Ten Lectures on Wavelets

INGRID DAUBECHIES
Rutgers University and
AT&T Bell Laboratories

Ten Lectures on Wavelets

SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS

Printed by Capital City Press, Montpelier, Vermont
Second Printing 1992

Library of Congress Cataloging-in-Publication Data

Daubechies, Ingrid.
Ten lectures on wavelets / Ingrid Daubechies
p. cm. — (CBMS-NSF regional conference series in applied mathematics ; 61)
Includes bibliographical references and index.
ISBN 0-89871-274-2
1. Wavelets (Mathematics)—Congresses. I. Title. II. Series.
QA403.3.D38 1992 92-13201
515'.2433—dc20

Introduction

Wavelets are a relatively recent development in applied mathematics. Their name itself was coined approximately a decade ago (Morlet, Arens, Fourgeau, and Giard (1982), Morlet (1983), Grossmann and Morlet (1984)); in the last ten years interest in them has grown at an explosive rate. There are several reasons for their present success. On the one hand, the concept of wavelets can be viewed as a synthesis of ideas which originated during the last twenty or thirty years in engineering (subband coding), physics (coherent states, renormalization group), and pure mathematics (study of Calderón–Zygmund operators). As a consequence of these interdisciplinary origins, wavelets appeal to scientists and engineers of many different backgrounds. On the other hand, wavelets are a fairly simple mathematical tool with a great variety of possible applications. Already they have led to exciting applications in signal analysis (sound, images) (some early references are Kronland-Martinet, Morlet and Grossmann (1987), Mallat (1989b), (1989c); more recent references are given later) and numerical analysis (fast algorithms for integral transforms in Beylkin, Coifman, and Rokhlin (1991)); many other applications are being studied. This wide applicability also contributes to the interest they generate.

This book contains ten lectures I delivered as the principal speaker at the CBMS conference on wavelets organized in June 1990 by the Mathematics Department at the University of Lowell, Massachusetts. According to the usual format of the CBMS conferences, other speakers (G. Battle, G. Beylkin, C. Chui, A. Cohen, R. Coifman, K. Gröchenig, J. Liandrat, S. Mallat, B. Torrésani, and A. Willsky) provided lectures on their work related to wavelets. Moreover, three workshops were organized, on applications to physics and inverse problems (chaired by B. DeFacio), group theory and harmonic analysis (H. Feichtinger), and signal analysis (M. Vetterli). The audience consisted of researchers active in the field of wavelets as well as of mathematicians and other scientists and engineers who knew little about wavelets and hoped to learn more. This second group constituted the largest part of the audience. I saw it as my task to provide a tutorial on wavelets to this part of the audience, which would then be a solid grounding for more recent work exposed by the other lecturers and myself. Consequently, about two thirds of my lectures consisted of "basic wavelet theory," the other third being devoted to more recent and unpublished work. This division is reflected in the present write-up as well. As a result, I believe that this book will be useful as an introduction to the subject, to be used either for individual reading, or for a seminar or graduate course. None of the other lectures or workshop papers presented at the CBMS conference have been incorporated here. As a result, this presentation is biased more toward my own work than the CBMS conference was. In many instances I have included pointers to references for further reading or a detailed exposition of particular applications, complementing the present text. Other books on wavelets published include *Wavelets and Time Frequency Methods* (Combes, Grossmann, and Tchamitchian (1987)), which contains the proceedings of the International Wavelet Conference held in Marseille, France, in December 1987, *Ondelettes*, by Y. Meyer (1990) (in French; English translation expected soon), which contains a mathematically more expanded treatment than the present lectures, with fewer forays into other fields however, *Les Ondelettes en* 1989, edited by P. G. Lemarié (1990), a collection of talks given at the Université Paris XI in the spring of 1989, and *An Introduction to Wavelets*, by C. K. Chui (1992b), an introduction from the approximation theory viewpoint. The proceedings of the International Wavelet Conference in May 1989, held again in Marseille, are due to come out soon (Meyer (1992)). Moreover, many of the other contributors to the CBMS conference, as well as some wavelet researchers who could not attend, were invited to write an essay on their wavelet work; the result is the essay collection *Wavelets and their Applications* (Ruskai et al. (1992)), which can be considered a companion book to this one. Another wavelet essay book is *Wavelets: A Tutorial in Theory and Applications*, edited by C. K. Chui (1992c); in addition, I know of several other wavelet essay books in preparation (edited by J. Benedetto and M. Frazier, another by M. Barlaud), as well as a monograph by M. Holschneider; there was a special wavelet issue of IEEE *Trans. Inform. Theory* in March of 1992; there will be another one, later in 1992, of *Constructive Approximation Theory*, and one in 1993, of IEEE *Trans. Sign. Proc.* In addition, several recent books include chapters on wavelets. Examples are *Multirate Systems and Filter Banks* by P. P. Vaidyanathan (1992) and *Quantum Physics, Relativity and Complex Spacetime: Towards a New Synthesis* by G. Kaiser (1990). Readers interested in the present lectures will find these books and special issues useful for many details and other aspects not fully presented here. It is moreover clear that the subject is still developing rapidly.

This book more or less follows the path of my lectures: each of the ten chapters stands for one of the ten lectures, presented in the order in which they were delivered. The first chapter presents a quick overview of different aspects of the wavelet transform. It sketches the outlines of a big fresco; subsequent chapters then fill in more detail. From there on, we proceed to the continuous wavelet transform (Chapter 2; with a short review of bandlimited functions and Shannon's theorem), to discrete but redundant wavelet transforms (frames; Chapter 3) and to a general discussion of time-frequency density and the possible existence of orthonormal basis (Chapter 4). Many of the results in Chapters 2–4 can be formulated for the windowed Fourier transform as well as the wavelet transform, and the two cases are presented in parallel, with analogies and differences pointed out as we go along. The remaining chapters all focus on orthonormal bases of wavelets: multiresolution analysis and a first general strategy for the construction of orthonormal wavelet bases (Chapter 5), orthonormal bases of compactly supported wavelets and their link to subband coding (Chapter 6), sharp regularity estimates for these wavelet bases (Chapter 7), symmetry for compactly supported wavelet bases (Chapter 8). Chapter 9 shows that orthonormal bases are "good" bases for many functional spaces where Fourier methods are not well adapted. This chapter is the most mathematical of the whole book; most of its material is not connected to the applications discussed in other chapters, so that it can be skipped by readers uninterested in this aspect of wavelet theory. I included it for several reasons: the kind of estimates used in the proof are very important for harmonic analysis, and similar (but more complicated) estimates in the proof of the "T(1)"-theorem of David and Journé have turned out to be the groundwork for the applications to numerical analysis in the work of Beylkin, Coifman, and Rokhlin (1991). Moreover, the Calderón–Zygmund theorem, explained in this chapter, illustrates how techniques using different scales, one of the forerunners of wavelets, were used in harmonic analysis long before the advent of wavelets. Finally, Chapter 10 sketches several extensions of the constructions of orthonormal wavelet bases: to more than one dimension, to dilation factors different from two (even noninteger), with the possibility of better frequency localization, and to wavelet bases on a finite interval instead of the whole line. Every chapter concludes with a section of numbered "Notes," referred to in the text of the chapter by superscript numbers. These contain additional references, extra proofs excised to keep the text flowing, remarks, etc.

This book is a mathematics book: it states and proves many theorems. It also presupposes some mathematical background. In particular, I assume that the reader is familiar with the basic properties of the Fourier transform and Fourier series. I also use some basic theorems of measure and integration theory (Fatou's lemma, dominated convergence theorem, Fubini's theorem; these can be found in any good book on real analysis). In some chapters, familiarity with basic Hilbert space techniques is useful. A list of the basic notions and theorems used in the book is given in the Preliminaries.

The reader who finds that he or she does not know all of these prerequisites should not be dismayed, however; most of the book can be followed with just the basic notions of Fourier analysis. Moreover, I have tried to keep a very pedestrian pace in almost all the proofs, at the risk of boring some mathematically sophisticated readers. I hope therefore that these lecture notes will interest people other than mathematicians. For this reason I have often shied away from the "Definition–Lemma–Proposition–Theorem–Corollary" sequence, and I have tried to be intuitive in many places, even if this meant that the exposition became less succinct. I hope to succeed in sharing with my readers some of the excitement that this interdisciplinary subject has brought into my scientific life.

I want to take this opportunity to express my gratitude to the many people who made the Lowell conference happen: the CBMS board, and the Mathematics Department of the University of Lowell, in particular Professors G. Kaiser and

M. B. Ruskai. The success of the conference, which unexpectedly turned out to have many more participants than customary for CBMS conferences, was due in large part to its very efficient organization. As experienced conference organizer I. M. James (1991) says, "every conference is mainly due to the efforts of a single individual who does almost all the work"; for the 1990 Wavelet CBMS conference, this individual was Mary Beth Ruskai. I am especially grateful to her for proposing the conference in the first place, for organizing it in such a way that I had a minimal paperwork load, while keeping me posted about all the developments, and for generally being the organizational backbone, no small task. Prior to the conference I had the opportunity to teach much of this material as a graduate course in the Mathematics Department of the University of Michigan, in Ann Arbor. My one-term visit there was supported jointly by a Visiting Professorship for Women from the National Science Foundation, and by the University of Michigan. I would like to thank both institutions for their support. I would also like to thank all the faculty and students who sat in on the course, and who provided feedback and useful suggestions. The manuscript was typeset by Martina Sharp, who I thank for her patience and diligence, and for doing a wonderful job. I wouldn't even have attempted to write this book without her. I am grateful to Jeff Lagarias for editorial comments. Several people helped me spot typos in the galley proofs, and I am grateful to all of them; I would like to thank especially Pascal Auscher, Gerry Kaiser, Ming-Jun Lai, and Martin Vetterli. All remaining mistakes are of course my responsiblity. I also would like to thank Jim Driscoll and Sharon Murrel for helping me prepare the author index. Finally, I want to thank my husband Robert Calderbank for being extremely supportive and committed to our two-career-track with family, even though it occasionally means that he as well as I prove a few theorems less.

Ingrid Daubechies
AT&T Bell Laboratories
and
Rutgers University

In this second printing several minor mistakes and many typographical errors have been corrected. I am grateful to everybody who helped me to spot them. I have also updated a few things: some of the previously unpublished references have appeared and some of the problems that were listed as open have been solved. I have made no attempt to include the many other interesting papers on wavelets that have appeared since the first printing; in any case, the list of references was not and is still not meant as a complete bibliography of the subject.

Ingrid Daubechies, Sept. 1992

Preliminaries and Notation

This preliminary chapter fixes notation conventions and normalizations. It also states some basic theorems that will be used later in the book. For those less familiar with Hilbert and Banach spaces, it contains a very brief primer. (This primer should be used mainly as a reference, to come back to in those instances when the reader comes across some Hilbert or Banach space language that she or he is unfamiliar with. For most chapters, these concepts are not used.)

Let us start by some notation conventions. For $x \in \mathbb{R}$, we write $\lfloor x \rfloor$ for the largest integer not exceeding x,

$$\lfloor x \rfloor = \max \{n \in \mathbb{Z};\ \ n \le x\} .$$

For example, $\lfloor 3/2 \rfloor = 1$, $\lfloor -3/2 \rfloor = -2$, $\lfloor -2 \rfloor = -2$. Similarly, $\lceil x \rceil$ is the smallest integer which is larger than or equal to x.

If $a \to 0$ (or ∞), then we denote by $O(a)$ any quantity that is bounded by a constant times a, by $o(a)$ any quantity that tends to 0 (or ∞) when a does.

The end of a proof is always marked with a ■; for clarity, many remarks or examples are ended with a □.

In many proofs, C denotes a "generic" constant, which need not have the same value throughout the proof. In chains of inequalities, I often use $C, C', C'', \cdots$ or $C_1, C_2, C_3, \cdots$ to avoid confusion.

We use the following convention for the Fourier transform (in one dimension):

$$(\mathcal{F}f)(\xi) = \hat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, e^{-ix\xi} f(x) . \qquad (0.0.1)$$

With this normalization, one has

$$\begin{aligned} \|\hat{f}\|_{L^2} &= \|f\|_{L^2} , \\ |\hat{f}(\xi)| &\le (2\pi)^{-1/2} \|f\|_{L^1} , \end{aligned}$$

where

$$\|f\|_{L^p} = \left[\int dx\, |f(x)|^p \right]^{1/p} . \qquad (0.0.2)$$

Inversion of the Fourier transform is then given by

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\xi \, e^{i\xi x} (\mathcal{F}f)(\xi) = (\mathcal{F}f)^{\vee}(x) \,, \\ \check{g}(x) &= \hat{g}(-x) \,. \end{aligned} \tag{0.0.3}$$

Strictly speaking, (0.0.1), (0.0.3) are well defined only if f, respectively $\mathcal{F}f$, are absolutely integrable; for general L^2-functions f, e.g., we should define $\mathcal{F}f$ via a limiting process (see also below). We will implicitly assume that the adequate limiting process is used in all cases, and write, with a convenient abuse of notation, formulas similar to (0.0.1) and (0.0.3) even when a limiting process is understood.

A standard property of the Fourier transform is:

$$\mathcal{F}\left(\frac{d^\ell}{dx^\ell} f\right) = (i\xi)^\ell \, (\mathcal{F}f)(\xi) \,,$$

hence

$$\int dx \, |f^{(\ell)}(x)|^2 < \infty \leftrightarrow \int d\xi \, |\xi|^{2\ell} \, |\hat{f}(\xi)|^2 < \infty \,,$$

with the notation $f^{(\ell)} = \frac{d^\ell}{dx^\ell} f$.

If a function f is compactly supported, i.e., $f(x) = 0$ if $x < a$ or $x > b$, where $-\infty < a < b < \infty$, then its Fourier transform $\hat{f}(\xi)$ is well defined also for complex ξ, and

$$\begin{aligned} |\hat{f}(\xi)| &\le (2\pi)^{-1/2} \int_a^b dx \, e^{(\mathrm{Im}\,\xi)x} \, |f(x)| \\ &\le (2\pi)^{-1/2} \, \|f\|_{L^1} \begin{cases} e^{b\,(\mathrm{Im}\,\xi)} & \text{if } \mathrm{Im}\,\xi \ge 0 \\ e^{a\,(\mathrm{Im}\,\xi)} & \text{if } \mathrm{Im}\,\xi \le 0 \,. \end{cases} \end{aligned}$$

If f is moreover infinitely differentiable, then the same argument can be applied to $f^{(\ell)}$, leading to bounds on $|\xi|^\ell \, |\hat{f}(\xi)|$. For a C^∞ function f with support $[a,b]$ there exist therefore constants C_N so that the analytic extension of the Fourier transform of f satisfies

$$|\hat{f}(\xi)| \le C_N (1 + |\xi|)^{-N} \begin{cases} e^{b\,\mathrm{Im}\,\xi} & \text{if } \mathrm{Im}\,\xi \ge 0 \\ e^{a\,\mathrm{Im}\,\xi} & \text{if } \mathrm{Im}\,\xi \le 0 \,. \end{cases} \tag{0.0.4}$$

Conversely, any entire function which satisfies bounds of the type (0.0.4) for all $N \in \mathbb{N}$ is the analytic extension of the Fourier transform of a C^∞ function with support in $[a,b]$. This is the Paley–Wiener theorem.

We will occasionally encounter (tempered) distributions. These are linear maps T from the set $\mathcal{S}(\mathbb{R})$ (consisting of all C^∞ functions that decay faster than any negative power $(1 + |x|)^{-N}$) to $\mathbb{C}$, such that for all $m, n \in \mathbb{N}$, there exists $C_{n,m}$ for which

$$|T(f)| \le C_{n,m} \sup_{x \in \mathbb{R}} |(1 + |x|)^n \, f^{(m)}(x)|$$

holds, for all $f \in \mathcal{S}(\mathbb{R})$. The set of all such distributions is called $\mathcal{S}'(\mathbb{R})$. Any polynomially bounded function F can be interpreted as a distribution, with $F(f) = \int dx\, \overline{F(x)}\, f(x)$. Another example is the so-called "δ-function" of Dirac, $\delta(f) = f(0)$. A distribution T is said to be supported in $[a,b]$ if $T(f) = 0$ for all functions f the support of which has empty intersection with $[a,b]$. One can define the Fourier transform $\mathcal{F}T$ or $\hat{T}$ of a distribution T by $\hat{T}(f) = T(\hat{f})$ (if T is a function, then this coincides with our earlier definition). There exists a version of the Paley Wiener theorem for distributions: an entire function $\hat{T}(\xi)$ is the analytic extension of the Fourier transform of a distribution T in $\mathcal{S}'(\mathbb{R})$ supported in $[a,b]$ if and only if, for some $N \in \mathbb{N}$, $C_N > 0$,

$$|\hat{T}(\xi)| \le C_N (1 + |\xi|)^N \begin{cases} e^{b\, \mathrm{Im}\, \xi} & \mathrm{Im}\, \xi \ge 0 \\ e^{a\, \mathrm{Im}\, \xi} & \mathrm{Im}\, \xi \le 0\,. \end{cases}$$

The only measure we will use is Lebesgue measure, on $\mathbb{R}$ and $\mathbb{R}^n$. We will often denote the (Lebesgue) measure of S by $|S|$; in particular, $|[a,b]| = b - a$ (where $b > a$).

Well-known theorems from measure and integration theory which we will use include

Fatou's lemma. *If $f_n \ge 0$, $f_n(x) \to f(x)$ almost everywhere (i.e., the set of points where pointwise convergence fails has zero measure with respect to Lebesgue measure), then*

$$\int dx\; f(x) \le \limsup_{n\to\infty} \int dx\; f_n(x)\,.$$

In particular, if this lim sup *is finite, then f is integrable.*

(The lim sup of a sequence is defined by

$$\limsup_{n\to\infty} \alpha_n = \lim_{n\to\infty} \left[\sup\{\alpha_k;\ k \ge n\}\right];$$

every sequence, even if it does not have a limit (such as $\alpha_n = (-1)^n$), has a lim sup (which may be ∞); for sequences that converge to a limit, the lim sup coincides with the limit.)

Dominated convergence theorem. *Suppose $f_n(x) \to f(x)$ almost everywhere. If $|f_n(x)| \le g(x)$ for all n, and $\int dx\; g(x) < \infty$, then f is integrable, and*

$$\int dx\; f(x) = \lim_{n\to\infty} \int dx\; f_n(x)\,.$$

Fubini's theorem. *If $\int dx [\int dy\; |f(x,y)|] < \infty$, then*

$$\begin{aligned} \int dx \int dy\; f(x,y) &= \int dx \left[\int dy\; f(x,y)\right] \\ &= \int dy \left[\int dx\; f(x,y)\right] \end{aligned}$$

i.e., the order of the integrations can be permuted.

In these three theorems the domain of integration can be any measurable subset of $\mathbb{R}$ (or $\mathbb{R}^2$ for Fubini).

When Hilbert spaces are used, they are usually denoted by $\mathcal{H}$, unless they already have a name. We will follow the mathematician's convention and use scalar products which are linear in the *first* argument:

$$\langle \lambda_1 u_1 + \lambda_2 u_2,\ v\rangle = \lambda_1 \langle u_1,\ v\rangle + \lambda_2 \langle u_2,\ v\rangle .$$

As usual, we have

$$\langle v, u\rangle = \overline{\langle u, v\rangle} ,$$

where $\bar{\alpha}$ denotes the complex conjugate of α, and $\langle u, u\rangle \geq 0$ for all $u \in \mathcal{H}$. We define the norm $\|u\|$ of u by

$$\|u\|^2 = \langle u, u\rangle . \tag{0.0.5}$$

In a Hilbert space, $\|u\| = 0$ implies $u = 0$, and all Cauchy sequences (with respect to $\|\ \ \|$) have limits within the space. (More explicitly, if $u_n \in \mathcal{H}$ and if $\|u_n - u_m\|$ becomes arbitrarily small if n, m are large enough—i.e., for all $\epsilon > 0$, there exists n_0, depending on ϵ, so that $\|u_n - u_m\| \leq \epsilon$ if $n, m \geq n_0$–, then there exists $u \in \mathcal{H}$ so that the u_n tend to u for $n \to \infty$, i.e., $\lim_{n\to\infty} \|u - u_n\| = 0$.)

A standard example of such a Hilbert space is $L^2(\mathbb{R})$, with

$$\langle f, g\rangle = \int dx\ f(x)\ \overline{g(x)} .$$

Here the integration runs from $-\infty$ to ∞; we will often drop the integration bounds when the integral runs over the whole real line.

Another example is $\ell^2(\mathbb{Z})$, the set of all square summable sequences of complex numbers indexed by integers, with

$$\langle c, d\rangle = \sum_{n=-\infty}^{\infty} c_n\ \overline{d_n} .$$

Again, we will often drop the limits on the summation index when we sum over all integers. Both $L^2(\mathbb{R})$ and $\ell^2(\mathbb{Z})$ are infinite-dimensional Hilbert bases. Even simpler are finite-dimensional Hilbert spaces, of which $\mathbb{C}^k$ is the standard example, with the scalar product

$$\langle u, v\rangle = \sum_{j=1}^{k} u_j\ \bar{v}_j ,$$

for $u = (u_1, \cdots, u_k)$, $v = (v_1, \cdots, v_k) \in \mathbb{C}^k$.

Hilbert spaces always have orthonormal bases, i.e., there exist families of vectors e_n in $\mathcal{H}$

$$\langle e_n, e_m\rangle = \delta_{n,m}$$

and

$$\|u\|^2 = \sum_n |\langle u, e_n\rangle|^2$$

for all $u \in \mathcal{H}$. (We only consider separable Hilbert spaces, i.e., spaces in which orthonormal bases are countable.) Examples of orthonormal bases are the Hermite functions in $L^2(\mathbb{R})$, the sequences e_n defined by $(e_n)_j = \delta_{n,j}$, with $n, j \in \mathbb{Z}$ in $\ell^2(\mathbb{Z})$ (i.e., all entries but the nth vanish), or the k vectors $e_1, \cdots, e_k$ in $\mathbb{C}^k$ defined by $(e_\ell)_m = \delta_{\ell,m}$, with $1 \leq \ell, m \leq k$. (We use Kronecker's symbol δ with the usual meaning: $\delta_{i,j} = 1$ if $i = j$, 0 if $i \neq j$.)

A standard inequality in a Hilbert space is the Cauchy–Schwarz inequality,

$$|\langle v, w\rangle| \leq \|v\| \; \|w\| \, , \tag{0.0.6}$$

easily proved by writing (0.0.5) for appropriate linear combinations of v and w. In particular, for $f, g \in L^2(\mathbb{R})$, we have

$$\left|\int dx \; f(x) \; \overline{g(x)}\right| \leq \left(\int dx \; |f(x)|^2\right)^{1/2} \left(\int dx \; |g(x)|^2\right)^{1/2} ,$$

and for $c = (c_n)_{n\in\mathbb{Z}}$, $d = (d_n)_{n\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$,

$$\sum_n c_n \overline{d_n} \leq \left(\sum_n |c_n|^2\right)^{1/2} \left(\sum_n |d_n|^2\right)^{1/2} .$$

A consequence of (0.0.6) is

$$\|u\| = \sup_{v,\; \|v\| \leq 1} |\langle u, v\rangle| = \sup_{v,\; \|v\|=1} |\langle u, v\rangle| \, . \tag{0.0.7}$$

"Operators" on $\mathcal{H}$ are linear maps from $\mathcal{H}$ to another Hilbert space, often $\mathcal{H}$ itself. Explicitly, if A is an operator on $\mathcal{H}$, then

$$A(\lambda_1 u_1 + \lambda_2 u_2) = \lambda_1 A u_1 + \lambda_2 A u_2 \, .$$

An operator is continuous if $Au - Av$ can be made arbitrarily small by making $u - v$ small. Explicitly, for all $\epsilon > 0$ there should exist δ (depending on ϵ) so that $\|u - v\| \leq \delta$ implies $\|Au - Av\| \leq \epsilon$. If we take $v = 0$, $\epsilon = 1$, then we find that, for some $b > 0$, $\|Au\| \leq 1$ if $\|u\| \leq b$. For any $w \in \mathcal{H}$ we can define $w' = \frac{b}{\|w\|} w$; clearly $\|w'\| \leq b$ and therefore $\|Aw\| = \frac{\|w\|}{b} \|Aw'\| \leq b^{-1}\|w\|$. If $\|Aw\|/\|w\|$ ($w \neq 0$) is bounded, then the operator A is called bounded. We have just seen that any continuous operator is bounded; the reverse is also true. The norm $\|A\|$ of A is defined by

$$\|A\| = \sup_{u\in\mathcal{H},\; \|u\|\neq 0} \|Au\|/\|u\| = \sup_{\|u\|=1} \|Au\| \, . \tag{0.0.8}$$

It immediately follows that, for all $u \in \mathcal{H}$,

$$\|Au\| \leq \|A\| \; \|u\| \, .$$

Operators from $\mathcal{H}$ to $\mathbb{C}$ are called "linear functionals." For bounded linear functionals one has Riesz' representation theorem: for any ℓ: $\mathcal{H} \to \mathbb{C}$, linear and bounded, i.e., $|\ell(u)| \le C\|u\|$ for all $u \in \mathcal{H}$, there exists a unique $v_\ell \in \mathcal{H}$ so that $\ell(u) = \langle u, v_\ell \rangle$.

An operator U from $\mathcal{H}_1$ to $\mathcal{H}_2$ is an isometry if $\langle Uv, Uw \rangle = \langle v, w \rangle$ for all $v, w \in \mathcal{H}_1$; U is unitary if moreover $U\mathcal{H}_1 = \mathcal{H}_2$, i.e., every element $v_2 \in \mathcal{H}_2$ can be written as $v_2 = Uv_1$ for some $v_1 \in \mathcal{H}_1$. If the e_n constitute an orthonormal basis in $\mathcal{H}_1$, and U is unitary, then the Ue_n constitute an orthonormal basis in $\mathcal{H}_2$. The reverse is also true: any operator that maps an orthonormal basis to another orthonormal basis is unitary.

A set D is called dense in $\mathcal{H}$ if every $u \in \mathcal{H}$ can be written as the limit of some sequence of u_n in D. (One then says that the closure of D is all of $\mathcal{H}$. The closure of a set S is obtained by adding to it all the v that can be obtained as limits of sequences in S.) If Av is only defined for $v \in D$, but we know that

$$\|Av\| \le C\|v\| \quad \text{for all } v \in D , \tag{0.0.9}$$

then we can extend A to all of $\mathcal{H}$ "by continuity." Explicitly: if $u \in \mathcal{H}$, find $u_n \in D$ so that $\lim_{n\to\infty} u_n = u$. Then the u_n are necessarily a Cauchy sequence, and because of (0.0.9), so are the Au_n; the Au_n have therefore a limit, which we call Au (it does not depend on the particular sequence u_n that was chosen).

One can also deal with unbounded operators, i.e., A for which there exists no finite C such that $\|Au\| \le C\|u\|$ holds for all $u \in \mathcal{H}$. It is a fact of life that these can usually only be defined on a dense set D in $\mathcal{H}$, and cannot be extended by the above trick (since they are not continuous). An example is $\frac{d}{dx}$ in $L^2(\mathbb{R})$, where we can take $D = C_0^\infty(\mathbb{R})$, the set of all infinitely differentiable functions with compact support, for D. The dense set on which the operator is defined is called its domain.

The adjoint A^* of a bounded operator A from a Hilbert space $\mathcal{H}_1$ to a Hilbert space $\mathcal{H}_2$ (which may be $\mathcal{H}_1$ itself) is the operator from $\mathcal{H}_2$ to $\mathcal{H}_1$ defined by

$$\langle u_1, A^* u_2 \rangle = \langle Au_1, u_2 \rangle ,$$

which should hold for all $u_1 \in \mathcal{H}_1$, $u_2 \in \mathcal{H}_2$. (The existence of A^* is guaranteed by Riesz' representation theorem: for fixed u_2, we can define a linear functional ℓ on $\mathcal{H}_1$ by $\ell(u_1) = \langle Au_1, u_2 \rangle$. It is clearly bounded, and corresponds therefore to a vector v so that $\langle u_1, v \rangle = \ell(u)$. It is easy to check that the correspondence $u_2 \to v$ is linear; this defines the operator A^*.) One has

$$\|A^*\| = \|A\|, \qquad \|A^*A\| = \|A\|^2 .$$

If $A^* = A$ (only possible if A maps $\mathcal{H}$ to itself), then A is called self-adjoint. If a self-adjoint operator A satisfies $\langle Au, u \rangle \ge 0$ for all $u \in \mathcal{H}$, then it is called a positive operator; this is often denoted $A \ge 0$. We will write $A \ge B$ if $A - B$ is a positive operator.

Trace-class operators are special operators such that $\sum_n |\langle Ae_n, e_n \rangle|$ is finite for *all* orthonormal bases in $\mathcal{H}$. For such a trace-class operator, $\sum_n \langle Ae_n, e_n \rangle$ is independent of the chosen orthonormal basis; we call this sum the trace of A,

$$\operatorname{tr} A = \sum_n \langle Ae_n, e_n \rangle .$$

If A is positive, then it is sufficient to check whether $\sum_n \langle Ae_n, e_n\rangle$ is finite for only one orthonormal basis; if it is, then A is trace-class. (This is not true for non-positive operators!)

The spectrum $\sigma(A)$ of an operator A from $\mathcal{H}$ to itself consists of all the $\lambda \in \mathbb{C}$ such that $A - \lambda\,\mathrm{Id}$ (Id stands for the identity operator, $\mathrm{Id}\,u = u$) does not have a bounded inverse. In a finite-dimensional Hilbert space, $\sigma(A)$ consists of the eigenvalues of A; in the infinite-dimensional case, $\sigma(A)$ contains all the eigenvalues (constituting the point spectrum) but often contains other λ as well, constituting the continuous spectrum. (For instance, in $L^2(\mathbb{R})$, multiplication of $f(x)$ with $\sin \pi x$ has no point spectrum, but its continuous spectrum is $[-1, 1]$.) The spectrum of a self-adjoint operator consists of only real numbers; the spectrum of a positive operator contains only non-negative numbers. The spectral radius $\rho(A)$ is defined by

$$\rho(A) = \sup\,\{|\lambda|;\ \lambda \in \sigma(A)\}\ .$$

It has the properties

$$\rho(A) \leq \|A\| \quad \text{and} \quad \rho(A) = \lim_{n\to\infty} \|A^n\|^{1/n}\ .$$

Self-adjoint operators can be diagonalized. This is easiest to understand if their spectrum consists only of eigenvalues (as is the case in finite dimensions). One then has

$$\sigma(A) = \{\lambda_n;\ n \in \mathbb{N}\}\ ,$$

with a corresponding orthonormal family of eigenvectors,

$$Ae_n = \lambda_n e_n\ .$$

It then follows that, for all $u \in \mathcal{H}$,

$$Au = \sum_n \langle Au, e_n\rangle e_n = \sum_n \langle u, Ae_n\rangle e_n = \sum_n \lambda_n \langle u, e_n\rangle e_n\ ,$$

which is the "diagonalization" of A. (The spectral theorem permits us to generalize this if part (or all) of the spectrum is continuous, but we will not need it in this book.) If two operators commute, i.e., $ABu = BAu$ for all $u \in \mathcal{H}$, then they can be diagonalized simultaneously: there exists an orthonormal basis such that

$$Ae_n = \alpha_n e_n \quad \text{and} \quad Be_n = \beta_n e_n\ .$$

Many of these properties for bounded operators can also be formulated for unbounded operators: adjoints, spectrum, diagonalization all exist for unbounded operators as well. One has to be very careful with domains, however. For instance, generalizing the simultaneous diagonalization of commuting operators requires a careful definition of commuting operators: there exist pathological examples where A, B are both defined on a domain D, where AB and BA both make sense on D and are equal on D, but where A and B nevertheless are not simultaneously diagonalizable (because D was chosen "too small"; see, e.g., Reed and Simon (1971) for an example). The proper definition of commuting for unbounded self-adjoint operators uses associated bounded operators: H_1 and H_2 commute if their associated unitary evolution operators commute. For a self-adjoint operator H, the associated unitary evolution operators U_t are defined as follows: for any $v \in D$, the domain of H (beware: the domain of a self-adjoint operator is not just any dense set on which H is well defined), $U_T v$ is the solution $v(t)$ at time $t = T$ of the differential equation

$$i \frac{d}{dt} v(t) = Hv(t) ,$$

with initial condition $v(0) = v$.

Banach spaces share many properties with but are more general than Hilbert spaces. They are linear spaces equipped with a norm (which need not be and generally is not derived from a scalar product), complete with respect to that norm (i.e., all Cauchy sequences converge; see above). Some of the concepts we reviewed above for Hilbert spaces also exist in Banach spaces; e.g., bounded operators, linear functionals, spectrum and spectral radius. An example of a Banach space that is not a Hilbert space is $L^p(\mathbf{R})$, the set of all functions f on $\mathbf{R}$ such that $\|f\|_{L^p}$ (see (0.0.2)) is finite, with $1 \leq p < \infty$, $p \neq 2$. Another example is $L^\infty(\mathbf{R})$, the set of all bounded functions on $\mathbf{R}$, with $\|f\|_{L^\infty} = s\sup_{x\in\mathbf{R}} |f(x)|$. The dual E^* of a Banach space E is the set of all bounded linear functionals on E; it is also a linear space, which comes with a natural norm (defined as in (0.0.7)), with respect to which it is complete: E^* is a Banach space itself. In the case of the L^p-spaces, $1 \leq p < \infty$, it turns out that elements of L^q, where p and q are related by $p^{-1} + q^{-1} = 1$, define bounded linear functionals on L^p. Indeed, one has Hölder's inequality,

$$\left| \int dx \; f(x) \, \overline{g(x)} \right| \leq \|f\|_{L^p} \, \|g\|_{L^q} \; .$$

It turns out that all bounded linear functionals on L^p are of this type, i.e., $(L^p)^* = L^q$. In particular, L^2 is its own dual; by Riesz' representation theorem (see above), every Hilbert space is its own dual. The adjoint A^* of an operator A from E_1 to E_2 is now an operator from E_2^* to E_1^*, defined by

$$(A^*\ell_2)(v_1) = \ell_2(Av_1) \; .$$

There exist different types of bases in Banach spaces. (We will again only consider separable spaces, in which bases are countable.) The e_n constitute a Schauder basis if, for all $v \in E$, there exist unique $\mu_n \in \mathbf{C}$ so that $v = \lim_{N\to\infty} \sum_{n=1}^N \mu_n e_n$ (i.e., $\|v - \sum_{n=1}^N \mu_n e_n\| \to 0$ as $N \to \infty$). The uniqueness requirement of the μ_n forces the e_n to be linearly independent, in the sense that no e_n can be in the closure of the linear span of all the others, i.e., there exist no γ_m so that $e_n = \lim_{N\to\infty} \sum_{m=1,\, m\neq n}^N \gamma_m e_m$. In a Schauder basis, the ordering of the e_n may be important. A basis is called unconditional if in addition it satisfies one of the following two equivalent properties:

- whenever $\sum_n \mu_n e_n \in E$, it follows that $\sum_n |\mu_n| e_n \in E$;
- if $\sum_n \mu_n e_n \in E$, and $\epsilon_n = \pm 1$, randomly chosen for every n, then $\sum_n \mu_n \epsilon_n e_n \in E$.

For an unconditional basis, the order in which the basis vectors are taken does not matter. Not all Banach spaces have unconditional bases: $L^1(\mathbb{R})$ and $L^\infty(\mathbb{R})$ do not.

In a Hilbert space $\mathcal{H}$, an unconditional basis is also called a Riesz basis. A Riesz basis can also be characterized by the following equivalent requirement: there exist $\alpha > 0$, $\beta < \infty$ so that

$$\alpha \|u\|^2 \le \sum_n |\langle u, e_n \rangle|^2 \le \beta \|u\|^2 , \qquad (0.0.10)$$

for all $u \in \mathcal{H}$. If A is a bounded operator with a bounded inverse, then A maps any orthonormal basis to a Riesz basis. Moreover, all Riesz bases can be obtained as such images of an orthonormal basis. In a way, Riesz bases are the next best thing to an orthonormal basis. Note that the inequalities in (0.0.10) are not sufficient to guarantee that the e_n constitute a Riesz basis: the e_n also need to be linearly independent!

Contents