{R}$ and leave the extension to the spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ to the reader (Exercise 3.2).

Let $\{c_k\}_{k=1}^{\infty}$ be a Cauchy sequence in $\mathbb{R}$. Using the definition with $\epsilon = 1$, there exists an $N \in \mathbb{N}$ such that

$$|c_k - c_\ell| \leq 1$$

whenever $k, \ell \geq N$. Thus, for $k \geq N$,

$$|c_k| = |c_k - c_N + c_N| \leq |c_k - c_N| + |c_N| \leq 1 + |c_N|,$$

which proves that $\{c_k\}_{k=N}^{\infty}$ is bounded. Therefore, also $\{c_k\}_{k=1}^{\infty}$ is bounded. By Lemma 1.5.9 the sequence $\{c_k\}_{k=1}^{\infty}$ has an accumulation point, to be denoted by c. We will show that $\{c_k\}_{k=1}^{\infty}$ converges toward c. In order to do so, let $\epsilon > 0$ be given. Choose $K \in \mathbb{N}$ such that

$$|c_k - c_\ell| \leq \epsilon/2 \text{ whenever } k, \ell \geq K.$$

Since c is an accumulation point for $\{c_k\}_{k=1}^{\infty}$, there exists $k \geq K$ such that

$$|c - c_k| \leq \epsilon/2.$$

Thus, for any $\ell \geq K$,

$$\begin{aligned} |c - c_\ell| = |(c - c_k) + (c_k - c_\ell)| &\leq |c - c_k| + |c_k - c_\ell| \\ &\leq \epsilon/2 + \epsilon/2 \\ &= \epsilon. \end{aligned}$$

This proves that $\lim_{k\to\infty} c_k = c$, and thus that $\mathbb{R}$ is a Banach space. □

One can actually prove a much stronger result than the one in Example 3.1.5: all normed *finite-dimensional* vector spaces are Banach spaces. Let us consider an important infinite-dimensional Banach space.

Theorem 3.1.6 (Continuous functions on bounded interval) *The vector space $C[a,b]$ is a Banach space with respect to the norm*

$$||f||_\infty = \max_{x\in[a,b]} |f(x)|. \tag{3.1}$$

Proof. That the expression in (3.1) defines a norm on $C[a,b]$ is shown in Example 2.1.4. In order to show that $C[a,b]$ is a Banach space, consider a Cauchy sequence of functions $\{f_k\}_{k=1}^{\infty}$ belonging to $C[a,b]$. According to the definition, this means that for any given $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$||f_k - f_\ell||_\infty \leq \epsilon \text{ whenever } k, \ell \geq N. \tag{3.2}$$

We need to show that the sequence $\{f_k\}_{k=1}^{\infty}$ converges to a function $f \in C[a,b]$. In order to find the relevant function f, fix an $x \in [a,b]$. Now, because

$$|f_k(x) - f_\ell(x)| \leq ||f_k - f_\ell||_\infty,$$

we know that $\{f_k(x)\}_{k=1}^{\infty}$ is a Cauchy sequence in $\mathbb{C}$; because $\mathbb{C}$ is a Banach space, this means that the sequence is convergent. That is, we can define a function f by

$$f(x) := \lim_{k\to\infty} f_k(x), \;\; x \in [a,b].$$

We have defined the function f as the *pointwise* limit of the functions f_k, but we have to prove that it even holds that $f_k \to f$ in $C[a,b]$, i.e., that

$$||f - f_k||_\infty \to 0 \text{ as } k \to \infty.$$

In order to do this, let $\epsilon > 0$ be given. It follows from (3.2) that we can choose $N \in \mathbb{N}$ such that

$$|f_k(x) - f_\ell(x)| \le \epsilon$$

for all $x \in [a,b]$ whenever $k, \ell \ge N$. Letting $\ell \to \infty$, it follows that

$$|f_k(x) - f(x)| \le \epsilon$$

for all $x \in [a,b]$ whenever $k \ge N$; this proves that $||f_k - f||_\infty \le \epsilon$ if $k \ge N$, i.e., that $f_k \to f$ in $C[a,b]$. In the terminology used in Section 1.6 we have proved that the sequence $\{f_k\}_{k=1}^\infty$ converges uniformly to f; that $f \in C[a,b]$ is thus a consequence of Theorem 1.6.6. □

3.2 The Banach spaces $\ell^1(\mathbb{N})$ and $\ell^p(\mathbb{N})$

With the notation used in Section 1.1, vectors $\mathbf{x}$ in $\mathbb{R}^n$ and $\mathbb{C}^n$ are finite sequences,

$$\mathbf{x} = (x_1, x_2, \dots, x_n).$$

In this section we discuss certain important Banach spaces, consisting of *infinite sequences.* For convenience, we will index the sequences by $\mathbb{N}$ and write them as

$$\mathbf{x} = (x_1, x_2, \dots, x_n, \dots).$$

Often we will also write a sequence as $\mathbf{x} = \{x_k\}_{k=1}^\infty$. We first define the vector space $\ell^1(\mathbb{N})$, consisting of all absolutely summable scalar sequences:

$$\ell^1(\mathbb{N}) := \left\{ \{x_k\}_{k=1}^\infty \mid x_k \in \mathbb{C}, \ \sum_{k=1}^\infty |x_k| < \infty \right\}.$$

Theorem 3.2.1 (The Banach space $\ell^1(\mathbb{N})$) *The space $\ell^1(\mathbb{N})$ is a Banach space with respect to the norm*

$$||\mathbf{x}||_1 = \sum_{k=1}^\infty |x_k|, \ \mathbf{x} \in \ell^1(\mathbb{N}). \tag{3.3}$$

Proof. We leave it to the reader as Exercise 3.3 to verify that the expression (3.3) defines a norm on $\ell^1(\mathbb{N})$. In order to show that $\ell^1(\mathbb{N})$ is a Banach space, assume that $\{\mathbf{x}^{(\ell)}\}_{\ell=1}^\infty$ is a Cauchy sequence of elements in $\ell^1(\mathbb{N})$. Writing $\mathbf{x}^{(\ell)} := \{x_k^{(\ell)}\}_{k=1}^\infty$, this means that for each $\epsilon > 0$, we can find $N \in \mathbb{N}$ such that

$$\left|\left|\mathbf{x}^{(m)} - \mathbf{x}^{(\ell)}\right|\right|_1 = \sum_{k=1}^\infty |x_k^{(m)} - x_k^{(\ell)}| \leq \epsilon \text{ for all } \ell, m > N. \tag{3.4}$$

Fix $k \in \mathbb{N}$, and consider the kth entry in the sequences $\mathbf{x}^\ell,\ \ell \in \mathbb{N}$. It follows from (3.4) that $\{x_k^{(\ell)}\}_{\ell=1}^\infty$ forms a Cauchy sequence in $\mathbb{C}$. Because $\mathbb{C}$ is a Banach space, each of these sequences is convergent; let

$$x_k := \lim_{\ell\to\infty} x_k^{(\ell)},\ \ k \in \mathbb{N}.$$

Now, put $\mathbf{x} = \{x_k\}_{k=1}^\infty$. In order to show that $\ell^1(\mathbb{N})$ is a Banach space, it is enough to show that $\mathbf{x} \in \ell^1(\mathbb{N})$ and that

$$||\mathbf{x} - \mathbf{x}^{(\ell)}||_1 \to 0 \text{ as } \ell \to \infty. \tag{3.5}$$

Note that

$$\begin{aligned} ||\mathbf{x} - \mathbf{x}^{(\ell)}||_1 &= \left|\left|\{x_k\}_{k=1}^\infty - \{x_k^{(\ell)}\}_{k=1}^\infty\right|\right|_1 \\ &= \sum_{k=1}^\infty |x_k - x_k^{(\ell)}| \\ &= \sum_{k=1}^\infty \lim_{m\to\infty} |x_k^{(m)} - x_k^{(\ell)}|. \end{aligned}$$

Since $\lim_{m\to\infty} |x_k^{(m)} - x_k^{(\ell)}|$ exists, Lemma 1.5.9 tells us that we can replace $\lim_{m\to\infty} |x_k^{(m)} - x_k^{(\ell)}|$ by $\liminf_{m\to\infty} |x_k^{(m)} - x_k^{(\ell)}|$. Now, given $\epsilon > 0$, choose $N \in \mathbb{N}$ as in (3.4). Then, for $\ell \geq N$, the above calculation together with Lemma 1.5.12 shows that

$$\begin{aligned} ||\mathbf{x} - \mathbf{x}^{(\ell)}||_1 &= \sum_{k=1}^\infty \liminf_{m\to\infty} |x_k^{(m)} - x_k^{(\ell)}| \\ &\leq \liminf_{m\to\infty} \sum_{k=1}^\infty |x_k^{(m)} - x_k^{(\ell)}| \\ &= \liminf_{m\to\infty} \left|\left|\mathbf{x}^{(m)} - \mathbf{x}^{(\ell)}\right|\right|_1 \\ &\leq \epsilon. \end{aligned}$$

This implies that $\mathbf{x} \in \ell^1(\mathbb{N})$ (why?), and that (3.5) holds. □

Example 3.2.2 (A basis for $\ell^1(\mathbb{N})$) Consider the vector space $\ell^1(\mathbb{N})$. For $k \in \mathbb{N}$, let $\delta_k \in \ell^1(\mathbb{N})$ be the vector with 1 at the kth entry and 0 otherwise, i.e.,

$$\delta_k = (0, \dots, 0, 1, 0, \dots 0, \dots). \tag{3.6}$$

We want to prove that $\{\delta_k\}_{k=1}^\infty$ is a basis for $\ell^1(\mathbb{N})$. According to Definition 2.5.4 we need to show that for any $\mathbf{u} \in \ell^1(\mathbb{N})$, there exist unique coefficients $c_k \in \mathbb{C}$ such that

$$\mathbf{u} = \sum_{k=1}^\infty c_k \delta_k. \tag{3.7}$$

Note that by definition of an infinite series in a normed vector space, (3.7) means that

$$\left|\left|\mathbf{u} - \sum_{k=1}^N c_k\delta_k\right|\right|_1 \to 0 \text{ as } N \to \infty. \tag{3.8}$$

Writing

$$\mathbf{u} = (u_1, u_2, \dots),$$

we see that

$$\begin{aligned}
&\left|\left|\mathbf{u} - \sum_{k=1}^N c_k\delta_k\right|\right|_1 \\
= \;& ||(u_1, u_2, \dots, u_N, u_{N+1}, u_{N+2}, \dots) - (c_1, c_2, \dots, c_N, 0, 0, \dots)||_1 \\
= \;& ||(u_1 - c_1, u_2 - c_2, \dots, u_N - c_N, u_{N+1}, u_{N+2}, \dots)||_1 \\
= \;& \sum_{k=1}^N |u_k - c_k| + \sum_{k=N+1}^\infty |u_k|. \qquad (3.9)
\end{aligned}$$

If we want (3.8) to hold, (3.9) shows that the only possibility is to let $c_k := u_k$ for all $k \in \mathbb{N}$. With this choice, (3.9) yields that

$$\begin{aligned}
\left|\left|\mathbf{u} - \sum_{k=1}^N u_k\delta_k\right|\right|_1 &= \sum_{k=N+1}^\infty |u_k| \\
&\to 0 \text{ as } N \to \infty.
\end{aligned}$$

This shows that

$$\mathbf{u} = \sum_{k=1}^\infty u_k\delta_k, \tag{3.10}$$

and that $\{\delta_k\}_{k=1}^\infty$ is a basis for $\ell^1(\mathbb{N})$. □

The statement in Theorem 3.2.1 can be extended to a result about a class of sequence spaces, called ℓ^p-spaces. Given any number $p \in [1, \infty[$, let

$$\ell^p(\mathbb{N}) := \left\{ \{x_k\}_{k=1}^{\infty} \mid x_k \in \mathbb{C}, \ \sum_{k=1}^{\infty} |x_k|^p < \infty \right\}.$$

Theorem 3.2.3 (The Banach spaces $\ell^p(\mathbb{N})$) *For $p \in [1, \infty[$, the space $\ell^p(\mathbb{N})$ is a Banach space with respect to the norm*

$$||\mathbf{x}||_p = \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p}, \ \mathbf{x} \in \ell^p(\mathbb{N}). \tag{3.11}$$

The proof of Theorem 3.2.3 is similar to the proof of Theorem 3.2.1 and is left to the reader (Exercise 3.4).

The sequence spaces $\ell^p(\mathbb{N})$ play an important role in applied mathematics. In fact, even if a concrete model is formulated in terms of functions, a computer-based implementation will always involve a transition from the setting of functions to the setting of discrete sequences of numbers, typically belonging to an ℓ^p-space.

Actually, the procedure goes one step further: before a computer-based calculation can take place, the model has to be formulated in terms of a *finite* sequence of numbers. This explains the role played by the following result, which shows that any element in $\ell^p(\mathbb{N})$ can be approximated by a sequence $\{x_k\}_{k=1}^{\infty}$ for which only finitely many entries x_k are nonzero:

Lemma 3.2.4 (Dense nonclosed subspace of $\ell^p(\mathbb{N})$) *Let*

$$V = \{ \{x_k\}_{k=1}^{\infty} \mid x_k \in \mathbb{C}, \ \forall\, k \in \mathbb{N}, \ \textit{and only finitely many } x_k \textit{ are nonzero} \}.$$

For any $p \in [1, \infty[$ the following hold:

(i) V is a subspace of $\ell^p(\mathbb{N})$.

(ii) V is dense in $\ell^p(\mathbb{N})$.

(iii) V does not form a closed subset of $\ell^p(\mathbb{N})$.

We leave the proof of Lemma 3.2.4 to the reader as Exercise 3.5.

Let us finally define the space $\ell^\infty(\mathbb{N})$, consisting of all bounded sequences:

$$\ell^\infty(\mathbb{N}) := \left\{ \{x_k\}_{k=1}^{\infty} \mid x_k \in \mathbb{C}, \ \sup_{k \in \mathbb{N}} |x_k| < \infty \right\}. \tag{3.12}$$

One can prove (Exercise 3.10) that $\ell^\infty(\mathbb{N})$ is a Banach space with respect to the norm

$$||\mathbf{x}||_\infty = \sup_{k \in \mathbb{N}} |x_k|. \tag{3.13}$$

For an explanation of the name $\ell^\infty(\mathbb{N})$ we refer to Exercise 3.11.

3.3 Linear operators on Banach spaces

We will now continue the analysis of bounded linear operators initiated in Section 2.4. Let us first consider an operator on $\ell^1(\mathbb{N})$ and some of its properties:

Example 3.3.1 (Bounded operator on $\ell^1(\mathbb{N})$) Consider the operator $T : \ell^1(\mathbb{N}) \to \ell^1(\mathbb{N})$ given by

$$T\{x_k\}_{k=1}^{\infty} = \{\frac{1}{k}\,x_k\}_{k=1}^{\infty}.$$

Written out in coordinates, we have

$$T(x_1, x_2, \dots, x_n, \dots) = \big(x_1, \frac{1}{2}x_2, \dots, \frac{1}{n}x_n, \dots\big).$$

We will check that T is a bounded linear operator. First, we show that T actually maps $\ell^1(\mathbb{N})$ into $\ell^1(\mathbb{N})$. Letting $\mathbf{x} = (x_1, x_2, \dots, x_n, \dots) \in \ell^1(\mathbb{N})$,

$$||T\{x_k\}_{k=1}^{\infty}||_1 = \sum_{k=1}^{\infty} \left|\frac{1}{k}\,x_k\right| \leq \sum_{k=1}^{\infty} |x_k| < \infty. \tag{3.14}$$

This shows that $T\mathbf{x} \in \ell^1(\mathbb{N})$, as claimed. By inspection it is clear that T is linear; we now want to show that T is bounded. The calculation in (3.14) shows that

$$||T\mathbf{x}||_1 \leq \sum_{k=1}^{\infty} |x_k| = ||\mathbf{x}||_1.$$

Thus, the operator T is bounded and $||T|| \leq 1$. In fact, since

$$||T(1, 0, \dots, 0, \dots)||_1 = ||(1, 0, \dots, 0, \dots)||_1,$$

we have $||T|| = 1$. The operator T is not isometric. For example, letting $\mathbf{x} = (0, 1, 0, 0, \dots, 0, \dots)$, we have $||\mathbf{x}||_1 = 1$, but

$$||T\mathbf{x}||_1 = \left|\left|\big(0, \frac{1}{2}, 0, 0, \dots, 0, \dots\big)\right|\right|_1 = \frac{1}{2}.$$

Thus, for this particular $\mathbf{x}$, we see that $||T\mathbf{x}||_1 \neq ||\mathbf{x}||_1$. □

Note that the first step in Example 3.3.1 was to show that T actually maps $\ell^1(\mathbb{N})$ into $\ell^1(\mathbb{N})$: after that, the calculation immediately gave that T is bounded. This is very typical for analysis of operators.

For technical reasons it is often difficult to define linear operators on infinite-dimensional vector spaces. For example, if we want to define an operator acting on $\ell^1(\mathbb{N})$ we might need to worry about convergence issues. In such cases it is an advantage first to define the operator on a convenient subspace, and *then* try to extend it to the entire space. For example, in the case of an operator acting on $\ell^1(\mathbb{N})$ the natural subspace to consider is the one defined in Lemma 3.2.4.

The next result shows that under certain circumstances, such an operator can be *extended* to a bounded operator acting on the entire Banach space:

Theorem 3.3.2 (Extension of bounded linear operator) *Let V_1 and V_2 be Banach spaces, W a dense subspace of V_1, and $T : W \to V_2$ a bounded linear operator. Then there exists a unique bounded linear operator*

$$\widetilde{T} : V_1 \to V_2$$

for which $\widetilde{T}\mathbf{v} = T\mathbf{v}$ for all $\mathbf{v} \in W$. The operator $\widetilde{T}$ satisfies that $||\widetilde{T}|| = ||T||$.

Proof. Let $\mathbf{v} \in V_1$. Because W is dense in V_1, we can select a sequence of elements $\{\mathbf{v}_k\}_{k=1}^{\infty}$ in W such that $\mathbf{v}_k \to \mathbf{v}$ as $k \to \infty$. Now, for $k, \ell \in \mathbb{N}$,

$$||T\mathbf{v}_k - T\mathbf{v}_\ell|| = ||T(\mathbf{v}_k - \mathbf{v}_\ell)|| \leq ||T||\,||\mathbf{v}_k - \mathbf{v}_\ell||.$$

This implies that $\{T\mathbf{v}_k\}_{k=1}^{\infty}$ is a Cauchy sequence in V_2 and therefore convergent. We now define

$$\widetilde{T}\mathbf{v} := \lim_{k\to\infty} T\mathbf{v}_k. \tag{3.15}$$

One can check that this definition is independent of the choice of the sequence $\{\mathbf{v}_k\}_{k=1}^{\infty}$, see Exercise 3.12. Next, we verify that the operator $\widetilde{T}$ is linear. Given any $\mathbf{v}, \mathbf{w} \in V_1$ and any $\alpha, \beta \in \mathbb{C}$, take sequences $\{\mathbf{v}_k\}_{k=1}^{\infty}$ and $\{\mathbf{w}_k\}_{k=1}^{\infty}$ in W such that $\mathbf{v}_k \to \mathbf{v}$ and $\mathbf{w}_k \to \mathbf{w}$ as $k \to \infty$. Then

$$\alpha\mathbf{v}_k + \beta\mathbf{w}_k \to \alpha\mathbf{v} + \beta\mathbf{w} \text{ as } k \to \infty,$$

so

$$\begin{aligned}\widetilde{T}(\alpha\mathbf{v} + \beta\mathbf{w}) &= \lim_{k\to\infty} T(\alpha\mathbf{v}_k + \beta\mathbf{w}_k)\\ &= \lim_{k\to\infty} (\alpha T\mathbf{v}_k + \beta T\mathbf{w}_k)\\ &= \alpha\widetilde{T}\mathbf{v} + \beta\widetilde{T}\mathbf{w}.\end{aligned}$$

This proves that $\widetilde{T}$ is linear. In order to show that $\widetilde{T}$ is bounded, let $\mathbf{v} \in V_1$ and take again a sequence $\{\mathbf{v}_k\}_{k=1}^{\infty}$ in W such that $\mathbf{v}_k \to \mathbf{v}$. Then, via the result in Exercise 2.1,

$$\begin{aligned}||\widetilde{T}\mathbf{v}|| &= ||\lim_{k\to\infty} T\mathbf{v}_k|| && (3.16)\\ &= \lim_{k\to\infty} ||T\mathbf{v}_k||\\ &\leq \lim_{k\to\infty} (||T||\,||\mathbf{v}_k||) && (3.17)\\ &= ||T||\,||\mathbf{v}||.\end{aligned}$$

This proves that $\widetilde{T}$ is bounded and that $||\widetilde{T}|| \leq ||T||$. We leave it to the reader to argue that $||\widetilde{T}|| \geq ||T||$, which concludes the proof. □

3.4 Exercises

3.1 Prove Lemma 3.1.3.

3.2 Let x_ℓ, $\ell \in \mathbb{N}$, be a Cauchy sequence of complex numbers.

(i) Show that the real part of the sequence as well as the imaginary part of the sequence form a Cauchy sequence.

(ii) Using that $\mathbb{R}$ is a Banach space, show that $\mathbb{C}$ is a Banach space.

For each $\ell \in \mathbb{N}$, let $\mathbf{x}^{(\ell)}$ denote a sequence in $\mathbb{R}^n$, with coordinates $(x_1^{(\ell)}, x_2^{(\ell)}, \dots, x_n^{(\ell)})$.

(iii) Show that $\mathbf{x}^{(\ell)}$, $\ell \in \mathbb{N}$, is a Cauchy sequence in $\mathbb{R}^n$ if and only if the coordinate sequences $x_k^{(\ell)}$, $\ell \in \mathbb{N}$, are Cauchy sequences in $\mathbb{R}$ for each $k = 1, \dots, n$.

(iv) Conclude that $\mathbb{R}^n$ and $\mathbb{C}^n$ are Banach spaces.

3.3 Prove that the expression (3.3) defines a norm on $\ell^1(\mathbb{N})$.

3.4 For $p \in [1, \infty[$, consider the space $\ell^p(\mathbb{N})$.

(i) Prove that (3.11) defines a norm on $\ell^p(\mathbb{N})$. (*Hint:* use Theorem 1.7.3.)

(ii) Prove that $\ell^p(\mathbb{N})$ is a Banach space with respect to the norm (3.11).

3.5 The goal of the exercise is to prove Lemma 3.2.4. Fix any $p \in [1, \infty[$.

(i) Show that V is a subset of $\ell^p(\mathbb{N})$.

(ii) Argue that V is a subspace of $\ell^p(\mathbb{N})$.

(iii) Show that V is dense in $\ell^p(\mathbb{N})$. (*Hint:* any sequence $\{x_k\}_{k=1}^\infty$ in $\ell^p(\mathbb{N})$ can be approximated arbitrarily well by a sequence obtained by replacing x_k by zero for sufficiently large indices.)

(iv) Show that V does not form a closed subset of $\ell^p(\mathbb{N})$.

3.6 Consider Example 3.2.2. What is the set $\text{span}\{\mathbf{v}_k\}_{k=1}^\infty$?

3.7 Let $p \in [1, \infty[$. Show that the vectors $\{\delta_k\}_{k=1}^{\infty}$ defined in (3.6) form a basis for $\ell^p(\mathbb{N})$.

3.8 Consider the set V defined in Lemma 3.2.4.

(i) Show that V is a subset of $\ell^{\infty}(\mathbb{N})$.

(ii) Show that V is a subspace of $\ell^{\infty}(\mathbb{N})$.

(iii) Show that V is not dense in $\ell^{\infty}(\mathbb{N})$.
Hint: consider $\mathbf{x} = \{x_k\}_{k=1}^{\infty} \in \ell^{\infty}(\mathbb{N})$ given by $x_k = 1$ for all $k \in \mathbb{N}$ and show that $||\mathbf{x} - \mathbf{y}||_{\infty} \geq 1$ for all $\mathbf{y} = \{y_k\}_{k=1}^{\infty} \in V$.

(iv) Show that V does not form a closed subset of $\ell^{\infty}(\mathbb{N})$.
Hint: consider $\mathbf{x} = \{x_k\}_{k=1}^{\infty} \in \ell^{\infty}(\mathbb{N})$ given by $x_k = 1/k$ for all $k \in \mathbb{N}$ and find sequences $\mathbf{y}_n \in V$ such that $||\mathbf{x} - \mathbf{y}_n||_{\infty} \to 0$ as $n \to \infty$.

3.9 Consider a sequence $\{w_k\}_{k=1}^{\infty}$ of positive real numbers, and define the weighted ℓ^1-space $\ell^1_w(\mathbb{N})$ by

$$\ell^1_w(\mathbb{N}) := \left\{ \{x_k\}_{k=1}^{\infty} \mid x_k \in \mathbb{C}, \quad \sum_{k=1}^{\infty} |x_k|\, w_k < \infty \right\}.$$

(i) Show that the expression $|| \cdot ||$ given by

$$||\{x_k\}_{k=1}^{\infty}|| := \sum_{k=1}^{\infty} |x_k|\, w_k$$

defines a norm on $\ell^1_w(\mathbb{N})$.
We now consider the special choice

$$w_k := 2^k, \ k \in \mathbb{N}.$$

(ii) Show that $\ell^1_w(\mathbb{N})$ is a subspace of $\ell^1(\mathbb{N})$.

(iii) Find a sequence $\{x_k\}_{k=1}^{\infty}$ belonging to $\ell^1(\mathbb{N})$, but not to $\ell^1_w(\mathbb{N})$.

(iv) Show that the left-shift operator

$$T(x_1, x_2, \dots) = (x_2, x_3, \dots)$$

is bounded from $\ell^1_w(\mathbb{N})$ into $\ell^1_w(\mathbb{N})$.

(v) Calculate the exact value of the norm of the operator T considered in (iv).

3.10 This exercise concerns the space $\ell^\infty(\mathbb{N})$ defined in (3.12).

(i) Show that $\ell^\infty(\mathbb{N})$ is a vector space and that (3.13) defines a norm.

(ii) Show that $\ell^\infty(\mathbb{N})$ is a Banach space with respect to the norm in (3.13).

3.11 The purpose of this exercise is to motivate the name $\ell^\infty(\mathbb{N})$ for the space in (3.12). We consider the vector space $\mathbb{C}^n$ equipped with various norms. For $\mathbf{x} \in \mathbb{C}^n$, write

$$\mathbf{x} = (x_1, x_2, \dots, x_n).$$

(i) Show that for any $p \in [1, \infty[$,

$$||\mathbf{x}||_p := \left(\sum_{k=1}^n |x_k|^p \right)^{1/p}$$

defines a norm on $\mathbb{C}^n$.

(ii) Show that

$$||\mathbf{x}||_\infty := \max_{k \in \{1,\dots,n\}} |x_k|$$

defines a norm on $\mathbb{C}^n$.

(iii) Show that for any $\mathbf{x} \in \mathbb{C}^n$,

$$||\mathbf{x}||_p \to ||\mathbf{x}||_\infty \text{ as } p \to \infty.$$

3.12 Complete the proof of Theorem 3.3.2 by showing that the limit in (3.15) is independent of the choice of the sequence $\{\mathbf{v}_k\}_{k=1}^\infty$. *Hint:* consider two sequences $\{\mathbf{v}_k\}_{k=1}^\infty$ and $\{\mathbf{w}_k\}_{k=1}^\infty$, both converging to $\mathbf{v}$, and a sequence obtained by mixing elements from $\{\mathbf{v}_k\}_{k=1}^\infty$ and $\{\mathbf{w}_k\}_{k=1}^\infty$.

3.13 Consider the mapping

$$T : C[0,2] \to C[0,2], \ (Tf)(x) := x^2 f(x), \ x \in [0,2].$$

(i) Show that T is well defined, i.e., T actually maps $C[0,2]$ into $C[0,2]$.

(ii) Show that T is linear and bounded.

3.14 This exercise deals with invertible operators on finite-dimensional and infinite-dimensional vector spaces.

(i) Assume that $T : \mathbb{C}^n \to \mathbb{C}^n$ is a linear operator. Show that if there exists an operator $S : \mathbb{C}^n \to \mathbb{C}^n$ such that $ST = I$, then it also holds that $TS = I$.

Now let $V = \ell^1(\mathbb{N})$, and consider the *right-shift operator*

$$T : \ell^1(\mathbb{N}) \to \ell^1(\mathbb{N}),\ T(x_1, x_2, \dots, x_n, \dots) = (0, x_1, x_2, \dots, x_n, \dots),$$

and the *left-shift operator*

$$S : \ell^1(\mathbb{N}) \to \ell^1(\mathbb{N}),\ S(x_1, x_2, \dots, x_n, \dots) = (x_2, x_3, \dots, x_n, \dots).$$

(ii) Show that T and S are bounded linear operators.

(iii) Show that $||T|| = ||S|| = 1$.

(iv) Is T an isometry? Is S an isometry?

(v) Is T injective? Is S injective?

(vi) Is T surjective? Is S surjective?

(vii) Show that $ST = I$, but $TS \neq I$.

3.15 Let $p \in [1, \infty[$ and consider the mapping

$$T : \ell^p(\mathbb{N}) \to \ell^p(\mathbb{N}),\ T\{x_k\}_{k=1}^{\infty} := \{x_k + x_{k+1}\}_{k=1}^{\infty}.$$

(i) Show that T actually maps $\ell^p(\mathbb{N})$ into $\ell^p(\mathbb{N})$.

(ii) Show that T is linear and bounded.

4

Hilbert Spaces

Hilbert spaces can be considered as infinite-dimensional analogues of $\mathbb{R}^n$ and $\mathbb{C}^n$: in fact, the structure imposed on Hilbert spaces imply that many of the properties of $\mathbb{R}^n$ and $\mathbb{C}^n$ (and the ways to deal with them) can be extended to Hilbert spaces. Most importantly, a Hilbert space is equipped with an inner product and an associated norm that makes the Hilbert space a Banach space.

The Hilbert spaces are introduced in Section 4.1. We prove several relationships between the inner product and the associated norm. A concrete Hilbert space, the sequence space $\ell^2(\mathbb{N})$, is considered in Section 4.2. Section 4.3 deals with the concept of orthogonality, and proves an important decomposition of a Hilbert space in terms of a closed subspace and its orthogonal complement. Section 4.4 introduces functionals on Hilbert spaces and their properties. In Section 4.5 we continue the analysis of linear operators initiated in Section 2.4, but now on Hilbert spaces. Section 4.6 introduces Bessel sequences, which allow one to obtain convergent series expansions in a convenient way. Orthonormal bases are considered in Section 4.7. Several characterizations of orthonormal bases are proved. In particular, it is shown that orthonormal bases lead to expansions that converge regardless how the vectors are ordered. Finally, Section 4.8 gives a very short introduction to series expansions in Hilbert spaces obtained via so-called frames, a concept that is more general than orthonormal bases.

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4980-7_4, © Springer Science+Business Media, LLC 2010

4.1 Inner product spaces

The norm on $\mathbb{C}^n$ considered in (1.3) arises from the inner product in (1.2) via $||\mathbf{x}|| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$. We will now introduce inner products on general vector spaces and show that they give rise to norms in the same fashion.

Definition 4.1.1 (Inner product space) *Let V be a (complex) vector space. An inner product on V is a mapping*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C} \tag{4.1}$$

for which

(i) $\langle \alpha \mathbf{v} + \beta \mathbf{w}, \mathbf{u} \rangle = \alpha \langle \mathbf{v}, \mathbf{u} \rangle + \beta \langle \mathbf{w}, \mathbf{u} \rangle, \ \forall \, \mathbf{v}, \mathbf{w}, \mathbf{u} \in V, \ \alpha, \beta \in \mathbb{C}$;

(ii) $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}, \ \forall \, \mathbf{v}, \mathbf{w} \in V$;

(iii) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0, \ \forall \, \mathbf{v} \in V$, *and* $\langle \mathbf{v}, \mathbf{v} \rangle = 0 \Leftrightarrow \mathbf{v} = \mathbf{0}$.

A vector space equipped with an inner product is called an inner product space.

The definition for a real vector space is similar, except that the inner product takes real values, the constants α and β in (i) are assumed to be real, and the complex conjugation in (ii) is superfluous.

The condition in Definition 4.1.1(i) is expressed by saying that the inner product $\langle \cdot, \cdot \rangle$ is linear in the first entry. It implies that the inner product is *antilinear* in the second entry; in fact, using the rules (i) and (ii),

$$\begin{aligned} \langle \mathbf{v}, \alpha \mathbf{w} + \beta \mathbf{u} \rangle &= \overline{\langle \alpha \mathbf{w} + \beta \mathbf{u}, \mathbf{v} \rangle} \\ &= \overline{\alpha \langle \mathbf{w}, \mathbf{v} \rangle + \beta \langle \mathbf{u}, \mathbf{v} \rangle} \\ &= \overline{\alpha} \langle \mathbf{v}, \mathbf{w} \rangle + \overline{\beta} \langle \mathbf{v}, \mathbf{u} \rangle, \ \forall \, \mathbf{v}, \mathbf{w}, \mathbf{u} \in V, \ \alpha, \beta \in \mathbb{C}. \end{aligned}$$

In various areas of science and engineering, inner products are defined slightly different: the inner products are linear in the second entry and antilinear in the first. See, e.g., the discussion in Example 6.1.3 in the context of quantum mechanics.

In any vector space V with an inner product $\langle \cdot, \cdot \rangle$, *Cauchy–Schwarz' inequality* holds:

Theorem 4.1.2 (Cauchy–Schwarz' inequality) *Let V be a vector space with an inner product $\langle \cdot, \cdot \rangle$. Then*

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \langle \mathbf{v}, \mathbf{v} \rangle^{1/2} \ \langle \mathbf{w}, \mathbf{w} \rangle^{1/2}, \ \ \forall \, \mathbf{v}, \mathbf{w} \in V.$$

Proof. The result certainly holds if $\mathbf{v} = \mathbf{0}$ (Exercise 4.1), so we can assume that $\mathbf{v} \neq \mathbf{0}$. We first prove the result for the case where $\langle \mathbf{v}, \mathbf{w} \rangle$ is real. In this case, the properties for the inner product imply that

$$\begin{aligned} 0 &\le \langle \mathbf{w} - \frac{\langle \mathbf{v}, \mathbf{w}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}, \mathbf{w} - \frac{\langle \mathbf{v}, \mathbf{w}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}\rangle \\ &= \langle \mathbf{w}, \mathbf{w}\rangle - 2\,\frac{\langle \mathbf{v}, \mathbf{w}\rangle^2}{\langle \mathbf{v}, \mathbf{v}\rangle} + \frac{\langle \mathbf{v}, \mathbf{w}\rangle^2}{\langle \mathbf{v}, \mathbf{v}\rangle^2}\langle \mathbf{v}, \mathbf{v}\rangle \\ &= \frac{\langle \mathbf{w}, \mathbf{w}\rangle\langle \mathbf{v}, \mathbf{v}\rangle - \langle \mathbf{v}, \mathbf{w}\rangle^2}{\langle \mathbf{v}, \mathbf{v}\rangle}. \end{aligned}$$

Because $\langle \mathbf{v}, \mathbf{v}\rangle > 0$ we obtain that

$$\langle \mathbf{v}, \mathbf{w}\rangle^2 \le \langle \mathbf{w}, \mathbf{w}\rangle\langle \mathbf{v}, \mathbf{v}\rangle;$$

this yields the result in case $\langle \mathbf{v}, \mathbf{w}\rangle$ is real.

We now consider the case where $\langle \mathbf{v}, \mathbf{w}\rangle$ is complex. Choose $\lambda \in \mathbb{C}$ such that $|\lambda| = 1$ and $\lambda\langle \mathbf{v}, \mathbf{w}\rangle \in [0, \infty[$; then

$$|\langle \mathbf{v}, \mathbf{w}\rangle| = |\lambda\langle \mathbf{v}, \mathbf{w}\rangle| = \langle \lambda\mathbf{v}, \mathbf{w}\rangle.$$

We can now apply the partial result just proved to the vectors $\lambda\mathbf{v}$ and $\mathbf{w}$; this leads to

$$\begin{aligned} |\langle \mathbf{v}, \mathbf{w}\rangle| &= \langle \lambda\mathbf{v}, \mathbf{w}\rangle \\ &\le \langle \lambda\mathbf{v}, \lambda\mathbf{v}\rangle^{1/2}\ \langle \mathbf{w}, \mathbf{w}\rangle^{1/2} \\ &= \langle \mathbf{v}, \mathbf{v}\rangle^{1/2}\ \langle \mathbf{w}, \mathbf{w}\rangle^{1/2}, \end{aligned}$$

as desired. □

We now prove that any vector space V with an inner product $\langle \cdot, \cdot\rangle$ can be equipped with a natural norm:

Lemma 4.1.3 (Norm in inner product space) *Let V be a vector space with an inner product $\langle \cdot, \cdot\rangle$. Then*

$$||\mathbf{v}|| := \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle},\ \ \mathbf{v} \in V, \tag{4.2}$$

defines a norm on V.

Proof. We verify that the expression for $||\cdot||$ in (4.2) satisfies the conditions in Definition 2.1.1. The property (i) in Definition 2.1.1 immediately follows from the property (iii) in Definition 4.1.1. Now, for any $\mathbf{v} \in V$ and $\alpha \in \mathbb{C}$,

$$||\alpha\mathbf{v}|| = \sqrt{\langle \alpha\mathbf{v}, \alpha\mathbf{v}\rangle} = \sqrt{\alpha\overline{\alpha}\langle \mathbf{v}, \mathbf{v}\rangle} = |\alpha|\,||\mathbf{v}||;$$

this verifies the condition (ii) in Definition 2.1.1. In order to check the condition (iii) in Definition 2.1.1, let $\mathbf{v}, \mathbf{w} \in V$; then, using the result in

Theorem 4.1.2,

$$\begin{aligned}
||\mathbf{v}+\mathbf{w}||^2 &= \langle \mathbf{v}+\mathbf{w}, \mathbf{v}+\mathbf{w}\rangle \\
&= \langle \mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{w}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle + \langle \mathbf{w}, \mathbf{v}\rangle \\
&\leq ||\mathbf{v}||^2 + ||\mathbf{w}||^2 + |\langle \mathbf{v}, \mathbf{w}\rangle| + |\langle \mathbf{w}, \mathbf{v}\rangle| \\
&\leq ||\mathbf{v}||^2 + ||\mathbf{w}||^2 + 2\,||\mathbf{v}||\,||\mathbf{w}|| \\
&= \left(||\mathbf{v}|| + ||\mathbf{w}||\right)^2.
\end{aligned}$$

This completes the proof. □

Due to the conditions imposed on an inner product, one can derive several general equalities and inequalities related to inner products and the associated norm. We collect some of them here.

Theorem 4.1.4 *Let V be a vector space with an inner product $\langle \cdot, \cdot \rangle$ and associated norm $||\cdot||$. Then the following hold:*

(i) **(Cauchy–Schwarz' inequality)** *For any* $\mathbf{v}, \mathbf{w} \in V$,

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

(ii) **(The parallelogram law)** *For all* $\mathbf{v}, \mathbf{w} \in V$,

$$||\mathbf{v}+\mathbf{w}||^2 + ||\mathbf{v}-\mathbf{w}||^2 = 2\left(||\mathbf{v}||^2 + ||\mathbf{w}||^2\right). \tag{4.3}$$

(iii) **(The polarization identity in a complex vector space)** *If V is a complex vector space and* $\mathbf{v}, \mathbf{w} \in V$,

$$\langle \mathbf{v}, \mathbf{w}\rangle = \frac{1}{4}\left(||\mathbf{v}+\mathbf{w}||^2 - ||\mathbf{v}-\mathbf{w}||^2 + i(||\mathbf{v}+i\mathbf{w}||^2 - ||\mathbf{v}-i\mathbf{w}||^2)\right).$$

(iv) **(The polarization identity in a real vector space)** *If V is a real vector space and* $\mathbf{v}, \mathbf{w} \in V$,

$$\langle \mathbf{v}, \mathbf{w}\rangle = \frac{1}{4}\left(||\mathbf{v}+\mathbf{w}||^2 - ||\mathbf{v}-\mathbf{w}||^2\right).$$

(v) **(Norm versus inner product)** *For any* $\mathbf{v} \in V$,

$$||\mathbf{v}|| = \sup\{\, |\langle \mathbf{v}, \mathbf{w}\rangle| \mid \mathbf{w} \in V,\ ||\mathbf{w}|| = 1\}. \tag{4.4}$$

Note that (i) is a restatement of Theorem 4.1.2 in terms of the norm associated with the inner product. We ask the reader to prove (ii) in Exercise 4.2, (iii) and (iv) in Exercise 4.3, and (v) in Exercise 4.4. The interpretation of (iii) and (iv) is that if we know the norm in a vector space with inner product, we are able to *reconstruct* the inner product based on the norm. On the other hand, (v) shows that the norm of an arbitrary element $\mathbf{v} \in \mathcal{H}$ can be recovered based on the inner product between $\mathbf{v}$ and the elements in the unit sphere in $\mathcal{H}$. The result in (v) is frequently written in a slightly shorter form as

$$||\mathbf{v}|| = \sup_{||\mathbf{w}||=1} |\langle \mathbf{v}, \mathbf{w} \rangle|.$$

In Example 2.1.4, we considered the norm $|| \cdot ||_\infty$ on the vector space $C[a, b]$. The norm $|| \cdot ||_\infty$ does not come from an inner product (Exercise 4.5). Frequently, a vector space can be equipped with different norms, with different properties; for example, in Exercise 6.1 we ask the reader to show that $C[a, b]$ can be equipped with another norm, which actually comes from an inner product. However, with respect to that norm, $C[a, b]$ is not a Banach space.

If an inner product space V is a Banach space with respect to the norm in (4.2), then V is called a *Hilbert space:*

Definition 4.1.5 (Hilbert space) *A vector space with an inner product* $\langle \cdot, \cdot \rangle$*, which is a Banach space with respect to the norm* (4.2)*, is called a Hilbert space.*

We reserve the letter $\mathcal{H}$ for Hilbert spaces. We will always assume that $\mathcal{H}$ is *nontrivial*, i.e., that $\mathcal{H} \neq \{\mathbf{0}\}$. The standard examples are the spaces $\ell^2(\mathbb{N})$ discussed in Section 4.2, and $L^2(\mathbb{R})$ discussed in Chapter 6.

Example 4.1.6 (The Hilbert spaces $\mathbb{R}^n$ and $\mathbb{C}^n$) In Example 3.1.5 we saw that $\mathbb{R}^n$ and $\mathbb{C}^n$ are Banach spaces when equipped with the norm

$$||\mathbf{x}|| = \left(\sum_{k=1}^{n} |x_k|^2 \right)^{1/2}, \ \mathbf{x} = (x_1, x_2, \dots, x_n).$$

In both cases this norm arises from an inner product, see (1.1) and (1.2). Thus, $\mathbb{R}^n$ and $\mathbb{C}^n$ are Hilbert spaces. □

We will say much more about general Hilbert spaces later in this chapter. But before doing so we want to discuss the Hilbert space $\ell^2(\mathbb{N})$.

4.2 The Hilbert space $\ell^2(\mathbb{N})$

In this section we give the first example of an infinite-dimensional Hilbert space; it is at the same time one of the most important examples.

In Section 3.2 we introduced the sequence spaces $\ell^p(\mathbb{N})$. One of them, $\ell^2(\mathbb{N})$, can be considered as a natural infinite-dimensional analogue of $\mathbb{C}^n$:

$$\ell^2(\mathbb{N}) := \left\{ \{x_k\}_{k=1}^\infty \ \middle| \ x_k \in \mathbb{C}, \ \sum_{k=1}^{\infty} |x_k|^2 < \infty \right\}.$$

Using the notation in Section 3.2, we will denote vectors in $\ell^2(\mathbb{N})$ by

$$\mathbf{x} = \{x_k\}_{k=1}^{\infty} = (x_1, x_2, \dots, x_n, \dots).$$

Theorem 4.2.1 (The Hilbert space $\ell^2(\mathbb{N})$) *The space $\ell^2(\mathbb{N})$ is a Hilbert space with respect to the inner product*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^{\infty} x_k \overline{y_k}, \ \mathbf{x}, \mathbf{y} \in \ell^2(\mathbb{N}). \tag{4.5}$$

Proof. Already in Theorem 3.2.3 we saw that $\ell^2(\mathbb{N})$ is a Banach space with respect to the norm

$$||\mathbf{x}||_2 = \left(\sum_{k=1}^{\infty} |x_k|^2 \right)^{1/2}.$$

Thus, it is enough to show that the expression in (4.5) actually defines an inner product on $\ell^2(\mathbb{N})$. The fact that $\langle \cdot, \cdot \rangle$ maps $\ell^2(\mathbb{N}) \times \ell^2(\mathbb{N})$ into $\mathbb{C}$ follows from Theorem 1.7.3, which actually shows that the infinite sum in (4.5) converges absolutely. The verification of the properties (i)–(iii) in Definition 4.1.1 is left to the reader. □

Note that for the space $\ell^2(\mathbb{N})$, Cauchy–Schwarz' inequality states that

$$\left| \sum_{k=1}^{\infty} x_k \overline{y_k} \right|^2 \leq \sum_{k=1}^{\infty} |x_k|^2 \sum_{k=1}^{\infty} |y_k|^2, \ \forall \{x_k\}_{k=1}^{\infty}, \{y_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N});$$

this is just a slightly different formulation of Hölder's inequality in Theorem 1.7.3.

4.3 Orthogonality and direct sum decomposition

One of the key concepts in vector spaces with an inner product is *orthogonality.* We collect some related definitions here:

Definition 4.3.1 (Orthogonality, orthonormal system) *Let $\mathcal{H}$ be a Hilbert space.*

(i) Two elements $\mathbf{v}, \mathbf{w} \in \mathcal{H}$ are orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$; in that case we write $\mathbf{v} \perp \mathbf{w}$.

(ii) A collection of vectors $\{\mathbf{v}_k\}_{k=1}^{\infty}$ in $\mathcal{H}$ is an orthogonal system if

$$\langle \mathbf{v}_k, \mathbf{v}_\ell \rangle = 0 \text{ for all } k \neq \ell.$$

(iii) An orthogonal system $\{\mathbf{v}_k\}_{k=1}^{\infty}$ for which $||\mathbf{v}_k|| = 1$ for all $k \in \mathbb{N}$ is called an orthonormal system.

We will also need a few definitions concerning vector spaces that are defined using the concept of orthogonality:

Definition 4.3.2 (Orthogonal complement, direct sum) *Let $\mathcal{H}$ be a Hilbert space.*

(i) The orthogonal complement of a subspace W of $\mathcal{H}$ consists of the vectors in $\mathcal{H}$ that are orthogonal to all elements in W:

$$W^\perp := \{\mathbf{v} \in \mathcal{H} \,|\, \langle \mathbf{v}, \mathbf{w} \rangle = 0, \ \forall \, \mathbf{w} \in W\}.$$

(ii) Given two subspaces U and W of $\mathcal{H}$, the sum $U + W$ is defined by

$$U + W := \{\mathbf{v} \in \mathcal{H} \mid \mathbf{v} = \mathbf{u} + \mathbf{w} \text{ for some } \mathbf{u} \in U, \mathbf{w} \in W\}.$$

(iii) In case every $\mathbf{v} \in U + W$ has a unique representation $\mathbf{v} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U, \mathbf{w} \in W$, the sum is called a direct sum, and we write $U \oplus W$ instead of $U + W$.

Note that the sum of two subspaces again is a subspace. In this book, when considering a direct sum $U \oplus W$, the spaces U and W will always be orthogonal to each other.

A closed subspace and its orthogonal complement induce a decomposition of the underlying Hilbert space. We will state the result in Theorem 4.3.5, but in order to prove the result we need to introduce *convex sets* first:

Definition 4.3.3 (Convex set) *A subset M of a vector space V is convex if*

$$\mathbf{v}, \mathbf{w} \in M \Rightarrow \lambda \mathbf{v} + (1 - \lambda)\mathbf{w} \in M, \ \forall \lambda \in [0, 1].$$

Intuitively, a set M is convex if the line segment connecting two arbitrary points in M lies entirely in M (Exercise 4.13). By Lemma 1.2.7, any subspace of a vector space is a convex set.

Lemma 4.3.4 (Nearest point to a convex set) *Let M be a closed, nonempty, and convex subset of the Hilbert space $\mathcal{H}$. For any $\mathbf{v} \in \mathcal{H}$ there exists a uniquely determined $\mathbf{w}_0 \in M$ such that*

$$||\mathbf{v} - \mathbf{w}_0|| \leq ||\mathbf{v} - \mathbf{w}|| \tag{4.6}$$

for all $\mathbf{w} \in M$.

Proof. Fix $\mathbf{v} \in \mathcal{H}$, and let

$$\delta := \inf\{\, ||\mathbf{v} - \mathbf{w}|| \,\big|\, \mathbf{w} \in M\}.$$

By definition we can find a sequence of elements $\{\mathbf{w}_k\}_{k=1}^\infty$ in M such that

$$||\mathbf{v} - \mathbf{w}_k|| \to \delta \text{ as } k \to \infty. \tag{4.7}$$

Using the parallelogram law in (4.3),

$$\begin{aligned} ||\mathbf{w}_k - \mathbf{w}_\ell||^2 &= ||(\mathbf{w}_k - \mathbf{v}) + (\mathbf{v} - \mathbf{w}_\ell)||^2 \\ &= 2\,||\mathbf{w}_k - \mathbf{v}||^2 + 2\,||\mathbf{v} - \mathbf{w}_\ell||^2 - ||\mathbf{w}_k + \mathbf{w}_\ell - 2\mathbf{v}||^2 \\ &= 2\,||\mathbf{w}_k - \mathbf{v}||^2 + 2\,||\mathbf{v} - \mathbf{w}_\ell||^2 - 4\,||\frac{1}{2}\,(\mathbf{w}_k + \mathbf{w}_\ell) - \mathbf{v}||^2. \end{aligned} \tag{4.8}$$

By the assumption that M is convex, $\frac{1}{2}\,(\mathbf{w}_k + \mathbf{w}_\ell) \in M$; thus,

$$\delta \leq ||\frac{1}{2}\,(\mathbf{w}_k + \mathbf{w}_\ell) - \mathbf{v}||.$$

It follows that

$$||\mathbf{w}_k - \mathbf{w}_\ell||^2 \leq 2\,||\mathbf{w}_k - \mathbf{v}||^2 + 2\,||\mathbf{v} - \mathbf{w}_\ell||^2 - 4\delta^2. \tag{4.9}$$

The left-hand side of (4.9) converges to 0 as $k, \ell \to \infty$, so $\{\mathbf{w}_k\}_{k=1}^\infty$ is a Cauchy sequence in $\mathcal{H}$. Since $\mathcal{H}$ is a Hilbert space, it follows that $\{\mathbf{w}_k\}_{k=1}^\infty$ is convergent. The limit, to be denoted by $\mathbf{w}_0$, belongs to M because M is assumed to be closed. It follows from (4.7) that $\delta = ||\mathbf{v} - \mathbf{w}_0||$, so (4.6) is satisfied.

In order to show that $\mathbf{w}_0$ is unique, assume that

$$|||\mathbf{v} - \mathbf{u}|| = ||\mathbf{v} - \mathbf{w}_0|| = \delta$$

for some $\mathbf{u} \in W$. Then, via the calculation in (4.8) with $\mathbf{w}_k$ and $\mathbf{w}_\ell$ replaced by $\mathbf{w}_0$ and $\mathbf{u}$,

$$\begin{aligned} ||\mathbf{w}_0 - \mathbf{u}||^2 &= 2\,||\mathbf{w}_0 - \mathbf{v}||^2 + 2\,||\mathbf{v} - \mathbf{u}||^2 - 4\,||\frac{1}{2}\,(\mathbf{w}_0 + \mathbf{u}) - \mathbf{v}||^2 \\ &\leq 2\delta^2 + 2\delta^2 - 4\delta^2 = 0. \end{aligned}$$

Thus $\mathbf{u} = \mathbf{w}_0$, which completes the proof. □

Theorem 4.3.5 (Direct sum decomposition) *For any closed subspace W of a Hilbert space $\mathcal{H}$,*

$$\mathcal{H} = W \oplus W^\perp.$$

Proof. If $W = \{\mathbf{0}\}$, then $W^\perp = \mathcal{H}$ and the result holds. We will therefore for the rest of the proof assume that $W \neq \{\mathbf{0}\}$. Fix any $\mathbf{v} \in \mathcal{H}$, and choose $\mathbf{w}_0 \in W$ as in Lemma 4.3.4. Then

$$\mathbf{v} = \mathbf{w}_0 + (\mathbf{v} - \mathbf{w}_0).$$

We will first show that $\mathbf{v} - \mathbf{w}_0 \in W^\perp$. In order to do this, assume that $\mathbf{v} - \mathbf{w}_0 \notin W^\perp$. Then there exists $\mathbf{u} \in W$ such that $\langle \mathbf{v} - \mathbf{w}_0, \mathbf{u} \rangle \neq 0$; by a scalar multiplication we can obtain that

$$\langle \mathbf{v} - \mathbf{w}_0, \mathbf{u} \rangle = 1.$$

For any $\lambda \in \mathbb{R}$,

$$\begin{aligned} & ||\mathbf{v} - \mathbf{w}_0 - \lambda \mathbf{u}||^2 \\ = \; & \langle \mathbf{v} - \mathbf{w}_0 - \lambda \mathbf{u}, \mathbf{v} - \mathbf{w}_0 - \lambda \mathbf{u} \rangle \\ = \; & \langle \mathbf{v} - \mathbf{w}_0, \mathbf{v} - \mathbf{w}_0 \rangle - \lambda \langle \mathbf{v} - \mathbf{w}_0, \mathbf{u} \rangle - \lambda \langle \mathbf{u}, \mathbf{v} - \mathbf{w}_0 \rangle + \lambda^2 \langle \mathbf{u}, \mathbf{u} \rangle \\ = \; & ||\mathbf{v} - \mathbf{w}_0||^2 - 2\lambda + \lambda^2 ||\mathbf{u}||^2. \end{aligned}$$

Note that

$$-2\lambda + \lambda^2 ||\mathbf{u}||^2 = -\lambda(2 - \lambda \, ||\mathbf{u}||^2),$$

which is negative for sufficiently small values of $\lambda > 0$. For $\lambda > 0$ sufficiently small, this implies that

$$||\mathbf{v} - \mathbf{w}_0 - \lambda \mathbf{u}|| < ||\mathbf{v} - \mathbf{w}_0||;$$

in other words, the point $\mathbf{w}_0 + \lambda \mathbf{u}$ in W is closer to $\mathbf{v}$ than $\mathbf{w}_0$. This contradicts the choice of $\mathbf{w}_0$. Thus, we conclude that $\mathbf{v} - \mathbf{w}_0 \in W^\perp$.

We have now proved that $\mathcal{H} = W + W^\perp$. We leave it to the reader to show that the sum is direct (Exercise 4.7). □

The following example gives a geometrical understanding of Theorem 4.3.5.

Example 4.3.6 (Direct sum decomposition) Consider the vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \; \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

in $\mathbb{R}^3$. Letting

$$W = \text{span}\{\mathbf{e}_1, \mathbf{e}_2\},$$

it follows that $W^\perp$ is the set of scalar multiples of the vector

$$\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The interpretation of Theorem 4.3.5 is that every vector in $\mathbb{R}^3$ in a unique way can be decomposed as the sum of a vector in the (x, y)-plane and a vector in the direction of the z-axis. □

4.4 Functionals on Hilbert spaces

In Section 4.5 we will continue our analysis of linear operators, now acting on Hilbert spaces. Before we can do so, we need to consider the particular case of linear operators mapping the Hilbert space into the set of complex numbers $\mathbb{C}$. Such operators have a special name:

Definition 4.4.1 (Functional) *Let $\mathcal{H}$ be a Hilbert space. A linear operator $\Phi : \mathcal{H} \to \mathbb{C}$ is called a functional.*

We are particularly interested in functionals Φ that are bounded in the sense of Definition 2.4.1. This means that there exists a constant K such that

$$|\Phi \mathbf{v}| \leq K \,||\mathbf{v}||, \ \forall \, \mathbf{v} \in \mathcal{H}.$$

Considering any $\mathbf{w} \in \mathcal{H}$, we can define a bounded functional

$$\Phi : \mathcal{H} \to \mathbb{C}, \quad \Phi \mathbf{v} := \langle \mathbf{v}, \mathbf{w} \rangle, \tag{4.10}$$

see Exercise 4.11. Different vectors $\mathbf{w}$ provide us with different functionals:

Lemma 4.4.2 (Uniqueness of vector associated with functional) *Let $\mathcal{H}$ be a Hilbert space. Assume that $\mathbf{u}, \mathbf{w} \in \mathcal{H}$ satisfy*

$$\langle \mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle, \ \forall \, \mathbf{v} \in \mathcal{H}.$$

Then $\mathbf{u} = \mathbf{w}$.

We ask the reader to provide the proof (Exercise 4.12). The famous *Riesz' Representation Theorem* states that all bounded functionals on $\mathcal{H}$ have the form (4.10):

Theorem 4.4.3 (Riesz' Representation Theorem) *Given any bounded functional $\Phi : \mathcal{H} \to \mathbb{C}$, there exists a unique vector $\mathbf{w} \in \mathcal{H}$ such that (4.10) holds.*

Proof. For the sake of clarity we will frequently write $\Phi(\mathbf{v})$ rather than just $\Phi \mathbf{v}$ in this proof. We first show the existence part. In case $\Phi \mathbf{v} = 0$ for all $\mathbf{v} \in \mathcal{H}$ we can take $\mathbf{w} = \mathbf{0}$. Now assume that there exists $\mathbf{v} \in \mathcal{H}$ such that $\Phi \mathbf{v} \neq 0$, and put

$$W = \{\mathbf{v} \in \mathcal{H} \mid \Phi \mathbf{v} = 0\}. \tag{4.11}$$

Then W is a closed subspace of $\mathcal{H}$ (Exercise 4.14). Since W is a proper subspace of $\mathcal{H}$, Theorem 4.3.5 implies that $W^\perp \neq \{\mathbf{0}\}$; thus, we can choose a vector $\mathbf{u} \in W^\perp$ with $||\mathbf{u}|| = 1$. Now, for any $\mathbf{v} \in \mathcal{H}$, consider the vector

$$\mathbf{z} := [\Phi(\mathbf{v})] \ \mathbf{u} - [\Phi(\mathbf{u})] \ \mathbf{v} \in \mathcal{H}.$$

Applying the functional Φ to the vector $\mathbf{z}$ and using the linearity yields

$$\Phi \mathbf{z} = \Phi([\Phi(\mathbf{v})] \ \mathbf{u} - [\Phi(\mathbf{u})] \ \mathbf{v}) = \Phi(\mathbf{v}) \, \Phi(\mathbf{u}) - \Phi(\mathbf{u}) \, \Phi(\mathbf{v}) = 0.$$

This shows that $\mathbf{z} \in W$. Since $\mathbf{u} \in W^\perp$ by assumption,

$$\langle \mathbf{z}, \mathbf{u} \rangle = \langle [\Phi(\mathbf{v})] \ \mathbf{u} - [\Phi(\mathbf{u})] \ \mathbf{v}, \mathbf{u} \rangle = 0.$$

Using the rules for calculations with inner products,

$$0 = \Phi(\mathbf{v})\,||\mathbf{u}||^2 - \Phi(\mathbf{u})\,\langle \mathbf{v}, \mathbf{u}\rangle.$$

Since $||\mathbf{u}|| = 1$, this leads to

$$\Phi(\mathbf{v}) = \Phi(\mathbf{u})\,\langle \mathbf{v}, \mathbf{u}\rangle = \langle \mathbf{v}, \overline{\Phi(\mathbf{u})}\,\mathbf{u}\rangle.$$

Thus, the claimed result holds with $\mathbf{w} := \overline{\Phi(\mathbf{u})}\,\mathbf{u}$. The uniqueness follows from Lemma 4.4.2. □

Let us state a result concerning Hilbert spaces that will be used repeatedly during the book. It is related to the operator Φ in (4.10). The proof is left to the reader as Exercise 4.15.

Lemma 4.4.4 (Completeness of $\{\mathbf{v}_k\}_{k=1}^\infty$) *For a sequence $\{\mathbf{v}_k\}_{k=1}^\infty$ in a Hilbert space $\mathcal{H}$ the following are equivalent:*

(i) $\{\mathbf{v}_k\}_{k=1}^\infty$ *is complete.*

(ii) If $\mathbf{v} \in \mathcal{H}$ *and* $\langle \mathbf{v}, \mathbf{v}_k\rangle = 0$ *for all* $k \in \mathbb{N}$*, then* $\mathbf{v} = \mathbf{0}$*.*

4.5 Linear operators on Hilbert spaces

Continuing our analysis of linear operators in Sections 2.4 and 3.3, we will now have a closer look at linear operators on Hilbert spaces. In particular, we will introduce the *adjoint* of an operator and discuss some of its properties.

For the sake of complete generality, we will consider two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, and a bounded linear operator $T : \mathcal{K} \to \mathcal{H}$. Denote the inner products on $\mathcal{H}$ and $\mathcal{K}$ by $\langle \cdot, \cdot\rangle_{\mathcal{H}}$ and $\langle \cdot, \cdot\rangle_{\mathcal{K}}$, respectively. In order to define the adjoint operator associated with T, fix any $\mathbf{w} \in \mathcal{H}$; then the linear mapping

$$\Phi : \mathcal{K} \to \mathbb{C},\ \Phi\mathbf{v} := \langle T\mathbf{v}, \mathbf{w}\rangle_{\mathcal{H}} \tag{4.12}$$

defines a bounded functional on the Hilbert space $\mathcal{K}$ (Exercise 4.16). By Theorem 4.4.3, there exists a unique element in $\mathcal{K}$, to be called $T^*\mathbf{w}$, such that

$$\Phi\mathbf{v} = \langle \mathbf{v}, T^*\mathbf{w}\rangle_{\mathcal{K}},\ \forall\, \mathbf{v} \in \mathcal{K}.$$

That is,

$$\langle T\mathbf{v}, \mathbf{w}\rangle_{\mathcal{H}} = \langle \mathbf{v}, T^*\mathbf{w}\rangle_{\mathcal{K}},\ \forall\, \mathbf{v} \in \mathcal{K}. \tag{4.13}$$

For each $\mathbf{w} \in \mathcal{H}$ we have now associated an element $T^*\mathbf{w} \in \mathcal{K}$. Thus, we can consider T^* as a mapping, $T^* : \mathcal{H} \to \mathcal{K}$. The mapping T^* is called the *adjoint operator* of T. We will now prove that T^* actually is linear and bounded:

Theorem 4.5.1 (Adjoint operator) *The mapping* $T^* : \mathcal{H} \to \mathcal{K}$ *defined by* (4.13) *is linear and bounded.*

Proof. Given any $\mathbf{w}, \mathbf{u} \in \mathcal{H}$ and $\alpha, \beta \in \mathbb{C}$, we need to show that

$$T^*(\alpha \mathbf{w} + \beta \mathbf{u}) = \alpha\, T^* \mathbf{w} + \beta\, T^* \mathbf{u}. \tag{4.14}$$

In order to do this, consider any $\mathbf{v} \in \mathcal{K}$; then, by definition of T^* and the properties of the inner product,

$$\begin{aligned} \langle \mathbf{v}, T^*(\alpha \mathbf{w} + \beta \mathbf{u}) \rangle_{\mathcal{K}} &= \langle T\mathbf{v}, \alpha \mathbf{w} + \beta \mathbf{u} \rangle_{\mathcal{H}} \\ &= \overline{\alpha}\, \langle T\mathbf{v}, \mathbf{w} \rangle_{\mathcal{H}} + \overline{\beta}\, \langle T\mathbf{v}, \mathbf{u} \rangle_{\mathcal{H}} \\ &= \overline{\alpha}\, \langle \mathbf{v}, T^*\mathbf{w} \rangle_{\mathcal{K}} + \overline{\beta}\, \langle \mathbf{v}, T^*\mathbf{u} \rangle_{\mathcal{K}} \\ &= \langle \mathbf{v}, \alpha T^* \mathbf{w} + \beta T^* \mathbf{u} \rangle_{\mathcal{K}}. \end{aligned}$$

By Lemma 4.4.2 this implies that (4.14) holds.

We now prove that T^* is bounded. For that purpose, consider any $\mathbf{w} \in \mathcal{H}$. Then Theorem 4.1.4(v) shows that

$$\begin{aligned} ||T^*\mathbf{w}|| &= \sup\{\, |\langle \mathbf{v}, T^*\mathbf{w} \rangle_{\mathcal{K}}| \,\big|\, \mathbf{v} \in \mathcal{K}, ||\mathbf{v}|| = 1\} \\ &= \sup\{\, |\langle T\mathbf{v}, \mathbf{w} \rangle_{\mathcal{H}}| \,\big|\, \mathbf{v} \in \mathcal{K}, ||\mathbf{v}|| = 1\}. \end{aligned}$$

Via Cauchy–Schwarz' inequality combined with the definition of the norm of T, this implies that

$$\begin{aligned} ||T^*\mathbf{w}|| &\leq \sup_{||\mathbf{v}||=1} (||T\mathbf{v}||\, ||\mathbf{w}||) \\ &\leq \sup_{||\mathbf{v}||=1} (||T||\, ||\mathbf{v}||\, ||\mathbf{w}||) \\ &= ||T||\, ||\mathbf{w}||. \end{aligned} \tag{4.15}$$

This proves that T^* is bounded. □

Note that the estimate (4.15) implies that

$$||T^*|| \leq ||T||.$$

The following result shows that T and T^* actually have the same norm; we ask the reader to provide the proof in Exercise 4.18.

Lemma 4.5.2 (Properties of the adjoint operator) *Let* $T : \mathcal{K} \to \mathcal{H}$ *be a bounded linear operator. Then the following hold:*

(i) $(T^*)^* = T$;

(ii) $||T|| = ||T^*||$.

Let us calculate T^* explicitly for a concrete operator:

Example 4.5.3 (Calculation of the adjoint operator) In Exercise 3.14 we considered the right-shift and left-shift operators on $\ell^1(\mathbb{N})$. We will now consider these operators on the Hilbert space $\ell^2(\mathbb{N})$. That is, we define

$$T : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}),\ T(x_1, x_2, \dots, x_n, \dots) := (0, x_1, x_2, \dots, x_n, \dots),$$

and

$$S : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}),\ S(x_1, x_2, \dots, x_n, \dots) := (x_2, x_3, \dots, x_n, \dots).$$

It is obvious that the operators T and S are linear; we will now show that the operators are bounded, and find the adjoint T^*.

In order to show that T is bounded, let $\mathbf{x} = (x_1, x_2, \dots) \in \ell^2(\mathbb{N})$; then

$$\begin{aligned} ||T\mathbf{x}||_2 &= ||(0, x_1, x_2, \dots, x_n, \dots)||_2 \\ &= \left(\sum_{k=1}^{\infty} |x_k|^2 \right)^{1/2} \\ &= ||\mathbf{x}||_2. \end{aligned}$$

This shows that T is bounded (and, in fact, an isometry). A similar argument shows that S is bounded but not an isometry, see Exercise 4.22.

In order to find the adjoint operator T^*, let $\mathbf{x} = (x_1, x_2, \dots)$ and $\mathbf{y} = (y_1, y_2, \dots)$ be vectors in $\ell^2(\mathbb{N})$. Using the definition of T^*,

$$\begin{aligned} \langle \mathbf{x}, T^*\mathbf{y} \rangle = \langle T\mathbf{x}, \mathbf{y} \rangle &= \langle \begin{pmatrix} 0 \\ x_1 \\ x_2 \\ \cdot \\ \cdot \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \end{pmatrix} \rangle \\ &= \sum_{k=1}^{\infty} x_k \overline{y_{k+1}} \\ &= \langle \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \end{pmatrix}, \begin{pmatrix} y_2 \\ y_3 \\ \cdot \\ \cdot \end{pmatrix} \rangle \\ &= \langle \mathbf{x}, S\mathbf{y} \rangle. \end{aligned}$$

Via Lemma 4.4.2 we conclude that

$$T^*\mathbf{y} = T^*(y_1, y_2, \dots) = (y_2, y_3, \dots).$$

Note that this actually shows that

$$T^* = S.$$

□

We now define some central concepts related to the interplay between the linear operators T and T^*:

Definition 4.5.4 (Self-adjoint and unitary operators) *Let $\mathcal{H}$ be a Hilbert space and $T : \mathcal{H} \to \mathcal{H}$ a bounded linear operator.*

(i) The operator T is self-adjoint if $T = T^*$.

(ii) The operator T is unitary if $TT^* = T^*T = I$.

Note that if a bounded operator T is self-adjoint, then

$$\langle T\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, T\mathbf{w}\rangle, \ \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{H};$$

if T is unitary, then

$$\langle T\mathbf{v}, T\mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{w}\rangle, \ \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{H}. \tag{4.16}$$

In particular, if $\langle \mathbf{v}, \mathbf{w}\rangle = 0$ for some $\mathbf{v}, \mathbf{w} \in \mathcal{H}$ and T is a unitary operator, then also $\langle T\mathbf{v}, T\mathbf{w}\rangle = 0$; that is, a unitary operator preserves orthogonality between two vectors. We also see directly from the definition that a unitary operator T is invertible, with

$$T^{-1} = T^*. \tag{4.17}$$

A special role is played by operators projecting elements in a Hilbert space onto a subspace. One might think, e.g., at the operator that maps a vector (x, y, z) in $\mathbb{R}^3$ onto the vector $(x, y, 0)$ belonging to a two-dimensional subspace of $\mathbb{R}^3$. The general definition of an *orthogonal projection* is based on the direct sum decomposition in Theorem 4.3.5:

Definition 4.5.5 (Orthogonal projections) *Let V be a closed subspace of a Hilbert space $\mathcal{H}$. Write $\mathbf{v} \in \mathcal{H}$ as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, where $\mathbf{v}_1 \in V, \ \mathbf{v}_2 \in V^\perp$. The orthogonal projection P onto V is defined by*

$$P\,\mathbf{v} := \mathbf{v}_1.$$

We leave the proof of the following properties of orthogonal projections to the reader (Exercise 4.24).

Lemma 4.5.6 (Properties of orthogonal projections) *Let $V \neq \{\mathbf{0}\}$ be a closed subspace of a Hilbert space $\mathcal{H}$. Then the orthogonal projection P of $\mathcal{H}$ onto V has the following properties:*

(i) P is linear and bounded, with $||P|| = 1$;

(ii) P is self-adjoint;

(iii) $P^2 = P$.

4.6 Bessel sequences in Hilbert spaces

The purpose of this section is to clarify convergence issues for infinite series in Hilbert spaces.

Let $\mathcal{H}$ be a separable Hilbert space. Recall from Section 2.1 that when speaking about a *sequence* $\{\mathbf{v}_k\}_{k=1}^{\infty}$ in $\mathcal{H}$, we mean an *ordered* set, i.e.,

$$\{\mathbf{v}_k\}_{k=1}^{\infty} = \{\mathbf{v}_1, \mathbf{v}_2, \dots\}.$$

In the formulation of the following results, we will use the hypothesis that

$$T : \ell^2(\mathbb{N}) \to \mathcal{H},\ T\{c_k\}_{k=1}^{\infty} := \sum_{k=1}^{\infty} c_k \mathbf{v}_k$$

defines a bounded linear operator. When doing so, we consider it as part of the condition that the infinite series $\sum_{k=1}^{\infty} c_k \mathbf{v}_k$ actually converges for all $\{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N})$.

Lemma 4.6.1 (Adjoint of the operator T) *Let $\{\mathbf{v}_k\}_{k=1}^{\infty}$ be a sequence in $\mathcal{H}$, and suppose that*

$$T : \ell^2(\mathbb{N}) \to \mathcal{H},\ T\{c_k\}_{k=1}^{\infty} := \sum_{k=1}^{\infty} c_k \mathbf{v}_k \tag{4.18}$$

defines a bounded linear operator. Then the adjoint operator is given by

$$T^* : \mathcal{H} \to \ell^2(\mathbb{N}), \quad T^*\mathbf{v} = \{\langle \mathbf{v}, \mathbf{v}_k \rangle_{\mathcal{H}}\}_{k=1}^{\infty}. \tag{4.19}$$

Furthermore,

$$\sum_{k=1}^{\infty} |\langle \mathbf{v}, \mathbf{v}_k \rangle_{\mathcal{H}}|^2 \le ||T||^2\, ||\mathbf{v}||_{\mathcal{H}}^2,\ \forall\, \mathbf{v} \in \mathcal{H}. \tag{4.20}$$

Proof. In order to find the expression for T^*, consider any $\mathbf{v} \in \mathcal{H}$ and $\{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N})$. Using the result in Exercise 4.25,

$$\langle \mathbf{v}, T\{c_k\}_{k=1}^{\infty} \rangle_{\mathcal{H}} = \langle \mathbf{v}, \sum_{k=1}^{\infty} c_k \mathbf{v}_k \rangle_{\mathcal{H}} = \sum_{k=1}^{\infty} \overline{c_k}\, \langle \mathbf{v}, \mathbf{v}_k \rangle_{\mathcal{H}}. \tag{4.21}$$

We know that T^* is a bounded linear operator from $\mathcal{H}$ to $\ell^2(\mathbb{N})$. We can write $T^*\mathbf{v}$ in terms of its coordinates as

$$T^*\mathbf{v} = \{(T^*\mathbf{v})_k\}_{k=1}^{\infty},\ \mathbf{v} \in \mathcal{H};$$

because T^* is bounded, we have that

$$||T^*\mathbf{v}||_2 = \left(\sum_{k=1}^{\infty} |(T^*\mathbf{v})_k|^2 \right)^{1/2} \le ||T^*||\, ||\mathbf{v}||_{\mathcal{H}},\ \forall\ \mathbf{v} \in \mathcal{H}.$$

This implies that for each $k \in \mathbb{N}$,

$$|(T^*\mathbf{v})_k| \leq ||T^*|| \, ||\mathbf{v}||_{\mathcal{H}}, \ \forall \ \mathbf{v} \in \mathcal{H},$$

i.e., that the kth coordinate function $\mathbf{v} \mapsto (T^*\mathbf{v})_k$ is bounded from $\mathcal{H}$ to $\mathbb{C}$. By Riesz' representation theorem, $(T^*\mathbf{v})_k$ therefore has the form

$$(T^*\mathbf{v})_k = \langle \mathbf{v}, \mathbf{w}_k \rangle_{\mathcal{H}}, \ \mathbf{v} \in \mathcal{H}$$

for some fixed $\mathbf{w}_k \in \mathcal{H}$. That is, $T^*\mathbf{v}$ has the form

$$T^*\mathbf{v} = \{\langle \mathbf{v}, \mathbf{w}_k \rangle_{\mathcal{H}}\}_{k=1}^{\infty}$$

for some fixed sequence $\{\mathbf{w}_k\}_{k=1}^{\infty}$ in $\mathcal{H}$. By definition of T^*,

$$\begin{aligned} \langle \mathbf{v}, T\{c_k\}_{k=1}^{\infty} \rangle_{\mathcal{H}} &= \langle T^*\mathbf{v}, \{c_k\}_{k=1}^{\infty} \rangle_{\ell^2(\mathbb{N})} \\ &= \langle \{\langle \mathbf{v}, \mathbf{w}_k \rangle_{\mathcal{H}}\}_{k=1}^{\infty}, \{c_k\}_{k=1}^{\infty} \rangle_{\ell^2(\mathbb{N})} \\ &= \sum_{k=1}^{\infty} \overline{c_k} \, \langle \mathbf{v}, \mathbf{w}_k \rangle_{\mathcal{H}}, \ \forall \{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N}), \ \mathbf{v} \in \mathcal{H}. \end{aligned}$$

Together with (4.21), this shows that

$$\sum_{k=1}^{\infty} \overline{c_k} \, \langle \mathbf{v}, \mathbf{w}_k \rangle_{\mathcal{H}} = \sum_{k=1}^{\infty} \overline{c_k} \, \langle \mathbf{v}, \mathbf{v}_k \rangle_{\mathcal{H}}, \quad \forall \{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N}), \ \mathbf{v} \in \mathcal{H}. \tag{4.22}$$

It follows from here that $\mathbf{v}_k = \mathbf{w}_k$, see Exercise 4.27. Thus, we have shown that T^* has the form announced in (4.19).

By Lemma 4.5.2, the adjoint of a bounded operator T is itself bounded, and $||T|| = ||T^*||$. Under the assumption in Lemma 4.6.1, we therefore have that

$$||T^*\mathbf{v}||^2 \leq ||T||^2 \, ||\mathbf{v}||^2, \ \forall \, \mathbf{v} \in \mathcal{H},$$

which leads to (4.20). □

Sequences $\{\mathbf{v}_k\}_{k=1}^{\infty}$ for which an inequality of the type (4.20) holds will play a crucial role in Sections 4.7 and 4.8, so it is convenient to give them a name:

Definition 4.6.2 (Bessel sequences) *A sequence $\{\mathbf{v}_k\}_{k=1}^{\infty}$ in $\mathcal{H}$ is called a Bessel sequence if there exists a constant $B > 0$ such that*

$$\sum_{k=1}^{\infty} |\langle \mathbf{v}, \mathbf{v}_k \rangle|^2 \leq B \, ||\mathbf{v}||^2, \ \forall \, \mathbf{v} \in \mathcal{H}. \tag{4.23}$$

Any number B satisfying (4.23) is called a Bessel bound for $\{\mathbf{v}_k\}_{k=1}^{\infty}$.

We can characterize Bessel sequences in terms of the operator T appearing in (4.18):

Theorem 4.6.3 (Characterization of Bessel sequences) *Let $\{\mathbf{v}_k\}_{k=1}^\infty$ be a sequence in $\mathcal{H}$ and $B > 0$ be given. Then $\{\mathbf{v}_k\}_{k=1}^\infty$ is a Bessel sequence with Bessel bound B if and only if*

$$T : \ell^2(\mathbb{N}) \to \mathcal{H}, \quad T\{c_k\}_{k=1}^\infty = \sum_{k=1}^\infty c_k \mathbf{v}_k$$

defines a bounded linear operator and $||T|| \le \sqrt{B}$.

Proof. First assume that $\{\mathbf{v}_k\}_{k=1}^\infty$ is a Bessel sequence with Bessel bound B. Let $\{c_k\}_{k=1}^\infty \in \ell^2(\mathbb{N})$. First we want to show that $T\{c_k\}_{k=1}^\infty$ is well-defined, i.e., that $\sum_{k=1}^\infty c_k \mathbf{v}_k$ is convergent. Consider $n, m \in \mathbb{N}, n > m$. Then

$$\left|\left|\sum_{k=1}^n c_k \mathbf{v}_k - \sum_{k=1}^m c_k \mathbf{v}_k\right|\right| = \left|\left|\sum_{k=m+1}^n c_k \mathbf{v}_k\right|\right|.$$

Using Theorem 4.1.4(v) and Cauchy–Schwarz' inequality, it follows that

$$\begin{aligned}
\left|\left|\sum_{k=1}^n c_k \mathbf{v}_k - \sum_{k=1}^m c_k \mathbf{v}_k\right|\right| &= \sup_{||\mathbf{w}||=1} \left|\langle \sum_{k=m+1}^n c_k \mathbf{v}_k, \mathbf{w}\rangle\right| \\
&\le \sup_{||\mathbf{w}||=1} \sum_{k=m+1}^n |c_k \langle \mathbf{v}_k, \mathbf{w}\rangle| \\
&\le \left(\sum_{k=m+1}^n |c_k|^2\right)^{1/2} \sup_{||\mathbf{w}||=1} \left(\sum_{k=m+1}^n |\langle \mathbf{v}_k, \mathbf{w}\rangle|^2\right)^{1/2} \\
&\le \sqrt{B}\left(\sum_{k=m+1}^n |c_k|^2\right)^{1/2}.
\end{aligned}$$

Since $\{c_k\}_{k=1}^\infty \in \ell^2(\mathbb{N})$, we know that $\left\{\sum_{k=1}^n |c_k|^2\right\}_{n=1}^\infty$ is a Cauchy sequence in $\mathbb{C}$. The above calculation now shows that $\left\{\sum_{k=1}^n c_k \mathbf{v}_k\right\}_{n=1}^\infty$ is a Cauchy sequence in $\mathcal{H}$ and therefore convergent. Thus, $T\{c_k\}_{k=1}^\infty$ is well-defined. Clearly T is linear; since

$$||T\{c_k\}_{k=1}^\infty|| = \sup_{||\mathbf{w}||=1} |\langle T\{c_k\}_{k=1}^\infty, \mathbf{w}\rangle|,$$

a calculation as above shows that T is bounded and that $||T|| \le \sqrt{B}$.

For the opposite implication, suppose that T defines a bounded operator with $||T|| \le \sqrt{B}$. Then Lemma 4.6.1 shows that $\{\mathbf{v}_k\}_{k=1}^\infty$ is a Bessel sequence with Bessel bound B. □

Let us consider an important example of a Bessel sequence:

Lemma 4.6.4 (Bessel's inequality for orthonormal systems) *Let $\{\mathbf{v}_k\}_{k=1}^\infty$ be an orthonormal system in $\mathcal{H}$. Then the following hold:*

(i) The infinite series $\sum_{k=1}^\infty c_k\mathbf{v}_k$ is convergent for all $\{c_k\}_{k=1}^\infty \in \ell^2(\mathbb{N})$.

(ii) For any $\mathbf{v} \in \mathcal{H}$,

$$\sum_{k=1}^\infty |\langle \mathbf{v}, \mathbf{v}_k\rangle|^2 \leq \ ||\mathbf{v}||^2.$$

(iii) With the operator T defined as in (4.18),

$$TT^*\mathbf{v} = \sum_{k=1}^\infty \langle \mathbf{v}, \mathbf{v}_k\rangle \mathbf{v}_k, \ \ \mathbf{v} \in \mathcal{H}.$$

Proof. In order to prove (i), let $\{c_k\}_{k=1}^\infty \in \ell^2(\mathbb{N})$. For any $m, n \in \mathbb{N}$ with $n > m$,

$$\begin{aligned} \left|\left|\sum_{k=1}^n c_k\mathbf{v}_k - \sum_{k=1}^m c_k\mathbf{v}_k\right|\right|^2 &= \left|\left|\sum_{k=m+1}^n c_k\mathbf{v}_k\right|\right|^2 \\ &= \langle \sum_{k=m+1}^n c_k\mathbf{v}_k, \sum_{\ell=m+1}^n c_\ell\mathbf{v}_\ell\rangle \\ &= \sum_{k=m+1}^n \sum_{\ell=m+1}^n c_k\overline{c_\ell}\langle \mathbf{v}_k, \mathbf{v}_\ell\rangle. \end{aligned}$$

Using that

$$\langle \mathbf{v}_k, \mathbf{v}_\ell\rangle = \delta_{k,\ell}, \ \ k, \ell \in \mathbb{N},$$

it follows that

$$\left|\left|\sum_{k=1}^n c_k\mathbf{v}_k - \sum_{k=1}^m c_k\mathbf{v}_k\right|\right|^2 = \sum_{k=m+1}^n c_k\overline{c_k} = \sum_{k=m+1}^n |c_k|^2. \tag{4.24}$$

Since $\{c_k\}_{k=1}^\infty \in \ell^2(\mathbb{N})$, we know that

$$\sum_{k=m+1}^n |c_k|^2 \to 0 \text{ as } m \to \infty.$$

Together with (4.24) this implies that $\left\{\sum_{k=1}^n c_k\mathbf{v}_k\right\}_{n=1}^\infty$ is a Cauchy sequence in $\mathcal{H}$, and therefore convergent. In order to prove (ii), we can repeat the above calculations (Exercise 4.31) with $\sum_{k=m+1}^n c_k\mathbf{v}_k$ replaced by $\sum_{k=1}^\infty c_k\mathbf{v}_k$, and hereby show that

$$\left|\left|\sum_{k=1}^\infty c_k\mathbf{v}_k\right|\right|^2 = \sum_{k=1}^\infty |c_k|^2, \ \ \{c_k\}_{k=1}^\infty \in \ell^2(\mathbb{N}).$$

By Theorem 4.6.3 this implies that $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a Bessel sequence with bound $B = 1$. Finally, (iii) is a direct consequence of (4.18) and (4.19). □

For later use we note that it is enough to check the Bessel condition (4.23) on a dense subset of $\mathcal{H}$:

Lemma 4.6.5 (Bessel sequences) *Suppose that $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a sequence of elements in $\mathcal{H}$ and that there exists a constant $B > 0$ such that*

$$\sum_{k=1}^{\infty} |\langle \mathbf{v}, \mathbf{v}_k \rangle|^2 \leq B \, ||\mathbf{v}||^2$$

for all $\mathbf{v}$ in a dense subset V of $\mathcal{H}$. Then $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a Bessel sequence with bound B.

We leave the proof to the reader (Exercise 4.26).

4.7 Orthonormal bases

In Definition 2.5.4 we defined bases in infinite-dimensional normed spaces. In Hilbert spaces the most important bases are the *orthonormal bases*:

Definition 4.7.1 (Orthonormal basis) *A basis $\{\mathbf{e}_k\}_{k=1}^{\infty}$ in a Hilbert space $\mathcal{H}$ is an orthonormal basis for $\mathcal{H}$ if $\{\mathbf{e}_k\}_{k=1}^{\infty}$ is an orthonormal system.*

The next theorem gives equivalent conditions for an orthonormal system $\{\mathbf{e}_k\}_{k=1}^{\infty}$ to be an orthonormal basis.

Theorem 4.7.2 (Characterization of orthonormal bases) *For an orthonormal system $\{\mathbf{e}_k\}_{k=1}^{\infty}$, the following are equivalent:*

(i) $\{\mathbf{e}_k\}_{k=1}^{\infty}$ *is an orthonormal basis.*

(ii) $\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k, \ \forall \, \mathbf{v} \in \mathcal{H}$.

(iii) $\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \langle \mathbf{e}_k, \mathbf{w} \rangle, \ \forall \, \mathbf{v}, \mathbf{w} \in \mathcal{H}$.

(iv) $\sum_{k=1}^{\infty} |\langle \mathbf{v}, \mathbf{e}_k \rangle|^2 = ||\mathbf{v}||^2, \ \forall \, \mathbf{v} \in \mathcal{H}$.

(v) $\overline{\text{span}}\{\mathbf{e}_k\}_{k=1}^{\infty} = \mathcal{H}$.

(vi) *If* $\mathbf{v} \in \mathcal{H}$ *and* $\langle \mathbf{v}, \mathbf{e}_k \rangle = 0, \ \forall k \in \mathbb{N}$, *then* $\mathbf{v} = \mathbf{0}$.

Proof. For the proof of (i) ⇒ (ii), let $\mathbf{v} \in \mathcal{H}$. If $\{\mathbf{e}_k\}_{k=1}^{\infty}$ is an orthonormal basis, there exist coefficients $\{c_k\}_{k=1}^{\infty}$ such that $\mathbf{v} = \sum_{k=1}^{\infty} c_k \mathbf{e}_k$. Given any $j \in \mathbb{N}$, we have

$$\langle \mathbf{v}, \mathbf{e}_j \rangle = \langle \sum_{k=1}^{\infty} c_k \mathbf{e}_k, \mathbf{e}_j \rangle = \sum_{k=1}^{\infty} c_k \delta_{k,j} = c_j,$$

and (ii) follows (note that the argument involved moving an infinite sum out from the inner product: this is justified by (4.35) in Exercise 4.25).

For the proof of (ii) ⇒ (iii), consider the inner product $\langle \mathbf{v}, \mathbf{w} \rangle$ between two arbitrary vectors in $\mathcal{H}$, and expand $\mathbf{v}$ as in (ii):

$$\begin{aligned}\langle \mathbf{v}, \mathbf{w} \rangle &= \langle \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k, \mathbf{w} \rangle \\ &= \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \langle \mathbf{e}_k, \mathbf{w} \rangle.\end{aligned}$$

This proves that (ii) ⇒ (iii). Letting $\mathbf{v} = \mathbf{w}$, we see that (iv) is a special case of (iii), i.e., (iii) ⇒ (iv).

In order to prove the implication (iv) ⇒ (v), assume that (iv) holds. If $\mathbf{v} \in \mathcal{H}$ is orthogonal to all $\mathbf{e}_k, k \in \mathbb{N}$, (iv) implies that $\mathbf{v} = \mathbf{0}$. By Lemma 4.4.4 we can now conclude that $\overline{\text{span}}\{\mathbf{e}_k\}_{k=1}^{\infty} = \mathcal{H}$, i.e., that (v) holds. The implication (v) ⇒ (vi) also follows from Lemma 4.4.4.

For the proof of (vi)⇒ (i), let $\mathbf{v} \in \mathcal{H}$. We know from Lemma 4.6.4 that

$$\mathbf{w} := \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k$$

is well defined. For any $j \in \mathbb{N}$,

$$\begin{aligned}\langle \mathbf{v} - \mathbf{w}, \mathbf{e}_j \rangle = \langle \mathbf{v}, \mathbf{e}_j \rangle - \langle \mathbf{w}, \mathbf{e}_j \rangle &= \langle \mathbf{v}, \mathbf{e}_j \rangle - \langle \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k, \mathbf{e}_j \rangle \\ &= \langle \mathbf{v}, \mathbf{e}_j \rangle - \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \langle \mathbf{e}_k, \mathbf{e}_j \rangle \\ &= 0.\end{aligned}$$

By (vi), this proves that

$$\mathbf{v} = \mathbf{w} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k.$$

To prove that $\{\mathbf{e}_k\}_{k=1}^{\infty}$ is a basis, we only need to show that no other linear combination of $\{\mathbf{e}_k\}_{k=1}^{\infty}$ can be equal to $\mathbf{v}$, and this follows by the argument we used to prove that (ii) follows from (i). □

The equality in Theorem 4.7.2(iv) is called *Parseval's equation.* The property in Theorem 4.7.2(ii) is the main reason for considering orthonormal bases: usually one refers to this property by saying that all elements $f \in \mathcal{H}$ can be *expanded* in terms of the vectors $\mathbf{e}_k$ in an orthonormal basis.

Later, we will encounter orthonormal bases that are not indexed by $\mathbb{N}$. For this reason it is important to realize that the expansion property is independent of the indexing of the vectors. In order to make this statement precise we need the next definition:

Definition 4.7.3 (Reordering) *Let* $\{\mathbf{v}_k\}_{k=1}^{\infty}$ *be a sequence in a vector space* V. *Given a bijective mapping* $\sigma : \mathbb{N} \to \mathbb{N}$, *the vectors*

$$\{\mathbf{v}_{\sigma(k)}\}_{k=1}^{\infty} = \{\mathbf{v}_{\sigma(1)}, \mathbf{v}_{\sigma(2)}, \dots, \mathbf{v}_{\sigma(n)}, \dots\}$$

are called a reordering of $\{\mathbf{v}_k\}_{k=1}^{\infty}$.

The vectors $\{\mathbf{v}_{\sigma(k)}\}_{k=1}^{\infty}$ are also called a *permutation* of $\{\mathbf{v}_k\}_{k=1}^{\infty}$: the two sequences consist of exactly the same vectors, but in a different order.

Example 4.7.4 (Reordering) The mapping $\sigma : \mathbb{N} \to \mathbb{N}$ given by

$$\sigma(k) = \begin{cases} k, & \text{if } k = 1, 5, 9, \dots, \\ k+1, & \text{if } k = 2, 6, 10, \dots, \\ k-1, & \text{if } k = 3, 7, 11, \dots, \\ k, & \text{if } k = 4, 8, 12, \dots, \end{cases}$$

is a bijection. By inspection,

$$\{\mathbf{v}_{\sigma(k)}\}_{k=1}^{\infty} = \{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_2, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_7, \mathbf{v}_6, \mathbf{v}_8, \dots\}.$$

□

Note that the property of being an orthonormal system is *independent* of how the vectors are ordered. The condition (vi) in Theorem 4.7.2 is also independent on the order! This shows that if $\{\mathbf{e}_k\}_{k=1}^{\infty}$ is an orthonormal basis, then $\{\mathbf{e}_{\sigma(k)}\}_{k=1}^{\infty}$ is an orthonormal basis for any permutation σ. In particular, the expansion

$$\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_{\sigma(k)} \rangle \mathbf{e}_{\sigma(k)} \tag{4.25}$$

holds for all $\mathbf{v} \in \mathcal{H}$. That is, the expansion property is independent on the order of the vectors $\mathbf{e}_k$. We say that the series (4.25) converges *unconditionally,* and state the result formally:

Corollary 4.7.5 (Expansion via orthonormal basis) *If* $\{\mathbf{e}_k\}_{k=1}^{\infty}$ *is an orthonormal basis, then each* $\mathbf{v} \in \mathcal{H}$ *has an expansion*

$$\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k. \tag{4.26}$$

The series in (4.26) *converges unconditionally.*

The unconditional convergence of the series in (4.26) implies that we can order the vectors $\mathbf{e}_k$ as we want. We can also choose to index the vectors by another index set than $\mathbb{N}$. This will be important in our analysis of wavelets in Chapter 8, where the index set is

$$\mathbb{Z} \times \mathbb{Z} = \{(j,k) \mid j, k \in \mathbb{Z}\}.$$

Orthonormal bases exist in all separable Hilbert spaces:

Theorem 4.7.6 (Existence of orthonormal bases) *Every separable Hilbert space $\mathcal{H}$ has an orthonormal basis.*

Proof. Since $\mathcal{H}$ is assumed separable, we can choose a sequence $\{\mathbf{v}_k\}_{k=1}^{\infty}$ in $\mathcal{H}$ such that $\overline{\text{span}}\{\mathbf{v}_k\}_{k=1}^{\infty} = \mathcal{H}$. By passing down to a subsequence if necessary, we can assume that for each $n \in \mathbb{N}, \mathbf{v}_{n+1} \notin \text{span}\{\mathbf{v}_k\}_{k=1}^{n}$. By applying the Gram–Schmidt process (see Exercise 4.9) to $\{\mathbf{v}_k\}_{k=1}^{\infty}$, we obtain an orthonormal system $\{\mathbf{e}_k\}_{k=1}^{\infty}$ in $\mathcal{H}$ for which

$$\overline{\text{span}}\{\mathbf{e}_k\}_{k=1}^{\infty} = \overline{\text{span}}\{\mathbf{v}_k\}_{k=1}^{\infty} = \mathcal{H}.$$

The result now follows from Theorem 4.7.2. □

In the Hilbert space $\ell^2(\mathbb{N})$, we have an orthonormal basis given by a particularly simple expression:

Example 4.7.7 (Orthonormal basis for $\ell^2(\mathbb{N})$) For $k \in \mathbb{N}$, let δ_k be the sequence in $\ell^2(\mathbb{N})$ whose kth entry is 1, and all other entries are 0. Then $\{\delta_k\}_{k=1}^{\infty}$ is an orthonormal basis for $\ell^2(\mathbb{N})$, see Exercise 4.28; it is called the *canonical orthonormal basis.* □

Based on Theorem 4.7.6, we will now prove that every separable Hilbert space $\mathcal{H}$ can be *identified* with $\ell^2(\mathbb{N})$. The exact meaning of this statement is that there exists a bijection $U : \mathcal{H} \to \ell^2(\mathbb{N})$ for which $||U\mathbf{v}|| = ||\mathbf{v}||$ for all $\mathbf{v} \in \mathcal{H}$. We say that the spaces $\mathcal{H}$ and $\ell^2(\mathbb{N})$ are *isometrically ismorphic.*

Theorem 4.7.8 (Identification of Hilbert spaces) *Every separable infinite-dimensional Hilbert space $\mathcal{H}$ is isometrically isomorphic to $\ell^2(\mathbb{N})$.*

Proof. Let $\{\mathbf{e}_k\}_{k=1}^{\infty}$ be an orthonormal basis for $\mathcal{H}$. In Lemma 4.6.4 we have observed that $\sum_{k=1}^{\infty} c_k \mathbf{e}_k$ is convergent for all $\{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N})$. Furthermore, Lemma 4.6.4 together with Theorem 4.7.2 imply that each $\mathbf{v} \in \mathcal{H}$ has a unique expansion with ℓ^2-coefficients, namely, $\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k$. Letting $\{\delta_k\}_{k=1}^{\infty}$ be the canonical orthonormal basis for $\ell^2(\mathbb{N})$, we can define the operator

$$U : \mathcal{H} \to \ell^2(\mathbb{N}),\ U\left(\sum_{k=1}^{\infty} c_k \mathbf{e}_k\right) := \sum_{k=1}^{\infty} c_k \delta_k,\ \forall \{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N}).$$

Then U maps $\mathcal{H}$ bijectively onto $\ell^2(\mathbb{N})$. For $\mathbf{v} \in \mathcal{H}, \mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k$, we have

$$
\begin{aligned}
||U\mathbf{v}||^2 &= ||\sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \delta_k||^2 \\
&= \sum_{k=1}^{\infty} |\langle \mathbf{v}, \mathbf{e}_k \rangle|^2 \\
&= ||\mathbf{v}||^2;
\end{aligned}
$$

thus, U is an isometry. □

The following theorem characterizes all orthonormal bases for $\mathcal{H}$ starting with one arbitrary orthonormal basis.

Theorem 4.7.9 (Characterization of orthonormal bases) *Let $\{\mathbf{e}_k\}_{k=1}^{\infty}$ be an orthonormal basis for $\mathcal{H}$. Then the orthonormal bases for $\mathcal{H}$ are precisely the sets $\{U\mathbf{e}_k\}_{k=1}^{\infty}$, where $U : \mathcal{H} \to \mathcal{H}$ is a unitary operator.*

Proof. Let $\{\mathbf{v}_k\}_{k=1}^{\infty}$ be an orthonormal basis for $\mathcal{H}$. Define the operator

$$U : \mathcal{H} \to \mathcal{H},\ U\left(\sum_{k=1}^{\infty} c_k \mathbf{e}_k\right) := \sum_{k=1}^{\infty} c_k \mathbf{v}_k,\ \ \forall \{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N}).$$

Then U maps $\mathcal{H}$ boundedly and bijectively onto $\mathcal{H}$, and $\mathbf{v}_k = U\mathbf{e}_k$. For $\mathbf{v}, \mathbf{w} \in \mathcal{H}$, write $\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k$ and $\mathbf{w} = \sum_{k=1}^{\infty} \langle \mathbf{w}, \mathbf{e}_k \rangle \mathbf{e}_k$; then, via the definition of U and Theorem 4.7.2,

$$
\begin{aligned}
\langle U^*U\mathbf{v}, \mathbf{w} \rangle &= \langle U\mathbf{v}, U\mathbf{w} \rangle \\
&= \left\langle \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{v}_k, \sum_{k=1}^{\infty} \langle \mathbf{w}, \mathbf{e}_k \rangle \mathbf{v}_k \right\rangle \\
&= \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \overline{\langle \mathbf{w}, \mathbf{e}_k \rangle} = \langle \mathbf{v}, \mathbf{w} \rangle.
\end{aligned}
$$

This implies that $U^*U = I$. Since U is bijective, we know that U^{-1} exists; and

$$U^* = U^*I = U^*UU^{-1} = IU^{-1} = U^{-1}.$$

This shows that U is unitary.

On the other hand, if U is a given unitary operator, then

$$\langle U\mathbf{e}_k, U\mathbf{e}_j \rangle = \langle U^*U\mathbf{e}_k, \mathbf{e}_j \rangle = \langle \mathbf{e}_k, \mathbf{e}_j \rangle = \delta_{k,j},$$

i.e., $\{U\mathbf{e}_k\}_{k=1}^{\infty}$ is an orthonormal system. To show that it is a basis, assume that $\langle \mathbf{v}, U\mathbf{e}_k \rangle = 0$ for all $k \in \mathbb{N}$. Then $\langle U^*\mathbf{v}, \mathbf{e}_k \rangle = 0$ for all $k \in \mathbb{N}$, so $U^*\mathbf{v} = \mathbf{0}$. But the operator $U^* = U^{-1}$ is invertible, so we conclude that $\mathbf{v} = \mathbf{0}$. Now the result follows from Theorem 4.7.2. □

In practice, orthonormal bases are certainly the most convenient bases to use: for other types of bases, the representation (4.26) has to be replaced by a more complicated expression. In fact, given any basis $\{\mathbf{v}_k\}_{k=1}^{\infty}$ for a Hilbert space $\mathcal{H}$ it is known that there exists another basis $\{\mathbf{w}_k\}_{k=1}^{\infty}$ such that

$$\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{w}_k \rangle \mathbf{v}_k, \ \forall \, \mathbf{v} \in \mathcal{H}.$$

See Exercise 4.30 for a discussion of a class of bases $\{\mathbf{v}_k\}_{k=1}^{\infty}$ and associated bases $\{\mathbf{w}_k\}_{k=1}^{\infty}$. Unfortunately, it might be complicated to calculate $\{\mathbf{w}_k\}_{k=1}^{\infty}$; see, e.g., [19] or [5] for more information.

4.8 Frames in Hilbert spaces

We will now give a short introduction to an active research area dealing with a generalization of the concept of an orthonormal basis. Formally, the concept of a *frame* is defined by considering Bessel sequences having an additional lower bound:

Definition 4.8.1 (Frames) *A sequence of elements $\{\mathbf{v}_k\}_{k=1}^{\infty}$ in a Hilbert space $\mathcal{H}$ is a frame if there exist two positive numbers A and B such that*

$$A \, ||\mathbf{v}||^2 \leq \sum_{k=1}^{\infty} |\langle \mathbf{v}, \mathbf{v}_k \rangle|^2 \leq B \, ||\mathbf{v}||^2, \ \forall \, \mathbf{v} \in \mathcal{H}.$$

The frame is said to be tight if we can choose $A = B$.

Note that by Theorem 4.7.2, any orthonormal basis $\{\mathbf{e}_k\}_{k=1}^{\infty}$ is a frame with $A = B = 1$; this shows that frames generalize the concept of an orthonormal basis. In more advanced literature, e.g., [19], [10], [5], it is proved that if $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a frame, there exists another frame $\{\mathbf{w}_k\}_{k=1}^{\infty}$ such that

$$\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{w}_k \rangle \mathbf{v}_k, \ \forall \, \mathbf{v} \in \mathcal{H}. \tag{4.27}$$

At a first glance, this expansion is very similar with the one we saw for orthonormal bases in Theorem 4.7.2(iii). The crucial difference is that the coefficients appearing in (4.27) are given as inner products between the element $\mathbf{v}$ and a *new* sequence $\{\mathbf{w}_k\}_{k=1}^{\infty}$. Unfortunately, in practice it might be difficult to find an appropriate sequence $\{\mathbf{w}_k\}_{k=1}^{\infty}$. This is the reason for considering tight frames: if $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a tight frame, one can prove that the complicated expansion (4.27) can be replaced by

$$\mathbf{v} = \frac{1}{A}\sum_{k=1}^{\infty}\langle \mathbf{v}, \mathbf{v}_k\rangle \mathbf{v}_k, \ \forall\, \mathbf{v} \in \mathcal{H}. \tag{4.28}$$

Except for the factor A^{-1}, this expansion looks exactly like the one we obtained for orthonormal bases! That is, tight frames can be used without any extra computational effort compared with the use of orthonormal bases.

One advantage of frames is that they might be *overcomplete:* one can prove that unless $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a basis, various choices of a family $\{\mathbf{w}_k\}_{k=1}^{\infty}$ satisfying (4.27) are possible. This implies that one has the option to choose the one that fits a certain application best! Also, since frames are more general than orthonormal bases, there are cases where one can construct a frame with features that cannot be combined with the orthonormal basis condition. It will go too far to discuss such cases, so we refer to the book [5] for more information.

4.9 Exercises

4.1 Let V be a vector space with an inner product $\langle\cdot,\cdot\rangle$. Show that

$$\langle \mathbf{0}, \mathbf{v}\rangle = 0, \ \forall\ \mathbf{v} \in V.$$

4.2 Prove Theorem 4.1.4(ii). See Exercises 6.3 and 4.5 for some important consequences.

4.3 Prove Theorem 4.1.4(iii)+(iv).

4.4 Prove Theorem 4.1.4(v).

4.5 Show that the norm $||\cdot||_\infty$ on $C[0,2]$ does not come from an inner product. *Hint:* show that the norm $||\cdot||_\infty$ does not satisfy the parallelogram law, e.g., for

$$f(x) = \begin{cases} x & \text{if } x\in[0,1], \\ 2-x & \text{if } x\in[1,2], \end{cases} \quad g(x) = \begin{cases} 1-x & \text{if } x\in[0,1], \\ x-1 & \text{if } x\in[1,2]. \end{cases}$$

The function f is a so-called B-spline, see Chapter 10.

4.6 Consider the vector space $\ell^p(\mathbb{N})$ for some $p \in [1, \infty[$, equipped with the usual norm

$$||\{x_k\}_{k=1}^{\infty}||_p = \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p}.$$

(i) Consider the vectors

$$\mathbf{x} = (1, 0, 0, \dots), \quad \mathbf{y} = (0, 1, 0, 0, \dots),$$

and show that

$$||\mathbf{x}||_p = ||\mathbf{y}||_p = 1, \quad ||\mathbf{x} + \mathbf{y}||_p = ||\mathbf{x} - \mathbf{y}||_p = 2^{1/p}.$$

(ii) Assume that $p \neq 2$. Show that the norm $|| \cdot ||_p$ does not come from an inner product.

4.7 Assume that U and W are two subspaces of a Hilbert space $\mathcal{H}$, and that $\mathbf{u} \perp \mathbf{w}$ for all $\mathbf{u} \in U, \mathbf{w} \in W$. Show that then the sum $U + W$ is direct.

4.8 Equip $\mathbb{R}^3$ with the usual inner product, and consider the vectors

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix}, \ \mathbf{v}_2 = \begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}.$$

(i) Compute the vectors $\mathbf{e}_1 := \frac{\mathbf{v}_1}{||\mathbf{v}_1||}$, and

$$\mathbf{e}_2 := \frac{\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{e}_1 \rangle \mathbf{e}_1}{||\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{e}_1 \rangle \mathbf{e}_1||}.$$

(ii) Plot the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, as well as $\mathbf{e}_1$ and $\mathbf{e}_2$.

(iii) Compute

$$\mathbf{e}_3 := \frac{\mathbf{v}_3 - \sum_{k=1}^{2} \langle \mathbf{v}_3, \mathbf{e}_k \rangle \mathbf{e}_k}{||\mathbf{v}_3 - \sum_{k=1}^{2} \langle \mathbf{v}_3, \mathbf{e}_k \rangle \mathbf{e}_k||}.$$

(iv) Plot the vectors $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$, as well as $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{e}_3$.

The procedure described here is called *Gram–Schmidt orthonormalization,* and is treated in Exercise 4.9.

4.9 Let $\{\mathbf{v}_k\}_{k=1}^{\infty}$ be a set of linearly independent vectors in a Hilbert space $\mathcal{H}$. We will show how one can construct an orthonormal set $\{\mathbf{e}_k\}_{k=1}^{\infty}$ for which

$$\overline{\text{span}}\{\mathbf{v}_k\}_{k=1}^{\infty} = \overline{\text{span}}\{\mathbf{e}_k\}_{k=1}^{\infty}. \tag{4.29}$$

Let $\mathbf{e}_1 := \frac{\mathbf{v}_1}{||\mathbf{v}_1||}$, and

$$\mathbf{e}_2 := \frac{\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{e}_1 \rangle \mathbf{e}_1}{||\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{e}_1 \rangle \mathbf{e}_1||}. \tag{4.30}$$

(i) Show that $\mathbf{e}_1$ and $\mathbf{e}_2$ are orthonormal, and that

$$\text{span}\{\mathbf{e}_k\}_{k=1}^{2} = \text{span}\{\mathbf{v}_k\}_{k=1}^{2}. \tag{4.31}$$

We will now proceed with an inductive construction. Assume that we for some $n \in \mathbb{N}$ have constructed an orthonormal system $\{\mathbf{e}_k\}_{k=1}^{n}$ such that $\text{span}\{\mathbf{e}_k\}_{k=1}^{n} = \text{span}\{\mathbf{v}_k\}_{k=1}^{n}$. Then we want to put

$$\mathbf{e}_{n+1} := \frac{\mathbf{v}_{n+1} - \sum_{k=1}^{n} \langle \mathbf{v}_{n+1}, \mathbf{e}_k \rangle \mathbf{e}_k}{||\mathbf{v}_{n+1} - \sum_{k=1}^{n} \langle \mathbf{v}_{n+1}, \mathbf{e}_k \rangle \mathbf{e}_k||}. \tag{4.32}$$

(ii) Show that $\mathbf{e}_{n+1}$ is well defined, i.e., that

$$\mathbf{v}_{n+1} - \sum_{k=1}^{n} \langle \mathbf{v}_{n+1}, \mathbf{e}_k \rangle \mathbf{e}_k \neq \mathbf{0}.$$

(iii) Show that $\{\mathbf{e}_k\}_{k=1}^{\infty}$ is an orthonormal system and that for all $n \in \mathbb{N}$,

$$\text{span}\{\mathbf{e}_k\}_{k=1}^{n} = \text{span}\{\mathbf{v}_k\}_{k=1}^{n}. \tag{4.33}$$

(iv) Conclude that (4.29) holds.

The above procedure is called *Gram–Schmidt orthonormalization.*

4.10 Let $\mathcal{H}$ be a Hilbert space and $\{\mathbf{v}_k\}_{k=1}^{\infty}$ a sequence of vectors in $\mathcal{H}$ with $||\mathbf{v}_k|| = 1$. Consider the mapping

$$\Phi : \mathcal{H} \to \mathbb{C}, \ \Phi \mathbf{v} := \sum_{k=1}^{\infty} \frac{1}{k^2} \langle \mathbf{v}, \mathbf{v}_k \rangle.$$

(i) Show that Φ is well defined, i.e., that $\sum_{k=1}^{\infty} \frac{1}{k^2} \langle \mathbf{v}, \mathbf{v}_k \rangle$ is convergent if $\mathbf{v} \in \mathcal{H}$.

(ii) Show that Φ is linear and bounded.

(iii) Determine the vector $\mathbf{w} \in \mathcal{H}$ such that

$$\Phi \mathbf{v} = \langle \mathbf{v}, \mathbf{w} \rangle, \ \ \mathbf{v} \in \mathcal{H}.$$

4.11 Show that Φ in (4.10) defines a bounded functional on $\mathcal{H}$.

4.12 Prove Lemma 4.4.2.

4.13 Define analytically an example of a convex subset as well as a nonconvex subset of $\mathbb{R}^2$. Make a draft of these subsets.

4.14 Under the assumptions in Theorem 4.4.3, prove that the set W in (4.11) is a closed subspace of $\mathcal{H}$.

4.15 Let $\{\mathbf{v}_k\}_{k=1}^{\infty}$ denote a sequence in a Hilbert space $\mathcal{H}$. We want to prove Lemma 4.4.4, which claims that (a) and (b) below are equivalent:

(a) $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is complete;

(b) If $\mathbf{v} \in \mathcal{H}$ and $\langle \mathbf{v}, \mathbf{v}_k \rangle = 0$ for all $k \in \mathbb{N}$, then $\mathbf{v} = \mathbf{0}$.

Proceed as follows:

(i) Show that if $\mathbf{v} \in \mathcal{H}$ and $\langle \mathbf{v}, \mathbf{v}_k \rangle = 0$ for all $k \in \mathbb{N}$, then $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{w} \in \overline{\text{span}}\{\mathbf{v}_k\}_{k=1}^{\infty}$.

(ii) Use (i) to show that if (a) holds, then (b) holds.

(iii) Prove that if (a) does not hold, then (b) does not hold. *Hint:* let $W := \overline{\text{span}}\{\mathbf{v}_k\}_{k=1}^{\infty}$, and use Theorem 4.3.5 to argue that if (a) does not hold, then $W^{\perp} \neq \{\mathbf{0}\}$.

4.16 Show that the mapping Φ in (4.12) defines a bounded functional on the Hilbert space $\mathcal{K}$.

4.17 Let S and T be bounded linear operators on a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$.

(i) Show that for all $\mathbf{v}, \mathbf{w} \in \mathcal{H}$,

$$\langle ST\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, T^*S^*\mathbf{w} \rangle.$$

(ii) Based on the result in (i), show that

$$(ST)^* = T^*S^*.$$

4.18 Prove Lemma 4.5.2. *Hint:* for (i), use formula (4.13) to show that

$$\langle T\mathbf{v}, \mathbf{w}\rangle = \langle (T^*)^*\mathbf{v}, \mathbf{w}\rangle, \ \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{H},$$

and use Lemma 4.4.2. For (ii), the inequality $||T^*|| \leq ||T||$ is a consequence of the proof of Theorem 4.5.1.

4.19 Let $\mathbf{A}$ be an $n \times n$ matrix,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & \cdot & a_{nn} \end{pmatrix}.$$

We assume that the entries a_{ij} are complex numbers, and consider $\mathbf{A}$ as a linear mapping from $\mathbb{C}^n$ into $\mathbb{C}^n$.

(i) Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be vectors in $\mathbb{C}^n$, and show that

$$\langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle = \sum_{k=1}^{n}\sum_{j=1}^{n} a_{kj} x_j \overline{y_k}.$$

(*Hint:* the kth entry in the vector $\mathbf{A}\mathbf{x}$ is $\sum_{j=1}^{n} a_{kj} x_j$.)

Let $\overline{\mathbf{A}}^{\,T}$ be the matrix obtained by transposing and complex conjugating of the matrix $\mathbf{A}$, i.e.,

$$\overline{\mathbf{A}}^{\,T} = \begin{pmatrix} \overline{a_{11}} & \overline{a_{21}} & \cdot & \cdot & \overline{a_{n1}} \\ \overline{a_{12}} & \overline{a_{22}} & \cdot & \cdot & \overline{a_{n2}} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \overline{a_{1n}} & \overline{a_{2n}} & \cdot & \cdot & \overline{a_{nn}} \end{pmatrix}.$$

(ii) Show that

$$\langle \mathbf{x}, \overline{\mathbf{A}}^{\,T}\mathbf{y}\rangle = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{kj} x_j \overline{y_k}.$$

(*Hint:* the jth entry in the vector $\overline{\mathbf{A}}^{\,T}\mathbf{y}$ is $\sum_{k=1}^{n} \overline{a_{kj}} y_k$.)

(iii) Conclude that $\mathbf{A}^* = \overline{\mathbf{A}}^{\,\mathbf{T}}$.

4.20 Consider the mapping

$$T : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}),\ T\{x_k\}_{k=1}^{\infty} := \{x_k + x_{k+1}\}_{k=1}^{\infty}.$$

In Exercise 3.15 it is proved that T is linear and bounded.

(i) Find the adjoint operator T^*.

(ii) Is T self–adjoint?

(iii) Is T unitary?

4.21 Consider the mapping $T : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ given by

$$T(x_1, x_2, x_3, x_4, x_5 \dots) := (3x_2, 2x_1, x_3, x_4, x_5, \dots).$$

(i) Show that T is linear and bounded.

(ii) Find the adjoint operator T^*.

4.22 Consider the left-shift operator on $\ell^2(\mathbb{N})$,

$$S : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}),\ S(x_1, x_2, \dots) := (x_2, x_3, \dots).$$

(i) Show that S is a bounded linear operator.

(ii) Show that S is not an isometry.

(iii) Find the adjoint operator S^*.

4.23 Let U be a unitary operator on a Hilbert space $\mathcal{H}$.

(i) Show that $\langle U\mathbf{u}, U\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle,\ \forall\, \mathbf{u}, \mathbf{v} \in \mathcal{H}$.

(ii) Show that U^j is unitary for all $j \in \mathbb{Z}$.

4.24 Prove Lemma 4.5.6.

4.25 Let $\{\mathbf{v}_k\}_{k=1}^{\infty}$ be a sequence in a Hilbert space $\mathcal{H}$, and let $\{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N})$. Assume that $\sum_{k=1}^{\infty} c_k \mathbf{v}_k$ is convergent. Prove that for all $\mathbf{v} \in \mathcal{H}$,

$$\langle \mathbf{v}, \sum_{k=1}^{\infty} c_k \mathbf{v}_k \rangle_{\mathcal{H}} = \sum_{k=1}^{\infty} \overline{c_k}\ \langle \mathbf{v}, \mathbf{v}_k \rangle_{\mathcal{H}} \tag{4.34}$$

and

$$\langle \sum_{k=1}^{\infty} c_k \mathbf{v}_k, \mathbf{v} \rangle_{\mathcal{H}} = \sum_{k=1}^{\infty} c_k\ \langle \mathbf{v}_k, \mathbf{v} \rangle_{\mathcal{H}}. \tag{4.35}$$

4.26 Prove Lemma 4.6.5.

4.27 Show that the condition in (4.22) implies that $\mathbf{v}_k = \mathbf{w}_k$ for all $k \in \mathbb{N}$.

4.28 Prove the result stated in Example 4.7.7.

4.29 Let V be a closed subspace of a Hilbert space $\mathcal{H}$ and $\{\mathbf{e}_k\}_{k=1}^{\infty}$ an orthonormal basis for V. Show that the orthogonal projection P of $\mathcal{H}$ onto V is given by

$$P\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{e}_k, \ \ \mathbf{v} \in \mathcal{H}.$$

4.30 Let $\mathcal{H}$ denote a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Let U be a bijective, bounded, and linear operator on $\mathcal{H}$, let $\{\mathbf{e}_k\}_{k=1}^{\infty}$ be an orthonormal basis for $\mathcal{H}$, and define the vectors $\{\mathbf{v}_k\}_{k=1}^{\infty}$ by

$$\mathbf{v}_k := U\mathbf{e}_k, \ \ k = 1, 2, \dots .$$

(i) Let $\mathbf{v} \in \mathcal{H}$. Show that

$$U\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, \mathbf{e}_k \rangle \mathbf{v}_k.$$

(ii) Let $\mathbf{v} \in \mathcal{H}$. Show that

$$\mathbf{v} = \sum_{k=1}^{\infty} \langle \mathbf{v}, (U^{-1})^* \mathbf{e}_k \rangle \mathbf{v}_k.$$

(*Hint:* use that $\mathbf{v} = UU^{-1}\mathbf{v}$.)

(iii) Show that $\{\mathbf{v}_k\}_{k=1}^{\infty}$ is a basis for $\mathcal{H}$.

A basis of the type considered here is called a *Riesz basis*. See, e.g., [19] or [5] for more information about such bases.

4.31 Let $\{\mathbf{e}_k\}_{k=1}^{\infty}$ be an orthonormal system in a Hilbert space $\mathcal{H}$. Show that for all $\{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N})$,

$$\left|\left| \sum_{k=1}^{\infty} c_k \mathbf{e}_k \right|\right|^2 = \sum_{k=1}^{\infty} |c_k|^2.$$

4.32 Let $\{\mathbf{e}_k\}_{k=1}^{\infty}$ be an orthonormal system in a Hilbert space $\mathcal{H}$. Show that

$$\sum_{k=1}^{\infty} c_k \mathbf{e}_k \text{ is convergent} \Leftrightarrow \ \{c_k\}_{k=1}^{\infty} \in \ell^2(\mathbb{N}).$$

5

The L^p-spaces

In this chapter we introduce and analyze an important class of Banach spaces consisting of functions, the so-called L^p-spaces, where p is a parameter that specifies a particular space. First, in Section 5.1 we discuss some vector spaces and concepts that play a role in the analysis of the L^p-spaces. In Section 5.2 we introduce the space $L^1(\mathbb{R})$, and show how to make it a normed space. Integration techniques in $L^1(\mathbb{R})$ are discussed in Section 5.3; in particular, we consider techniques for interchanging an integral and an infinite sum. For $p \in]1, \infty]$, the spaces $L^p(\mathbb{R})$ are introduced in Section 5.4; their counterparts $L^p(a, b)$ concerning functions defined on a subinterval $]a, b[\subset \mathbb{R}$ are discussed in Section 5.5.

In the definition and analysis of the L^p-spaces, we are at some points forced to skip the discussion of certain subtle details. In fact, a complete mathematical treatment of these spaces would require us to deal with the concept of measurable functions; it is outside the scope of the book to do that (see, e.g., [17] for a detailed presentation). Fortunately, all "natural" functions, e.g., all piecewise continuous functions and even all functions that we can just write down "by hand," are measurable. Thus, this shortcoming will not have any influence in practice.

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4980-7_5,

5.1 Vector spaces consisting of continuous functions

In this chapter, we are mainly interested in functions defined on the set $\mathbb{R}$. For technical reasons it is often necessary to approximate such functions by continuous functions that are zero outside a bounded interval, or, at least, decay to zero as the variable tends to $\pm\infty$. We will therefore introduce vector spaces consisting of such functions first.

Definition 5.1.1 (Support, classes of continuous functions) *Consider a function $f : \mathbb{R} \to \mathbb{C}$.*

(i) The support of the function f is the smallest closed set outside which the function is equal to zero:

$$\text{supp}\, f = \overline{\{x \in \mathbb{R} \mid f(x) \neq 0\}}.$$

(ii) If $\text{supp}\, f$ is a bounded set, i.e., contained in an interval of the form $[a, b]$ for some $a, b \in \mathbb{R}$, we say that f has compact support.

(iii) The vector space $C_c(\mathbb{R})$ consists of all continuous functions having compact support:

$$C_c(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{C} \mid f \text{ is continuous and has compact support}\}.$$

(iv) The vector space $C_0(\mathbb{R})$ consists of all continuous functions that tends to zero as $x \to \pm\infty$:

$$C_0(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{C} \mid f \text{ is continuous and } f(x) \to 0 \text{ as } x \to \pm\infty\}.$$

The verification that the spaces $C_c(\mathbb{R})$ and $C_0(\mathbb{R})$ actually are vector spaces with respect to the usual operations of addition and scalar multiplication is left to the reader. Note that a function having compact support is equivalent to the function being zero outside some finite interval. We illustrate the definitions with some examples:

Example 5.1.2 (The vector spaces $C_c(\mathbb{R})$ and $C_0(\mathbb{R})$) We suggest that the reader make a sketch of the following functions.

(i) For the function

$$f(x) = \begin{cases} x & \text{if } x \in [0, 1[, \\ 2 - x & \text{if } x \in [1, 2], \\ 0 & \text{otherwise,} \end{cases}$$

we see that

$$\{x \in \mathbb{R} \mid f(x) \neq 0\} =]0, 2[.$$

Thus, $\operatorname{supp} f = [0,2]$, i.e., f has compact support. Because f is continuous, we conclude that $f \in C_c(\mathbb{R})$. The function f is called a B-spline, see Chapter 10.

(ii) The function

$$f(x) = \begin{cases} 2x & \text{if } x \in [0,1[, \\ 2-x & \text{if } x \in [1,2], \\ 0 & \text{otherwise,} \end{cases}$$

is not continuous, so it does not belong to $C_c(\mathbb{R})$. However, it has compact support, $\operatorname{supp} f = [0,2]$.

(iii) The function

$$f(x) = \begin{cases} 1 & \text{if } x \in [-1,1], \\ \frac{1}{|x|} & \text{if } x \in \mathbb{R} \setminus [-1,1], \end{cases}$$

does not have compact support, so it does not belong to $C_c(\mathbb{R})$. However, f is continuous and $f(x) \to 0$ as $x \to \pm\infty$, i.e., $f \in C_0(\mathbb{R})$.

□

We now state some of the central properties for the spaces $C_c(\mathbb{R})$ and $C_0(\mathbb{R})$. We begin with $C_0(\mathbb{R})$:

Lemma 5.1.3 (Norm on $C_0(\mathbb{R})$) *The expression*

$$||f||_\infty = \max_{x\in\mathbb{R}} |f(x)| \tag{5.1}$$

defines a norm on $C_0(\mathbb{R})$ that makes it a Banach space.

Proof. We first show that the expression in (5.1) makes sense and is finite for all $f \in C_0(\mathbb{R})$. The result is obvious for the function $f = 0$, so consider any function $f \in C_0(\mathbb{R}) \setminus \{0\}$. Choose $x_0 \in \mathbb{R}$ such that $\delta := |f(x_0)| > 0$. By definition of $C_0(\mathbb{R})$, there exists a number $K > 0$ such that

$$|f(x)| \le \delta/2 \text{ whenever } |x| \ge K.$$

Also, according to Theorem 1.6.3, there exists an $x_1 \in [-K, K]$ such that

$$|f(x_1)| = \max_{x\in[-K,K]} |f(x)|.$$

Due to the choice of K, we necessarily have that $x_0 \in]-K, K[$. Therefore, $|f(x_1)| \ge |f(x_0)| = \delta$; thus, the maximum value for $|f(x)|$ for $x \in \mathbb{R}$ exists and equals $|f(x_1)|$.

Now, that (5.1) is a norm on $C_0(\mathbb{R})$ is proved exactly like the corresponding result for $C[a,b]$ in Example 2.1.4. In order to show that $C_0(\mathbb{R})$ is a

Banach space, let $\{f_k\}_{k=1}^\infty$ be a Cauchy sequence of functions belonging to $C_0(\mathbb{R})$. Thus, for any $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$||f_k - f_\ell||_\infty \leq \epsilon \text{ whenever } k, \ell \geq N.$$

Fix any $x \in [a, b]$. Since

$$|f_k(x) - f_\ell(x)| \leq ||f_k - f_\ell||_\infty,$$

we infer that $\{f_k(x)\}_{k=1}^\infty$ is a Cauchy sequence in $\mathbb{C}$. Since $\mathbb{C}$ is a Banach space the sequence is convergent, and we can define a function f by

$$f(x) := \lim_{k\to\infty} f_k(x).$$

Exactly like in the proof of Theorem 3.1.6 it follows that f is continuous and

$$||f - f_k||_\infty \to 0 \text{ as } k \to \infty. \tag{5.2}$$

Thus, we only need to show that

$$f(x) \to 0 \text{ as } x \to \pm\infty. \tag{5.3}$$

The exact meaning of (5.3) is that for each $\epsilon > 0$ there exists an $M > 0$ such that

$$|f(x)| \leq \epsilon \text{ whenever } |x| \geq M.$$

Now, given $\epsilon > 0$, it follows from the above considerations that we can choose $N \in \mathbb{N}$ such that

$$||f - f_k||_\infty \leq \epsilon/2 \text{ whenever } k \geq N.$$

For the chosen N, consider the function f_N: since this function belongs to $C_0(\mathbb{R})$, we can find an $M > 0$ such that

$$|f_N(x)| \leq \epsilon/2 \text{ whenever } |x| \geq M.$$

Thus, for $|x| \geq M$,

$$\begin{aligned} |f(x)| = |f(x) - f_N(x) + f_N(x)| &\leq |f(x) - f_N(x)| + |f_N(x)| \\ &\leq ||f - f_N||_\infty + |f_N(x)| \\ &\leq \epsilon/2 + \epsilon/2 \\ &= \epsilon. \end{aligned}$$

This conclude the proof. □

The norm

$$||f||_\infty = \max_{x\in\mathbb{R}} |f(x)| \tag{5.4}$$

is called the *supremum-norm,* or *infinity-norm.* It also defines a norm on $C_c(\mathbb{R})$, but the resulting normed space is not a Banach space:

Lemma 5.1.4 (Norm on $C_c(\mathbb{R})$) *The vector space $C_c(\mathbb{R})$ is a subspace of $C_0(\mathbb{R})$. It is a normed vector space with respect to the norm*

$$||f||_\infty = \max_{x\in\mathbb{R}} |f(x)|,$$

but not a Banach space.

Proof. That $C_c(\mathbb{R})$ is a subset of $C_0(\mathbb{R})$ follows directly from the definition; and that it is a subspace follows from Lemma 1.2.7. Since $||\cdot||_\infty$ is a norm on $C_0(\mathbb{R})$, it is also a norm on the subspace $C_c(\mathbb{R})$.

We will now show that $C_c(\mathbb{R})$ cannot be a Banach space with respect to the norm $||\cdot||_\infty$. In order to do so, we use the characteristic functions defined in (1.20), and define a collection of functions $\{f_k\}_{k=1}^\infty$ by

$$f_k(x) := \begin{cases} \dfrac{\sin x}{x}\chi_{[-k\pi,k\pi]}(x) & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases} \tag{5.5}$$

Note that the function f_k has compact support, $\operatorname{supp} f_k = [-k\pi, k\pi]$. Because $\sin(k\pi) = 0$ for all $k \in \mathbb{Z}$ and $x^{-1}\sin x \to 1$ as $x \to 0$, we see that f_k is continuous, i.e., $f_k \in C_c(\mathbb{R})$ for all $k \in \mathbb{N}$. Define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) := \begin{cases} \dfrac{\sin x}{x} & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases}$$

Then $f \in C_0(\mathbb{R}) \setminus C_c(\mathbb{R})$. Note further that for all $x \in \mathbb{R}$ and all $k \in \mathbb{N}$,

$$\begin{aligned} |f(x) - f_k(x)| &= \left| \frac{\sin x}{x}\chi_{\mathbb{R}\setminus[-k\pi,k\pi]}(x) \right| \\ &\leq \frac{1}{k\pi}; \end{aligned}$$

thus,

$$||f - f_k||_\infty \leq \frac{1}{k\pi}.$$

We conclude that

$$||f - f_k||_\infty \to 0 \text{ as } k \to \infty.$$

The same argument (Exercise 5.1) shows that $\{f_k\}_{k=1}^\infty$ is a Cauchy sequence in $C_c(\mathbb{R})$. Thus, we have a Cauchy sequence $\{f_k\}_{k=1}^\infty$ of functions in $C_c(\mathbb{R})$ that converges to a function not belonging to $C_c(\mathbb{R})$; this shows that $C_c(\mathbb{R})$ cannot be a Banach space. □

In Exercise 5.4, we ask the reader to consider the vector space $C_c(\mathbb{R})$ with another norm than the $||\cdot||_\infty$-norm. This alternative norm comes from an inner product, but $C_c(\mathbb{R})$ is not a Hilbert space with respect to that norm.

5.2 The vector space $L^1(\mathbb{R})$

We will now introduce one of the central vector spaces in mathematical analysis. Unfortunately, a completely axiomatic treatment would require that we introduce measure theory and Lebesgue integration, two topics that are outside the scope of this book. Thus, at a few points, we are forced to explain the theory in words rather than providing a completely exact presentation. In particular, we will not explain the meaning of a *measurable* function; since every function one can write down "by hand," e.g., each piecewise continuous function, actually is measurable, this shortcoming is not critical. In fact, the existence of nonmeasurable functions is completely nonconstructive, and as a practical matter the measurability will never be an issue.

As integrability issues will play a role at several instances, we do need to spend a few words on the meaning of the formal expression

$$\int_{-\infty}^{\infty} f(x)\, dx. \tag{5.6}$$

Whenever $f : \mathbb{R} \to \mathbb{C}$ is a piecewise continuous function, the integral appearing in (5.6) is the *improper Riemann integral,* discussed shortly in Section 1.7. The theory for measurable functions allows us to integrate a larger class of functions in terms of the so-called *Lebesgue integral.* Since the Riemann integral and the Lebesgue integral coincide for piecewise continuous functions this does not lead to any ambiguity in the definition of (5.6). We will not discuss the exact meaning of the integral for functions that are Lebesgue integrable but not Riemann integrable.

All concrete calculations in this book will be performed on piecewise continuous functions, i.e., on functions that are Riemann integrable. Just to satisfy the curiosity of the reader, we mention that the function

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \notin \mathbb{Q}, \end{cases} \tag{5.7}$$

is an example of a measurable function that is not piecewise continuous; the function is Lebesgue integrable, but not Riemann integrable.

Definition 5.2.1 (Integrable functions) *A (measurable) function $f : \mathbb{R} \to \mathbb{C}$ is said to be integrable if*

$$\int_{-\infty}^{\infty} |f(x)|\, dx < \infty. \tag{5.8}$$

The set consisting of all integrable functions on $\mathbb{R}$ is denoted by $L^1(\mathbb{R})$:

$$L^1(\mathbb{R}) := \left\{ f : \mathbb{R} \to \mathbb{C} \;\middle|\; \int_{-\infty}^{\infty} |f(x)|\, dx < \infty \right\}.$$

We will only consider integrals $\int_{-\infty}^{\infty} f(x)\,dx$ under the assumption that $f \in L^1(\mathbb{R})$. We leave it to the reader to verify that $L^1(\mathbb{R})$ is a vector space (Exercise 5.5). For piecewise continuous functions f, we verify the condition (5.8) using the Riemann integral:

Example 5.2.2 (The $L^1(\mathbb{R})$-condition) We consider various functions and check whether they belong to $L^1(\mathbb{R})$.

(i) For the function

$$f(x) \quad = \quad \begin{cases} 2x & \text{if } x \in [0,1[, \\ 2-x & \text{if } x \in [1,2], \\ 0 & \text{otherwise,} \end{cases}$$

(considered in Example 5.1.2(ii)) we see that

$$\begin{aligned} \int_{-\infty}^{\infty} |f(x)|\,dx \quad &= \quad \int_0^1 |2x|\,dx + \int_1^2 |2-x|\,dx \\ &= \quad \frac{3}{2} < \infty; \end{aligned}$$

thus, $f \in L^1(\mathbb{R})$. The same type of argument shows that the B-spline in Example 5.1.2(i) belongs to $L^1(\mathbb{R})$.

(ii) Consider the function

$$f(x) \quad = \quad \begin{cases} 1 & \text{if } x \in [-1,1], \\ \frac{1}{|x|} & \text{if } x \in \mathbb{R} \setminus [-1,1], \end{cases}$$

also appearing in Example 5.1.2(iii). First, we note that for any $\alpha > 1$,

$$\int_1^{\alpha} \frac{1}{|x|}\,dx = \int_1^{\alpha} \frac{1}{x}\,dx = \ln \alpha \to \infty \text{ as } \alpha \to \infty;$$

this shows that

$$\int_1^{\infty} |f(x)|\,dx = \infty.$$

It follows that $f \notin L^1(\mathbb{R})$.

(iii) Consider the function

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}} & \text{if } x \in]0,4], \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\int_{-\infty}^{\infty} |f(x)|\;dx = \int_0^4 \frac{1}{2\sqrt{x}}\,dx = \left[\sqrt{x}\right]_0^4 = 2 < \infty;$$

thus, $f \in L^1(\mathbb{R})$. See Figure 5.1.

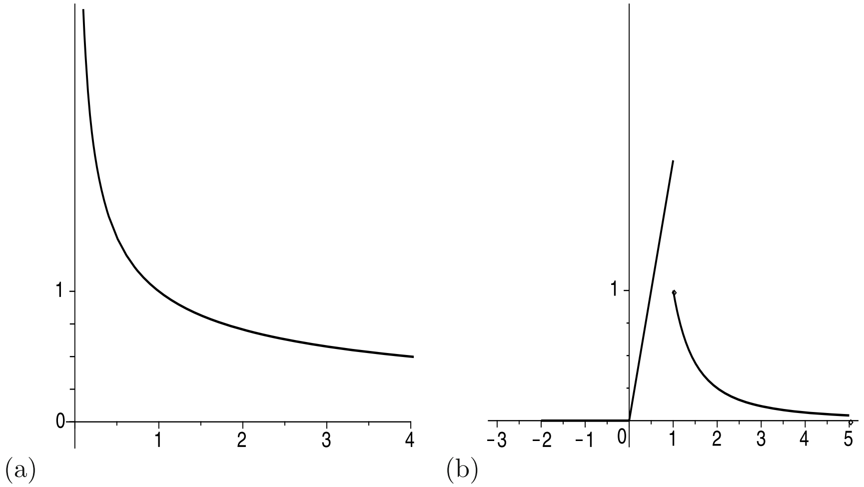

Figure 5.1. (a): The graph of the function f in Example 5.2.2(iii). (b): The function f in Example 5.2.2(iv).

(iv) For the function

$$f(x) = \begin{cases} 0 & \text{if } x \in]-\infty, 0], \\ 2x & \text{if } x \in [0,1], \\ x^{-2} & \text{if } x \in]1, \infty[, \end{cases}$$

we see that

$$\int_{-\infty}^{\infty} |f(x)|\, dx = \int_0^1 2x\, dx + \int_1^{\infty} \frac{1}{x^2}\, dx = 1 + \int_1^{\infty} \frac{1}{x^2}\, dx.$$

Now,

$$\int_1^{\alpha} \frac{1}{x^2}\, dx = \left[\frac{-1}{x}\right]_1^{\alpha} = 1 - \frac{1}{\alpha} \to 1 \text{ as } \alpha \to \infty;$$

thus,

$$\int_{-\infty}^{\infty} |f(x)|\, dx = 2,$$

i.e., $f \in L^1(\mathbb{R})$. See Figure 5.1. □

We have now introduced three vector spaces of functions, namely, $C_c(\mathbb{R}), C_0(\mathbb{R})$, and $L^1(\mathbb{R})$. One might wonder what the advantages of the various spaces are. At a first glance, the space $C_c(\mathbb{R})$ is the most convenient space to use: in practical implementations one can only consider functions on a bounded interval, so the compact support of the functions in $C_c(\mathbb{R})$ is convenient. However, the space $C_c(\mathbb{R})$ also has severe disadvantages:

(i) The space $C_c(\mathbb{R})$ is not a Banach space with respect to the natural norm $||\cdot||_\infty$;

(ii) The functions in $C_c(\mathbb{R})$ are continuous, but discontinuous functions appear in all areas of science and engineering.

The comment in (ii) also explains why $C_0(\mathbb{R})$ should not be chosen as our favorite space. Motivated by these considerations, we will consider the space $L^1(\mathbb{R})$. Later, in Theorem 5.4.2, we will see that our "favorite" vector space $C_c(\mathbb{R})$ is closely related with $L^1(\mathbb{R})$: it actually forms a dense subspace of $L^1(\mathbb{R})$.

Our goal now is to show how $L^1(\mathbb{R})$ can be equipped with a norm that makes it a Banach space. We first note that the norm $||\cdot||_\infty$ in (5.4) on the spaces $C_c(\mathbb{R})$ and $C_0(\mathbb{R})$ does not define a norm on $L^1(\mathbb{R})$:

Example 5.2.3 ($||\cdot||_\infty$ does not form a norm on $L^1(\mathbb{R})$) The function $f \in L^1(\mathbb{R})$ in Example 5.2.2(iii) is unbounded, so

$$\sup_{x\in\mathbb{R}} |f(x)| = \infty.$$

But one of the requirements in Definition 2.1.1 is that the norm of any vector is a finite real number; thus, $||\cdot||_\infty$ does not define a norm on $L^1(\mathbb{R})$. □

We will now consider an alternative "guess" for how a norm can be chosen. We define the mapping $||\cdot||_1 : L^1(\mathbb{R}) \to \mathbb{R}$ by

$$||f||_1 := \int_{-\infty}^{\infty} |f(x)|\, dx, \ \ f \in L^1(\mathbb{R}). \tag{5.9}$$

The advantage of this definition is that $||\cdot||_1$ takes finite and nonnegative real values by definition of $L^1(\mathbb{R})$. Unfortunately, the expression $||\cdot||_1$ *does not* define a norm on $L^1(\mathbb{R})$: it does not satisfy condition (i) in Definition 2.1.1! In fact, it is clear that the function

$$f = 0$$

satisfies that $||f||_1 = 0$; but there exist functions f that are not identically zero and still have the property that $||f||_1 = 0$. For example, the function $\tilde{f}$ given by

$$\tilde{f}(x) = \begin{cases} 1 \text{ if } x = 0, \\ 0 \text{ if } x \in \mathbb{R} \setminus \{0\}, \end{cases} \tag{5.10}$$

also satisfies that $||\tilde{f}||_1 = 0$.

Mathematically, there is a way to circumvent this problem. The exact way of doing it is described in books on measure theory (see, e.g., [17]). We define a so-called *equivalence relation* $\sim$ between the functions in $L^1(\mathbb{R})$ by

$$f \sim g \Leftrightarrow \int_{-\infty}^{\infty} |f(x) - g(x)|\, dx = 0. \tag{5.11}$$

We will say that two functions $f, g \in L^1(\mathbb{R})$ are *equivalent* if $f \sim g$ in the sense of (5.11); in that case we will also say that f and g are identical *almost everywhere*. This is in accordance with the use of this wording in measure theory.

We will now *identify* functions that are equivalent; that is, we will not distinguish between functions f and g for which $f \sim g$. Strictly speaking, the identification in (5.11) implies that the elements in $L^1(\mathbb{R})$ are no longer functions: they are (equivalence) classes of functions.

The following important result presents a criterion implying equivalence of two given functions. It is based on the concept of countable sets, which is explained in Exercise 5.6.

Lemma 5.2.4 (Equivalence of functions) *Assume that $\Gamma \in \mathbb{R}$ is a countable set, and let f and g be two functions from $\mathbb{R}$ to $\mathbb{C}$. If*

$$f(x) = g(x) \text{ for all } x \in \mathbb{R} \setminus \Gamma,$$

then $f \sim g$.

For the purpose in the current book it is enough to know that the set $\Gamma = \mathbb{Z}$ and any subset hereof are countable (Exercise 5.6). The set $\mathbb{R}$ is *uncountable.*

Example 5.2.5 (Equivalence of functions) We illustrate the identification of functions with some examples.

(i) The function $\tilde{f}$ in (5.10) differs from the zero function only for $x = 0$; thus, $\tilde{f} \sim 0$. That is, we will not distinguish between the zero function and the function $\tilde{f}$.

(ii) Let

$$f(x) = x, \quad g(x) = \begin{cases} x \text{ if } x \in \mathbb{R} \setminus \{1, 2\}, \\ 0 \text{ if } x \in \{1, 2\}. \end{cases}$$

Then $f(x) = g(x)$, except for $x \in \{1, 2\}$. Thus, $f \sim g$.

(iii) The sets $\mathbb{N}, \mathbb{Z}$, and $\mathbb{Q}$ are countable (Exercise 5.6). In particular, the functions

$$f(x) = x \text{ and } h(x) = \begin{cases} x \text{ if } x \in \mathbb{R} \setminus \mathbb{Z}, \\ 0 \text{ if } x \in \mathbb{Z}, \end{cases}$$

are equivalent, $f \sim h$.

(iv) Consider the function $\chi_{\mathbb{Q}}$ in (5.7). It follows from the results in Exercise 5.6 that $\chi_{\mathbb{Q}} \sim 0$.

□

With the identification in (5.11), there is exactly *one* element f satisfying $||f||_1 = 0$, and $|| \cdot ||_1$ defines a norm:

Theorem 5.2.6 (Norm on $L^1(\mathbb{R})$) *Identifying functions which are equivalent in the sense of (5.11), the expression*

$$||f||_1 := \int_{-\infty}^{\infty} |f(x)|\, dx \tag{5.12}$$

defines a norm on $L^1(\mathbb{R})$.

Proof. It is clear that $|| \cdot ||_1$ satisfies the condition (ii) in Definition 2.1.1. Also, for $f, g \in L^1(\mathbb{R})$, the triangle inequality shows that

$$\begin{aligned} ||f+g||_1 &= \int_{-\infty}^{\infty} |f(x)+g(x)|\, dx \\ &\leq \int_{-\infty}^{\infty} (|f(x)| + |g(x)|)\, dx \\ &= \int_{-\infty}^{\infty} |f(x)|\, dx + \int_{-\infty}^{\infty} |g(x)|\, dx \\ &= ||f||_1 + ||g||_1, \end{aligned}$$

i.e., condition (iii) in Definition 2.1.1 is also satisfied. As we have seen, the equation $||f||_1 = 0$ only holds for the functions in the equivalence class with $f = 0$; this concludes the proof. □

5.3 Integration in $L^1(\mathbb{R})$

Integration is a central tool in mathematical analysis. At several instances we need to integrate functions in $L^1(\mathbb{R})$ that are given in terms of infinite series; in such a case, the integral can often be evaluated by an exchange of the order of integration and summation. In the literature, several techniques have been introduced in order to do so. We state the results in the full generality of measurable functions, and suggest that the reader without knowledge about measure theory simply assumes the functions to be piecewise continuous instead. Since the proofs are measure-theoretic by nature, we skip them.

Before we state the main result in Theorem 5.3.5, we mention *Fatou's Lemma*:

Lemma 5.3.1 (Fatou's lemma) *Let* $f_n : \mathbb{R} \to [0, \infty[$, $n \in \mathbb{N}$, *be a sequence of integrable functions. Then the function* $\liminf_{n\to\infty} f_n$ *is integrable, and*

$$\int_{-\infty}^{\infty} \liminf_{n\to\infty} f_n(x)\,dx \le \liminf_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,dx. \tag{5.13}$$

Observe the inequality in (5.13)! The next example exhibits a case where strict inequality appears:

Example 5.3.2 (Strict inequality in Fatou's lemma) For $n \in \mathbb{N}$, let

$$f_n(x) = \begin{cases} nx & \text{if } x \in [0, 1/n], \\ 0 & \text{if } x \notin [0, 1/n]. \end{cases}$$

We suggest that the reader makes a draft of a few of these functions. Note that $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$. It follows that

$$\liminf_{n\to\infty} f_n(x) = 0, \ \ x \in \mathbb{R},$$

so

$$\int_{-\infty}^{\infty} \liminf_{n\to\infty} f_n(x)\,dx = 0.$$

On the other hand, for all $n \in \mathbb{N}$,

$$\int_{-\infty}^{\infty} f_n(x)\,dx = \frac{1}{2}, \tag{5.14}$$

so

$$\liminf_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,dx = \frac{1}{2}.$$

Thus, in this particular case,

$$\int_{-\infty}^{\infty} \liminf_{n\to\infty} f_n(x)\,dx < \liminf_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,dx. \qquad \square$$

As stated in the introduction to this section, we want to find conditions on a sequence of functions $\{h_k\}_{k=1}^{\infty}$ on $\mathbb{R}$ implying that

$$\int_{-\infty}^{\infty} \sum_{k=1}^{\infty} h_k(x)\,dx = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} h_k(x)\,dx. \tag{5.15}$$

For a *monotone increasing* sequence of positive functions such conditions follow from *Lebesgue's Monotone Convergence Theorem:*

Theorem 5.3.3 (Lebesgue's monotone convergence theorem) *Suppose that $f_n : \mathbb{R} \to \mathbb{R},\ n \in \mathbb{N}$, is an increasing sequence of (measurable) functions, i.e., that*

$$f_1(x) \le f_2(x) \le \cdots, \text{ for all } x \in \mathbb{R}. \tag{5.16}$$

Then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\, dx = \int_{-\infty}^{\infty} \lim_{n\to\infty} f_n(x)\, dx.$$

Theorem 5.3.3 is often used to justify an exchange of the order of summation and integration for positive functions:

Example 5.3.4 (Exchange of the order of integral and sum) For any sequence of (measurable) positive-valued functions $h_n,\ n \in \mathbb{N}$, Theorem 5.3.3 implies that (Exercise 5.7)

$$\int_{-\infty}^{\infty} \sum_{k=1}^{\infty} h_k(x)\, dx = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} h_k(x)\, dx. \qquad \square$$

For functions f_n that are not necessarily positive, we need extra conditions in order for (5.15) to hold. Such conditions can be derived based on *Lebesgue's Dominated Convergence Theorem:*

Theorem 5.3.5 (Lebesgue's dominated convergence theorem) *Let $f : \mathbb{R} \to \mathbb{C}$ be given. Suppose that $f_n : \mathbb{R} \to \mathbb{C},\ n \in \mathbb{N}$, is a sequence of (measurable) functions such that*

(i) $f_n(x) \to f(x)$ pointwise as $n \to \infty$;

(ii) There exists a positive, integrable function g such that

$$|f_n(x)| \le g(x) \text{ for all } n \in \mathbb{N},\ x \in \mathbb{R}. \tag{5.17}$$

Then f is integrable, and

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx.$$

Before we show how to apply Theorem 5.3.5, we state an alternative version, where the *sequence* of functions f_n is replaced by a family of functions f_δ, parametrized by $\delta > 0$:

Theorem 5.3.6 (Lebesgue's dominated convergence theorem) *Let $f : \mathbb{R} \to \mathbb{C}$ be given. Suppose that for each $\delta > 0$ a (measurable) function $f_\delta : \mathbb{R} \to \mathbb{C}$ is given and that*

(i) $f_\delta(x) \to f(x)$ *pointwise as* $\delta \to 0$*;*

(ii) There exists a positive, integrable function g such that

$$|f_\delta(x)| \leq g(x) \text{ for all } \delta > 0, \ x \in \mathbb{R}.$$

Then f is integrable, and

$$\lim_{\delta \to 0} \int_{-\infty}^{\infty} f_\delta(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx.$$

Let us now return to the issue of how to verify (5.15). There is a standard procedure to do that, based on Theorems 5.3.3 and 5.3.5:

Theorem 5.3.7 (Exchange of the order of integral and sum) *Assume that $h_n, \ n \in \mathbb{N}$, is a sequence of (measurable) functions and that*

$$\sum_{k=1}^{\infty} \int_{-\infty}^{\infty} |h_k(x)| \, dx < \infty. \tag{5.18}$$

Then

$$\int_{-\infty}^{\infty} \sum_{k=1}^{\infty} h_k(x)\, dx = \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} h_k(x)\, dx. \tag{5.19}$$

Proof. We will demonstrate how the result follows from Theorems 5.3.3 and 5.3.5, but skip the measure-theoretic parts. Put

$$f(x) := \sum_{k=1}^{\infty} h_k(x).$$

Let the functions $f_n, \ n \in \mathbb{N}$, be defined by

$$f_n(x) := \sum_{k=1}^{n} h_k(x), \ x \in \mathbb{R}.$$

Then

$$f_n(x) \to f(x) \text{ as } n \to \infty,$$

and

$$\begin{aligned} |f_n(x)| = \left| \sum_{k=1}^{n} h_k(x) \right| &\leq \sum_{k=1}^{n} |h_k(x)| \\ &\leq \sum_{k=1}^{\infty} |h_k(x)|. \end{aligned}$$

Putting

$$g(x) := \sum_{k=1}^{\infty} |h_k(x)|\,, \tag{5.20}$$

this means that (5.17) is satisfied. That g actually defines a function (i.e., that the infinite sum in (5.20) is convergent for a.e. $x \in \mathbb{R}$) can be shown to be a consequence of (5.18). Using the result in Example 5.3.4 and (5.18) again,

$$\int_{-\infty}^{\infty} |g(x)|\,dx = \int_{-\infty}^{\infty} \sum_{k=1}^{\infty} |h_k(x)|\ dx \quad = \quad \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} |h_k(x)|\ dx < \infty,$$

i.e., $g \in L^1(\mathbb{R})$. It now follows from Theorem 5.3.5 that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx,$$

i.e., that

$$\begin{aligned}
\int_{-\infty}^{\infty} \sum_{k=1}^{\infty} h_k(x)\,dx &= \lim_{n\to\infty} \int_{-\infty}^{\infty} \sum_{k=1}^{n} h_k(x)\,dx \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \int_{-\infty}^{\infty} h_k(x)\,dx \\
&= \sum_{k=1}^{\infty} \int_{-\infty}^{\infty} h_k(x)\,dx.
\end{aligned}$$

This completes the proof. □

Let us apply the procedure in Theorem 5.3.7 on a concrete case:

Example 5.3.8 (Exchange of the order of integral and sum) Consider the function

$$f(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^2} \chi_{[k,k+1[}(x).$$

Looking carefully at the definition of that function (do that!) we see that

$$f(x) = \begin{cases} \frac{(-1)^{k+1}}{k^2} & \text{if } x \in [k, k+1[,\ \ k \in \mathbb{N}, \\ 0 & \text{if } x \notin [1, \infty[. \end{cases}$$

We suggest that the reader makes a draft of the function. Note that

$$\begin{aligned}\int_{-\infty}^{\infty}\sum_{k=1}^{\infty}\left|\frac{(-1)^{k+1}}{k^2}\chi_{[k,k+1[}(x)\right|\,dx &= \int_{-\infty}^{\infty}\sum_{k=1}^{\infty}\frac{1}{k^2}\chi_{[k,k+1[}(x)\,dx\\ &= \sum_{k=1}^{\infty}\int_{-\infty}^{\infty}\frac{1}{k^2}\chi_{[k,k+1[}(x)\,dx\\ &= \sum_{k=1}^{\infty}\int_{k}^{k+1}\frac{1}{k^2}\,dx\\ &= \sum_{k=1}^{\infty}\frac{1}{k^2}. \qquad (5.21)\end{aligned}$$

Any series of the form $\sum_{k=1}^{\infty} k^{-\alpha}$ with $\alpha > 1$ is convergent, so we conclude that (5.21) is finite. By Theorem 5.3.7 we now conclude that the function f is integrable, and that

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,dx &= \int_{-\infty}^{\infty}\lim_{n\to\infty} f_n(x)\,dx\\ &= \lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)\,dx\\ &= \lim_{n\to\infty}\int_{-\infty}^{\infty}\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k^2}\chi_{[k,k+1[}(x)\,dx\\ &= \lim_{n\to\infty}\sum_{k=1}^{n}\int_{-\infty}^{\infty}\frac{(-1)^{k+1}}{k^2}\chi_{[k,k+1[}(x)\,dx\\ &= \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k^2}.\end{aligned}$$

□

Note that there are cases where the dominating function g does not belong to $L^1(\mathbb{R})$, i.e., cases where the above procedure does not work:

Example 5.3.9 (Exchange of the order of integral and sum) With the functions f_n defined in Example 5.3.2, we have

$$f(x) := \lim_{n\to\infty} f_n(x) = 0, \ \forall x \in \mathbb{R}.$$

Thus,

$$\int_{-\infty}^{\infty} f(x)\,dx = 0.$$

Together with (5.14) this shows that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,dx \neq \int_{-\infty}^{\infty} f(x)\,dx.$$

This does not contradict Theorems 5.3.3 and 5.3.5. In fact, the sequence f_n is not increasing in the sense of (5.16) — and no dominating integrable function g as in (5.17) exists. □

Later, e.g., in the context of the Fourier transform and convolution, we will see cases where one has to integrate a function of two variables. Very often, it is a computational simplification to switch the order of integration. Technically, this can be done via *Fubini's Theorem:*

Theorem 5.3.10 (Fubini's theorem) *Given a function* $f : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$, *assume that*

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(x,y)|\,dx \right) dy < \infty.$$

Then

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x,y)\,dx \right) dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x,y)\,dy \right) dx.$$

5.4 The spaces $L^p(\mathbb{R})$

We are now ready to define the Banach spaces $L^p(\mathbb{R})$ for $1 < p < \infty$. For $1 < p < \infty$, $L^p(\mathbb{R})$ is the space of functions f for which $|f|^p$ is integrable:

$$L^p(\mathbb{R}) := \left\{ f : \mathbb{R} \to \mathbb{C} \;\middle|\; \int_{-\infty}^{\infty} |f(x)|^p\,dx < \infty \right\}. \tag{5.22}$$

As in the special case of $L^1(\mathbb{R})$, we can define a norm on $L^p(\mathbb{R})$. Define the equivalence relation between the functions in $L^p(\mathbb{R})$ by

$$f \sim g \Leftrightarrow \int_{-\infty}^{\infty} |f(x) - g(x)|^p\,dx = 0.$$

Identifying equivalent functions similarly to what we did for $L^1(\mathbb{R})$, one can show the following important result (Exercise 5.8):

Theorem 5.4.1 (Norm on $L^p(\mathbb{R})$) *For each* $p \in [1, \infty[$, *the expression*

$$||f||_p = \left(\int_{-\infty}^{\infty} |f(x)|^p\,dx \right)^{1/p} \tag{5.23}$$

defines a norm on $L^p(\mathbb{R})$ *that makes* $L^p(\mathbb{R})$ *a Banach space.*

Technically, it is complicated to deal with functions in $L^p(\mathbb{R})$: they might be unbounded, and they do not need to have compact support. Therefore, it is important that all functions in $L^p(\mathbb{R})$ can be approximated arbitrarily well by continuous functions with compact support:

Theorem 5.4.2 ($C_c(\mathbb{R})$ dense in $L^p(\mathbb{R})$) *For each $p \in [1, \infty[$, the vector space $C_c(\mathbb{R})$ is a dense subspace of $L^p(\mathbb{R})$.*

Theorem 5.4.2 can be proved based on properties of the approximating identities discussed in Section A.2, see [18]. Readers who want to understand how to approximate functions in $L^1(\mathbb{R})$ by functions in $C_c(\mathbb{R})$ are referred to Exercise 5.14. Also, in Exercise 5.15 we ask the reader to show that $C_c(\mathbb{R})$ is a subspace of $L^p(\mathbb{R})$ for all $p \in [1, \infty[$.

The conclusion of Theorem 5.4.2 means that if $g \in L^p(\mathbb{R})$ for some $p \in [1, \infty[$, there exists a sequence $\{g_k\}_{k=1}^{\infty}$ of continuous functions with compact support, for which

$$||g - g_k||_p = \left(\int_{-\infty}^{\infty} |g(x) - g_k(x)|^p \, dx \right)^{1/p} \to 0 \text{ as } k \to \infty. \tag{5.24}$$

It should be noted that (5.24) *does not* imply that the function values $g_k(x)$ converge to $g(x)$ for all $x \in \mathbb{R}$! We come back to this issue in the context of Fourier series in Section 6.4.

So far, we have defined the Banach spaces $L^p(\mathbb{R})$ for all $p \in [1, \infty[$. In the mathematical analysis one also encounters the space $L^\infty(\mathbb{R})$, which we introduce now. Let

$$L^\infty(\mathbb{R}) := \big\{ f : \mathbb{R} \to \mathbb{C} \;\big|\; f \text{ is bounded} \big\}. \tag{5.25}$$

The vector space $L^\infty(\mathbb{R})$ is a Banach space with respect to the norm

$$||f||_\infty = \sup_{x \in \mathbb{R}} |f(x)|. \tag{5.26}$$

Note that a completely exact definition of the space $L^\infty(\mathbb{R})$ and the associated norm would require us to replace "bounded functions" by *functions that are bounded almost everywhere*, and the supremum-norm by the *essential supremum-norm*. An understanding of these concepts requires knowledge of measure theory, so we are not able to go into the details here.

5.5 The spaces $L^p(a, b)$

In this chapter we have only dealt with function spaces consisting of functions defined on all of $\mathbb{R}$. Similar results can be derived for functions defined on subintervals of $\mathbb{R}$. Given any interval $]a,b[\subseteq \mathbb{R}$ and any $p \in [1,\infty[$, we can consider the vector space

$$L^p(a,b) := \left\{ f :]a,b[\to \mathbb{C} \;\middle|\; \int_a^b |f(x)|^p \, dx < \infty \right\}. \tag{5.27}$$

The vector space $L^p(a,b)$ is a Banach space with respect to the norm

$$||f||_{L^p(a,b)} = \left(\int_a^b |f(x)|^p \, dx \right)^{1/p}. \tag{5.28}$$

In cases where it is clear from the context what the interval $]a,b[$ is, the norm will frequently just be denoted by $||\cdot||_p$.

5.6 Exercises

5.1 This exercise concerns the proof of Lemma 5.1.4. Consider the functions $\{f_k\}_{k=1}^{\infty}$ defined in (5.5).

(i) Make a draft of the functions f_1, f_2, and f_3.

(ii) Make a draft of the functions $f_2 - f_1$ and $f_3 - f_2$.

(iii) Show that $\{f_k\}_{k=1}^{\infty}$ is a Cauchy sequence in $C_c(\mathbb{R})$.

5.2 Consider the functions

$$f_1(x) = e^{-|x|}, \quad f_2(x) = \min(0, x^2 - 1).$$

Make a sketch of the functions f_1 and f_2, and check for each of the functions whether it belongs to the vector spaces $C_c(\mathbb{R})$, $C_0(\mathbb{R})$, or $L^1(\mathbb{R})$.

5.3 Consider the following functions defined on $\mathbb{R}$:

$$\begin{aligned} f_1(x) &= e^{-x^2}, \\ f_2(x) &= e^{-x}, \\ f_3(x) &= x^3 + 2x + 4, \\ f_4(x) &= \sin(x), \\ f_5(x) &= \sin(x)\chi_{[-2,2]}(x), \\ f_6(x) &= \sin(x)\chi_{[-2\pi,2\pi]}(x), \\ f_7(x) &= \frac{1}{1+x^2}, \\ f_8(x) &= \begin{cases} x & \text{if } x \in]-1,1], \\ 2-x & \text{if } x \in [1,3], \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

(i) Make a rough sketch of the graph of each of the functions.

(ii) Determine the support for each of the functions. Which functions have compact support?

(iii) Which functions belong to $C_0(\mathbb{R})$?

(iv) Which functions belong to $C_c(\mathbb{R})$?

(v) Which functions belong to $L^1(\mathbb{R})$?

5.4 We consider the vector space $C_c(\mathbb{R})$, equipped with the $||\cdot||_2$-norm.

(i) Show that

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

defines an inner product on $C_c(\mathbb{R})$.

(ii) Show that $C_c(\mathbb{R})$ does not form a Hilbert space with respect to the norm $||f||_2 = \sqrt{\langle f, f\rangle}$.

5.5 Verify that $L^1(\mathbb{R})$ is a vector space.

5.6 A set $\Gamma \subset \mathbb{R}$ is *countable* if its elements can be written as a list,

$$\Gamma = \{x_1, x_2, \dots\}.$$

Note that in order to show that a set Γ is countable, we need to specify a procedure guaranteeing that each element in Γ appears somewhere in the list. Show that the sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are countable.

5.7 Prove the result in Example 5.3.4.

5.8 Show that for $p \in [1, \infty[$, the expression (5.23) defines a norm on $L^p(\mathbb{R})$. (*Hint:* use Minkowski's inequality, see Theorem 1.7.1.)

5.9 This exercise relates $L^2(\mathbb{R})$ and $L^1(\mathbb{R})$.

(i) Show that $L^1(\mathbb{R})$ is not a subspace of $L^2(\mathbb{R})$ (*Hint:* find a concrete function belonging to $L^1(\mathbb{R})$ but not to $L^2(\mathbb{R})$.)

(ii) Show that $L^2(\mathbb{R})$ is not a subspace of $L^1(\mathbb{R})$ (*Hint:* find a concrete function belonging to $L^2(\mathbb{R})$ but not to $L^1(\mathbb{R})$.)

(iii) Assume that $f \in L^2(\mathbb{R})$ has compact support. Show that $f \in L^1(\mathbb{R})$; in particular, this shows that

$$L^2(\mathbb{R}) \cap C_c(\mathbb{R}) \subset L^1(\mathbb{R}).$$

5.10 This exercise relates $L^1(0,1)$ and $L^2(0,1)$.

(i) Show that $L^2(0,1) \subset L^1(0,1)$.

(ii) Show that if a sequence of functions f_k in $L^2(0,1)$ converges to 0 in $L^2(0,1)$ as $k \to \infty$, then $f_k \to 0$ in $L^1(0,1)$ as $k \to \infty$.

5.11 Consider the functions $f_k,\ k \in \mathbb{N}$, defined by

$$f_k(x) = \begin{cases} 1/k & \text{if } x \in [0,k], \\ 0 & \text{otherwise.} \end{cases}$$

(i) Make a sketch of the first few functions f_k.

(ii) Show that $f_k \to 0$ in $L^\infty(\mathbb{R})$ as $k \to \infty$.

(iii) Show that $f_k \in L^1(\mathbb{R})$ for all k.

(iv) Does it hold that

$$f_k \to 0 \text{ in } L^1(\mathbb{R}) \text{ as } k \to \infty?$$

5.12 Consider a subinterval $]a,b[\subseteq \mathbb{R}$, and functions $f, f_1, f_2, f_3, \dots$ defined on $]a,b[$. Assume that

$$||f - f_k||_{L^\infty(a,b)} \to 0 \text{ as } k \to \infty. \tag{5.29}$$

(i) Assume that the interval $]a,b[$ is finite. Show that for all $p \in [1,\infty[$, the assumption (5.29) implies that

$$||f - f_k||_{L^p(a,b)} \to 0 \text{ as } k \to \infty. \tag{5.30}$$

(ii) Assume that $]a,b[=\mathbb{R}$. Show that regardless of the choice of $p \in [1,\infty[$, the assumption (5.29) does not imply that (5.30) holds.

5.13 Consider the following functions defined on $\mathbb{R}$:

$$\begin{aligned} f_1(x) &= e^x \chi_{[0,\infty[}(x), \\ f_2(x) &= e^{-x} \chi_{[0,\infty[}(x), \\ f_3(x) &= \frac{1}{\sqrt{x}} \chi_{[1,\infty[}(x), \\ f_4(x) &= \frac{1}{\sqrt{x}} \chi_{]0,1[}(x). \end{aligned}$$

(i) Which functions belong to $L^\infty(\mathbb{R})$?

(ii) For each of the functions $f_k,\ k = 1, \dots, 4$, determine the exact range of parameters $p \in [1,\infty[$ for which $f_k \in L^p(\mathbb{R})$.

5.14 The purpose of the exercise is show how certain functions in $L^1(\mathbb{R})$ can be approximated by functions in $C_c(\mathbb{R})$.

(i) Let $f(x) := \chi_{[0,1]}(x)$. Argue that for any $\epsilon > 0$ there exists a function $g \in C_c(\mathbb{R})$ such that $||f - g||_1 \leq \epsilon$.

(ii) The function $f(x) := (1 + x^2)^{-1}$ belongs to $L^1(\mathbb{R})$ by the result in Exercise 5.3. Show that for any $\epsilon > 0$ there exists a function $g \in C_c(\mathbb{R})$ such that $||f - g||_1 \leq \epsilon$.

(iii) Argue (in words) how any bounded piecewise continuous function in $L^1(\mathbb{R})$ can be approximated by a function in $C_c(\mathbb{R})$.

5.15 This exercise relates the spaces $C_0(\mathbb{R}), C_c(\mathbb{R})$, and $L^p(\mathbb{R})$.

(i) Show that $C_c(\mathbb{R})$ is a subspace of $L^p(\mathbb{R})$ for all $p \in [1,\infty[$.

(ii) Show that $C_0(\mathbb{R})$ is not a subspace of $L^p(\mathbb{R})$ for any $p \in [1,\infty[$. (*Hint:* show first the result for $p = 1$.)

5.16 (**Weighted L^p-spaces**) Let $w : \mathbb{R} \to]0, \infty[$ be a continuous function, and define for $p \in]1, \infty[$ the vector space $L^p_w(\mathbb{R})$ by

$$L^p_w(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} \mid \int_{-\infty}^{\infty} |f(x)|^p w(x)\, dx < \infty \right\}.$$

(i) Show that

$$||f||_{L^p_w(\mathbb{R})} := \left(\int_{-\infty}^{\infty} |f(x)|^p w(x)\, dx \right)^{1/p} \tag{5.31}$$

defines a norm on $L^p_w(\mathbb{R})$.

(ii) Use the fact that $L^p(\mathbb{R})$ is a Banach space to show that $L^p_w(\mathbb{R})$ equipped with the norm in (5.31) is a Banach space.

5.17 Let $p \in [1, \infty[$ and consider the mapping

$$T : L^p(\mathbb{R}) \to L^p(\mathbb{R}),\ (Tf)(x) := f(3x + 2).$$

(i) Show that T indeed maps $L^p(\mathbb{R})$ into $L^p(\mathbb{R})$.

(ii) Show that T is linear and bounded.

(iii) Consider the function

$$f(x) := \frac{1}{x^2}\, \chi_{[1,\infty[}(x),$$

and show that $f \in L^p(\mathbb{R})$ for all $p \in [1, \infty[$.

5.18 Let $p \in [1, \infty[$ and consider the dilation operator

$$D : L^p(\mathbb{R}) \to L^p(\mathbb{R}),\ (Df)(x) = 2^{1/2} f(2x).$$

(i) Show that D actually maps $L^p(\mathbb{R})$ into $L^p(\mathbb{R})$.

(ii) Show that D is linear and bounded.

(iii) Find, as a function of p, the norm of the operator D.

5.19 Consider the mapping

$$T : L^1(0, 2) \to L^1(0, 2),\ (Tf)(x) := \int_0^x tf(t)dt.$$

(i) Show that T indeed maps $L^1(0, 2)$ into $L^1(0, 2)$.

(ii) Show that T is linear and bounded.

5.20 Let $p \in [1, \infty[$ and consider the mapping

$$T : L^p(-2,2) \to L^p(-2,2), \ (Tf)(x) := xf(x).$$

(i) Show that T indeed maps $L^p(-2,2)$ into $L^p(-2,2)$.

(ii) Show that T is linear and bounded.

(iii) Calculate the norm of the operator T.
Hint: let $f = \chi_{[2-\epsilon,2]}$ for some $\epsilon > 0$ and consider

$$\frac{||Tf||_p}{||f||_p}$$

as $\epsilon \to 2$.

5.21 Let $c > 0$ be given and consider the mapping

$$D_c : C_c(\mathbb{R}) \to C_c(\mathbb{R}), \ (D_c f)(x) := \frac{1}{\sqrt{c}} f(\frac{x}{c}), \ x \in \mathbb{R}.$$

(i) Argue that D_c actually maps $C_c(\mathbb{R})$ into $C_c(\mathbb{R})$.

(ii) Show that D_c is linear.

(iii) Show that D_c is bounded as operator from $C_c(\mathbb{R})$ into $C_c(\mathbb{R})$ (as usual, $C_c(\mathbb{R})$ is equipped with the $|| \cdot ||_\infty$-norm).

In Section 6.2 the operator D_c will be considered on $L^2(\mathbb{R})$.

6

The Hilbert Space L^2

In Chapter 5, we introduced the general $L^p(\mathbb{R})$-spaces. Among the $L^p(\mathbb{R})$-spaces, the case $p = 2$ has a very special status: $L^2(\mathbb{R})$ is a Hilbert space, and in fact the only $L^p(\mathbb{R})$-space with that property. The space $L^2(\mathbb{R})$ is discussed in Section 6.1. As a continuation and specialization of the previous sections on operators, some fundamental operators on $L^2(\mathbb{R})$ are considered in Section 6.2. The considered operators will play important roles in the later chapters on the Fourier transform and wavelets. Section 6.3 deals with the Hilbert space $L^2(a, b)$; in particular it is shown that the polynomials form a dense subspace of $L^2(a, b)$. Section 6.4 discusses Fourier expansions in the framework of the Hilbert space $L^2(-\pi, \pi)$.

6.1 The Hilbert space $L^2(\mathbb{R})$

In (5.22) we defined the Banach spaces $L^p(\mathbb{R})$ for all $p \in [1, \infty[$. The case $p = 2$ plays a very special role. We first prove that the space

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} \mid \int_{-\infty}^{\infty} |f(x)|^2 \, dx < \infty \right\}$$

can be equipped with an inner product:

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis,
DOI 10.1007/978-0-8176-4980-7_6,

Theorem 6.1.1 (The Hilbert space $L^2(\mathbb{R})$) *The vector space $L^2(\mathbb{R})$ is a Hilbert space with respect to the inner product*

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\overline{g(x)}\,dx, \;\; f, g \in L^2(\mathbb{R}). \tag{6.1}$$

Proof. The general result in Theorem 5.4.1 shows that $L^2(\mathbb{R})$ is a Banach space with respect to the norm

$$||f||_2 = \left(\int_{-\infty}^{\infty} |f(x)|^2\,dx\right)^{1/2}. \tag{6.2}$$

Thus, we only have to show that (6.1) actually defines an inner product on $L^2(\mathbb{R})$. The first step is to verify that the integral in (6.1) is well-defined, i.e., that the function

$$x \mapsto f(x)\overline{g(x)} \tag{6.3}$$

belongs to $L^1(\mathbb{R})$. Using Hölder's inequality (1.19) with $p = q = 2$, we see that for all $f, g \in L^2(\mathbb{R})$,

$$\begin{aligned}\int_{-\infty}^{\infty} |f(x)\overline{g(x)}|\,dx &\le \left(\int_{-\infty}^{\infty} |f(x)|^2\,dx\right)^{1/2}\left(\int_{-\infty}^{\infty} |g(x)|^2\,dx\right)^{1/2} \\ &< \infty.\end{aligned}$$

Thus, the function in (6.3) is in $L^1(\mathbb{R})$, and the expression in (6.1) is well-defined. The next step is to verify that $\langle \cdot, \cdot\rangle$ actually defines an inner product on $L^2(\mathbb{R})$. This is left to the reader (Exercise 6.2). □

The status of $L^2(\mathbb{R})$ is really outstanding among the $L^p(\mathbb{R})$-spaces: for none of the other spaces, the norm comes from an inner product, see Exercise 6.3.

In manipulations involving integrals of functions in $L^2(\mathbb{R})$, Cauchy–Schwarz' inequality plays a prominent role. The result in $L^2(\mathbb{R})$ is a special case of the general inequality in Theorem 4.1.2:

Theorem 6.1.2 (Cauchy–Schwarz' inequality) *For all $f, g \in L^2(\mathbb{R})$,*

$$\left|\int_{-\infty}^{\infty} f(x)\overline{g(x)}\,dx\right| \le \left(\int_{-\infty}^{\infty} |f(x)|^2\,dx\right)^{1/2}\left(\int_{-\infty}^{\infty} |g(x)|^2\,dx\right)^{1/2}.$$

When dealing with functions in $L^2(\mathbb{R})$, convergence issues shall in general be understood in the sense of that norm; thus, to say that a sequence of functions $\{g_k\}_{k=1}^{\infty}$ in $L^2(\mathbb{R})$ converges to $g \in L^2(\mathbb{R})$ means that

$$||g - g_k||_2 = \left(\int_{-\infty}^{\infty} |g(x) - g_k(x)|^2\,dx\right)^{1/2} \to 0 \text{ as } k \to \infty. \tag{6.4}$$

Convergence in L^2 is different from pointwise convergence: that is, (6.4) does not imply that $g_k(x) \to g(x)$ for all $x \in \mathbb{R}$. We will discuss this issue in the context of Fourier series in Section 6.4. However, one can prove that any convergent sequence in $L^2(\mathbb{R})$ has a *subsequence* that converges pointwise for almost all $x \in \mathbb{R}$; this fundamental result is called *Riesz' Subsequence Theorem.*

The Hilbert space $L^2(\mathbb{R})$ plays a central role in many areas of science, e.g., electrical engineering and physics. In signal processing, many time-varying signals are interpreted as functions in $L^2(\mathbb{R})$; and in quantum mechanics, the term "Hilbert space" often simply means $L^2(\mathbb{R})$. The notation used in quantum mechanics differs slightly from the one used here, so in the next example we connect the notation used in quantum mechanics with the one used here. We refer to [9] for an introduction to quantum mechanics.

Example 6.1.3 (Bracket notation) The theory for quantum mechanics is formulated using the so-called *Dirac bracket notation.* The inner product between two elements $f, g \in L^2(\mathbb{R})$ is written as

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

We note that physicists often define the inner product to be linear in the second entry rather than the first, so they put the complex conjugation on the first function f in the inner product instead of the second function g. A more significant difference between the notation used in quantum mechanics and mathematics is that physicists associate independent meaning to the symbols $\langle f|$ and $|g\rangle$; in fact, these symbols are used to denote *vectors* (that is, functions in $L^2(\mathbb{R})$). We will now explain how these symbols should be interpreted.

In quantum mechanics, the symbol $|g\rangle$ is used to denote a vector, exactly like we have used the symbol $\mathbf{v}$ in Chapter 4. The symbol $\langle f|$ is also used to denote a vector; however, now the symbol is interpreted as a *functional*, acting on a vector $|g\rangle$ by taking inner product between $\langle f|$ and $|g\rangle$. That is, the operator $P := \langle f|$, acting on $L^2(\mathbb{R})$ and taking values in $\mathbb{C}$, is given by

$$P|g\rangle := \langle f|g\rangle, \quad |g\rangle \in L^2(\mathbb{R}).$$

Recall from Riesz' representation theorem that all bounded functionals P on $L^2(\mathbb{R})$ can be written this way for an appropriate choice of the vector $\langle f|$. Thus, the bracket notation $\langle f|$ is simply a convenient way to describe all bounded functionals on $L^2(\mathbb{R})$ in a short way.

In particular, in quantum mechanical terms an orthonormal system is a collection of vectors $\{\, |g_k\rangle \,\}$ for which

$$\langle g_\ell|g_k\rangle = \begin{cases} 1 \text{ if } \ell = k, \\ 0 \text{ if } \ell \neq k. \end{cases}$$

Similarly, the symbol $P := |f\rangle\langle g|$ is used to denote the linear operator from $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$ which acts on a vector $|h\rangle$ via

$$P|h\rangle = |f\rangle\langle g|h\rangle = \langle g|h\rangle\,|f\rangle, \quad |h\rangle \in L^2(\mathbb{R});$$

that is, we obtain the image of the vector $|h\rangle$ by forming the inner product with $\langle g|$ and multiplying the resulting scalar by the vector $|f\rangle$. More generally, $P := \sum |f_k\rangle\langle g_k|$ denotes the linear operator from $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$ given by

$$P|h\rangle = \sum |f_k\rangle\langle g_k|h\rangle = \sum \langle g_k|h\rangle\,|f_k\rangle, \quad |h\rangle \in L^2(\mathbb{R}).$$

Thus, the notation

$$\sum |f_k\rangle\langle f_k| = 1 \tag{6.5}$$

simply means that for all $|h\rangle \in L^2(\mathbb{R})$,

$$|h\rangle = \sum |f_k\rangle\langle f_k|h\rangle = \sum \langle f_k|h\rangle\ |f_k\rangle. \tag{6.6}$$

In other words: if $\{\,|f_k\rangle\,\}$ is an orthonormal system and (6.5) holds, then, according to Theorem 4.7.2, $\{\,|f_k\rangle\,\}$ is an orthonormal basis for $L^2(\mathbb{R})$. □

6.2 Linear operators on $L^2(\mathbb{R})$

There are only a few natural ways of defining linear operators on general Hilbert spaces. With a concrete Hilbert space at hand it is much easier! In this section, we consider some important operators on $L^2(\mathbb{R})$.

Definition 6.2.1 (Translation, modulation, dilation) *Consider the following classes of linear operators on* $L^2(\mathbb{R})$:

(i) For $a \in \mathbb{R}$, the operator T_a, called translation by a, is defined by

$$(T_a f)(x) := f(x-a), \ x \in \mathbb{R}. \tag{6.7}$$

(ii) For $b \in \mathbb{R}$, the operator E_b, called modulation by b, is defined by

$$(E_b f)(x) := e^{2\pi i b x} f(x), \ x \in \mathbb{R}. \tag{6.8}$$

(iii) For $c > 0$, the operator D_c, called dilation by c, is defined by

$$(D_c f)(x) := \frac{1}{\sqrt{c}} f(\frac{x}{c}), \ x \in \mathbb{R}. \tag{6.9}$$

A comment about notation: we will often skip the parentheses and simply write $T_a f(x)$ instead of $(T_a f)(x)$, and similarly for the other operators. Frequently, we will also let E_b denote the function $x \mapsto e^{2\pi i b x}$; that is,

$$E_b(x) := e^{2\pi i b x}.$$

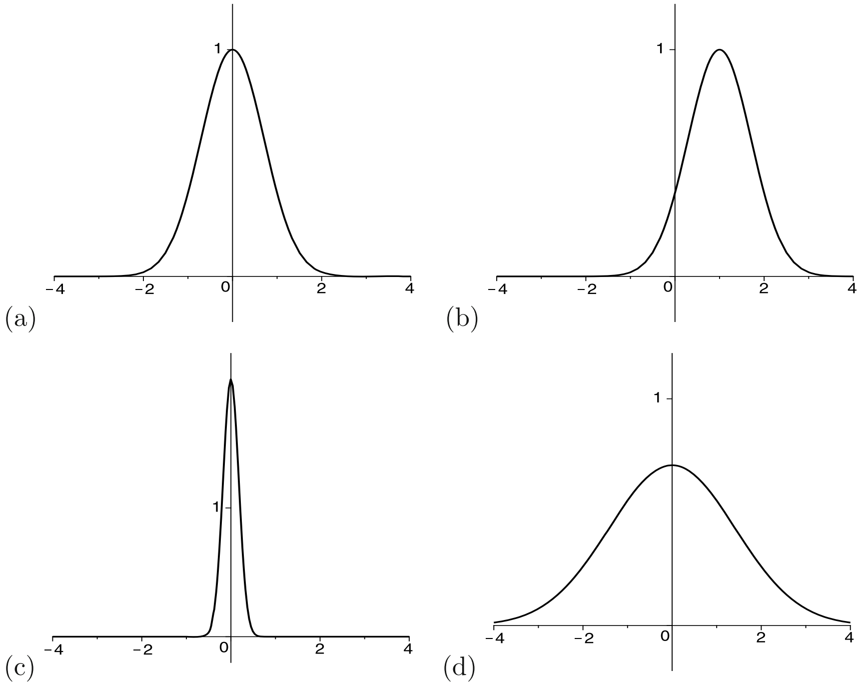

Figure 6.1. The figures illustrate the action of the operators T_a and D_c. (a) $\psi(x) = e^{-x^2}$; (b) $T_1\psi(x) = e^{-(x-1)^2}$; (c) $D_{1/4}\psi(x) = 2\psi(2^2x) = 2e^{-16x^2}$; and (d) $D_2\psi(x) = 2^{-1/2}\psi(2^{-1}x) = \frac{1}{\sqrt{2}}e^{-x^2/4}$.

Figure 6.1 illustrates how the translation operators and the dilation operators act. We collect some of the most important properties for the operators in (6.7)–(6.9):

Lemma 6.2.2 (Translation, modulation, dilation) *The operators T_a, E_b, and D_c are unitary linear operators of $L^2(\mathbb{R})$ onto $L^2(\mathbb{R})$, and the following relations hold:*

(i) $T_a^{-1} = T_{-a} = (T_a)^*$,

(ii) $E_b^{-1} = E_{-b} = (E_b)^*$,

(iii) $D_c^{-1} = D_{1/c} = (D_c)^*$.

Proof. We give a complete proof for the operator T_a and leave the cases of E_b and D_c to the reader (Exercise 6.4).

We first prove that T_a is a bounded linear operator mapping $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$. Given any $f, g \in L^2(\mathbb{R})$ and $\alpha, \beta \in \mathbb{C}$,

$$\begin{aligned} T_a(\alpha f + \beta g)(x) &= (\alpha f + \beta g)(x-a) \\ &= \alpha f(x-a) + \beta g(x-a) \\ &= \alpha T_a f(x) + \beta T_a g(x). \end{aligned}$$

This shows that

$$T_a(\alpha f + \beta g) = \alpha T_a f + \beta T_a g,$$

i.e., that T_a is linear. Also, if $f \in L^2(\mathbb{R})$, the change of variable $z = x - a$ shows that

$$\int_{-\infty}^{\infty} |T_a f(x)|^2 \, dx = \int_{-\infty}^{\infty} |f(x-a)|^2 \, dx = \int_{-\infty}^{\infty} |f(z)|^2 \, dz < \infty;$$

this proves that T_a actually maps $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$. As a consequence of this calculation, we see that

$$||T_a f||_2 = ||f||_2, \ \forall f \in L^2(\mathbb{R}),$$

i.e., T_a is bounded.

We will now prove that T_a is unitary. First, for all $f, g \in L^2(\mathbb{R})$, the change of variable $z = x - a$ yields that

$$\begin{aligned} \langle T_a f, g\rangle = \int_{-\infty}^{\infty} f(x-a)\overline{g(x)} \, dx &= \int_{-\infty}^{\infty} f(z)\overline{g(z+a)} \, dz \\ &= \langle f, T_{-a} g\rangle. \end{aligned}$$

From the definition of the adjoint operator T_a^* we know that

$$\langle T_a f, g\rangle = \langle f, T_a^* g\rangle;$$

thus, via Lemma 4.4.2, it follows that $T_a^* = T_{-a}$. Thus,

$$T_a T_a^* = T_a T_{-a} = I,$$

and

$$T_a^* T_a = T_{-a} T_a = I.$$

By definition, we conclude that T_a is unitary. Furthermore, the calculations show that

$$(T_a)^{-1} = T_a^* = T_{-a}.$$

This concludes the proof for the operator T_a. □

Operators defined by composition of some of the operators T_a, E_b, and D_c appear in several branches of mathematics and engineering. For example, the so-called Schrödinger representation essentially consists of a composition of translation operators and modulation operators; and wavelet systems, to be considered in Chapter 8, consist of scaled and translated versions of a fixed function. For analysis of such cases, the following *commutation relations* are useful (Exercise 6.5):

Lemma 6.2.3 (Commutation relations) *For all $a, b \in \mathbb{R}$ and $c > 0$, the following commutation relations hold:*

(i) $(T_a E_b f)(x) = e^{2\pi i b(x-a)} f(x-a) = e^{-2\pi i b a}(E_b T_a f)(x), \ x \in \mathbb{R}$;

(ii) $(T_a D_c f)(x) = \frac{1}{\sqrt{c}} f(\frac{x}{c} - \frac{a}{c}) = (D_c T_{a/c} f)(x), \ x \in \mathbb{R}$;

(iii) $(D_c E_b f)(x) = \frac{1}{\sqrt{c}} e^{2\pi i x b/c} f(\frac{x}{c}) = (E_{\frac{b}{c}} D_c f)(x), \ x \in \mathbb{R}$.

In wavelet analysis, to be considered in Chapter 8, the dilation operator $D_{1/2}$ plays a special role; for this reason we will simply denote this operator by D, i.e.,

$$(Df)(x) := 2^{1/2} f(2x), \ x \in \mathbb{R}. \tag{6.10}$$

This implies (Exercise 6.7) that for any $j \in \mathbb{Z}$,

$$(D^j f)(x) = 2^{j/2} f(2^j x), \ x \in \mathbb{R}. \tag{6.11}$$

With this notation, the commutation relations in Lemma 6.2.3 in particular imply the following result (Exercise 6.6):

Lemma 6.2.4 (Commutation relations) *For all $j, k \in \mathbb{Z}$, the following commutation relations hold:*

(i) $T_k D^j = D^j T_{2^j k}$;

(ii) $D^j T_k = T_{2^{-j} k} D^j$;

(iii) $(D^j)^* = D^{-j}$.

We state one more property for the operators T_a, E_b, and D_c that will turn out to be needed later:

Lemma 6.2.5 (Translation, modulation, dilation) *Let $f \in L^2(\mathbb{R})$. Then*

$$||T_a f - f||_2 \to 0 \text{ as } a \to 0.$$

Similar statements hold for the operators $E_b, b \in \mathbb{R}$, and $D_c, \ c > 0$.

Proof. We first prove the result for functions f that are continuous and have compact support, say, contained in the bounded interval $[c, d]$. We will prove that for any $\epsilon > 0$, an estimate of the form

$$||T_a f - f||_2 \leq \epsilon$$

holds if a is sufficiently close to 0. First, for $y \in]-\frac{1}{2}, \frac{1}{2}[$ the function

$$\phi(x) := (T_y f)(x) - f(x) = f(x-y) - f(x), \ x \in \mathbb{R},$$

has support in the interval $[-\frac{1}{2}+c, d+\frac{1}{2}]$. By Theorem 1.6.2, the function f is uniformly continuous; thus, for a given $\epsilon > 0$ we can find $\delta > 0$ such that

$$|f(x-y) - f(x)| \leq \epsilon \text{ for all } x \in \mathbb{R} \text{ whenever } |y| \leq \delta.$$

With this choice of δ, we obtain that for $|a| \leq \delta$,

$$\begin{aligned} ||T_a f - f||_2 &= \left(\int_{-\infty}^{\infty} |f(x-a) - f(x)|^2 \, dx \right)^{1/2} \\ &= \left(\int_{-\frac{1}{2}+c}^{\frac{1}{2}+d} |f(x-a) - f(x)|^2 \, dx \right)^{1/2} \\ &\leq \epsilon \sqrt{d-c+1}. \end{aligned}$$

This proves the result for functions $f \in C_c(\mathbb{R})$. The case of an arbitrary function $f \in L^2(\mathbb{R})$ follows by an approximation argument, using the fact that the continuous functions with compact support are dense in $L^2(\mathbb{R})$; this part of the argument is left to the reader as Exercise 6.8. The proofs of the statements for E_b and D_c are similar. □

The result in Lemma 6.2.5 is often formulated by saying that the mapping $y \mapsto T_y f$ is continuous from $\mathbb{R}$ to $L^2(\mathbb{R})$; see Exercise 6.9 for an explanation of this terminology.

6.3 The space $L^2(a, b)$

So far, we have been dealing with square-integrable functions defined on all of $\mathbb{R}$. We will now consider functions that are square-integrable on a subinterval $]a, b[\subseteq \mathbb{R}$. Let

$$L^2(a,b) := \left\{ f :]a,b[\to \mathbb{C} \;\middle|\; \int_a^b |f(x)|^2 \, dx < \infty \right\}.$$

Like in the case of $L^2(\mathbb{R})$, one can prove that $L^2(a, b)$ is a Hilbert space when equipped with the inner product

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)} \, dx, \;\; f, g \in L^2(a,b).$$

The associated norm is

$$||f||_{L^2(a,b)} = \sqrt{\int_a^b |f(x)|^2 \, dx}, \;\; f \in L^2(a,b).$$

There are two reasons to consider spaces $L^2(a, b)$ in cases where $]a, b[$ is a finite interval. First, many applications naturally lead to functions on a finite interval: for example, we can think of the function f as a measurement, related to an experiment starting at the time $x = a$ and running until $x = b$. The second reason is that the space $L^2(a,b)$ forms a convenient framework for dealing with periodic signals; in that case, a function defined on $]a,b[$ will usually be extended to a periodic function on $\mathbb{R}$.

Already in Theorem 5.4.2 we saw that the continuous functions with compact support are dense in $L^2(\mathbb{R})$. A consequence of this fact is that if $]a,b[$ is a finite interval, then the set of continuous functions on $]a,b[$ are dense in $L^2(a,b)$; see Exercise 6.13. We will use this to derive that the set of polynomials on $]a,b[$ is dense in $L^2(a,b)$.

Theorem 6.3.1 *The set of polynomials is dense in* $L^2(a,b)$.

Proof. We have to prove that for each $f \in L^2(a,b)$ and each $\epsilon > 0$ we can find a polynomial P such that

$$||f - P||_{L^2(a,b)} = \left(\int_a^b |f(x) - P(x)|^2 \, dx \right)^{1/2} \leq \epsilon.$$

In order to do so, we notice that we can extend any function $f \in L^2(a,b)$ to a function in $L^2(\mathbb{R})$ by putting $f(x) = 0$ for $x \notin [a,b]$. Now, according to Theorem 5.4.2, we can find a function $g \in C_c(\mathbb{R})$ such that

$$||f - g||_{L^2(\mathbb{R})} \leq \epsilon/2.$$

This implies that

$$\begin{aligned} ||f - g||_{L^2(a,b)} &= \sqrt{\int_a^b |f(x) - g(x)|^2 \, dx} \\ &\leq \sqrt{\int_{-\infty}^{\infty} |f(x) - g(x)|^2 \, dx} \\ &\leq \epsilon/2. \end{aligned}$$

The restriction of the function g to the interval $[a,b]$ is still a continuous function; thus, according to Theorem 2.3.4, there exists a polynomial P such that

$$|g(x) - P(x)| \leq \frac{\epsilon}{2\sqrt{b-a}}, \ \forall \ x \in [a,b]. \tag{6.12}$$

The choice of P implies that

$$\begin{aligned} ||g - P||_{L^2(a,b)} &= \sqrt{\int_a^b |g(x) - P(x)|^2 \, dx} \\ &\leq \sqrt{\int_a^b \left(\frac{\epsilon}{2\sqrt{b-a}} \right)^2 dx} \\ &\leq \epsilon/2. \end{aligned}$$

Thus, putting everything together,

$$\begin{aligned} ||f - P||_{L^2(a,b)} &= ||(f-g) + (g-P)||_{L^2(a,b)} \\ &\le ||f-g||_{L^2(a,b)} + ||g-P||_{L^2(a,b)} \\ &\le \epsilon. \end{aligned}$$

This completes the proof. □

6.4 Fourier series revisited

Fourier series are a useful tool for representation and approximation of periodic functions via trigonometric functions. We expect the reader to have a basic knowledge of Fourier series and their properties, e.g., concerning pointwise convergence. The purpose of this section is to place Fourier series in the context of the Hilbert space $L^2(-\pi,\pi)$ and clarify the exact convergence properties of Fourier series in that setting. The starting point is the fact that $L^2(-\pi,\pi)$ is a Hilbert space when equipped with the inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx, \;\; f, g \in L^2(-\pi,\pi). \tag{6.13}$$

In the context of Fourier series, one thinks about functions in $L^2(-\pi,\pi)$ as *periodic functions* with *period* 2π. That is, the functions $f \in L^2(-\pi,\pi)$ are functions on $\mathbb{R}$ for which

$$f(x+2\pi) = f(x), \;\; x \in \mathbb{R}.$$

The Fourier series of a 2π-periodic function f expands the function in terms of the trigonometric functions $\cos kx, \sin kx, \;\; k \in \mathbb{N}_0$. Formally, the Fourier series of f is defined by

$$f \sim \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx), \tag{6.14}$$

where the Fourier coefficients are given by

$$\begin{aligned} a_k &:= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\; dx, \;\; k = 0,1,2,\dots, \\ b_k &:= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)\; dx, \;\; k = 1,2,\dots. \end{aligned}$$

We note that so far, our definition of the Fourier series is purely formal: we have not shown any relationship between the function f and the associated Fourier series yet! This is the reason for the use of the symbol

"$\sim$" in (6.14). From elementary Fourier analysis we know that the question of convergence of the Fourier series is quite complicated: for example, the Fourier series might not converge pointwise for all $x \in \mathbb{R}$. We come back to the issue of convergence for the Fourier series later in this section.

All the trigonometric functions appearing in the Fourier series have period 2π. It is natural to think of a Fourier series as a decomposition of the given function f into harmonic oscillations, i.e., sine and cosine functions, with frequencies $\frac{k}{2\pi}, k = 0, 1, ...$; the size of the contributions at these frequencies are given by the Fourier coefficients a_k and b_k.

The Fourier series of a function f can be rewritten in *complex form* as

$$f \sim \sum_{k=-\infty}^{\infty} c_k e^{ikx}; \tag{6.15}$$

here, the coefficients c_k are given by

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} \, dx.$$

For later use we remind the reader about *Parseval's equation:*

Lemma 6.4.1 (Parseval's equation) *If the function $f \in L^2(\mathbb{R})$ has the Fourier coefficients c_k, $k \in \mathbb{Z}$, then*

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx = \sum_{k \in \mathbb{Z}} |c_k|^2.$$

We will now show that the expression in (6.15) actually is the expansion of $f \in L^2(-\pi, \pi)$ in terms of a certain orthonormal basis for $L^2(\mathbb{R})$. In the proof we use the Nth *partial sum* of the Fourier series, given by

$$S_N(x) = \sum_{k=-N}^{N} c_k e^{ikx}.$$

Theorem 6.4.2 (Orthonormal basis for $L^2(-\pi, \pi)$) *The functions*

$$\left\{ \frac{1}{\sqrt{2\pi}} e^{ikx} \right\}_{k \in \mathbb{Z}} \tag{6.16}$$

form an orthonormal basis for $L^2(-\pi, \pi)$.

Proof. A complete proof is technical, so we will not present all details. However, we will provide the key steps. The idea is to show that

(i) The functions in (6.16) form an orthonormal system in $L^2(-\pi, \pi)$;

(ii) The functions in (6.16) are dense in $L^2(-\pi, \pi)$.

After proving (i) and (ii), the result follows from the equivalence between (i) and (v) in Theorem 4.7.2.

In order to show that the functions in (6.16) form an orthonormal system in $L^2(-\pi,\pi)$, we need to show that

$$\int_{-\pi}^{\pi} e^{imx}e^{-ikx}\,dx = 0 \text{ if } k \neq m, \tag{6.17}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |e^{-ikx}|^2\,dx = 1 \text{ for all } k \in \mathbb{Z}. \tag{6.18}$$

This is left to the reader (Exercise 6.14).

In order to show that the functions in (6.16) are dense in $L^2(-\pi,\pi)$, we must show that for any function $f \in L^2(-\pi,\pi)$ and any $\epsilon > 0$, there exists a function $g \in \text{span}\left\{\frac{1}{\sqrt{2\pi}}e^{ikx}\right\}_{k\in\mathbb{Z}}$ such that

$$||f-g||_{L^2(-\pi,\pi)} \leq \epsilon.$$

One can prove that the 2π-periodic, continuous and *piecewise differentiable functions* are dense in $L^2(-\pi,\pi)$; note that this result is stronger than the fact that the continuous functions are dense in $L^2(-\pi,\pi)$. For any 2π-periodic, continuous and piecewise differentiable function h, it is known that the Fourier series converges uniformly to the function; that is, given any $\epsilon > 0$, we can find an $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$,

$$|h(x) - S_N(x)| \leq \frac{\epsilon}{\sqrt{8\pi}} \text{ for all } x \in]-\pi,\pi[.$$

It follows that for all $N \geq N_0$,

$$\int_{-\pi}^{\pi} |h(x) - S_N(x)|^2\,dx \leq \int_{-\pi}^{\pi} \frac{\epsilon^2}{8\pi}\,dx = \frac{\epsilon^2}{4}.$$

This proves that

$$||h - S_N||_{L^2(-\pi,\pi)} \leq \frac{\epsilon}{2}.$$

Since $S_N \in \text{span}\left\{\frac{1}{\sqrt{2\pi}}e^{ikx}\right\}_{k\in\mathbb{Z}}$, we conclude that any continuous and piecewise differentiable function on $]-\pi,\pi[$ can be approximated arbitrarily well by a function in $\text{span}\left\{\frac{1}{\sqrt{2\pi}}e^{ikx}\right\}_{k\in\mathbb{Z}}$. Now, take an arbitrary function $f \in L^2(-\pi,\pi)$, and let $\epsilon > 0$ be given. Then we can find a continuous and piecewise differentiable function h such that $||f-h||_{L^2(-\pi,\pi)} \leq \epsilon/2$. By the result we just proved, we can find a trigonometric polynomial S_N such that $||h - S_N||_{L^2(-\pi,\pi)} \leq \epsilon/2$. Putting these results together, we arrive at

$$\begin{aligned} ||f - S_N||_{L^2(-\pi,\pi)} &= ||(f-h) + (h - S_N)||_{L^2(-\pi,\pi)} \\ &\leq ||f-h||_{L^2(-\pi,\pi)} + ||h - S_N||_{L^2(-\pi,\pi)} \\ &\leq \epsilon. \end{aligned}$$

This implies that the complex exponentials in (6.16) are dense in $L^2(-\pi,\pi)$, and completes the sketch of the proof. □

For notational convenience, define the functions $\{\mathbf{e}_k\}_{k\in\mathbb{Z}}$ by

$$\mathbf{e}_k(x) = e^{ikx}, \ x \in \mathbb{R}.$$

With this notation, the orthonormal basis in (6.16) can be written as $\{\frac{1}{\sqrt{2\pi}}\mathbf{e}_k\}_{k\in\mathbb{Z}}$. Using (4.26) on this orthonormal basis leads to

$$\begin{aligned} f &= \sum_{k\in\mathbb{Z}} \langle f, \frac{1}{\sqrt{2\pi}}\mathbf{e}_k\rangle \frac{1}{\sqrt{2\pi}}\mathbf{e}_k \\ &= \sum_{k\in\mathbb{Z}} c_k \mathbf{e}_k, \end{aligned} \tag{6.19}$$

where

$$c_k := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}\,dx.$$

This is exactly the expression that is used as definition of the Fourier series in complex form! This shows that the Fourier series for a function f actually is the expansion of f in terms of the orthonormal basis (6.16). This coincidence explains the troubles one has with pointwise convergence of Fourier series: viewed as a Hilbert space-identity, the exact meaning of the identity (6.19) is that

$$\left\| f - \sum_{k=-N}^{N} c_k e^{ikx} \right\|_{L^2(-\pi,\pi)} \to 0 \text{ as } N \to \infty,$$

i.e., that

$$\int_{-\pi}^{\pi} \left| f(x) - \sum_{k=-N}^{N} c_k e^{ikx} \right|^2 dx \to 0 \text{ as } N \to \infty; \tag{6.20}$$

this is different from requiring pointwise convergence of the Fourier series to $f(x)$.

We will frequently need Fourier series on other intervals than $]-\pi,\pi[$. For later reference we state the Fourier series explicitly for 1-periodic functions:

Example 6.4.3 (Orthonormal basis for $L^2(0,1)$) By a scaling of the interval, the reader can check that the functions $\{e^{2\pi ikx}\}_{k\in\mathbb{Z}}$ form an orthonormal basis for $L^2(-1/2,1/2)$ and for $L^2(0,1)$. Writing

$$\mathbf{e}_k(x) := e^{2\pi ikx},$$

the Fourier series of $f \in L^2(0,1)$ is

$$f \sim \sum_{k\in\mathbb{Z}} c_k \mathbf{e}_k,$$

where the Fourier coefficients are

$$c_k = \int_{-1/2}^{1/2} f(x)e^{-2\pi ikx}\,dx = \int_0^1 f(x)e^{-2\pi ikx}\,dx.$$

The Fourier series converges toward f in the sense that

$$\int_0^1 \left| f(x) - \sum_{k=-N}^{N} c_k e^{2\pi ikx} \right|^2 dx \to 0 \text{ as } N \to \infty. \tag{6.21}$$

□

6.5 Exercises

6.1 We continue the analysis of the vector space $C[a,b]$ considered in Exercise 1.3 and Example 2.1.4. For convenience, let $a = 0,\ b = 2$.

(i) Show that

$$\langle f, g\rangle := \int_0^2 f(x)\overline{g(x)}\,dx$$

defines an inner product on $C[0,2]$.

We now consider the norm associated with the inner product in (i), i.e.,

$$||f|| = \sqrt{\int_0^2 |f(x)|^2\,dx}.$$

We will show that $C[0,2]$ equipped with this norm does not satisfy the condition for being a Hilbert space. Consider the functions

$$f_n(x) = \min(x^n, 1),\ n \in \mathbb{N}, \text{ and } f(x) = \begin{cases} 0 \text{ if } x \in [0,1[, \\ 1 \text{ if } x \in [1,2]. \end{cases}$$

(ii) Show that $||f - f_n|| \to 0$ as $n \to \infty$.

(iii) Conclude that $C[0,2]$ equipped with the norm $||\cdot||$ cannot be a Hilbert space.

6.2 Complete the proof of Theorem 6.1.1, showing that the expression (6.1) defines an inner product on $L^2(\mathbb{R})$.

6.3 Consider the Banach space $L^p(\mathbb{R})$ for some $p \geq 1$, and the functions

$$f = \chi_{[0,1[}, \; g = \chi_{[1,2[}.$$

(i) Calculate the numbers

$$||f||_p \; , \; ||g||_p \; , \; ||f+g||_p \; , \; ||f-g||_p.$$

(ii) Use Theorem 4.1.4 to conclude that $L^p(\mathbb{R})$ does not form a Hilbert space for $p \neq 2$.

6.4 Prove Lemma 6.2.2 for the operators $E_b, \; b \in \mathbb{R}$, and D_c, $c > 0$.

6.5 Prove Lemma 6.2.3.

6.6 Prove Lemma 6.2.4.

6.7 Consider the operator D introduced in (6.10). The purpose of the exercise is to give a rigorous proof of the formula

$$(D^j f)(x) = 2^{j/2} f(2^j x), \; j \in \mathbb{Z}. \tag{6.22}$$

(i) Show that (6.22) holds for $j \in \mathbb{N}$.

(ii) For any bounded operator T on a Hilbert space, one define

$$T^0 := I,$$

where I is the identity operator. Show that this is in accordance with (6.22) for $j = 0$.

(iii) Show that the inverse operator D^{-1} is given by the expression in (6.22) with $j = -1$.

(iv) For any invertible operator T on a Hilbert space and any $k \in \mathbb{N}$, the operator T^{-k} is defined by $T^{-k} := (T^{-1})^k$. Show that this definition leads to (6.22) for $j \in \{-1, -2, \dots\}$.

6.8 Complete the proof of Lemma 6.2.5 by showing the result for general functions $f \in L^2(\mathbb{R})$.

6.9 Let $f \in L^2(\mathbb{R})$. Show that if $\{y_n\}_{n=1}^\infty$ is a sequence of real numbers that converges to $y \in \mathbb{R}$, then

$$||T_{y_n} f - T_y f||_2 \to 0 \text{ as } n \to \infty.$$

In analogy with the wording used for functions, the result is expressed by saying that the function $y \mapsto T_y f$ is continuous from $\mathbb{R}$ into $L^2(\mathbb{R})$.

6.10 Consider mapping

$$U : L^2(\mathbb{R}) \to L^2(\mathbb{R}),\ (Uf)(x) = f(2x - 2),\ x \in \mathbb{R}.$$

(i) Show that U indeed maps $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$.

(ii) Show that U is linear and bounded.

(iii) Compute the adjoint operator U^*.

6.11 Consider the linear mapping

$$(Tf)(x) := xf(x).$$

Show that T does not map $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$.

6.12 (**Weighted L^2-spaces**) Let $r : \mathbb{R} \to]0, \infty[$ be a continuous function, and define the vector space $L_r^2(\mathbb{R})$ by

$$L_r^2(\mathbb{R}) := \left\{ f : \mathbb{R} \to \mathbb{C} \mid \int_{-\infty}^{\infty} |f(x)|^2 r(x)\, dx < \infty \right\}.$$

(i) Show that

$$\langle f, g \rangle_{L_r^2(\mathbb{R})} := \int_{-\infty}^{\infty} f(x)\overline{g(x)} r(x)\, dx \tag{6.23}$$

defines an inner product on $L_r^2(\mathbb{R})$.

(ii) Using that $L^2(\mathbb{R})$ is a Hilbert space, show that $L_r^2(\mathbb{R})$ equipped with the inner product in (6.23) is a Hilbert space.

6.13 Show that if $]a, b[$ is a finite interval, then the set of continuous functions on $]a, b[$ is dense in $L^2(a, b)$.

6.14 Prove (6.17) and (6.18).

6.15 For $f \in L^2(-\pi, \pi)$, the complex Fourier coefficients are defined by

$$c_k := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-ikx}\, dx, \;\; k \in \mathbb{Z}.$$

Show that the integral defining c_k is well defined for $f \in L^2(-\pi, \pi)$, i.e., that

$$\int_{-\pi}^{\pi} \left| f(x)e^{-ikx} \right| dx < \infty.$$

7
The Fourier Transform

The Fourier transform is one of the main tools for analyzing functions in $L^2(\mathbb{R})$. It appears in all contexts where one wants to extract the frequencies appearing in a given signal. The definition and main properties of the Fourier transform of functions in $L^1(\mathbb{R})$ are considered in Section 7.1. An extension of the Fourier transform to a unitary operator on $L^2(\mathbb{R})$ is discussed in Section 7.2. Convolution and its interplay with the Fourier transform is described in Section 7.3. Section 7.4 introduces the sampling problem and the Paley–Wiener space. In particular, it is shown how to recover arbitrary functions in the Paley–Wiener space based on their function values on the discrete set $\mathbb{Z}$. Finally, we relate the Fourier transform to the discrete Fourier transform in Section 7.5.

7.1 The Fourier transform on $L^1(\mathbb{R})$

We have already seen that Fourier series are useful for representation and approximation of periodic functions via trigonometric functions. For the case of $L^2(-\pi,\pi)$, the Fourier series is an infinite sum of functions $\sin nx$ and $\cos nx$, $n \in \mathbb{N}_0$, i.e., oscillations with frequencies $\frac{n}{2\pi}$. For aperiodic functions we need to search for other methods. The classical tool is the Fourier transform, which we introduce here.

If we want to expand aperiodic functions, the situation is more complicated than for periodic functions. All frequencies can appear in the signal, and the Fourier series must be replaced by an integral over all frequencies.

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis,
DOI 10.1007/978-0-8176-4980-7_7,

The so-called inversion formula, to be presented later in this section, explains how to represent aperiodic functions in terms of their frequency content. The frequency content itself is described via the Fourier transform:

Definition 7.1.1 (Fourier transform) *The Fourier transform associates to each function $f \in L^1(\mathbb{R})$ a new function $\hat{f} : \mathbb{R} \to \mathbb{C}$ given by*

$$\hat{f}(\gamma) := \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \gamma} \, dx, \ \gamma \in \mathbb{R}. \tag{7.1}$$

The Fourier transform of f is also denoted by

$$(\mathcal{F}f)(\gamma) := \hat{f}(\gamma). \tag{7.2}$$

Note that the Fourier transform is well defined: due to the assumption $f \in L^1(\mathbb{R})$, Lemma 1.7.2 implies that

$$\int_{-\infty}^{\infty} \left| f(x) e^{-2\pi i x \gamma} \right| \, dx = \int_{-\infty}^{\infty} |f(x)| \, dx < \infty.$$

The notation (7.2) indicates that we can look at the Fourier transform as an *operator* $\mathcal{F}$ that maps the function f to the function $\hat{f}$. For the moment we do not specify the range for the operator $\mathcal{F}$; we come back to this issue in Corollary 7.1.6. It is clear that the operator $\mathcal{F}$ is linear. So far, the Fourier transform is defined on $L^1(\mathbb{R})$; later, we will see that it has an extension to a bounded operator that maps $L^2(\mathbb{R})$ onto $L^2(\mathbb{R})$.

As already mentioned, the Fourier transform contains information about the frequency content of the function f. We will demonstrate this in Example 7.1.4, based on a chain of rules for calculations with the Fourier transform to be presented now. All these rules are based on the translation operators and modulation operators

$$T_a f(x) = f(x-a), \quad E_b f(x) = e^{2\pi i b x} f(x), \ x \in \mathbb{R},$$

considered in Section 6.2. We ask the reader to provide the proofs of the following result in Exercise 7.4:

Theorem 7.1.2 (Calculations with the Fourier transform) *Given $f \in L^1(\mathbb{R})$, the following hold:*

(i) If f is an even function, then

$$\hat{f}(\gamma) = 2 \int_0^{\infty} f(x) \cos(2\pi x \gamma) dx.$$

(ii) If f is an odd function, then

$$\hat{f}(\gamma) = -2i \int_0^{\infty} f(x) \sin(2\pi x \gamma) dx.$$

(iii) The Fourier transform of the function $T_a f,\ a \in \mathbb{R}$, *is*

$$(\mathcal{F}T_a f)(\gamma) = \hat{f}(\gamma)e^{-2\pi i a\gamma} = E_{-a}\hat{f}(\gamma).$$

In operator language this is expressed as $\mathcal{F}T_a = E_{-a}\mathcal{F}$.

(iv) The Fourier transform of the function $E_b f,\ b \in \mathbb{R}$, *is*

$$(\mathcal{F}E_b f)(\gamma) = \hat{f}(\gamma - b).$$

In operator language this is expressed as $\mathcal{F}E_b = T_b\mathcal{F}$.

(v) If $f(x) \to 0$ *as* $x \to \pm\infty$, f *is differentiable, and* $f' \in L^1(\mathbb{R})$, *then*

$$(\mathcal{F}f')(\gamma) = 2\pi i\gamma\hat{f}(\gamma).$$

Note in particular the rules (iii) and (iv); in words rather than symbols, they say that

- taking the Fourier transform of a translated version of f is done by multiplying $\hat{f}$ with a complex exponential function;
- taking the Fourier transform of a function f which is multiplied with a complex exponential function, corresponds to a translation of $\hat{f}$.

We will also need a commutation relation for the Fourier transform and the dilation operators D_c and D, introduced in (6.9) and (6.10). We leave the proof to the reader (Exercise 7.5).

Lemma 7.1.3 (Commutation relation) *For each* $c \neq 0$,

$$\mathcal{F}D_c = D_{1/c}\mathcal{F};$$

in particular,

$$\mathcal{F}D = D^{-1}\mathcal{F}.$$

Let us show how we can use some of these rules to find the Fourier transform of a cosine function on an interval:

Example 7.1.4 (Fourier transform of oscillation) Given constants $a, \omega > 0$, we want to calculate the Fourier transform of the function

$$f(x) = \cos(2\pi\omega x)\chi_{[-\frac{a}{2},\frac{a}{2}]}(x). \tag{7.3}$$

This signal corresponds to an oscillation which starts at the time $x = -a/2$ and lasts till $x = a/2$. If x is measured in seconds, we have ω oscillations per second, i.e., the frequency is $\nu = \omega$. Note that we can write

$$\begin{aligned} f(x) &= \frac{e^{2\pi i\omega x} + e^{-2\pi i\omega x}}{2}\chi_{[-\frac{a}{2},\frac{a}{2}]}(x) \\ &= \frac{1}{2}e^{2\pi i\omega x}\chi_{[-\frac{a}{2},\frac{a}{2}]}(x) + \frac{1}{2}e^{-2\pi i\omega x}\chi_{[-\frac{a}{2},\frac{a}{2}]}(x) \\ &= \frac{1}{2}E_\omega\chi_{[-\frac{a}{2},\frac{a}{2}]}(x) + \frac{1}{2}E_{-\omega}\chi_{[-\frac{a}{2},\frac{a}{2}]}(x). \end{aligned}$$

That is, in order to find $\hat{f}$, it is enough to calculate the Fourier transform of the function $\chi_{[-\frac{a}{2},\frac{a}{2}]}$ and apply Theorem 7.1.2(iv). Now, because $\chi_{[-\frac{a}{2},\frac{a}{2}]}$ is an even function, Theorem 7.1.2(i) shows that for $\gamma \neq 0$,

$$\begin{aligned}
\mathcal{F}\chi_{[-\frac{a}{2},\frac{a}{2}]}(\gamma) &= 2\int_0^\infty \chi_{[-\frac{a}{2},\frac{a}{2}]}(x)\cos(2\pi x\gamma)\,dx \\
&= 2\int_0^{\frac{a}{2}} \cos(2\pi x\gamma)\,dx \\
&= \frac{2}{2\pi\gamma}\left[\sin(2\pi x\gamma)\right]_{x=0}^{x=\frac{a}{2}} \\
&= \frac{\sin \pi a\gamma}{\pi\gamma}.
\end{aligned}$$

Via Theorem 7.1.2(iv) it follows that

$$\begin{aligned}
\hat{f}(\gamma) &= \mathcal{F}\left(\frac{1}{2}E_\omega\chi_{[-\frac{a}{2},\frac{a}{2}]} + \frac{1}{2}E_{-\omega}\chi_{[-\frac{a}{2},\frac{a}{2}]}\right)(\gamma) \\
&= \frac{1}{2}\mathcal{F}\chi_{[-\frac{a}{2},\frac{a}{2}]}(\gamma-\omega) + \frac{1}{2}\mathcal{F}\chi_{[-\frac{a}{2},\frac{a}{2}]}(\gamma+\omega) \\
&= \frac{1}{2}\left(\frac{\sin(\pi a(\gamma-\omega))}{\pi(\gamma-\omega)} + \frac{\sin(\pi a(\gamma+\omega))}{\pi(\gamma+\omega)}\right).
\end{aligned}$$

Figure 7.1 shows the function $\hat{f}$ for $\omega = 10$ and different values of a. A larger value of a corresponds to the oscillation $\cos(2\pi\omega x)$ being present in the signal over a larger time interval; we see that this increases the peak of $\hat{f}$ at the frequency $\gamma = \nu = 10$.

The figures show that other frequencies than just $\gamma = 10$ appear in the signal. This is perhaps surprising, because the cosine function by itself only contains the frequency $\gamma = 10$. The explanation is that the multiplication with the characteristic function in (7.3) introduces other frequencies in the signal as well. □

We will now describe some of the important properties of the Fourier transform. The first is called *Riemann–Lebesgue's Lemma:*

Theorem 7.1.5 (Riemann–Lebesgue's lemma) *For $f \in L^1(\mathbb{R})$, $\hat{f}$ is a continuous function which tends to zero as $\gamma \to \pm\infty$; that is, $\hat{f} \in C_0(\mathbb{R})$.*

Proof. Let $\gamma \in \mathbb{R}$. Using Lemma 1.7.2, for any $\delta \in \mathbb{R}$ we have that

$$\begin{aligned}
|\hat{f}(\gamma+\delta) - \hat{f}(\gamma)| &= \left|\int_{-\infty}^\infty f(x)\left(e^{-2\pi ix(\gamma+\delta)} - e^{-2\pi ix\gamma}\right)dx\right| \\
&\leq \int_{-\infty}^\infty \left|f(x)\left(e^{-2\pi ix(\gamma+\delta)} - e^{-2\pi ix\gamma}\right)\right| dx \\
&= \int_{-\infty}^\infty |f(x)|\left|e^{-2\pi ix\delta} - 1\right| dx. \qquad (7.4)
\end{aligned}$$

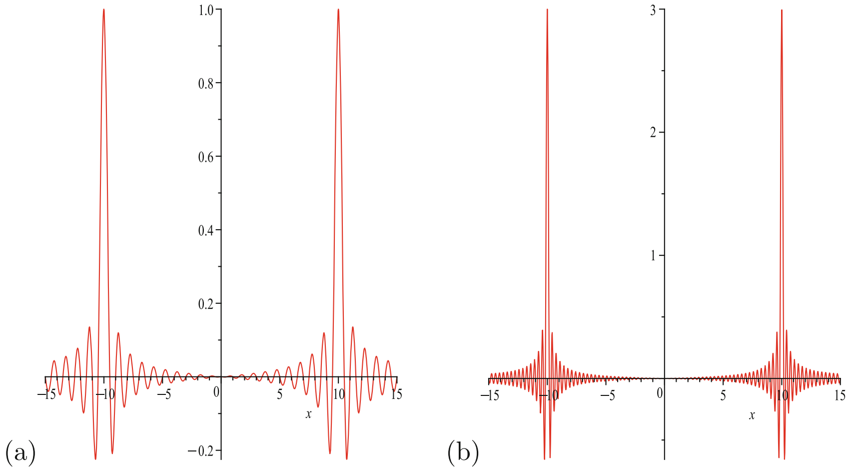

Figure 7.1. (a): The Fourier transform of the function f in (7.3) for $\omega = 10, a = 2$, corresponding to a signal with frequency $\nu = 10$ which is present during a time interval of length 2. (b): The Fourier transform of the function f in (7.3) for $\omega = 10, a = 6$, corresponding to a signal with frequency $\nu = 10$ which is present during a time interval of length 6.

We want to show that

$$\hat{f}(\gamma + \delta) \to \hat{f}(\gamma) \text{ as } \delta \to \infty.$$

For that purpose, consider the functions

$$f_\delta(x) = |f(x)| \left| e^{-2\pi i x \delta} - 1 \right|.$$

Note that

$$f_\delta(x) \to 0 \text{ as } \delta \to 0,$$

and that for all $\delta > 0$,

$$|f_\delta(x)| \leq 2\,|f(x)|.$$

By Theorem 5.3.6 this implies that

$$\lim_{\delta \to 0} \int_{-\infty}^{\infty} |f(x)| \left| e^{-2\pi i x \delta} - 1 \right| dx = \int_{-\infty}^{\infty} 0\, dx = 0.$$

Using (7.4), we conclude that

$$\hat{f}(\gamma + \delta) - \hat{f}(\gamma) \to 0 \text{ as } \delta \to 0,$$

i.e., the function $\hat{f}$ is continuous at the arbitrarily chosen point $\gamma \in \mathbb{R}$.

In order to prove the second part of the result, let $\gamma \in \mathbb{R} \setminus \{0\}$. We first perform the change of variable $y = x - \frac{1}{2\gamma}$ in the definition of the Fourier transform:

$$\begin{aligned}\hat{f}(\gamma) &= \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\gamma}\,dx\\ &= \int_{-\infty}^{\infty} f(y+\frac{1}{2\gamma})e^{-2\pi i(y+\frac{1}{2\gamma})\gamma}\,dy\\ &= e^{-\pi i}\int_{-\infty}^{\infty} f(y+\frac{1}{2\gamma})e^{-2\pi i y\gamma}\,dy\\ &= -\int_{-\infty}^{\infty} f(y+\frac{1}{2\gamma})e^{-2\pi i y\gamma}\,dy.\end{aligned}$$

Combined with the definition of $\hat{f}(\gamma)$, the above calculation yields that

$$\begin{aligned}\hat{f}(\gamma) &= \frac{1}{2}\big(\hat{f}(\gamma)+\hat{f}(\gamma)\big)\\ &= \frac{1}{2}\left(\int_{-\infty}^{\infty} f(y)e^{-2\pi i y\gamma}\,dy - \int_{-\infty}^{\infty} f(y+\frac{1}{2\gamma})e^{-2\pi i y\gamma}\,dy\right)\\ &= \frac{1}{2}\int_{-\infty}^{\infty}\left(f(y)-f(y+\frac{1}{2\gamma})\right)e^{-2\pi i y\gamma}\,dy.\end{aligned}$$

Via Lemma 1.7.2 this implies that

$$\begin{aligned}|\hat{f}(\gamma)| &\le \frac{1}{2}\int_{-\infty}^{\infty}\left|f(y)-f(y+\frac{1}{2\gamma})\right|\,dy\\ &= \frac{1}{2}\,||f-T_{\frac{1}{2\gamma}}f||_1. \qquad (7.5)\end{aligned}$$

Exactly as in the proof of Lemma 6.2.5, but with the L^1-norm instead of the L^2-norm, one can show that

$$||f-T_x f||_1 \to 0 \text{ as } x\to 0;$$

we leave this part of the argument to the reader as Exercise 7.6. Via (7.5) we now conclude that $\hat{f}(\gamma)\to 0$ as $\gamma\to\pm\infty$, as desired. □

Note that Theorem 7.1.5 immediately shows that the Fourier transform can be considered as an operator mapping $L^1(\mathbb{R})$ into $C_0(\mathbb{R})$:

Corollary 7.1.6 *The Fourier transform can be considered as a bounded linear operator,*

$$\mathcal{F}: L^1(\mathbb{R})\to C_0(\mathbb{R}),$$

and

$$||\mathcal{F}f||_\infty \le ||f||_1,\ f\in L^1(\mathbb{R}). \qquad (7.6)$$

Proof. That the Fourier transform maps $L^1(\mathbb{R})$ into $C_0(\mathbb{R})$ follows from Theorem 7.1.5. The norm estimate (7.6) is a consequence of the definition of the Fourier transform and Lemma 1.7.2. □

It turns out that under certain assumptions, the values of the Fourier transform $\hat{f}$ contain enough information to *reconstruct* the function f: that is, based on knowledge of the Fourier transform $\hat{f}$ we can determine the function f. In operator language, this says that the Fourier transform is invertible. The result is known as the *inversion formula.* We state a version of this result here, and outline a proof in Appendix A.2:

Theorem 7.1.7 (Inversion formula for $L^1(\mathbb{R})$-functions) *Assume that $f \in L^1(\mathbb{R})$ and that also $\hat{f} \in L^1(\mathbb{R})$. Then*

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\gamma) e^{2\pi i x \gamma} d\gamma \quad \text{for almost all } x \in \mathbb{R}. \tag{7.7}$$

If f is continuous, the formula (7.7) holds pointwise for all $x \in \mathbb{R}$.

Let us discuss the role of the Fourier transform in a concrete case:

Example 7.1.8 The current running in the loudspeaker cable as a recording is played can be considered as a function f. If the playback starts at the time $x = 0$ and lasts till $x = a$, we can consider f as a continuous function with compact support on $[0, a]$. The function f represents the recording in the *time-domain:* by looking, e.g., at the graph of the function f we can see how the signal changes as a function of the time, but we cannot immediately see which frequencies are present. Look, e.g., at the speech signal depicted in the prologue.

Looking at the Fourier transform $\hat{f}$, we obtain information about the frequencies appearing in the signal. However, $\hat{f}$ does not tell at what time the frequencies appear! Look, e.g., at the figure in the prologue that shows the absolute value of the Fourier transform of the speech signal. In that sense, the Fourier transform yields a representation in the *frequency-domain.*

The (surprising) content of the inversion formula is that we can reconstruct the signal f based on information on the Fourier transform. That is, just by knowing the frequency content, we can recover the piece of music that is played.

Note, however, that in order for the inversion formula to apply, we need knowledge of $\hat{f}(\gamma)$: it is not enough to know the amplitude of the frequencies, i.e., the function $|\hat{f}(\gamma)|$. The interpretation of this is that it is the *phase* of the Fourier transform that contains the information about what time the various frequencies appear. □

7.2 The Fourier transform on $L^2(\mathbb{R})$

So far, we have only defined the Fourier transform of functions belonging to $L^1(\mathbb{R})$. It turns out that the Fourier transform can be considered on several other spaces as well. We limit ourselves to a discussion of an important extension to the Hilbert space $L^2(\mathbb{R})$.

We begin with a lemma concerning the Fourier transformation on the subspace $C_c(\mathbb{R})$ of $L^1(\mathbb{R})$.

Lemma 7.2.1 (The Fourier transform on $C_c(\mathbb{R})$**)** *For any* $f \in C_c(\mathbb{R})$,

$$\int_{-\infty}^{\infty} |\hat{f}(\gamma)|^2 \, d\gamma = \int_{-\infty}^{\infty} |f(x)|^2 \, dx. \tag{7.8}$$

Proof. The starting point is Parseval's equation for Fourier series, see Lemma 6.4.1. Let us first assume that the considered function f is supported on $[-\pi, \pi]$. Then the Fourier coefficients on complex form for f are

$$\begin{aligned} c_k &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}\, dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{-ikx}\, dx \\ &= \frac{1}{2\pi}\hat{f}(\frac{k}{2\pi}). \end{aligned}$$

Thus, by Lemma 6.4.1,

$$\begin{aligned} \int_{-\infty}^{\infty} |f(x)|^2 \, dx = \int_{-\pi}^{\pi} |f(x)|^2 \, dx &= 2\pi \sum_{k\in\mathbb{Z}} |c_k|^2 \\ &= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \left|\hat{f}(\frac{k}{2\pi})\right|^2. \end{aligned} \tag{7.9}$$

Let E_b denote the modulation operator, see (6.8). Then $||f||_2 = ||E_b f||_2$. Using (7.9) with f replaced by $E_b f$ for some $b \in \mathbb{R}$, Theorem 7.1.2 gives that

$$\begin{aligned} \int_{-\infty}^{\infty} |f(x)|^2 \, dx &= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \left|\mathcal{F}E_b f(\frac{k}{2\pi})\right|^2 \\ &= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \left|\hat{f}(\frac{k}{2\pi} - b)\right|^2. \end{aligned} \tag{7.10}$$

Note that

$$\int_{-\infty}^{\infty} |f(x)|^2 \, dx = \int_0^{(2\pi)^{-1}} 2\pi \left(\int_{-\infty}^{\infty} |f(x)|^2 \, dx\right) db;$$

applying (7.10) to the inner integral followed by an interchange of the integral and the sum (justified by the result in Example 5.3.4) yields that

$$\begin{aligned}\int_{-\infty}^{\infty} |f(x)|^2\, dx &= \int_0^{(2\pi)^{-1}} \sum_{k\in\mathbb{Z}} \left|\hat{f}(\frac{k}{2\pi}-b)\right|^2 db \\ &= \sum_{k\in\mathbb{Z}} \int_0^{(2\pi)^{-1}} \left|\hat{f}(\frac{k}{2\pi}-b)\right|^2 db. \qquad (7.11)\end{aligned}$$

The change of variable $\gamma = \frac{k}{2\pi} - b$ shows that

$$\begin{aligned}\int_0^{(2\pi)^{-1}} \left|\hat{f}(\frac{k}{2\pi}-b)\right|^2 db &= -\int_{k(2\pi)^{-1}}^{(k-1)(2\pi)^{-1}} \left|\hat{f}(\gamma)\right|^2 d\gamma \\ &= \int_{(k-1)(2\pi)^{-1}}^{k(2\pi)^{-1}} \left|\hat{f}(\gamma)\right|^2 d\gamma.\end{aligned}$$

Inserting this in (7.11), we finally arrive at

$$\begin{aligned}\int_{-\infty}^{\infty} |f(x)|^2\, dx &= \sum_{k\in\mathbb{Z}} \int_{(k-1)(2\pi)^{-1}}^{k(2\pi)^{-1}} \left|\hat{f}(\gamma)\right|^2 d\gamma \\ &= \int_{-\infty}^{\infty} |\hat{f}(\gamma)|^2\, d\gamma\end{aligned}$$

as desired. The proof of Lemma 7.2.1 for the case where f is supported outside $[-\pi, \pi]$ is left to the reader (Exercise 7.7). □

Lemma 7.2.1 shows that if we equip $C_c(\mathbb{R})$ with the $L^2(\mathbb{R})$-norm, the Fourier transform is an isometry from $C_c(\mathbb{R})$ into $L^2(\mathbb{R})$. Using that the vector space $C_c(\mathbb{R})$ is dense in $L^2(\mathbb{R})$ we can prove the following important result:

Theorem 7.2.2 (Fourier transform as unitary operator) *The Fourier transform can be extended to a unitary mapping of $L^2(\mathbb{R})$ onto $L^2(\mathbb{R})$. In particular, this extension (also called the Fourier transform and denoted by $\mathcal{F}f = \hat{f}$) satisfies the following:*

(i) For all $f \in L^2(\mathbb{R})$,

$$||\hat{f}||_2 = ||f||_2. \qquad (7.12)$$

(ii) For all $f, g \in L^2(\mathbb{R})$,

$$\langle \hat{f}, \hat{g}\rangle = \langle f, g\rangle. \qquad (7.13)$$

Proof. As we have seen in Lemma 7.2.1,

$$||\mathcal{F}f|| = ||f||_2, \ \forall f \in C_c(\mathbb{R}). \tag{7.14}$$

By Theorem 3.3.2 and the fact that $C_c(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, this implies that the Fourier transform in a unique way can be extended to a bounded operator $\mathcal{F}$ from $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$. It follows from the construction in Theorem 3.3.2 that the equality in (7.14) holds for all $f \in L^2(\mathbb{R})$, see Exercise 7.8. This proves (i).

The result in (ii) follows from (i) and the polarization identity in Theorem 4.1.4 (Exercise 7.8). That the considered extension of the Fourier transform is unitary follows from the fact that it is surjective and an isometry; we skip the proofs of these facts. □

Recall that a unitary operator is invertible (see (4.17)). Thus, the result in Theorem 7.2.2 allows us to speak about the *inverse Fourier transform,* $\mathcal{F}^{-1}$, as a mapping from $L^2(\mathbb{R})$ onto $L^2(\mathbb{R})$.

The equation (7.13) is called *Parseval's equation,* and the equation (7.12) is called *Plancherel's equation.*

The inversion formula in Theorem 7.1.7 for the Fourier transform has a similar version for $L^2(\mathbb{R})$-functions:

Theorem 7.2.3 (Inversion formula for $L^2(\mathbb{R})$-functions) *Assume that $f \in L^2(\mathbb{R})$ and that $\hat{f} \in L^1(\mathbb{R})$. Then*

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\gamma) e^{2\pi i x\gamma} d\gamma \ \ \text{for almost all } x \in \mathbb{R}. \tag{7.15}$$

If f is continuous, the formula (7.15) holds pointwise for all $x \in \mathbb{R}$.

The proof of Theorem 7.2.3 is almost identical with the proof of the inversion formula for $L^1(\mathbb{R})$-functions outlined in Appendix A.2: the only difference is that the result in Lemma A.2.4 needs to be replaced by a similar statement for $f \in L^2(\mathbb{R})$ rather than $f \in L^1(\mathbb{R})$.

In the literature, one finds several extensions of the Fourier transform to other spaces, e.g., to $L^p(\mathbb{R})$. The extension to $L^2(\mathbb{R})$ is particularly convenient because $L^2(\mathbb{R})$ is a Hilbert space, and because of the special properties stated in Theorem 7.2.2. Due to the various possible domains for the Fourier transform, it is necessary to pay close attention to the assumptions when dealing with the properties of the Fourier transform. For example, Theorem 7.1.5 does not hold if the assumption $f \in L^1(\mathbb{R})$ is replaced by $f \in L^2(\mathbb{R})$.

7.3 Convolution

In manipulations with the Fourier transform, the concept of *convolution* is very useful. Given two functions $f, g \in L^1(\mathbb{R})$, the convolution is defined as a new function, denoted by $f * g$:

Definition 7.3.1 (Convolution) *For two functions $f, g \in L^1(\mathbb{R})$, the convolution $f * g : \mathbb{R} \to \mathbb{C}$ is defined by*

$$(f * g)(y) := \int_{-\infty}^{\infty} f(y - x)g(x)\, dx, \;\; y \in \mathbb{R}. \tag{7.16}$$

In order for this to be a valid definition, we have to check that the integral appearing in (7.16) exists. This is the case if $f, g \in L^1(\mathbb{R})$, but also under other assumptions as well, e.g., if $f \in L^2(\mathbb{R}), \;\; g \in L^1(\mathbb{R})$:

Lemma 7.3.2 *For $g \in L^1(\mathbb{R})$, the following hold:*

(i) If $f \in L^1(\mathbb{R})$, the convolution $(f * g)(y)$ is well defined for all $y \in \mathbb{R}$, and defines a function in $L^1(\mathbb{R})$. Furthermore,

$$||f * g||_1 \le ||f||_1 \, ||g||_1. \tag{7.17}$$

(ii) If $f \in L^2(\mathbb{R})$, the convolution $(f * g)(y)$ is well defined for almost all $y \in \mathbb{R}$, and defines a function in $L^2(\mathbb{R})$. Furthermore,

$$||f * g||_2 \le ||f||_2 \, ||g||_1. \tag{7.18}$$

Proof. The arguments below actually imply that the convolution $f * g$ is well defined under the stated assumptions, but this will only be clear for a reader with knowledge of measure theory. Thus, for the sake of the reader without that knowledge, we note that at least if $f, g \in C_c(\mathbb{R})$, the convolution $(f * g)(y)$ is well defined for all $y \in \mathbb{R}$. In fact, for any fixed $y \in \mathbb{R}$, the function $x \mapsto f(y - x)g(x)$ is continuous and has compact support, and therefore belongs to $L^1(\mathbb{R})$.

Now let $f, g \in L^1(\mathbb{R})$. Assuming for the moment that $f * g$ is well-defined, we need to show that $f * g \in L^1(\mathbb{R})$, i.e., that

$$\int_{-\infty}^{\infty} |(f * g)(y)|\, dy < \infty. \tag{7.19}$$

Now,

$$|(f * g)(y)| = \left| \int_{-\infty}^{\infty} f(y-x)g(x)\,dx \right| \leq \int_{-\infty}^{\infty} |f(y-x)g(x)|\,dx. \tag{7.20}$$

If we can show that

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(y-x)g(x)|\,dx \right) dy < \infty, \tag{7.21}$$

measure theory will tell us that the expression

$$\int_{-\infty}^{\infty} |f(y-x)g(x)|\,dx$$

is finite for (almost all) $y \in \mathbb{R}$, i.e., that the mapping $x \to f(y-x)g(x)$ is integrable. This means that the convolution $f * g$ is well defined. Due to (7.20), (7.21) will also give us that (7.19) holds. Thus, all we have to do is to verify (7.21).

In order to do that we will use Fubini's theorem and first consider the expression (7.21) with the order of the integrals interchanged. Using the change of variable $z = y - x$ and the fact that

$$\int_{-\infty}^{\infty} |f(y-x)|\,dy = \int_{-\infty}^{\infty} |f(z)|\,dz = ||f||_1,$$

we see that

$$\begin{aligned} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(y-x)g(x)|\,dy \right) dx &= \int_{-\infty}^{\infty} |g(x)| \left(\int_{-\infty}^{\infty} |f(y-x)|\,dy \right) dx \\ &= \int_{-\infty}^{\infty} |g(x)|\,||f||_1\,dx \\ &= ||f||_1\,||g||_1 \\ &< \infty. \end{aligned}$$

Using first (7.20) and then Fubini's theorem, we arrive at

$$\begin{aligned} \int_{-\infty}^{\infty} |(f * g)(y)|\,dy &= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} f(y-x)g(x)\,dx \right| dy \\ &\leq \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(y-x)g(x)|\,dx \right) dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(y-x)g(x)|\,dy \right) dx \\ &= ||f||_1\,||g||_1 \\ &< \infty. \end{aligned}$$

That is, $f * g \in L^1(\mathbb{R})$, and (7.17) holds. This proves (i). In order to prove (ii), assume that $f \in L^2(\mathbb{R})$ and $g \in L^1(\mathbb{R})$. We will prove that

$$\int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} |f(y-x)|\, |g(x)|\, dx \right|^2 dy \le ||g||_1^2\, ||f||_2^2. \tag{7.22}$$

As in (i), this implies that

$$\int_{-\infty}^{\infty} |f(y-x)|\, |g(x)|\, dx$$

is finite for a.e. $y \in \mathbb{R}$, i.e., that the convolution $f * g$ is well defined. The estimate (7.22) will also imply that (7.18) holds.

Let us first consider the inner term in (7.22). By Cauchy–Schwarz' inequality,

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} |f(y-x)|\, |g(x)|\, dx\right)^2 &= \left(\int_{-\infty}^{\infty} |f(y-x)|\, |g(x)|^{1/2}\, |g(x)|^{1/2}\, dx\right)^2 \\
&\le \int_{-\infty}^{\infty} |f(y-x)|^2\, |g(x)|\, dx \int |g(x)|\, dx \\
&= ||g||_1 \int_{-\infty}^{\infty} |f(y-x)|^2\, |g(x)|\, dx.
\end{aligned}$$

Inserting this and using a calculation like in the proof of (i),

$$\begin{aligned}
&\int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} |f(y-x)|\, |g(x)|\, dx \right|^2 dy \\
\le\ & \int_{-\infty}^{\infty} ||g||_1 \left(\int_{-\infty}^{\infty} |f(y-x)|^2\, |g(x)|\, dx\right) dy \\
=\ & ||g||_1 \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(y-x)|^2\, |g(x)|\, dy\right) dx \\
\le\ & ||g||_1 \int |g(x)| \left(\int_{-\infty}^{\infty} |f(y-x)|^2\, dy\right) dx \\
\le\ & ||g||_1^2\, ||f||_2^2.
\end{aligned}$$

This implies that (7.22) holds, as desired. □

Lemma 7.3.2 has an extension to functions $f \in L^p(\mathbb{R})$, see Exercise 7.10.

For later reference we state the following result, which tells us that convolution is commutative (Exercise 7.11):

Lemma 7.3.3 *If* $f, g \in L^1(\mathbb{R})$, *then* $f * g = g * f$.

The Fourier transform and convolution are related by the following important result.

Theorem 7.3.4 (Fourier transform and convolution)

(i) If $f, g \in L^1(\mathbb{R})$, the formula

$$\widehat{f * g}(\gamma) = \hat{f}(\gamma)\hat{g}(\gamma) \tag{7.23}$$

holds for all $\gamma \in \mathbb{R}$;

(ii) If $f \in L^2(\mathbb{R})$ and $g \in L^1(\mathbb{R})$, the formula (7.23) holds for almost all $\gamma \in \mathbb{R}$.

Proof. Let $f, g \in L^1(\mathbb{R})$. In the proof of Lemma 7.3.2, we saw that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(y-x)g(x)|\, dx\, dy < \infty.$$

Because $|e^{-2\pi i y\gamma}| = 1$, this implies that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|f(y-x)g(x)e^{-2\pi i y\gamma}\right|\, dx\, dy < \infty.$$

Via Fubini's theorem,

$$\begin{aligned}
\widehat{f * g}(\gamma) &= \int_{-\infty}^{\infty} (f*g)(y)e^{-2\pi i y\gamma}\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(y-x)g(x)\, dx\right) e^{-2\pi i y\gamma}\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(y-x)e^{-2\pi i y\gamma}\, dy\right) g(x)\, dx.
\end{aligned}$$

Writing

$$e^{-2\pi i y\gamma} = e^{-2\pi i (y-x)\gamma} e^{-2\pi i x\gamma},$$

it follows that

$$\begin{aligned}
\widehat{f * g}(\gamma) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(y-x)e^{-2\pi i (y-x)\gamma} e^{-2\pi i x\gamma}\, dy\right) g(x)\, dx \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(y-x)e^{-2\pi i (y-x)\gamma}\, dy\right) g(x)e^{-2\pi i x\gamma}\, dx.
\end{aligned}$$

Now, by the change of variable $z = y - x$ in the inner integral,

$$\begin{aligned}
\widehat{f * g}(\gamma) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(z)e^{-2\pi i z\gamma}\, dz\right) g(x)e^{-2\pi i x\gamma}\, dx \\
&= \left(\int_{-\infty}^{\infty} f(z)e^{-2\pi i z\gamma}\, dz\right)\left(\int_{-\infty}^{\infty} g(x)e^{-2\pi i x\gamma}\, dx\right) \\
&= \hat{f}(\gamma)\hat{g}(\gamma).
\end{aligned}$$

This proves (i). We skip the proof of (ii). □

7.4 The sampling theorem

A short and not yet precise formulation of the *sampling problem* is: How can we recover a function $f : \mathbb{R} \to \mathbb{C}$ if we only know a countable set of function values $\{f(\lambda_k)\}_{k\in I}$? Here, "recover" means to be able to determine completely what function we are dealing with. Formulated this way the problem is ill-posed: there are infinitely many functions that take the same prescribed values on a given countable set, so we need to impose some condition on the function f for the problem to make sense. Traditionally, this is done by requiring f to belong to a certain function space. A classical example is to consider a space of *band-limited* functions:

Definition 7.4.1 (Band-limited functions, Paley–Wiener space)

(i) A function $f \in L^2(\mathbb{R})$ is band-limited if the Fourier transform $\hat{f}$ has compact support.

(ii) The Paley–Wiener space PW is the subspace of $L^2(\mathbb{R})$ defined by

$$PW := \left\{ f \in L^2(\mathbb{R}) \mid supp\, \hat{f} \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right] \right\}. \tag{7.24}$$

We begin with a lemma that will be used repeatedly. We ask the reader to provide the proof in Exercise 7.12:

Lemma 7.4.2 *If $f \in PW$, then $\hat{f} \in L^1(\mathbb{R})$.*

As always when dealing with L^2-functions, the Paley–Wiener space really consists of equivalence classes of functions. However, due to the fact that the Fourier transform of functions in PW has compact support, each of these equivalence classes contains a continuous function:

Theorem 7.4.3 (Continuity of functions in PW) *Assume that $f \in PW$. Then f is equivalent to a continuous function.*

Proof. Lemma 7.4.2 shows that $\hat{f} \in L^1(\mathbb{R})$ if $f \in PW$. Thus, by Theorem 7.1.5, the function

$$\mathcal{F}\hat{f}(x) = \int_{-\infty}^{\infty} \hat{f}(\gamma) e^{-2\pi i\gamma x}\, d\gamma$$

is continuous. This clearly implies that the function

$$g(x) := \mathcal{F}\hat{f}(-x) = \int_{-\infty}^{\infty} \hat{f}(\gamma) e^{2\pi i\gamma x}\, d\gamma$$

is continuous as well. By the inversion formula for $L^2(\mathbb{R})$-functions, Theorem 7.2.3,

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\gamma) e^{2\pi i\gamma x}\, d\gamma$$

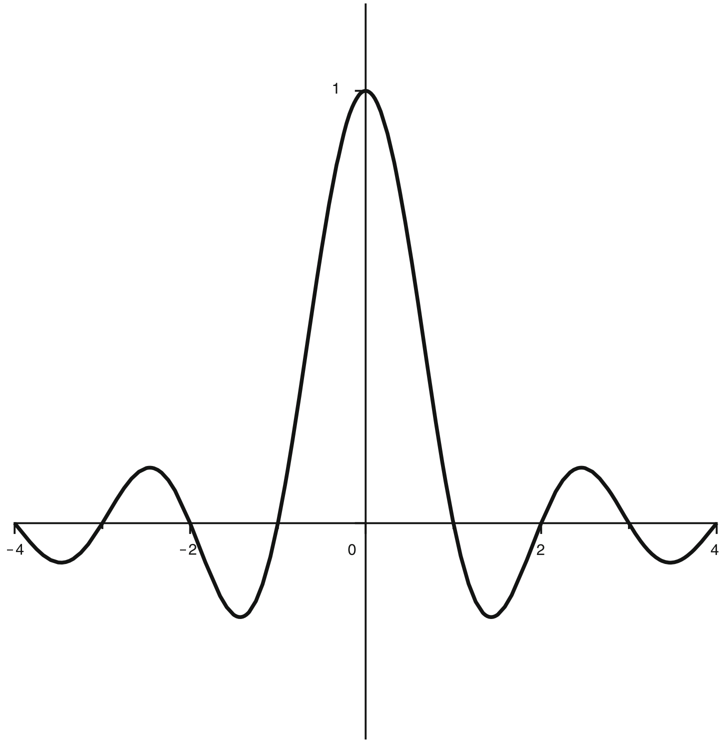

Figure 7.2. The sinc-function defined in (7.25).

holds for almost all x; that is, f is equal to the continuous function g almost everywhere. □

We will now show that the Paley–Wiener space has an orthonormal basis consisting of translates of a single function. The relevant function is called the *sinc-function*:

Definition 7.4.4 (Sinc-function) *The sinc-function is given by*

$$sinc(x) := \begin{cases} \frac{\sin(\pi x)}{\pi x} & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases} \tag{7.25}$$

Shannon's sampling theorem states that any continuous function f in the Paley–Wiener space can be fully recovered from its samples at the integers. That is, if we know the values $f(k)$ for all $k \in \mathbb{Z}$, we can determine the function f completely:

Theorem 7.4.5 (Shannon's sampling theorem) *The functions* $\{sinc(\cdot - k)\}_{k\in\mathbb{Z}}$ *form an orthonormal basis for PW. If* $f \in PW$ *is continuous, then*

$$f(x) = \sum_{k\in\mathbb{Z}} f(k)\, sinc(x-k), \tag{7.26}$$

with two interpretations of the convergence of the infinite series:

(i) The symmetric partial sums converge pointwise, i.e.,

$$\lim_{N\to\infty} \sum_{k=-N}^{N} f(k)\, sinc(x-k) \quad = \quad f(x), \ \forall x \in \mathbb{R};$$

(ii) The symmetric partial sums converge in $L^2(\mathbb{R})$*, i.e.,*

$$\int_{-\infty}^{\infty} \left| f(x) - \sum_{k=-N}^{N} f(k)\, sinc(x-k) \right|^2 dx \quad \to \quad 0 \text{ as } N \to \infty.$$

Proof. The proof is based on classical Fourier analysis. Because of our definition of the Paley–Wiener space, it will be convenient to work with Fourier series in the space $L^2(-1/2, 1/2)$.

We first show that the functions $\{\mathrm{sinc}(\cdot - k)\}_{k\in\mathbb{Z}}$ form an orthonormal sequence in $L^2(\mathbb{R})$. As noted in Example 6.4.3 the functions $\{e^{2\pi i k(\cdot)}\chi_{]-1/2,1/2[}(\cdot)\}_{k\in\mathbb{Z}}$ form an orthononormal sequence in $L^2(\mathbb{R})$. Taking the Fourier transform of these functions, we arrive at

$$\begin{aligned}
\mathcal{F}\left(e^{2\pi i k(\cdot)}\chi_{]-1/2,1/2[(\cdot)}\right)(\gamma) &= \int_{-1/2}^{1/2} e^{2\pi i k x} e^{-2\pi i x\gamma} dx \\
&= \int_{-1/2}^{1/2} e^{-2\pi i(\gamma-k)x} dx \\
&= \left[\frac{1}{-2\pi i(\gamma-k)} e^{-2\pi i(\gamma-k)x} \right]_{x=-1/2}^{1/2} \\
&= \mathrm{sinc}(\gamma-k).
\end{aligned}$$

Because the Fourier transformation is unitary, (4.16) implies that the functions $\{\mathrm{sinc}(\cdot - k)\}_{k\in\mathbb{Z}}$ are orthonormal as well.

Now let $f \in PW$ be the continuous representative for a given equivalence class. On the interval $]-1/2, 1/2[$ we can apply the results in Example 6.4.3 to expand $\hat{f}$ in a Fourier series,

$$\hat{f}(\cdot) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi i k(\cdot)}, \tag{7.27}$$

where

$$c_k = \int_{-1/2}^{1/2} \hat{f}(\gamma) e^{-2\pi i k\gamma} d\gamma. \tag{7.28}$$

According to Example 6.4.3 the partial sums of the Fourier series converge in the norm of $L^2(-1/2, 1/2)$, i.e.,

$$\int_{-1/2}^{1/2} \left| \hat{f}(\gamma) - \sum_{n=-N}^{N} c_k e^{2\pi i k\gamma} \right|^2 d\gamma \to 0 \text{ as } N \to \infty.$$

Because we are dealing with a finite interval, convergence in $L^2(-1/2, 1/2)$ implies convergence in $L^1(-1/2, 1/2)$, see Exercise 5.10. So

$$\int_{-1/2}^{1/2} \left| \hat{f}(\gamma) - \sum_{n=-N}^{N} c_k e^{2\pi i k\gamma} \right| d\gamma \to 0 \text{ as } N \to \infty. \tag{7.29}$$

By Lemma 7.4.2, we know that $\hat{f} \in L^1(\mathbb{R})$. Thus, the expression for c_k in (7.28) implies by Theorem 7.2.3 that

$$c_k = \int_{-1/2}^{1/2} \hat{f}(\gamma) e^{-2\pi i k\gamma} d\gamma = \int_{-\infty}^{\infty} \hat{f}(\gamma) e^{-2\pi i k\gamma} d\gamma = f(-k).$$

Thus, via (7.27),

$$\hat{f}(\cdot) = \sum_{k\in\mathbb{Z}} f(-k) e^{2\pi i k(\cdot)}.$$

Using Theorem 7.2.3 once more, we arrive at the following formula, valid pointwise for all $x \in \mathbb{R}$:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\gamma) e^{2\pi i x\gamma} d\gamma \quad = \quad \int_{-1/2}^{1/2} \left(\sum_{k\in\mathbb{Z}} f(-k) e^{2\pi i k\gamma} \right) e^{2\pi i x\gamma} d\gamma.$$

Note that via Lemma 1.7.2 and (7.29),

$$\begin{aligned}
&\left| f(x) - \sum_{k=-N}^{N} f(-k) \int_{-1/2}^{1/2} e^{2\pi i (x+k)\gamma} d\gamma \right| \\
&\left| \int_{-\infty}^{\infty} \hat{f}(\gamma) e^{2\pi i x\gamma} d\gamma - \sum_{k=-N}^{N} f(-k) \int_{-1/2}^{1/2} e^{2\pi i (x+k)\gamma} d\gamma \right| \\
= \; &\left| \int_{-\infty}^{\infty} \left(\hat{f}(\gamma) - \sum_{k=-N}^{N} f(-k) e^{2\pi i k\gamma} \right) e^{2\pi i x\gamma} d\gamma \right| \\
\leq \; &\int_{-\infty}^{\infty} \left| \left(\hat{f}(\gamma) - \sum_{k=-N}^{N} f(-k) e^{2\pi i k\gamma} \right) e^{2\pi i x\gamma} \right| d\gamma \\
= \; &\int_{-\infty}^{\infty} \left| \hat{f}(\gamma) - \sum_{k=-N}^{N} f(-k) e^{2\pi i k\gamma} \right| d\gamma \\
&\to 0 \text{ as } N \to \infty.
\end{aligned}$$

This shows that

$$
\begin{aligned}
f(x) &= \lim_{N\to\infty} \sum_{k=-N}^{N} f(-k) \int_{-1/2}^{1/2} e^{2\pi i(x+k)\gamma} d\gamma \\
&= \sum_{k\in\mathbb{Z}} f(-k) \int_{-1/2}^{1/2} e^{2\pi i(x+k)\gamma} d\gamma \\
&= \sum_{k\in\mathbb{Z}} f(-k) \left[\frac{1}{2\pi i(x+k)} e^{2\pi i(x+k)\gamma} \right]_{\gamma=-1/2}^{1/2} \\
&= \sum_{k\in\mathbb{Z}} f(-k) \,\mathrm{sinc}(x+k) \\
&= \sum_{k\in\mathbb{Z}} f(k) \,\mathrm{sinc}(x-k).
\end{aligned}
$$

We have now proved (i). In order to prove (ii), we rewrite the result obtained in (i) as

$$
f(x) - \sum_{n=-N}^{N} f(k) \,\mathrm{sinc}(x-k) = \sum_{|n|>N} f(k) \,\mathrm{sinc}(x-k).
$$

Note that because $\{f(k)\}_{k\in\mathbb{Z}}$ are Fourier coefficients, we know that $\{f(k)\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$. Since $\{\mathrm{sinc}(\cdot - k)\}_{k\in\mathbb{Z}}$ is an orthonormal system, the result in Exercise 4.31 shows that

$$
\begin{aligned}
\left\| f - \sum_{n=-N}^{N} f(k) \,\mathrm{sinc}(\cdot - k) \right\|_2 &= \left\| \sum_{|n|>N} f(k) \,\mathrm{sinc}(\cdot - k) \right\|_2 \\
&= \sqrt{\sum_{|n|>N} |f(k)|^2},
\end{aligned}
$$

which converges to 0 as $N \to \infty$ because $\{f(k)\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$. This implies that

$$
\overline{\mathrm{span}}\{T_k \,\mathrm{sinc}\}_{k\in\mathbb{Z}} = PW.
$$

We have already seen that $\{\mathrm{sinc}(\cdot - k)\}_{k\in\mathbb{Z}}$ is an orthonormal sequence, so by Theorem 4.7.2 we can now also conclude that $\{\mathrm{sinc}(\cdot - k)\}_{k\in\mathbb{Z}}$ forms an orthonormal basis for PW. □

Via an appropriate scaling, the result in Theorem 7.1.7 can be extended to functions whose Fourier transform has support in an arbitrary fixed interval (Exercise 7.13). In fact, if $\mathrm{supp}\,\hat{f} \subseteq [-\alpha/2, \alpha/2]$, Shannon's sampling theorem takes the form

$$
f(x) = \sum_{k\in\mathbb{Z}} f(\frac{k}{\alpha}) \,\mathrm{sinc}(\alpha x - k), \; x \in \mathbb{R}. \tag{7.30}
$$

The principle in Shannon's sampling theorem is the basis for modern communication technology. Most signals appearing in practice depend on a continuous variable (very often, the time). Processing of such a signal is facilitated greatly if it can be stored and handled in terms of a sequence of samples. Under the assumptions in Theorem 7.4.5 this is possible: all information about a function $f \in PW$ is contained in the scalar sequence $\{f(k)\}_{k\in\mathbb{Z}}$.

Example 7.4.6 Consider a recording of a piece of music. In principle, all frequencies might appear, but the human ear can only hear frequencies belonging to a certain range (at most up to 20000 Hz). Thus, we can remove the high frequencies and consider the resulting signal f as band-limited, e.g., with $\operatorname{supp} \hat{f} \subseteq [-\frac{40000}{2}, \frac{40000}{2}] = [-20000, 20000]$. Formula (7.30) shows that this signal can be recovered from its samples at the points $\frac{k}{40000}$, $k \in \mathbb{Z}$. Thus, all information about the signal is contained in a discrete sequence of numbers! This principle is used in CD players and other places where a conversion of an analog signal to a digital signal is needed. □

7.5 The discrete Fourier transform

We have now introduced the Fourier transform on $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$, and seen a few cases of its use. Unfortunately, the definition of the Fourier transform shows that it is difficult to calculate $\hat{f}$ explicitly for most functions f; usually, we need to use numerical methods. We will now shortly discuss the role of the discrete Fourier transform considered in Example 1.1.3 in this context. The presentation is inspired by the book [4], to which we also refer for a much more detailed description of the DFT and its applications.

In order to calculate or approximate the Fourier transform of a real-life signal f, we need to assume that f has compact support. Assuming that $\operatorname{supp} f \subseteq [0, A]$ for some $A > 0$, we have that

$$\begin{aligned} \hat{f}(\gamma) &= \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\gamma}\,dx \\ &= \int_0^A f(x)e^{-2\pi i x\gamma}\,dx. \end{aligned} \tag{7.31}$$

Let $n \in \mathbb{N}$. We will now split the interval $[0, A]$ into n subintervals of length $\Delta x = A/n$; the *grid points,* i.e., the end points of the subintervals, are

$$x_\ell = (\ell - 1)\Delta x, \;\; \ell = 1, \dots, n+1.$$

Fix $\gamma \in \mathbb{R}$, and let

$$g(x) := f(x)e^{-2\pi i x\gamma}.$$

We will assume that $g(0) = g(A)$. If this is not the case, an extra correction term has to be added below; however, as $n \to \infty$, this term will disappear. The trapezoid rule for approximation of the integral in (7.31) gives that

$$\hat{f}(\gamma) \approx \frac{\Delta x}{2} \left(g(0) + 2\sum_{\ell=2}^{n} g(x_\ell) + g(A) \right) = \Delta x \sum_{\ell=1}^{n} g(x_\ell).$$

We are considering functions f supported on $[0, A]$. If we want such a function to be periodic, with a period fitting exactly in $[0, A]$, the *largest possible period* is exactly A; that is, the *smallest possible frequency* that can appear in the function f is $1/A$. For this reason we will consider the frequency content in f with frequencies $\gamma_k := (k-1)/A,\ k = 1, \dots, n$. Observing that

$$x_\ell y_k = (\ell - 1)\frac{A}{n}\frac{k-1}{A} = \frac{(\ell-1)(k-1)}{n},$$

we see that

$$\begin{aligned} \hat{f}(\gamma_k) &\approx \frac{A}{n} \sum_{\ell=1}^{n} f(x_\ell)\, e^{-2\pi i x_\ell \gamma_k} \\ &= \frac{A}{n} \sum_{\ell=1}^{n} f(x_\ell)\, e^{-2\pi i(\ell-1)(k-1)/n}. \end{aligned} \tag{7.32}$$

Recall from Example 1.1.3 that the discrete Fourier transform of a sequence $\mathbf{v} \in \mathbb{C}^n$ is defined as $\widehat{\mathbf{v}} := \{\langle \mathbf{v}, \mathbf{e}_k \rangle\}_{k=1}^{n}$ with $\mathbf{e}_k$ as in (1.6). Writing

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_n \end{pmatrix},$$

the coordinates of $\widehat{\mathbf{v}}$ are

$$\langle \mathbf{v}, \mathbf{e}_k \rangle = \frac{1}{\sqrt{n}} \sum_{\ell=1}^{n} v_\ell\, e^{-2\pi i(\ell-1)(k-1)/n},\ k = 1, \dots, n.$$

Thus, up to the multiplicative factor $A/\sqrt{n}$, the term in (7.32) used to approximate the sequence $\{\hat{f}(\gamma_k)\}_{k=1}^{n}$ is precisely the discrete Fourier transform of the sequence $\{f(x_k)\}_{k=1}^{n}$. That is, we can approximate the sequence $\{\hat{f}(\gamma_k)\}_{k=1}^{n}$ via a discrete Fourier transform. Letting $n \to \infty$, the exact values for $\hat{f}(\gamma_k)$ appear in the limit.

In practice, the discrete Fourier transform is usually calculated using an algorithm known as the *fast Fourier transform,* abbreviated FFT.

7.6 Exercises

7.1 Calculate the Fourier transform of the function $\chi_{[0,1]}$.

7.2 Calculate the Fourier transform of the function

$$f(x) = e^{-x}\chi_{[0,1]}(x).$$

7.3 Calculate the Fourier transform of the functions

$$g(x) = e^{-x}\chi_{[0,\infty[}(x)$$

and

$$f(x) = e^{-|x|}.$$

7.4 Prove Theorem 7.1.2.

7.5 Prove Lemma 7.1.3.

7.6 Complete the proof of Theorem 7.1.5 by showing that

$$||T_y f - f||_1 \to 0 \text{ as } y \to 0$$

for all $f \in L^1(\mathbb{R})$.

7.7 Complete the proof of Lemma 7.2.1 by proving (7.8) for functions $f \in C_c(\mathbb{R})$ that are not necessarily supported on $[-\pi, \pi]$.
Hint: let as usual D_c denote the dilation operator. For sufficiently small values of $c > 0$ the function $g := D_c f$ will be supported on $[-\pi, \pi]$. Argue for each of the following equalities:

$$||f||_2 = ||D_c f||_2 = ||\mathcal{F}g||_2 = ||D_c \mathcal{F}g||_2 = ||\mathcal{F}D_{c^{-1}}g||_2 = ||\mathcal{F}f||_2.$$

7.8 This exercise concerns some of the steps in the proof of Theorem 7.2.2.

(i) Consider the extension of the Fourier transform from $C_c(\mathbb{R})$ to $L^2(\mathbb{R})$, as described in the proof of Theorem 7.2.2. Argue that (7.14) holds for all $f \in L^2(\mathbb{R})$.
Hint: consider the first part of (3.16) with T being the Fourier transform, and apply the result in Exercise 2.1.

(ii) Prove Theorem 7.2.2(ii).

7.9 We consider a convolution of two characteristic functions.

(i) Calculate the function $\chi_{[0,1]} * \chi_{[0,2]}$.

(ii) Make a sketch of the function $\chi_{[0,1]} * \chi_{[0,2]}$.

(iii) Calculate the Fourier transform of the function $\chi_{[0,1]} * \chi_{[0,2]}$.

7.10 (Young's inequality) Generalize the proof of Lemma 7.3.2 to show that if $f \in L^p(\mathbb{R}),\ p \in]1,\infty[$, and $g \in L^1(\mathbb{R})$, the convolution $f * g$ defines a function in $L^p(\mathbb{R})$, and

$$||f * g||_p \leq ||f||_p \, ||g||_1. \tag{7.33}$$

7.11 Prove Lemma 7.3.3.

7.12 Prove Lemma 7.4.2.

7.13 Let $f \in L^2(\mathbb{R})$ be a continuous function for which

$$\text{supp}\, \hat{f} \subseteq [-\alpha/2, \alpha/2]$$

for some $\alpha > 0$. Show that f can be recovered from its samples $\{f(k/\alpha)\}_{k\in\mathbb{Z}}$ via

$$f(x) = \sum_{k\in\mathbb{Z}} f(\frac{k}{\alpha})\ \text{sinc}(\alpha x - k),\ \ x \in \mathbb{R}.$$

7.14 Prove Lemma A.2.2.

7.15 Consider the setup in Lemma A.2.4.

(i) Prove that the integral in (A.20) is finite.

(ii) Prove Lemma A.2.4(ii).

8

An Introduction to Wavelet Analysis

In Section 4.7 we introduced orthonormal bases in general Hilbert spaces. The purpose of the current chapter is to present a general way of constructing orthonormal bases with a particular structure in the Hilbert space $L^2(\mathbb{R})$. In contrast to the other topics treated in the book, wavelet analysis is a quite new topic: although the first constructions appeared about 100 years ago, the systematic analysis began around 1982. In 1987, the key concept of a multiresolution analysis was introduced, and shortly hereafter Daubechies used it to construct a special class of orthonormal bases with attractive properties, e.g., in the context of data compression.

The present chapter gives a quick overview of the key ideas in wavelet analysis. The subsequent Chapter 9 provides the technical details in the construction of a wavelet orthonormal basis for $L^2(\mathbb{R})$.

The basic definitions in wavelet analysis are considered in Section 8.1. Section 8.2 is devoted to the main tool in wavelet analysis, the multiresolution analysis. Section 8.3 describes certain properties that are desirable for concrete wavelet constructions, as well as the construction of Daubechies' wavelets. Section 8.4 discusses a few aspects of the role of these wavelets in applications.

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4980-7_8, © Springer Science+Business Media, LLC 2010

8.1 Wavelets

Wavelet theory provides us with a way of constructing orthonormal bases in $L^2(\mathbb{R})$. Compared to our description of general orthonormal bases in Section 4.7 there are two new issues:

- We now deal with a *concrete* Hilbert space consisting of functions, namely, $L^2(\mathbb{R})$;
- We want the functions in the basis to have a special structure: all of them will be scaled and translated versions of a fixed function.

That the analysis shall take place in $L^2(\mathbb{R})$ is motivated by the fact that many applications, e.g., in signal processing, actually take place within this framework. The special structure of the functions in the basis is also motivated by applications. In fact, if $\{\mathbf{e}_k\}$ is an orthonormal basis for $L^2(\mathbb{R})$, we know from Theorem 4.7.2 that all functions $f \in L^2(\mathbb{R})$ have an expansion

$$f = \sum c_k \mathbf{e}_k$$

for suitable coefficients c_k. In general, the index set of the sum is infinite. However, in order for this representation to be of practical use, it is important that the relevant functions f can be approximated well by finite partial sums, preferably with just a few nonzero coefficients c_k. In Section 8.3 we will see that it often is possible to construct bases with wavelet structure that satisfies this requirement; see Theorem 8.3.3.

The technical definition of a wavelet basis is as follows:

Definition 8.1.1 (Wavelet) *Let* $\psi \in L^2(\mathbb{R})$.

(i) For $j, k \in \mathbb{Z}$, *define the function* $\psi_{j,k}$ *by*

$$\psi_{j,k}(x) := 2^{j/2}\psi(2^j x - k), \ x \in \mathbb{R}. \tag{8.1}$$

(ii) The function ψ *is called a wavelet if the functions* $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ *form an orthonormal basis for* $L^2(\mathbb{R})$.

In terms of the translation operators T_k and the dilation operator D introduced in Section 6.2, we can write the functions $\psi_{j,k}$ as (Exercise 6.7)

$$\psi_{j,k} = D^j T_k \psi, \ j, k \in \mathbb{Z}.$$

The systematic study of wavelet bases began around 1985, but the first example of a wavelet appeared much earlier:

Example 8.1.2 (Haar wavelet) The *Haar function* is defined by

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \le x < \frac{1}{2}, \\ -1 & \text{if } \frac{1}{2} \le x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8.2}$$

Already in 1910 it was proved by Haar that the functions $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ constitute an orthonormal basis for $L^2(\mathbb{R})$ for this choice of ψ. The proof of the basis property is quite technical and will be skipped (see [8] or [12]). For the orthonormality one can argue as follows. We first consider $\psi_{j,k}$ and $\psi_{j,k'}$, i.e., elements with the same dilation parameter. By Lemma 6.2.2 we know that the dilation operator D is unitary; by Exercise 4.23 this implies that D^j is unitary for all $j \in \mathbb{Z}$. Using (4.16) and that $\operatorname{supp}\psi = [0,1]$,

$$\begin{aligned}\langle \psi_{j,k}, \psi_{j,k'}\rangle = \langle D^jT_k\psi, D^jT_{k'}\psi\rangle &= \langle T_k\psi, T_{k'}\psi\rangle \\ &= \int_{-\infty}^{\infty} \psi(x-k)\overline{\psi(x-k')}\,dx \\ &= \delta_{k,k'}.\end{aligned}$$

Now assume that $j' \neq j$, say, $j' > j$. Using Lemma 6.2.2 and the commutation relations in Lemma 6.2.4,

$$\begin{aligned}\langle \psi_{j,k}, \psi_{j',k'}\rangle = \langle D^jT_k\psi, D^{j'}T_{k'}\psi\rangle &= \langle (D^{j'})^*D^jT_k\psi, T_{k'}\psi\rangle \\ &= \langle D^{j-j'}T_k\psi, T_{k'}\psi\rangle \\ &= \langle T_{-k'}D^{j-j'}T_k\psi, \psi\rangle \\ &= \langle D^{j-j'}T_{-k'2^{j-j'}+k}\psi, \psi\rangle.\end{aligned}$$

The function ψ has support on the interval $[0,1]$, so $T_{-k'2^{j-j'}+k}\psi$ has support on the interval $[-k'2^{j-j'}+k, -k'2^{j-j'}+k+1]$; this implies that $D^{j-j'}T_{-k'2^{j-j'}+k}\psi$ has support on the interval

$$\begin{aligned}I: &= [2^{j'-j}(-k'2^{j-j'}+k), 2^{j'-j}(-k'2^{j-j'}+k+1)] \\ &= [-k'+2^{j'-j}k, -k'+2^{j'-j}(k+1)].\end{aligned}$$

The length of I is $2^{j'-j}$, which can take the values $2, 4, 8, ...$ Now, the support of ψ has length 1, and is contained in an interval on which $D^{j-j'}T_{-k'2^{j-j'}+k}\psi$ is constant (make a picture!); it follows that

$$\langle \psi_{j',k'}, \psi_{j,k}\rangle = \int_{-\infty}^{\infty} \left(D^{j-j'}T_{-k'2^{j-j'}+k}\psi\right)(x)\psi(x)\,dx = 0.$$

This concludes the proof of the orthonormality of $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$. □

Example 8.1.2 illustrates how complicated it is to verify directly that a function ψ is a wavelet: just to show the orthonormality of the functions $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is quite involved. In the next section we present a general method for construction of such a function.

8.2 Multiresolution analysis

In 1986, Mallat and Meyer introduced *multiresolution analysis* as a general tool to construct wavelet orthonormal bases. A multiresolution analysis consists of a collection of conditions on certain subspaces of $L^2(\mathbb{R})$ and an associated function $\phi \in L^2(\mathbb{R})$. We note already now that it will take some work before we can show how to use the concept to construct an orthonormal basis for $L^2(\mathbb{R})$ in Theorem 8.2.7.

Definition 8.2.1 (Multiresolution analysis) *A multiresolution analysis for $L^2(\mathbb{R})$ consists of a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ and a function $\phi \in V_0$, such that the following conditions hold:*

(i) *The spaces V_j are nested, i.e.,*

$$\cdots V_{-1} \subset V_0 \subset V_1 \cdots .$$

(ii) $\overline{\cup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$ *and* $\cap_{j\in\mathbb{Z}} V_j = \{0\}$.

(iii) *For all $j \in \mathbb{Z}$, $V_{j+1} = D(V_j)$.*

(iv) $f \in V_0 \Rightarrow T_k f \in V_0, \ \forall k \in \mathbb{Z}$.

(v) $\{T_k\phi\}_{k\in\mathbb{Z}}$ *is an orthonormal basis for V_0.*

The condition $\overline{\cup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$ in Definition 8.2.1(ii) means that $\cup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$, i.e., that for any $f \in L^2(\mathbb{R})$ and any $\epsilon > 0$ we can find a function $g \in \cup_{j\in\mathbb{Z}} V_j$ such that $||f - g|| \le \epsilon$. The function g will belong to some space $V_J, \ J \in \mathbb{Z}$, and therefore by condition (i) to all spaces V_j with $j \ge J$.

We also note that the condition (iii) means that

$$V_{j+1} = \{Df \mid f \in V_j\}.$$

A closer look at the conditions in Definition 8.2.1 reveals that the choice of the function ϕ in a multiresolution analysis actually determines the spaces V_j uniquely:

Lemma 8.2.2 (The spaces V_j) *Assume that the conditions (iii) and (iv) in Definition 8.2.1 are satisfied. Then the following hold:*

(i) $V_j = D^j(V_0)$ *for all* $j \in \mathbb{Z}$.

(ii) $V_j = \overline{\text{span}}\{D^j T_k \phi\}_{k\in\mathbb{Z}}$ *for all* $j \in \mathbb{Z}$.

Proof. For $j \in \mathbb{N}$, an iteration of (iii) in Definition 8.2.1 yields that

$$V_j = D(V_{j-1}) = DD(V_{j-2}) = \cdots = D^j(V_0).$$

For the case where $j \in \{-1, -2, \dots\}$,

$$V_j = D^{-1}(V_{j+1}) = D^{-1}D^{-1}(V_{j+2}) = \cdots = \left(D^{-1}\right)^{-j}(V_0) = D^j(V_0).$$

This proves (i). For the proof of (ii), the condition (v) in Definition 8.2.1 implies by Theorem 4.7.2 (with $\mathcal{H} := V_0$) that $V_0 = \overline{\text{span}}\{T_k\phi\}_{k\in\mathbb{Z}}$. Now, using (i), we conclude by Exercise 2.13(ii) that

$$V_j = D^j(V_0) = D^j(\overline{\text{span}}\{T_k\phi\}_{k\in\mathbb{Z}}) = \overline{\text{span}}\{D^jT_k\phi\}_{k\in\mathbb{Z}}. \tag{8.3}$$

This concludes the proof. □

Lemma 8.2.2(ii) shows that the spaces V_j in a multiresolution analysis are uniquely determined by the function ϕ. For this reason we say that the function ϕ *generates the multiresolution analysis.* But we have to stress the fact that only very special functions ϕ can generate a multiresolution analysis. We will come back to this issue in Theorem 8.2.11.

Our first example of a multiresolution analysis is closely related with the Haar wavelet in Example 8.1.2:

Example 8.2.3 (Haar multiresolution analysis) We can define a multiresolution analysis by

$$\begin{cases} \phi := \chi_{[0,1[}; \\ V_j := \{f \in L^2(\mathbb{R}) : f \text{ is constant on } [2^{-j}k, 2^{-j}(k+1)[,\ \forall k \in \mathbb{Z}\}. \end{cases}$$

We leave the verification of the details to the reader (Exercise 8.1).

It turns out that (up to an irrelevant multiplicative factor of minus 1) the wavelet associated with this multiresolution analysis is the Haar wavelet in (8.2). The technical tools to show this are presented later in this section, see Exercise 8.2. Note that the Haar wavelet can be written as

$$\begin{aligned} \psi(x) &= \chi_{[0,1/2[}(x) - \chi_{[1/2,1[}(x) \\ &= \chi_{[0,1[}(2x) - \chi_{[0,1[}(2x-1) \\ &= \frac{1}{\sqrt{2}}\left(D\chi_{[0,1[}(x) - DT_1\chi_{[0,1[}(x)\right). \end{aligned} \tag{8.4}$$

Thus, the Haar wavelet is a finite linear combination of scaled and translated versions of the function $\phi = \chi_{[0,1[}$. In Theorem 8.2.7 and Proposition 8.2.8 we will see that a similar result holds for any wavelet derived from a multiresolution analysis. The multiresolution analysis considered here is naturally called the *Haar multiresolution analysis.* The function $\phi = \chi_{[0,1[}$ is called the *Haar scaling function.* □

We will now describe how a multiresolution analysis can be used to construct an orthonormal basis for $L^2(\mathbb{R})$. For this purpose we need to consider a class of vector spaces associated with $\{V_j\}_{j\in\mathbb{Z}}$:

Definition 8.2.4 (The spaces W_j) *Assume that V_j is a sequence of closed subspaces of $L^2(\mathbb{R})$ and that the condition (i) in Definition 8.2.1 is satisfied. For any $j \in \mathbb{Z}$, let W_j denote the orthogonal complement of V_j with respect to V_{j+1}, i.e.,*

$$W_j := \{f \in V_{j+1} \mid \langle f, g \rangle = 0, \ \forall g \in V_j\}.$$

We denote the orthogonal projection of $L^2(\mathbb{R})$ onto W_j by Q_j.

It turns out that the space W_0 plays a very special role in wavelet analysis. In fact, the next result shows that in order to obtain an orthonormal basis $\{D^j T_k \psi\}_{j,k \in \mathbb{Z}}$ for $L^2(\mathbb{R})$, it is enough to find a function $\psi \in W_0$ such that $\{\psi(\cdot - k)\}_{k \in \mathbb{Z}}$ is an orthonormal basis for W_0. The proof requires some preparation, and will be given on page 187 in Chapter 9.

Proposition 8.2.5 *Assume that the function $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis. Let $\psi \in L^2(\mathbb{R})$ and suppose that $\{T_k \psi\}_{k \in \mathbb{Z}}$ is an orthonormal basis for W_0. Then the following hold:*

(i) For each $j \in \mathbb{Z}$, the functions $\{D^j T_k \psi\}_{k \in \mathbb{Z}}$ form an orthonormal basis for W_j.

(ii) The functions $\{D^j T_k \psi\}_{j,k \in \mathbb{Z}}$ form an orthonormal basis for $L^2(\mathbb{R})$, i.e., ψ is a wavelet.

(iii) The functions $\{T_k \phi\}_{k \in \mathbb{Z}} \cup \{D^j T_k \psi\}_{j \in \mathbb{N}, k \in \mathbb{Z}}$ form an orthonormal basis for $L^2(\mathbb{R})$.

By Proposition 8.2.5 we can find a wavelet by constructing a function $\psi \in L^2(\mathbb{R})$ such that $\{T_k \psi\}_{k \in \mathbb{Z}}$ is an orthonormal basis for the space W_0. Conceptually this is a simplification: the wavelet system $\{D^j T_k \psi\}_{j,k \in \mathbb{Z}}$ involves the operations of scaling and translation, while $\{T_k \psi\}_{k \in \mathbb{Z}}$ just consists of translates. The following result is a key step in the construction of an appropriate function ψ:

Proposition 8.2.6 (Scaling equation) *Assume that the function $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis. Then there exists a 1-periodic function $H_0 \in L^2(0,1)$ such that*

$$\hat{\phi}(2\gamma) = H_0(\gamma)\hat{\phi}(\gamma), \ \gamma \in \mathbb{R}. \tag{8.5}$$

Proof. By assumption, the functions $\{T_k \phi\}_{k \in \mathbb{Z}}$ form an orthonormal basis for V_0; in particular, $\phi \in V_0$. Since $V_0 \subset V_1 = D(V_0)$, this means that $\phi \in D(V_0)$, i.e., that $D^{-1}\phi \in V_0$. Because V_0 is a vector space, it follows that

$$\frac{1}{\sqrt{2}}D^{-1}\phi \in V_0.$$

Since $\{T_{-k}\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0, Theorem 4.7.2 shows that there exist coefficients $\{c_k\}_{k\in\mathbb{Z}}$ such that

$$\frac{1}{\sqrt{2}}D^{-1}\phi = \sum_{k\in\mathbb{Z}} c_k T_{-k}\phi. \tag{8.6}$$

By Exercise 4.32 the coefficients $\{c_k\}_{k\in\mathbb{Z}}$ belong to $\ell^2(\mathbb{Z})$. Applying the Fourier transform and the result in Exercise 2.14(ii),

$$\begin{aligned}\frac{1}{\sqrt{2}}\mathcal{F}D^{-1}\phi &= \mathcal{F}\sum_{k\in\mathbb{Z}} c_k T_{-k}\phi \\ &= \sum_{k\in\mathbb{Z}} c_k \mathcal{F}T_{-k}\phi.\end{aligned}$$

Now, the commutation relations in Theorem 7.1.2(iii) and Lemma 7.1.3 imply that

$$\frac{1}{\sqrt{2}}D\hat{\phi}(\gamma) = \sum_{k\in\mathbb{Z}} c_k E_k(\gamma)\hat{\phi}(\gamma).$$

Thus,

$$\hat{\phi}(2\gamma) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi ik\gamma}\hat{\phi}(\gamma).$$

Defining the 1-periodic function

$$H_0(\gamma) := \sum_{k\in\mathbb{Z}} c_k e^{2\pi ik\gamma}, \tag{8.7}$$

this shows that (8.5) is satisfied. □

The equation (8.5) is called a *scaling equation* or *refinement equation.* A function ϕ that satisfies a scaling equation is called a *scaling function,* or said to be *refinable.* Formulated in this language, Proposition 8.2.6 says that a necessary condition for a function ϕ to generate a multiresolution analysis is that ϕ is a scaling function. In Theorem 8.2.11 we will see that two extra conditions on a scaling function imply that it actually generates a multiresolution analysis.

We are now ready to present the main result for construction of an orthonormal basis via a multiresolution analysis. The result itself can be formulated based on knowledge of the function H_0 in the scaling equation. The proof is given in Section 9.3.

Theorem 8.2.7 (Wavelet orthonormal basis) *Assume that* $\phi \in L^2(\mathbb{R})$ *generates a multiresolution analysis, and let* $H_0 \in L^2(0,1)$ *be a 1-periodic function satisfying the scaling equation* (8.5). *Define the 1-periodic function* H_1 *by*

$$H_1(\gamma) := \overline{H_0(\gamma + \frac{1}{2})}\, e^{-2\pi i\gamma}. \tag{8.8}$$

Also, define the function ψ *via*

$$\hat{\psi}(2\gamma) := H_1(\gamma)\hat{\phi}(\gamma). \tag{8.9}$$

Then the following hold:

(i) $\{T_k\psi\}_{k\in\mathbb{Z}}$ *is an orthonormal basis for* W_0.

(ii) $\{D^jT_k\psi\}_{j,k\in\mathbb{Z}}$ *is an orthonormal basis for* $L^2(\mathbb{R})$, *i.e.,* ψ *is a wavelet.*

The definition in (8.9) is quite indirect: it defines the function ψ in terms of its Fourier transform, so we have to apply the inverse Fourier transform in order to obtain an expression for ψ. This actually leads to an explicit expression of the function ψ in terms of the given function ϕ:

Proposition 8.2.8 (Explicit expression for the wavelet) *Assume that* (8.9) *holds for a 1-periodic function* $H_1 \in L^2(0,1)$,

$$H_1(\gamma) = \sum_{k\in\mathbb{Z}} d_k e^{2\pi ik\gamma}. \tag{8.10}$$

Then

$$\psi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} d_k DT_{-k}\phi(x) = 2\sum_{k\in\mathbb{Z}} d_k\phi(2x+k),\ \ x \in \mathbb{R}. \tag{8.11}$$

Proof. We can rewrite (8.9) as

$$\hat{\psi}(\gamma) = H_1(\gamma/2)\hat{\phi}(\gamma/2);$$

formulated in terms of the dilation operator D and the modulation operators, this means that

$$\mathcal{F}\psi(\gamma) = \sum_{k\in\mathbb{Z}} d_k e^{\pi ik\gamma}\mathcal{F}\phi(\gamma/2) = \sqrt{2}\sum_{k\in\mathbb{Z}} d_k E_{k/2}D^{-1}\mathcal{F}\phi(\gamma).$$

Using Lemma 7.1.3 this yields that

$$\mathcal{F}\psi(\gamma) = \sqrt{2}\sum_{k\in\mathbb{Z}} d_k E_{k/2}\mathcal{F}D\phi(\gamma),$$

Invoking the commutation relations in Theorem 7.1.2 and Lemma 6.2.4 and using the linearity of the Fourier transform now leads to

$$\begin{aligned}\mathcal{F}\psi(\gamma) &= \sqrt{2}\sum_{k\in\mathbb{Z}} d_k \mathcal{F}T_{-k/2}D\phi(\gamma)\\ &= \sqrt{2}\mathcal{F}\sum_{k\in\mathbb{Z}} d_k T_{-k/2}D\phi(\gamma)\\ &= \sqrt{2}\mathcal{F}\sum_{k\in\mathbb{Z}} d_k DT_{-k}\phi(\gamma).\end{aligned}$$

Applying the inverse Fourier transform finally yields the desired result. □

The results obtained so far show that it is easy to find the wavelet ψ whenever the function H_0 in the scaling equation has been calculated. Indeed, based on the expression (8.7) for the function H_0 we use (8.8) to find the function H_1 and bring it on the form (8.10). Hereafter we simply insert the coefficients d_k in (8.11).

In most cases of practical interest, H_0 is actually a trigonometric polynomial,

$$H_0(\gamma) = \sum_{k=-N}^{N} c_k e^{2\pi i k\gamma}. \tag{8.12}$$

The explicit expression for the wavelet in (8.11) immediately leads to a criterion for how to obtain a compactly supported wavelet:

Corollary 8.2.9 (Compactly supported wavelet) *Assume that the function $\phi \in L^2(\mathbb{R})$ is compactly supported and generates a multiresolution analysis. Assume further that the function H_0 in the scaling equation (8.5) is a trigonometric polynomial. Then the wavelet ψ in (8.11) is compactly supported.*

Proof. If H_0 is a trigonometric polynomial, the function H_1 in (8.8) is also a trigonometric polynomial. Thus, we can write

$$H_1(\gamma) = \sum_{k=N_1}^{N_2} d_k e^{2\pi i k\gamma}$$

for suitable coefficients $\{d_k\}_{k\in\mathbb{Z}}$. The proof of Proposition 8.2.8 now shows that the wavelet is

$$\psi(x) = 2\sum_{k=N_1}^{N_2} d_k \phi(2x+k),\ \ x\in\mathbb{R},$$

i.e., a compactly supported function. □

Example 8.2.10 (Haar wavelet) Theorem 8.2.7 and Proposition 8.2.8 lead to the Haar wavelet (up to an irrelevant multiplicative factor of minus 1) whenever they are applied to the Haar multiresolution analysis. We ask the reader to provide the proof in Exercise 8.2. □

As we have seen on page 163, the choice of a scaling function ϕ characterizes a multiresolution analysis uniquely. This makes it natural to examine how to formulate the multiresolution analysis conditions directly in terms of conditions on the function ϕ. Such conditions are presented now:

Theorem 8.2.11 (Construction of multiresolution analysis) *Let $\phi \in L^2(\mathbb{R})$. Define the spaces V_j by (8.3), and assume that the following conditions hold:*

(i) $\inf_{\gamma \in]-\epsilon,\epsilon[} |\hat{\phi}(\gamma)| > 0$ *for some* $\epsilon > 0$;

(ii) The scaling equation

$$\hat{\phi}(2\gamma) = H_0(\gamma)\hat{\phi}(\gamma), \tag{8.13}$$

is satisfied for a bounded 1-periodic function H_0;

(iii) $\{T_k\phi\}_{k\in\mathbb{Z}}$ *is an orthonormal system.*

Then ϕ generates a multiresolution analysis.

The proof of Theorem 8.2.11 is given in Section 9.4. We see that we still need a characterization of the functions $\phi \in L^2(\mathbb{R})$ for which $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal system. Such a characterization is given now:

Theorem 8.2.12 (Characterization of orthonormal system $\{T_k\phi\}_{k\in\mathbb{Z}}$) *Let $\phi \in L^2(\mathbb{R})$. Then $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal system if and only if*

$$\sum_{k\in\mathbb{Z}} |\hat{\phi}(\gamma+k)|^2 = 1, \ \gamma \in \mathbb{R}.$$

A guide to a proof of Theorem 8.2.12 can be found in Exercise 8.11.

8.3 Vanishing moments and the Daubechies' wavelets

The purpose of this section is to discuss certain properties that make wavelets useful in signal processing, and to present the construction by Daubechies.

Assume that the wavelet ψ comes from a multiresolution analysis generated by the function ϕ. By Proposition 8.2.5, the functions

$$\{T_k\phi\}_{k\in\mathbb{Z}} \cup \{D^jT_k\psi\}_{j\in\mathbb{N},k\in\mathbb{Z}}$$

form an orthonormal basis for $L^2(\mathbb{R})$, so we know from Theorem 4.7.2 that any $f \in L^2(\mathbb{R})$ has the representation

$$f = \sum_{k\in\mathbb{Z}} \langle f, T_k\phi\rangle T_k\phi + \sum_{j=1}^{\infty}\sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}. \tag{8.14}$$

All information about the function f is stored in the coefficients

$$\{\langle f, T_k\phi\rangle\}_{k\in\mathbb{Z}} \cup \{\langle f, \psi_{j,k}\rangle\}_{j\in\mathbb{N},k\in\mathbb{Z}}, \tag{8.15}$$

and (8.14) tells us how to reconstruct f based on these coefficients. In practice one cannot store an infinite sequence of nonzero numbers, so one has to select a finite number of the coefficients to keep. In most situations of practical interest, only a finite number of entries in the sequence $\{\langle f, T_k\phi\rangle\}_{k\in\mathbb{Z}}$ are nonzero (see Exercise 8.4). We express this by saying that $\{\langle f, T_k\phi\rangle\}_{k\in\mathbb{Z}}$ is a *finite sequence.* Thus, the problem is how to deal with the infinite sequence $\{\langle f, \psi_{j,k}\rangle\}_{j\in\mathbb{N},k\in\mathbb{Z}}$. This is usually done by *thresholding*: that is, we choose a certain $\epsilon > 0$ and keep only the coefficients in (8.15) for which

$$|\langle f, \psi_{j,k}\rangle| \geq \epsilon. \tag{8.16}$$

By Exercise 4.32 the coefficients $\{\langle f, \psi_{j,k}\rangle\}_{j,k\in\mathbb{Z}}$ belong to ℓ^2. Thus, only a *finite* number of indices $(j,k) \in \mathbb{Z}\times\mathbb{Z}$ will satisfy (8.16). A key feature of wavelet theory is that it often is possible to choose the wavelet ψ such that many of the coefficients $\{\langle f, \psi_{j,k}\rangle\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ are small for the relevant signals f, i.e., a relatively small number of coefficients will satisfy (8.16). Let us for the moment consider the Haar wavelet:

Example 8.3.1 (Haar wavelet) Let $\phi = \chi_{[0,1[}$. By Exercise 8.2, the multiresolution analysis generated by ϕ leads to the Haar wavelet considered in Example 8.1.2. Consider the terms in the second infinite sum in (8.14),

$$\langle f, \psi_{j,k}\rangle \psi_{j,k}(x) = 2^{j/2}\langle f, \psi_{j,k}\rangle \psi(2^j x - k).$$

Direct calculation (Exercise 8.7) shows that

$$\begin{aligned} d_{j,k} &:= 2^{j/2}\langle f, \psi_{j,k}\rangle \\ &= \frac{1}{2}\big(\text{average of } f \text{ over } 2^{-j}[k, k+1/2[\\ &\quad - \text{average of } f \text{ over } 2^{-j}[k+1/2, k+1[\big). \end{aligned} \tag{8.17}$$

Thus, the coefficient $d_{j,k}$ is directly related to the behavior of the function f on the interval $[2^{-j}k, 2^{-j}(k+1)[$. If the function f is continuous, the calculation (8.17) implies that many coefficients $d_{j,k}$ are small: since

$$\langle f, \psi_{j,k}\rangle = 2^{-j/2}d_{j,k},\ \ j \geq 1, k \in \mathbb{Z},$$

this implies that only few coefficients $\{\langle f, \psi_{j,k}\rangle\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ satisfy (8.16).

Slight modifications extend the above argument to discontinuous functions. Assume that f is continuous, except at a point x_0. Then if j, k are chosen such that $x_0 \notin 2^{-j}[k, k+1[$, the above argument still works. Thus, many of the associated coefficients $d_{j,k}$ will be small. On the other hand, if j, k are chosen such that $x_0 \in 2^{-j}[k, k+1[$, then $d_{j,k}$ can be expected to be approximately half the size of the jump

$$f(x_0^+) - f(x_0^-),$$

at least whenever j is large (Exercise 8.7). Thus, large values for the coefficients $d_{j,k}$ that persist through the scales, indicate a discontinuity in the function f. In other words: just by looking at the coefficients $d_{j,k}$ we can *detect* that there is a discontinuity in the function f! □

As explained above, it is desirable to consider a wavelet ψ for which the sequence $\{\langle f, \psi_{j,k}\rangle\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ contains many small entries for the relevant signals f. There exist wavelets that perform much better than the Haar wavelet in that regard. The key property turns out to be the *vanishing moments:*

Definition 8.3.2 (Vanishing moments) *Let $N \in \mathbb{N}$. A function ψ has N vanishing moments if*

$$\int_{-\infty}^{\infty} x^\ell \psi(x)\, dx = 0 \quad \textit{for} \quad \ell = 0, 1, \dots, N-1.$$

The Haar wavelet has only one vanishing moment (Exercise 8.6). If ψ has a large number of vanishing moments, the following result shows that only relatively few coefficients $\langle f, \psi_{j,k}\rangle$ will be large. A proof can be found, e.g., in [18].

Theorem 8.3.3 (Decay of wavelet coefficients) *Assume that the function $\psi \in L^2(\mathbb{R})$ is compactly supported and has N vanishing moments. Then, for any N times differentiable function $f \in L^2(\mathbb{R})$ for which the Nth derivative $f^{(N)}$ is bounded, there exists a constant $C > 0$ such that*

$$|\langle f, \psi_{j,k}\rangle| \leq C\, 2^{-jN} 2^{-j/2}, \ \forall j \geq 1, \ k \in \mathbb{Z}. \tag{8.18}$$

Looking at the estimate (8.18), we see that a high number of vanishing moments implies that the numbers $\langle f, \psi_{j,k}\rangle$ decay quickly as $j \to \infty$: the higher number of vanishing moments a wavelet has, the fewer coefficients $\{\langle f, \psi_{j,k}\rangle\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ in (8.15) remain after the thresholding.

It turns out that a condition for obtaining wavelets with a certain number of vanishing moments can be expressed in terms of the function H_0 appearing in the scaling equation (8.5). The proof of the following result can be found in [18, Theorem 9.11].

Theorem 8.3.4 (Vanishing moments) *Let ϕ be a compactly supported scaling function associated with a multiresolution analysis, and let ψ be the associated wavelet as in (8.9). Then the following are equivalent:*

(i) ψ has N vanishing moments;

(ii) The function H_0 can be factorized

$$H_0(\gamma) = \left(\frac{1+e^{-2\pi i\gamma}}{2}\right)^N L(\gamma) \tag{8.19}$$

for some 1-periodic trigonometric polynomial L.

Combining Theorems 8.2.11 and 8.3.4, we see that one approach to construct a wavelet ψ having N vanishing moments is to search for a 1-periodic function $H_0 \in L^2(0,1)$ for which

- The function H_0 is associated with a multiresolution analysis, generated by a compactly supported function ϕ; in particular, the scaling equation (8.5) should be satisfied.
- The condition (8.19) is satisfied for some 1-periodic trigonometric polynomial L.

As soon as we have the function H_0 and the associated scaling function ϕ at hand, we define the wavelet ψ as in Theorem 8.2.7. Observe that if we have determined a function H_0, we can actually find the only (up to a scalar multiplication) candidate for the associated scaling function ϕ. In fact, the scaling equation shows that

$$\begin{aligned}\hat{\phi}(\gamma) = H_0(\gamma/2)\hat{\phi}(\gamma/2) &= H_0(\gamma/2)H_0(\gamma/4)\hat{\phi}(\gamma/4) \\ &= H_0(\gamma/2)H_0(\gamma/4)H_0(\gamma/8)\hat{\phi}(\gamma/8).\end{aligned}$$

Iterating this equation shows that for any $K \in \mathbb{N}$,

$$\hat{\phi}(\gamma) = H_0(\gamma/2)H_0(\gamma/4)H_0(\gamma/8)\cdots H_0(\gamma/2^K)\hat{\phi}(\gamma/2^K).$$

Letting

$$\prod_{j=1}^{K} H_0(\gamma/2^j) := H_0(\gamma/2)H_0(\gamma/4)H_0(\gamma/8)\cdots H_0(\gamma/2^K),$$

this result can be written

$$\hat{\phi}(\gamma) = \hat{\phi}(\gamma/2^K)\,\prod_{j=1}^{K} H_0(\gamma/2^j).$$

One can prove that $\hat{\phi}$ is continuous at zero and that $\prod_{j=1}^{K} H_0(\gamma/2^j)$ has a limit as $K \to \infty$. Writing

$$\prod_{j=1}^{\infty} H_0(\gamma/2^j) := \lim_{K\to\infty} \prod_{j=1}^{K} H_0(\gamma/2^j),$$

it follows that

$$\begin{aligned}\hat{\phi}(\gamma) &= \lim_{K\to\infty}\left(\hat{\phi}(\gamma/2^K)\prod_{j=1}^{K}H_0(\gamma/2^j)\right)\\ &= \hat{\phi}(0)\prod_{j=1}^{\infty}H_0(\gamma/2^j).\end{aligned}$$

This shows that as soon as we have fixed a value for $\hat{\phi}(0)$, all other values for the function $\hat{\phi}$ are determined; that is, up to a scalar multiplication, the function ϕ is unique. The remaining work now consists in a proof of the fact that the function ϕ actually generates a multiresolution analysis, and this can be done using Theorem 8.2.11.

The *Daubechies' wavelets* are the best known constructions based on the above idea. Except for $N = 1$, the Daubechies' wavelets are not given by an explicit formula, but it is known that the smoothness of the wavelets increases with N. The construction of wavelets by Daubechies is based on a family of polynomials $P_{N-1},\ N \in \mathbb{N}$, given by

$$P_{N-1}(y) = \sum_{k=0}^{N-1}\frac{(2N-1)!}{k!(2N-1-k)!}\,y^k(1-y)^{N-1-k}. \tag{8.20}$$

The construction works as follows:

Theorem 8.3.5 (Daubechies' wavelets) *For any $N \in \mathbb{N}$, there exists a trigonometric polynomial L such that*

$$|L(\gamma)|^2 = P_{N-1}(\sin^2\pi\gamma). \tag{8.21}$$

With such a choice for L, the following hold:

(i) The function H_0 given by (8.19) is associated with a multiresolution analysis;

(ii) With H_0 as in (i), the wavelet ψ given in Theorem 8.2.7 has N vanishing moments, and support in $[0, 2N-1]$.

The proof of Theorem 8.3.5 can be found in [8].

Example 8.3.6 It is easy to calculate the first few polynomials P_{N-1} explicitly:

- $P_0(y) = 1$
- $P_1(y) = 1 + 2y$
- $P_2(y) = 1 + 3y + 6y^2$. □

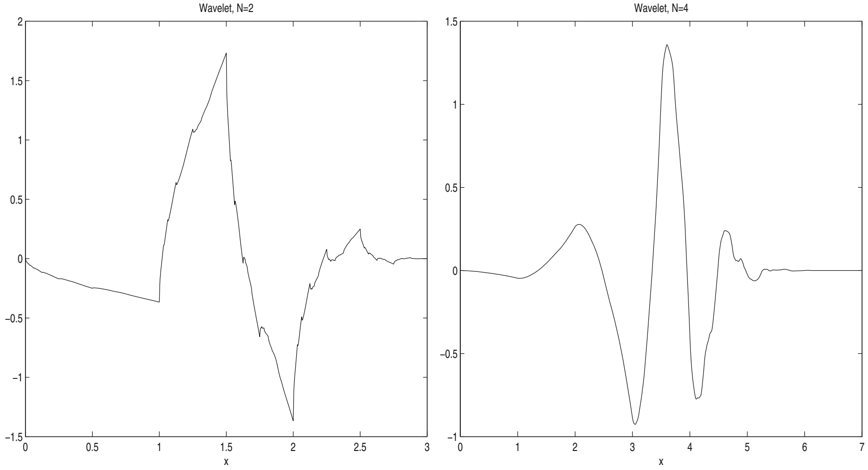

Figure 8.1. Some of Daubechies' wavelets.

In order to apply Theorem 8.3.5, we need to be able to find the trigonometric polynomial L satisfying (8.21). There exists a procedure, called *spectral factorization,* to do this; see [8]. The special case $N = 1$ is easy:

Example 8.3.7 For $N = 1$, the condition (8.21) means that $|L(\gamma)|^2 = 1$, which is satisfied for $L(\gamma) = -1$ (we could also have taken $L(\gamma) = 1$). Via (8.19) this leads to

$$H_0(\gamma) = \frac{1 + e^{-2\pi i\gamma}}{2}\, L(\gamma) = \frac{-1 - e^{-2\pi i\gamma}}{2}.$$

Following Theorem 8.2.7, we now calculate

$$\begin{aligned} H_1(\gamma) &= \overline{H_0(\gamma + \frac{1}{2})}\, e^{-2\pi i\gamma} \\ &= \overline{\frac{-1 - e^{-2\pi i(\gamma+1/2)}}{2}}\, e^{-2\pi i\gamma} \\ &= \frac{1 - e^{-2\pi i\gamma}}{2}. \end{aligned}$$

That is,

$$H_1(\gamma) = \sum_{k=-1}^{0} d_k e^{2\pi i k x},$$

where $d_0 = 1/2, d_{-1} = -1/2$. By Proposition 8.2.8, the associated wavelet is

$$\begin{aligned}\psi(x) = \sqrt{2}\sum_{k=-1}^{0} d_k DT_{-k}\phi(x) &= \frac{1}{\sqrt{2}}\left(\sqrt{2}\phi(2x) - \sqrt{2}\phi(2x-1)\right)\\ &= \phi(2x) - \phi(2x-1).\end{aligned}$$

This is exactly the expression for the Haar wavelet in Examples 8.1.2 and 8.2.3. □

8.4 Wavelets and signal processing

We will now shortly discuss how the wavelet results discussed in the previous sections play a role in various areas of signal processing.

Example 8.4.1 (Image compression) In Example 8.3.1 we saw how the wavelet representation (8.14) can be used to detect singularities in the signal f: large coefficients $d_{j,k} := 2^{j/2}\langle f, \psi_{j,k}\rangle$ that persist for all scales j suggest the presence of a discontinuity in f.

The above principle is applied in image compression. Black/white images are defined by the collection of their pixels, together with the values for the color intensity at each pixel; the intensity measures the color on a scale going from for example 0 (completely white) to $512 = 2^9$ (completely black). When performing a wavelet analysis on an image, it is done separately on each row of pixels, one by one. Images have large parts of almost constant intensity that can be interpreted as the smooth part of the signal, separated by edges, that can be interpreted as the non-smooth part. Performing a wavelet analysis with a wavelet ψ that has a high number of vanishing moments, considerations like the ones we did in Example 8.3.1 show that the wavelet representation of the pixels in each row of the image will have a few large coefficients corresponding to the edges, and many small coefficients corresponding to regions with small changes. This implies that many coefficients can be discarded by thresholding, i.e., that an efficient compression is obtained. □

Due to the efficiency of wavelets as described in Example 8.4.1, wavelets are implemented in, e.g., digital cameras. In fact, wavelets are part of the so-called *JPEG2000 standard* for image compression. They are also used by the FBI (Federal Bureau for Investigation) to store fingerprints electronically:

Example 8.4.2 (Fingerprints) With reasonable accuracy, it uses about 10 Mb to store a fingerprint electronically. The FBI has more than 30 million sets of fingerprints (each consisting of 10 fingers) and receives about

30000 new fingerprints each day. We are speaking about enormous data sets, and it is necessary to do some compression in order to be able to handle them. This has to be done in a way such that the structure of the fingerprints is kept. The FBI uses a variant of Daubechies' wavelets discussed in Section 8.3 for that purpose. In the concrete case it is enough to represent a fingerprint using about 8% of the original information. The compression method has the impressive name *The Wavelet Scalar Quantization Gray-scale Fingerprint Image Compression Algorithm,* usually abbreviated WSQ. □

We note that compression of fingerprints (or any other image) in principle also can be performed using Fourier analysis. However, Fourier methods are bad at representing edges (think, e.g., of the Gibb's phenomena for the Fourier series of a function having a discontinuity). As a consequence, this classical method is less efficient: at the rate of compression the FBI uses for the wavelet method, it would no longer be possible to follow the contours in a reconstructed fingerprint, and the result would be useless in this special context.

Example 8.4.3 (Noise reduction) Assume that the signal f represents a piece of music. Looking again at the representation (8.14), it is natural to consider the functions $\psi_{j,k}$ to contain the high-frequency information of f for j large and the low-frequency information for j small. Noise is often contained in the high-frequency range. Thus, by replacing (8.14) by

$$f = \sum_{k\in\mathbb{Z}} \langle f, T_k\phi\rangle T_k\phi + \sum_{j=1}^{J}\sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}$$

for an appropriately chosen value of J, an efficient noise reduction of the signal f is obtained. Observe that J needs to be chosen with great care: if J is too small, a substantial part of the information in the signal (i.e., music piece) is removed as well!

An application of this principle led to an amazing conclusion within music science. In 1889, the composer Johannes Brahms "recorded" his playing of one of his piano pieces using piano rolls. Later, the recording was transferred to an LP-disc; unfortunately, the result was so noisy that it was impossible to hear that it was a piano recording! Using wavelet methods, J. Berger and C. Nichols managed to get rid of so much noise that it became possible to hear the music and obtain an idea about how Brahms played his own compositions. Somewhat surprisingly it turned out that he did not follow his own score very closely, but rather took the score as a starting point for an improvisation. □

8.5 Exercises

8.1 (Haar multiresolution analysis) Let

$$\begin{cases} \phi := \chi_{[0,1[}; \\ V_0 := \{f \in L^2(\mathbb{R}) : f \text{ is constant on } [k, k+1[, \ \forall k \in \mathbb{Z}\}. \end{cases}$$

(i) Describe the spaces $V_j := D^j(V_0)$.

(ii) Show that the spaces V_j and the function ϕ satisfy the conditions (i) and (iii)–(v) in the definition of a multiresolution analysis.

(iii) Use Corollary 9.1.3 and Proposition 9.4.3 to argue that the spaces V_j satisfy condition (ii) in the definition of a multiresolution analysis.

8.2 (Haar multiresolution analysis) For the multiresolution analysis in Exercise 8.1, calculate the functions H_0 and H_1 in Theorem 8.2.7 and find the expression for the wavelet ψ.

8.3 Assume that two subspaces V_0 and V_1 of $L^2(\mathbb{R})$ satisfy that

(i) $V_1 = D(V_0)$;

(ii) $f \in V_0 \Rightarrow T_k f \in V_0, \ \forall k \in \mathbb{Z}$.

Show that if $g \in V_1$, then $T_k g \in V_1$ for all $k \in \mathbb{Z}$.

8.4 Assume that a compactly supported function $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis with associated wavelet ψ, and consider the expansion (8.14) of functions $f \in L^2(\mathbb{R})$. Argue that if f has compact support, the term

$$\sum_{k \in \mathbb{Z}} \langle f, T_k \phi \rangle T_k \phi$$

is actually a finite sum.

8.5 Assume that $\phi \in L^2(\mathbb{R})$ satisfies the scaling equation

$$\hat{\phi}(2\gamma) = H_0(\gamma)\hat{\phi}(\gamma)$$

with

$$H_0(\gamma) = \sum_{k \in \mathbb{Z}} c_k e^{2\pi i k \gamma}.$$

According to Proposition 8.2.8 this means that

$$\phi(x) = 2\sum_{k\in\mathbb{Z}} c_k \phi(2x+k).$$

Now define the function ψ as in Theorem 8.2.7, and show that

$$\psi(x) = 2\sum_{\ell\in\mathbb{Z}} c_{-1-\ell}(-1)^{\ell+1}\phi(2x+\ell).$$

8.6 (Haar wavelet) Show that the number of vanishing moments for the Haar wavelet is $N = 1$.

8.7 (Haar wavelet) Let $\phi := \chi_{[0,1]}$ and let ψ denote the Haar wavelet. We consider the second term in the representation of functions $f \in L^2(\mathbb{R})$ in (8.14),

$$\sum_{j=1}^{\infty}\sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}(x) = \sum_{j=1}^{\infty}\sum_{k\in\mathbb{Z}} 2^{j/2}\langle f, \psi_{j,k}\rangle \psi(2^j x - k). \quad (8.22)$$

Assume that f is real-valued, and let

$$d_{j,k} := 2^{j/2}\langle f, \psi_{j,k}\rangle. \quad (8.23)$$

(i) For $j = 0, 1, 2$, make a sketch of the intervals

$$I_{j,k} := 2^{-j}[k, k+1[,\ \ k \in \mathbb{Z}.$$

(ii) Given an arbitrary $x \in \mathbb{R}$, we can only obtain a nonzero contribution in the sum in (8.22) for the values of $j \in \mathbb{N}$ and $k \in \mathbb{Z}$ for which $2^j x - k \in [0, 1[$. Show that for any $j \in \mathbb{N}$ this happens for exactly one value of $k \in \mathbb{Z}$.

(iii) Show that

$$\begin{aligned} d_{j,k} &= \frac{1}{2}\big(\text{average of } f \text{ over } 2^{-j}[k, k+1/2[\\ &\quad - \text{average of } f \text{ over } 2^{-j}[k+1/2, k+1[\big)\,. \end{aligned} \quad (8.24)$$

Hint: the average of a real function f over an interval $[a, b]$ is

$$\frac{1}{b-a}\int_a^b f(x)\,dx.$$

(iv) For $j = 0, 1, 2$, make a sketch of the intervals

$$M_{j,k} := 2^{-j}[k, k+1/2[,\ \ \tilde{M}_{j,k} := 2^{-j}[k+1/2, k+1[,\ \ k \in \mathbb{Z}.$$

The exercise continues!

(v) Assume that f is continuous at a given point $x \in \mathbb{R}$, and consider the values of $(j,k) \in \mathbb{N} \times \mathbb{Z}$ that give contributions in (8.22). Argue that

$$d_{j,k} \to 0 \text{ as } j \to \infty. \tag{8.25}$$

(vi) Explain in words how fast the convergence in (8.25) is for functions f that oscillate fast around x, compared to the speed of convergence for functions that are almost constant around x.

(vii) Assume that f is discontinuous at a given point $x \in \mathbb{R}$, and that $x = 2^{-j}(k + 1/2)$ for some $k \in \mathbb{Z}$ and a large value of $j \in \mathbb{N}$. Argue that we can expect that

$$|d_{j,k}| \approx \frac{1}{2} \left| f(x^+) - f(x^-) \right|.$$

(viii) Generalize (vii) as follows. Assume that f is discontinuous at a given point $x \in \mathbb{R}$ and assume that $2^j x - k \in [0,1[$ for some $k \in \mathbb{Z}$ and a large value of $j \in \mathbb{N}$. Argue that we can still expect $d_{j,k}$ to be "large," i.e., of the size

$$|d_{j,k}| \approx C \left| f(x^+) - f(x^-) \right|.$$

8.8 (Haar wavelet) We continue Exercise 8.7, now with the aim of a concrete analysis of the function

$$f(x) := \begin{cases} x & \text{if } x \in [0, \frac{3}{4}], \\ 0 & \text{if } x \notin [0, \frac{3}{4}]. \end{cases}$$

(i) Let $x = 1/2$, and consider the values of $j \in \mathbb{N}$ and $k \in \mathbb{Z}$ for which $2^j x - k \in [0,1[$ (see Exercise 8.7(i)). Show via (8.24) that for $j \geq 1$, the coefficients $d_{j,k}$ in (8.23) satisfy that

$$|d_{j,k}| = \frac{1}{4} 2^{-j}.$$

(ii) Let $x = 3/4$. Consider the values of $j \in \mathbb{N}$ and $k \in \mathbb{Z}$ for which $2^j x - k \in [0,1[$. Calculate $d_{j,k}$ in (8.23) for $j = 1, 2, 3$.

8.9 (Haar wavelet) We continue Exercise 8.7, now with the aim of a concrete analysis of the function

$$f(x) := \begin{cases} x & \text{if } x \in [0, \frac{\pi}{4}], \\ 0 & \text{if } x \notin [0, \frac{\pi}{4}]. \end{cases}$$

Let $x = \pi/4$. Consider the values of $j \in \mathbb{N}$ and $k \in \mathbb{Z}$ for which $2^j x - k \in [0,1[$. Calculate $d_{j,k}$ for $j = 1, 2, 3$.

8.10 Let $\phi \in L^2(\mathbb{R})$, and denote its Fourier transform by $\hat{\phi}$. Define the function Φ by

$$\Phi(\gamma) := \sum_{n \in \mathbb{Z}} |\hat{\phi}(\gamma + n)|^2.$$

The purpose of the exercise is to show that

$$\langle \phi, T_k \phi \rangle = \int_0^1 \Phi(\gamma) e^{2\pi i k \gamma} \, d\gamma, \ \ k \in \mathbb{Z}. \tag{8.26}$$

(i) Argue that $\Phi \in L^1(0,1)$.
Hint: use the result in Example 5.3.4.

(ii) Show that

$$\langle \phi, T_k \phi \rangle = \int_{-\infty}^{\infty} e^{2\pi i k \gamma} |\hat{\phi}(\gamma)|^2 \, d\gamma, \ \ k \in \mathbb{Z}.$$

Hint: use formula (7.13).

(iii) Show that

$$\int_{-\infty}^{\infty} e^{2\pi i k \gamma} |\hat{\phi}(\gamma)|^2 d\gamma = \sum_{n \in \mathbb{Z}} \int_0^1 e^{2\pi i k \gamma} |\hat{\phi}(\gamma + n)|^2 d\gamma.$$

Hint: use that

$$\int_{-\infty}^{\infty} e^{2\pi i k \gamma} |\hat{\phi}(\gamma)|^2 d\gamma = \sum_{n \in \mathbb{Z}} \int_n^{n+1} e^{2\pi i k \gamma} |\hat{\phi}(\gamma)|^2 d\gamma.$$

(iv) Show that for any $k \in \mathbb{Z}$,

$$\sum_{n \in \mathbb{Z}} \int_0^1 e^{2\pi i k \gamma} |\hat{\phi}(\gamma + n)|^2 d\gamma = \int_0^1 \Phi(\gamma) e^{2\pi i k \gamma} \, d\gamma, \ \ k \in \mathbb{Z}.$$

Hint: fix $k \in \mathbb{Z}$. Let

$$f_N(\gamma) := e^{2\pi i k \gamma} \sum_{n=-N}^{N} |\hat{\phi}(\gamma + n)|^2$$

and

$$f(\gamma) := e^{2\pi i k \gamma} \sum_{n \in \mathbb{Z}} |\hat{\phi}(\gamma + n)|^2,$$

and apply Theorem 5.3.5.

(v) Combine (ii)–(iv) to conclude the proof of (8.26).

8.11 Let $\phi \in L^2(\mathbb{R})$, and denote its Fourier transform by $\hat{\phi}$. The purpose of the exercise is to prove Theorem 8.2.12.

(i) Show that $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal system if and only if

$$\langle \phi, T_k\phi\rangle = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k \neq 0. \end{cases}$$

Let

$$\Phi(\gamma) := \sum_{n\in\mathbb{Z}} |\hat{\phi}(\gamma + n)|^2.$$

One can show (Exercise 8.10) that

$$\langle \phi, T_k\phi\rangle = \int_0^1 \Phi(\gamma) e^{2\pi i k\gamma}\, d\gamma.$$

(ii) Show that $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal system if and only if

$$\Phi(\gamma) = 1, \ \gamma \in \mathbb{R}.$$

Hint: by Example 6.4.3, the numbers $\int_0^1 \Phi(\gamma)e^{2\pi ik\gamma}\, d\gamma$ are Fourier coefficients for the function Φ.

8.12 Assume that the function $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis. Show that the 1-periodic function $H_0 \in L^2(0,1)$ satisfying the scaling equation (8.5) is uniquely determined. *Hint:* by the result in Theorem 8.2.12 there exists for any $\gamma \in \mathbb{R}$ a $k \in \mathbb{Z}$ such that $\hat{\phi}(\gamma + k) \neq 0$. Now apply the scaling equation with γ replaced by $\gamma + k$.

9

A Closer Look at Multiresolution Analysis

In this chapter we provide the technical details in the construction of a wavelet orthonormal basis for $L^2(\mathbb{R})$. In Section 9.1 we prove a few immediate consequences of the conditions in a multiresolution analysis. Section 9.2 proves several results concerning the spaces V_j and W_j, and the construction of a wavelet is presented in Section 9.3. Finally, Section 9.4 proves that the conditions in Theorem 8.2.11 are sufficient for a function ϕ to generate a multiresolution analysis.

9.1 Basic properties of multiresolution analysis

On page 163 we saw that a multiresolution analysis is uniquely defined by the choice of the function ϕ. For this reason we will use the following definition throughout the chapter:

Definition 9.1.1 (The spaces V_j) *Given a function $\phi \in L^2(\mathbb{R})$, define the spaces $V_j,\ j \in \mathbb{Z}$, by*

$$V_j = \overline{span}\{D^j T_k \phi\}_{k\in\mathbb{Z}}. \tag{9.1}$$

We will denote the orthogonal projection of $L^2(\mathbb{R})$ onto V_j by P_j.

Recall that orthogonal projections are defined in Definition 4.5.5 and that their main properties are stated in Lemma 4.5.6.

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis,
DOI 10.1007/978-0-8176-4980-7_9,

Lemma 9.1.2 *Let $\phi \in L^2(\mathbb{R})$, and define the spaces V_j by (9.1). Assume that $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0, and fix any $j \in \mathbb{Z}$. Then the following hold:*

(i) $\{D^jT_k\phi\}_{k\in\mathbb{Z}}$ *is an orthonormal basis for V_j.*

(ii) $\lim_{j\to-\infty} P_jf = 0, \ \forall f \in L^2(\mathbb{R})$.

Proof. Fix $j \in \mathbb{Z}$. In order to prove (i), we first show that $\{D^jT_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal system. By Lemma 6.2.2, the operator D is unitary. For any $k, \ell \in \mathbb{Z}$, the results in Exercise 4.23 imply that

$$\langle D^jT_k\phi, D^jT_\ell\phi\rangle = \langle T_k\phi, T_\ell\phi\rangle = \delta_{k,\ell},$$

where we used that $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal system.

We have now verified that $\{D^jT_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal system. In order to show that $\{D^jT_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_j, it is enough to verify that the condition (ii) in Theorem 4.7.2 is satisfied with $\mathcal{H} := V_j$. Let $f \in V_j$. By Lemma 8.2.2, $f = D^jg$ for some $g \in V_0$. Because $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0, Theorem 4.7.2 shows that

$$g = \sum_{k\in\mathbb{Z}} \langle g, T_k\phi\rangle T_k\phi.$$

It follows that

$$f = D^jg = D^j\sum_{k\in\mathbb{Z}} \langle g, T_k\phi\rangle T_k\phi = \sum_{k\in\mathbb{Z}} \langle g, T_k\phi\rangle D^jT_k\phi,$$

where the result in Exercise 2.14 was used to move the operator D^j under the sum sign. Applying the results in Exercise 4.23 again yields that

$$\begin{aligned} f &= \sum_{k\in\mathbb{Z}} \langle D^jg, D^jT_k\phi\rangle D^jT_k\phi \\ &= \sum_{k\in\mathbb{Z}} \langle f, D^jT_k\phi\rangle D^jT_k\phi. \end{aligned}$$

Thus, condition (ii) in Theorem 4.7.2 is satisfied, which proves (i).

In order to prove (ii), we will first consider a function $f \in L^2(\mathbb{R})$ with compact support, $\operatorname{supp} f \subseteq [-N, N]$ for some $N > 0$. By (i) and the result in Exercise 4.29, the orthogonal projection of f on V_j is

$$P_jf = \sum_{k\in\mathbb{Z}} \langle f, D^jT_k\phi\rangle D^jT_k\phi.$$

Using the result in Exercise 4.31, it follows that

$$\begin{aligned} ||P_j f||_2^2 &= \sum_{k\in\mathbb{Z}} |\langle f, D^j T_k \phi\rangle|^2 \\ &= \sum_{k\in\mathbb{Z}} \left| \int_{-\infty}^{\infty} f(x)\overline{2^{j/2}\phi(2^j x-k)}\,dx \right|^2 \\ &= \sum_{k\in\mathbb{Z}} \left| \int_{-N}^{N} f(x)\overline{2^{j/2}\phi(2^j x-k)}\,dx \right|^2 . \end{aligned}$$

Using Cauchy–Schwarz' inequality and the change of variable $y = 2^j x - k$ it follows that

$$\begin{aligned} ||P_j f||_2^2 &\le \sum_{k\in\mathbb{Z}} \left(\int_{-N}^{N} |f(x)|^2\,dx \right) \left(\int_{-N}^{N} 2^j\,|\phi(2^j x-k)|^2\,dx \right) \\ &= ||f||_2^2 \sum_{k\in\mathbb{Z}} \int_{-2^j N-k}^{2^j N-k} |\phi(y)|^2\,dy. \end{aligned} \tag{9.2}$$

We want to show that the expression in (9.2) tends to 0 as $j \to -\infty$. Given $\epsilon > 0$, choose $K \in \mathbb{N}$ such that

$$\int_{|y|\ge K-1/2} |\phi(y)|^2\,dy \le \frac{\epsilon}{2}. \tag{9.3}$$

We can write

$$\begin{aligned} &\sum_{k\in\mathbb{Z}} \int_{-2^j N-k}^{2^j N-k} |\phi(y)|^2\,dy \\ &= \sum_{|k|<K} \int_{-2^j N-k}^{2^j N-k} |\phi(y)|^2\,dy + \sum_{|k|\ge K} \int_{-2^j N-k}^{2^j N-k} |\phi(y)|^2\,dy. \end{aligned} \tag{9.4}$$

We will analyze the two terms in (9.4) separately. First, for j chosen small enough, i.e., such that $2^j N < \frac{1}{2}$, we note that

$$[-2^j N - k, 2^j N - k] \subset [-\frac{1}{2} - k, \frac{1}{2} - k].$$

Thus, the choice of K in (9.3) implies that

$$\begin{aligned} \sum_{|k|\ge K} \int_{-2^j N-k}^{2^j N-k} |\phi(y)|^2\,dy &\le \sum_{|k|\ge K} \int_{-1/2-k}^{1/2-k} |\phi(y)|^2\,dy \\ &= \int_{|y|\ge K-1/2} |\phi(y)|^2\,dy \le \frac{\epsilon}{2}. \end{aligned}$$

We now look at the first term in (9.4), where we only need to worry about $k = -K+1, -K+2, \ldots, K-1$. First we note that for each $k \in \mathbb{Z}$, the integration interval $[-2^jN-k, 2^jN-k]$ is a symmetric interval around k. As $j \to -\infty$, the length of the interval tends to 0, so

$$\lim_{j\to-\infty} \int_{-2^jN-k}^{2^jN-k} |\phi(y)|^2\,dy = 0.$$

Thus, we can choose $j \in \mathbb{Z}$ sufficiently small (still such that $2^jN < \frac{1}{2}$) such that

$$\sum_{|k|<K} \int_{-2^jN-k}^{2^jN-k} |\phi(y)|^2\,dy \le \frac{\epsilon}{2}.$$

With this choice for j, (9.4) shows that

$$\sum_{k\in\mathbb{Z}} \int_{-2^jN-k}^{2^jN-k} |\phi(y)|^2\,dy \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

It now follows from the estimate (9.2) that

$$||P_jf||_2^2 \le \epsilon\,||f||_2^2.$$

The above proof implies that $||P_{j'}f||_2^2 \le \epsilon\,||f||_2^2$ for all $j' \le j$. Since $\epsilon > 0$ was arbitrary, it follows that

$$\lim_{j\to-\infty} ||P_jf||_2 = 0.$$

This proves the result for compactly supported functions f. We now consider an arbitrary function $f \in L^2(\mathbb{R})$. By Theorem 5.4.2 the set of functions in $L^2(\mathbb{R})$ with compact support forms a dense subspace of $L^2(\mathbb{R})$. Given any $\epsilon > 0$, we can therefore find a function $g \in L^2(\mathbb{R})$ with compact support such that $||f-g||_2 \le \epsilon/2$. By the result we just proved, we can find $J \in \mathbb{Z}$ such that $||P_jg||_2 \le \epsilon/2$ if $j \le J$. Using the triangle inequality and that any orthogonal projection has norm at most 1, it follows that for $j \le J$,

$$\begin{aligned} ||P_jf||_2 &= ||P_j(f-g) + P_jg||_2 \\ &\le ||P_j(f-g)||_2 + ||P_jg||_2 \\ &\le ||f-g||_2 + ||P_jg||_2 \\ &\le \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

As before, this implies that $\lim_{j\to-\infty} ||P_jf||_2 = 0$. □

Lemma 9.1.2 implies that the second condition in Definition 8.2.1(ii) is superfluous – it is automatically satisfied if the conditions in (i), (iii), and (v) hold:

Corollary 9.1.3 *Let $\phi \in L^2(\mathbb{R})$, and define the spaces V_j by (9.1). Assume that condition (v) in Definition 8.2.1 holds. Then*

$$\cap_{j\in\mathbb{Z}} V_j = \mathbf{0}.$$

Proof. Assume that $f \in \cap_{j\in\mathbb{Z}} V_j$. Then $f \in V_j$ for all $j \in \mathbb{Z}$, i.e., $P_j f = f$ for all $j \in \mathbb{Z}$. Thus, by Lemma 9.1.2(ii),

$$||f||_2 = \lim_{j\to-\infty} ||P_j f||_2 = 0.$$

We conclude that $f = 0$. □

The proof of the next result is similar to the last part of the proof for Lemma 9.1.2, and is left to the reader (Exercise 9.3).

Lemma 9.1.4 *Assume that the conditions (i) and (ii) in Definition 8.2.1 are satisfied. Then* $\lim_{j\to\infty} P_j f = f,\ \forall f \in L^2(\mathbb{R})$.

9.2 The spaces V_j and W_j

In the entire section we assume that the function $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis and that the spaces V_j are given by (9.1). Based on the condition (i) in Definition 8.2.1, let W_j denote the orthogonal complement of V_j relative to V_{j+1} (see Definition 8.2.4). We note that by Theorem 4.3.5,

$$V_{j+1} = V_j \oplus W_j. \tag{9.5}$$

Recall also from Definition 8.2.4 that we denote the orthogonal projection of $L^2(\mathbb{R})$ onto W_j by Q_j.

We will now provide a detailed analysis of the spaces V_j and W_j. Along the way we prove Proposition 8.2.5. The final goal for the section is to obtain a characterization of the space W_0. Hereby we will be able to find a function $\psi \in L^2(\mathbb{R})$ such that $\{T_k\psi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_0, which by Proposition 8.2.5 means that ψ is a wavelet.

Lemma 9.2.1 (Decomposition of V_K) *Assume that condition (i) in Definition 8.2.1 is satisfied. For any $J, K \in \mathbb{Z}$ with $K > J$, the following hold:*

(i) The vector space V_K can be decomposed as an orthogonal direct sum,

$$V_K = V_J \oplus W_J \oplus W_{J+1} \oplus \cdots \oplus W_{K-1}. \tag{9.6}$$

(ii) In terms of the orthogonal projections P_j onto V_j and Q_j onto W_j,

$$P_K = P_J + \sum_{j=J}^{K-1} Q_j.$$

Proof. The result in (i) follows by iteration of (9.5):

$$\begin{aligned} V_K &= V_{K-1} \oplus W_{K-1} \\ &= V_{K-2} \oplus W_{K-2} \oplus W_{K-1} \\ &= \cdots \\ &= V_J \oplus W_J \oplus W_{J+1} \oplus \cdots \oplus W_{K-1}. \end{aligned}$$

Similarly, iterating the result in Exercise 9.2,

$$\begin{aligned} P_K &= P_{K-1} + Q_{K-1} \\ &= P_{K-2} + Q_{K-2} + Q_{K-1} \\ &= \cdots \\ &= P_J + \sum_{j=J}^{K-1} Q_j. \end{aligned}$$

This concludes the proof. □

The spaces W_j satisfy the same dilation relationship as V_j:

Lemma 9.2.2 *For each* $j \in \mathbb{Z}$,

$$W_j = D^j(W_0). \tag{9.7}$$

Proof. Fix $j \in \mathbb{Z}$. Since $W_0 \subset V_1$, we have

$$D^j(W_0) \subseteq D^j(V_1) = V_{j+1}. \tag{9.8}$$

Also, because $W_0 \perp V_0$, and D is unitary, we have $D^j(W_0) \perp D^j(V_0)$, i.e., $D^j(W_0) \perp V_j$. Put together, these two observations show that

$$D^j(W_0) \subseteq W_j.$$

In order to complete the proof, we need to show that $W_j \subseteq D^j(W_0)$. Thus, let $f \in W_j$. Since $W_j \subseteq V_{j+1} = D^j(V_1)$, we can write $f = D^j g$ for some $g \in V_1$. Write $g = v + w$ for some $v \in V_0,\ w \in W_0$. Then

$$f = D^j g = D^j v + D^j w.$$

Since $D^j v \in V_j$ and $f \in W_j$, we have

$$\begin{aligned} 0 = \langle f, D^j v\rangle = \langle D^j v + D^j w, D^j v\rangle &= \langle D^j v, D^j v\rangle + \langle D^j w, D^j v\rangle \\ &= ||D^j v||^2, \end{aligned}$$

i.e., $D^j v = 0$. Thus, $f = D^j v + D^j w = D^j w \in D^j(W_0)$, as desired. □

We will now show that the spaces W_j are the key to obtain an orthogonal decomposition of $L^2(\mathbb{R})$:

Lemma 9.2.3 (Decomposition of $L^2(\mathbb{R})$) *Assume that $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis. Define the spaces W_j and the orthogonal projection Q_j as in Definition 8.2.4. Then the following hold:*

(i) For any $j, j' \in \mathbb{Z}$ with $j \neq j'$, the spaces W_j and $W_{j'}$ are orthogonal.

(ii) Each $f \in L^2(\mathbb{R})$ has a representation $f = \sum_{j\in\mathbb{Z}} Q_j f$.

Proof. Let $j, j' \in \mathbb{Z}$ with $j \neq j'$ be given; we can assume that $j > j'$. Lemma 9.2.1 implies that

$$V_{j+1} = V_{j'} \oplus W_{j'} \oplus \cdots \oplus W_j,$$

which proves (i).

In order to prove (ii), we use that for $K > J$,

$$P_K f = P_J f + \sum_{j=J}^{K-1} Q_j f$$

by Lemma 9.2.1(ii). Letting $J \to -\infty$, Lemma 9.1.2 implies that

$$P_K f = \sum_{j=-\infty}^{K-1} Q_j f.$$

Finally, letting $K \to \infty$, Lemma 9.1.4 leads to the result stated in (ii). □

In brief, the properties of the spaces W_j described in (i) and (ii) in Lemma 9.2.3 are often expressed as

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j. \tag{9.9}$$

We are now ready for the announced proof of Proposition 8.2.5.

Proof of Proposition 8.2.5: The proof of (i) is identical to the proof of Lemma 9.1.2(i).

For the proof of (ii), we first show that the functions $\{D^j T_k \psi\}_{j,k\in\mathbb{Z}}$ form an orthonormal system. By the result in (i) the functions $\{D^j T_k \psi\}_{k\in\mathbb{Z}}$ form an orthonormal basis for W_j for any given $j \in \mathbb{Z}$. Also, for $j \neq j'$, the functions $\{D^j T_k \psi\}_{k\in\mathbb{Z}}$ and $\{D^{j'} T_k \psi\}_{k\in\mathbb{Z}}$ are orthogonal by Lemma 9.2.3(i). Thus, $\{D^j T_k \psi\}_{j,k\in\mathbb{Z}}$ form an orthonormal system. In order to prove that $\{D^j T_k \psi\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis for $L^2(\mathbb{R})$ we will now use Theorem 4.7.2. The result in (i) together with Exercise 4.29 implies that for any $f \in L^2(\mathbb{R})$,

$$Q_j f = \sum_{k\in\mathbb{Z}} \langle f, D^j T_k \psi \rangle D^j T_k \psi.$$

Thus, by Lemma 9.2.3,

$$f = \sum_{j\in\mathbb{Z}} Q_j f = \sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \langle f, D^j T_k\psi\rangle D^j T_k\psi.$$

The conclusion in (ii) now follows from Theorem 4.7.2.

By Lemma 9.2.1, all the spaces W_j, $j \geq 1$, are orthogonal with V_0. This implies that the functions $\{T_k\phi\}_{k\in\mathbb{Z}} \cup \{D^jT_k\psi\}_{j\in\mathbb{N},k\in\mathbb{Z}}$ form an orthonormal system. By an argument like in the proof of Lemma 9.2.3(ii), the reader can show (Exercise 9.4) that for any $f \in L^2(\mathbb{R})$,

$$f = P_0 f + \sum_{j=0}^{\infty} Q_j f \tag{9.10}$$

$$= \sum_{k\in\mathbb{Z}} \langle f, T_k\phi\rangle T_k\phi + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} \langle f, D^jT_k\psi\rangle D^jT_k\psi. \tag{9.11}$$

Again, we conclude from Theorem 4.7.2 that the functions

$$\{T_k\phi\}_{k\in\mathbb{Z}} \cup \{D^jT_k\psi\}_{j\in\mathbb{N},k\in\mathbb{Z}}$$

form an orthonormal basis for $L^2(\mathbb{R})$. □

In Proposition 8.2.6 we saw that if the function $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis, there exists a 1-periodic function $H_0 \in L^2(0,1)$ satisfying the *scaling equation*

$$\hat{\phi}(2\gamma) = H_0(\gamma)\hat{\phi}(\gamma), \ \gamma \in \mathbb{R}. \tag{9.12}$$

Note that the result in Exercise 8.12 shows that the function H_0 is uniquely determined. We will now show that H_0 satisfies an equation that will be crucial in the sequel:

Lemma 9.2.4 *Assume that $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0 and that the 1-periodic function $H_0 \in L^2(0,1)$ satisfies the scaling equation* (9.12)*. Then*

$$|H_0(\gamma)|^2 + |H_0(\gamma + \frac{1}{2})|^2 = 1, \ \gamma \in \mathbb{R}. \tag{9.13}$$

Proof. The scaling relation (9.12) implies that

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \left|\hat{\phi}(\gamma+k)\right|^2 \\
= &\sum_{k\in\mathbb{Z}} \left|H_0(\frac{\gamma+k}{2})\hat{\phi}(\frac{\gamma+k}{2})\right|^2 \\
= &\sum_{k \text{ even}} \left|H_0(\frac{\gamma+k}{2})\hat{\phi}(\frac{\gamma+k}{2})\right|^2 + \sum_{k \text{ odd}} \left|H_0(\frac{\gamma+k}{2})\hat{\phi}(\frac{\gamma+k}{2})\right|^2.
\end{aligned}$$

Writing the even indices as $2k,\ k \in \mathbb{Z}$, the periodicity of H_0 shows that

$$H_0(\frac{\gamma+2k}{2})\,\hat{\phi}(\frac{\gamma+2k}{2}) = H_0(\frac{\gamma}{2}+k)\,\hat{\phi}(\frac{\gamma}{2}+k) = H_0(\frac{\gamma}{2})\,\hat{\phi}(\frac{\gamma}{2}+k).$$

Similarly, writing the odd indices as $2k+1,\ k \in \mathbb{Z}$,

$$H_0(\frac{\gamma+2k+1}{2})\,\hat{\phi}(\frac{\gamma+2k+1}{2}) = H_0(\frac{\gamma}{2}+\frac{1}{2})\,\hat{\phi}(\frac{\gamma}{2}+\frac{1}{2}+k).$$

Thus, the above calculation implies that

$$\begin{aligned}&\sum_{k\in\mathbb{Z}} \left|\hat{\phi}(\gamma+k)\right|^2 \\ = \quad & |H_0(\frac{\gamma}{2})|^2 \sum_{k\in\mathbb{Z}} \left|\hat{\phi}(\frac{\gamma}{2}+k)\right|^2 + |H_0(\frac{\gamma}{2}+\frac{1}{2})|^2 \sum_{k\in\mathbb{Z}} \left|\hat{\phi}(\frac{\gamma}{2}+\frac{1}{2}+k)\right|^2.\end{aligned}$$

By the result in Theorem 8.2.12(ii), the assumption that $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0 implies that

$$\sum_{k\in\mathbb{Z}} \left|\hat{\phi}(\gamma+k)\right|^2 = 1,\ \gamma \in \mathbb{R}.$$

Thus, the above calculation implies that

$$1 = |H_0(\frac{\gamma}{2})|^2 + |H_0(\frac{\gamma}{2}+\frac{1}{2})|^2,\ \gamma \in \mathbb{R};$$

replacing γ by 2γ leads to the desired result. □

By Proposition 8.2.5 we can construct an orthonormal basis $\{D^jT_k\psi\}_{j,k\in\mathbb{Z}}$ for $L^2(\mathbb{R})$ by finding a function $\psi \in W_0$ such that $\{T_k\psi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_0. Our aim now is to characterize the space W_0 and hereby find such a function ψ. The starting point is to characterize the spaces V_0 and V_{-1}:

Lemma 9.2.5 (Characterization of V_{-1} and V_0) *Assume that the function $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis with scaling function H_0. Then the following hold:*

(i) The space V_{-1} can be characterized as

$$V_{-1} = \{f \in L^2(\mathbb{R}) \mid \hat{f}(\gamma) = m(2\gamma)H_0(\gamma)\hat{\phi}(\gamma) \text{ for some 1-periodic } m \in L^2(0,1)\}.$$

(ii) The space V_0 can be characterized as

$$V_0 = \{f \in L^2(\mathbb{R}) \mid \hat{f}(\gamma) = m(\gamma)\hat{\phi}(\gamma) \text{ for some 1-periodic } m \in L^2(0,1)\}.$$

Proof. By Lemma 9.1.2 the functions $\{D^{-1}T_k\phi\}_{k\in\mathbb{Z}}$ form an orthonormal basis for V_{-1}. Thus, given any $f \in V_{-1}$ we can write

$$f = \sum_{k\in\mathbb{Z}} c_k D^{-1}T_k\phi$$

for some coefficients $\{c_k\}_{k\in\mathbb{Z}}$. Note that by Exercise 4.32 we know that $\{c_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$. Using the commutation relations in Theorem 7.1.2 and Lemma 7.1.3,

$$\begin{aligned}
\hat{f}(\gamma) &= \mathcal{F}\sum_{k\in\mathbb{Z}} c_k D^{-1}T_k\phi(\gamma) \\
&= \sum_{k\in\mathbb{Z}} c_k \mathcal{F} D^{-1}T_k\phi(\gamma) \\
&= \sum_{k\in\mathbb{Z}} c_k D E_{-k}\mathcal{F}\phi(\gamma) \\
&= \sqrt{2}\sum_{k\in\mathbb{Z}} c_k e^{-4\pi i k\gamma}\hat{\phi}(2\gamma). \qquad (9.14)
\end{aligned}$$

Via the scaling equation (9.12) this implies that

$$\hat{f}(\gamma) = \sqrt{2}\sum_{k\in\mathbb{Z}} c_k e^{-4\pi i k\gamma}\, H_0(\gamma)\,\hat{\phi}(\gamma). \qquad (9.15)$$

Defining the function m by

$$m(\gamma) := \sqrt{2}\sum_{k\in\mathbb{Z}} c_k e^{-2\pi i k\gamma}, \qquad (9.16)$$

we have that m is a 1-periodic function belonging to $L^2(0,1)$, and

$$\hat{f}(\gamma) = m(2\gamma)\, H_0(\gamma)\,\hat{\phi}(\gamma), \qquad (9.17)$$

as desired. On the other hand, assume that f is any function for which $\hat{f}$ can be written on the form (9.17) for some 1-periodic function $m \in L^2(0,1)$. We want to show that $f \in V_{-1}$. Observe that for such a function m, the function

$$\gamma \mapsto m(2\gamma)H_0(\gamma)$$

belongs to $L^2(0,1)$ because $m \in L^2(0,1)$ and H_0 is bounded by (9.13). We will now show that therefore the function f defined by (9.17) actually belongs to $L^2(\mathbb{R})$. In order to do this, we first calculate

$$\begin{aligned}\int_{-\infty}^{\infty} |\hat{f}(\gamma)|^2 \, d\gamma &= \int_{-\infty}^{\infty} |m(2\gamma)\, H_0(\gamma)\, \hat{\phi}(\gamma)|^2 \, d\gamma \\ &= \sum_{k\in\mathbb{Z}} \int_{k}^{k+1} |m(2\gamma)\, H_0(\gamma)\, \hat{\phi}(\gamma)|^2 \, d\gamma \\ &= \sum_{k\in\mathbb{Z}} \int_{0}^{1} |m(2(\gamma+k))\, H_0(\gamma+k)\, \hat{\phi}(\gamma+k)|^2 \, d\gamma.\end{aligned}$$

Using that the functions m and H_0 are 1-periodic and the result in Example 5.3.4, it follows that

$$\begin{aligned}\int_{-\infty}^{\infty} |\hat{f}(\gamma)|^2 \, d\gamma &= \sum_{k\in\mathbb{Z}} \int_{0}^{1} |m(2\gamma)\, H_0(\gamma)\, \hat{\phi}(\gamma+k)|^2 \, d\gamma \\ &= \int_{0}^{1} |m(2\gamma)\, H_0(\gamma)|^2 \sum_{k\in\mathbb{Z}} |\hat{\phi}(\gamma+k)|^2 \, d\gamma.\end{aligned}$$

Because $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal system, the result in Theorem 8.2.12 shows that

$$\sum_{k\in\mathbb{Z}} |\hat{\phi}(\gamma+k)|^2 \, d\gamma = 1.$$

Therefore, we conclude that

$$\int_{-\infty}^{\infty} |\hat{f}(\gamma)|^2 \, d\gamma = \int_{0}^{1} |m(2\gamma)\, H_0(\gamma)|^2 \, d\gamma < \infty,$$

showing that $\hat{f} \in L^2(\mathbb{R})$. Because the Fourier transform is a bijection from $L^2(\mathbb{R})$ onto $L^2(\mathbb{R})$ we conclude that $f \in L^2(\mathbb{R})$, as claimed. Now, taking (9.17) as the starting point and writing the function m as in (9.16), we can reverse the calculation leading to (9.14) and conclude that $f \in V_{-1}$. This completes the proof of (i). The proof for (ii) is similar (but does not use the scaling equation). In fact, we get immediately that for any $f \in V_0$ we can write

$$f = \sum_{k\in\mathbb{Z}} c_k T_k \phi \tag{9.18}$$

for some coefficients $\{c_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$, and therefore

$$\hat{f} = \sum_{k\in\mathbb{Z}} c_k E_{-k} \hat{\phi} = m\, \hat{\phi} \tag{9.19}$$

with

$$m(\gamma) := \sum_{k\in\mathbb{Z}} c_k e^{-2\pi i k\gamma}. \tag{9.20}$$

Again, reversing the calculations leads to the announced result. □

The result in Lemma 9.2.5(ii) shows that for any function $f \in V_0$ we can find a 1-periodic function $m_f \in L^2(0,1)$ such that

$$\hat{f}(\gamma) = m_f(\gamma)\hat{\phi}(\gamma), \gamma \in \mathbb{R}. \tag{9.21}$$

The function m_f associated with f is actually uniquely determined. In fact, the result in Theorem 8.2.12 implies that for any $\gamma \in \mathbb{R}$ we can find $k \in \mathbb{Z}$ such that $\hat{\phi}(\gamma + k) \neq 0$; thus, the relation (9.21) implies that

$$m_f(\gamma) = m_f(\gamma + k) = \frac{\hat{f}(\gamma + k)}{\hat{\phi}(\gamma + k)}.$$

The above consideration leads to the definition of an operator, which we state formally:

Definition 9.2.6 (The operator U) *Given an orthonormal basis $\{T_k\phi\}_{k\in\mathbb{Z}}$ for V_0, define the mapping $U : V_0 \to L^2(0,1)$ by*

$$Uf := m_f. \tag{9.22}$$

Writing

$$f = \sum_{k\in\mathbb{Z}} c_k T_k \phi,$$

the function m_f is given explicitly by

$$m_f(\gamma) = \sum_{k\in\mathbb{Z}} c_k e^{-2\pi i k\gamma}. \tag{9.23}$$

Let us state a few properties of the operator U:

Lemma 9.2.7 (Properties of the operator U) *The mapping U defined by (9.22) is linear and isometric. In particular,*

$$\langle Uf, Ug\rangle_{L^2(0,1)} = \langle f, g\rangle_{L^2(\mathbb{R})}, \ \forall f, g \in V_0. \tag{9.24}$$

Proof. It is clear that U is linear. Using that $\{e^{-2\pi i k\gamma}\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(0,1)$ and $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0, the result in Exercise 4.31 shows that with $f = \sum_{k\in\mathbb{Z}} c_k T_k\phi$,

$$\begin{aligned} ||Uf||^2_{L^2(0,1)} &= ||m_f||^2_{L^2(0,1)} \\ &= \left|\left| \sum_{k\in\mathbb{Z}} c_k e^{-2\pi i k\gamma} \right|\right|^2_{L^2(0,1)} \\ &= \sum_{k\in\mathbb{Z}} |c_k|^2 \\ &= \left|\left| \sum c_k T_k \phi \right|\right|_2 \\ &= ||f||_2. \end{aligned}$$

This proves that U is an isometry. The result in (9.24) now follows from the polarization identity in Theorem 4.1.4. □

Our characterizations of the spaces V_{-1} and V_0 together with the properties of the operator U make it possible to characterize W_{-1}:

Lemma 9.2.8 (Characterization of W_{-1}) *If $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis with scaling function H_0, then*

$$W_{-1} =$$

$$\{f \in V_0 | m_f(\gamma) = e^{-2\pi i\gamma} m(2\gamma)\overline{H_0(\gamma + \frac{1}{2})} \text{ for some 1-periodic} m \in L^2(0,1)\}.$$

Proof. By definition, the space W_{-1} is the orthogonal complement of V_{-1} with respect to V_0. Thus, a function $f \in V_0$ belongs to W_{-1} if and only if

$$\langle f, g\rangle = 0, \ \forall g \in V_{-1},$$

or, by (9.24), if and only if

$$\langle Uf, Ug\rangle = 0, \ \forall g \in V_{-1}. \tag{9.25}$$

Note that the characterization of V_{-1} in Lemma 9.2.5 shows that

$$\{Ug \mid g \in V_{-1}\} = \{m(2\cdot)\, H_0 \mid m \in L^2(0,1) \text{ is 1-periodic}\}.$$

Let m_f be the function in $L^2(0,1)$ associated with f as in Definition 9.2.6, i.e., $m_f = Uf$. It follows that the condition (9.25) is equivalent with

$$0 = \langle m_f, m(2\cdot)\, H_0\rangle_{L^2(0,1)} = 0, \ \forall m \in L^2(0,1),$$

i.e.,

$$\int_0^1 m_f(\gamma)\overline{m(2\gamma)\, H_0(\gamma)}\, d\gamma = 0, \ \forall m \in L^2(0,1). \tag{9.26}$$

The function $m(2\cdot)$ is periodic with period 1/2. We rewrite (9.26) as

$$\begin{aligned}
0 &= \int_0^{1/2} m_f(\gamma)\,\overline{m(2\gamma)\, H_0(\gamma)}\, d\gamma + \int_{1/2}^1 m_f(\gamma)\,\overline{m(2\gamma)\, H_0(\gamma)}\, d\gamma \\
&= \int_0^{1/2} m_f(\gamma)\,\overline{m(2\gamma)\, H_0(\gamma)}\, d\gamma \\
&\quad + \int_0^{1/2} m_f(\gamma+1/2)\,\overline{m(2(\gamma+1/2))\, H_0(\gamma+1/2)}\, d\gamma \\
&= \int_0^{1/2} m_f(\gamma)\,\overline{m(2\gamma)\, H_0(\gamma)}\, d\gamma \\
&\quad + \int_0^{1/2} m_f(\gamma+1/2)\,\overline{m(2\gamma)\, H_0(\gamma+1/2)}\, d\gamma \\
&= \int_0^{1/2} \overline{m(2\gamma)} \left(m_f(\gamma)\overline{H_0(\gamma)} + m_f(\gamma+1/2)\overline{H_0(\gamma+1/2)} \right) d\gamma.
\end{aligned}$$

Note that

$$M(\gamma) := m_f(\gamma)\overline{H_0(\gamma)} + m_f(\gamma+1/2)\overline{H_0(\gamma+1/2)}$$

is a $\frac{1}{2}$-periodic function in $L^2(0,1/2)$, and that

$$L^2(0,1/2) = \{m(2\cdot) \mid m \in L^2(0,1)\}.$$

Thus, the above calculation shows that the function M is orthogonal to all functions in $L^2(0,1/2)$. It now follows from Lemma 4.4.2 applied with $\mathcal{H} := L^2(0,1/2)$ that

$$m_f(\gamma)\overline{H_0(\gamma)} + m_f(\gamma+1/2)\overline{H_0(\gamma+1/2)} = 0, \ \gamma \in [0,1/2].$$

Due to periodicity,

$$m_f(\gamma)\overline{H_0(\gamma)} + m_f(\gamma+1/2)\overline{H_0(\gamma+1/2)} = 0, \ \gamma \in \mathbb{R}. \tag{9.27}$$

In other words, for $\gamma \in \mathbb{R}$, the vectors

$$\begin{pmatrix} m_f(\gamma) \\ m_f(\gamma+1/2) \end{pmatrix}, \quad \begin{pmatrix} H_0(\gamma) \\ H_0(\gamma+1/2) \end{pmatrix}$$

are orthogonal in $\mathbb{C}^2$. The vector space $\mathbb{C}^2$ is 2-dimensional, and one vector that is orthogonal on $\begin{pmatrix} H_0(\gamma) \\ H_0(\gamma+1/2) \end{pmatrix}$ is given by $\begin{pmatrix} \overline{H_0(\gamma+1/2)} \\ -\overline{H_0(\gamma)} \end{pmatrix}$. Thus, for $\gamma \in \mathbb{R}$ there exists a constant $\lambda(\gamma) \in \mathbb{C}$ such that

$$\begin{pmatrix} m_f(\gamma) \\ m_f(\gamma+1/2) \end{pmatrix} = \lambda(\gamma) \begin{pmatrix} \overline{H_0(\gamma+1/2)} \\ -\overline{H_0(\gamma)} \end{pmatrix}. \tag{9.28}$$

Because the functions m_f and H_0 are 1-periodic, the function λ is 1-periodic as well. Replacing γ by $\gamma+1/2$ in (9.28) and using that $m_f(\gamma+1) = m_f(\gamma)$ and $H_0(\gamma+1) = H_0(\gamma)$, we see that

$$\begin{pmatrix} m_f(\gamma+1/2) \\ m_f(\gamma) \end{pmatrix} = \lambda(\gamma+1/2) \begin{pmatrix} \overline{H_0(\gamma)} \\ -\overline{H_0(\gamma+1/2)} \end{pmatrix}. \tag{9.29}$$

The results in (9.28) and (9.29) imply that

$$m_f(\gamma) = \lambda(\gamma)\,\overline{H_0(\gamma+1/2)} = -\lambda(\gamma+1/2)\,\overline{H_0(\gamma+1/2)} \tag{9.30}$$

and

$$m_f(\gamma+1/2) = -\lambda(\gamma)\,\overline{H_0(\gamma)} = \lambda(\gamma+1/2)\,\overline{H_0(\gamma)}. \tag{9.31}$$

From Lemma 9.2.4 we know that for $\gamma \in \mathbb{R}$ either $H_0(\gamma) \neq 0$ or $H_0(\gamma+1/2) \neq 0$; thus, from (9.30) and (9.31) we conclude that

$$\lambda(\gamma) = -\lambda(\gamma+1/2), \ \gamma \in \mathbb{R}. \tag{9.32}$$

Letting

$$m(\gamma) := e^{\pi i \gamma}\lambda(\gamma/2), \tag{9.33}$$

we can write

$$\lambda(\gamma) = e^{-2\pi i\gamma} m(2\gamma).$$

We leave it to the reader (Exercise 9.7) to show that the function m belongs to $L^2(0,1)$. Via (9.33) and (9.32),

$$\begin{aligned} m(\gamma+1) = e^{\pi i(\gamma+1)}\lambda(\frac{\gamma+1}{2}) &= e^{\pi i\gamma}e^{\pi i}\left(-\lambda(\frac{\gamma}{2})\right) \\ &= e^{\pi i\gamma}\lambda(\frac{\gamma}{2}) \\ &= m(\gamma). \end{aligned}$$

This shows that the function m is 1-periodic. From (9.30) we see that

$$\begin{aligned} m_f(\gamma) &= \lambda(\gamma)\,\overline{H_0(\gamma+1/2)} \\ &= e^{-2\pi i\gamma}\, m(2\gamma)\,\overline{H_0(\gamma+1/2)}. \end{aligned} \tag{9.34}$$

We have now shown that if $f \in W_{-1}$, then m_f has the form in (9.34) for some 1-periodic function m. On the other hand, any function $f \in L^2(\mathbb{R})$ for which m_f has the form in (9.34) for some 1-periodic function $m \in L^2(0,1)$, will belong to V_0 by Lemma 9.2.5(ii). Note that (9.34) also implies that

$$\begin{aligned} & m_f(\gamma)\,\overline{H_0(\gamma)} + m_f(\gamma+1/2)\,\overline{H_0(\gamma+1/2)} \\ = \; & e^{-2\pi i\gamma}\, m(2\gamma)\,\overline{H_0(\gamma+1/2)H_0(\gamma)} \\ & + e^{-2\pi i(\gamma+1/2)}\, m(2(\gamma+1/2))\,\overline{H_0(\gamma+1)H_0(\gamma+1/2)} \\ = \; & e^{-2\pi i\gamma}\, m(2\gamma)\,\overline{H_0(\gamma+1/2)H_0(\gamma)} - e^{-2\pi i\gamma}\, m(2\gamma)\,\overline{H_0(\gamma+1/2)H_0(\gamma)} \\ = \; & 0, \end{aligned}$$

i.e., (9.27) is satisfied. Reversing the calculation at the beginning of the proof now shows that $\langle f, g\rangle = 0$ for all $g \in V_{-1}$, i.e., that $f \in W_{-1}$. Thus, we have proved the result. □

The result in Lemma 9.2.8 together with $W_0 = D(W_{-1})$ finally leads to the desired characterization of the space W_0:

Proposition 9.2.9 (Characterization of W_0) *If $\phi \in L^2(\mathbb{R})$ generates a multiresolution analysis with scaling function H_0, then*

$$\begin{aligned} W_0 \;=\; & \{f \in L^2(\mathbb{R}) \,|\, \hat{f}(2\gamma) = e^{-2\pi i\gamma} m(2\gamma)\overline{H_0(\gamma+\frac{1}{2})}\hat{\phi}(\gamma) \\ & \textit{for some 1-periodic } m \in L^2(0,1)\}. \end{aligned}$$

The proof of Proposition 9.2.9 is left to the reader as Exercise 9.5.

9.3 Proof of Theorem 8.2.7

We are now ready to combine all the obtained results and prove how to construct a wavelet.

Proof of Theorem 8.2.7: By Proposition 9.2.9, the function ψ defined in Theorem 8.2.11 belongs to W_0. By Proposition 8.2.5 we can show that $\{D^jT_k\psi\}_{j,k\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$ by showing that $\{T_k\psi\}_{k\in\mathbb{Z}}$ forms an orthonormal basis for W_0. We first prove that $\{T_k\psi\}_{k\in\mathbb{Z}}$ is an orthonormal system. Using the definition of ψ,

$$\sum_{k\in\mathbb{Z}}|\hat{\psi}(\gamma+k)|^2 \quad = \quad \sum_{k\in\mathbb{Z}}|H_0(\frac{\gamma+k}{2}+\frac{1}{2})\hat{\phi}(\frac{\gamma+k}{2})|. \tag{9.35}$$

Splitting into even and odd values of $k\in\mathbb{Z}$ in (9.35), the fact that $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal system together with Theorem 8.2.12 and Lemma 9.2.4 implies that

$$\sum_{k\in\mathbb{Z}}|\hat{\psi}(\gamma+k)|^2 = 1,\ \gamma\in\mathbb{R}, \tag{9.36}$$

i.e., that $\{T_k\psi\}_{k\in\mathbb{Z}}$ forms an orthonormal system (Exercise 9.6). In order to show that $\{T_k\psi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_0 we now only need to show that $\{T_k\psi\}_{k\in\mathbb{Z}}$ spans W_0. For that purpose, let $g\in W_0$. By Proposition 9.2.9 there exists a 1-periodic function $m\in L^2(0,1)$ such that

$$\hat{g}(2\gamma) = e^{-2\pi i\gamma}m(2\gamma)\overline{H_0(\gamma+\frac{1}{2})}\hat{\phi}(\gamma);$$

thus, replacing γ by $\gamma/2$ and using the definition of ψ,

$$\begin{aligned}\hat{g}(\gamma) \quad &= \quad e^{-\pi i\gamma}m(\gamma)\overline{H_0(\frac{\gamma}{2}+\frac{1}{2})}\hat{\phi}(\gamma/2)\\ &= \quad m(\gamma)H_1(\gamma/2)\hat{\phi}(\gamma/2)\\ &= \quad m(\gamma)\hat{\psi}(\gamma).\end{aligned}$$

Writing

$$m(\gamma) = \sum_{k\in\mathbb{Z}}c_ke^{-2\pi ik\gamma},$$

this implies that

$$g = \sum_{k\in\mathbb{Z}}c_kT_k\psi.$$

Since $\{T_k\psi\}_{k\in\mathbb{Z}}$ is an orthonormal system,

$$\begin{aligned}\langle g, T_\ell\psi\rangle &= \langle\sum_{k\in\mathbb{Z}} c_k T_k\psi, T_\ell\psi\rangle \\ &= \sum_{k\in\mathbb{Z}} c_k\langle T_k\psi, T_\ell\psi\rangle \\ &= c_\ell,\end{aligned}$$

i.e.,

$$g = \sum_{k\in\mathbb{Z}} c_k T_k\psi = \sum_{k\in\mathbb{Z}} \langle g, T_k\psi T_k\psi.$$

By Theorem 4.7.2 this confirms that $\{T_k\psi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for W_0, as desired. □

9.4 Proof of Theorem 8.2.11

The purpose of this section is to prove that the conditions in Theorem 8.2.11 are sufficient for a function $\phi \in L^2(\mathbb{R})$ to generate a multiresolution analysis. We first provide a weak condition for condition (i) in Definition 8.2.1 to hold:

Lemma 9.4.1 *Assume that $\phi \in L^2(\mathbb{R})$ and that $\{T_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal sequence. Define the spaces V_j by*

$$V_j = \overline{span}\{D^j T_k\phi\}_{k\in\mathbb{Z}}.$$

Then the following hold:

(i) If $\psi \in L^2(\mathbb{R})$ and there exists a bounded 1-periodic function H_1 such that $\hat{\psi}(2\gamma) = H_1(\gamma)\hat{\phi}(\gamma)$, then $\psi \in V_1$.

(ii) If there exists a bounded 1-periodic function H_0 such that

$$\hat{\phi}(2\gamma) = H_0(\gamma)\hat{\phi}(\gamma), \tag{9.37}$$

then $V_j \subseteq V_{j+1}$ for all $j \in \mathbb{Z}$.

Proof. For convenience we will assume that H_0 and H_1 are trigonometric polynomials. If the conditions in (i) are satisfied, the expression for the function ψ in Proposition 8.2.8 shows that $\psi \in V_1$. This proves (i). For the proof of (ii), we note that, via (i), $\phi \in V_1$. It follows that $T_k\phi \in V_1$ for all $k \in \mathbb{Z}$ (Exercise 8.3). Now, because V_1 is a closed vector space, we can conclude that

$$\overline{\text{span}}\{T_k\phi\}_{k\in\mathbb{Z}} \subseteq V_1,$$

i.e., $V_0 \subseteq V_1$. Applying the operator D^j now implies that $V_j \subseteq V_{j+1}$ for all $j \in \mathbb{Z}$. □

Lemma 9.4.2 *Let* $\phi \in L^2(\mathbb{R})$, *and define as usual*

$$V_j = \overline{span}\{D^j T_k \phi\}_{k \in \mathbb{Z}}.$$

Assume that $V_j \subseteq V_{j+1}$ *for all* $j \in \mathbb{Z}$ *and that condition (v) in Definition 8.2.1 holds. Then the space*

$$V := \overline{\bigcup_{j \in \mathbb{Z}} V_j}$$

has the property that

$$f \in V \Rightarrow T_y f \in V, \ \forall y \in \mathbb{R}.$$

Proof. Let $f \in V$. We first show that $T_{2^{-\ell}m} f \in V$ for all $\ell, m \in \mathbb{Z}$, i.e., that for any given $\epsilon > 0$ there exists a function g belonging to some space $V_j, \ j \in \mathbb{Z}$, such that

$$||T_{2^{-\ell}m} f - g||_2 \leq \epsilon. \tag{9.38}$$

Given $\epsilon > 0$, there exists a function $h \in \cup_{j \in \mathbb{Z}} V_j$ such that $||f - h||_2 \leq \epsilon$. Then $h \in V_J$ for some $J \in \mathbb{Z}$, and thus $h \in V_j$ for all $j \geq J$. By Lemma 9.1.2, the sequence $\{D^j T_k \phi\}_{k \in \mathbb{Z}}$ is an orthonormal basis for V_j. Thus, for any $j \geq J$ there exists by Theorem 4.7.2 coefficients $\{c_k\}_{k=1}^{\infty}$ such that

$$h(x) = \sum_{k \in \mathbb{Z}} c_k \phi(2^j x - k).$$

Thus,

$$\begin{aligned} T_{2^{-\ell}m} h(x) &= \sum_{k \in \mathbb{Z}} c_k \phi(2^j (x - 2^{-\ell} m) - k) \\ &= \sum_{k \in \mathbb{Z}} c_k \phi(2^j x - 2^{j-\ell} m - k). \end{aligned} \tag{9.39}$$

If $j \geq \ell$, then $2^{j-\ell} m - k \in \mathbb{Z}$, so the functions

$$x \mapsto \phi(2^j x - k - 2^{j-\ell} m)$$

belong to V_j. By (9.39) this shows that $T_{2^{-\ell}m} h \in V_j$ if $j \geq \max(J, \ell)$. Now,

$$||T_{2^{-\ell}m} f - T_{2^{-\ell}m} h||_2 = ||f - h||_2 \leq \epsilon.$$

Thus, the condition (9.38) is satisfied with $g := T_{2^{-\ell}m} h$.

So far, we have proved the result for $y \in \mathbb{R}$ of the special form $y = 2^{-\ell} m, \ \ell, m \in \mathbb{Z}$. Now consider an arbitrary $y \in \mathbb{R}$. Given any $\delta > 0$, there exists $\ell, m \in \mathbb{Z}$ such that

$$|2^{-\ell} m - y| \leq \delta. \tag{9.40}$$

Consider again any $f \in V$. Now,

$$\begin{aligned} ||T_y f - T_{2^{-\ell}m} f||_2 &= ||T_y(f - T_{2^{-\ell}m-y} f)||_2 \\ &= ||f - T_{2^{-\ell}m-y} f||_2. \end{aligned} \tag{9.41}$$

For any $\epsilon > 0$ there exists by Lemma 6.2.5 a $\delta > 0$ such that

$$||f - T_{2^{-\ell}m-y} f||_2 \le \epsilon \ \text{ if } \ |2^{-\ell}m - y| \le \delta.$$

Thus, by (9.41),

$$||T_y f - T_{2^{-\ell}m} f||_2 \le \epsilon \ \text{ if } \ |2^{-\ell}m - y| \le \delta.$$

By the result we just proved, $T_{2^{-\ell}m} f \in V$, so we conclude that $T_y f \in V$. □

Via Lemma 9.4.2 we can now derive a sufficient condition for the spaces

$$V_j = \overline{span}\{D^j T_k \phi\}_{k \in \mathbb{Z}} \tag{9.42}$$

to satisfy the condition (ii) in Definition 8.2.1:

Proposition 9.4.3 *Let $\phi \in L^2(\mathbb{R})$, and define as usual the spaces V_j by (9.42). Assume that the function $\phi \in L^2(\mathbb{R})$ satisfies that $\hat{\phi}$ is continuous at 0 and that $\hat{\phi}(0) \neq 0$. Assume furthermore that $V_j \subseteq V_{j+1}$ for all $j \in \mathbb{Z}$ and that condition (v) in Definition 8.2.1 is satisfied. Then*

$$\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R}).$$

Proof. Let $V := \overline{\bigcup_{j\in\mathbb{Z}} V_j}$. To show that $V = L^2(\mathbb{R})$ amounts to show that $V^\perp = \{0\}$, see Theorem 4.3.5. Thus, let $g \in V^\perp$. Given any function $f \in V$, Lemma 9.4.2 shows that $T_{-y} f \in V$ for all $y \in V$. By Theorem 7.2.2 together with the commutation relation in Theorem 7.1.2(iii),

$$\begin{aligned} 0 = \langle T_{-y} f, g \rangle &= \langle \mathcal{F} T_{-y}, \mathcal{F} g \rangle \\ &= \langle E_y \hat{f}, \hat{g} \rangle \\ &= \int_{-\infty}^{\infty} e^{2\pi i x y} \hat{f}(x) \overline{\hat{g}(x)}\, dx. \end{aligned}$$

This shows that

$$\mathcal{F}(\hat{f}\,\overline{\hat{g}})(y) = 0, \ y \in \mathbb{R}.$$

By Cauchy–Schwarz' inequality, the function $\hat{f}\overline{\hat{g}}$ belongs to $L^1(\mathbb{R})$. Since the Fourier transform is injective on $L^1(\mathbb{R})$, we conclude that

$$\hat{f}(x)\overline{\hat{g}(x)} = 0, \ x \in \mathbb{R}. \tag{9.43}$$

We will now consider a special choice of f, namely, $f := D^j \phi$ for some $j \in \mathbb{R}$. By the commutation relations,

$$\hat{f}(x) = \mathcal{F} D^j \phi(x) = D^{-j} \mathcal{F} f(x) = 2^{-j/2} \hat{\phi}(2^{-j} x). \tag{9.44}$$

By the assumptions on the function ϕ, there exists an $\epsilon > 0$ such that

$$\hat{\phi}(x) \neq 0 \text{ for } x \in]-\epsilon, \epsilon[.$$

With $f = D^j\phi$, this implies by (9.44) that

$$\hat{f}(x) \neq 0 \text{ for } x \in]-2^j\epsilon, 2^j\epsilon[.$$

Thus, via (9.43) we can now conclude that $\hat{g}(x) = 0, \ x \in]-2^j\epsilon, 2^j\epsilon[$. Letting $j \to \infty$ now shows that $\hat{g} = 0$, i.e., $g = 0$. Thus, we have shown that $V^\perp = \{0\}$, as desired. □

Via Lemma 9.4.1, Corollary 9.1.3, and Proposition 9.4.3 we obtain the criterion in Theorem 8.2.11 for a function ϕ to generate a multiresolution analysis:

Proof of Theorem 8.2.11: Assume that the conditions in Theorem 8.2.11 are satisfied, and let as usual V_j be defined by (9.42). By Lemma 9.4.1(ii) we have that $V_j \subseteq V_{j+1}$. The definition of V_j also implies that condition (iii) in Definition 8.2.1 holds. Corollary 9.1.3 and Proposition 9.4.3 show that condition (ii) in Definition 8.2.1 holds as well. This clearly also implies that $V_j \subset V_{j+1}$ for all $j \in \mathbb{Z}$. Finally, condition (iv) in Definition 8.2.1 holds by the choice of V_j. □

9.5 Exercises

9.1 Let V and W be closed subspaces of a Hilbert space $\mathcal{H}$, and assume that $V \subset W$. Let P_V denote the orthogonal projection of $\mathcal{H}$ onto V, and P_W the orthogonal projection onto W. Show that

$$||P_V\mathbf{v}|| \leq ||P_W\mathbf{v}||, \ \forall \, \mathbf{v} \in \mathcal{H}.$$

9.2 Assume that V_{j+1}, V_j, and W_j are subspaces of a Hilbert space $\mathcal{H}$ and satisfy that

$$V_{j+1} = V_j \oplus W_j,$$

where the sum is an orthogonal direct sum. Let P_j denote the orthogonal projection onto V_j and Q_j the orthogonal projection onto W_j. Show that

$$P_{j+1} = P_j + Q_j.$$

9.3 Prove Lemma 9.1.4.

9.4 Prove the result in (9.10).

9.5 Prove Proposition 9.2.9.

9.6 Derive (9.36) as outlined in the proof of Theorem 8.2.11.

9.7 The purpose of this exercise is to show that the function m in (9.33) belongs to $L^2(0,1)$.

(i) Show via (9.30) and (9.31) that the function λ introduced in the proof of Proposition 9.2.9 belongs to $L^2(0,1)$.

(ii) Conclude that the function m in (9.33) belongs to $L^2(0,1)$.

10
B-splines

Splines on $\mathbb{R}$ are functions which are piecewise polynomials. We will not discuss general splines, but only some special splines, called *B-splines.* These splines have attractive features: for example, they have compact support and are given by explicit and quite simple formulas, in the time domain as well as in the frequency domain. For these (and other) reasons, B-splines play an important role in applied mathematics, geometric modelling, and many other areas. The theoretical properties of B-splines and the broad range of applications make them natural to include here.

In Section 10.1 we introduce B-splines supported on bounded subintervals of the positive real axis. We derive explicit expressions for these functions and their Fourier transforms. Furthermore, we prove that the integer-translates of any B-spline possess a so-called partition of unity property. In some cases it is more convenient to consider B-splines that are supported on subsets of $\mathbb{R}$ that are symmetric around the y-axis. Such B-splines are obtained by a simple translation of the B-splines in Section 10.1, and they are treated in Section 10.2. In Sections 10.3 and 10.4 we briefly discuss wavelet constructions based on B-splines.

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis,
DOI 10.1007/978-0-8176-4980-7_10,

10.1 The B-splines N_m

A *spline* on $\mathbb{R}$ is a function $f : \mathbb{R} \to \mathbb{R}$ for which one can split $\mathbb{R}$ into intervals in such a way that f is a polynomial on each interval. The points where the function changes from one polynomial to another polynomial are called *knots.*

Example 10.1.1 The function (make a draft!)

$$f(x) \quad := \quad \begin{cases} 0 & \text{if } x^2 \in]-\infty, 0], \\ 2x^2 & \text{if } x^2 \in]0, 1], \\ 2 - x & \text{if } x \in]1, 4], \\ \frac{1}{16}x^3 & \text{if } x \in]4, \infty[, \end{cases}$$

is a spline. The knots are $x = 0, x = 1$, and $x = 4$. □

As we see, splines are not necessarily continuous and might not have compact support. We will now introduce a class of splines with very attractive features — the *B-splines* $N_m,\ m \in \mathbb{N}$. We note in advance that the definition does not immediately show that the B-splines actually are splines, but this will be shown in Corollary 10.1.7. The B-splines are defined inductively:

Definition 10.1.2 (B-splines) *The B-spline N_1 is defined as the characteristic function for the interval $[0, 1]$:*

$$N_1(x) := \chi_{[0,1]}(x), \ x \in \mathbb{R}. \tag{10.1}$$

Assuming that we have defined N_m for some $m \in \mathbb{N}$, the B-spline N_{m+1} is defined by the convolution

$$N_{m+1}(x) := (N_m * N_1)(x), \ x \in \mathbb{R}. \tag{10.2}$$

The B-spline N_m is said to have order m.

The definition of convolution shows that

$$N_{m+1}(x) = \int_{-\infty}^{\infty} N_m(x-t)N_1(t)\,dt = \int_0^1 N_m(x-t)\,dt. \tag{10.3}$$

See Figure 10.1 for graphs of the first few B-splines. We now collect some of their fundamental properties. In particular, the result shows that the support of the B-splines N_m increases with m:

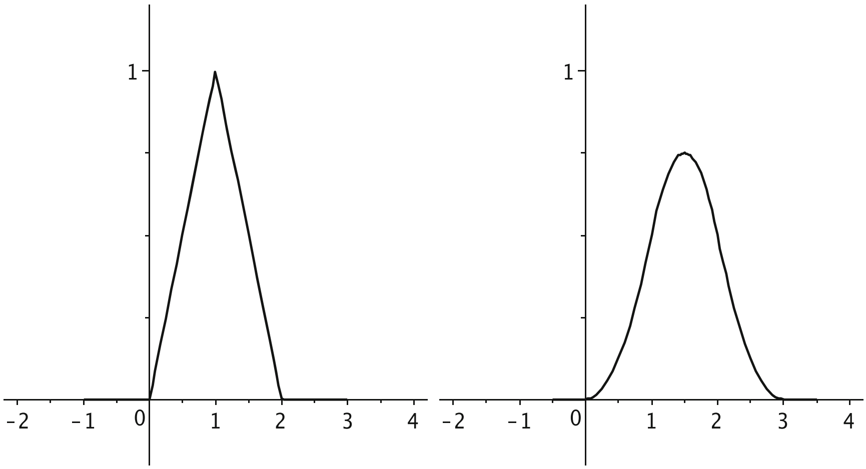

Figure 10.1. The B-splines N_2 and N_3, respectively.

Theorem 10.1.3 (Properties of B-splines) *Given $m \in \mathbb{N}$, the B-spline N_m has the following properties:*

(i) supp $N_m = [0, m]$ and $N_m(x) > 0$ for $x \in]0, m[$.

(ii) $\int_{-\infty}^{\infty} N_m(x)\, dx = 1$.

(iii) For $m \geq 2$,

$$\sum_{k \in \mathbb{Z}} N_m(x-k) = 1 \text{ for all } x \in \mathbb{R}; \tag{10.4}$$

for $m = 1$, the formula (10.4) holds for all $x \in \mathbb{R} \setminus \mathbb{Z}$.

Proof. All the statements can be proved by induction based on (10.3); we prove (i), and leave the rest to the reader (Exercise 10.1). It is clear that (i) holds for $m = 1$. Assuming that (i) holds for some $m \in \mathbb{N}$, we now consider the B-spline N_{m+1}. Whenever $t \in [0, 1]$, the induction hypothesis shows that it only is possible for $N_m(x-t)$ to be nonzero if $x \in [0, m+1]$, so (10.3) implies that supp $N_{m+1} \subseteq [0, m+1]$. On the other hand, if $x \in]0, m+1[$, then there exists $t \in]0, 1[$ such that $x-t \in]0, m[$; by the induction hypothesis this implies that $N_m(x-t) > 0$. We would like to conclude from this that $N_{m+1}(x) > 0$. For $m = 1$ this follows from (10.3) and the definition of N_1; for $m \geq 2$, it also follows from (10.3), using that N_m is nonnegative and continuous.

We have now proved that $N_{m+1}(x) > 0$ for $x \in [0, m+1]$. In the above induction argument we already saw that supp $N_{m+1} \subseteq [0, m+1]$, so we can now conclude that actually supp $N_{m+1} = [0, m+1]$. This completes the proof of (i). □

The formula (10.4) shows that the integer-translates of any B-spline pointwise add up to 1; we say that the integer-translates of N_m form a *partition of unity.*

Iteration of the formula (10.2) defining N_m shows that for $m \geq 2$,

$$N_m = N_1 * N_1 * \cdots * N_1, \tag{10.5}$$

with m terms in the convolution. Via Theorem 7.3.4 this leads to an expression for the Fourier transform of N_m:

Corollary 10.1.4 (Fourier transform of B-splines) *For any $m \in \mathbb{N}$,*

$$\widehat{N_m}(\gamma) = \left(\frac{1 - e^{-2\pi i\gamma}}{2\pi i\gamma}\right)^m. \tag{10.6}$$

Proof. Using the definition of the Fourier transform, we see that

$$\begin{aligned}
\widehat{N_1}(\gamma) &= \int_{-\infty}^{\infty} N_1(x) e^{-2\pi i x\gamma}\, dx \\
&= \int_0^1 e^{-2\pi i x\gamma}\, dx \\
&= \frac{1 - e^{-2\pi i\gamma}}{2\pi i\gamma}.
\end{aligned}$$

This shows that the desired result holds for the B-spline N_1. Using Theorem 7.3.4 and (10.5) we now obtain the result for the general B-spline N_m:

$$\widehat{N_m}(\gamma) = \left(\widehat{N_1}(\gamma)\right)^m = \left(\frac{1 - e^{-2\pi i\gamma}}{2\pi i\gamma}\right)^m. \qquad \square$$

We will now state an alternative expression for the B-splines N_m. For a real-valued function f, let

$$f(x)_+ := \max(0, f(x)).$$

Also, for any $n \in \mathbb{N}_0$, let

$$f(x)_+^n := \left(f(x)_+\right)^n.$$

Finally, for $n \in \mathbb{N}$ and $j = 0, 1, \ldots, n$, we need the binomial coefficient

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}.$$

Theorem 10.1.5 (Explicit formula for the B-splines) *For each $m = 2, 3, \dots,$ the B-spline N_m can be written*

$$N_m(x) = \frac{1}{(m-1)!} \sum_{j=0}^{m} (-1)^j \binom{m}{j} (x-j)_+^{m-1}, \ x \in \mathbb{R}. \tag{10.7}$$

The proof of Theorem 10.1.5 is quite lengthy and is given in Appendix A.3. Theorem 10.1.5 has direct consequences concerning differentiability of the functions N_m. In order to explain this we consider the following example.

Example 10.1.6 (Differentiability of x_+^m) For $m = 2, 3, \dots,$ consider the function

$$x_+^{m-1} = (\max(0,x))^{m-1} = \begin{cases} 0 & \text{if } x \leq 0, \\ x^{m-1} & \text{if } x \geq 0. \end{cases} \tag{10.8}$$

For the first few values of m we obtain the following:

- For $m = 2$,

  $$x_+^1 = \max(0,x) = \begin{cases} 0 \ \text{ if } & x \leq 0, \\ x \ \text{ if } & x \geq 0; \end{cases}$$

  this function is continuous for all $x \in \mathbb{R}$, but not differentiable for $x = 0$.

- For $m = 3$,

  $$x_+^2 = (\max(0,x))^2 = \begin{cases} 0 & \text{if } x \leq 0, \\ x^2 & \text{if } x \geq 0; \end{cases}$$

  this function is differentiable for all $x \in \mathbb{R}$, and

  $$\left(x_+^2\right)' = \begin{cases} 0 & \text{if } x \leq 0, \\ 2x & \text{if } x \geq 0. \end{cases}$$

  So $\left(x_+^2\right)'$ is continuous for all $x \in \mathbb{R}$, but not differentiable for $x = 0$.

- In general, for any $m = 2, 3, \dots,$ the function x_+^{m-1} has derivatives up to order $m-2$, and the $(m-2)$th derivative is continuous. □

Via Theorem 10.1.5 and the considerations in Example 10.1.6, we see that the B-spline N_2 is continuous for all $x \in \mathbb{R}$, but not differentiable for $x \in \{0, 1, 2\}$; and N_3 is differentiable, but N_3' is a continuous function that is not differentiable for $x \in \{0, 1, 2, 3\}$. See Figure 10.1 and Exercise 10.2. The general statement is as follows:

Corollary 10.1.7 (Regularity of B-splines) *For $m = 2, 3, \ldots,$ the B-spline N_m has the following properties:*

(i) $N_m \in C^{m-2}(\mathbb{R})$.

(ii) The restriction of N_m to each interval $[k, k+1]$, $k \in \mathbb{Z}$, is a polynomial of degree at most $m-1$.

Proof. The formula in Theorem 10.1.5 shows that N_m is a linear combination of translated versions of the functions x_+^{m-1}; thus, the result in (i) follows immediately from Example 10.1.6.

We now prove (ii). The expression (10.8) implies that for $j = 0, 1, \ldots, m$,

$$(x-j)_+^{m-1} = (\max(0, x-j))^{m-1} = \begin{cases} 0 & \text{if } x \leq j, \\ (x-j)^{m-1} & \text{if } x \geq j. \end{cases} \tag{10.9}$$

The result in (ii) is an immediate consequence of (10.9) and the formula in Theorem 10.1.5. □

10.2 The centered B-splines B_m

We will now consider a centered version of the B-splines discussed in Section 10.1. That is, we will translate the B-splines such that they get centered at the y-axis. For that purpose we will employ the translation operator introduced in (6.7):

Definition 10.2.1 (Centered B-splines) *For $m \in \mathbb{N}$, the (centered) B-spline B_m is defined by*

$$B_m(x) := T_{-\frac{m}{2}} N_m(x) = N_m(x + \frac{m}{2}). \tag{10.10}$$

Alternatively, the B-splines B_m can be defined by

$$B_1 := \chi_{[-1/2, 1/2]}, \quad B_{m+1} := B_m * B_1, \; m \in \mathbb{N}; \tag{10.11}$$

this definition leads to the same functions, see Exercise 10.5. Thus, for any $m \in \mathbb{N}$, we have that

$$B_{m+1}(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} B_m(x-t)\, dt.$$

Note that an explicit expression for B_m can be obtained via the definition (10.10) together with Theorem 10.1.5.

We state the following consequences of Theorem 10.1.3 and Corollary 10.1.4 and ask the reader to provide the proof in Exercise 10.7:

Corollary 10.2.2 (Properties of centered B-splines) *For $m \in \mathbb{N}$, the B-spline B_m has the following properties:*

(i) *supp* $B_m = [-m/2, m/2]$.

(ii) *For* $m \geq 2$,

$$\sum_{k\in\mathbb{Z}} B_m(x-k) = 1 \quad \text{for all } x \in \mathbb{R}.$$

For $m = 1$, *the formula holds for all* $x \in \mathbb{R} \setminus \{\pm\frac{1}{2}, \pm\frac{3}{2}, \dots\}$.

(iii) $\widehat{B_m}(\gamma) = \left(\frac{e^{\pi i\gamma} - e^{-\pi i\gamma}}{2\pi i\gamma}\right)^m = \left(\frac{\sin(\pi\gamma)}{\pi\gamma}\right)^m$.

10.3 B-splines and wavelet expansions

In Example 8.2.3, we saw that the Haar wavelet can be derived based on a multiresolution analysis with the scaling function $\phi = \chi_{[0,1]}$. With our knowledge from the current chapter we recognize the function ϕ as the B-spline N_1. This raises the natural question whether we can generate wavelets using higher order B-splines. Indeed, the Haar wavelet is a special case of a *spline wavelet:* one can consider splines N_m of arbitrary order and define associated multiresolution analyses, which lead to wavelets of the type

$$\psi(x) = \sum_{k\in\mathbb{Z}} c_k N_m(2x-k). \tag{10.12}$$

We will now describe how to do that. We have outlined the necessary technical steps in a series of exercises, so the purpose of the description is just to connect the results.

The starting point is Theorem 8.2.11, which provides us with convenient conditions for a function $\phi \in L^2(\mathbb{R})$ to generate a multiresolution analysis. The central condition in Theorem 8.2.11 is that the function ϕ should satisfy a scaling equation

$$\hat{\phi}(2\gamma) = H_0(\gamma)\hat{\phi}(\gamma)$$

for a bounded 1-periodic function H_0. For the case $\phi = N_m$, Exercise 10.8 shows that such an equation is available.

However, Theorem 8.2.11 also requires that the functions $\{T_k\phi\}_{k\in\mathbb{Z}}$ form an orthonormal system. For $\phi = N_m$ with $m \geq 2$ this condition is not satisfied (why?). Exercise 10.11 shows how to circumvent this problem. In fact, introducing the function

$$G(\gamma) := \sum_{k\in\mathbb{Z}} |\widehat{N_m}(\gamma+k)|^2,$$

it is shown that the function φ defined via its Fourier transform $\hat{\varphi}$ by

$$\hat{\varphi}(\gamma) := \frac{1}{\sqrt{G(\gamma)}} \widehat{N_m}(\gamma) \tag{10.13}$$

satisfies all the requirements in Theorem 8.2.11. According to the general theory outlined in Theorem 8.2.7 and Proposition 8.2.8 this leads to a wavelet of the form

$$\psi = \sum_{k \in \mathbb{Z}} d_k \varphi(2x + k). \tag{10.14}$$

In order to come from (10.14) to (10.12) we need to show that the function φ is a linear combination of integer-translated versions of N_m; this follows from (10.13) with a calculation that is similar to the one we used in the proof of Proposition 8.2.8. Putting everything together, we have arrived at the following result:

Theorem 10.3.1 (B-spline wavelets) *For any $m \in \mathbb{N}$, there exists a wavelet of the form*

$$\psi(x) = \sum_{k \in \mathbb{Z}} c_k N_m(2x - k). \tag{10.15}$$

The wavelets constructed above are known under the name *Battle–Lemarié wavelets.* The coefficients $\{c_k\}_{k \in \mathbb{Z}}$ in (10.15) are calculated, e.g., in [8]. Except for the case $m = 1$, all coefficients c_k are nonzero, which implies that all these wavelets have support equal to $\mathbb{R}$. However, the wavelets decay very fast: one can prove that for some $C, \alpha > 0$,

$$|\psi(x)| \leq Ce^{-\alpha|x|}, \ x \in \mathbb{R}.$$

Decay of that type is called *exponential decay.*

10.4 Frames generated by B-splines

In Section 10.3 we saw that one can construct wavelets of the form (10.12) for any mth order B-spline N_m. For concrete applications it is desirable to have wavelets with compact support, but unfortunately this is not the case for the Battle–Lemarié wavelets. Note that a *finite* linear combination of the form (10.12), i.e., a function

$$\psi(x) = \sum_{k=-K}^{K} c_k N_m(2x - k) \tag{10.16}$$

for some $K \in \mathbb{N}$, does have compact support. Thus, it is natural to ask whether a function of the form (10.16) can be a wavelet. Unfortunately, the answer is no for $m > 1$.

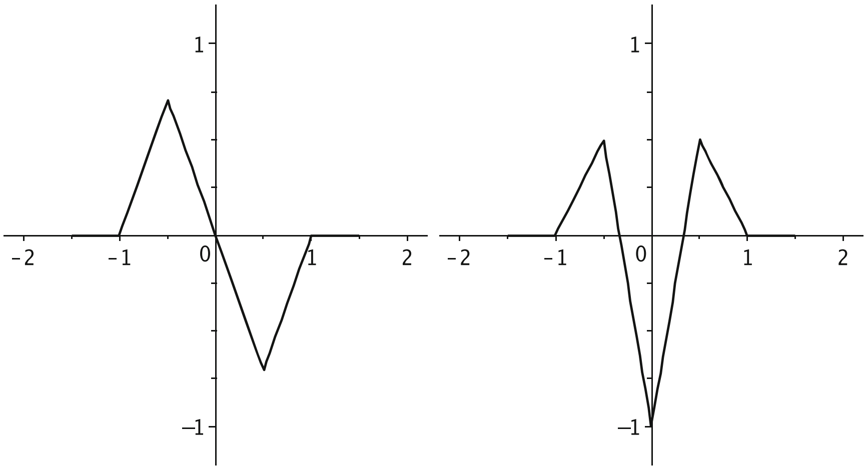

Figure 10.2. The function ψ_1 given by (10.17), and the function ψ_2 given by (10.18).

In Section 4.8 we saw that tight frames lead to expansions that are similar to the one we know from orthonormal bases; see (4.26) and (4.28). With this in mind it is natural to examine whether we can construct tight frames $\{D^j T_k \psi\}_{j,k\in\mathbb{Z}}$ for $L^2(\mathbb{R})$, for functions ψ of the form (10.12). The answer is again no for $m > 1$. However, one can prove that for any choice of $m \in \mathbb{N}$ it is possible to construct a tight frame generated by *two* such functions. That is, there exist functions ψ_1 and ψ_2 of the form

$$\psi_1(x) = \sum_{k=-K}^{K} c_k N_m(2x-k), \;\; \psi_2(x) = \sum_{k=-K}^{K} d_k N_m(2x-k),$$

with the property that the functions

$$\{D^j T_k \psi_1\}_{j,k\in\mathbb{Z}} \cup \{D^j T_k \psi_2\}_{j,k\in\mathbb{Z}}$$

form a tight frame. The construction is too complicated to be presented here; we refer to the original article [16] or the book [5] for a more detailed description.

If m is even, the same procedure works for the centered B-spline B_m. In the case $m = 2$, one choice of the functions ψ_1 and ψ_2 is

$$\psi_1(x) = \frac{1}{\sqrt{2}}(B_2(2x+1) - B_2(2x-1)) \tag{10.17}$$

and

$$\psi_2(x) = \frac{1}{2}\left(B_2(2x+1) - 2B_2(2x) + B_2(2x-1)\right). \tag{10.18}$$

The graphs of the functions ψ_1 and ψ_2 are shown on Figure 10.2. Similar expressions can be obtained for any B-spline B_m: increasing m will increase the smoothness of the functions ψ_1 and ψ_2, but also the size of the support.

10.5 Exercises

10.1 Prove Theorem 10.1.3(ii)–(iii).

10.2 (i) Show via the definition that the B-spline N_2 is given by

$$N_2(x) = \begin{cases} x & \text{if } x \in [0,1], \\ 2-x & \text{if } x \in [1,2], \\ 0 & \text{otherwise.} \end{cases}$$

(ii) Use this to show that (10.7) holds for $m = 2$.

(iii) Show that

$$N_3(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } x \in [0,1], \\ -x^2 + 3x - \frac{3}{2} & \text{if } x \in [1,2], \\ \frac{1}{2}x^2 - 3x + \frac{9}{2} & \text{if } x \in [2,3], \\ 0 & \text{otherwise.} \end{cases}$$

(iv) Calculate the derivative $N_3'(x)$ and plot it.

10.3 Show that (10.7) holds for $x > m + 1$.

10.4 Show that the B-splines B_2 and B_3 are given by

$$B_2(x) = \begin{cases} 1+x & \text{if } x \in [-1,0], \\ 1-x & \text{if } x \in [0,1], \\ 0 & \text{otherwise,} \end{cases}$$

$$B_3(x) = \begin{cases} \frac{1}{2}x^2 + \frac{3}{2}x + \frac{9}{8} & \text{if } x \in [-\frac{3}{2}, -\frac{1}{2}], \\ -x^2 + \frac{3}{4} & \text{if } x \in [-\frac{1}{2}, \frac{1}{2}], \\ \frac{1}{2}x^2 - \frac{3}{2}x + \frac{9}{8} & \text{if } x \in [\frac{1}{2}, \frac{3}{2}], \\ 0 & \text{otherwise.} \end{cases}$$

10.5 Show that the definitions of the centered B-splines B_m in (10.10) and (10.11) coincide.

10.6 Consider the centered B-splines B_2 and B_3. Find an expression for the function $\widehat{B_2 * B_3}$.

10.7 Prove Corollary 10.2.2.

10.8 (B-spline scaling equation) Consider the B-spline $N_m,\ m \in \mathbb{N}$.

(i) Show that the scaling equation

$$\widehat{N_m}(2\gamma) = H_0(\gamma)\widehat{N_m}(\gamma),\ \forall \gamma\ \in \mathbb{R}$$

is satisfied with

$$H_0(\gamma) = \left(\frac{1 + e^{-2\pi i\gamma}}{2}\right)^m.$$

(ii) Show that H_0 is periodic with period 1.

10.9 (B-spline scaling equation) Let $m \in \mathbb{N}$ and consider the centered B-spline $B_m,\ m \in \mathbb{N}$.

(i) Determine the function H_0 such that

$$\widehat{B_m}(2\gamma) = H_0(\gamma)\widehat{B_m}(\gamma),\ \forall \gamma\ \in \mathbb{R}.$$

(ii) Is H_0 periodic with period 1?

10.10 Consider the B-spline $N_m,\ m \in \mathbb{N}$. The purpose of the exercise is to show that there exist constants $A, B > 0$ such that

$$A \leq \sum_{k \in \mathbb{Z}} |\widehat{N_m}(\gamma + k)|^2 \leq B,\ \gamma \in \mathbb{R}. \tag{10.19}$$

(i) Show that (10.19) holds for $m = 1$, with $A = B = 1$.
Hint: use the result in Theorem 8.2.12.

(ii) Show that for any $m \in \mathbb{N}$,

$$\sum_{k \in \mathbb{Z}} |\widehat{N_m}(\gamma + k)|^2 \leq 1,\ \gamma \in \mathbb{R}.$$

Hint: use the result in (i) together with Corollary 10.1.4.

The exercise continues!

(iii) Show that the function

$$G(\gamma) := \sum_{k\in\mathbb{Z}} |\widehat{N_m}(\gamma+k)|^2$$

is 1-periodic.

(iv) Show that for $\gamma \in \mathbb{R}$,

$$\sum_{k\in\mathbb{Z}} |\widehat{N_m}(\gamma+k)|^2 \geq \inf_{\gamma\in[-1/2,1/2]} |\widehat{N_m}(\gamma)|^2.$$

(v) Show that for any $m \in \mathbb{N}$,

$$\left(\frac{2}{\pi}\right)^{2m} \leq \sum_{k\in\mathbb{Z}} |\widehat{N_m}(\gamma+k)|^2, \ \gamma \in \mathbb{R}.$$

10.11 (B-spline multiresolution analysis) Let $m \in \mathbb{N}$, and consider the B-spline N_m. One can show (Exercise 10.10) that there exist constants $A, B > 0$ such that

$$A \leq \sum_{k\in\mathbb{Z}} |\widehat{N_m}(\gamma+k)|^2 \leq B, \ \gamma \in \mathbb{R}.$$

(i) Show that the function

$$G(\gamma) := \sum_{k\in\mathbb{Z}} |\widehat{N_m}(\gamma+k)|^2$$

is 1-periodic.

Define the function $\varphi \in L^2(\mathbb{R})$ by its Fourier transform $\hat{\varphi}$ via

$$\hat{\varphi}(\gamma) := \frac{1}{\sqrt{G(\gamma)}}\widehat{N_m}(\gamma).$$

(ii) Show that

$$\sum_{k\in\mathbb{Z}} |\widehat{\varphi}(\gamma+k)|^2 = 1, \ \gamma \in \mathbb{R}.$$

Hint: use the result in (i).

(iii) Show that there exists a 1-periodic function M_0 such that

$$\hat{\varphi}(2\gamma) = M_0(\gamma)\hat{\varphi}(\gamma), \ \gamma \in \mathbb{R}.$$

Hint: use the result in Exercise 10.8.

(iv) Use Theorem 8.2.11 to verify that the function φ generates a multiresolution analysis.

11
Special Functions

In this chapter we introduce some classes of special functions that play important roles in applied mathematics, physics, etc. Most of these functions are polynomials, and they appear as solutions to various differential equations. The analysis of these special functions is intimately connected with the main theme of the book: in fact, for each of the considered classes of differential equations, the associated polynomial solutions form an orthonormal basis for a related L^2-space. The study of the differential equations and their solutions can easily cover an entire book. We do not aim at a complete description with full proofs, but will focus on the relationship with the theory we have developed for series expansions in Hilbert spaces.

Section 11.1 gives a general introduction to the relevant differential equations, and presents some of the fundamental results concerning the solutions. We focus on regular Sturm–Liouville problems. In Section 11.2, we consider Legendre's differential equation and derive a class of polynomial solutions, the Legendre polynomials. It is shown that these polynomials form an orthonormal basis for $L^2(-1,1)$. Sections 11.3 and 11.4 give short presentations of Laguerre polynomials, respectively, Hermite polynomials.

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis,
DOI 10.1007/978-0-8176-4980-7_11,

11.1 Regular Sturm–Liouville problems

The special functions introduced in the following sections are solutions to certain differential equations, typically appearing in physics and chemistry. Most of the functions are actually polynomials. For a given differential equation, such solutions can always be found by searching for general solutions in terms of power series, i.e., functions that can be written on the form

$$u(x) = \sum_{k=0}^{\infty} c_k x^k \tag{11.1}$$

for some scalar coefficients c_k. For functions of the type (11.1) it is well known that there exists a number $\rho \in [0, \infty]$ such that the infinite series is convergent for $x \in]-\rho, \rho[$ and divergent for $x \notin [-\rho, \rho]$; that number is called the *radius of convergence.*

Functions with a representation of the form (11.1) are called *analytic functions.* The class of analytic functions is very convenient: it contains all polynomials, and the standard trigonometric functions are analytic. Thus, a fundamental question to be answered is the following: for what kind of differential equations can we be sure that analytic solutions exist? The result below gives a partial answer to that question.

Theorem 11.1.1 (Analytic solutions to differential equations) *Assume that the functions a_1 and a_2 are analytic. Then the differential equation*

$$\frac{d^2u}{dx^2} + a_1(x)\frac{du}{dx} + a_2(x)u = 0$$

has a nonzero solution of the form (11.1)*; denoting the radius of convergence for the functions a_i by ρ_i, $i = 1, 2$, the radius of convergence for the solution is at least $\rho = \min(\rho_1, \rho_2)$.*

We will now define the type of differential equations to be considered.

Definition 11.1.2 (Sturm–Liouville differential equation) *Let p, q, and r be functions on $\mathbb{R}$ or a subinterval of $\mathbb{R}$; assume that the function p is differentiable, and that the functions p', q, and r are continuous. A differential equation that can be written on the form*

$$[p(x)u']' + [q(x) + \lambda r(x)]\, u = 0 \tag{11.2}$$

for some parameter $\lambda \in \mathbb{R}$, is called a Sturm–Liouville differential equation.

We note that the Sturm–Liouville differential equation (11.2) also can be written on the form

$$p(x)u'' + p'(x)u' + [q(x) + \lambda r(x)]\, u = 0. \tag{11.3}$$

At first glance, it might seem strange that the function to be multiplied with the function u in (11.2) is written as $q(x)+\lambda r(x)$ rather than just $q(x)$. The reason is that we often will consider *classes* of differential equations of the form (11.2), with the same functions q and r, but various values for the parameter λ.

For reasons to become clear soon, the function r is called a *weight function.* We will usually impose extra conditions on the Sturm–Liouville equations:

Definition 11.1.3 (Regular Sturm–Liouville problem) *A regular Sturm–Liouville problem consists of a differential equation of the type in Definition 11.1.2, considered on a finite closed interval* $[a, b]$*, with the following extra constraints:*

(i) $p(x) > 0$ *and* $r(x) > 0$ *for all* $x \in [a, b]$*;*

(ii) We search for a solution u *which, for given* $(c_1, c_2) \neq (0, 0)$ *and* $(d_1, d_2) \neq (0, 0)$*, satisfies the boundary value conditions*

$$\begin{cases} c_1 u(a) + c_2 u'(a) = & 0, \\ d_1 u(b) + d_2 u'(b) = & 0. \end{cases}$$

We will search for solutions to the Sturm–Liouville equation (11.2) that belong to the Hilbert space

$$L^2_r(a, b) := \left\{ f : \mathbb{R} \to \mathbb{C} \mid \int_a^b |f(x)|^2 r(x)\, dx < \infty \right\}.$$

The space $L^2_r(a, b)$ corresponds to the space $L^2_r(\mathbb{R})$ considered in Exercise 6.12, except that we now deal with functions on a finite interval. The inner product in $L^2_r(a, b)$ is given by

$$\langle f, g \rangle_{L^2_r(a,b)} = \int_a^b f(x)\overline{g(x)} r(x)\, dx, \;\; f, g \in L^2_r(\mathbb{R}).$$

We will exclusively consider real solutions, so at many instances we can drop the complex conjugation appearing in the inner product.

It turns out that there is a close connection between Sturm–Liouville differential equations and the theory for orthonormal bases in Hilbert spaces. In fact, under certain assumptions to be discussed in Theorem 11.1.10, a special collection of solutions to a Sturm–Liouville problem forms an orthonormal basis for the Hilbert space $L^2_r(a, b)$.

We need the concepts of *eigenvalues* and *eigenfunctions* for Sturm–Liouville problems:

Definition 11.1.4 (Eigenvalues and eigenfunctions for SL-problem) *If a Sturm–Liouville differential equation has a nonzero solution $u \in L_r^2(a,b)$, the number λ appearing in (11.2) is called an eigenvalue; the corresponding solution u is called an eigenfunction.*

Let us relate the use of the word "eigenvalues" to the one that is known in linear algebra:

Example 11.1.5 (Eigenvalues and eigenfunctions) Consider a Sturm-Liouville differential equation with weight function $r(x) = 1$. Then a parameter λ is an eigenvalue with associated eigenfunction $u \neq 0$ if

$$-\left(p(x)u'' + p'(x)u' + q(x)u\right) = \lambda u. \tag{11.4}$$

Let $C^2[a,b]$ denote the set of functions on $[a,b]$ that are twice differentiable with a continuous second derivative. The expression

$$\mathcal{D} : C^2[a,b] \to C[a,b],\ \mathcal{D}u := -\left(p(x)u'' + p'(x)u' + q(x)u\right)$$

defines a linear operator $\mathcal{D}$. Using the operator $\mathcal{D}$, the Sturm-Liouville problem (11.4) takes the form

$$\mathcal{D}u = \lambda u. \tag{11.5}$$

The formulation in (11.5) shows that the concepts of eigenvalues and eigenvectors in Definition 11.1.4 are similar to the ones considered in linear algebra. The differences are that in (11.5), u is a function instead of a vector in $\mathbb{R}^n$, and that $\mathcal{D}$ is an operator on an infinite-dimensional vector space instead of $\mathbb{R}^n$. □

Let us study a concrete example, where a Sturm-Liouville equation appears.

Example 11.1.6 (Vibrating string) We consider a string of length L, e.g., a piano string. We fix the string at the end points $x = 0$ and $x = L$. Assume that we act on the string with a force, pulling it away from the equilibrium position. Releasing the string at the time $t = 0$, it will start performing small vibrations: for $x \in [0, L]$, we will denote the displacement from the equilibrium position at time t by $u(x,t)$. Since we have fixed the string for $x = 0$ and $x = L$, we have

$$u(0,t) = u(L,t) = 0,\ t \geq 0. \tag{11.6}$$

One can show that for an appropriate choice of the units, the function u will satisfy the equation

$$\frac{\partial^2 u}{\partial t^2}(x,t) = \frac{\partial^2 u}{\partial x^2}(x,t). \tag{11.7}$$

We will search for solutions of (11.6) on the form

$$u(x,t) = f(x)g(t) \tag{11.8}$$

for suitable functions f and g. Direct calculations (outlined in Exercise 11.1) show that for some constant $\lambda \in \mathbb{R}$, the functions f and g have to satisfy the equations

$$f'' + \lambda f = 0, \tag{11.9}$$

$$g'' + \lambda g = 0. \tag{11.10}$$

Note also that (11.6) implies that

$$f(0) = f(L) = 0. \tag{11.11}$$

Together, (11.9) and (11.11) form a regular Sturm–Liouville problem. Direct calculations (Exercise 11.1) show that we only have nonzero solutions for λ of the form

$$\lambda_n = \frac{\pi^2 n^2}{L^2}, \; n \in \mathbb{N}, \tag{11.12}$$

and that the real solutions corresponding to λ_n are

$$f_n(x) = c \sin(\frac{\pi n x}{L}), \; c \in \mathbb{R}.$$

Thus, the eigenvalues are exactly the numbers λ_n in (11.12). With the choice $\lambda = \lambda_n$, the equation (11.10) has the real solutions

$$g_n(t) = a_n \cos(\frac{\pi n t}{L}) + b_n \sin(\frac{\pi n t}{L}), \; a_n, b_n \in \mathbb{R}.$$

Putting everything together, we obtain that (11.7) with the boundary conditions (11.6) has the solutions

$$u_n(x,t) = c \sin(\frac{\pi n x}{L}) \left(a_n \cos(\frac{\pi n t}{L}) + b_n \sin(\frac{\pi n t}{L}) \right), \; n \in \mathbb{N}.$$

Note that the outcome so far is a *class* of solutions: for each $n \in \mathbb{N}$ we have found a solution u_n. With knowledge about the force that pulls the string away from the equilibrium position it is now possible to find the corresponding unique solution of (11.7) as a linear combination of the functions u_n. We will not go into the discussion about how to do that. □

We will now prove that for a regular Sturm–Liouville problem, eigenfunctions corresponding to different eigenvalues are orthogonal in $L^2_r(a,b)$.

Theorem 11.1.7 (Orthogonality of solutions, regular case) *Let u_m and u_n be real-valued eigenfunctions for a regular Sturm–Liouville problem corresponding to different eigenvalues λ_m and λ_n. Then u_n and u_m are orthogonal in $L^2_r(a,b)$, i.e.,*

$$\int_a^b u_m(x)u_n(x)r(x)\,dx = 0. \tag{11.13}$$

Proof. By assumption,

$$\begin{aligned} [p(x)u_m']' + [q(x) + \lambda_m r(x)]u_m &= 0,\\ [p(x)u_n']' + [q(x) + \lambda_n r(x)]u_n &= 0. \end{aligned}$$

Multiplying the first equation by u_n and the second by u_m, this implies that

$$\left([p(x)u_m']' + [q(x) + \lambda_m r(x)]u_m\right) u_n - \left([p(x)u_n']' + [q(x) + \lambda_n r(x)]u_n\right) u_m = 0,$$

or

$$(\lambda_n - \lambda_m)r(x)u_m u_n = [p(x)u_m']'u_n - [p(x)u_n']'u_m. \tag{11.14}$$

By direct calculation of the term on the right-hand side of (11.14), this leads to

$$(\lambda_n - \lambda_m)r(x)u_m u_n = [(pu_m')u_n - (pu_n')u_m]'.$$

Integrating from a to b yields

$$\begin{aligned} &(\lambda_n - \lambda_m)\int_a^b r(x)u_m(x)u_n(x)\,dx\\ =\;& \big[(p(x)u_m'(x))u_n(x) - (p(x)u_n'(x))u_m(x)\big]_a^b\\ =\;& p(b)\Big(u_m'(b)u_n(b) - u_n'(b)u_m(b)\Big) \qquad (11.15)\\ & -p(a)\Big(u_m'(a)u_n(a) - u_n'(a)u_m(a)\Big). \qquad (11.16) \end{aligned}$$

By the boundary-value conditions, we know that

$$\begin{aligned} d_1 u_n(b) + d_2 u_n'(b) &= 0, \qquad (11.17)\\ d_1 u_m(b) + d_2 u_m'(b) &= 0, \qquad (11.18) \end{aligned}$$

for some $(d_1, d_2) \neq (0,0)$. Let us assume that $d_1 \neq 0$. Multiplying (11.17) with $u_m'(b)$ and (11.18) with $u_n'(b)$, a subtraction leads to

$$(d_1 u_n(b) + d_2 u_n'(b))u_m'(b) - (d_1 u_m(b) + d_2 u_m'(b))u_n'(b) = 0,$$

i.e., that

$$d_1(u_n(b)u_m'(b) - u_m(b)u_n'(b)) = 0.$$

Since $d_1 \neq 0$, we conclude that the term in (11.15) vanishes. In a similar way, one can prove that the term in (11.16) vanishes. Because $\lambda_n \neq \lambda_m$, this leads to the result in (11.13). □

The proof of Theorem 11.1.7 shows that a similar result holds under modified assumptions:

Corollary 11.1.8 (Orthogonality of solutions) *Consider a Sturm–Liouville differential equation as in Definition 11.1.2, on a finite interval* $[a, b]$. *Assuming that* $p(x) > 0$ *and* $r(x) > 0$ *for all* $x \in]a, b[$, *the following hold:*

(i) If $p(a) = p(b) = 0$, *the conclusion in Theorem 11.1.7 holds.*

(ii) If $p(a) = p(b)$ *and we add the boundary conditions*

$$u(a) = u(b) \quad \textit{and} \quad u'(a) = u'(b),$$

the conclusion in Theorem 11.1.7 holds.

We leave the proof of Corollary 11.1.8 to the reader (Exercise 11.2).

So far, we have not discussed the question of *existence* of eigenvalues for regular Sturm–Liouville problems. With more advanced tools from functional analysis than we have developed here, one can show the following general result about the existence and properties of the eigenvalues:

Lemma 11.1.9 *For any regular Sturm–Liouville problem, the set of eigenvalues form a discrete set* $\{\lambda_n\}_{n=1}^{\infty}$ *of real numbers. The sequence can be ordered increasingly,*

$$\ldots \lambda_1 < \lambda_2 < \ldots < \lambda_n < \ldots,$$

and $\lambda_n \to \infty$ *as* $n \to \infty$.

For a proof of Lemma 11.1.9 we refer to [3]. We note that the properties described in Lemma 11.1.9 correspond to the eigenvalues in (11.12) for the case of the vibrating string.

For each eigenvalue λ_n associated with a regular Sturm–Liouville problem, we now choose an eigenfunction u_n that is *normalized* in the Hilbert space $L_r^2(a, b)$: this means that

$$||u_n||_{L_r^2(a,b)} = \left(\int_a^b |u_n(x)|^2 r(x)\, dx \right)^{1/2} = 1.$$

The orthonormality of the functions u_n in $L_r^2(a, b)$ makes it natural to ask whether they form an orthonormal basis for $L_r^2(a, b)$. By Theorem 4.7.2

this is equivalent to

$$\overline{\text{span}}\{u_n\}_{n=1}^{\infty} = L_r^2(a,b). \tag{11.19}$$

Indeed, (11.19) holds! We state the result formally, and refer again to [3] for a proof.

Theorem 11.1.10 (Orthonormal basis of eigenfunctions) *Consider a regular Sturm–Liouville problem on a finite interval* $[a,b]$. *Let* $\{\lambda_n\}_{n=1}^{\infty}$ *denote the collection of eigenvalues and* $\{u_n\}_{n=1}^{\infty}$ *the associated eigenfunctions, normalized in* $L_r^2(a,b)$. *Then* $\{u_n\}_{n=1}^{\infty}$ *constitute an orthonormal basis for* $L_r^2(a,b)$. *In particular, each function* $f \in L_r^2(a,b)$ *has an expansion*

$$f = \sum_{n=1}^{\infty} \langle f, u_n \rangle_{L_r^2(a,b)} u_n, \tag{11.20}$$

with convergence in $L_r^2(a,b)$. *For piecewise smooth functions* f*, the expansion* (11.20) *holds pointwise at all points of continuity of* f*; at a point* x *of discontinuity, the series in* (11.20) *converges to*

$$\frac{1}{2}\Big(\lim_{t\to x^+} f(t) + \lim_{t\to x^-} f(t)\Big).$$

Note that the result about pointwise convergence for the expansion (11.20) is similar to the well-known result for pointwise convergence of Fourier series.

In Sections 11.2–11.4, we will consider various differential equations appearing in physics and engineering. Some of them are on the Sturm–Liouville form; and some of them can be rewritten into differential equations on Sturm–Liouville form that have the same solutions.

11.2 Legendre polynomials

As our first example of a Sturm–Liouville equation, we consider the so-called *Legendre's differential equation.* It is actually a collection of differential equations: for a given and fixed parameter $\lambda \in \mathbb{R}$, the equation is

$$(1-x^2)\frac{d^2u}{dx^2} - 2x\frac{du}{dx} + \lambda u = 0. \tag{11.21}$$

The differential equation can also be written on the form

$$[(1-x^2)u']' + \lambda u = 0. \tag{11.22}$$

This shows that Legendre's differential equation is a Sturm–Liouville differential equation, with parameter λ and

$$p(x) = 1 - x^2,\ q(x) = 0,\ r(x) = 1. \tag{11.23}$$

As starting point, we will relate Legendre's differential equation to the concepts and results in Section 11.1.

Example 11.2.1 We have not imposed any boundary conditions like the ones in Definition 11.1.3(ii) on Legendre's differential equation, so we are not dealing with a regular Sturm–Liouville problem. On the other hand, the positivity condition in Definition 11.1.3(i) is satisfied on $]a, b[$ if we let $a = -1, b = 1$. Since the weight function is $r(x) = 1$, this suggests that the analysis of Legendre's differential equation shall take place in the Hilbert space

$$L_1^2(-1,1) = L^2(-1,1).$$

Note that $p(-1) = p(1) = 0$; thus, Corollary 11.1.8(i) implies that the eigenfunctions for (11.21) corresponding to different eigenvalues are orthogonal in $L^2(-1,1)$. □

One can prove that Legendre's differential equation only has nonzero solutions in $L^2(-1,1)$ if $\lambda = \ell(\ell+1)$ for some $\ell \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For this reason, we will now concentrate on the class of differential equations given by

$$(1-x^2)\frac{d^2u}{dx^2} - 2x\frac{du}{dx} + \ell(\ell+1)u = 0 \tag{11.24}$$

for some $\ell \in \mathbb{N}_0$. Using the method of power series, we can find a solution of (11.24) for each value of $\ell \in \mathbb{N}_0$. In the statement and proof of the result we need to split into two cases, depending on ℓ being even or odd:

Theorem 11.2.2 (Legendre polynomials) *Let $\ell \in \mathbb{N}_0$; put $m = \ell/2$ if ℓ is even, and $m = (\ell-1)/2$ if ℓ is odd. Then (11.24) has the solution*

$$P_\ell(x) = \frac{1}{2^\ell}\sum_{k=0}^{m} \frac{(-1)^k}{k!}\frac{(2\ell-2k)!}{(\ell-2k)!(\ell-k)!}x^{\ell-2k}. \tag{11.25}$$

The proof of Theorem 11.2.2 is quite lengthy and is given in Appendix A.4.

The polynomial P_ℓ in (11.25) is called the *ℓth Legendre polynomial.* 13 contains a list of the first few Legendre polynomials. Note that the factor $2^{-\ell}$ in (11.25) appears as a matter of convenience: we would also have a solution of the Legendre equation without that factor. We collect some of the properties of the polynomials P_ℓ:

Lemma 11.2.3 (Properties of Legendre polynomials) *The polynomials P_ℓ, $\ell \in \mathbb{N}_0$, have the following properties:*

(i) *P_ℓ is a polynomial of degree ℓ.*

(ii) *If ℓ is even, then P_ℓ is an even function; if ℓ is odd, then P_ℓ is an odd function. That is,*

$$P_\ell(-x) = (-1)^\ell P_\ell(x), \ x \in \mathbb{R}, \ \ell \in \mathbb{N}_0.$$

(iii) *For each $m \in \mathbb{N}_0$, the polynomials $\{P_\ell\}_{\ell=0}^m$ are linearly independent. In particular, the span of the polynomials $\{P_\ell\}_{\ell=0}^m$ equals the vector space of all polynomials of degree at most m:*

$$span\{P_\ell(x)\}_{\ell=0}^m = span\{x^\ell\}_{\ell=0}^m.$$

Proof. The statement (i) follows from (11.25) by considering the term corresponding to $k = 0$. If ℓ is even, the expression (11.25) shows that P_ℓ is a linear combination of polynomials $x^\ell, x^{\ell-2}, \ldots, x^2, 1$, and hence an even function. A similar consideration proves the second statement in (ii). In order to prove (iii), we first note that because P_ℓ is a polynomial of degree ℓ, we have that

$$\text{span}\{P_\ell\}_{\ell=0}^m \subseteq \text{span}\{x^\ell\}_{\ell=0}^m. \tag{11.26}$$

From the fact that P_ℓ is a polynomial of degree ℓ, it follows that the polynomials $\{P_\ell\}_{\ell=0}^m$ are linearly independent (Exercise 1.6). That is, both spaces in (11.26) have dimension $m + 1$; therefore, they are equal. □

Rodrigues' formula provides us with an alternative expression for the Legendre polynomials:

Theorem 11.2.4 (Rodrigues' formula) *For each $\ell \in \mathbb{N}_0$,*

$$P_\ell(x) = \frac{1}{2^\ell \ell!} \frac{d^\ell}{dx^\ell} (x^2 - 1)^\ell.$$

Proof. Applying the binomial formula, see Lemma 1.9.2, we obtain that

$$\begin{aligned} (x^2 - 1)^\ell &= \sum_{k=0}^{\ell} \binom{\ell}{k} (-1)^k (x^2)^{\ell-k} \\ &= \sum_{k=0}^{\ell} \binom{\ell}{k} (-1)^k x^{2(\ell-k)}. \end{aligned} \tag{11.27}$$

It follows that

$$\frac{d}{dx}(x^2 - 1)^\ell = \sum_{k=0}^{\ell-1} \binom{\ell}{k} (-1)^k (2\ell - 2k) x^{2(\ell-k)-1}.$$

Notice the change in the number of terms in the summation, due to the fact that the power series for $(x^2-1)^\ell$ in (11.27) contains a constant term corresponding to $k=\ell$. Differentiating once more, we obtain that

$$\frac{d^2}{dx^2}(x^2-1)^\ell = \sum_{k=0}^{\ell-1}\left(\begin{array}{c}\ell\\k\end{array}\right)(-1)^k(2\ell-2k)(2\ell-2k-1)x^{2(\ell-k)-2};$$

because no constant term appears in the power series for $\frac{d}{dx}(x^2-1)^\ell$, no change in the number of terms in the summation appears. Continuing differentiating, and letting $m=\ell/2$ if ℓ is even, and $m=(\ell-1)/2$ if ℓ is odd (like in Theorem 11.2.2), we finally obtain that

$$\begin{aligned}&\frac{d^\ell}{dx^\ell}(x^2-1)^\ell\\ =&\ \sum_{k=0}^{m}\left(\begin{array}{c}\ell\\k\end{array}\right)(-1)^k(2\ell-2k)(2\ell-2k-1)\cdots(2\ell-2k-\ell+1)x^{2(\ell-k)-\ell}\\ =&\ \sum_{k=0}^{m}(-1)^k\frac{\ell!}{k!(\ell-k)!}\frac{(2\ell-2k)!}{(\ell-2k)!}x^{\ell-2k}.\end{aligned}$$

Thus,

$$\begin{aligned}\frac{1}{2^\ell\ell!}\frac{d^\ell}{dx^\ell}(x^2-1)^\ell &= \frac{1}{2^\ell\ell!}\sum_{k=0}^{m}(-1)^k\frac{\ell!}{k!(\ell-k)!}\frac{(2\ell-2k)!}{(\ell-2k)!}x^{\ell-2k}\\ &= \frac{1}{2^\ell}\sum_{k=0}^{m}\frac{(-1)^k}{k!}\frac{(2\ell-2k)!}{(\ell-2k)!(\ell-k)!}x^{\ell-2k}.\end{aligned}$$

This corresponds exactly to the expression for P_ℓ in Theorem 11.2.2, and the proof is completed. □

Via the explicit formula for P_ℓ in Theorem 11.2.2 or Rodrigues' formula it is easy to calculate the first few Legendre polynomials — see Exercise 11.3.

With an elementary but tedious calculation based on Rodrigues' formula, one can show that the Legendre polynomials satisfy a recursion formula:

Lemma 11.2.5 (Recursion formula for Legendre polynomials) *For each $\ell\in\mathbb{N}$,*

$$(\ell+1)P_{\ell+1}(x)-(2\ell+1)xP_\ell(x)+\ell P_{\ell-1}(x)=0.$$

Thus, as soon as we have calculated P_0 and P_1, the formula tells us how to construct all the following polynomials. A formula of the type in Lemma 11.2.5 is typical for the polynomial solutions of Sturm–Liouville equations.

We will now connect the Legendre polynomials with the theory for the Hilbert space $L^2(-1,1)$ considered in Section 6.3. Our goal is to show that the collection of all Legendre polynomials (with an appropriate normalization) forms an orthonormal basis for $L^2(-1,1)$. We start with a lemma:

Lemma 11.2.6 (Orthogonality of Legendre polynomials) *The Legendre polynomials $\{P_\ell\}_{\ell=0}^{\infty}$ satisfy the following relations:*

(i) For each $\ell \neq \ell'$,

$$\int_{-1}^{1} P_\ell(x) P_{\ell'}(x)\, dx = 0. \tag{11.28}$$

(ii) For each $\ell \in \mathbb{N}_0$,

$$\int_{-1}^{1} |P_\ell(x)|^2\, dx = \frac{2}{2\ell+1}. \tag{11.29}$$

Proof. We already saw in Example 11.2.1 that Corollary 11.1.8(i) implies the orthogonality result in (i).

In order to prove (ii), we will use an induction argument. We leave it to the reader to verify the result for $\ell = 0$ and $\ell = 1$. Now, note that by Lemma 11.2.5, the following two equations hold for all $\ell \geq 2$:

$$\begin{aligned} (\ell+1)P_{\ell+1}(x) + \ell\, P_{\ell-1}(x) &= (2\ell+1)xP_\ell(x), \\ \ell P_\ell(x) + (\ell-1)P_{\ell-2}(x) &= (2\ell-1)xP_{\ell-1}(x). \end{aligned}$$

Multiplying the first equation by $P_{\ell-1}(x)$ and the second by $P_\ell(x)$ yields that

$$\begin{aligned} (\ell+1)P_{\ell+1}(x)P_{\ell-1}(x) + \ell\, |P_{\ell-1}(x)|^2 &= (2\ell+1)xP_\ell(x)P_{\ell-1}(x), \\ \ell\, |P_\ell(x)|^2 + (\ell-1)P_{\ell-2}(x)P_\ell(x) &= (2\ell-1)xP_{\ell-1}(x)P_\ell(x). \end{aligned}$$

This implies that

$$\begin{aligned} &\frac{1}{2\ell+1}\left((\ell+1)P_{\ell+1}(x)P_{\ell-1}(x) + \ell\, |P_{\ell-1}(x)|^2\right) \\ = \; &\frac{1}{2\ell-1}\left(\ell\, |P_\ell(x)|^2 + (\ell-1)P_{\ell-2}(x)P_\ell(x)\right). \end{aligned}$$

Integrating this result from -1 to 1, the orthogonality relations in (i) imply that

$$\frac{1}{2\ell+1}\int_{-1}^{1} |P_{\ell-1}(x)|^2\, dx = \frac{1}{2\ell-1}\int_{-1}^{1} |P_\ell(x)|^2\, dx,$$

or,

$$\int_{-1}^{1} |P_\ell(x)|^2\, dx = \frac{2\ell-1}{2\ell+1}\int_{-1}^{1} |P_{\ell-1}(x)|^2\, dx. \tag{11.30}$$

One can now complete the proof by induction (Exercise 11.4). □

Note that, due to the fact that the Legendre polynomials are real-valued, we have omitted the complex conjugation on $P_{\ell'}(x)$ in (11.28). The result shows that an appropriate normalization of the Legendre polynomials forms an orthonormal set in $L^2(-1,1)$. In fact, these functions are even complete in $L^2(-1,1)$:

Theorem 11.2.7 (Orthonormal basis of Legendre polynomials) *The functions* $\left\{\sqrt{\frac{2\ell+1}{2}}P_\ell\right\}_{\ell=0}^{\infty}$ *form an orthonormal basis for* $L^2(-1,1)$. *In particular,*

$$f = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{2} \langle f, P_\ell \rangle P_\ell, \ \forall f \in L^2(-1,1), \tag{11.31}$$

with convergence in the norm of $L^2(-1,1)$.

Proof. It follows from Lemma 11.2.6 that the functions $\left\{\sqrt{\frac{2\ell+1}{2}}P_\ell\right\}_{\ell=0}^{\infty}$ form an orthonormal set in $L^2(-1,1)$; invoking Theorem 4.7.2, we just need to show that these functions are complete in $L^2(-1,1)$, i.e., that

$$\overline{\text{span}}\left\{\sqrt{\frac{2\ell+1}{2}}P_\ell\right\}_{\ell=0}^{\infty} = L^2(-1,1).$$

For that purpose, we notice that by Lemma 11.2.3, the vector space spanned by all the Legendre polynomials equals the vector space of all polynomials. According to Theorem 6.3.1, the set of polynomials is dense in $L^2(-1,1)$. Thus,

$$\begin{aligned}\overline{\text{span}}\left\{\sqrt{\frac{2\ell+1}{2}}P_\ell\right\}_{\ell=0}^{\infty} = \overline{\text{span}}\{P_\ell\}_{\ell=0}^{\infty} &= \overline{\text{span}}\{x^\ell\}_{\ell=0}^{\infty} \\ &= L^2(-1,1),\end{aligned}$$

as desired. In particular, each $f \in L^2(-1,1)$ has an expansion

$$\begin{aligned} f &= \sum_{\ell=0}^{\infty} \langle f, \sqrt{\frac{2\ell+1}{2}}P_\ell \rangle \sqrt{\frac{2\ell+1}{2}}P_\ell \\ &= \sum_{\ell=0}^{\infty} \frac{2\ell+1}{2} \langle f, P_\ell \rangle P_\ell, \end{aligned}$$

with convergence in the norm of $L^2(-1,1)$. This concludes the proof. □

Note that the general result in Theorem 11.1.10 shows that for piecewise differentiable functions f, the *pointwise* expansion

$$f(x) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{2} \langle f, P_\ell \rangle P_\ell(x)$$

holds at all points $x \in [-1,1]$ where f is continuous.

The Legendre equation in (11.24) is actually a special case of a class of differential equations, called *associated Legendre equations.* These equations contain an extra parameter $n \in \{0,1,\ldots,\ell\}$ compared to the Legendre equation:

$$(1-x^2)\frac{d^2u}{dx^2} - 2x\frac{du}{dx} + \left(\ell(\ell+1) - \frac{n^2}{1-x^2}\right)u = 0,\ x \in]-1,1[. \quad (11.32)$$

The theory for the associated Legendre equation is very similar to the theory for the Legendre equation, and the proofs are parallel (although slightly more involved for the associated Legendre equation, due to the extra term). We will state a few central results without proofs:

- For given parameters ℓ and $n \in \{0,1,\ldots,\ell\}$, the equation (11.32) has the solution

$$P_{\ell,n}(x) = (1-x^2)^{n/2}\frac{d^nP_\ell}{dx^n}(x), \quad (11.33)$$

  where P_ℓ is the ℓth Legendre polynomial.

- For any given $n \in \mathbb{N}_0$, the functions $\{\sqrt{\frac{(2\ell+1)(\ell-n)!}{2(\ell+n)!}}P_{\ell,n}\}_{\ell=n}^{\infty}$ form an orthonormal basis for $L^2(-1,1)$.

The functions $P_{\ell,n}$ in (11.33) are called *associated Legendre functions.* Note that $P_{\ell,n}$ is a polynomial if n is even; if n is odd, $P_{\ell,n}$ is a polynomial multiplied with $\sqrt{1-x^2}$. 13 contains a list of the first few associated Legendre functions.

11.3 Laguerre polynomials

For a given parameter $\ell \in \mathbb{N}_0$, the *Laguerre equation* is

$$x\frac{d^2u}{dx^2} + (1-x)\frac{du}{dx} + \ell u = 0,\ x \in]0,\infty[. \quad (11.34)$$

Note that a function u is solution to the Laguerre equation if and only if

$$xe^{-x}\frac{d^2u}{dx^2} + (1-x)e^{-x}\frac{du}{dx} + \ell e^{-x}u = 0,\ x \in]0,\infty[,$$

or, if and only if

$$\left(xe^{-x}u'\right)' + \ell e^{-x}u = 0,\ x \in]0,\infty[. \quad (11.35)$$

The equation (11.35) is on Sturm–Liouville form with

$$p(x) = xe^{-x},\ q(x) = 0,\ r(x) = e^{-x},\ \lambda = \ell.$$

Thus, we can also consider the solutions to the Laguerre equation as solutions to a Sturm–Liouville problem and apply the general theory. Note, however, that because the interval $]0,\infty[$ is infinite, we are not dealing with a regular Sturm–Liouville problem.

Using the method of power series we can derive the following expression for a solution to the Laguerre equation (Exercise 11.6):

Theorem 11.3.1 (Laguerre polynomials) *Given $\ell \in \mathbb{N}_0$, the equation (11.34) has the solution*

$$Q_\ell(x) = \ell! \sum_{k=0}^{\ell} \frac{(-1)^k}{(k!)^2(\ell-k)!} x^k. \tag{11.36}$$

The function Q_ℓ in (11.36) is called the *ℓth Laguerre polynomial.* 13 contains a list of the first few Laguerre polynomials. Analogous to the proof of Rodrigues' formula for Legendre polynomials, one can prove that Q_ℓ also can be written on the form (Exercise 11.5)

$$Q_\ell(x) = \frac{e^x}{\ell!} \frac{d^\ell}{dx^\ell}(x^\ell e^{-x}). \tag{11.37}$$

The Sturm–Liouville problem (11.35) is not regular, so we cannot immediately use Theorem 11.1.7 to conclude that the Laguerre polynomials Q_ℓ are orthogonal on the positive axis with respect to the weight function $r(x) = e^{-x}$. However, via a limit-argument and a closer look at the proof of Theorem 11.1.7 one can show that this actually holds. We state the result formally, and ask the reader to provide a direct proof in Exercise 11.6:

Lemma 11.3.2 (Orthogonality of Laguerre polynomials)
The Laguerre polynomials $\{Q_\ell\}_{\ell=0}^\infty$ satisfy the following relations:

(i) For $\ell \neq \ell'$,

$$\int_0^\infty Q_\ell(x) Q_{\ell'}(x)\, e^{-x}\, dx = 0.$$

(ii) For any $\ell \in \mathbb{N}_0$,

$$\int_0^\infty |Q_\ell(x)|^2 e^{-x}\, dx = 1.$$

Lemma 11.3.2 implies that the functions $\{Q_\ell(x)e^{-x/2}\}_{\ell=0}^\infty$ form an orthonormal set in $L^2(0,\infty)$. These functions are called *Laguerre functions.* With more advanced tools at hand than developed here, one can show that they even form an orthonormal basis for $L^2(0,\infty)$:

Theorem 11.3.3 (Orthonormal basis of Laguerre polynomials) *The Laguerre functions $\{Q_\ell(x)\, e^{-x/2}\}_{\ell=0}^\infty$ form an orthonormal basis for $L^2(0,\infty)$.*

A proof of Theorem 11.3.3 can be found in [3].

11.4 Hermite polynomials

The *Hermite differential equation* is given by

$$\frac{d^2w}{dx^2} - 2x\frac{dw}{dx} + (\lambda - 1)w = 0, \tag{11.38}$$

where $\lambda \in \mathbb{R}$. We will mainly consider the Hermite differential equation in the special case $\lambda = 2\ell+1,\ \ell \in \mathbb{N}_0$; in that case, an appropriate substitution transfers the Hermite differential equation into a Sturm–Liouville problem (Exercise 11.7).

In Exercise 11.8, we consider solutions to (11.38) of the form

$$w(x) = \sum_{k=0}^{\infty} c_k x^k.$$

In particular, the exercise shows the existence of polynomial solutions in the special case $\lambda = 2\ell + 1,\ \ell \in \mathbb{N}_0$:

Theorem 11.4.1 (Hermite polynomials) *Let $\lambda = 2\ell + 1$ for some $\ell \in \mathbb{N}_0$ and put $m = \ell/2$ if ℓ is even, and $m = (\ell - 1)/2$ if ℓ is odd. Then* (11.38) *has the solution*

$$H_\ell(x) = \ell! \sum_{k=0}^{m} \frac{(-1)^k}{k!(\ell - 2k)!} 2^{\ell-2k} x^{\ell-2k}. \tag{11.39}$$

The polynomials H_ℓ are called *Hermite polynomials*; 13 contains a list of the first ones. We mention some of their important properties, and ask the reader to provide the proof in Exercises 11.9 and 11.10.

Theorem 11.4.2 (Orthogonality of Hermite polynomials) *The Hermite polynomials H_ℓ have the following properties:*

(i) For $\ell \in \mathbb{N}_0$, H_ℓ is given by

$$H_\ell(x) = (-1)^\ell e^{x^2} \frac{d^\ell}{dx^\ell} e^{-x^2}.$$

(ii) Letting $r(x) := e^{-x^2}$, the polynomials $\{H_\ell\}_{\ell=0}^\infty$ are orthogonal in $L_r^2(\mathbb{R})$, i.e.,

$$\int_{-\infty}^{\infty} H_\ell(x) H_{\ell'}(x)\, e^{-x^2}\, dx = 0 \text{ if } \ell \neq \ell'. \tag{11.40}$$

(iii) For $\ell \in \mathbb{N}_0$,

$$\int_{-\infty}^{\infty} |H_\ell(x)|^2 e^{-x^2}\, dx = \sqrt{\pi} 2^\ell \ell!$$

Note that because the Hermite polynomials are real, we omit a complex conjugation on $H_{\ell'}(x)$ in (11.40).

As in the similar case for the Laguerre polynomials, Theorem 11.4.2 implies that the functions $\left\{\pi^{-1/4}\left(2^\ell \ell!\right)^{-1/2} H_\ell(x)e^{-x^2/2}\right\}_{\ell=0}^{\infty}$ form an orthonormal system in $L^2(\mathbb{R})$. These functions are called *Hermite functions.* We refer to [3] for a proof of the fact that the Hermite functions are complete in $L^2(\mathbb{R})$:

Theorem 11.4.3 (Hermite functions) *The functions*

$$\left\{\pi^{-1/4}\left(2^\ell \ell!\right)^{-1/2} H_\ell(x)e^{-x^2/2}\right\}_{\ell=0}^{\infty}$$

form an orthonormal basis for $L^2(\mathbb{R})$.

The Hermite polynomials appear naturally in the context of quantum mechanics:

Example 11.4.4 (Quantum-mechanical harmonic oscillator) In appropriate units, the wave function $u(x,t)$ describing the quantum-mechanical harmonic oscillator in one dimension is determined by the equation

$$-i\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - x^2 u. \tag{11.41}$$

The equation (11.41) is a special case of the *time-dependent one-dimensional Schrödinger wave equation*, corresponding to the case where the potential is proportional to x^2. We will search for a solution of the form

$$u(x,t) := T(t)v(x)$$

for some functions T and v. Hereby, the equation (11.41) turns into

$$-iT'(t)v(x) = T(t)v''(x) - x^2 T(t)v(x),$$

or, whenever $T(t) \neq 0$ and $v(x) \neq 0$,

$$-i\frac{T'(t)}{T(t)} = \frac{v''(x) - x^2 v(x)}{v(x)}. \tag{11.42}$$

Thus, the expression (11.42) is forced to be constant, independently of the choice of x and t. Denoting this constant by $-\lambda$, the functions T and v will satisfy the equations

$$\begin{cases} -iT'(t) = -\lambda T(t), \\ v''(x) - x^2 v(x) = -\lambda v(x), \end{cases}$$

or,

$$T'(t) - i\lambda T(t) = 0, \tag{11.43}$$

$$v''(x) + (\lambda - x^2)v(x) = 0. \tag{11.44}$$

The equation (11.43) has the solutions

$$T(t) = Ce^{i\lambda t}, \tag{11.45}$$

where $C \in \mathbb{C}$.

The equation (11.44) is more complicated. We will first perform a substitution that turns the equation into the Hermite equation. We note that an arbitrary function v can be written on the form

$$v(x) := w(x)e^{-x^2/2}, \tag{11.46}$$

for an appropriate function w; now, a direct verification (Exercise 11.8) shows that v solves (11.44) if and only if the function w solves the Hermite equation (11.38).

Let us now again consider the case where $\lambda = 2\ell + 1$ for some $\ell \in \mathbb{N}_0$. Theorem 11.4.1 shows that the Hermite equation has the solution H_ℓ, so we conclude that the equation (11.44) has a solution

$$v_\ell(x) = H_\ell(x)e^{-x^2/2}. \tag{11.47}$$

Up to a normalization factor, the functions v_ℓ in (11.47) equal the Hermite functions. In the context of quantum mechanics they are known under the name *harmonic oscillator wave functions.*

Finally, using (11.45) with $C = 1$, we see that equation (11.41) has the solutions

$$u_\ell(x,t) = T(t)v_\ell(x) = e^{i(2\ell+1)t}H_\ell(x)e^{-x^2/2}, \ \ell \in \mathbb{N}_0. \qquad \square$$

11.5 Exercises

11.1 (Vibrating string) We consider Example 11.1.6.

(i) Show that with the choice of u in (11.8), (11.7) takes the form

$$f(x)g''(t) = f''(x)g(t).$$

(ii) Show that for some constant $\lambda \in \mathbb{R}$,

$$\frac{f''(x)}{f(x)} = \frac{g''(t)}{g(t)} = -\lambda, \tag{11.48}$$

which is equivalent with the two equations (11.9) and (11.10).

The exercise continues!

(iii) Assume that $\lambda < 0$. Show that the real solutions of (11.9) are

$$f(x) = c_1 e^{\sqrt{\lambda}\,x} + c_2 e^{-\sqrt{\lambda}\,x}, \tag{11.49}$$

where $c_1, c_2 \in \mathbb{R}$.

(iv) Show that the only function of the form in (11.49) that satisfies the boundary conditions $f(0) = f(L) = 0$ is $f = 0$.

(v) Assume that $\lambda = 0$. Show that the only solution of (11.9) satisfying the boundary conditions $f(0) = f(L) = 0$ is $f = 0$.

(vi) Assume that $\lambda > 0$, and write $\lambda = k^2$ for some $k > 0$. Show that the real solutions of (11.9) are

$$f(x) = c_1 \cos(k\,x) + c_2 \sin(k\,x), \tag{11.50}$$

where $c_1, c_2 \in \mathbb{R}$.

(vii) Show that the only functions of the form in (11.50) that satisfies the boundary conditions $f(0) = 0$ are the functions

$$f(x) = c_2 \sin(k\,x), \tag{11.51}$$

where $c_2 \in \mathbb{R}$.

(viii) Consider the functions of the form (11.51) with $k > 0$. Show that there exists a nonzero function of that form satisfying $f(L) = 0$ if and only if $k = \frac{n\pi}{L}$ for some $n \in \mathbb{N}$.

(ix) Conclude that (11.9) has a nonzero solution f with $f(0) = f(L) = 0$ if and only if $\lambda = \frac{n^2\pi^2}{L^2}$ for some $n \in \mathbb{N}$; and that the corresponding solutions are

$$f(x) = c\sin(\frac{\pi n}{L}\,x).$$

11.2 Prove Corollary 11.1.8.

11.3 (Legendre polynomials) Show that the first few Legendre polynomials are given by

$$\begin{aligned}
P_0(x) &= 1,\\
P_1(x) &= x,\\
P_2(x) &= \frac{1}{2}(3x^2 - 1),\\
P_3(x) &= \frac{1}{2}(5x^3 - 3x).
\end{aligned}$$

11.4 **(Legendre polynomials)** Using that the first Legendre polynomial is $P_0(x) = 1$, complete the proof of Lemma 11.2.6(ii) based on (11.30).

11.5 (**Laguerre polynomials**) Prove that the Laguerre polynomials are given by (11.37).

11.6 (**Laguerre polynomials**) Denote the Laguerre polynomials by $Q_\ell,\ \ell \in \mathbb{N}_0$.

(i) Derive the expression (11.36), e.g., as follows. First, show that the coefficients in any power series solution $u(x) = \sum_{k=0}^{\infty} c_k x^k$ of (11.34) will satisfy that

$$c_{k+1} = \frac{k-\ell}{(k+1)^2}\, c_k, \ \ k \in \mathbb{N}_0.$$

Conclude that $c_{\ell+1} = c_{\ell+2} = \cdots = 0$; choose $c_0 = 1$, and show by induction that

$$c_k = \frac{\ell!\,(-1)^k}{(k!)^2\,(\ell-k)!}, \ \ k = 0, 1, \ldots, \ell.$$

(ii) Show that

$$\begin{aligned} Q_1(x) = 1 - x, \ Q_2(x) &= 1 - 2x + \frac{x^2}{2}, \\ Q_3(x) &= 1 - 3x + 3\frac{x^2}{2} - \frac{x^3}{6}. \end{aligned}$$

(iii) Prove Lemma 11.3.2. *Hint:* for $\ell' < \ell, \ \ Q_{\ell'}(x)$ is a linear combination of polynomials of degree smaller than ℓ. Show via k integrations by parts that for $k < \ell$,

$$\int_0^\infty e^{-x} x^k Q_\ell(x)\, dx = 0.$$

11.7 (**Hermite's differential equation**) Show that a function w is a solution to Hermite's equation (11.55) if and only if w is a solution to the Sturm–Liouville equation

$$\left(e^{-x^2} w'\right)' + 2\ell e^{-x^2} w = 0.$$

11.8 (**Hermite polynomials**) Consider Hermite's differential equation,

$$\frac{d^2w}{dx^2} - 2x\frac{dw}{dx} + (\lambda - 1)w = 0. \tag{11.52}$$

We search for solutions of the form

$$w(x) = \sum_{k=0}^{\infty} c_k x^k. \tag{11.53}$$

(i) Show that the function w in (11.53) is a solution to (11.52) if and only if

$$c_{k+2}(k+2)(k+1) = (2k+1-\lambda)c_k, \ \forall\, k \in \mathbb{N}_0. \tag{11.54}$$

We will now consider the case where $\lambda = 2\ell + 1$ for some $\ell \in \mathbb{N}_0$, i.e., the equation

$$\frac{d^2w}{dx^2} - 2x\frac{dw}{dx} + 2\ell w = 0. \tag{11.55}$$

(ii) Show that when $\ell \in \mathbb{N}_0$ is even, the equation (11.55) has an even polynomial solution; and that if $\ell \in \mathbb{N}_0$ is odd, the equation (11.55) has an odd polynomial solution. (*Hint:* in case ℓ is even, consider (11.54) with the initial condition $c_1 = 0$.)

(iii) Put $m = \ell/2$ if ℓ is even, and $m = (\ell-1)/2$ if ℓ is odd. Follow the approach from the proof of Theorem 11.2.2 and show that (11.55) has the solution

$$H_\ell(x) = \ell! \sum_{k=0}^{m} \frac{(-1)^k}{k!(\ell-2k)!} 2^{\ell-2k} x^{\ell-2k}. \tag{11.56}$$

(iv) Show that

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= 2x, \\ H_2(x) &= 4x^2 - 2, \\ H_3(x) &= 8x^3 - 12x. \end{aligned}$$

11.9 (**Hermite polynomials**) Prove Theorem 11.4.2(i). You can, e.g., do the following. First, check that the result holds for $\ell = 0$. Now, assume that the result in Theorem 11.4.2(i) holds for some $\ell \in \mathbb{N}_0$. Then

$$\frac{d^\ell}{dx^\ell} e^{-x^2} = (-1)^\ell H_\ell(x) e^{-x^2},$$

which implies that

$$\frac{d^{\ell+1}}{dx^{\ell+1}}e^{-x^2} = (-1)^\ell \left(H'_\ell(x) - 2xH_\ell(x)\right) e^{-x^2}.$$

Use this expression and (11.39) to calculate

$$(-1)^{\ell+1}e^{x^2}\frac{d^{\ell+1}}{dx^{\ell+1}}e^{-x^2},$$

and check that the derived result corresponds to the expression for $H_{\ell+1}(x)$ obtained via (11.39).

11.10 (**Hermite polynomials**) Prove Theorem 11.4.2(ii) and (iii), e.g., as follows. Assume that $\ell' \leq \ell$, and let $\phi(x) = e^{-x^2}$. Via the formula in Theorem 11.4.2(i) and partial integration, check that

$$\begin{aligned}
I(\ell, \ell') &:= \int_{-\infty}^{\infty} H_\ell(x)H_{\ell'}(x)e^{-x^2}\, dx \\
&= \int_{-\infty}^{\infty} (-1)^\ell H_{\ell'}(x)\frac{d^\ell \phi}{dx^\ell}(x)\, dx \\
&= (-1)^{\ell+1}\int_{-\infty}^{\infty} H'_{\ell'}(x)\frac{d^{\ell-1}\phi}{dx^{\ell-1}}(x)\, dx,
\end{aligned}$$

and more generally, for $k \leq \ell$,

$$I(\ell, \ell') = (-1)^{\ell+k}\int_{-\infty}^{\infty} H_{\ell'}^{(k)}(x)\frac{d^{\ell-k}\phi}{dx^{\ell-k}}(x)\, dx. \tag{11.57}$$

Now, for $\ell' < \ell$, use (11.57) with $k = \ell'$ and the fact that $H_{\ell'}^{(\ell')}(x)$ is constant to show that $I(\ell, \ell') = 0$.

For $\ell = \ell'$, use (11.57) with $k = \ell$; show that $H_\ell^{(\ell)}(x) = 2^\ell \ell!$, and calculate $I(\ell, \ell)$ using that

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}.$$

11.11 (**Hermite polynomials**) Assume that a function v of the form $v(x) = w(x)e^{-x^2/2}$ solves the differential equation (11.44). Show that w satisfies (11.38).

11.12 (**Chebyshev polynomials of the first kind**) For $\ell \in \mathbb{N}_0$, let

$$T_\ell(x) := \cos(\ell \arccos(x)),\ \ x \in [-1,1].$$

(i) Show that

$$T_0(x) = 1, T_1(x) = x,\ \ T_2(x) = 2x^2 - 1.$$

(ii) Show that for $\ell = 2, 3, \ldots,$

$$T_\ell(x) = 2xT_{\ell-1}(x) - T_{\ell-2}(x).$$

Hint: use that

$$2\cos(y)\ \cos(z) = \cos(y+z) + \cos(y-z) \tag{11.58}$$

with $y = (\ell - 1)\arccos(x),\ \ z = \arccos(x)$.

(iii) Show that for any $\ell \in \mathbb{N}_0$ the function T_ℓ is a polynomial of degree ℓ.

(iv) Show that for $\ell, \ell' \in \mathbb{N}_0$,

$$\int_{-1}^{1} T_\ell(x)T_{\ell'}(x)\frac{1}{\sqrt{1-x^2}}\,dx = \begin{cases} 0 & \text{if } \ell \neq \ell', \\ \pi & \text{if } \ell = \ell' = 0, \\ \frac{\pi}{2} & \text{if } \ell = \ell' \neq 0. \end{cases}$$

Hint: use again (11.58), followed by the change of variable

$$u = \arccos(x), \quad du = \frac{1}{\sqrt{1-x^2}}\,dx.$$

(v) Show that for $\ell \in \mathbb{N}_0$, the polynomial T_ℓ satisfies the equation

$$(1-x^2)\frac{d^2u}{dx^2} - x\frac{du}{dx} + \ell^2 u = 0.$$

11.13 (**Chebyshev polynomials of the second kind**) For $\ell \in \mathbb{N}_0$, let

$$U_\ell(x) := \frac{\sin((\ell+1)\arccos(x))}{\sqrt{1-x^2}}.$$

Show that

$$U_0(x) = 1,\ U_1(x) = 2x,\ U_2(x) = 4x^2 - 1,\ U_3(x) = 8x^3 - 4x.$$

Appendix A

A.1 Proof of Weierstrass' theorem, Theorem 2.3.4

For the sake of simplicity we consider the interval $[0, 1]$; the general case can be obtained from this case via a scaling. We define a sequence of polynomials $Q_N, N \in \mathbb{N}$, by

$$Q_N(x) = \sum_{n=0}^{N} f(\frac{n}{N}) \begin{pmatrix} N \\ n \end{pmatrix} x^n (1-x)^{N-n}, \tag{A.1}$$

where

$$\begin{pmatrix} N \\ n \end{pmatrix} = \frac{N!}{n!(N-n)!}.$$

In order to estimate the quantity $|f(x) - Q_N(x)|$ we first rewrite $f(x)$ with help of the binomial formula in Lemma 1.9.2. The trick is to multiply $f(x)$ by the number 1, which we can write as

$$\begin{aligned} 1 &= (x + (1-x))^N \\ &= \sum_{n=0}^{N} \begin{pmatrix} N \\ n \end{pmatrix} x^n (1-x)^{N-n}. \end{aligned}$$

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis,
DOI 10.1007/978-0-8176-4980-7_12,

Hereby we get

$$\begin{aligned} f(x) &= f(x)\ (x+(1-x))^N \\ &= f(x)\ \sum_{n=0}^{N} \binom{N}{n} x^n(1-x)^{N-n} \\ &= \sum_{n=0}^{N} f(x) \binom{N}{n} x^n(1-x)^{N-n}. \end{aligned}$$

Thus,

$$\begin{aligned} & f(x) - Q_N(x) \\ = & \sum_{n=0}^{N} f(x) \binom{N}{n} x^n(1-x)^{N-n} - \sum_{n=0}^{N} f(\frac{n}{N}) \binom{N}{n} x^n(1-x)^{N-n} \\ = & \sum_{n=0}^{N} (f(x) - f(\frac{n}{N})) \binom{N}{n} x^n(1-x)^{N-n}. \qquad \text{(A.2)} \end{aligned}$$

For a given $\epsilon > 0$, we will show that the degree of polynomial Q_N can be chosen such that $|f(x) - Q_N(x)| \leq \epsilon$ for all $x \in [0,1]$. We will use the fact that the function f is uniformly continuous on the interval $[0,1]$, see Theorem 1.6.2; by (1.18) this means that we can find a $\delta > 0$ such that

$$x, y \in [0,1], |x-y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\epsilon}{2}. \qquad \text{(A.3)}$$

Fix $x \in [0,1]$. We first divide the index set for the sum (A.2) into two sets:

$$\begin{aligned} f(x) - Q_N(x) &= \sum_{n=0}^{N} (f(x) - f(\frac{n}{N})) \binom{N}{n} x^n(1-x)^{N-n} \qquad \text{(A.4)} \\ &= \sum_{|x-\frac{n}{N}|<\delta} (f(x) - f(\frac{n}{N})) \binom{N}{n} x^n(1-x)^{N-n} \qquad \text{(A.5)} \\ &\quad + \sum_{|x-\frac{n}{N}|\geq\delta} (f(x) - f(\frac{n}{N})) \binom{N}{n} x^n(1-x)^{N-n}. \qquad \text{(A.6)} \end{aligned}$$

In the following we investigate (A.5) and (A.6) separately. We begin by considering (A.5). By the triangle inequality

$$\begin{aligned} & \left| \sum_{|x-\frac{n}{N}|<\delta} (f(x) - f(\frac{n}{N})) \binom{N}{n} x^n(1-x)^{N-n} \right| \\ \leq & \sum_{|x-\frac{n}{N}|<\delta} \left| f(x) - f(\frac{n}{N}) \right| \binom{N}{n} x^n(1-x)^{N-n}. \end{aligned}$$

It follows by our choice of δ in (A.3) that

$$
\begin{aligned}
&\sum_{|x-\frac{n}{N}|<\delta} \left| f(x) - f(\frac{n}{N}) \right| \binom{N}{n} x^n (1-x)^{N-n} \\
\leq\ & \frac{\epsilon}{2} \sum_{|x-\frac{n}{N}|<\delta} \binom{N}{n} x^n (1-x)^{N-n} \\
\leq\ & \frac{\epsilon}{2} \sum_{n=0}^{N} \binom{N}{n} x^n (1-x)^{N-n} \\
=\ & \frac{\epsilon}{2} (x + (1-x))^N \\
=\ & \frac{\epsilon}{2}.
\end{aligned}
$$

Therefore, we obtain the following estimate for (A.5):

$$
\left| \sum_{|x-\frac{n}{N}|<\delta} (f(x) - f(\frac{n}{N})) \binom{N}{n} x^n (1-x)^{N-n} \right| \leq \frac{\epsilon}{2}. \tag{A.7}
$$

Next, we consider (A.6). First we notice that

$$
\begin{aligned}
&\left| \sum_{|x-\frac{n}{N}|\geq\delta} (f(x) - f(\frac{n}{N})) \binom{N}{n} x^n (1-x)^{N-n} \right| \\
\leq\ & \sum_{\frac{(x-\frac{n}{N})^2}{\delta^2}\geq 1} \left| f(x) - f(\frac{n}{N}) \right| \binom{N}{n} x^n (1-x)^{N-n}.
\end{aligned}
$$

Let $M = \max_{x\in[0,1]} |f(x)|$. By the triangle inequality,

$$
|f(x) - f(\frac{n}{N})| \leq |f(x)| + |f(\frac{n}{N})| \leq 2M. \tag{A.8}
$$

Using (A.8) it follows that

$$
\begin{aligned}
&\sum_{\frac{(x-\frac{n}{N})^2}{\delta^2}\geq 1} \left| f(x) - f(\frac{n}{N}) \right| \binom{N}{n} x^n (1-x)^{N-n} \\
\leq\ & 2M \sum_{\frac{(x-\frac{n}{N})^2}{\delta^2}\geq 1} \binom{N}{n} x^n (1-x)^{N-n} \\
\leq\ & 2M \sum_{\frac{(x-\frac{n}{N})^2}{\delta^2}\geq 1} \frac{(x-\frac{n}{N})^2}{\delta^2} \binom{N}{n} x^n (1-x)^{N-n} = (*);
\end{aligned}
$$

in the last inequality we used that $\frac{(x-\frac{n}{N})^2}{\delta^2} \geq 1$ for all the values of n which appear in the summation. Furthermore, the sum increases if we sum over all $n = 0, 1, \ldots, N$ rather than just the values of n for which $\frac{(x-\frac{n}{N})^2}{\delta^2} \geq 1$; thus,

$$(*) \quad \leq \quad 2M \sum_{n=0}^{N} \frac{(x-\frac{n}{N})^2}{\delta^2} \begin{pmatrix} N \\ n \end{pmatrix} x^n (1-x)^{N-n}$$

$$\leq \quad \frac{2M}{\delta^2} \sum_{n=0}^{N} \left(x - \frac{n}{N}\right)^2 \begin{pmatrix} N \\ n \end{pmatrix} x^n (1-x)^{N-n} = (**).$$

Note that $\sum_{n=0}^{N} (n - Nx)^2 \begin{pmatrix} N \\ n \end{pmatrix} x^n (1-x)^{N-n}$ is the variance for the binomial distribution with parameter N and probability parameter x; thus, it equals $Nx(1-x)$. Thus,

$$\sum_{n=0}^{N} \left(x - \frac{n}{N}\right)^2 \begin{pmatrix} N \\ n \end{pmatrix} x^n (1-x)^{N-n} = \frac{1}{N} x(1-x),$$

and

$$(**) \quad = \quad \frac{2M}{N\delta^2} x(1-x).$$

Since we only consider $x \in [0, 1]$,

$$\frac{2M}{N\delta^2} x(1-x) \leq \frac{M}{2N\delta^2}.$$

Altogether we now have that the term in (A.6) can be estimated by

$$\left| \sum_{|x-\frac{n}{N}| \geq \delta} f(x) - f(\frac{n}{N}) \begin{pmatrix} N \\ n \end{pmatrix} (1-x)^{N-n} \right| \leq \frac{M}{2N\delta^2}. \tag{A.9}$$

Let us now return to the expression for $f(x) - Q_N(x)$ in (A.4). According to our estimates for the terms in (A.5) and (A.6) in (A.7) and (A.9), we have that

$$|f(x) - Q_N(x)| \quad \leq \quad \frac{\epsilon}{2} + \frac{M}{2N\delta^2}.$$

This holds for all values of x. Now choose $N \in \mathbb{N}$ such that $\frac{M}{2N\delta^2} < \frac{\epsilon}{2}$, i.e., $N > \frac{M}{\delta^2 \epsilon}$. Then

$$|f(x) - Q_N(x)| < \epsilon \text{ for all } x \in [0, 1]. \qquad \square$$

A.2 Proof of Theorem 7.1.7

The purpose of this section is to outline a proof of the inversion theorem for the Fourier transform. As usual we will have to skip the measure-theoretic parts, but the proof is complete for continuous functions. Our presentation is inspired by the book by Katznelson [14]. The first step is to introduce the so-called *approximate identities.*

Definition A.2.1 (Approximate identity) *A collection of functions* $k_\lambda \in L^1(\mathbb{R})$, *indexed by* $\lambda > 0$, *is an approximate identity if the following conditions hold:*

(i) For all $\lambda > 0$, $\int_{-\infty}^{\infty} k_\lambda(x)\,dx = 1$.

(ii) There exists a constant $C > 0$ *such that*

$$\int_{-\infty}^{\infty} |k_\lambda(x)|\,dx \le C, \ \forall\, \lambda > 0.$$

(iii) For every $\delta \in]0, \infty[$,

$$\lim_{\lambda \to 0} \int_{|x|>\delta} |k_\lambda(x)|\,dx = 0.$$

There is a standard procedure to construct an approximate identity, based on a single function:

Lemma A.2.2 (Construction of approximate identity) *Let* $k \in L^1(\mathbb{R})$ *and assume that* $\int_{-\infty}^{\infty} k(x)\,dx = 1$. *Then the functions* k_λ *defined by*

$$k_\lambda(x) := \frac{1}{\lambda} k(\frac{x}{\lambda}) \tag{A.10}$$

form an approximate identity .

We leave the proof of Lemma A.2.2 to the reader (Exercise 7.14). We will consider the following choice of an approximate identity, known in the literature under the name *Fejér kernel:*

Example A.2.3 (Fejér kernel) Define the function k by

$$k(x) := \int_{-(2\pi)^{-1}}^{(2\pi)^{-1}} (1 - 2\pi\,|y|)\, e^{2\pi i x y}\, dy, \ x \in \mathbb{R}. \tag{A.11}$$

Via the change of variable $z = 2\pi y$, we see that

$$k(x) = \frac{1}{2\pi} \int_{-1}^{1} (1 - |z|)\, e^{ixz}\, dz = \frac{1}{2\pi} \int_{-\infty}^{\infty} (1 - |z|)\, \chi_{[-1,1]}(z)\, e^{ixz}\, dz. \tag{A.12}$$

There are many ways to show that $k \in L^1(\mathbb{R})$ and that $||k||_1 = 1$. Instead of direct calculations we will use some of the results about B-splines that are derived in Chapter 10.

First, (A.12) and Corollary 10.2.2 show that in terms of the B-spline B_2,

$$k(x) = \frac{1}{2\pi}\widehat{B_2}(-\frac{x}{2\pi}) \tag{A.13}$$

$$= \frac{1}{2\pi}\left(\frac{\sin\frac{x}{2}}{\frac{x}{2}}\right)^2. \tag{A.14}$$

The expression (A.14) immediately shows that $k \in L^1(\mathbb{R})$. Furthermore, using (A.13) and the change of variable $y = \frac{-x}{2\pi}$,

$$\int_{-\infty}^{\infty} k(x)\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty}\widehat{B_2}(-\frac{x}{2\pi})\,dx$$

$$= \int_{-\infty}^{\infty}\widehat{B_2}(y)\,dy. \tag{A.15}$$

By definition of B_2 and Theorem 7.3.4,

$$\widehat{B_2}(y) = \widehat{B_1 * B_1}(y) = (\widehat{B_1}(y))^2.$$

Using (A.15) and Theorem 7.2.2, this implies that

$$\int_{-\infty}^{\infty} k(x)\,dx = \int_{-\infty}^{\infty}|\widehat{B_1}(y)|^2\,dy = \int_{-\infty}^{\infty}|B_1(y)|^2\,dy = 1.$$

By Lemma A.2.2 we conclude that the functions

$$k_\lambda(x) = \frac{1}{\lambda}k(\frac{x}{\lambda}) = \frac{1}{\lambda}\int_{-(2\pi)^{-1}}^{(2\pi)^{-1}}(1-2\pi\,|y|)\,e^{2\pi ixy/\lambda}\,dy,\ \ \lambda > 0, \tag{A.16}$$

form an approximate identity. Note that by the change of variable $z = y/\lambda$,

$$k_\lambda(x) = \int_{-\infty}^{\infty}(1-2\pi\lambda\,|z|)\,\chi_{[-(2\pi\lambda)^{-1},(2\pi\lambda)^{-1}]}(z)\,e^{2\pi ixz}\,dz,\ \ \lambda > 0. \tag{A.17}$$

□

As said, the full proof of the inversion theorem requires measure theory, so we will not give all the details. But (ii) and (iii) in Lemma A.2.4 below give a complete proof for continuous functions $f \in L^1(\mathbb{R})$. The full statement in Theorem 7.1.7 is a consequence of the (unproven) result in Lemma A.2.4(iv) together with (ii).

Lemma A.2.4 *Assume that $f, \hat{f} \in L^1(\mathbb{R})$ and that $k_\lambda,\ \lambda > 0$, form an approximate identity. Then the following hold:*

(i) $(k_\lambda * f)(x) = \int_{-\infty}^{\infty}(1-2\pi\lambda\,|z|)\,\chi_{[-(2\pi\lambda)^{-1},(2\pi\lambda)^{-1}]}(z)\,\hat{f}(z)e^{2\pi ixz}\,dz.$

(ii) For all $x \in \mathbb{R}$,

$$(k_\lambda * f)(x) \to \int_{-\infty}^{\infty} \hat{f}(z) e^{2\pi i x z}\, dz \ \text{as}\ \lambda \to 0. \tag{A.18}$$

(iii) If f *is bounded and continuous at* x*, then*

$$(k_\lambda * f)(x) \to f(x) \ \text{as}\ \lambda \to 0.$$

(iv) As $\lambda \to 0$,

$$k_\lambda * f \to f \tag{A.19}$$

in $L^1(\mathbb{R})$.

Proof. Using (A.17) and the definition of convolution,

$$(k_\lambda * f)(x) = \int_{-\infty}^{\infty} k_\lambda(x-y) f(y)\, dy$$
$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (1 - 2\pi\lambda\,|z|)\, \chi_{[-(2\pi\lambda)^{-1},(2\pi\lambda)^{-1}]}(z)\, e^{2\pi i(x-y)z}\, dz \right) f(y)\, dy.$$

In order to continue the calculation, we would like to switch the order of integration. By Fubini's theorem, we know that this is possible if the integral

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \left| (1 - 2\pi\lambda\,|z|)\, \chi_{[-(2\pi\lambda)^{-1},(2\pi\lambda)^{-1}]}(z)\, e^{2\pi i(x-y)z} f(y)\, dy \right| \right) dz \tag{A.20}$$

is finite. We leave it to the reader as Exercise 7.15 to show that this actually is the case. Thus, by Theorem 5.3.10,

$$(k_\lambda * f)(x)$$
$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(y) e^{-2\pi i y z}\, dy \right) (1 - 2\pi\lambda\,|z|)\, \chi_{[-(2\pi\lambda)^{-1},(2\pi\lambda)^{-1}]}(z)\, e^{2\pi i x z}\, dz$$
$$= \int_{-\infty}^{\infty} (1 - 2\pi\lambda\,|z|)\, \chi_{[-(2\pi\lambda)^{-1},(2\pi\lambda)^{-1}]}(z)\, e^{2\pi i x z} \hat{f}(z)\, dz.$$

This proves (i). The proof of (ii) follows from Lebesgue's theorem on dominated convergence (Exercise 7.15).

For the proof of (iii), we use the property (i) in Definition A.2.1 and that $k_\lambda * f = f * k_\lambda$ to see that

$$\begin{aligned} |f(x) - (k_\lambda * f)(x)| &= \left| f(x) \int_{-\infty}^{\infty} k_\lambda(y)\, dy - \int_{-\infty}^{\infty} k_\lambda(y) f(x-y)\, dy \right| \\ &= \left| \int_{-\infty}^{\infty} (f(x) - f(x-y))\, k_\lambda(y)\, dy \right| \\ &\le \int_{-\infty}^{\infty} |f(x) - f(x-y)|\, |k_\lambda(y)|\, dy. \end{aligned} \tag{A.21}$$

Let C denote the constant in Definition A.2.1(ii). Given any $\epsilon > 0$, we can use the continuity of f at x to find $\delta > 0$ such that

$$|f(x) - f(x-y)| \leq \frac{\epsilon}{2C} \quad \text{whenever} \quad |y| \leq \delta.$$

Using the estimate in (A.21) and splitting the integral in two,

$$\begin{aligned}
|f(x) - (k_\lambda * f)(x)| &\leq \int_{-\delta}^{\delta} |f(x) - f(x-y)|\, |k_\lambda(y)|\, dy \\
&\quad + \int_{|y|\geq\delta} |f(x) - f(x-y)|\, |k_\lambda(y)|\, dy \\
&\leq \frac{\epsilon}{2C} \int_{-\delta}^{\delta} |k_\lambda(y)|\, dy + 2\, ||f||_\infty \int_{|y|\geq\delta} |k_\lambda(y)|\, dy \\
&\leq \frac{\epsilon}{2} + 2\, ||f||_\infty \int_{|y|\geq\delta} |k_\lambda(y)|\, dy.
\end{aligned}$$

Using property (iii) in Definition A.2.1 it follows that for λ sufficiently close to 0,

$$|f(x) - (k_\lambda * f)(x)| \leq \epsilon,$$

and (iii) follows. For the proof of (iv) we refer to [14].

A.3 Proof of Theorem 10.1.5

We prove (10.7) by induction. For $m = 2$ the result can be proved by a direct calculation (Exercise 10.2). Now, assume that (10.7) holds for the B-spline N_m for some $m \in \mathbb{N}$, and consider the B-spline N_{m+1}; we want to show that

$$N_{m+1}(x) = \frac{1}{m!} \sum_{j=0}^{m+1} (-1)^j \binom{m+1}{j} (x-j)_+^m, \; x \in \mathbb{R}. \tag{A.22}$$

First we notice that for $x < 0$, we have $N_{m+1}(x) = 0$ and $(x-j)_+ = 0$ for all $j = 0, \ldots, m+1$; thus, the equation in (A.22) holds. Let us now consider $x \in [0, m+1]$. Via the induction hypothesis we derive that

$$\begin{aligned}
N_{m+1}(x) &= \int_0^1 N_m(x-t)\, dt \\
&= \frac{1}{(m-1)!} \sum_{j=0}^{m} (-1)^j \binom{m}{j} \int_0^1 (x-t-j)_+^{m-1}\, dt. \quad \text{(A.23)}
\end{aligned}$$

For technical reasons we will now split the interval $[0, m+1]$ into subintervals and consider $x \in [J, J+1]$ for some arbitrary but fixed $J \in \{0, 1, \ldots, m\}$; if we can prove (A.22) for such x, the result holds for all

$x \in [0, m+1]$. In order to calculate the integrals in (A.23) we split the index set $j = 0, 1, \ldots, m$ into three groups:

- For $j = J+1, J+2, \ldots, m$,

$$\int_0^1 (x - t - j)_+^{m-1} \, dt = 0.$$

- For $j = J$,

$$\begin{aligned} \int_0^1 (x - t - J)_+^{m-1} \, dt &= \int_0^{x-J} (x - t - J)^{m-1} \, dt \\ &= \frac{1}{m}(x - J)^m. \end{aligned}$$

- For $j = 0, 1, \ldots, J-1$,

$$\begin{aligned} \int_0^1 (x - t - j)_+^{m-1} \, dt &= \int_0^1 (x - t - j)^{m-1} \, dt \\ &= \frac{1}{m}\left((x - j)^m - (x - 1 - j)^m\right). \end{aligned}$$

We now have all the information needed to calculate the sum in (A.23). Let us first consider the partial sum corresponding to $j = 0, \ldots, J-1$:

$$\begin{aligned} &\sum_{j=0}^{J-1} (-1)^j \binom{m}{j} \int_0^1 (x - t - j)_+^{m-1} dt \\ = \;& \frac{1}{m} \sum_{j=0}^{J-1} (-1)^j \binom{m}{j} \left((x - j)^m - (x - 1 - j)^m\right) \\ = \;& \frac{1}{m} \sum_{j=0}^{J-1} (-1)^j \binom{m}{j} (x - j)^m - \frac{1}{m} \sum_{j=0}^{J-1} (-1)^j \binom{m}{j} (x - 1 - j)^m = (*). \end{aligned}$$

Splitting of the sum into two and reordering of the terms leads to

$$\begin{aligned} (*) \;&=\; \frac{1}{m} \sum_{j=0}^{J-1} (-1)^j \binom{m}{j} (x - j)^m + \frac{1}{m} \sum_{j=1}^{J} (-1)^j \binom{m}{j-1} (x - j)^m \\ &=\; \frac{1}{m} x^m + \frac{1}{m} \sum_{j=1}^{J-1} (-1)^j \left(\binom{m}{j} + \binom{m}{j-1} \right) (x - j)^m \\ &\quad + \frac{1}{m} (-1)^J \binom{m}{J-1} (x - J)^m. \end{aligned}$$

Using the identity (Exercise 1.26)

$$\binom{m}{j} + \binom{m}{j-1} = \binom{m+1}{j}, \tag{A.24}$$

this implies that

$$\begin{aligned}
&\sum_{j=0}^{J-1}(-1)^j\binom{m}{j}\int_0^1 (x-t-j)_+^{m-1}\,dt \\
= \;& \frac{1}{m}x^m + \frac{1}{m}\sum_{j=1}^{J-1}(-1)^j\binom{m+1}{j}(x-j)^m + \frac{1}{m}(-1)^J\binom{m}{J-1}(x-J)^m.
\end{aligned}$$

We can now find N_{m+1} using (A.23):

$$\begin{aligned}
&N_{m+1}(x) \\
= \;& \frac{1}{(m-1)!}\sum_{j=0}^{m}(-1)^j\binom{m}{j}\int_0^1 (x-t-j)_+^{m-1}\,dt \\
= \;& \frac{1}{(m-1)!}\sum_{j=0}^{J}(-1)^j\binom{m}{j}\int_0^1 (x-t-j)_+^{m-1}\,dt \\
= \;& \frac{1}{(m-1)!}\left(\frac{1}{m}x^m + \frac{1}{m}\sum_{j=1}^{J-1}(-1)^j\binom{m+1}{j}(x-j)^m\right) \\
& + \frac{1}{(m-1)!}\frac{1}{m}(-1)^J\binom{m}{J-1}(x-J)^m \\
& + \frac{1}{(m-1)!}\frac{1}{m}(-1)^J\binom{m}{J}(x-J)^m \\
= \;& \frac{1}{m!}x^m + \frac{1}{m!}\sum_{j=1}^{J-1}(-1)^j\binom{m+1}{j}(x-j)^m \\
& + \frac{1}{m!}(-1)^J\left(\binom{m}{J-1}+\binom{m}{J}\right)(x-J)^m.
\end{aligned}$$

Using (A.24) again, this leads to

$$\begin{aligned}
N_{m+1}(x) &= \frac{1}{m!}\sum_{j=0}^{J}(-1)^j\binom{m+1}{j}(x-j)^m \\
&= \frac{1}{m!}\sum_{j=0}^{m+1}(-1)^j\binom{m+1}{j}(x-j)_+^m.
\end{aligned}$$

This proves (A.22) for $x \in [0, m+1]$. The proof that (A.22) holds for $x > m+1$ is left to the reader (Exercise 10.3). □

A.4 Proof of Theorem 11.2.2

Fix $\ell \in \mathbb{N}_0$. We search for a solution P_ℓ in terms of a power series,

$$P_\ell(x) = \sum_{k=0}^{\infty} c_k x^k. \tag{A.25}$$

Inserting (A.25) in (11.24), we obtain the equation

$$\begin{aligned}\sum_{k=2}^{\infty} c_k k(k-1)x^{k-2} \quad &- \quad \sum_{k=2}^{\infty} c_k k(k-1)x^k \\ &- \quad 2\sum_{k=1}^{\infty} c_k k x^k + \ell(\ell+1)\sum_{k=0}^{\infty} c_k x^k = 0,\end{aligned}$$

which can be rewritten as

$$\begin{aligned}\sum_{k=0}^{\infty} c_{k+2}(k+2)(k+1)x^k - \sum_{k=2}^{\infty} c_k k(k-1)x^k \quad &- \quad 2\sum_{k=1}^{\infty} c_k k x^k \\ &+ \quad \ell(\ell+1)\sum_{k=0}^{\infty} c_k x^k = 0.\end{aligned}$$

Collecting terms corresponding to the power $x^k,\ k \in \mathbb{N}_0$, yields the equation

$$\begin{aligned}[2c_2 \quad &+ \quad \ell(\ell+1)c_0] + [6c_3 - c_1(2-\ell(\ell+1))]x \\ &+ \quad \sum_{k=2}^{\infty}[c_{k+2}(k+2)(k+1) - c_k(k(k+1)-\ell(\ell+1))]x^k = 0.\end{aligned}$$

In order for this to hold for all x in some interval, we must have

$$c_2 = c_0\frac{-\ell(\ell+1)}{2},\ c_3 = c_1\frac{2-\ell(\ell+1)}{6}, \tag{A.26}$$

and, for $k \geq 2$,

$$c_{k+2} = c_k\frac{k(k+1)-\ell(\ell+1)}{(k+1)(k+2)}. \tag{A.27}$$

Note that the conditions in (A.26) correspond to the expression in (A.27) with $k=0$ and $k=1$; thus, we can put the requirements together as the condition

$$c_{k+2} = c_k\frac{k(k+1)-\ell(\ell+1)}{(k+1)(k+2)},\ k \in \mathbb{N}_0. \tag{A.28}$$

At this point, the proof has to be split into two cases, depending on ℓ being even or odd; we give the proof in the case where ℓ is even and leave the modifications for the case of odd values of ℓ to the reader.

The condition (A.28) shows that in order to determine a candidate for a solution, we have to fix choices for the parameters c_0 and c_1; as soon as this is done, the condition (A.28) determines the rest of the coefficients c_k. We choose $c_1 = 0$; this implies by (A.28) that $c_k = 0$ for all odd values of $k \in \mathbb{N}_0$.

Now, notice that any set of coefficients c_k satisfying (A.28) actually yields a polynomial solution P_ℓ of the differential equation: in fact, taking $k = \ell$, the recursion formula (A.28) shows that $c_{\ell+2} = 0$, and therefore $c_{\ell+2n} = 0$ for all $n \in \mathbb{N}$. Keeping in mind that all coefficients c_k with k odd are zero, we thus search for a solution

$$P_\ell(x) = \sum_{k=0}^{\ell} c_{2k} x^{2k} = c_0 + c_2 x^2 + \cdots + c_\ell x^\ell;$$

with $m = \ell/2$, such a solution can also be written as

$$P_\ell(x) = \sum_{k=0}^{m} c_{\ell-2k} x^{\ell-2k} = c_\ell x^\ell + c_{\ell-2} x^{\ell-2} + \cdots + c_0. \tag{A.29}$$

Comparing with (11.25), we have to show that there is a solution P_ℓ determined by

$$c_{\ell-2k} = \frac{1}{2^\ell} \frac{(-1)^k}{k!} \frac{(2\ell-2k)!}{(\ell-2k)!(\ell-k)!}, \quad k = 0, \ldots, m. \tag{A.30}$$

In order to do so, we have to show that the coefficients $c_{\ell-2k}$ in (A.30) satisfy (A.28) for any $k = 1, \ldots, m$, i.e., that

$$c_{\ell-2k+2} = \frac{(\ell-2k)(\ell-2k+1) - \ell(\ell+1)}{(\ell-2k+1)(\ell-2k+2)} c_{\ell-2k}. \tag{A.31}$$

Now, using (A.30),

$$\begin{aligned} c_{\ell-2k+2} &= c_{\ell-2(k-1)} \\ &= \frac{1}{2^\ell} \frac{(-1)^{k-1}}{(k-1)!} \frac{(2\ell-2(k-1))!}{(\ell-2(k-1))!(\ell-(k-1))!} \\ &= \frac{1}{2^\ell} \frac{(-1)^{k-1}}{(k-1)!} \frac{(2\ell-2k+2)!}{(\ell-2k+2)!(\ell-k+1)!}. \end{aligned}$$

We now rewrite this expression, in a way such that the terms in $c_{\ell-2k}$ appear:

$$\begin{aligned}
c_{\ell-2k+2} &= \frac{1}{2^\ell}\frac{(-1)^k(-k)}{k!} \\
&\quad \times \frac{(2\ell-2k+2)(2\ell-2k+1)(2\ell-2k)!}{(\ell-2k+2)(\ell-2k)(\ell-2k)!(\ell-k+1)(\ell-k)!} \\
&= \frac{1}{2^\ell}\frac{(-1)^k}{k!}\frac{(2\ell-2k)!}{(\ell-2k)!(\ell-k)!} \\
&\quad \times \frac{-k(2\ell-2k+2)(2\ell-2k+1)}{(\ell-2k+2)(\ell-2k+1)(\ell-k+1)} \\
&= c_{\ell-2k}\frac{1}{(\ell-2k+2)(\ell-2k+1)}\frac{-k(2\ell-2k+2)(2\ell-2k+1)}{(\ell-k+1)}.
\end{aligned}$$

In order to complete the proof of (A.31), we have to show that

$$\frac{-k(2\ell-2k+2)(2\ell-2k+1)}{(\ell-k+1)} = (\ell-2k)(\ell-2k+1)-\ell(\ell+1);$$

this can be done by direct calculation. □

Appendix B

B.1 List of vector spaces

Vector spaces consisting of continuous functions:

$$C[a,b] = \{f : [a,b] \to \mathbb{C} \mid f \text{ is continuous}\};$$
Banach space w.r.t. the norm $||f||_\infty = \max_{x\in[a,b]} |f(x)|$;
The norm does not come from an inner product.

$$C[a,b] = \{f : [a,b] \to \mathbb{C} \mid f \text{ is continuous}\};$$
Inner product space w.r.t. $\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}\,dx$;
Not complete w.r.t. the norm $||f|| = \sqrt{\int_a^b |f(x)|^2\,dx}$.

$$C_0(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{C} \mid f \text{ is continuous and } f(x) \to 0 \text{ as } x \to \pm\infty\};$$
Banach space w.r.t. the norm $||f||_\infty = \max_{x\in[a,b]} |f(x)|$.

$$C_c(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{C} \mid f \text{ is continuous and has compact support}\};$$
Not a Banach space w.r.t. the norm $||f||_\infty = \max_{x\in[a,b]} |f(x)|$.

O. Christensen, *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4980-7_13,

L^p-spaces:

$$L^p(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} \mid \int_{-\infty}^{\infty} |f(x)|^p \, dx < \infty \right\}, \ 1 \le p < \infty;$$

Banach space w.r.t. the norm $||f||_p = \left(\int_{-\infty}^{\infty} |f(x)|^p \, dx \right)^{1/p}$;

For $p \neq 2$, the norm does not come from an inner product.

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} \mid \int_{-\infty}^{\infty} |f(x)|^2 \, dx < \infty \right\};$$

Inner product space w.r.t. $\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)\overline{g(x)} \, dx$;

Hilbert space w.r.t. the norm $||f||_2 = \sqrt{\int_{-\infty}^{\infty} |f(x)|^2 \, dx}$.

$$L^\infty(\mathbb{R}) = \{ f : \mathbb{R} \to \mathbb{C} \mid f \text{ is bounded} \};$$

Banach space w.r.t. the norm $||f||_\infty = \sup_{x \in \mathbb{R}} |f(x)|$;

The norm does not come from an inner product.

Discrete spaces:

$$\ell^p(\mathbb{N}) = \left\{ \{x_k\}_{k=1}^{\infty} | x_k \in \mathbb{C} \text{ for all } k \in \mathbb{N} \text{ and } \sum_{k \in \mathbb{N}} |x_k|^p < \infty \right\}, 1 \le p < \infty;$$

Banach space w.r.t. the norm $||\{x_k\}_{k=1}^{\infty}||_p = \left(\sum_{k \in \mathbb{N}} |x_k|^p \right)^{1/p}$;

For $p \neq 2$, the norm does not come from an inner product.

$$\ell^2(\mathbb{N}) = \left\{ \{x_k\}_{k=1}^{\infty} \mid x_k \in \mathbb{C} \text{ for all } k \in \mathbb{N} \text{ and } \sum_{k \in \mathbb{N}} |x_k|^2 < \infty \right\};$$

Inner product space w.r.t. $\langle \{x_k\}_{k=1}^{\infty}, \{y_k\}_{k=1}^{\infty} \rangle = \sum_{k \in \mathbb{N}} x_k \overline{y_k}$;

Hilbert space w.r.t. the norm $||\{x_k\}_{k=1}^{\infty}||_2 = \left(\sum_{k \in \mathbb{N}} |x_k|^2 \right)^{1/2}$.

$$\ell^\infty(\mathbb{N}) = \{ \{x_k\}_{k=1}^{\infty} \mid x_k \in \mathbb{C} \text{ for all } k \in \mathbb{N} \text{ and } \sup_{k \in \mathbb{N}} |x_k| < \infty \};$$

Banach space w.r.t. the norm $||\{x_k\}_{k=1}^{\infty}||_\infty = \sup_{k \in \mathbb{N}} |x_k|$;

The norm does not come from an inner product.

B.2 List of special polynomials

Legendre polynomials:

$$
\begin{aligned}
P_\ell(x) &= \frac{1}{2^\ell \ell!}\frac{d^\ell}{dx^\ell}(x^2-1)^\ell \\
&= \frac{1}{2^\ell}\sum_{k=0}^{m}\frac{(-1)^k}{k!}\frac{(2\ell-2k)!}{(\ell-2k)!(\ell-k)!}x^{\ell-2k}, \\
&\quad \text{where } m=\ell/2 \text{ if } \ell \text{ is even, and } m=(\ell-1)/2 \text{ if } \ell \text{ is odd;} \\
P_0(x) &= 1, \\
P_1(x) &= x, \\
P_2(x) &= \frac{1}{2}(3x^2-1), \\
P_3(x) &= \frac{1}{2}(5x^3-3x).
\end{aligned}
$$

Associated Legendre functions:

$$
\begin{aligned}
P_{\ell,n}(x) &= (1-x^2)^{n/2}\frac{d^n P_\ell}{dx^n}(x) \\
&= \frac{1}{2^\ell \ell!}(-1)^\ell(1-x^2)^{n/2}\frac{d^{n+\ell}}{dx^{n+\ell}}(1-x^2)^\ell, \\
P_{1,1}(x) &= (1-x^2)^{1/2}, \\
P_{2,1}(x) &= 3x(1-x^2)^{1/2}, \\
P_{2,2}(x) &= 3(1-x^2), \\
P_{3,1}(x) &= \frac{3}{2}(5x^2-1)(1-x^2)^{1/2}, \\
P_{3,2}(x) &= 15x(1-x^2), \\
P_{3,3}(x) &= 15(1-x^2)^{3/2}.
\end{aligned}
$$

Laguerre polynomials:

$$
\begin{aligned}
Q_\ell(x) &= \frac{e^x}{\ell!}\frac{d^\ell}{dx^\ell}(x^\ell e^{-x}), \\
Q_0(x) &= 1, \\
Q_1(x) &= 1-x, \\
Q_2(x) &= 1-2x+\frac{x^2}{2}, \\
Q_3(x) &= 1-3x+3\frac{x^2}{2}-\frac{x^3}{6}.
\end{aligned}
$$

Hermite polynomials:

$$
\begin{aligned}
H_k(x) &= (-1)^k e^{x^2} \frac{d^k}{dx^k} e^{-x^2} \\
&= \ell! \sum_{k=0}^{m} \frac{(-1)^k}{k!(\ell-2k)!} 2^{\ell-2k} x^{\ell-2k}, \\
&\quad \text{where } m = \ell/2 \ \text{ if } \ell \text{ is even, and } m = (\ell-1)/2 \text{ if } \ell \text{ is odd,} \\
H_0(x) &= 1, \\
H_1(x) &= 2x, \\
H_2(x) &= 4x^2 - 2, \\
H_3(x) &= 8x^3 - 12x.
\end{aligned}
$$

Chebyshev polynomials of the first kind:

$$
\begin{aligned}
T_\ell(x) &= \cos(\ell \arccos x), \\
T_0(x) &= 1, \\
T_1(x) &= x, \\
T_2(x) &= 2x^2 - 1, \\
T_3(x) &= 4x^3 - 3x.
\end{aligned}
$$

Chebyshev polynomials of the second kind:

$$
\begin{aligned}
U_\ell(x) &= \frac{\sin((\ell+1)\arccos x)}{\sqrt{1-x^2}}, \\
U_0(x) &= 1, \\
U_1(x) &= 2x, \\
U_2(x) &= 4x^2 - 1, \\
U_3(x) &= 8x^3 - 4x.
\end{aligned}
$$

List of Symbols

$\forall$:	Logical sign, meaning "for all."
$\exists$:	Logical sign, meaning "there exists."
$\mathbb{R}$:	The real numbers.
$\mathbb{R}^+$:	The strictly positive real numbers.
$\mathbb{N}$:	The natural numbers: 1,2,3,....
$\mathbb{N}_0$:	The nonnegative integers: 0,1,2,3,....
$\mathbb{Z}$:	The integers.
$\mathbb{Q}$:	The rational numbers.
$\mathbb{C}$:	The complex numbers.
$\overline{x}$:	The complex conjugate of $x \in \mathbb{C}$.
X, Y :	Banach spaces.
$\mathcal{H}, \mathcal{K}$:	Hilbert spaces.
$\oplus$:	Direct sum.
$\prod_{k=1}^{\infty}$:	Infinite product.
$L^p(\mathbb{R})$:	For $p \in [1, \infty[$, the space of (measurable) functions $f : \mathbb{R} \mapsto \mathbb{C}$ for which $\int_{\mathbb{R}} \|f(x)\|^p dx < \infty$.
$L^\infty(\mathbb{R})$:	The set of bounded functions on $\mathbb{R}$.
$C^k(\mathbb{R})$:	The space of k times differentiable functions with a continuous kth derivative.
$C[a, b]$:	The space of continuous functions $f : [a, b] \to \mathbb{C}$.
$C_0(\mathbb{R})$:	The space of continuous functions $f : \mathbb{R} \to \mathbb{C}$ for which $f(x) \to 0$ as $x \to \pm\infty$.
$C_c(\mathbb{R})$:	The space of continuous functions $f : \mathbb{R} \to \mathbb{C}$ with compact support.
$\mathcal{F}f(\gamma) = \hat{f}(\gamma)$:	The Fourier transform, for $f \in L^1(\mathbb{R})$ given by $\hat{f}(\gamma) = \int_{\mathbb{R}} f(x) e^{-2\pi i x \gamma} dx$.
$\ell^p(\mathbb{N})$:	For $p \in [1, \infty[$, the space of p-summable sequences, indexed by $\mathbb{N}$.
$\ell^\infty(\mathbb{N})$:	The set of bounded sequences, indexed by $\mathbb{N}$.

χ_A : The characteristic function for a set A, $\chi_A(x) = 1$ if $x \in A$, otherwise 0.
$\overline{A}$: The closure of a set A.
$A \cap B$: The set of elements belonging to A and B.
$A \cup B$: The set of elements belonging to at least one of the sets A and B.
$A \setminus B$: The set of elements belonging to A but not to B.
A^c : The complement of a set A.
$A^\perp$: The orthogonal complement of a subset A in a Hilbert space.
$\text{supp} f$: The support of the function f: $\text{supp} f = \overline{\{x \in \mathbb{R} : f(x) \neq 0\}}$.
$\delta_{k,j}$: The Kronecker delta: $\delta_{k,j} = 1$ if $k = j$, $\delta_{k,j} = 0$ if $k \neq j$.
T_a : The translation operator $(T_a f)(x) = f(x - a)$.
E_b : The modulation operator $(E_b f)(x) = e^{2\pi i b x} f(x)$.
D_a : The dilation operator $(D_a f)(x) = \frac{1}{\sqrt{a}} f(\frac{x}{a}), \quad a > 0$.
D : The dilation operator $(Df)(x) = 2^{1/2} f(2x)$.
$\psi_{j,k}$: $\psi_{j,k}(x) = D^j T_k \psi(x) = 2^{j/2} \psi(2^j x - k)$.
N_m : B-spline of order m, supported on $[0, m]$.
B_m : Centered B-spline of order m, supported on $[-m/2, m/2]$.
P_ℓ : Legendre polynomial of order ℓ.
$P_{\ell,n}$: Associated Legendre functions.
Q_ℓ : Laguerre polynomial of order ℓ.
H_ℓ : Hermite polynomial of order ℓ.

References

[1] **(B)** N. H. Asmar: *Partial differential equations with Fourier series and boundary value problems.* Pearson Education, Upper Saddle River, NJ, 2005.

[2] **(B-C)** G. Bachman, L. Narici, and E. Beckenstein: *Fourier and wavelet analysis.* Springer, New York, 2000.

[3] **(C)** G. Birkhoff and G. Roya: *Ordinary differential equations.* Third Edition. Wiley & Sons, New York, 1978.

[4] **(B-C)** W. L. Briggs: *The DFT — an owner's manual for the discrete Fourier transform.* SIAM, Philadelphia, 1995.

[5] **(C)** O. Christensen: *Frames and bases.* Birkhäuser, Boston, 2008.

[6] **(A-B)** O. Christensen and K. Laghrida Christensen: *Linear independence and series expansions in function spaces.* Amer. Math. Monthly **113**, 7 (2006), 611–627.

[7] **(A-B)** O. Christensen and K. Laghrida Christensen: *Approximation theory.* Birkhäuser, Boston, 2004.

[8] **(C)** I. Daubechies: *Ten lectures on wavelets.* SIAM, Philadelphia, 1992.

[9] **(B)** D. J. Griffiths: *An introduction to quantum mechanics.* Prentice Hall, Englewood Cliffs, NJ, 2005.

[10] **(C)** K. Gröchenig: *Foundations of time-frequency analysis.* Birkhäuser, Boston, 2000.

[11] **(H)** A. Haar: *Zur Theorie der Orthogonalen Funktionen-Systeme.* Math. Ann. **69** (1910), 331–371.

[12] **(C)** E. Hernandez and G. Weiss: *A first course on wavelets.* CRC Press, Boca Raton, 1996.

[13] **(B)** E. W. Kamen and B. S. Heck: *Fundamentals of signals and systems using the web and Matlab.* Prentice Hall, Englewood Cliffs, NJ, 2000.

[14] **(C)** Y. Katznelson: *An introduction to harmonic analysis.* Cambridge University Press, London, 2004.

[15] **(B)** E. Kreyszig: *Advanced engineering mathematics.* Wiley & Sons, New York, 2006.

[16] **(D)** A. Ron and Z. Shen: *Compactly supported tight affine spline frames in* $L_2(\mathbb{R}^d)$. Math. Comp. **67** (1998), 191–207.

[17] **(B-C)** W. Rudin: *Real and complex analysis.* McGraw Hill, New York, 1966.

[18] **(B-C)** D. Walnut: *An introduction to wavelet analysis.* Birkhäuser, Boston, 2001.

[19] **(C)** Young, R.: *An introduction to nonharmonic Fourier series.* Academic Press, New York, 1980 (revised first edition 2001).

Index