Dimensionality Reduction

Dimensionality Reduction

An Empirical Approach

Michael Kirby

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

Preface

Patterns may be found everywhere where there is not total disorder. In the simplest sense, a pattern may be viewed simply as one or more *relations* present in a signal or collection of signals. Patterns may extend spatially or temporally, or both. For example, the diurnal temperature variation of the earth at any given location is a temporal pattern. Clearly there are also temporal patterns occurring at longer time scales, i.e., seasonal variations. On the other hand, a snapshot of the temperature distribution of the entire earth is an example of a spatial pattern with warmer temperatures at the equator growing cooler towards the poles. The evolution of temperatures over the planet is an example of a complicated spatio-temporal pattern which is the subject of intense study.

To apply a mathematical procedure for analyzing a collection of patterns it is first necessary to obtain some measurement, or signal, which *quantifies* the information in some form. One of the primary sources of patterns representing physical phenomena is the high-speed computer. It is a major dilemma that these numerical simulations often produce more information than can be digested or understood. Furthermore, as computers continue to become faster, this situation only promises to become worse.

Hence, this text is intended for students and researchers whose work involves analyzing patterns in high-dimensional raw data sets. In particular, we are concerned with data the produced by the modeling of complicated physical phenomena. It has grown out of a course entitled "Pattern Analysis" which has been taught several times at Colorado State University and once at the

Technical University of Vienna. The typical audience has ranged from first year graduate students to faculty from a broad variety of disciplines including Computer Science, Physics, Engineering, Medicine, Atmospheric Sciences and Mathematics. It offers a selection of approaches, classical and contemporary, all of which are actively applied in the literature.

The main emphasis in the following pages is on developing an array of techniques for the low-dimensional representation of data. The somewhat unusual collection of topics was motivated by the idea that it is by comparison that the virtues and short-comings of any given technique can be fully understood and appreciated. It is the aim of the author to demonstrate that the methodologies presented, when taken as a whole, provide an even more powerful toolkit for extracting information out of data sets. Essentially every technique that we consider is an approach for decomposing such a signal in a manner which simplifies its study.

The methods are compared in terms of their primary mathematical characteristics. We shall take great efforts to distinguish methods in these terms. We will be primarily interested the properties of *global* and *local* representation. In addition, we characterize methods as linear or nonlinear and consider the problem of deciding which class of technique is most suitable for a given problem. Thus, one objective of this text is to encourage the reader to abandon the one tool for all problems approach and to replace it, hopefully, with some mathematical insight into the methodology selection procedure.

In the brief first Chapter the underlying concepts and the associated mathematical framework of dimensionality reduction are presented. Data sources such as numerical simulations of physical models, laboratory experiments and digital imaging systems are discussed as are general issues concerning the nature of data.

In Chapter 2 the mathematical background required for the later chapters is established. Coordinate transformations, change of bases, inner product spaces, subspace operations and finally the spectral theorem are presented. Even though some of the material is probably familiar to the reader it is hoped that this summary can be read, or returned to, with benefit, as it emphasizes the presentation within the context of applications to pattern analysis. Old questions may be asked in new ways, such as: Does a collection of digital images of human faces form a vector space?

In Chapter 3.1 we introduce one of the most important tools for dimensionality reduction, namely, the Karhunen-Loève (K-L) transform. We begin the discussion of an example from Pearson's 1901 paper. After a detailed theoretical presentation which develops the procedure for high-dimensional multivariate data sets, we present an application high-resolution image representation in the context of human face identification. The benefits of the exploitation of symmetry are also discussed and a methodology for computing symmetric eigenpictures is presented and applied. The main approaches developed i.e., the *snapshot* method and the *direct* method, are discussed in

terms of the singular value decomposition, This Chapter concludes with a local extension of the K-L procedure.

Chapter 4 continues the discussion of the K-L procedure with more applications. In particular we consider the problem of reducing the dimensionality of the dynamical description of physical models. We also present an application to the processing of sequences of time-dependent high-resolution digital images.

Part II begins with Chapter 5 which introduces one of the most important tools in applied mathematics and pattern analysis, i.e., the discrete Fourier transform. The presentation is done within the framework of finite orthogonal expansions and is shown to be a special case of an optimal expansion. The short-time Fourier transform is developed to show its poor time-frequency resolution properties and motivate the discussion of an adapted window transform, or wavelets. The continuous wavelet transform and the idea of time-scale analysis is presented.

Chapter 6 introduces the discrete wavelet transform and multi-resolution analysis. It begins with a presentation of the Haar wavelet and develops the multi-resolution idea in this setting. The pyramidal decomposition procedure is also presented and applications to signal analysis using several different types of wavelet are given.

Part III concerns biologically motivated computational paradigms, i.e., artificial neural networks. Chapter ?? is an introduction to the supervised approximation of non-linear functions. We focus primarily on feed-forward networks and the well-known back-propagation training procedure. The techniques in this Chapter are global in nature.

Chapter ?? presents more computational tools for function approximation and clustering. Local transfer functions for neural networks, i.e., radial basis functions, are introduced. The self-organizing feature map is presented as a clustering routine based on a competitive learning algorithm. The techniques in this Chapter are local in nature.

The last Chapter entitled *Hybridology* consists of several sections, each of which is a "multiple methodology" application. Hopefully, the pay-off of having developed so many disparate approaches is now apparent. We combine the ideas of optimal expansions, wavelets and neural networks by means of several examples. It is seen that unsupervised neural networks can *learn* the K-L basis. Constrainted nonlinear reduction networks provide an excellent new tool for pattern analysis. Radial basis functions can be implemented efficiently using the clustering techniques developed in Chapter ??. Neural networks can be implemented to construct data dependent wavelets.

Most of the algorithms presented require a computer for implementation and students have been required to complete an extensive application of one of the techniques on the computer. This approach allows students to become more familiar with a methodology than solving routine exercises would permit. Some suggestions for projects are included at the end of each Chapter.

The text is essentially self-contained with an extensive review of the elements of linear algebra which are relevant to the analysis of patterns. Some knowledge of Hilbert spaces is helpful for the Sections on the continuous Karhunen-Loève expansion and for the Chapter on Wavelets. It is possible, however, to investigate most of the basic ideas within the context of discrete data which makes a detailed development of Hilbert spaces less critical. While it is hoped that the presentation of any single methodology will be sufficient to get the reader started, extensive references to recent literature are included to aid in supplementing the material in the text.

MICHAEL KIRBY

Fort Collins

Acknowledgments

text...

Contents

Part I

Introduction

1

Pattern Analysis as Data Reduction

2

Linear Spaces and Transformations

Part II

Optimal Orthogonal Pattern Representations

3

The Karhunen-Loève (KL) Expansion

4

Additional Theory and Applications of the KL Expansion

Part III

Time, Frequency and Scale Analysis

5

Fourier Analysis

In this Chapter we will be concerned with representing data in terms of its frequency content. We will be interested in the analysis of signals with both continuous- and discrete-time representations. In addition, we will consider the transition from one representation to the other in terms of *Shannon's Sampling Theorem.* We will investigate continuous and discrete Fourier representations and develop the ideas required to apply transformations to signals in either the time- or frequency domain.

One of the characterizing features of Fourier analysis is its global nature. In particular, if the signal under investigation is changed locally in the time (frequency) domain then the frequency (time) domain representation is effected globally. We shall consider the pros and cons of this property and contrast this feature with the wavelet transform in Chapter 6. We begin our discussion by introducing some necessary classical definitions from the field of digital signal processing.

5.1 THE DISCRETE FOURIER TRANSFORM

We return to our approach of considering a pattern $\mathbf{x}$ as point in a vector space V where $\mathbf{x} = (x_1, \ldots, x_N)$. This can be expressed in almost a trivial way as a finite orthogonal expansion with respect to the standard basis vectors $\{\mathbf{e}_j\}_{j=1}^N$, i.e.,

$$\mathbf{x} = x_1\mathbf{e}_1 + \ldots x_N\mathbf{e}_N.$$

In this instance we view the decomposition of V as

$$V = W_1 \oplus \cdots \oplus W_N$$

where

$$W_i = \text{Span}(\mathbf{e}_i).$$

There is nothing special about this representation given that the expansion coefficients x_i do not reflect any specific features of the data. In addition, the coefficients may share information about the pattern and in fact they generally are redundant. For example, consider a digital image which has uniform pixel values. There may be thousands of these pixels but if we knew they all had the same value we would only need to store one number. This is a trivial example of information shared by coefficients. However this can be the case in much more subtle situations.

One of the main goals of the K-L analysis was to develop alternative subspace decompositions $V = W_1 \oplus \cdots \oplus W_N$ which package the information content of a pattern vector $\mathbf{x}$ in an *informative* and systematic way. The K-L procedure resulted in what we referred as an *empirical* basis since the eigenvectors were seen to be data dependent. In this section we will see that for special types of data the K-L basis is a collection of well known functions (sinusoids). We will be especially interested in the discrete case which will lead us to the discrete Fourier transform.

5.1.1 Finite Fourier Expansion

In this section we will be considering a complex inner product space $\mathbb{C}^N$ with inner product

$$(\mathbf{f}, \mathbf{g}) = \sum_{j=0}^{N-1} f_j \overline{g_j}$$

where $\overline{g_j}$ stands for the complex conjugate of g_j. We consider in detail the finite orthogonal expansion

$$\mathbf{f} = \sum_{j=0}^{N-1} \hat{f}_j \mathbf{v}^{(j)}$$

where the $\mathbf{v}^{(j)}$ are the Fourier vectors defined as

$$\mathbf{v}^{(j)} = (1, z^j, z^{2j}, z^{3j}, \ldots, z^{(N-1)j})$$

with $z = e^{2\pi i/N}$ and $i = \sqrt{-1}$. Using the relationship

$$z^m = e^{2\pi i m/N} = \cos(2\pi m/N) + i \sin(2\pi m/N)$$

we see that the z^k, $k = 0, \ldots, N-1$ are in fact the N-N'th roots of unity.

Proposition 5.1. *The set of N Fourier vectors $\{\mathbf{v}^{(j)}\}_{j=0}^{N-1}$ forms a basis for $\mathbb{C}^N$.*

Proof. It is enough to show that the Fourier vectors are orthogonal, i.e., $(\mathbf{v}^{(j)}, \mathbf{v}^{(k)}) = \alpha \delta_{jk}$ where α is some constant to be determined. First, we write the lth component of the jth Fourier vector as $v_l^{(j)} = z^{lj}$. Then

$$S = (\mathbf{v}^{(j)}, \mathbf{v}^{(k)}) = \sum_{l=0}^{N-1} z^{lj} \overline{z^{lk}} = \sum_{l=0}^{N-1} z^{l(j-k)}.$$

In other words, $S = \sum_{j=0}^{N-1} z^{lp}$ with $p = j - k$. This is just a geometric sum and can be computed using standard techniques, i.e.,

$$\begin{aligned} S &= 1 + z^p + z^{2p} + \cdots + z^{(N-1)p} \\ z^p S &= \quad z^p + z^{2p} + \cdots + z^{Np} \end{aligned}$$

Subtracting the second equation above from the first,

$$S(1 - z^p) = 1 - z^{Np}.$$

Using the fact $z^{Np} = 1$ we have the result

$$S(1 - z^p) = 0.$$

If $1 - z^p \neq 0$ then we must conclude $S = 0$. Thus S will be nonzero iff $z^p = 1$. This is true for $p = 0, \pm N, \pm 2N, \ldots$. Thus the inner product S is nonzero iff $j = k \pm mN$, where m is any integer. Note, in this case $S = \sum_{l=0}^{N-1} 1 = N$ so we can conclude

$$(\mathbf{v}^{(j)}, \mathbf{v}^{(k)}) = N\delta_{j,k+mN}.$$

□

The orthogonality of the Fourier vectors and the fact that we have N of them means we a constructed a basis for $\mathbb{C}^N$. Next we show how this basis can be used. In view of the above, let us now reconsider the expansion of a complex N-tuple in terms of Fourier vectors which we now write

$$\mathbf{f} = \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}_k \mathbf{v}^{(k)}.$$

To compute the value of the expansion coefficients we apply the orthogonality property of the Fourier vectors. Taking the inner product of $\mathbf{f}$ with $\mathbf{v}^{(j)}$ we have

$$\begin{aligned}(\mathbf{f}, \mathbf{v}^{(j)}) &= \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}_k (\mathbf{v}^{(k)}, \mathbf{v}^{(j)}) \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}_k N\delta_{j,k} \\ &= \hat{f}_j\end{aligned}$$

From which we have

$$\hat{f}_j = (\mathbf{f}, \mathbf{v}^{(j)})$$

which we interpret as the projection of the discrete data vector $\mathbf{f}$ onto the jth Fourier vector. Evaluating the complex inner product above we write

$$\hat{f}_j = \sum_{k=0}^{N-1} f_k \overline{v_k^{(j)}} = \sum_{k=0}^{N-1} f_k e^{-2\pi ijk/N}$$

We note that using this normalization for the Fourier expansion we have *Plancherel's relation*

$$(\mathbf{f}, \mathbf{g}) = \frac{1}{N}(\hat{\mathbf{f}}, \hat{\mathbf{g}})$$

and in particular

$$\|\mathbf{f}\| = \frac{1}{N}\|\hat{\mathbf{f}}\|,$$

a fact know as *Parseval's equality.* It says that the magnitude of the original vector is related to the magnitude of the Fourier transformed sequence by a fixed scalar multiple.

In summary we define the discrete Fourier transform pair:

Definition 5.1. *The discrete Fourier transform (DFT) of the sequence* $\{f_n\}$ *of length* N *is given by*

$$\hat{f}_j = \sum_{k=0}^{N-1} f_k e^{-2\pi ijk/N}. \tag{5.1}$$

where $j = 0, \ldots, N-1$. *The inverse transform is*

$$f_k = \frac{1}{N} \sum_{j=0}^{N-1} \hat{f}_j e^{2\pi ijk/N}. \tag{5.2}$$

with $k = 0, \ldots, N-1$.

The DFT is generally referred to as a global transform since each coefficient the transform space depends on all of the data in the periodic data sequence. Similarly, all the inverse coefficients are a function of all of the transformed coefficients. This situation is contrasted later with function expansions, i.e., wavelets, with have coefficients which have a local rather than global dependecy.
Sometimes we will find it convenient to use the notation

$$\mathcal{F}(f_k) = \hat{f}_j$$

and the inverse transform

$$\mathcal{F}^{-1}(\hat{f}_j) = f_k.$$

Note that the summation for both the forward and backward DFT is over a finite number of points. However, by examining the sums in equations (5.1-5.2) we observe that the quantities f_k and $\hat{f}_j$ are defined for all k and j in $\mathbb{Z}$.

5.1.2 Properties of the DFT

Property 5.1.1. *Periodicity. The Fourier transform and inverse Fourier transform of a sequence of length* N *is a periodic sequence of period* N, *i.e.,*

$$f_k = f_{k+ln} \qquad \hat{f}_j = \hat{f}_{j+ln}$$

This may be established directly from the definition in equation (5.2) by substituting the index $k + lN$, i.e.,

$$\begin{aligned} f_{k+ln} &= \frac{1}{N} \sum_{j=0}^{N-1} \hat{f}_j e^{2\pi ij_{k+ln}/N} \\ &= \frac{1}{N} \sum_{j=0}^{N-1} \hat{f}_j e^{2\pi ijk/N} e^{2\pi ijl} \\ &= f_k \end{aligned}$$

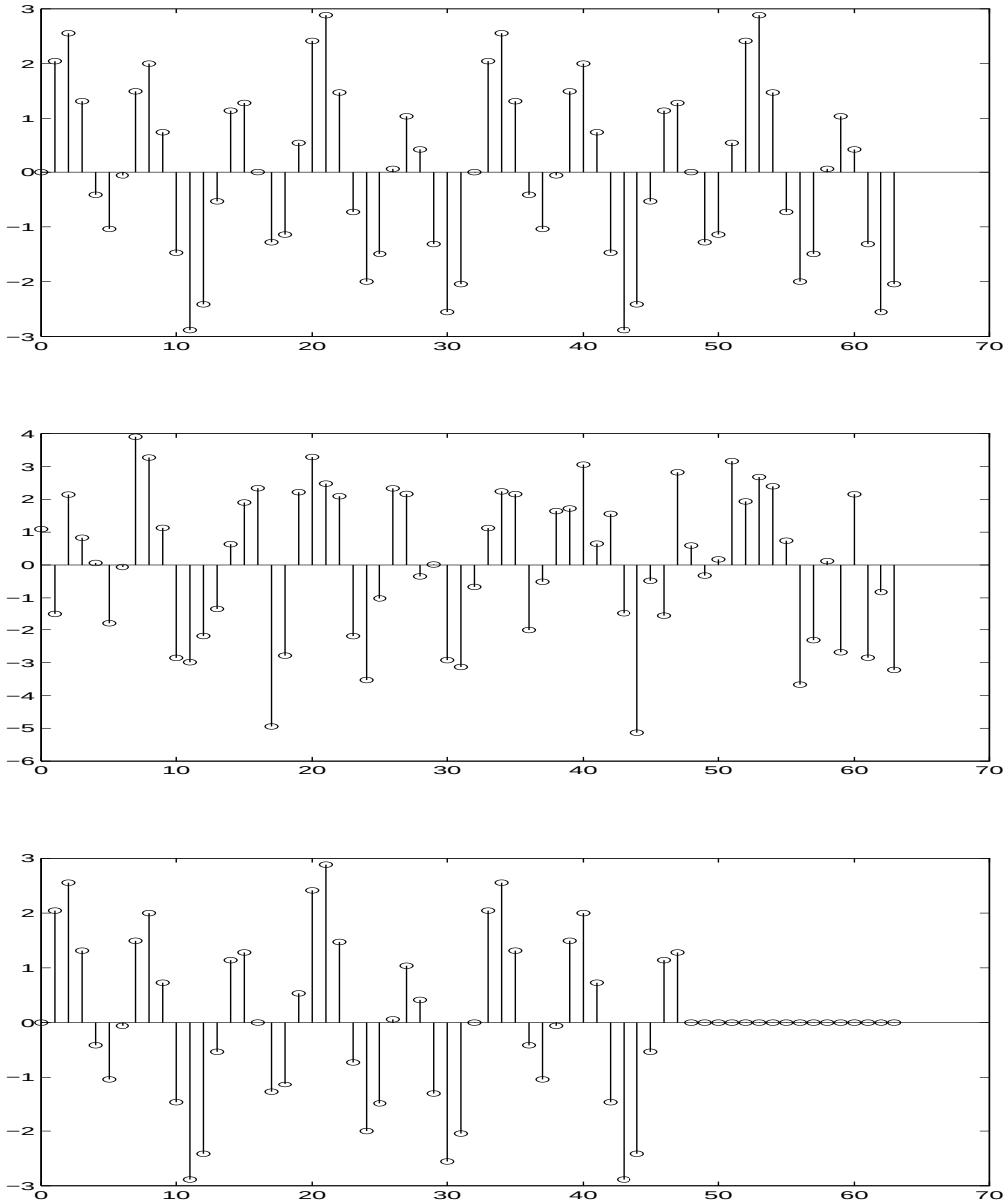

Fig. 5.1 The top figure is the discrete time function $x(n) = \sin(\pi n/8) + 2\sin(25\pi n/8)$ where $n = 0, \ldots, 63$. The middle figure is the same discrete time function as in the top figure, but with noise added. The bottom figure is the same as the top but with one quarter of the entries set to zero.

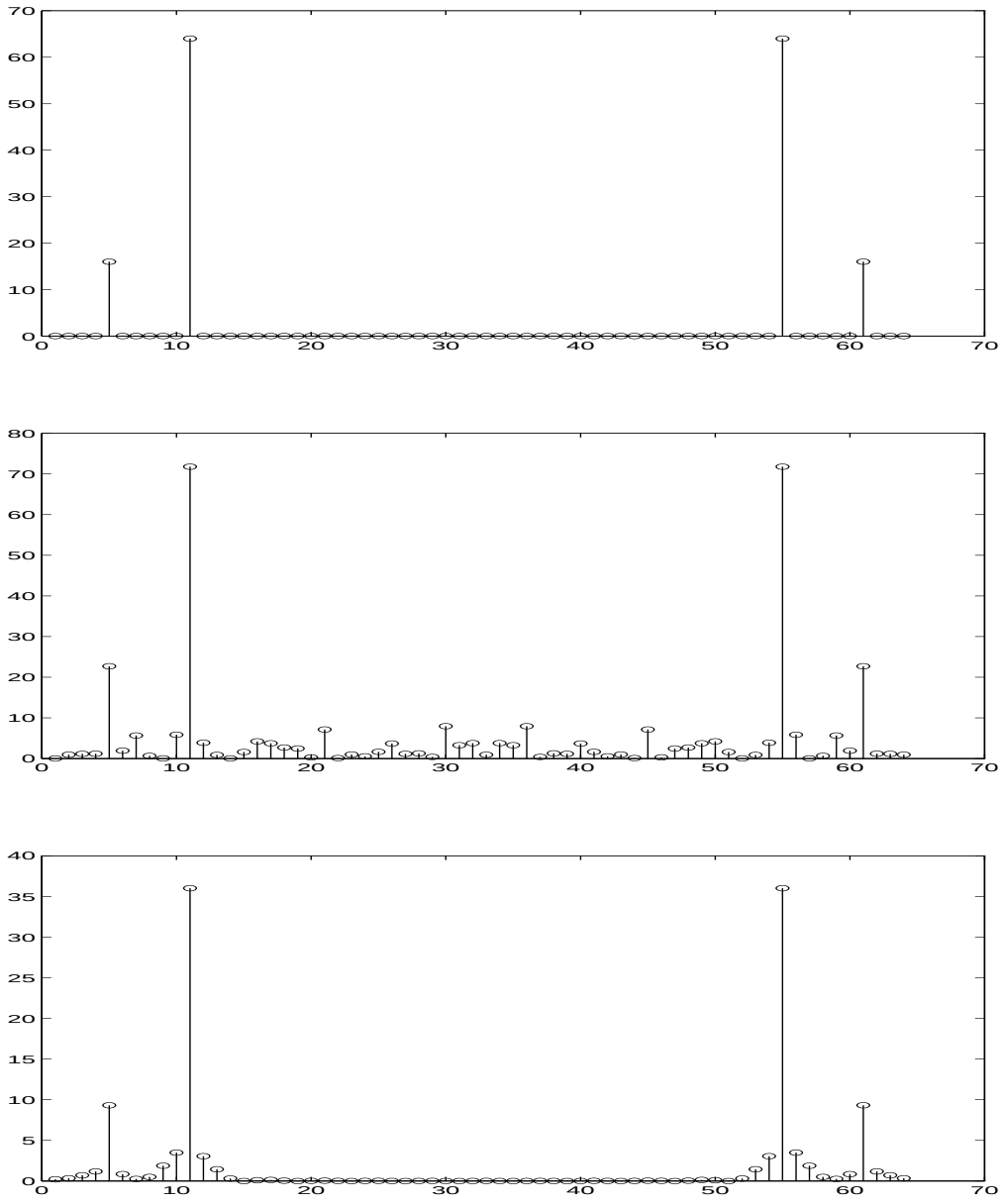

Fig. 5.2 This figure shows the discrete Fourier transforms corresponding to the discrete sequences displayed in Figure 5.1. The 2 frequencies appear as spikes in the top figure. The added noise contributes additional low amplitude frequencies as shown in the middle figure. The bottom transform shows the effect of replacing the tail of the 2 frequency data with zeros. The additional components are referred to as *leakage*.

Also, the transformed sequence may be seen to be periodic proceeding as above but using equation (5.1)

$$\begin{aligned}\hat{f}_{j+ln} &= \sum_{k=0}^{N-1} f_k e^{-2\pi i j+ln\, k/N} \\ &= \sum_{k=0}^{N-1} f_k e^{-2\pi ijk/N} e^{-2\pi ilk} \\ &= \hat{f}_j\end{aligned}$$

In fact, any finite sequence may be viewed as an infinite sequence by extending it periodically. This extension occurs automatically for sequences computed via a Fourier transform or its inverse. We will assume that by convention that the periodic extensions of the DFT are taken to be zero. We will refer to the periodically extended transform as the discrete Fourier series (DFS) of a finite sequence.

Property 5.1.2. *Shifting Properties*

$$\mathcal{F}(f_{n-m}) = z^{-mk}\hat{f}_k$$

Proof.

$$\begin{aligned}f_{n-m} &= \frac{1}{N}\sum_{k=0}^{N-1} \hat{f}_k e^{2\pi i(n-m)k/N} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} \hat{f}_k e^{2\pi ink/N} e^{-2\pi imk/N} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} (\hat{f}_k e^{-2\pi imk/N}) e^{2\pi ink/N}\end{aligned}$$

In other words,

$$\mathcal{F}(f_{n-m}) = z^{-mk}\hat{f}_k$$

□

In alternate notation,

$$\mathbf{f} = (f_0, \ldots, f_{N-1}) \leftrightarrow (\hat{f}_0, \ldots, \hat{f}_{N-1}) = \hat{\mathbf{f}}$$
$$\mathbf{f}^s = (f_{-m}, \ldots, f_{N-1-m}) \leftrightarrow (z^{-m}\hat{f}_0, \ldots, z^{-(N-1)m}\hat{f}_{N-1}) = \hat{\mathbf{f}}^s$$

Using the same approach it can also be shown that

$$\mathcal{F}(f_k z^{nk}) = \hat{f}_{j-n}.$$

Specifically,

$$\hat{\mathbf{f}} = (\hat{f}_0, \ldots, \hat{f}_{N-1}) \leftrightarrow (f_0, \ldots, f_{N-1}) = \mathbf{f}$$
$$(\hat{f}_{-m}, \ldots, \hat{f}_{N-1-m}) \leftrightarrow (z^m f_0, \ldots, z^{(N-1)m} f_{N-1})$$

Property 5.1.3. *The circularity property states*

$$\sum_{k=1}^{N} a_k e^{\pm 2\pi i j k/N} = \sum_{k=1+m}^{N+m} a_k e^{\pm 2\pi i j k/N}$$

where $\{a_k\}$ *is periodic sequence, for any integer* $m \in \mathbb{Z}$.

5.1.3 The DFT as an Optimal Basis

We saw in Chapter 2 that certain kinds of operators could be used to determine a complete set of basis vectors which spanned invariant subspaces. These subspaces collectively described the vector space and the basis allow us to decompose the operator into the superposition of operations on the subspaces. When have also seen in the section how Fourier vectors form a complete basis for $\mathbb{C}^N$. We might now ask, is there an operator from which this basis can be determined ? As you might suspect, the answer is yes.

Let C be a circulant matrix

$$C = \begin{pmatrix} c_0 & c_{N-1} & c_{N-2} & \cdots & c_1 \\ c_1 & c_0 & c_{N-1} & \cdots & c_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{N-1} & c_{N-2} & c_{N-3} & \cdots & c_0 \end{pmatrix} \tag{5.3}$$

Note that $(C)_{mn} = c_{m-n}$ where $c_n = c_{n+lN}$.

Proposition 5.2. *The eigenvectors of any circulant matrix* C *are the Fourier vectors, and the eigenvalues are the discrete Fourier coefficients* $\hat{c}_k$ *defined by the DFT of* $\mathbf{c} = (c_1, c_2, \ldots, c_N)$.

Proof. The N Fourier vectors are given by

$$\mathbf{v}^{(k)} = (1, z^k, \ldots, z^{(N-1)k})$$

As before, write $v_j^{(k)} = z^{jk}$ where $z^N = 1$. Then

$$\begin{aligned} (C\mathbf{v}^{(k)})_n &= \sum_{j=0}^{N-1} c_{nj} v_j^{(k)} \\ &= \sum_{j=0}^{N-1} c_{n-j} z^{jk} \\ &= z^{nk} \sum_{j=0}^{N-1} c_{n-j} z^{-k(n-j)}. \end{aligned}$$

Substituting the exponential notation for z,

$$\begin{aligned}(C\mathbf{v}^{(k)})_n &= e^{2\pi ink/N}\sum_{j=0}^{N-1} c_{n-j}e^{-2\pi ik(n-j)/N} \\ &= \lambda_k v_n^{(k)}\end{aligned}$$

where we have defined λ_k as

$$\lambda_k \equiv \sum_{j=0}^{N-1} c_{n-j}e^{-2\pi ik(n-j)/N}.$$

□

We leave it as an exercise to show that $\lambda_k = \mathcal{F}(\mathbf{c})$ where $\mathbf{c} = (c_1, c_2, \ldots, c_N)$.

5.1.4 The Convolution Theorem.

Let $\{a_n\}$ and $\{b_n\}$ be periodic sequences each with period N.

Definition 5.2. *The periodic convolution of these sequences is given by*

$$(a * b)_n = \sum_{m=0}^{N-1} a_{n-m}b_m.$$

Theorem 5.1. *The Convolution Theorem: The discrete Fourier transform of the periodic convolution is the product of the discrete Fourier transforms of the individual sequences.*

Proof.

$$\begin{aligned}\mathcal{F}(a * b) &= \sum_{k=0}^{N-1}(a * b)_k e^{-2\pi ijk/N} \\ &= \sum_{k=0}^{N-1}(\sum_{m=0}^{N-1} a_{k-m}b_m)e^{-2\pi ijk/N} \\ &= \sum_{m=0}^{N-1} b_m e^{-2\pi ijm/N}\sum_{k=0}^{N-1} a_{k-m}e^{-2\pi ij(k-m)/N}.\end{aligned}$$

Using the circularity property (5.1.3) We rewrite the above sum, shifting by m, viz.:

$$\sum_{k=0}^{N-1} a_{k-m}e^{-2\pi ij(k-m)/N} = \sum_{k=m}^{N-1+m} a_{k-m}e^{-2\pi ij(k-m)/N} = \sum_{l=0}^{N-1} a_l e^{-2\pi ijl/N}$$

where in the last sum we have put $l = k - m$. So,

$$\mathcal{F}(a * b) = \sum_{m=0}^{N-1} b_m e^{-2\pi ijm/N} \sum_{l=0}^{N-1} a_l e^{-2\pi ijl/N}$$

$$= \hat{b}_j \hat{a}_j$$

□

The discrete Fourier transform of a function, even a real function, is generally complex and we write it

$$\hat{f}_j = \hat{r}_j + i\hat{s}_j$$

i.e., $\hat{r}_j = \mathrm{Re}(\hat{f}_j)$ and $\hat{s}_j = \mathrm{Im}(\hat{f}_j)$.

It will also be useful to express $\hat{f}_j$ in complex exponential form:

$$\hat{f}_j = |\hat{f}_j| e^{i\phi_j}$$

where the Fourier spectrum is given by

$$|\hat{f}_j| = ((\hat{r}_j)^2 + (\hat{s}_j)^2)^{1/2} = \hat{f}_j \overline{\hat{f}_j}$$

and

$$\phi_j = \tan^{-1}(\frac{\hat{s}_j}{\hat{r}_j}).$$

Definition 5.3. *The sequence $P_j = |\hat{f}_j|^2$ is called the* **discrete power spectrum** *of f_j and ϕ_j is the phase angle.*

Proposition 5.3. *The power spectrum is periodic.*

5.1.4.1 **Real Sequences** If the sequences in question are made up of real numbers instead of complex numbers we may make the observation

$$f_n = \overline{f_n}$$

which is often referred to as the *reality* condition.

Proposition 5.4. *If f_k is a real sequence then*

$$\hat{f}_j = \overline{\hat{f}_{-j}}.$$

Proof.

$$\hat{f}_j = \sum_{k=0}^{N-1} f_k e^{-2\pi ijk/N}.$$

$$\overline{\hat{f}_{-j}} = \sum_{k=0}^{N-1} \overline{f_k} e^{-2\pi ijk/N}.$$

but since $f_k = \overline{f_k}$ due to the reality of the sequence we conclude $\hat{f}_j = \overline{\hat{f}_{-j}}$. □

5.1.4.2 Even and odd transforms.

Definition 5.4. *A sequence* $\{f_n\}$ *is said to be* **even** *if*

$$f_n = f_{-n}$$

and **odd** *if*

$$f_n = -f_{-n}.$$

The DFT of a sequence can be simplified if the discrete periodic sequence possesses either even or odd symmetry. If $\{f_n\}$ is an even sequence then $f_n = (f_n + f_{-n})/2$. Using this relation combined with the definition of the DFT

$$f_k = \frac{1}{N}\sum_{j=0}^{N-1} \hat{f}_j e^{2\pi ijk/N}$$

from which it follows

$$f_{-k} = \frac{1}{N}\sum_{j=0}^{N-1} \hat{f}_j e^{-2\pi ijk/N}$$

thus,

$$f_k + f_{-k} = \frac{1}{N}\sum_{j=0}^{N-1} \hat{f}_j (e^{2\pi ijk/N} + e^{-2\pi ijk/N}).$$

Now using $f_{ek} = (f_k + f_{-k})/2$ we get

$$f_{ek} = \frac{1}{N}\sum_{j=0}^{N-1} \hat{f}_j \cos(\frac{2\pi jk}{N})$$

Similarly, if $\{f_{on}\}$ is an odd sequence one can show

$$f_{ok} = \frac{i}{N}\sum_{j=0}^{N-1} \hat{f}_j \sin(\frac{2\pi jk}{N})$$

For real sequences one can show

$$f_n = \sum_{k=0}^{N-1} a_k \cos(\frac{2\pi nk}{N}) + b_k \sin(\frac{2\pi nk}{N})$$

Proposition 5.5. *The power spectrum of a real sequence is even.*

$$P_j = \hat{f}_j \overline{\hat{f}_j}$$

$$P(-j) = \hat{f}(-j)\overline{\hat{f}(-j)}$$

Taking the complex conjugate of the reality condition $\hat{f}_j = \overline{\hat{f}(-j)}$ gives $\overline{\hat{f}_j} = \hat{f}(-j)$ in other words,

$$P_j = \overline{\hat{f}(-j)}\hat{f}(-j) = P(-j).$$

5.1.5 Fourier Descriptors.

The discrete Fourier transform can be used for describing the shape of a boundary on a quantitative basis. Consider a closed curve in the plane which is given (either initially, or after sampling) as a set of discrete (x_j, y_j) pairs. Consider repackaging the data as

$$(x_j, y_j) \to x_j + iy_j = f_j$$

Now we have created a complex sequence of points from the closed planar curve. When we take the DFT we may look at the curve in terms of its frequency domain representation. This removes dependence on position, size, and orientation.

The results of an experiment with filtering the Fourier transform of the Fourier descriptors corresponding to a sampling of points on the closed curve representing the letter C (or U!) is shown in Figure 5.1.5. Since the data is not translationally invariant, the DFT is not an optimal transform in the energy capturing sense. Also, we note that the energies corresponding to each coefficient, as displayed in the the power spectrum, are not monotonically decreasing so it is not advisable to simply truncate the tail of this expansion. As we see when the 15th coefficient is set to zero, it is possible to throw away potentially valuable information.

5.1.6 The 2-D DFT

Transforms such as the DFT can be used for image coding as well as signal analysis. It might seem at first glance that all we need do to apply the DFT to images is to make a big vector out of the array of pixels. This would work, but we would be failing to exploit the natural extension of the DFT to higher dimensions which can be computationally more efficient.

First, we briefly consider the application of general transforms to higher dimensional arrays. In what follows let f_{ij} be a rectangular array with $i = 0, \ldots, M-1$ and $j = 0, \ldots, N-1$.

Definition 5.5. *The* **forward transformation** *of a* $2-$*dimensional array* f_{ij} *is given by*

$$\hat{f}_{mn} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f_{ij} s_{ij,mn}$$

and the **inverse transformation** *by*

$$f_{ij} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \hat{f}_{mn} t_{ij,mn}$$

The four-dimensional arrays $s_{ij,mn}$ and $t_{ij,mn}$ are referred to as the **forward** and **inverse** transformation kernels.

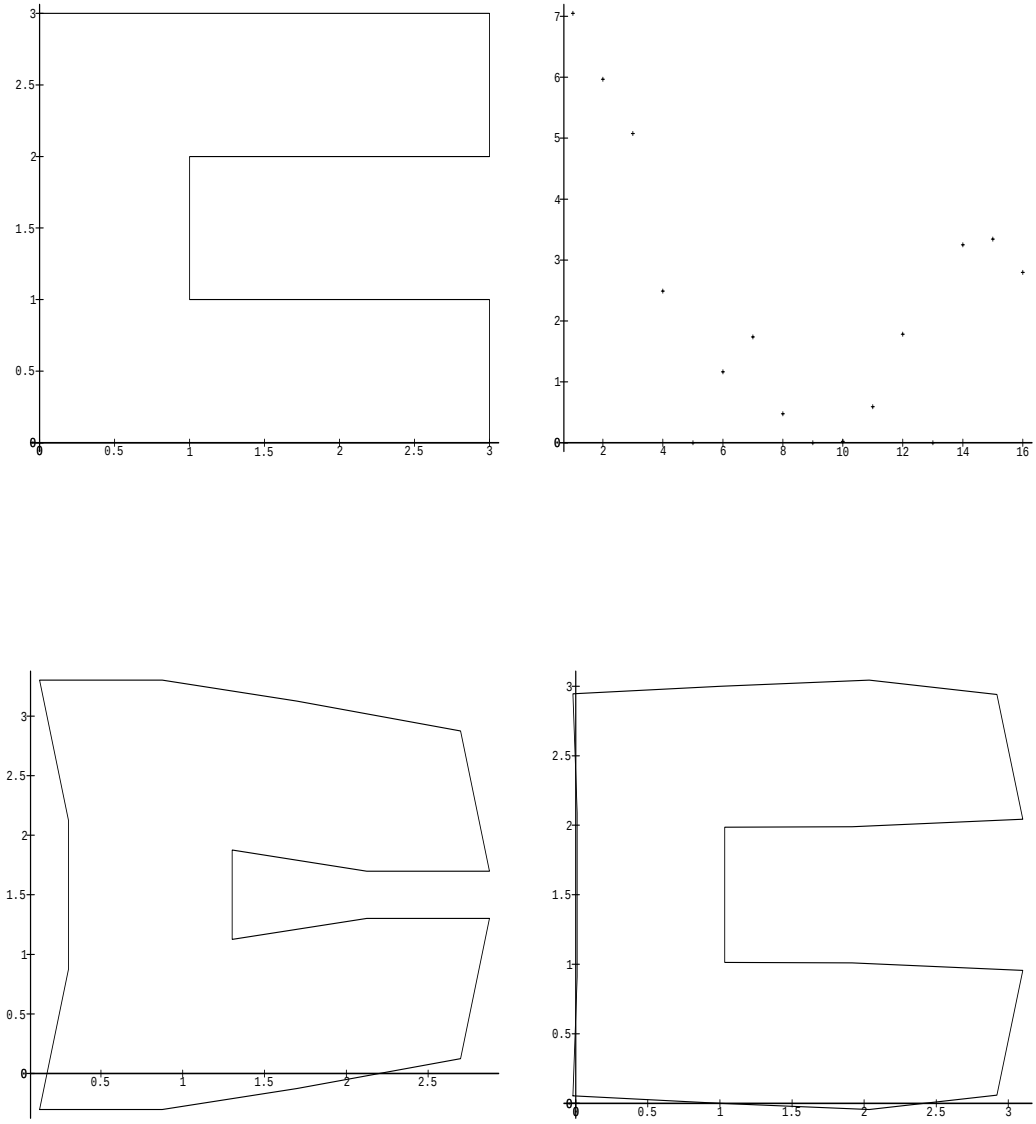

Fig. 5.3 Top left: The letter C before it is discretely sampled. Top right: The power spectrum of the Fourier descriptors of the letter C. Bottom left: The inverse transform of the DFT of the Fourier descriptors, with the 15'th coefficient (in transform space) set to zero. Bottom right: The inverse transform of the DFT of the Fourier descriptors, with the low energy 8,9,10 and 11 coefficients set to zero.

In general, a kernel $k_{ij,mn}$ is said to be **separable** if

$$k_{ij,mn} = k^1_{im} k^2_{jn}$$

and **symmetric** if it is separable and

$$k^1_{ij} = k^2_{ij}.$$

We observe that a 2-dimensional transform with a separable kernel can be computed via repeated applications of the one-dimensional version of the transform.

The transform, with s separable, is

$$\hat{f}_{mn} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f_{ij} s^1_{im} s^2_{jn}$$

$$= \sum_{i=0}^{N-1} s^1_{im} (\sum_{j=0}^{N-1} f_{ij} s^2_{jn}).$$

Definition 5.6. *The discrete Fourier transform (DFT) pair in 2-D is given by*

$$f_{jk} = \frac{1}{MN} \sum_{j'=0}^{M-1} \sum_{k'=0}^{N-1} \hat{f}_{j'k'} e^{2\pi i (jj'/M + kk'/N)}$$

$$\hat{f}_{j'k'} = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f_{jk} e^{-2\pi i (jj'/M + kk'/N)}.$$

Proposition 5.6. *The 2-D DFT is separable.*

As mentioned previously, this separability property means that the application of a 2D DFT is mathematically equivalent to computation of M DFTs with variable length N, or N DFTs with variable length M.

To see this consider rewriting

$$\hat{f}_{j'k'} = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f_{jk} e^{-2\pi i (jj'/M + kk'/N)}$$

as

$$\hat{f}_{j'k'} = \sum_{j=0}^{M-1} e^{-2\pi i jj'/M} \sum_{k=0}^{N-1} f_{jk} e^{-2\pi i kk'/N}$$

$$\hat{f}_{j'k'} = \sum_{j=0}^{M-1} e^{-2\pi i jj'/M} g_{jk'}.$$

Definition 5.7. *The* **power spectrum** *of a 2-dimensional DFT is given by*

$$P_{jk} = |\hat{f}_{jk}|^2$$

As in the 1D case, we may express the transform in polar form

$$\hat{f}_{jk} = |\hat{f}_{jk}|e^{i\theta_{jk}}.$$

Property 5.1.4. *The 2D transform is also periodic, i.e.,*

$$f_{jk} = f_{j+nM,k+nN}$$

and

$$\hat{f}_{jk} = \hat{f}_{j+nM,k+nN}.$$

Property 5.1.5. *In analogue with the 1D case we have the 2D shifting property*

$$\mathcal{F}(f_{j-m,k-n}) = \hat{f}_{j'k'}e^{-2\pi i(j'm/M+k'n/N)}.$$

5.1.7 The Fast Fourier Transform

Symmetry and periodicity may be exploited to accelerate the computation of the DFT. We consider the case that the length of the sequence to be transformed is of the form $N = 2^m$ where m is some positive integer.

The fast algorithm proceeds by splitting the sequence $f(n)$ into 2 $N/2$ point subsequences made up of the odd and even indices of $\mathbf{f}$. To see why this is helpful let us write again the DFT of $\mathbf{f}$ as

$$\hat{f}_k = \sum_{n=0}^{N-1} f_n W_N^{nk} \tag{5.4}$$

where $k = 0, \ldots, N-1$, $W_N^{nk} = exp(-2\pi ikn/N)$. This sum may now be split into two sums over the even and odd indices, i.e.,

$$\hat{f}_k = \sum_{r=0}^{\frac{N}{2}-1} f_{2r} W_N^{2rk} + \sum_{r=0}^{\frac{N}{2}-1} f_{2r+1} W_N^{(2r+1)k} \tag{5.5}$$

$$= \sum_{r=0}^{\frac{N}{2}-1} f_{2r} (W_N^2)^{rk} + W_N^k \sum_{r=0}^{\frac{N}{2}-1} f_{2r+1} (W_N^2)^{rk} \tag{5.6}$$

$$= \sum_{r=0}^{\frac{N}{2}-1} f_{2r} W_{N/2}^{rk} + W_N^k \sum_{r=0}^{\frac{N}{2}-1} f_{2r+1} W_{N/2}^{rk} \tag{5.7}$$

$$= \hat{g}_k + W_N^k \hat{h}_k \tag{5.8}$$

where we have used the fact that

$$W_N^2 = W_{N/2}$$

and defined

$$\hat{g}_k = \sum_{r=0}^{\frac{N}{2}-1} f_{2r} W_{N/2}^{rk}$$

and

$$\hat{h}_k = \sum_{r=0}^{\frac{N}{2}-1} f_{2r} W_{N/2}^{rk} + W_N^k \sum_{r=0}^{\frac{N}{2}-1} f_{2r+1} W_{N/2}^{rk}.$$

Problems

5.1 The (symmetrical) discrete Fourier transform (DFT) pair is given by

$$f_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \hat{f}_j e^{2\pi i j k/N}$$

$$\hat{f}_j = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} f_k e^{-2\pi i j k/N}.$$

Show that

a) this definition follows from equation (5.1) using the appropriate substitutions.

b) $\|\mathbf{f}\| = \|\hat{\mathbf{f}}\|$ (assume the standard Euclidean norm).

5.2 Find the discrete Fourier transforms of $\mathbf{f} = (-1, 0, 1)$ and $\mathbf{g} = (-1, 0, 1, 0)$ and compare.

5.3 Let f_n be an odd periodic sequence with integer period N. Namely,

$$f_n = f_{n+jN}$$

$$f_n = -f_{-n}$$

where j is any integer. Simplify the DFST for this case. What more can you say if the sequence is real?

5.4 Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$. We define the finite convolution product as

$$(\mathbf{u} * \mathbf{v})_n = \sum_{m=0}^{N-1} u_{n-m} v_m.$$

Prove

$$(\mathbf{u} * \mathbf{v})_n = (\mathbf{v} * \mathbf{u})_n.$$

5.5 Let $\mathbf{f} \in \mathbb{C}^N$ and let $\hat{\mathbf{f}}$ be the DFT of $\mathbf{f}$. If we write the DFT and its inverse as matrix operations, i.e.,

$$\mathbf{f} = \mathbf{A}\hat{\mathbf{f}}$$

and

$$\hat{\mathbf{f}} = \mathbf{B}\mathbf{f}$$

what are $\mathbf{A}$ and $\mathbf{B}$?

5.6 Represent the FFT of a 4-point sequence in terms of matrix multiplication and compare with the DFT.

5.7 Let $\mathbf{C}$ be an $N \times N$ circulant matrix. Show that the eigenvalues of $\mathbf{C}$ are given by the DFT of $\mathbf{c}$ with $\mathbf{c} = (c_1, c_2, \ldots, c_N)$.

5.8 Assume $\mathbf{f} \in \mathbb{R}^N$, i.e., it is a real vector. Starting from the expansion

$$f_k = \frac{1}{N} \sum_{j=0}^{N-1} \hat{f}_j e^{2\pi i j k/N}$$

and using the reality of the sequence, show that expansion can be reformulated as

$$f_n = \sum_{k=0}^{N-1} a_k \cos(\frac{2\pi n k}{N}) + b_k \sin(\frac{2\pi n k}{N})$$

where the a_k, b_k are real. Find a formula for the a_k and b_k in terms of $\hat{f}_k$.

5.9 As we have shown, N- Fourier vectors span an N-dimensional complex vector space. To help in visualizing the expansions better, draw the complex and real parts of the eight Fourier vectors $\{\mathbf{v}^{(j)}\}_{j=1}^{8}$. Also draw the eight Fourier vectors which span $\mathbb{R}^8$ (noting which of these are odd and which are even.)

5.10 Show the following:

a) if $\hat{f}_j$ is a real sequence, then $f_k = \overline{f_{-k}}$

b) if f_k is an imaginary sequence, then $\hat{f}_k = -\overline{\hat{f}_{-k}}$

c) if $\hat{f}_j$ is an imaginary sequence, then $f_k = -\overline{f_{-k}}$

5.11 Consider the complex vector $\mathbf{f} = (f_0, \ldots, f_{N-1}) \in \mathbb{C}^N$ where $\mathbb{C}^N$ denotes the N-dimensional complex vector space. Define the shifted vector, $\mathbf{f}^s = (f_{-m}, \ldots, f_{N-1-m})$ and assume that the vectors may be extended periodically so that $f_m = f_{m+nN}, f_m^s = f_{m+nN}^s$ where $n = 0, \pm 1, \pm 2, \ldots$. Show that the power spectra of $\mathbf{f}$ and $\mathbf{f}^s$ are the same.

5.12 If the f_{jk} are real numbers show that

$$\hat{f}_{jk} = \overline{\hat{f}_{-j,-k}}.$$

From this conclude that

$$P_{jk} = P_{-j,-k}.$$

5.13 Show that the rotation of an image leaves the power spectrum invariant.

5.14 Prove the circularity property, i.e., show

$$\sum_{k=1}^{N} a_k e^{\pm 2\pi i j k/N} = \sum_{k=1+m}^{N+m} a_k e^{\pm 2\pi i j k/N}$$

for any integer m where $\{a_n\}$ a periodic sequence with integer period N.

5.15 Show that

$$f_{jk} = f_{j+lM,k+lN}$$

and

$$\hat{f}_{jk} = \hat{f}_{j+lM,k+lN}$$

where $l \in \mathbb{Z}$

5.16 Show that

$$f_{j-m,k-n} \leftrightarrow \hat{f}_{j'k'} e^{-2\pi i(j'm/M+k'n/N)}$$

What can you conclude about the 2-dimensional power spectrum of a shifted image?

5.17 Given that two complex vectors $\mathbf{f}, \mathbf{s} \in \mathbb{C}^N$ have the same power spectrum must there exist parameters θ, m such that

$$\hat{\mathbf{s}} = e^{i\theta} S^m \hat{\mathbf{f}}?$$

If not, can you specify conditions which would make the above statement true?

5.2 CONTINUOUS-TIME FOURIER TRANSFORM

In this section we assume that the domain of the function $x(t)$ to be transformed is continuous and furthermore, that the range is at least piecewise continuous, so both $x, t \in \mathbb{R}$. This is in contrast to a discrete-time system where $t \in \mathbb{Z}$ as considered in the previous section. The fully continuous analogue to the DFT is given by the Fourier integral pair which consists of the forward transform

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt \tag{5.9}$$

and the inverse transform

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega \tag{5.10}$$

In these equations we generally think of t as representing time and ω as representing the continuous frequency. Hence we will often refer to this transform pair as the continuous-time Fourier transform (CTFT). We will also employ the alternative notations

$$X(\omega) = \mathcal{F}(x(t))$$

and

$$x(t) = \mathcal{F}^{-1}(X(\omega))$$

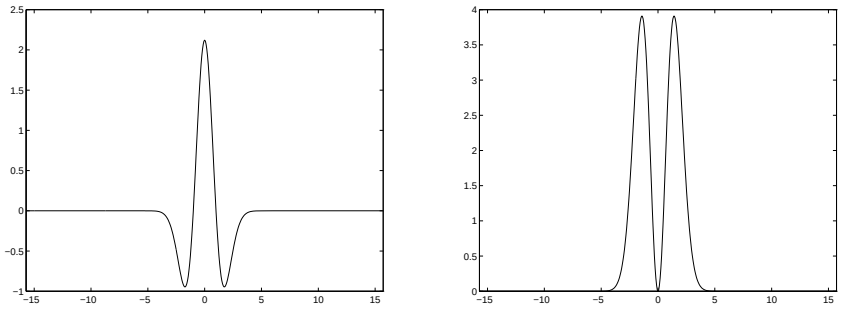

Fig. 5.4 Left: The Mexican hat function. Right: The CTFT of the Mexican hat.

and often denote a transform pair as

$$x(t) \leftrightarrow X(\omega)$$

Example 5.1. Compute the Fourier transform of the function

$$h_L(t) = \begin{cases} 1 & |t| \leq T \\ 0 & T < |t| \end{cases} \tag{5.11}$$

Applying equation (5.9) we obtain

$$\begin{aligned} H_L(\omega) &= \int_{-\infty}^{\infty} h_L(t) e^{-i\omega t} dt \\ &= \int_{-T}^{T} e^{-i\omega t} dt \\ &= \frac{2\sin(\omega T)}{\omega} \end{aligned}$$

The subscript L is employed because, as will be discussed later, this function acts as a low-pass filter.

Example 5.2. The Mexican hat function

$$x(t) = \frac{3}{\sqrt{2}}(1 - t^2)e^{-t^2/2} \tag{5.12}$$

has the Fourier transform

$$X(\omega) = 3\sqrt{\pi}\omega^2 e^{-\omega^2/2} \tag{5.13}$$

The function and its transform are shown in Figure 5.2.

Example 5.3. The Morlet *mother wavelet*

$$x(t) = \frac{1}{\sqrt{2\pi}} e^{-i\omega_0 t} e^{-t^2/2} \tag{5.14}$$

has the Fourier transform

$$X(\omega) = e^{-\frac{1}{2}(\omega-\omega)^2} \tag{5.15}$$

The function and its transform are shown in Figure 5.2 with $\omega_0 = \pi\sqrt{2/\ln 2}$.

Example 5.4. The Fourier transform of a Gaussian

$$x(t) = e^{-\alpha t^2} \tag{5.16}$$

is given by

$$X(\omega) = \sqrt{\frac{\pi}{\alpha}} e^{-\omega^2/4\alpha} \tag{5.17}$$

The fact that the Fourier transform of a Gaussian is also a Gaussian is a unique property. Figure 5.2.3 shows the Gaussian and its CTFT for various values of α.

5.2.1 Connection With the Fourier Series

One motivation for this transform pair is the extension of the Fourier series representation of a periodic function to a function which is not periodic, or whose period may be viewed as infinitely long. Hence, we first consider the representation of a function $x(t)$ which is assumed to have a fintite period T. The Fourier series representation for this is

$$x(t) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n t/T} \tag{5.18}$$

where the Fourier coefficients are given by

$$a_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-2\pi i n t/T} dt. \tag{5.19}$$

Now examine the effect of letting the period $T \to \infty$. First define

$$\Delta\omega = \frac{2\pi}{T}.$$

Multiplying by the period T and taking the limit of equation (5.19) as $T \to \infty$ we obtain

$$\lim_{T\to\infty} T a_n = \int_{-\infty}^{\infty} x(t) e^{-in\Delta\omega t} dt$$

From equation (5.9) we then conclude

$$\lim_{T\to\infty} T a_n = X(n\Delta\omega).$$

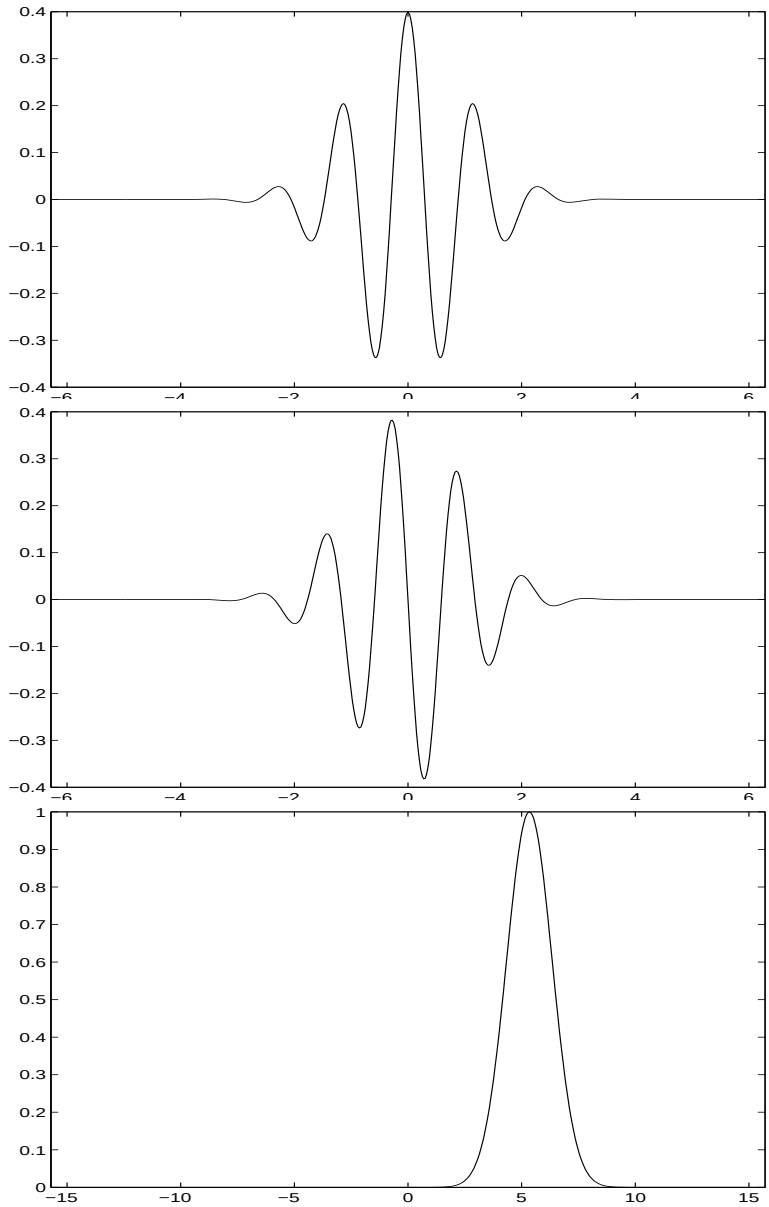

Fig. 5.5 Top: The real part of the Morlet mother wavelet. Middle: The imaginary part of the Morlet mother wavelet. Bottom: The CTFT of the Morlet mother wavelet.

Also, computing the limit of equation (5.18) and inserting the result from above

$$\begin{aligned} x(t) &= \lim_{T\to\infty} \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} T a_n e^{in\Delta\omega t} \Delta\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega \end{aligned}$$

5.2.2 Properties of the CTFT

5.2.2.1 Real-valued functions If $x(t)$ is a real valued function then its Fourier transform satisfies the equation

$$X(\omega) = \overline{X}(-\omega) \tag{5.20}$$

This is analogous to the reality condition for the DFT.

5.2.2.2 Even and Odd Functions If $x(t)$ is an even function then the forward transform is

$$X(\omega) = 2 \int_0^{\infty} x(t) \cos(\omega t) dt$$

and the inverse transform is

$$x(t) = \frac{1}{\pi} \int_0^{\infty} X(\omega) \cos(\omega t) d\omega$$

If $x(t)$ is odd then the forward transform is

$$X(\omega) = -2i \int_0^{\infty} x(t) \sin(\omega t) dt$$

and the inverse transform is

$$x(t) = -\frac{1}{\pi} \int_0^{\infty} X(\omega) \sin(\omega t) d\omega$$

5.2.2.3 Shifting Properties The shifting and scaling properties of the Fourier transform are frequently used in Fourier analysis and are analogous to the properties of the discrete Fourier transform.

- Time-Shifting property:

$$x(t-\alpha) \leftrightarrow e^{-i\alpha\omega} X(\omega). \tag{5.21}$$

To show this write

$$\begin{aligned} \mathcal{F}(x(t-\alpha)) &= \int_{-\infty}^{\infty} x(t-\alpha) e^{-i\omega t} dt \\ &= e^{-i\alpha\omega} \int_{-\infty}^{\infty} x(t-\alpha) e^{-i(t-\alpha)\omega} d(t-\alpha) \\ &= e^{-i\alpha\omega} X(\omega) \end{aligned}$$

- Frequency-Shifting property:

$$e^{i\omega_0 t}x(t) \leftrightarrow X(\omega - \omega_0) \tag{5.22}$$

5.2.2.4 Scaling Property The scaling property states that

$$x(\alpha t) \leftrightarrow \frac{1}{|\alpha|}X(\frac{\omega}{\alpha}) \tag{5.23}$$

This property is also referred to as the time-dilation property. First consider the case $\alpha > 0$. Then

$$\mathcal{F}(x(\alpha t)) = \int_{-\infty}^{\infty} x(\alpha t)e^{-i\omega t}dt$$

letting $\tau = \alpha t$ we have

$$\begin{aligned} &= \int_{-\infty}^{\infty} x(\tau)e^{-i\omega\tau/\alpha}\frac{d\tau}{\alpha} \\ &= \frac{1}{\alpha}X(\frac{\omega}{\alpha}) \end{aligned}$$

A similar argument applies for $\alpha < 0$ where care must be taken to appropriately reverse the limits of integration.

5.2.2.5 Transform of a Derivative This property may be stated both in the time and frequency domain.

- Time-domain:

$$\int_{-\infty}^{\infty} x'(t)e^{-i\omega t}dt = e^{-i\omega t}x(t)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} x(t)(-i\omega t)e^{-i\omega t}dt$$

hence

$$\mathcal{F}(x'(t)) = i\omega X(\omega) \tag{5.24}$$

where we have assumed that the function to be transformed vanishes at ∞. For the n'th derivative we have the formula

$$\frac{d^n x}{dt^n} \leftrightarrow (i\omega)^n X(\omega)$$

- Frequency-domain:

$$(-it)^n x(t) \leftrightarrow \frac{d^n X(\omega)}{d\omega^n} \tag{5.25}$$

5.2.2.6 Parseval's Formula This important formula establishes the relationship between the magnitude of a signal and the magnitude of its transform as

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \tag{5.26}$$

5.2.2.7 Time-Frequency Symmetry The following property reveals the symmetry between the time and frequency domains in the CTFT pair.

Theorem 5.2. *If*

$$x(t) \leftrightarrow X(\omega),$$

then

$$X(t) \leftrightarrow 2\pi x(-\omega).$$

proof.

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega$$
$$2\pi x(-t) = \int_{-\infty}^{\infty} X(\omega) e^{-i\omega t} d\omega$$

switching t and ω

$$2\pi x(-\omega) = \int_{-\infty}^{\infty} X(t) e^{-i\omega t} dt$$

Thus, the Fourier transform of the function $X(t)$ is $2\pi x(-\omega)$.

Example 5.5. To demonstrate the utility if this theorem, we consider its application to a previous example

$$h_L(t) \leftrightarrow H_L(\omega) = \frac{2\sin(\omega T)}{\omega}$$

thus,

$$2\pi h_L(-\omega) \leftrightarrow H_L(t)$$

from which we conclude

$$2\pi h_L(-\omega) = \begin{cases} 2\pi & |\omega| \leq T \\ 0 & T < |\omega| \end{cases} \leftrightarrow \frac{2\sin(tT)}{t} = H_L(t) \tag{5.27}$$

or equivalently,

$$\begin{cases} 1 & |\omega| \leq T \\ 0 & T < |\omega| \end{cases} \leftrightarrow \frac{\sin(tT)}{\pi t} \tag{5.28}$$

5.2.2.8 Convolution The definition of the continuous-time convolution is

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau$$

which we also write

$$y(t) = x(t) * h(t) \tag{5.29}$$

The following convolution theorems are especially important:

Theorem 5.3. *Let the functions $y(t), x_1(t)$ and $x_2(t)$ have the Fourier transforms $Y(\omega), X_1(\omega)$ and $X_2(\omega)$, respectively.*

- *Time convolution: If*

$$y(t) = x_1(t) * x_2(t) \tag{5.30}$$

then

$$Y(\omega) = X_1(\omega)X_2(\omega). \tag{5.31}$$

- *Frequency convolution: If*

$$y(t) = x_1(t)x_2(t) \tag{5.32}$$

then

$$Y(\omega) = \frac{1}{2\pi}X_1(\omega) * X_2(\omega) \tag{5.33}$$

5.2.3 The Uncertainty Principle

It is a property of the Fourier transform of a function that the wider the domain on which the function is non-zero the narrower the width of the transform of the function. Similarly, functions which are narrow in the time-domain have correspondingly wider transforms. This tradeoff is quantified by the *Uncertainty Principle*. To begin, we need a suitable definition for signal width, or duration. Following Papoulis [61] we define the time-duration Δ_t of a signal as

$$\Delta_t^2 \equiv \frac{\int_{-\infty}^{\infty} t^2|x(t)|^2dt}{\int_{-\infty}^{\infty}|x(t)|^2dt} \tag{5.34}$$

where the denominator is referred to as the *energy* of the function. Similarly, the frequency-duration, or width, is defined as

$$\Delta_\omega^2 \equiv \frac{\int_{-\infty}^{\infty} \omega^2|X(\omega)|^2d\omega}{\int_{-\infty}^{\infty}|X(\omega)|^2d\omega}. \tag{5.35}$$

Now we may state more precisely this relationship between the widths of the Fourier transform pair [61].

Theorem 5.4. *If a function $f(t)$ vanishes faster at infinity than $t^{-1/2}$ then*

$$\Delta_t \Delta_\omega \geq \sqrt{\frac{\pi}{2}} \tag{5.36}$$

where equality holds only for functions of the form

$$f(t) = \sqrt{\frac{\alpha}{\pi}} e^{-\alpha t^2}$$

i.e., signals which are Gaussians.

The proof is a consequence of Schwarz's inequality which states

$$|\int_{-\infty}^{\infty} g(t)h(t)dt|^2 \leq \int_{-\infty}^{\infty} |g(t)|^2 dt \int_{-\infty}^{\infty} |h(t)|^2 dt$$

with equality only if g and h are linearly dependent, i.e.,

$$h(t) = kg(t)$$

Now take $g(t) = tf(t)$ and $h(t) = \frac{df}{dt} = f'$. Then substituting into Schwarz's inequality we obtain

$$|\int_{-\infty}^{\infty} tf(t)f'(t)dt|^2 \leq \int_{-\infty}^{\infty} |tf(t)|^2 dt \int_{-\infty}^{\infty} |f'(t)|^2 dt \tag{5.37}$$

To prepare for integration by parts observe that

$$f(t)f'(t) = \frac{1}{2}\frac{df^2(t)}{dt}$$

Now the quantity inside the absolute value sign on the left-hand side of the inequality becomes

$$\int_{-\infty}^{\infty} t\frac{1}{2}(f^2)'dt = \frac{t(f^2(t))}{2}|_{-\infty}^{\infty} - \frac{1}{2}\int_{-\infty}^{\infty} f^2 dt$$

Since

$$\lim_{t\to\pm\infty} \sqrt{t}f(t) = 0$$

the first term on the right-hand side above is zero. For simplicity, we make the further assumption that

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = 1$$

Given this we now have

$$|\int_{-\infty}^{\infty} tf(t)f'(t)dt|^2 = \frac{1}{4}$$

Invoking Parseval's Formula given in equation (5.26) and the formula for the transform of a derivative in equation (5.24) we deduce

$$\int_{-\infty}^{\infty} |f'(t)|^2 dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 |F(\omega)|^2 d\omega \tag{5.38}$$

$$= \frac{1}{2\pi}\Delta_\omega^2 \tag{5.39}$$

Hence equation (5.37) becomes

$$\frac{1}{4} \leq \Delta_t^2 \frac{1}{2\pi}\Delta_\omega^2$$

from which the Uncertainty Principle in equation (5.36) follows.

We have equality if

$$\frac{df}{dt} = ktf(t)$$

which has the solution

$$f(t) = e^{kt^2/2}C$$

which has the form of a Gaussian. Figure 5.2.3 demonstrates the consequences of this uncertainty principle. Observe how a signal which is narrow, or localized in time has poorer localization in the frequency domain and vice versa.

5.2.4 The Delta Function

In this section we introduce the Dirac delta function $\delta(t)$ which facilitates the definition of the Fourier integral transform for many important functions.

The Dirac delta function is not a function in the traditional sense and if it is viewed as such, the formula which define it do not make sense. It is referred to as a *generalized* function, or *functional* inasmuch as it is defined *by its properties* which we now present.

Property 5.2.1.

$$\int_{-\infty}^{\infty} x(t)\delta(t-\tau)dt = x(\tau)$$

where $x(t)$ is arbitrary, but continuous at the point $t = \tau$.

Property 5.2.2.

$$\int_{-\infty}^{\infty} \delta(t)dt = 1 \quad \delta(t) = 0 \text{ if } t \neq 0.$$

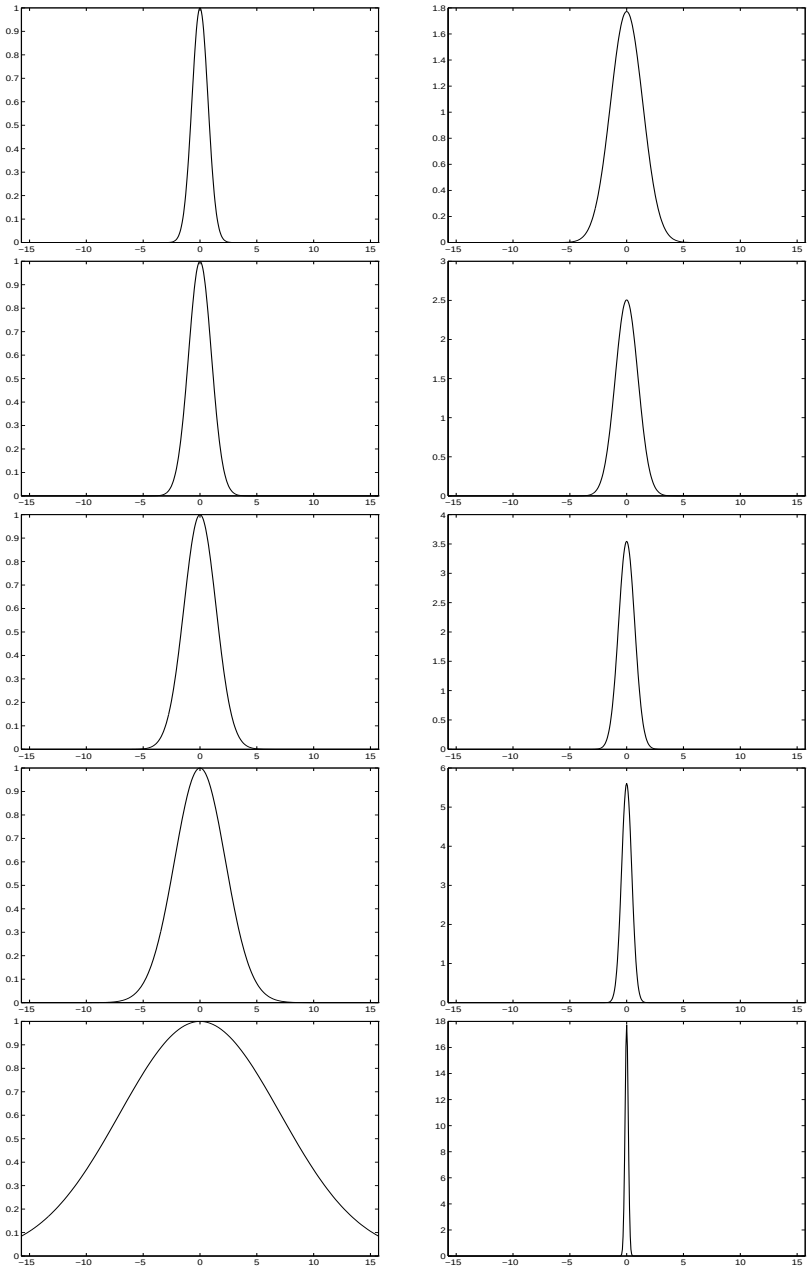

Fig. 5.6 The left column, top to bottom, are Gaussians with $\alpha = 1, .5, .25, .1, .01$. The right column shows the corresponding CTFTs of the figures on the left.

Property 5.2.3.

$$\delta(t) = \lim_{n\to\infty} f_n(t)$$

where

$$\int_{-\infty}^{\infty} f_n(t)dt = 1, \quad \lim_{n\to\infty} f_n(t) = 0 \text{ if } t \neq 0.$$

For example, such a sequence of (discontinuous) functions may be defined

$$f_n(t) = \begin{cases} n/2 & |t| \leq 1/n \\ 0 & 1/n < |t| \end{cases} \tag{5.40}$$

Property 5.2.4.

$$f(t)\delta(t) = f(0)\delta(t)$$

We remark that properties 5.2.1-5.2.3 serve as alternative definitions of the delta function.

As already stated, the motivation for introducing the delta function was to permit the computation of Fourier transforms of otherwise non-transformable functions. To begin, the Fourier transform of the Delta function is

$$\begin{aligned} \mathcal{F}(\delta(t)) &= \int_{-\infty}^{\infty} \delta(t)e^{-i\omega t}dt \\ &= 1 \end{aligned}$$

We also have the related transform

$$1 \leftrightarrow 2\pi\delta(\omega) \tag{5.41}$$

Since we know the transform of the delta function is given by $\mathcal{F}(\delta(t)) = 1$, we can easily compute the inverse transform of 1 as

$$\delta(t) = \int_{-\infty}^{\infty} e^{i\omega t}d\omega.$$

Example 5.6.

$$e^{i\alpha t} \leftrightarrow 2\pi\delta(\omega - \alpha)$$

and

$$\delta(t - \tau) \leftrightarrow e^{-i\omega\tau}$$

Example 5.7.

$$\cos(\alpha t) \leftrightarrow \pi(\delta(\omega + \alpha) + \delta(\omega - \alpha))$$

Example 5.8.

$$\sin(\alpha t) \leftrightarrow i\pi(\delta(\omega + \alpha) - \delta(\omega - \alpha))$$

Equipped with the generalized function $\delta(t)$ we may formally show the transform pair (5.10-5.9) actually performs an inverse operation. Taking $\tilde{x}(t)$

as the inverse of the transform, we would like to show that this is in fact $x(t)$. The inverse is provided by equation (5.10) as

$$\tilde{x}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega$$

Inserting the expression for the transform from equation (5.9)

$$\tilde{x}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} x(\tau) e^{-i\omega\tau} d\tau) e^{i\omega t} d\omega$$

Interchanging the order of integration

$$\tilde{x}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(\tau) (\int_{-\infty}^{\infty} e^{i\omega(t-\tau)} d\omega) d\tau$$

Since the integral $\int_{-\infty}^{\infty} x(\tau) e^{i\omega(t-\tau)} d\omega = 2\pi\delta(t-\tau)$ we have

$$\tilde{x}(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t-\tau) d\tau = x(t)$$

In other words, the function and the inverse of its transform are equal at points of continuity. We note that this proof is strictly only valid at points where $x(t)$ is continuous since this assumed in invoking the delta function property. However, the inversion formula may be shown to hold at points where there is a jump discontinuity if we assume that

$$f(t) = \frac{f(t_+) + f(t_-)}{2}$$

5.2.5 Continuous-Time Systems

This section develops the ideas of linear systems within the framework of continuous-time systems. A physical system is modeled as a transformation, or operator T, where both the range and the domain of the operator are functions with values in $\mathbb{R}$. Again, we write

$$y(t) = T\{x(t)\}$$

5.2.5.1 Linearity A system is said to be linear if

$$T\{c_1 x_1(t) + c_2 x_2(t)\} = c_1 T\{x_1(t)\} + c_2 T\{x_2(t)\} \tag{5.42}$$

The system output $h_\tau(t)$ due to the continuous unit (area) impulse, occurring at time τ, is referred to as the impulse response and is defined as

$$h_\tau(t) = T\{\delta(t-\tau)\}. \tag{5.43}$$

We see that in the general case, the impulse response is actually a family of functions which depends on the parameter τ. It will be shown that this impulse response fully characterizes linear time-invariant systems.

5.2.5.2 Continuous-Time Invariance A system is said to be time invariant if

$$y(t) = T\{x(t)\},$$

then it follows that

$$y(t-\tau) = T\{x(t-\tau)\}.$$

In terms of the impulse response, we have that if

$$h_0(t) = T\{\delta(t)\},$$

then

$$h_0(t-\tau) = T\{\delta(t-\tau)\}.$$

We will write the zero displacement impulse function as $h(t) = h_0(t)$.

Now we show that the operation of a linear time invariant system is equivalent to continuous-time convolution. We begin by expressing the input function as the sum of its impulses, i.e.,

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau \tag{5.44}$$

Applying the linear operator T to this representation gives

$$T\{x(t)\} = T\{\int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau\}$$

which by linearity becomes

$$\begin{aligned} &= \int_{-\infty}^{\infty} x(\tau)T\{\delta(t-\tau)\}d\tau \\ &= \int_{-\infty}^{\infty} x(\tau)h_\tau(t)d\tau \end{aligned}$$

Therefore

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h_\tau(t)d\tau$$

which for a time invariant linear becomes

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau$$

which we recognize as the continuous-time convolution equation. This result is especially important in view of the convolution theorems stated in equations (5.30) and (??).

5.2.5.3 Frequency Analysis and Eigenfunctions The Fourier transform $H(\omega)$ of the impulse response $h(t)$ again plays a special role in the frequency analysis of linear time-invariant systems. Referred to as the frequency response, or *system function*, we write this as the integral

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-i\omega t}dt \tag{5.45}$$

The continuous-time exponentials are the eigenfunctions of the continuous-time linear time invariant systems and that the associated eigenvalues are the corresponding frequency response functions. To show this first write

$$y(t) = x(t) * h(t)$$

from which it follows that

$$Y(\omega) = X(\omega)H(\omega).$$

Upon taking the inverse transform we have

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)H(\omega)e^{i\omega t}d\omega \tag{5.46}$$

If we take as our input

$$x(t) = e^{i\alpha t}$$

then

$$X(\omega) = 2\pi\delta(\omega - \alpha)$$

Thus,

$$Y(\omega) = 2\pi\delta(\omega - \alpha)H(\omega).$$

Inverting this shows that the output of the system is a constant multiple of the input, i.e.,

$$y(t) = H(\alpha)e^{i\alpha t}$$

from which we conclude that the functions $e^{i\alpha t}$ are eigenfunctions of the linear time invariant operators with eigenvalues $H(\alpha)$.

5.2.6 Sampling Continuous-Time Data

We may view a discrete sampling of the continuous function $x(t)$ in terms of the application of the Dirac delta function discussed in section 5.2.4. To start, it is useful to define the impulse train

$$s_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT) \tag{5.47}$$

where T is the time between impulses. It can be shown that the Fourier transform $S_T(\omega)$ of the continuous-time function $s_T(t)$ is given by

$$S_T(\omega) = \frac{2\pi}{T} \sum_{n=-\infty}^{\infty} \delta(\omega - n\frac{2\pi}{T})$$

Defining the sampling frequency $\omega_T = \frac{2\pi}{T}$, we then write this as

$$S_T(\omega) = \omega_T \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_T)$$

Now we define the continuous-time sampled function $x_s(t)$ as a product of the impulse train (5.47) and the continuous-time function $x(t)$ to be sampled, i.e.,

$$x_s(t) = x(t)s_T(t).$$

By the frequency convolution theorem (5.32) we have

$$X_s(\omega) = \frac{1}{2\pi} X(\omega) * S(\omega).$$

Writing the Fourier integrals of the functions $x_s(t), x(t)$ as $X_s(\omega), X(\omega)$ we evaluate the convolution

$$\begin{aligned} X_s(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega - \tau) S(\tau) d\tau \\ &= \frac{1}{T} \int_{-\infty}^{\infty} X(\omega - \tau) \sum_{n=-\infty}^{\infty} \delta(\tau - n\omega_T) d\tau \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} X(\omega - \tau) \delta(\tau - n\omega_T) d\tau \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} X(\omega - n\omega_T) \end{aligned}$$

Thus, the sampled time-frequency is equal to the superposition of an infinite number of shifted replicas of the continuous-time frequency spectrum.

At this point we address the important question of when this inversion of the transformed result of sampling will lead to a lossless reconstruction of the continuous signal. The transformed quantity to be inverted must clearly look like $X(\omega)$. This would be possible if we could eliminate all of the translated replicas of $X(\omega)$ which make up $X_s(\omega)$ so as to leave exactly one version of $X(\omega)$ centered about the origin. Geometrically, we see that this will always be possible as long as the shifted transforms

$$\ldots, X(\omega - \omega_T), X(\omega), X(\omega + \omega_T), \ldots$$

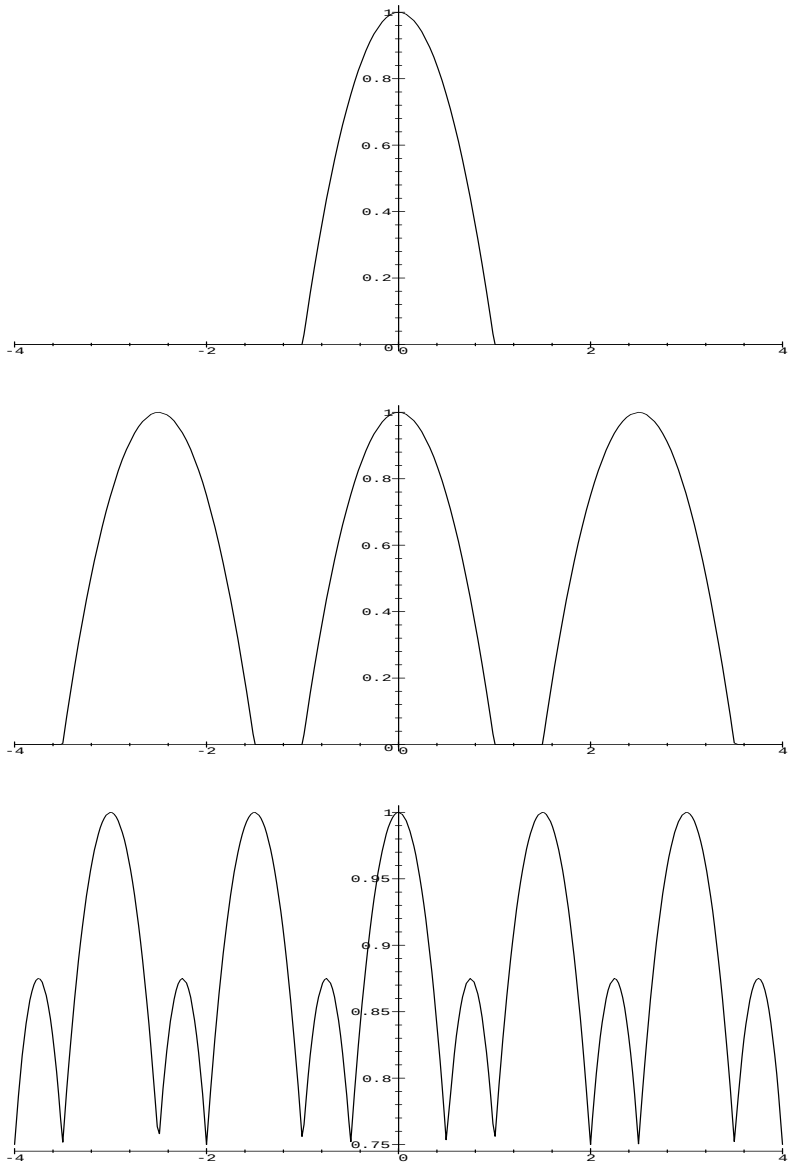

Fig. 5.7 Top: $X(\omega)$; middle $X_s(\omega)$ with sampling above the Nyquist frequency; bottom: $X_s(\omega)$ with sampling below the Nyquist frequency.

do not overlap.

The condition for non-overlapping X's is a property of the width of the spectrum of $X(\omega)$ and the sampling rate ω_T. It is clear that this will be possible if, and only if the Fourier transform $X(\omega)$ has finite extent. Such a function is referred to as band-limited.

Definition 5.8. *A function $x(t)$ with Fourier transform $X(\omega)$ is said to be band-limited if*

$$X(\omega) = 0 \quad \textit{for } |\omega| > \omega_N$$

Thus, we see that the non-overlapping condition is satisfied if

$$\omega_N < \omega_T - \omega_N.$$

This result is summarized in the following important theorem:

Theorem 5.5. *The* **Nyquist sampling theorem** *states that a band-limited continuous function $x(t)$ where*

$$X(\omega) = 0 \quad \textit{for } |\omega| > \omega_N$$

may be reconstructed from its samples if the sampling frequency $\omega_T > 2\omega_N$

This also implies that a function sampled at a frequency lower than $2\omega_N$ will not be reconstructable (without error). The error introduced by under-sampling a signal is called *aliasing.*

5.2.6.1 Reconstruction of Sampled Signals This section discusses the reconstruction of the sampled function by first posing the problem as a continuous-time frequency selective filter and then inverse transforming the result into the time-domain. We saw in the previous section that if the continuous function is sampled enough then we may reconstruct the signal without loss. This was done by applying an ideal low-pass filter to the function $X_s(\omega)$. Define the low-pass-filter as

$$H_L(\omega) = \begin{cases} 1 & |\omega| < \omega_T/2 \\ 0 & \text{else} \end{cases} \tag{5.48}$$

Then the inverse transform is

$$h_L(t) = \frac{\sin(\pi t/T)}{\pi t/T}.$$

The result of this filtering procedure in the frequency domain is

$$X_R(\omega) = H_L(\omega) X_S(\omega) \tag{5.49}$$

If

$$X(\omega) = X_R(\omega)$$

then no aliasing has occurred. If we write $x_n \equiv x_{nT}$, inverting equation (5.49) gives us

$$x_R(t) = x_s(t) * h_L(t) \tag{5.50}$$

where

$$x_s(t) = \sum_{n=-\infty}^{\infty} x_n \delta(t - nT) \tag{5.51}$$

and $h_L(t)$ is the inverse of the low-pass filter. Carring out the integration in equation (5.50) we obtain

$$\begin{aligned} x_R(t) &= \int_{-\infty}^{\infty} x_s(\tau) h_L(t-\tau) d\tau \\ &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x_n \delta(\tau - nT) h_L(t-\tau) d\tau \\ &= \sum_{n=-\infty}^{\infty} x_n \int_{-\infty}^{\infty} \delta(\tau - nT) h_L(t-\tau) d\tau \\ &= \sum_{n=-\infty}^{\infty} x_n h_L(t - nT) \end{aligned}$$

Inserting the expression for the impulse response gives the reconstruction as

$$x_R(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)/T}$$

Problems

5.18 Derive equation (5.13).

5.19 Derive equation (5.15).

5.20 Derive equation (5.17).

5.21 Derive equation (5.20).

5.22 Let the continuous-time function $x(t) = \sin(\alpha t)$ be sampled at the frequency $\omega_T = 2\pi/T$. If an ideal low-pass filter

$$H_L(\omega) = \begin{cases} 1 & |\omega| \leq \omega_T/2 \\ 0 & \text{else} \end{cases} \tag{5.52}$$

is used to filter the transformed sampled function $x_s(t)$ determine the reconstruction of the signal assuming aliasing and no aliasing.

5.23 The Fourier transform of a function $f(x) \in L^2$ may alternatively be written as

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

with the inverse transform

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{f}(\xi) e^{ix\xi} d\xi.$$

Show, as a consequence of this normalization, that

$$\|\tilde{f}\|_{L^2} = \|f\|_{L^2}. \tag{5.53}$$

5.24 Show that

$$\mathcal{F}(x(2t-k)) = \frac{1}{2} e^{-ik\omega/2} X(\frac{\omega}{2})$$

This transform is frequently used in the analysis of wavelets.

5.25 Prove equation (5.22).

5.26 Compute the CTFT of

$$x(t) = e^{-\alpha|t|}$$

Plot the function and the transform for various values of α to qualitatively establish the Uncertainty Principle.

5.27 Compute Δ_t and Δ_ω for the Guassian $x(t) = \exp(-\alpha t^2)$ and plot them and their product $\Delta_t \Delta_\omega$ as a function of α on the same graph.

5.2.7 Windowing in the Continuous-Time Domain

The examination of finite lengths of data from a long or infinite function may affect the frequency content of the Fourier transform. In order to interpret transforms of these smaller time series in terms of the larger series it is important to understand the impact of data *windowing*.

We assume that the data to be Fourier analyzed is a continuous function of time $x(t)$ with $x, t \in \mathbb{R}$. In addition, we assume a finite window $w(t) = 0$ if $|t| > T$. Then we define the fixed windowed function

$$x_{\mathrm{w}}(t) = x(t)w(t)$$

where now $x_{\mathrm{w}}(t)$ has finite duration. The transform of the above product is determined by applying equation (5.32) to be

$$X_{\mathrm{w}}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\theta)W(\omega - \theta)d\theta. \tag{5.54}$$

The interpretation of the above windowing procedure in the frequency domain is central to understanding the impact of the windowing procedure on the Fourier analysis. Depending on the shape of $W(\omega)$ the resulting windowed transform $X_{\mathrm{w}}(\omega)$ is a contaminated approximation of $X(\omega)$. The frequency resolution can be reduced and the fact that $W(\omega)$ has infinite duration gives rise to *leakage*. We will consider both these topics in some detail before we proceed to the main application of the windowed transform, i.e., the *short-time* Fourier transform.

From equation (5.54) the action of windowing data in the time-domain is equivalent to convolving the true (and unknown) spectrum with the transform of the window function. We begin to examine the effect of this convolution by considering the special case of the window with infinitely narrow spectrum, i.e.,

$$W(\theta) = 2\pi\delta(\theta)$$

Plugging this window spectrum into equation (5.54) we conclude that the transform and the windowed transform are the same, i.e.,

$$X_{\mathrm{w}}(\omega) = X(\omega).$$

However, we may expect this result in view of the fact that in this case

$$w(t) = \mathcal{F}^{-1}(2\pi\delta(\theta)) = 1$$

which means we have done no windowing at all! Clearly the narrowness of the window spectrum is a desirable quality, although we cannot expect to obtain the limiting case for a practical window.

Now recall the rectangular window

$$w(t) = \begin{cases} \frac{1}{2T} & |t| < T \\ 0 & \text{else} \end{cases} \tag{5.55}$$

with associated Fourier transform

$$W(\omega) = \frac{\sin(T\omega)}{T\omega}. \tag{5.56}$$

To test the effect of this window on an infinite function, consider the complex exponential

$$x(t) = e^{i\omega_0 t}$$

which has as its Fourier transform

$$X(\omega) = 2\pi\delta(\omega - \omega_0).$$

If we construct the windowed function $x_{\mathrm{w}}(t)$ it is interesting to see how the Fourier transform is affected by this rectangular windowing procedure.

$$\begin{aligned} X_{\mathrm{w}}(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\theta)W(\omega-\theta)d\theta \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} 2\pi\delta(\theta-\omega_0)\frac{\sin T(\omega-\theta)}{T(\omega-\theta)}d\theta \\ &= \frac{\sin T(\omega-\omega_0)}{T(\omega-\omega_0)} \end{aligned}$$

This is the same graph as for the transform of the window, but now centered at ω_0. The above transform gives considerable insight into the effect of applying a window. We define the width of the *main-lobe* of the transformed window to be the distance between the first 2 zeros, and will refer to the oscillations away from the main-lobe as *side-lobes*. See the top of Figure 5.2.7. The roles of the main-lobe and the side-lobes may be considered independently. The more narrow the main-lobe, the better the frequency resolution, since we may view the action of $W(\omega - \theta)$ as favoring, or passing via convolution, the frequencies near ω. Note that the side-lobes add up contributions to the integral at frequencies further removed from the center of the window and consequently distort the spectrum of the true signal. The resulting transform of the windowed complex exponential is seen to be wider than the transform of the infinite signal (i.e., the delta function). We interpret this broadening of the spectrum as reduced frequency resolution (adjacent frequencies which are distinct may become indistinguishable). The undesirable oscillatory tail produced by the rectangular window will be referred to as *leakage*.

Now we consider the effect of *tapering* the window, i.e., reducing the values of the window gradually to zero at its edges. A simple example of a tapered window is the *triangular* window given by

$$w(t) = \begin{cases} \frac{1}{T}(1 - \frac{|t|}{T}) & |t| < T \\ 0 & \text{else} \end{cases} \tag{5.57}$$

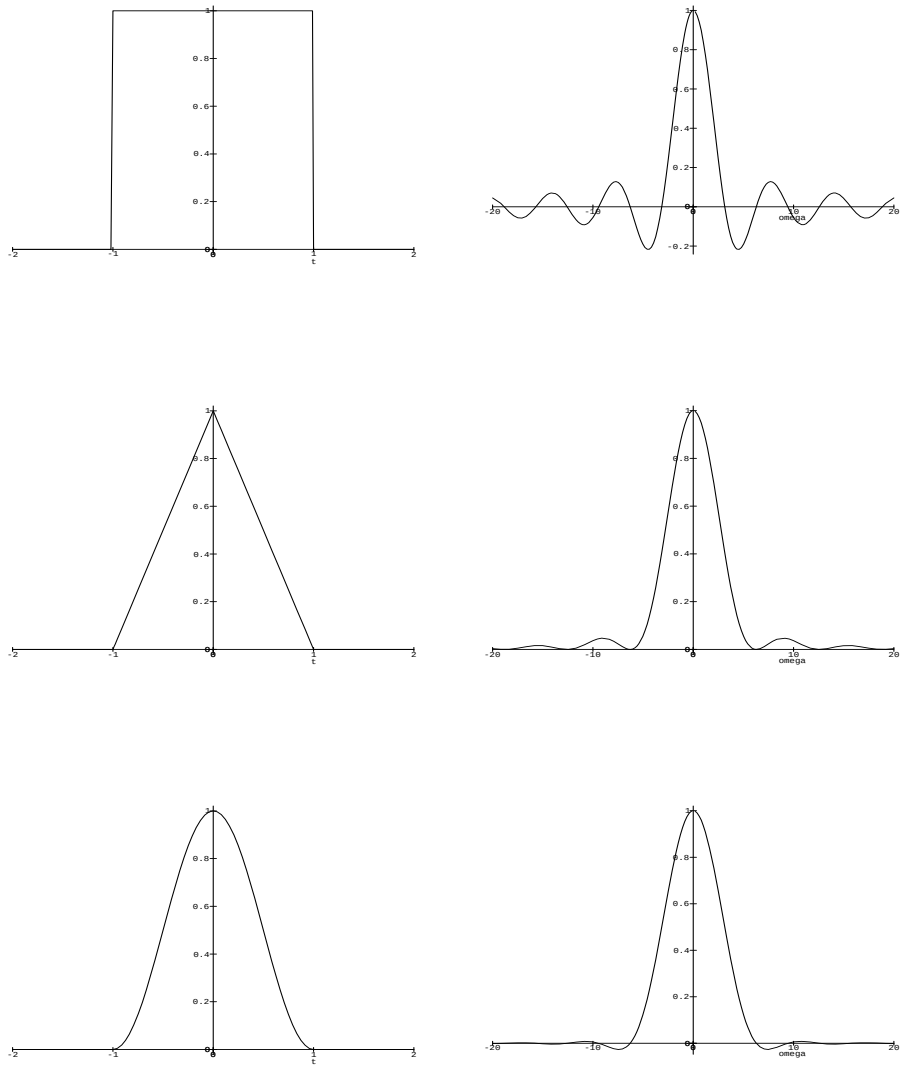

Fig. 5.8 The left column shows the boxcar, triangle and hanning windows. The right column shows the corresponding continuous Fourier transforms.

The corresponding Fourier transform of this window is

$$W(\omega) = \frac{\sin^2(\omega T/2)}{(\frac{\omega T}{2})^2}.$$

We notice that the main-lobe has width $4\pi/T$ which is actually wider than the width of the rectangular transform $2\pi/T$. This wider main-lobe is offset by the fact that the oscillations have a greatly reduced amplitude. As such, we expect that for a given T the frequency resolution will be reduced (due to the wider main-lobe) and that the leakage will be diminished due to the smaller side-lobes.

It is useful to summarize the effects of tapering the window:

- tapering the window in the time domain decreases the amplitudes of the side-lobes in the CTFT which in turn reduces leakage.
- tapering increases the width of the main lobe and consequently reduces frequency resolution (compared to other windows of the same duration).

5.2.7.1 The Uncertainty Principle Revisited Of course in addition to the tapering effects on the spectrum it is clear that the width $2T$ of the window plays an important role. Due to the uncertainty principle, the wider $w(t)$, the narrower $W(\omega)$ and conversely, the narrower $w(t)$ the wider the main-lobe of $W(\omega)$. Consider the following desired window attributes:

1. The main-lobe of $W(\omega)$ should be as narrow as possible to improve frequency resolution.
2. The window should be as short as possible to reduce filter length. Also, for the time dependent Fourier transform a shorter window improves time localization as will be discussed in the next section.

Since the first item requires that the window be wide and the second item requires that the window be narrow we see that there is a fundamental trade-off between frequency resolution and time resolution. Mathematically this is a consequence of the Uncertainty Principle.

5.3 SHORT-TIME FOURIER TRANSFORM

We have already remarked that the CTFT is a global transformation of a given function. For many purposes this is sufficient. For example, consider a signal which has registered the sounds of several different species of birds, each of which has distinct frequency content. If our goal is to identify which birds are present in a given recording, then the global transform is sufficient. However, if we seek to answer the modified question: when did a particular bird sing, then the global transform is not enough.

We show this using a simple analytical example. Let $x(t) = \delta(t - t_0)$ be the function to be transformed. The associated Fourier power spectrum is $|X(\omega)| = 1$. The delta function has equal energy in all frequencies. But the information concerning *when* a particular frequency occurs is clearly washed out.

Now apply a finite length window to the delta funtion and transform giving

$$X_w(\omega) = \int_{-\infty}^{\infty} \delta(t - t_0) w(t) e^{-i\omega t} dt$$

hence

$$X_w(\omega) = w(t_0) e^{-i\omega t_0} dt$$

Now we see that the spectrum is non-zero only inside the window, i.e.,

$$X_w(\omega) = \begin{cases} w(t_0) & t_0 \text{ in the window} \\ 0 & \text{else} \end{cases} \tag{5.58}$$

The implication is that we may detect a pulse only to within the time-resolution of the window. In particular, it is not possible to say where in the window the pulse occurred. This is however, a major improvement over the global transform where no spectrum based estimate would indicate the time dependence.

Hence we are motivated to apply a window to the signal explicitly, i.e., construct the family of windowed functions

$$x_w(\tau, t) = x(t) w(t - \tau)$$

with parameter τ which indicates the location of the window. This leads to the so-called short-time Fourier transform (STFT)

$$X_w(\tau, \omega) = \int_{-\infty}^{\infty} x(t) w(t - \tau) e^{-i\omega t} dt$$

To better understand the action of this window on the signal we examine the windowing procedure in the frequency domain. To assist with this analysis we introduce the reflected window $h(t) = w(-t)$. Then we can show

$$\mathcal{F}(e^{i\omega\tau} X_w(\tau, \omega)) = X(\theta) H(\theta - \omega)$$

Thus, we interpret the action of the STFT as applying a window of fixed length in the frequency domain to the spectrum of the function $X(\theta)$ about the frequency ω.

If the definition of the window is modified to be

$$x_w(t) = x(t) \frac{1}{|a|} \psi(\frac{t - b}{a})$$

then a similar calculation shows (see Section 6.1.5)

$$\mathcal{F}(x(b) * h_\omega(b-t)) = X(\theta)\tilde{\psi}(-a(\theta-\omega))$$

We see the action of this adaptive window is to apply a filter in the frequency domain of varying widths.

6

Wavelet Expansions

6.1 THE CONTINUOUS WAVELET TRANSFORM

The continuous wavelet transform (CWT) is given by

$$X(b,a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\overline{\psi(\frac{t-b}{a})}dt \tag{6.1}$$

One interpretation of the transform $X(b,a)$ is that it provides a measure of similarity between the signal $x(t)$ and the continuously translated and dilated function $\psi(t)$. The inverse wavelet transform is then provided by

$$x(t) = \frac{1}{\sqrt{|a|}C_\psi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X(b,a)\psi(\frac{t-b}{a})\frac{dadb}{a^2} \tag{6.2}$$

where $\mathcal{F}(\psi(t)) = \tilde{\psi}(\omega)$ and

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\tilde{\psi}(\omega)|^2}{|\omega|}d\omega \tag{6.3}$$

For the inverse transform to exist, it is required that the *admissibility condition* $C_\psi < \infty$ hold. Together, equations (6.1) and (6.2) form the continuous wavelet transform pair.

We may denote the CWT transform pair either as

$$\mathcal{W}(x(t)) = X(b,a)$$

or as

$$x(t) \leftrightarrow X(b,a)$$

For future reference we define the scaled and translated window

$$g_{ab}(t) = \frac{1}{\sqrt{|a|}}\psi(\frac{t-b}{a})$$

where a is the *scale* parameter and b is the *shift* parameter. The factor $1/\sqrt{|a|}$ ensures that the window is normalized at all scales, i.e.,

$$\|\psi(t)\|^2 = 1, \quad \Rightarrow \quad \|g_{ab}(t)\|^2 = 1 \tag{6.4}$$

See Figure 6.1 for a display of the dilated and shifted Gaussian window.

6.1.1 Wavelet Analysis in the Fourier Domain

It will be useful to know that the Fourier transform of the scaled and dilated mother is given by

$$\frac{1}{\sqrt{|a|}}\psi(\frac{t-b}{a}) \leftrightarrow \sqrt{|a|}e^{-ib\omega}\tilde{\psi}(a\omega) \tag{6.5}$$

In addition, the Fourier transform of the forward CWT is found using Parseval's equation, i.e.,

$$X(b,a) = (x(t), g_{ab}(t)) \tag{6.6}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\overline{G_{ab}(\omega)}d\omega \tag{6.7}$$

Substituting equation (6.5) into the above gives the desired representation of the CWT in the Fourier domain

$$X(b,a) = \frac{\sqrt{|a|}}{2\pi}\int_{-\infty}^{\infty} X(\omega)\overline{\tilde{\psi}(a\omega)}e^{i\omega b}d\omega \tag{6.8}$$

6.1.2 The Resolution of the Identity and the Admissibility Condition

In this section we reconsider the CWT pair and verify that the application of the transform, followed by its inverse, indeed preforms the identity mapping. In addition, we consider the criterion a function must satisfy to be considered admissible as a wavelet. For further mathematical details we refer the reader to the text [16].

To begin, we seek to evaluate the integral

$$\int\int X(b,a)g_{ab}(t)\frac{dbda}{a^2}$$

To this end, substitute the expression for $X(b,a)$ in the Fourier domain provided by equation (6.8) to give

$$\int X(b,a)g_{ab}(t)db = \frac{\sqrt{|a|}}{2\pi}\int\left(\int X(\omega)\overline{\tilde{\psi}(a\omega)}e^{i\omega b}d\omega\right)g_{ab}db \tag{6.9}$$

interchanging orders of integration

$$= \frac{\sqrt{|a|}}{2\pi}\int X(\omega)\overline{\tilde{\psi}(a\omega)}\left(\int g_{ab}e^{i\omega b}db\right)d\omega \tag{6.10}$$

$$\tag{6.11}$$

It can be shown that

$$\int g_{ab}(t)e^{i\omega b}db = \sqrt{|a|}e^{i\omega t}\tilde{\psi}(a\omega) \tag{6.12}$$

Hence, using this result above, we now have

$$\int X(b,a)g_{ab}(t)db = \frac{\sqrt{|a|}}{2\pi}\int X(\omega)|\tilde{\psi}(a\omega)|^2e^{i\omega t}d\omega$$

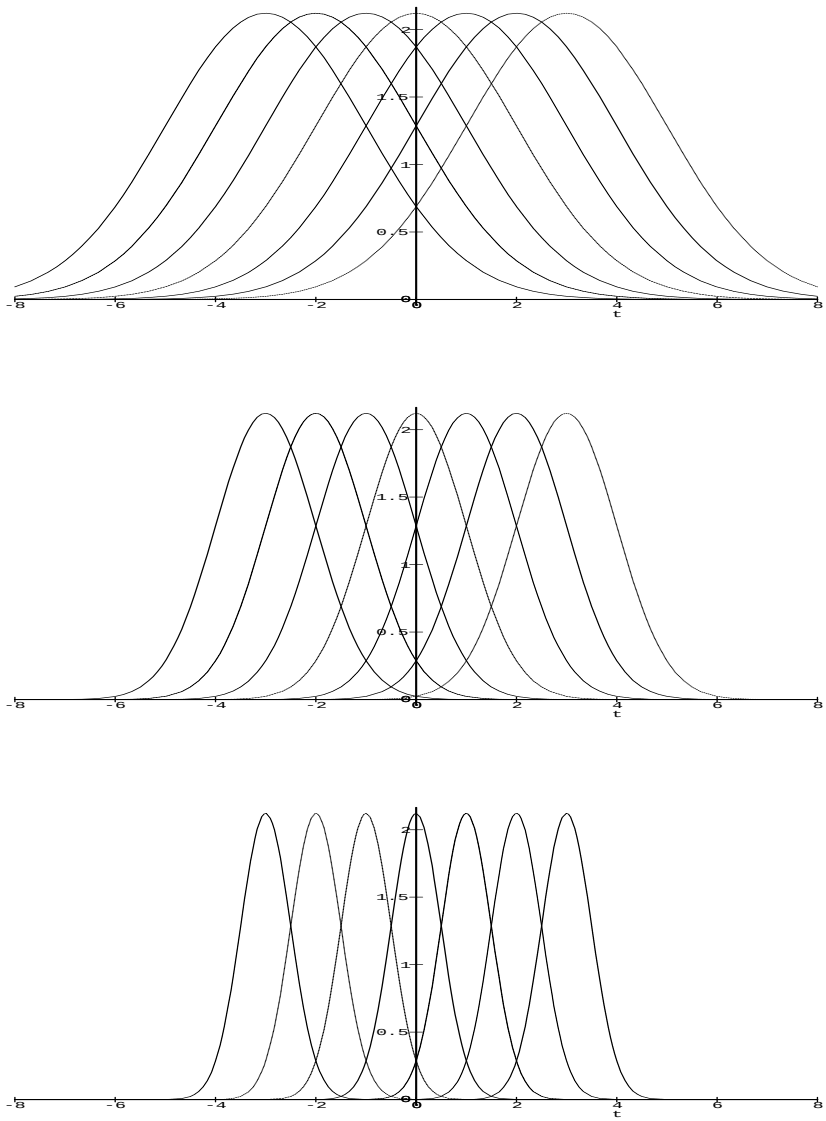

Fig. 6.1 This figure displays dilations and shifts of the Gaussian window $\psi((t-b)/a)$ for discrete values of $b = -3, -2, -1, 0, 1, 2, 3$. Top: The dilation corresponding to $a = 2$. Middle: $a = 1$. Bottom: $a = 1/2$.

Now integrating this equation with respect to a and employing the differential element da/a^2 it follows that

$$\int\int X(b,a)g_{ab}(t)\frac{dbda}{a^2} = \int\left(\frac{\sqrt{|a|}}{2\pi}\int X(\omega)|\tilde{\psi}(a\omega)|^2 e^{i\omega t}d\omega\right)\frac{da}{a^2} \tag{6.13}$$

again interchanging orders of integration

$$= \frac{1}{2\pi}\int X(\omega)e^{i\omega t}\left(\int\frac{|\tilde{\psi}(a\omega)|^2}{|a|}da\right)d\omega \tag{6.14}$$

A standard substition of variables permits the rewriting of the inner integral as

$$\int\frac{|\tilde{\psi}(a\omega)|^2}{|a|}da = \int\frac{|\tilde{\psi}(a')|^2}{|a'|}da' \tag{6.15}$$

which we recognize as the definition of C_ψ.

Thus,

$$\int\int X(b,a)g_{ab}(t)\frac{dbda}{a^2} = \frac{C_\psi}{2\pi}\int X(\omega)e^{i\omega t}d\omega \tag{6.16}$$

$$= C_\psi x(t) \tag{6.17}$$

since the last equation is just the inverse Fourier transform of $x(t)$. This is exactly the inverse transform of equation (6.2).

The admissibility condition

$$C_\psi = \int_{-\infty}^{\infty}\frac{|\tilde{\psi}(\omega)|^2}{|\omega|}d\omega < \infty \tag{6.18}$$

is required so that the function $x(t)$ may be reconstructed via the inverse CWT equation.

The value C_ψ will in general not be finite if the integrand blows up at $\omega = 0$. Hence it is required that

$$\tilde{\psi}(0) = 0 \tag{6.19}$$

This requirement may be interpreted further by inverting the Fourier transformation. Recall that

$$\tilde{\psi}(\omega) = \int_{-\infty}^{\infty}\psi(t)e^{-i\omega t}dt$$

Substituting $\omega = 0$ into the above results in

$$\int_{-\infty}^{\infty}\psi(t) = 0 \tag{6.20}$$

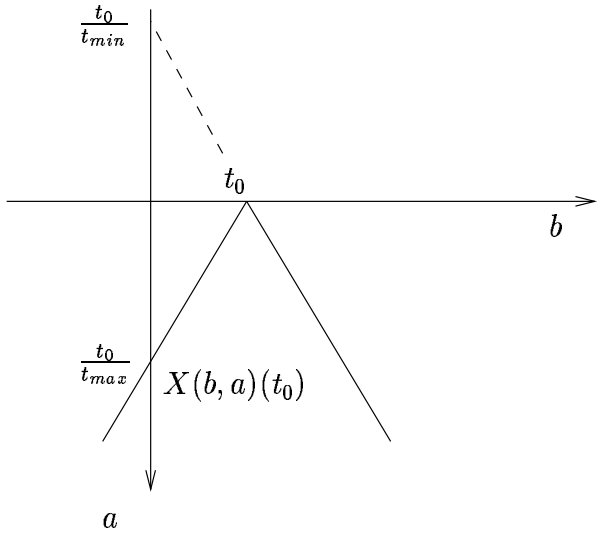

Fig. 6.2 The triangular region is the domain of influence of the point t_0 on the continuous wavelet transform $X(b,a)$

- Given $\tilde{\psi}(0) = 0$ and that the spectrum of $\tilde{\psi}(\omega)$ decays for high-frequencies the wavelet $\psi(t)$ may be interpreted as a band-pass filter.
- An admissible wavelet has zero area by equation (6.20).
- $\psi(t) \in L^1(\mathbb{R})$ implies that $\tilde{\psi}(\omega)$ is continous.
- *Almost* admissibility. For the Morlet wavelet, $\tilde{\psi}(0) \approx 0$

6.1.3 Properties of the Continuous Wavelet Transform

Proposition 6.1. *Linearity of the CWT.*

$$\mathcal{W}(x(t) + y(t)) = \mathcal{W}(x(t)) + \mathcal{W}(y(t)) \tag{6.21}$$
$$\mathcal{W}(cx(t)) = c\mathcal{W}(x(t)) \tag{6.22}$$

Proposition 6.2. *Shifting Property.*

$$x(t) \leftrightarrow X(b,a) \tag{6.23}$$
$$x(t - t_0) \leftrightarrow X(b - t_0, a) \tag{6.24}$$

Proposition 6.3. *Scaling Property.*

$$x(\frac{t}{\alpha}) \leftrightarrow \sqrt{\alpha} X(\frac{b}{\alpha}, \frac{a}{\alpha}) \tag{6.25}$$

Proposition 6.4. *Energy Property 1.*

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |X(b,a)|^2 \frac{dadb}{a^2} \tag{6.26}$$

Proposition 6.5. *Energy Property 2.*

$$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X(b,a)\overline{Y(b,a)} \frac{dadb}{a^2} \tag{6.27}$$

Proposition 6.6. *Linearity of the Wavelet. If the collection of functions* $\{\psi_i(t)\}$ *are wavelets, i.e., they satisfy the admissibility condition, then the linear combination*

$$\psi(t) = \sum_i c_i \psi_i(t)$$

is also a wavelet (since it also satisfies the admissibility condition).

6.1.4 Time-Scale Analysis

In contrast to the STFT which produced a transformation of data to the time-frequency domain (τ, ω) the wavelet transform transforms a 1-dimensional signal to a 2-dimensional time-scale plane (b, a). In the former case, it is standard to examine the graph of the magnitude of the STFT $|X(\tau, \omega)|$. This time-frequency diagram is referred to as the spectogram. In the latter case, the spectrogram is replaced by the *scalogram* which consists of the graph of the magnitude of the CWT, i.e., $|X(b,a)|$. In both cases there is an associated phase diagram, e.g., the polar form of the CWT is

$$X(b,a) = |X(b,a)|e^{\phi(b,a)}$$

In what follows we consider *localization* features of the CWT which aid in the interpretation of the scaleogram.

6.1.4.1 Time-Localization of the CWT For the moment, we assume that the wavelet $\psi(t)$ vanishes identically outside the time domain $[t_{\min}, t_{\max}]$ where $t_{\min} < 0$ and $t_{\max} > 0$. Although this condition may not be true in practice, the magnitude of the wavelet will in general be very small, and approximately zero on this interval.

- $X(b,a)$ *dependence on* t_0. We define the domain of influence of the signal $x(t)$ at the point $t = t_0$ to be the region in the (b,a)-plane such that the values of $X(b,a)$ are dependent on t_0.

  To demonstrate the time-localization property of the CWT we consider the signal $x(t) = \delta(t - t_0)$. Substituting this choice for $x(t)$ into equation (6.1) gives

$$X(b,a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \delta(t - t_0)\overline{\psi(\frac{t-b}{a})}dt$$

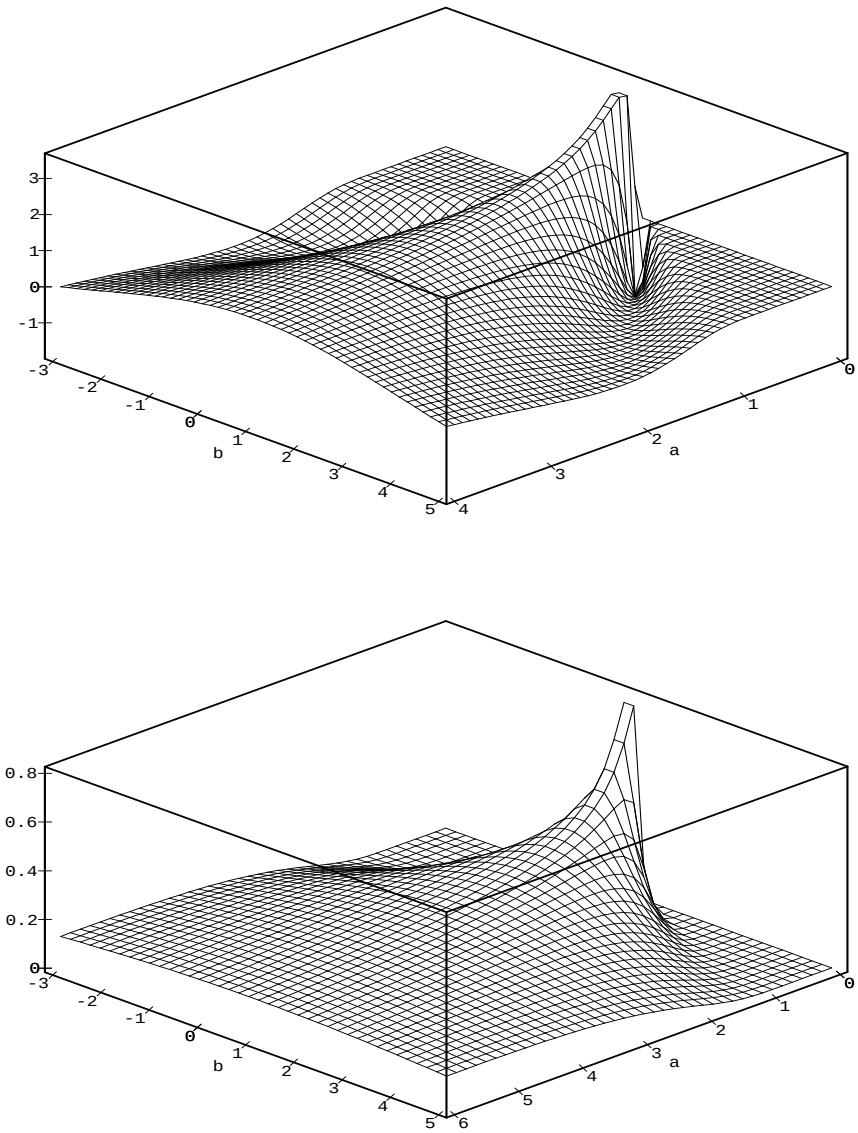

Fig. 6.3 Top: The scalogram of the Mexican hat wavelet transform of the delta function centered at $t = 1$. Bottom: The scalogram of the Morlet wavelet transform of the same delta function.

which integrates to

$$X(b,a) = \frac{1}{\sqrt{|a|}}\overline{\psi(\frac{t_0-b}{a})}$$

Note that $\overline{\psi(\frac{t_0-b}{a})} = 0$ unless

$$\frac{t_0-b}{a} \in [t_{\min}, t_{\max}]$$

This condition defines 2 lines, $at_{\min} + b = t_0$ and $at_{\max} + b = t_0$, which intersect at the point t_0 and delimit the region in the (b,a) plane for which the transform $X(b,a)$ depends on t_0. This situation is depicted in Figure 6.2.

The CWT $X(b,a)$ of the delta function $\delta(t-1)$ is shown in Figure 6.3 for both the mexican hat wavelet and the Morlet wavelet. Observe that the width of the transform is narrower for smaller scale parameter a. The projection of the non-zero values of this this transform fall into the domain of influence of the pulse originating at $t = 1$.

- $X(b_0, a_0)$ *dependence on* t. Another helpful question to consider is for which values of t does the signal $x(t)$ contribute to the value of $X(b,a)$ at the point (b_0, a_0)? By examining Figure **??** we conclude that if

$$a_0 t_{\min} + b_0 \leq t \leq a_0 t_{\max} + b_0$$

then the value of $x(t)$ may influence the CWT at the point (b_0, a_0).

6.1.4.2 Frequency-Localization of the CWT In the previous section we considered the how a localized signal $x(t)$ in the time domain could influence the the values of the CWT $X(b,a)$. Now we examine the same question but focus on the influence of a pure sinusoid on the transform.

- $X(b,a)$ *dependence on* ω_0. Now we consider for which portion of the (b,a) plane the CWT $X(b,a)$ is influenced by the frequency ω_0. To examine this let $x(t) = \exp(i\omega_0 t)$, the complex monochromatic signal with frequency ω_0. Given that the Fourier transform of $\exp(i\omega_0 t)$ is $2\pi\delta(\omega - \omega_0)$ we may compute the CWT of $x(t)$ using equation (6.8), i.e.,

$$X(b,a) = \frac{\sqrt{|a|}}{2\pi}\int_{-\infty}^{\infty} 2\pi\delta(\omega-\omega_0)\overline{\tilde{\psi}(a\omega)}e^{i\omega b}d\omega$$

from which we conclude that

$$X(b,a) = \sqrt{|a|}\overline{\tilde{\psi}(a\omega_0)}e^{i\omega_0 b}$$

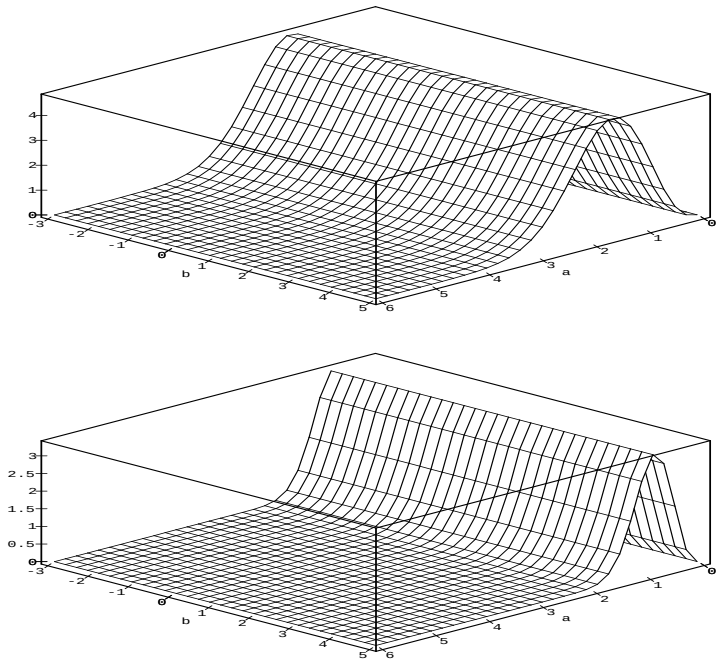

Fig. 6.4 Top: The scalogram of the Mexican hat wavelet transform of the monochromatic complex exponential $x(t) = \exp(it)$. Bottom: The scalogram of the Mexican hat wavelet transform of the monochromatic complex exponential $x(t) = \exp(i2t)$.

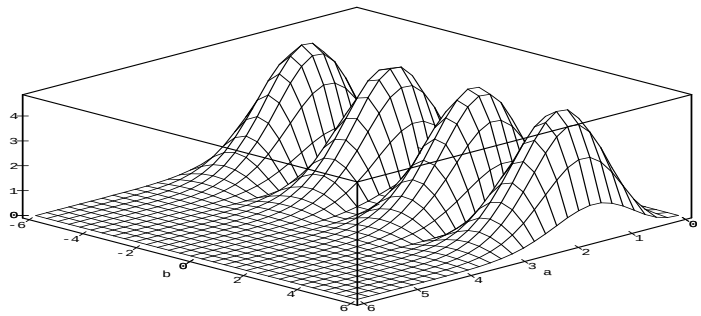

Fig. 6.5 The scalogram of the Mexican hat wavelet transform of the monochromatic signal $x(t) = \sin(t)$.

If we now assume that the Fourier transform of the wavelet is band-limited, i.e., $\tilde{\psi}(\omega) = 0$ whenever $\omega \notin [\omega_{\min}, \omega_{\max}]$. Hence, in general,

$$\tilde{\psi}(a\omega) = 0 \quad \text{if} \quad \omega_0 \in [\frac{\omega_{\min}}{a}, \frac{\omega_{\max}}{a}]$$

Hence,

$$X(b,a) = 0 \quad \text{unless} \quad \omega_0 \notin [\frac{\omega_{\min}}{a}, \frac{\omega_{\max}}{a}]$$

In other words, the domain of influence of ω_0 on $X(b,a)$ is the horizontal strip defined by

$$\frac{\omega_{\min}}{\omega_0} \leq a \leq \frac{\omega_{\max}}{\omega_0} \tag{6.28}$$

The CWT of $\exp(i\omega_0 t)$ is shown in Figure 6.4 for 2 different values of ω_0. Compare this with Figure 6.1.6 which displays the CWT of the signal $\sin(\omega_0)$. The magnitude is no longer constant along lines a =constant. This effect can be problematic when interpreting scalograms. However, it may be avoided by employing *progressive* wavelets; see Problem 6.10 and reference [25] for further details.

It is also interesting to note that the actual value of a for which $|X(b,a)|$ is a maximum (see Figure 6.4) when the signal has the form $\exp(i\omega_0 t)$ is wavelet dependent. For instance, for the Mexican hat wavelet this value is

$$a = \frac{1}{\omega_0}\sqrt{\frac{5}{2}} \tag{6.29}$$

while for Morlet's wavelet

$$a = \frac{1}{2\omega_0}(\alpha + \sqrt{\alpha^2 + 2}) \tag{6.30}$$

See exercise 6.9 for more details.

- $X(b_0, a_0)$ *dependence on* ω. Finally, it may be similarly argued that the Fourier components ω of $X(\omega)$ which influence the value of the CWT at the point (b_0, a_0) are determined by the relation

$$\frac{\omega_{\min}}{a_0} \leq \omega \leq \frac{\omega_{\max}}{a_0} \tag{6.31}$$

We observe that the bandwidth of the frequencies which influence $X(b, a)$ increases as the scale a decreases.

6.1.5 The Wavelet Transform as an Adaptive Filter

In this section we continue our discussion of the action of the wavelet transform in the frequency domain. The ideas parallel those of Section **??** where we considered the application of the STFT in the Fourier domain. Now we write the windowed signal as

$$x_w(t) = x(t)\overline{g_{ab}}(t)$$

The Fourier transform of this windowed signal is then

$$X_w(\omega) = \frac{1}{\sqrt{|a|}} \int x(t)\overline{\psi(\frac{t-b}{a})}e^{-i\omega t}dt$$

We seek to represent this transformation as a convolution. Thus define the impulse responses $h(t) = \overline{\psi(-t/a)}$ and $h_\omega(t) = \overline{\psi(-t/a)}\exp(i\omega t)$. Now we may write

$$X_w(\omega) = \frac{e^{-i\omega b}}{\sqrt{|a|}} \int x(t)h(b-t)e^{i\omega(b-t)}dt$$

or, using convolution notation,

$$X_w(\omega) = \frac{e^{-i\omega b}}{\sqrt{|a|}} x(b) * h_\omega(b)$$

Now computing the transform taking b as the time-variable we have

$$\mathcal{F}(x(b) * h_\omega(b)) = X(\theta)|a|\tilde{\psi}(-a(\theta - \omega))$$

From this equation we may conclude that the wavelet transform acts as a band-pass filter which has width proportional to the frequency. The benefit

of this adaptation is that the correponding width of the window in the time-domain will be narrow for high frequencies and wide for low frequencies, rather than a constant width window as for the STFT.

The change in width of the transform of the window with the center or mean frequency of the window is a property referred to as constant relative bandwidth. The *quality* factor, or Q of a transform may be defined as

$$Q = \frac{\text{center frequency}}{\text{bandwidth}}$$

The wavelet transform is often referred to as constant Q frequency analysis.

6.1.6 Discretization of the CWT

The number of instances for which the CWT of a signal may be computed analtically is obviously very small. In general, the evaluation of the CWT pair is done discretely. In this Section we consider the discretization of the forward transform and postpone discussion of the inverse to Chapter 6. Using the terminolgy of [25], we consider the restriction of $X(b,a)$ to a collection of fixed values, or *voices* $a = a_m$ where $a_{m+1} = \kappa a_m$, where κ is a constant and $a \in \mathbb{R}^+$. In general, this is done by writing

$$a_m = a_0^m \quad \text{where} \quad a_0 > 1 \quad \text{and} \quad m \in \mathbb{Z} \tag{6.32}$$

Observe that the width of the function

$$\psi(\frac{t-b}{a_0^m})$$

is dependent on the magnitude of a_0^m (it is in fact a_0^m times the width of $\psi(\frac{t-b}{a_0^m})$). This fact must be taken into account when discretizing the shift variable b since it makes good sense to use a discretized shift size which is proportional to the width of the discretely dilated wavelet. This may be accomplished by taking

$$b = nb_0a_0^m \text{ where } \quad b_0 > 0 \tag{6.33}$$

Putting this all together we write the wavelet on the discretized grid as

$$\psi_{mn} = a_0^{-\frac{m}{2}}\psi(a_0^{-m}t - nb_0)$$

Example 6.1. Let $a_0 = 2^{1/l}$. The positive integer l is often referred as the number of voices per octave. A choice of $l = 12$ represents a fine sampling of the (b,a)-plane. See the top of Figure 6.6. The choice of discrete time shift depends heavily on the nature of the data but typical values are between .1 and 1.

Example 6.2. Another importany choice of voice for the discrete grid is $a_0 = 2$ and $b_0 = 1$. This coarser sampling of the (b,a)-plane produces a *dyadic* grid. See the middle of Figure 6.6. It is also often convenient to display the grid using the coordinates $(b, -\ln a)$. See the bottom of Figure.

6.2 THE DISCRETE WAVELET TRANSFORM

Wavelet expansions are *local* expansions of the form

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_k^j \psi_k^j(x) \tag{6.34}$$

where the localized *little wave* functions $\psi_k^j(x)$

$$\psi_k^j(x) = 2^{-j/2}\psi(2^{-j}x - k) \tag{6.35}$$

are generated by the translations and dilations of the function $\psi(x)$ on the dyadic grid. The integers $j, k \in \mathbb{Z}$ correspond to the scale and location of the center of the function, respectively. Given the direct dependence of each basis function on it, $\psi(x)$ is sometimes referred to as the *mother wavelet.*

One manifestation of the local nature of the set of functions $\{\psi_k^j(x)\}$ is their orthogonality across both scale and shift, i.e., they are chosen to satisfy

$$(\psi_k^j, \psi_{k'}^{j'}) = \int_{-\infty}^{\infty} \psi_k^j(x)\overline{\psi_{k'}^{j'}(x)}dx = \delta_{k,k'}\delta_{j,j'} \tag{6.36}$$

With the wavelet orthogonality property given by equation (6.36) the expansion, or wavelet, coefficients $\{d_k^j\}$ may be calculated directly by

$$d_k^j = \int_{-\infty}^{\infty} f(x)\overline{\psi_k^j(x)}dx \tag{6.37}$$

The direct application of this formula for the computation of the coefficients $\{d_k^j\}$ is generally avoided in favor of a much faster recursion scheme called multiresolution pyramidal decomposition, or *Mallat's algorithm* which is derived in general in Section 6.5.

The wavelet expansion is local in the sense that only a few of the expansion coefficients $\{d_k^j\}$ contribute to the sum of the series $\sum_{j,k} d_k^j \psi_k^j(x)$ around any given point $x = x_o$. Note that the localization property of the expansion is scale dependent and better time-localization is achieved for coefficients with small j, while better frequency localization is achieved for coefficients with large j. This trade-off between time-frequency localization is discussed in more detail in Sections 5.2 and 5.2.

Wavelet analysis on a dyadic grid is a form of *multi-resolution* analysis (MRA). The MRA procedes by splitting a function into nested subspaces of ever descreasing scale. The portion of the function which is removed at each level is projected into a wavelet subspace. With the MRA perspective, the construction of wavelets is based on first solving the *dilation* (or scaling) equation

$$\phi(x) = \sqrt{2}\sum h_k \phi(2x - k) \tag{6.38}$$

for the *scaling* function $\phi(x)$ and then the associated *wavelet* equation

$$\psi(x) = \sqrt{2}\sum g_k \phi(2x - k). \tag{6.39}$$

for the wavelet $\psi(x)$. A solution to the pair of equations (6.38) and (6.39) is constituted by appropriate coefficients $\{g_k\}$ and $\{h_k\}$ which determine the functions of $\phi(x)$ and $\psi(x)$. The structure of the MRA determines the relationship between the g_k and the h_k and, as shown in Section, 6.4, one need in general only solve the scaling equation (6.38) for the coefficients $\{h_k\}$ from which the scaling function $\phi(x)$ may be computed. We shall see that the new wavelets derived from an MRA may not actually be determined in closed form, e.g., the compactly supported family of orthonornal wavelets proposed by Daubechies.

In addition, we may apply constraints to these coefficients g_k, h_k which impact the properties of the wavelet expansion. In fact, the number of vanishing moments p of $\psi(x)$, i.e.,

$$\int_{-\infty}^{\infty} x^m \psi(x) dx = 0$$

for $m = 0, \ldots, p-1$ determines the order of accuracy of the wavelet expansion.

6.3 THE HAAR WAVELET BASIS

The Haar wavelet is defined as

$$\psi(x) = \begin{cases} 1 & \text{if } x \in [0, \frac{1}{2}), \\ -1 & \text{if } x \in [\frac{1}{2}, 1), \\ 0 & \text{otherwise.} \end{cases} \tag{6.40}$$

It is not appropriate for many applications due to the fact it is not smooth, and low order of accuracy of its approximations (it has only one vanishing moment). However, it is an excellent introductory example and many of the ideas central to wavelet analysis are readily apparent in the context of the Haar wavelet.

As already mentioned, the wavelet basis is formed by the dilations and translations of $\psi(x)$, i.e., $\psi_k^j(x) = 2^{-j/2}\psi(2^{-j}x - k)$.

For instance, a simple dilation is

$$\psi(2x) = \begin{cases} 1 & \text{if } 2x \in [0, \frac{1}{2}), \\ -1 & \text{if } 2x \in [\frac{1}{2}, 1), \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} 1 & \text{if } x \in [0, \frac{1}{4}), \\ -1 & \text{if } x \in [\frac{1}{4}, \frac{1}{2}), \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 6.7. *the collection of translations and dilations of the Haar wavelet form an orthonormal basis for* $L^2(\mathbb{R})$.

Proof.

- Step 1. Show
$$(\psi_k^j, \psi_{k'}^{j'}) = \delta_{k,k'}\delta_{j,j'}$$
- Step 2. Show ψ_k^j are dense in $L^2(\mathbb{R})$, i.e., any $f \in L^2$ can be expressed as the superposition of the ψ_k^j to arbitrary precision.

The first part is easy but we need some preliminary definitions.

Definition 6.1. *The support of a function* $f(x)$, *denoted* $supp[f(x)]$, *is the closure of the domain on which* $f(x) \neq 0$.

We shall first show that

$$\text{supp}[\psi_k^j(x)] = [2^j k, 2^j(k+1)].$$

To see this we apply the definition $\psi_k^j(x) = 2^{-j/2}\psi(2^{-j}x - k)$ which may be further evaluated as

$$\psi_k^j(x) = \begin{cases} 2^{-j/2} & \text{if } 2^{-j}x - k \in [0, \frac{1}{2}), \\ -2^{-j/2} & \text{if } 2^{-j}x - k \in [\frac{1}{2}, 1), \\ 0 & \text{otherwise.} \end{cases}$$
$$= \begin{cases} 2^{-j/2} & \text{if } x \in [2^j k, 2^j(k+\frac{1}{2})), \\ -2^{-j/2} & \text{if } x \in [2^j(k+\frac{1}{2}), 2^j(k+1)), \\ 0 & \text{otherwise.} \end{cases}$$

From this we conclude that two Haar wavelets at the same scale don't overlap, i.e.,

$$\text{supp}[\psi_k^j] \cap \text{supp}[\psi_{k'}^j] = \{\emptyset\}$$

At two different scales, $j \neq j'$, overlap is clearly possible. We leave it as an exercise to show that $\text{supp}[\psi_k^j] \subset \text{supp}[\psi_{k'}^{j'}]$ where $j < j'$ and that $\psi_{k'}^{j'}(x) =$ constant for $x \in \text{supp}[\psi_k^j]$. We can view this result graphically by considering the two functions

$$\psi_0^2(x) = \frac{1}{2}\psi(\frac{x}{4}) = \begin{cases} \frac{1}{2} & \text{if } \frac{x}{4} \in [0, \frac{1}{2}), \\ -\frac{1}{2} & \text{if } \frac{x}{4} \in [\frac{1}{2}, 1), \\ 0 & \text{otherwise.} \end{cases}$$
$$= \begin{cases} \frac{1}{2} & \text{if } x \in [0, 2), \\ -\frac{1}{2} & \text{if } x \in [2, 4), \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\psi_1^0(x) = \psi(x-1) = \begin{cases} 1 & \text{if } x-1 \in [0, \frac{1}{2}), \\ -1 & \text{if } x-1 \in [\frac{1}{2}, 1), \\ 0 & \text{otherwise.} \end{cases}$$
$$= \begin{cases} 1 & \text{if } x \in [1, \frac{3}{2}), \\ -1 & \text{if } x \in [\frac{1}{2}, 2), \\ 0 & \text{otherwise.} \end{cases}$$

6.3.1 The Haar Multiresolution Framework

Consider the sequence of *Haar* subspaces V_j where

$$V_j = \{f(x) \in L^2(\mathbb{R}); f(x) = \text{constant on } x \in [2^j k, 2^j(k+1)) \ \forall k \in \mathbb{Z}.\} \tag{6.41}$$

In other words, the V_j are spanned by piecewise constant functions at different scales. Let us consider in particular V_0 where

$$V_0 = \{f(x) \in L^2(\mathbb{R}); f(x) = \text{constant on } x \in [k, k+1), \forall k \in \mathbb{Z}\}$$

In addition, define the *box* function

$$\phi(x) = \begin{cases} 1 & \text{if } x \in [0, 1), \\ 0 & \text{otherwise.} \end{cases}$$

The first observation we can make is that

$$V_0 = \text{span } \{\phi(x-k); k \in \mathbb{Z}\}.$$

In fact, the translates of $\phi(x)$ are an orthonormal basis for V_0 as can be readily seen graphically.

The next observation is that each subspace scales to any other, i.e.,

$$f(x) \in V_0 \Leftrightarrow f(2x) \in V_{-1}$$

This is easily verified:

$$f(x) \in V_0 \rightarrow f(x) = \text{constant for } x \in [k, k+1)$$

$$f(2x) = \text{constant for } 2x \in [k, k+1)$$

$$f(2x) = \text{constant for } x \in [\frac{k}{2}, \frac{k+1}{2})$$

therefore $f(2x) \in V_{-1}$. This generalizes to give the following:

Property 6.3.1. *The Haar subspaces satisfy the scaling property*

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j-1}.$$

We also observe that the subspaces V_j are nested. For instance, since

$$f(x) \in V_0 \rightarrow f(x) = \sum_{k\in\mathbb{Z}} \alpha_k \phi(2x-k) \in V_{-1},$$

so

$$V_0 \subset V_{-1}.$$

This property provides a connection, or link between adjacent Haar subspaces.

Property 6.3.2. *The Haar subspaces are nested, i.e.,*

$$V_j \subset V_{j-1}.$$

with the coarser subspace being a subset of the finer scaled subspace. We also note that the translates of $\phi(2x)$ form an orthonormal basis for V_{-1}. This follows from the scaling property 6.3.1 and the fact that $\phi(x)$ generates a basis for $f(x) \in V_0$. To see this, consider any function $f(x) \in V_{-1}$. Then by the scaling property $f(\frac{x}{2}) \in V_0$ and since the translates $\{\phi(x-k)\}$ form a basis for V_0 we can write

$$f(\frac{x}{2}) = \sum_{k\in\mathbb{Z}} \alpha_k \phi(x-k).$$

Hence, it follows that

$$f(x) = \sum_{k\in\mathbb{Z}} \alpha_k \phi(2x-k)$$

from which we conclude that the translates $\{\phi(2x-k); k \in \mathbb{Z}\}$ form a basis for V_{-1}. This fact is true in general for adjacent Haar subspaces, i.e., the translates $\phi_k^j(x) = 2^{-j/2}\phi(2^{-j}x-k)$ form an orthonormal basis for V_j.

6.3.2 The Haar Wavelet Subspaces

We have seen that piecewise constant functions produce a sequence of nested subspaces which are essentially scaled replicates of each other. We now investigate the *difference* between V_j and V_{j+1} .

To start we recall that $V_0 \subset V_{-1}$. The object is to construct a basis for W_0 where

$$V_{-1} = V_0 \oplus W_0.$$

We recall the the scaling (box) function $\phi(x)$ formed a basis for V_0. Now we determine the wavelet $\psi(x)$ by requiring that its translates generate a basis for W_0 which is orthogonal to V_0. Recall that the box function is given by

$$\phi(x) = \begin{cases} 1 & \text{if } x \in [0,1), \\ 0 & \text{otherwise.} \end{cases}$$

Hence we can write

$$\phi(x) = \phi(2x) + \phi(2x-1)$$

where we note that $\phi(x) \in V_0$ and $\phi(2x), \phi(2x-1) \in V_{-1}$. This is an example of a solution to the dilation equation (6.38) with $h_0 = h_1 = 1/\sqrt{2}$. It relates the box function at different scales and provides the important connection between all the nested Haar subspaces V_j. We will see that it determines the basis for W_0 as well.

The function $\psi(x)$ which generates the basis for W_0 is constructed by the requirement that

$$(\phi(x), \psi(x)) = 0$$

which is satisfied if

$$\psi(x) = \phi(2x) - \phi(2x-1).$$

It is easily verified that this relation reproduces the Haar wavelet previously considered. It is also a solution to the (6.39) with $g_0 = -g_1 = 1/\sqrt{2}$. It

6.3.3 The Haar Multiresolution Analysis

In this Section we demonstrate how the Haar multiresolution framework introduced in the preceding section may be used for efficient computation of the Haar wavelet decomposition and reconstruction.

We now define the projection operators P_j, Q_j such that for any $f(x) \in L^2(\mathbb{R})$

$$P_j f \in V_j \quad \text{and} \quad Q_j f \in W_j.$$

Then we have

$$P_{j-1} f(x) = P_j f(x) + Q_j f(x)$$
$$f^{j-1}(x) = f^j(x) + s^j(x)$$

where $f^j \in V_j$ and $s^j \in W_j$. It is common to view P_j as a *low-pass* filter since it removes the finest scale and Q_j as a *high-pass* filter since all but the finest *detail* is removed. Specifically, we view s^j as the detail component of f^{j-1}.

As expansions we have

$$f^j(x) = P_j f(x) = \sum_{k \in \mathbb{Z}} c_k^j \phi_k^j(x) \tag{6.42}$$

where the $\{c_k^j\}$ are referred to as the *scaling* coefficients and

$$s^j(x) = Q_j f(x) = \sum_{k \in \mathbb{Z}} d_k^j \psi_k^j(x) \tag{6.43}$$

where the $\{d_k^j\}$ are referred to as the *wavelet* coefficients. Next we shall see how all of these coefficients may be calculated very efficiently using recursion once we are given the $\{c_k^j; \forall k \in \mathbb{Z}\}$ for some fixed resolution j.

To compute this first level of scaling coefficients we must in general perform a numerical integration. This is done using the identity

$$c_k^j = (f, \phi_k^j) = \int_{-\infty}^{\infty} f(x) 2^{-j/2} \phi(2^{-j}x - k) dx$$

since

$$\phi(2^{-j}x - k) = \begin{cases} 1 & \text{if } x \in [2^j k, 2^j(k+1)), \\ 0 & \text{otherwise.} \end{cases}$$

which gives

$$c_k^j = 2^{-j/2} \int_{2^j k}^{2^j(k+1)} f(x) dx.$$

6.3.4 Haar Pyramidal Decomposition

The starting point of the multiresolution decomposition are these scaling coefficients $\{c_k^j; \forall k \in \mathbb{Z}\}$ for some fixed resolution j. To determine a recursion relation first write

$$\begin{aligned} P_j f(x) &= \sum_k c_k^j \phi_k^j(x) \\ &= \sum_k c_{2k}^j \phi_{2k}^j(x) + c_{2k+1}^j \phi_{2k+1}^j(x) \end{aligned}$$

where the sum has been split into even and odd parts.

To connect the scaling and wavelet functions at different resolutions we will need the following fact:

Proposition 6.8. *The box scaling function (associated with the Haar wavelet) relates two adjacent Haar subspaces via the relation*

$$\phi_k^j(x) = \frac{1}{\sqrt{2}} (\phi_{2k}^{j-1}(x) + \phi_{2k+1}^{j-1}(x)). \tag{6.44}$$

where $\phi_k^j(x) = 2^{-j/2}\phi(2^{-j}x - k)$. *Also, the Haar wavelet relates adjacent wavelet subspaces via the relation*

$$\psi_k^j(x) = \frac{1}{\sqrt{2}} (\phi_{2k}^{j-1}(x) - \phi_{2k+1}^{j-1}(x)). \tag{6.45}$$

proof: This follows from the dilation equation $\phi(x) = \phi(2x) + \phi(2x-1)$ and the relation $\phi_k^j(x) = 2^{-j/2}\phi(2^{-j}x - k)$. Setting $\xi = 2^{-j}x - k$ we have

$$\phi(\xi) = \phi(2\xi) + \phi(2\xi - 1)$$

from which it follows

$$2^{1/2}2^{-j/2}\phi(2^{-j}x-k) = 2^{-(j-1)/2}\phi(2^{-(j-1)}x-2k) + 2^{-(j-1)/2}\phi(2^{-(j-1)}x-(2k+1))$$

which is exactly the statement of the proposition.

We now use these results to construct a recursion relation between the expansion coefficients at the known resolution and the coefficients at the next lower resolution. Recall

$$P_{j+1}f(x) \in V_{j+1} \subset V_j$$

and that

$$P_{j+1}f(x) = \sum_k c_k^{j+1}\phi_k^{j+1}(x).$$

By orthogonality

$$\begin{aligned} c_k^{j+1} &= (f, \phi_k^{j+1}) \\ &= (f, \frac{\phi_{2k}^j + \phi_{2k+1}^j}{\sqrt{2}}) \\ &= \frac{1}{\sqrt{2}}(f, \phi_{2k}^j) + \frac{1}{\sqrt{2}}(f, \phi_{2k+1}^j) \end{aligned}$$

but these are just the expansion coefficients from before

$$= \frac{1}{\sqrt{2}}c_{2k}^j + \frac{1}{\sqrt{2}}c_{2k+1}^j.$$

The recursion formula

$$c_k^{j+1} = \frac{c_{2k}^j + c_{2k+1}^j}{\sqrt{2}} \tag{6.46}$$

then provides all the scaling coefficients $\{c_k^{j+1}\}$ at the next coarser scaling subspace. The coefficients at a given level $j+1$ are seen to be *smoothed* versions of the coefficients at the higher resolution level j.

We need the corresponding recursion relations for the wavelet coefficients $\{d_k^{j+1}\}$. Now

$$Q_{j+1}f(x) \in W_{j+1} \subset V_j$$

and

$$Q_{j+1}f(x) = \sum_k d_k^{j+1}\psi_k^{j+1}(x).$$

In addition, we have

$$Q_{j+1}f(x) = P_j f(x) - P_{j+1}f(x).$$

Splitting the expansion for the projection onto P_j into its even and odd parts as

$$P_j f(x) = \sum_k c_{2k}^j \phi_{2k}^j(x) + \sum_k c_{2k+1}^j \phi_{2k+1}^j(x)$$

and

$$P_{j+1} f(x) = \sum_k c_k^{j+1} \phi_k^{j+1}(x).$$

Directly substituting the recursion relations given by equations (6.44) and (6.46) into this above equation gives

$$\begin{aligned} P_{j+1} f(x) &= \sum_k \left(\frac{c_{2k}^j + c_{2k+1}^j}{\sqrt{2}}\right)\left(\frac{\phi_{2k}^j(x) + \phi_{2k+1}^j(x)}{\sqrt{2}}\right) \\ &= \frac{1}{2}\sum_k c_{2k}^j \phi_{2k}^j(x) + c_{2k}^j \phi_{2k+1}^j(x) + c_{2k+1}^j \phi_{2k}^j(x) + c_{2k+1}^j \phi_{2k+1}^j(x) \end{aligned}$$

Since

$$\begin{aligned} Q_{j+1} f(x) &= P_j f - P_{j+1} f(x) \\ &= \frac{1}{2}\sum_k c_{2k}^j \phi_{2k}^j(x) + c_{2k+1}^j \phi_{2k+1}^j(x) - c_{2k}^j \phi_{2k+1}^j(x) - c_{2k+1}^j \phi_{2k}^j(x) \\ &= \sum_k \frac{1}{2} c_{2k}^j (\phi_{2k}^j(x) - \phi_{2k+1}^j(x)) - \frac{1}{2} c_{2k+1}^j (\phi_{2k}^j(x) - \phi_{2k+1}^j(x)) \\ &= \sum_k \left(\frac{c_{2k}^j - c_{2k+1}^j}{\sqrt{2}}\right)\left(\frac{\phi_{2k}^j(x) - \phi_{2k+1}^j(x)}{\sqrt{2}}\right) \\ &= \sum_k d_k^{j+1} \psi_k^{j+1}(x) \end{aligned}$$

Thus the recursion equation for the wavelet coefficients is given by

$$d_k^{j+1} = \frac{c_{2k}^j - c_{2k+1}^j}{\sqrt{2}}. \tag{6.47}$$

6.3.5 Haar Pyramidal Reconstruction

The previous Section considers an algorithm for taking a function and computing its Haar wavelet decomposition. In particular, after computing the initial projection of the function onto the Haar scaling subspace of the desired resolution, the algorithm recursively computes the detail of the function as represented by its wavelet coefficients in the corresponding wavelet subspaces. The projection of the original function may be reconstructed from the wavelet coefficients (and the scaling coefficients from the most reduced level) via the formulae

$$c_{2n}^{j-1} = \frac{c_n^j + d_n^j}{\sqrt{2}} \qquad c_{2n+1}^{j-1} = \frac{c_n^j - d_n^j}{\sqrt{2}} \tag{6.48}$$

This is a special case of a more general formula developed later.

An example

We are now in the position to compute a wavelet decomposition of a function. We take as an example the decomposition of the vector $f = (9, 1, 2, 0)$ treated in detail in Strang's paper. Our presentation stresses somewhat different points.

So how do we compute the $\{c_k^j\}$ and $\{d_k^j\}$ for this f? We start by viewing $f \in V_j$ where the V_j are the Haar multiresolution nested subspaces. We need to arbitrarily specify the size of the smallest scale to start the pyramidal decomposition algorithm. We will choose $\dim f = 4 = 2^{-j}$. Thus the finest resolution required is at the level V_{-2}.

Hence we view f as

$$f(x) = \begin{cases} 9 & \text{if } x \in [0, \frac{1}{4}), \\ 1 & \text{if } x \in [\frac{1}{4}, \frac{1}{2}), \\ 2 & \text{if } x \in [\frac{1}{2}, \frac{3}{4}), \\ 0 & \text{if } x \in [\frac{3}{4}, 1). \end{cases}$$

We will see that the decomposition in this case will involve the subspaces

$$V_{-2} = \text{span } \{2\phi(4x - k); k \in \mathbb{Z}\}$$

$$V_{-1} = \text{span } \{\sqrt{2}\phi(2x - k); k \in \mathbb{Z}\}$$

$$V_0 = \text{span } \{\phi(x - k); k \in \mathbb{Z}\}$$

where the coefficients normalize the functions such that orthonormal.

At the finest resolution, $j = -2$, the scaling coefficients $\{c_k^j\}$ are found by projecting f onto the basis for V_{-2}. I.e., since $P_{-2}f \in V_{-2}$ we write

$$P_{-2}f(x) = \sum_{k \in \mathbb{Z}} c_k^{-2} \phi_k^{-2}(x)$$

where

$$c_k^j = 2^{-j/2} \int_{2^j k}^{2^j (k+1)} f(x) dx$$

From which we have

$$c_k^{-2} = (f, \phi_k^{-2}) = 2 \int_{k/4}^{(k+1)/4} f(x) dx.$$

Evaluating the integrals

$$c_0^{-2} = 2 \int_0^{1/4} 9 dx = \frac{9}{2},$$

$$c_1^{-2} = 2\int_{1/4}^{1/2} 1dx = \frac{1}{2},$$

$$c_2^{-2} = 2\int_{1/2}^{3/4} 2dx = 1,$$

$$c_3^{-2} = 2\int_{3/4}^{1} 0dx = 0.$$

Also, $c_k^{-2} = 0$ for $k < 0, k > 3$.

Now project f onto V_{-1} via $P_{-1}f(x)$ and onto W_{-1} via $Q_{-1}f(x)$. We have

$$P_{-1}f(x) = \sum_k c_k^{-1}\phi_k^{-1}(x)$$

where using the recursion equation (6.46)

$$c_k^{-1} = \frac{1}{\sqrt{2}}(c_{2k}^{-2} + c_{2k+1}^{-2})$$

we get

$$c_0^{-1} = \frac{1}{\sqrt{2}}(c_0^{-2} + c_1^{-2}) = \frac{5}{\sqrt{2}}$$

$$c_1^{-1} = \frac{1}{\sqrt{2}}(c_2^{-2} + c_3^{-2}) = \frac{1}{\sqrt{2}}$$

and $c_2^{-1} = c_3^{-1} = 0$. Therefore,

$$\begin{aligned} P_{-1}f(x) &= c_0^{-1}\phi_0^{-1}(x) + c_1^{-1}\phi_1^{-1}(x) \\ &= 5\phi(2x) + \phi(2x-1) \end{aligned}$$

The *detail* is given by

$$Q_{-1}f(x) = \sum_k d_k^{-1}\psi_k^{-1}(x)$$

where $\psi(x)$ is the Haar wavelet. Using equation (6.47) at the level $j = -2$

$$d_k^{-1} = \frac{1}{\sqrt{2}}(c_{2k}^{-2} - c_{2k+1}^{-2}).$$

Hence

$$d_0^{-1} = \frac{1}{\sqrt{2}}(c_0^{-2} - c_1^{-2}) = \frac{4}{\sqrt{2}}$$

$$d_1^{-1} = \frac{1}{\sqrt{2}}(c_2^{-2} - c_3^{-2}) = \frac{1}{\sqrt{2}}$$

Therefore,

$$\begin{aligned} Q_{-1}f(x) &= d_0^{-1}\psi_0^{-1}(x) + d_1^{-1}\psi_1^{-1}(x) \\ &= \frac{4}{\sqrt{2}}\psi_0^{-1} + \frac{1}{\sqrt{2}}\psi_1^{-1} \end{aligned}$$

We can write these in terms of the box function using the relations

$$\psi_0^{-1} = \frac{1}{\sqrt{2}}(\phi_0^{-2} - \phi_1^{-2}) = \frac{1}{\sqrt{2}}(2\phi(4x) - 2\phi(4x-1))$$

and

$$\psi_1^{-1} = \frac{1}{\sqrt{2}}(\phi_2^{-2} - \phi_3^{-2}) = \frac{1}{\sqrt{2}}(2\phi(4x-2) - 2\phi(4x-3))$$

Thus we have the projection onto W_{-1} as

$$Q_{-1}f(x) = 4(\phi(4x) - \phi(4x-1)) + (\phi(4x-2) - \phi(4x-3)).$$

which is the portion of $f(x)$ contained in the wavelet subspace at level $j = -1$.

The projection onto V_0 proceeds similarly via the computation of $P_0 f(x)$. The scaling coefficients at the level $j = 0$ are found using

$$c_k^0 = \frac{1}{\sqrt{2}}(c_{2k}^{-1} + c_{2k+1}^{-1})$$

The single non-zero scaling coefficient is given by

$$c_0^0 = \frac{1}{\sqrt{2}}(c_0^{-1} + c_1^{-1}) = 3$$

and the associated wavelet coefficient is

$$d_0^0 = \frac{1}{\sqrt{2}}(c_0^{-1} - c_1^{-1}) = 2.$$

Now we have

$$P_0 f(x) = c_0^0 \phi_0^0(x) = 3\phi(x)$$

and

$$\begin{aligned} Q_0 f(x) &= d_0^0 \psi_0^0(x) \\ &= 2(\frac{1}{\sqrt{2}}(\phi_0^{-1} - \phi_1^{-1})) \\ &= \frac{2}{\sqrt{2}}(\sqrt{2}\phi(2x) - \sqrt{2}\phi(2x-1)) \\ &= 2(\phi(2x) - \phi(2x-1)) \end{aligned}$$

Note, $f(x) \in V_{-2}$ without approximation. In summary,

$$\begin{aligned} f(x) &= P_{-2}f(x) \in V_{-2} \\ &= P_{-1}f(x) + Q_{-1}f(x) \in V_{-1} \oplus W_{-1} \\ &= P_0 f(x) + Q_0 f(x) + Q_{-1}f(x) \in V_0 \oplus W_0 \oplus W_{-1} \end{aligned}$$

6.4 THE GENERAL MULTIRESOLUTION FRAMEWORK

In the previous sections we saw how the box function and the related Haar wavelet fit naturally into a multiresolution analysis framework. Such a description is distinguished by the fact that a function is represented by a set of subspaces, each of which is just a scaled replication of the other. Now we see that this is a general framework which extends to other types of scaling functions and their corresponding wavelets.

In general, a multiresolution analysis consists of a sequence of closed subspaces with the following six properties:

Property 6.4.1. *The subspaces are nested*

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \ldots \tag{6.49}$$

Property 6.4.2. *The closure of the subspaces is identified with all square integrable functions*

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}) \tag{6.50}$$

Property 6.4.3. *The intersection of the subspaces contains only zero*

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\} \tag{6.51}$$

Property 6.4.4. *Each subspace is a scaled replica of the other*

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j-1} \tag{6.52}$$

Property 6.4.5. *The subspace V_0 is invariant under integer translations,*

$$f(x) \in V_0 \rightarrow f(x-k) \in V_0 \ \ \forall k \in \mathbb{Z}. \tag{6.53}$$

Property 6.4.6. *There exists a scaling function $\phi(x)$ s.t. $\{\phi(x-k); \forall k \in \mathbb{Z}\}$ is an o.n. basis in V_0.*

The power of a multiresolution analysis is that it expresses a function $f \in L^2(\mathbb{R})$ as a limit of successive approximations with increasing resolution.

We will see that a multiresolution analysis generates an o.n. wavelet basis in the sense

$$P_{j-1}f = P_j f + \sum_{k\in\mathbb{Z}} (f, \psi_k^j)\psi_k^j. \tag{6.54}$$

By the completeness property we have

$$\lim_{j\rightarrow-\infty} P_j f = f. \tag{6.55}$$

We shal see that the multiresolution analysis framework provides a general procedure for constructing wavelets. We will write W_j for the wavelet subspace which is the orthogonal complement to V_j in V_{j-1}, i.e.,

$$V_{j-1} = V_j \oplus W_j \tag{6.56}$$

and $W_j \perp W_j'$ if $j \neq j'$. Then the subspace V_j may be written

$$V_j = V_J \oplus W_J \oplus W_{J-1} \oplus \cdots \oplus W_{j+1} \tag{6.57}$$

where J denotes the coarsest resolution subspace. The remarkable aspect of this decomposition is the fact that we have the wavelet decomposition

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j. \tag{6.58}$$

6.5 THE GENERAL PYRAMIDAL ALGORITHM

We have seen how the pyramidal decomposition and reconstruction algorithm works for the specific case of the Haar wavelet. Now we consider the algorithm in its general form. To start we assume that $f(x) \in V_0$. Specifically,

$$f(x) = \sum_k c_k^0 \phi_k^0(x)$$

where $c_k^0 = (f, \phi_k^0)$.

6.5.1 Pyramidal Decomposition

As before, we will project f onto V_1, i.e., the nested subspace at the next lower resolution and contained in V_0. We project the *detail*, i.e., the information contained in V_0 but not in V_1 onto the wavelet subspace W_1, i.e.,

$$V_0 = V_1 \oplus W_1,$$

and

$$f(x) = P_1 f(x) + Q_1 f(x)$$

where $P_1 f \in V_1$ and $Q_1 f \in W_1$. The projection operators are defined as

$$P_1 f(x) = \sum_k c_k^1 \phi_k^1(x)$$

$$Q_1 f(x) = \sum_k d_k^1 \psi_k^1(x)$$

also with $c_k^1 = (f, \phi_k^1)$ and $d_k^1 = (f, \psi_k^1)$.

We now seek to determine the recursion relations for the scaling and wavelet coefficients. Using the definition of the scaling coefficients and orthogonality of the translates of $\phi(x)$ we have

$$c_k^1 = \frac{1}{\sqrt{2}} \int f(x)\phi(\frac{x}{2} - k)dx.$$

Given the dilation equation

$$\phi(x) = \sqrt{2} \sum_l h_l \phi(2x - l)$$

it follows that

$$\phi(\frac{x}{2} - k) = \sqrt{2} \sum_l h_l \phi(x - 2k - l).$$

Therefore

$$c_k^1 = \sum_l h_l \int f(x)\phi(x - 2k - l)dx.$$

Writing $m = 2k + l$ we obtain

$$c_k^1 = \sum_m h_{m-2k} \int f(x)\phi(x - m)dx.$$

The integral is seen to be c_m^0 giving

$$c_k^1 = \sum_m h_{m-2k} c_m^0 \tag{6.59}$$

which is the desired recursion.

To obtain a recursion for the wavelet coefficients $\{d_k^1\}$ we start with

$$\begin{aligned} d_k^1 &= (f, \psi_k^1) \\ &= \frac{1}{\sqrt{2}} \int f(x)\psi(\frac{x}{2} - k)dx \end{aligned}$$

Recalling the wavelet equation

$$\psi(x) = \sqrt{2} \sum_k g_k \phi(2x - k)$$

we have

$$d_k^1 = \sum_l g_l \int f(x)\phi(x - 2k - l)dx$$

substituting $l = m - 2k$

$$d_k^1 = \sum_m g_{m-2k} \int f(x)\phi(x - m)dx$$

Again, recognizing the above integral as c_m^0 it follows that

$$d_k^1 = \sum_m g_{m-2k} c_m^0$$

These recursions are valid between any two neighboring levels of resolution so we can write

$$c_k^{j+1} = \sum_m h_{m-2k} c_m^j \tag{6.60}$$

$$d_k^{j+1} = \sum_m g_{m-2k} c_m^j. \tag{6.61}$$

In the above derivations, we have made extensive use of the fact that the coefficients $\{g_k\}, \{h_k\}$ are determined by the relations $\phi(x) = \sqrt{2}\sum_k h_k \phi(2x-k)$ and $\psi(x) = \sqrt{2}\sum_k g_k \psi(2x-k)$. Later we will consider these in more detail as well as the remarkable relationship between the wavelet and the scaling function

$$g_k = (-)^k h_{1-k+2N}$$

where N is a suitably chosen integer.

6.5.2 Pyramidal Reconstruction

Now we derive the recursion relations going the other way. That is, we start with the function at its coarsest level and add on the detail from each of the wavelet subspaces. At each level we have

$$\begin{aligned} f^{j-1}(x) &= f^j(x) + s^j(x) \\ &= \sum_k c_k^j \phi_k^j(x) + \sum_k d_k^j \psi_k^j(x) \end{aligned}$$

$$\begin{aligned} c_n^{j-1} &= (f^{j-1}, \phi_n^{j-1}) \\ &= (\sum_k c_k^j \phi_k^j + \sum_k d_k^j \psi_k^j, \phi_n^{j-1}) \\ &= \sum_k c_k^j (\phi_k^j, \phi_n^{j-1}) + \sum_k d_k^j (\psi_k^j, \phi_n^{j-1}) \end{aligned}$$

To evaluate the inner products we make use of the dilation equation

$$\phi(2^{-j}x - k) = \sum_l h_l \sqrt{2} \phi(2^{-(j-1)}x - 2k - l)$$

which upon putting $m = 2k + l$ gives

$$\phi(2^{-j}x - k) = \sum_m h_{m-2k} \sqrt{2} \phi(2^{-(j-1)}x - m)$$

from which we conclude

$$\phi_k^j(x) = \sum_m h_{m-2k}\phi_m^{j-1}(x).$$

Now we have all we need to evaluate the inner products above, i.e.,

$$\begin{aligned}(\phi_k^j, \phi_n^{j-1}) &= (\sum_m h_{m-2k}\phi_m^{j-1}, \phi_n^{j-1}) \\ &= \sum_m h_{m-2k}\delta_{mn} \\ &= h_{n-2k}\end{aligned}$$

Using the same approach it can also be shown that

$$(\psi_k^j, \phi_n^{j-1}) = g_{n-2k}.$$

Thus we have the general reconstruction formula

$$c_n^{j-1} = \sum_k h_{n-2k}c_k^j + \sum_k g_{n-2k}d_k^j \tag{6.62}$$

6.5.2.1 The Cascade Algorithm The scaling function $\phi(x)$ may be computed by initializing the general reconstruction equation (6.62) with the values

$$c_n^0 = \delta_{0,n} \quad d_n^j = 0 \; \forall n, j \in \mathbb{Z}$$

It can be shown [16] that the recursion

$$c_n^{j-1} = \sum_k h_{n-2k}c_k^j$$

converges exponentially fast to $\phi(x)$ at the dyadic grid points. We further discuss the computation of the scaling functions and wavelets in the next Section.

6.6 THE DILATION EQUATION

One of the main approaches for constructing a multiresolution analysis is to determine the scaling function $\phi(x)$ which satisfies the dilation equation

$$\phi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k\phi(2x-k)$$

with appropriate side constraints. For example, a particularly attractive solution $\phi(x)$ will generate an o.n. family directly. It is possible, as we shall see, to compute solutions which do not have this property. In these cases the orthonormality must be achieved via an orthogonalization trick and with the penalty that the resulting o.n. family no longer has compact support. It is demonstrated later that any solution $\phi(x)$ associated with a finite number of non-zero $\{h_k\}, k = 0..N$ will have support on the compact interval $[0, N]$.

6.6.1 Normalization of $\phi(x)$

By convention we require that

$$\int_{-\infty}^{\infty} \phi(x)dx = 1 \tag{6.63}$$

In practice it is sufficient that the right-hand side be non-zero and finite. As we shall see, this restriction ensures that for a given collection of coefficients $\{h_k\}$ there is a unique $\phi(x)$. Hence, we may view a solution of the dilation equation to consist of these coefficients which uniquely determine the scaling function as long as equation (6.63) holds.

As a consequence of the normalization given by equation (6.63), the scaling function $\phi(x) \in L^1(\mathbb{R})$. Hence the Fourier transform $\tilde{\phi}(\omega)$ is continuous. Also, the Fourier transform evaluated at zero is given by $\tilde{\phi}(0) = \int \phi(x)dx$ from which we conclude

$$\tilde{\phi}(0) = 1 \tag{6.64}$$

We will often require this fact, in particular the property $\tilde{\phi}(0) \neq 0$.

6.6.1.1 Normalization of the $\{h_k\}$ The properties and constraints of the scaling function readily translate into equations involving the coefficients $\{h_k\}$. The normalization constraint given in equation (6.63) provides us with our first example. Integrating the dilation equation

$$\int \phi(x)dx = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \int \phi(2x-k)dx$$

Letting $x' = 2x - k$ and substituting equation (6.63) gives

$$1 = \sqrt{2}\sum_{k} \int \phi(x')\frac{dx'}{2}$$

After substituting equation (6.63) again, we obtain the normalization property in terms of the filter coefficients

$$\sum_{k\in\mathbb{Z}} h_k = \sqrt{2} \tag{6.65}$$

6.6.1.2 Examples Here are several examples of solutions to the dilation equation.

Example 6.3.

$$h_0 = \sqrt{2}, \quad h_j = 0 \ \ \forall j \neq 0$$

In other words $\phi(x) = 2\phi(2x)$. This function $\phi(x)$ should be non-zero only at $x = 0$. We see in Section 6.68 that this function is in fact the delta function.

Example 6.4.

$$h_0 = h_1 = \frac{1}{\sqrt{2}}, \quad h_j = 0 \ \ \forall j \neq \{0, 1\} \tag{6.66}$$

This is the Haar scaling function.

Example 6.5.

$$h_0 = \frac{1}{2\sqrt{2}}, \quad h_1 = \frac{1}{\sqrt{2}}, \quad h_2 = \frac{1}{2\sqrt{2}}$$

and the remaining $h_j = 0 \ \ \forall j \neq \{0, 1, 2\}$. This scaling function corresponds to an equilateral triangle of unit height of width 2, centered at $x = 1$. This does not directly produce an o.n. family.

Example 6.6.

$$h_0 = \frac{1}{4}, \quad h_1 = h_2 = \frac{3}{4}, \quad h_3 = \frac{1}{4}$$

and the remaining $h_j = 0 \forall j \neq \{0, 1, 2, 3\}$. This scaling function corresponds to the quadratic cardinal B-spline. It has support $[0, 3]$ and does not directly produce an o.n. family.

Example 6.7. This important example, due to Dabechies, has filter coefficients

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \quad h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \quad h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, \quad h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$

and the remaining $h_j = 0 \forall j \neq \{0, 1, 2, 3\}$ The corresponding function $\phi(x)$ has compact support $[0, 3]$ and produces an o.n. family and compactly supported wavelets. It is continuous but only weakly differentiable.

6.6.2 Othogonality Constraint on the $\{h_k\}$

The orthogonality of the $\phi(x-k)$ gives another condition that the $\{h_k\}$ must satisfy. Recall that

$$\phi(x) = \sqrt{2} \sum_k h_k \phi(2x - k)$$

so the translates may be represented as

$$\phi(x - m) = \sqrt{2} \sum_k h_k \phi(2x - 2m - k).$$

Substituting these equations into the orthogonality condition

$$\int \phi(x)\phi(x - m)dx = \delta_{0m}$$

we have

$$\int \phi(x)\phi(x-m)dx = \int \sqrt{2}\sum_k h_k \phi(2x-k)\sqrt{2}\sum_{k'} h_{k'}\phi(2x-2m-k')dx$$
$$= 2\sum_{kk'} h_k h_{k'} \frac{\delta_{k,2m+k'}}{2}$$

Thus the orthogonality property of the scaling function is expressed in terms of the coefficients solving the dilation equation as

$$\sum_k h_k h_{k-2m} = \delta_{m0}. \tag{6.67}$$

Note that out of all the previous examples, only the Haar and Daubechies' scaling functions satisfy the o.n. constraint.

6.6.3 Zero Moment Constraints on the $\{h_k\}$

The following theorem relates how the number of zero moments of the wavelet basis is formulated as a condition on the $\{h_k\}$.

Theorem 6.1. *Let $\phi(x)$ and $\psi(x)$ be a scaling and wavelet function which generate an MRA $\{V_j\}$ as well as the orthogonal subspaces W_j. Then*

$$\int x^m \psi(x)dx = 0$$

for $m = 0, \ldots, p-1$ iff

$$\sum_k (-)^k k^m h_k = 0$$

The proof of this theorem (see Strang) requires that $\phi(x)$ and $\psi(x)$ decay faster than $O(|x|^{-m-1})$. In practice this is no problem since $\phi(x)$ and $\psi(x)$ have either compact support or decay exponentially.

Furthermore, if the above theorem holds for $m = 0, 1, \ldots, p-1$, then

- $1, x, \ldots, x^{p-1}$ are spanned by the $\{\phi(x-k)\}$.
- $O(h^p)$ accuracy, i.e.,

$$\|f - \sum_k a_k \phi(2^j - k)\| \leq C 2^{-jp} \|f^{(p)}\|$$

- The wavelet coefficients decay as

$$\int f(x)\psi(2^j x)dx \leq C 2^{-jp}$$

6.6.4 Daubechies Compactly Supported Orthonormal Wavelets

The main drawback of the Haar wavelet is that it is not smooth. Let's look for a multiresolution analysis based on having four non-zero coefficients h_0, h_1, h_2, h_3 which has 2 non-zero moments. The orthonormality condition given in equation (6.65) with 4 non-zero coefficients requires that

$$h_0^2 + h_1^2 + h_2^2 + h_3^2 = 1$$

with $m = 0$ and

$$h_0 h_2 + h_3 h_1 = 0$$

with $m = -1$. Further, if we require an order of approximation $p = 2$ we have

$$h_0 - h_1 + h_2 - h_3 = 0$$

$$-h_1 + 2h_2 - 3h_3 = 0$$

We leave it as an exercise to verify that the coefficients in example 6.7 satisfy the above constraints.

6.6.5 Fourier Analysis of $\phi(x)$

The Fourier transform of the dilation equation $\phi(x) = \sqrt{2}\sum_k h_k \phi(2x - k)$ is given by

$$\tilde{\phi}(\omega) = \sqrt{2}\sum_k h_k \frac{1}{2}\tilde{\phi}(\frac{\omega}{2})e^{-ik\omega/2}$$

Introducing the *symbol* $P(\omega)$ defined as

$$P(\omega) = \frac{1}{\sqrt{2}}\sum_k e^{-ik\omega} \tag{6.68}$$

hence

$$\tilde{\phi}(\omega) = \tilde{\phi}(\frac{\omega}{2})P(\frac{\omega}{2}) \tag{6.69}$$

By direct substitution we see that $P(0) = 1$.

Proposition 6.9.

$$\tilde{\phi}(\omega) = \tilde{\phi}(\frac{\omega}{2^N})\Pi_{j=1}^N P(\frac{\omega}{2^j}) \tag{6.70}$$

Proposition 6.10.

$$\tilde{\phi}(\omega) = \lim \Pi_{j=1}^N P(\frac{\omega}{2^j}) \tag{6.71}$$

if the limit exists.

This is a consequence of the continuity of $\tilde{\phi}(\omega)$ in addition to the fact that $\tilde{\phi}(0) = 1$.

6.6.6 Iteration of the Dilation Equation

Consider the algorithm

$$\phi^{(n)}(x) = \sqrt{2}\sum_{k=0}^{N} h_k \phi^{(n-1)}(x)(2x-k) \tag{6.72}$$

where $\phi^{(0)}(x) \in L^1$, in particular $\int \phi^{(0)}(x) = 1$.

Proposition 6.11.

$$\phi(x) = \lim_{n\to\infty} \phi^{(n)}(x)$$

The Fourier transform of the algorithm reveals how it is working.

Proposition 6.12.

$$\tilde{\phi}^{(n)}(\omega) = \tilde{\phi}^{(0)}(\frac{\omega}{2^n})\Pi_{j=1}^{n} P(\frac{\omega}{2^j}) \tag{6.73}$$

Then

$$\lim \tilde{\phi}^{(0)}(\frac{\omega}{2^n}) = 1$$

From proposition 6.71 we know that the limit of the product of symbols is the transform of the scaling function.

6.6.6.1 Support of the Scaling Function The above iteration algorithm provides a method to compute the support of the scaling function $\phi(x)$ associated with a finite length filter.

Proposition 6.13. *If the solution to the scaling equation has all zero coefficients except possibly for $\{h_k\}$ where $k = 0, \ldots, N$ then supp$(\phi(x)) = [0, N]$.*

6.6.7 Computation of $\phi(x)$ on the Dyadic Grid

The scaling equation may have solutions which are not representable as functions in closed form. In these situations, the dilation equation can be used for the computation of $\phi(x)$ at the points $x/2$ given that the values $\phi(x)$ are known. Recursively, ϕ may be computed at all the points $x/2^j$. We demonstrate this idea be means of an example.

Let

$$\phi(\frac{x}{2}) = \sqrt{2}(h_0\phi(x) + h_1\phi(x-1) + h_2\phi(x-2) + h_3\phi(x-3))$$

Then the support of $\phi(x)$ is $[0, 3]$. Also, by continuity, we have that $\phi(0) = \phi(3) = 0$.

Now observe that given the values of the scaling function at the integers $\{\phi(j), j \in \mathbb{Z}\}$ the above formula may be used to compute ϕ at the half-integers, i.e., the set $\{\phi(\frac{j}{2}), j \in \mathbb{Z}\}$.

6.7 WAVELETS FROM THE SCALING FUNCTION

In this section we outline the connection between the scaling function $\phi(x)$ of a multiresolution analysis and the mother wavelet $\psi(x)$. We have alluded to the fact that the scaling function can be used to construct the wavelet which generates the bases for all the W_j. Most of the deliberations take place in Fourier space. We begin by recalling that in the Fourier transform of the scaling equation (6.38) is given by

$$\tilde{\phi}(\omega) = \tilde{\phi}(\frac{\omega}{2})P(\frac{\omega}{2}) \tag{6.74}$$

where

$$P(\omega) = \frac{1}{\sqrt{2}}\sum_k h_k e^{-ik\omega} \tag{6.75}$$

Also, the Fourier transform of the two-scale wavelet equation is (6.39)

$$\tilde{\psi}(\omega) = \tilde{\phi}(\frac{\omega}{2})Q(\frac{\omega}{2}) \tag{6.76}$$

where

$$Q(\omega) = \frac{1}{\sqrt{2}}\sum_k g_k e^{-ik\omega} \tag{6.77}$$

6.7.1 Orthonormality Conditions in Fourier Space

In this Section we derive the orthonormality conditions on the scaling and wavelet families in the Fourier domain.

Proposition 6.14. *A set of functions* $\{\phi(x-k); \forall k \in \mathbb{Z}\}$ *form an o.n. family iff*

$$\sum_{k\in\mathbb{Z}} |\tilde{\phi}(\omega + 2\pi k)|^2 = 1. \tag{6.78}$$

proof: By Parseval's theorem the orthonormality condition

$$\int_{-\infty}^{\infty} \phi(x)\phi(x-m)dx = \delta_{m0}$$

may be written in the Fourier domain as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{\psi}(\omega)\overline{e^{-im\omega}\tilde{\phi}(\omega)}d\omega = \delta_{m0}$$

Breaking this integral into 2π intervals

$$\frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\int_{2\pi k}^{2\pi(k+1)} e^{im\omega}|\tilde{\phi}(\omega)|^2 d\omega = \delta_{m0}$$

Now substituting $\omega' = \omega - 2\pi k$ we obtain

$$\frac{1}{2\pi}\int_0^{2\pi} e^{im\omega}\sum_{k\in\mathbb{Z}}|\tilde{\phi}(\omega'+2\pi k)|^2 d\omega = \delta_{m0}$$

This is an equation for the Fourier series expansion coefficients $x(n) = \frac{1}{2\pi}\int_0^{2\pi} X(\omega)e^{im\omega}d\omega$ where the 2π periodic function $|\tilde{\psi}(\omega'+2\pi k)|^2 = X(\omega) = \sum_n x(n)e^{-in\omega}$. Also, $x(n) = \delta_{n0}$. Substituting this coefficient into the expansion provides the desired result, i.e.,

$$\sum_{k\in\mathbb{Z}}|\tilde{\phi}(\omega+2\pi k)|^2 = 1.$$

We have largely ignored the fine print in this derivation. In particular we note that if $\phi(x) \in L^1$ then we may infer that the transform $\tilde{\psi}(\omega)$ is continuous and equation (6.78) holds for all ω. If this is not the case then it holds almost everywhere, i.e., a.e. Also, the existence of a converging Fourier series representation is linked to decay of the scaling function, see Chui for details.

Note that proposition 6.14 actually holds for any o.n. family, not just solutions to the scaling equation. Hence the orthonormality condition for the translates of the wavelet is given by

Proposition 6.15. *A set of functions $\{\psi(x-k); \forall k \in \mathbb{Z}\}$ form an o.n. family iff*

$$\sum_{k\in\mathbb{Z}}|\tilde{\psi}(\omega+2\pi k)|^2 = 1. \tag{6.79}$$

6.7.1.1 Orthonormalization Trick Proposition 6.14 actually holds for any o.n. family, not just solutions to the scaling equation. If we let

$$N(\omega) = \sum_{k\in\mathbb{Z}}|\tilde{\phi}(\omega+2\pi k)|^2$$

we may conclude that the family of functions is o.n. iff $N(\omega) = 1$. If $N(\omega) \neq 1$ the we may define a new family of functions

$$\tilde{\phi}^{\perp}(\omega) = \frac{\tilde{\phi}(\omega)}{\sqrt{N(\omega)}} \tag{6.80}$$

which is o.n. as long as $N(\omega) \neq 0 \ \forall\omega$. Note that if the original non o.n. family has compact support then the new family obtained via this orthonormalization procedure will not.

Proposition 6.16. *The set of functions* $\{\phi(x-k); \forall k \in \mathbb{Z}\}$ *is orthogonal to the set* $\{\phi(x-k); \forall k \in \mathbb{Z}\}$ *iff*

$$\sum_{k\in\mathbb{Z}} \tilde{\psi}(\omega + 2\pi k) = 0. \tag{6.81}$$

6.7.2 O.N. Families and 2-Scale Equations

Prepared with the results of the previous Section we can now derive 3 fundamental relationships concerning the filters $P(\omega)$ and $Q(\omega)$. These will be combined with the prefect reconstruction constraint to ultimately derive the relationship between $\{g_k\}$ and $\{h_k\}$. We first consider a result concerning $P(\omega)$.

Proposition 6.17. *If the set of functions* $\{\phi(x-k); \forall k \in \mathbb{Z}\}$ *form an o.n. family and are solutions to the dilation equation* $\phi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \phi(2x-k)$ *then*

$$|P(\omega)|^2 + |P(\omega+\pi)|^2 = 1 \tag{6.82}$$

where $P(\omega)$ *is defined by equation (6.75).*

Substituting equation (6.74) into equation (6.78) gives

$$\sum_{k\in\mathbb{Z}} |P(\frac{\omega+2\pi k}{2})|^2 |\tilde{\phi}(\frac{\omega+2\pi k}{2})|^2 = 1.$$

We can employ the periodicity of $P(\omega)$ by splitting this last sum into even and odd parts,

$$\begin{aligned}&\sum_{k\in\mathbb{Z}} |P(\omega+2k\pi)|^2 |\tilde{\phi}(\omega+2k\pi)|^2 \\ &+ \sum_{k\in\mathbb{Z}} |P(\omega+(2k+1)\pi)|^2 |\tilde{\phi}(\omega+(2k+1)\pi)|^2 = 1\end{aligned}$$

where we have replaced $\omega/2$ with ω. Using $P(\omega) = P(\omega+2\pi k)$ and $P(\omega+\pi) = P(\omega+\pi+2\pi k)$ we have

$$\begin{aligned}&|P(\omega)|^2 \sum_{k\in\mathbb{Z}} |\tilde{\phi}(\omega+2k\pi)|^2 \\ &+ |P(\omega+\pi)|^2 \sum_{k\in\mathbb{Z}} |\tilde{\phi}(\omega+(2k+1)\pi)|^2 = 1.\end{aligned}$$

But each of these sums is equal to 1 by equation (6.78) thus the above simplifies to the desired result

$$|P(\omega)|^2 + |P(\omega+\pi)|^2 = 1.$$

We compare this with the (equivalent) constraint

$$\sum_k h_k h_{k-2m} = \delta_{m0}$$

on the coefficients of the dilation equation.

Proposition 6.18. *If the set of functions* $\{\psi(x-k); \forall k \in \mathbb{Z}\}$ *form an o.n. family and are solutions to the wavelet equation* $\psi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} g_k \phi(2x-k)$ *then*

$$|Q(\omega)|^2 + |Q(\omega+\pi)|^2 = 1 \tag{6.83}$$

where $Q(\omega)$ *is defined by equation (6.77).*

The proof of this proposition follows the previous one and is left as an exercise.

Proposition 6.19. *If the functions* $\{\phi(x-k)\}$ *and* $\{\psi(x-k)\}$ *are orthogonal* $\forall k \in \mathbb{Z}$ *and satisfy the two-scale equations (6.38) and (6.39), respectively, then*

$$P(\omega)\overline{Q(\omega)} + P(\omega+\pi)\overline{Q(\omega+\pi)} = 0 \tag{6.84}$$

where $P(\omega), Q(\omega)$ *are defined by equations (6.75, 6.77).*

The proof of this proposition follows the previous two and is left as an exercise.

6.7.3 The Wavelet Decomposition in the Fourier Domain

After laying the ground-work in the previous Sections, we are now ready to derive the relationship between the functions $\phi(x)$ and $\psi(x)$ which generate an MRA and the orthogonal wavelet subspaces. This result, as you might now anticipate, is derived in the Fourier domain and is then inverted.

Let

$$a(x) = f(x) + s(x)$$

where $a(x) \in V_{-1}, f(x) \in V_0$ and $s(x) \in W_0$. As expansions we have

$$\sqrt{2}\sum_k a_k \phi(2x-k) = \sum_k c_k \phi(x-k) + \sum_k d_k \psi(x-k)$$

Upon taking the Fourier transform

$$\sqrt{2}\sum_k a_k \frac{1}{2} e^{-ik\frac{\omega}{2}} \tilde{\phi}(\frac{\omega}{2}) = \sum_k c_k e^{-ik\omega}\tilde{\phi}(\omega) + \sum_k d_k e^{-ik\omega}\tilde{\psi}(\omega)$$

Hence,

$$A(\frac{\omega}{2})\tilde{\phi}(\frac{\omega}{2}) = C(\frac{\omega}{2})\tilde{\phi}(\omega) + D(\frac{\omega}{2})\tilde{\psi}(\omega)$$

where

$$A(\omega) = \frac{1}{\sqrt{2}} \sum_k a_k e^{-i\omega k}$$

$$C(\omega) = \sum_k c_k e^{-i2\omega k}$$

and

$$D(\omega) = \sum_k d_k e^{-i2\omega k}$$

Note that $A(\omega)$ is 2π periodic and $C(\omega)$, $D(\omega)$ are π periodic.

Substituting equations (6.76)) and (6.74) into

$$A(\omega)\tilde{\phi}(\omega) = C(\omega)\tilde{\phi}(2\omega) + D(\omega)\tilde{\psi}(2\omega)$$

gives

$$A(\omega)\tilde{\phi}(\omega) = C(\omega)\tilde{\phi}(\omega)P(\omega) + D(\omega)\tilde{\phi}(\omega)Q(\omega)$$

which upon simplication becomes

$$A(\omega) = C(\omega)P(\omega) + D(\omega)Q(\omega)$$

Evaluating this at $\omega + \pi$ as well and putting the system into matrix form gives

$$\begin{pmatrix} P(\omega) & Q(\omega) \\ P(\omega+\pi) & Q(\omega+\pi) \end{pmatrix} \begin{pmatrix} C(\omega) \\ D(\omega) \end{pmatrix} = \begin{pmatrix} A(\omega) \\ A(\omega+\pi) \end{pmatrix}$$

Given the requirement that the transformation be unitary we conclude

$$|P(\omega)|^2 + |Q(\omega)|^2 = 1 \tag{6.85}$$

This equation states that the original function $A(\omega)$ may be mapped to $C(\omega), D(\omega)$, and that this process can be reversed without loss.

Note that equations (6.85), (6.82) and (6.84) are compatible for

$$Q(\omega) = e^{-i\omega(1+2N)}\overline{P(\omega+\pi)} \tag{6.86}$$

where $N \in \mathbb{Z}$ Hence, we conclude

$$g_k = (-1)^k h_{1-k+2N}$$

6.7.3.1 The Haar wavelet revisited We have seen a given wavelet may be written in terms of its corresponding scaling function as

$$\psi(x) = \sum_k (-)^k h_{1-k} \phi_k^{-1}(x).$$

Taking $\phi(x)$ as the box scaling function we seek a solution to the dilation equation

$$\phi(x) = \sqrt{2} \sum_k h_k \phi(2x - k).$$

Since

$$(\phi_k^{-1}, \phi_m^{-1}) = \delta_{km}$$

$$h_k = (\phi, \phi_k^{-1}) = \sqrt{2}\int_{-\infty}^{\infty} \phi(x)\phi(2x-k)dx$$

so

$$h_0 = h_1 = \frac{1}{\sqrt{2}}$$

and $h_k = 0$ if $k \neq 0, 1$. Since

$$\psi(x) = \sum_k (-)^k h_{1-k}\phi_k^{-1}(x)$$

the only non-zero terms are given by

$$h_0 = h_{1-k} \rightarrow k = 1$$

$$h_1 = h_{1-k} \rightarrow k = 0$$

therefore,

$$\psi(x) = -h_0\phi_1^{-1} + h_1\phi_0^{-1}.$$

Recalling that $\phi_k^j(x) = 2^{-j/2}\phi(2^{-j}x - k)$ we have

$$\begin{aligned}\psi(x) &= \frac{1}{\sqrt{2}}(-\sqrt{2}\phi(2x-1) + \sqrt{2}\phi(2x)) \\ &= \phi(2x) - \phi(2x-1)\end{aligned}$$

as expected.

Proposition 6.20. *Assume the Fourier transform of the function $\psi(x)$ is given by*

$$\tilde{\psi}(\omega) = e^{i\omega}\overline{P(\frac{\omega}{2}+\pi)}\tilde{\phi}(\frac{\omega}{2}).$$

Then the set of functions $\{\psi(x-k); \forall k \in \mathbb{Z}\}$ form an o.n. basis for W_0.

Again, this proof can be broken in to two parts. First we must show the $\psi_k^0(x)$ are o.n.. To see this

$$\begin{aligned}(\psi(x), \psi_k^0(x)) &= (\tilde{\psi}(\omega), \tilde{\psi}_k^0(\omega)) \\ &= \int_{-\infty}^{\infty} \tilde{\psi}(\omega)e^{ik\omega}\overline{\tilde{\psi}(\omega)}d\omega \\ &= \int_{-\infty}^{\infty} e^{ik\omega}|\tilde{\psi}(\omega)|^2 d\omega \\ &= \int_0^{2\pi} e^{ik\omega}\sum_{l\in\mathbb{Z}}|\tilde{\psi}(\omega+2l\pi)|^2 d\omega\end{aligned}$$

Hence,

$$\sum_{l\in\mathbb{Z}} |\tilde{\psi}(\omega + 2l\pi)|^2 d\omega = 1$$

To show completeness of the wavelets see Daubechies [16].

Problems

6.1 Derive equation (6.5).

6.2 Show proposition (6.4).

6.3 Establish the identity (6.12).

6.4 Compute the integral identity (6.15).

6.5 Prove proposition (6.4).

6.6 Prove proposition (6.6).

6.7 Find the CWT of $x(t) = \cos(\omega_0 t) + \delta(t - t_0)$.

6.8 Derive the condition given by equation (6.31) which determines the influence of an interval of frequencies on the value CWT $X(b, a)$ at the point (b_0, a_0). Draw this region of influence in the (b, a)-plane.

6.9 This problem concerns computing the maximum value of the magnitude of $X(b, a)$ for an input signal of the form $x(t) = \exp(i\omega_0 t)$. For simplicity assume that $a, \omega_0 > 0$

a) Show that for the mexican hat wavelet $\psi(t) = \frac{2}{\sqrt{3}}\pi^{-\frac{1}{4}}(1-t^2)\exp(-t^2/2)$ that the value of a corresponding to a peak in the transform is given by $a = \frac{1}{\omega_0}\sqrt{\frac{5}{2}}$.

b) Show that for Morlet's wavelet $\psi(t) = \frac{1}{\sqrt{2\pi}}\exp(i\alpha t - t^2/2)$ that $a = \frac{1}{2\omega_0}(\alpha + \sqrt{\alpha^2 + 2})$.

6.10 A wavelet is said to be progressive if its Fourier transform has the property that

$$\tilde{\psi}(\omega) = 0 \quad if \quad \omega < 0$$

Consider the signal $x(t) = \sin(\omega_0 t)$. Show that if the CWT $X(b, a)$ of this signal is computed using a progressive wavelet then the magnitude of the CWT $|X(b, a)|$ is independent of b.

6.11 *Constant Q analysis.* Assume that the Fourier transform of a wavelet has finite bandwidth which is non-zero on the interval $\omega \in [\omega_{\min}, \omega_{\max}]$. Noting that the width of the window is a function of the scale a of the wavelet and that the center frequency may be defined as geometric mean of the edges of the window compute the (*constant*) Q factor of the CWT.

Computer Projects

6.12 Write a computer code to implement the discretization of the CWT as described in Section 6.1.6. Compute the approximate CWT by evaluating the integrals numerically using Simpson's rule with a step size of $h = .01$. Find the CWT of $x(t) = \sin(5t) + \sin(\sqrt{26}t)$ and compare with the analytically calculated result. In this problem you may employ either Morlet's wavelet or the mexican hat wavelet. Take $a_0 = 2^{1/12}, b_0 = 1$ and compute $\|X_{mn}\|$ for enough scales and time steps to develop a good picture of the scalogram.

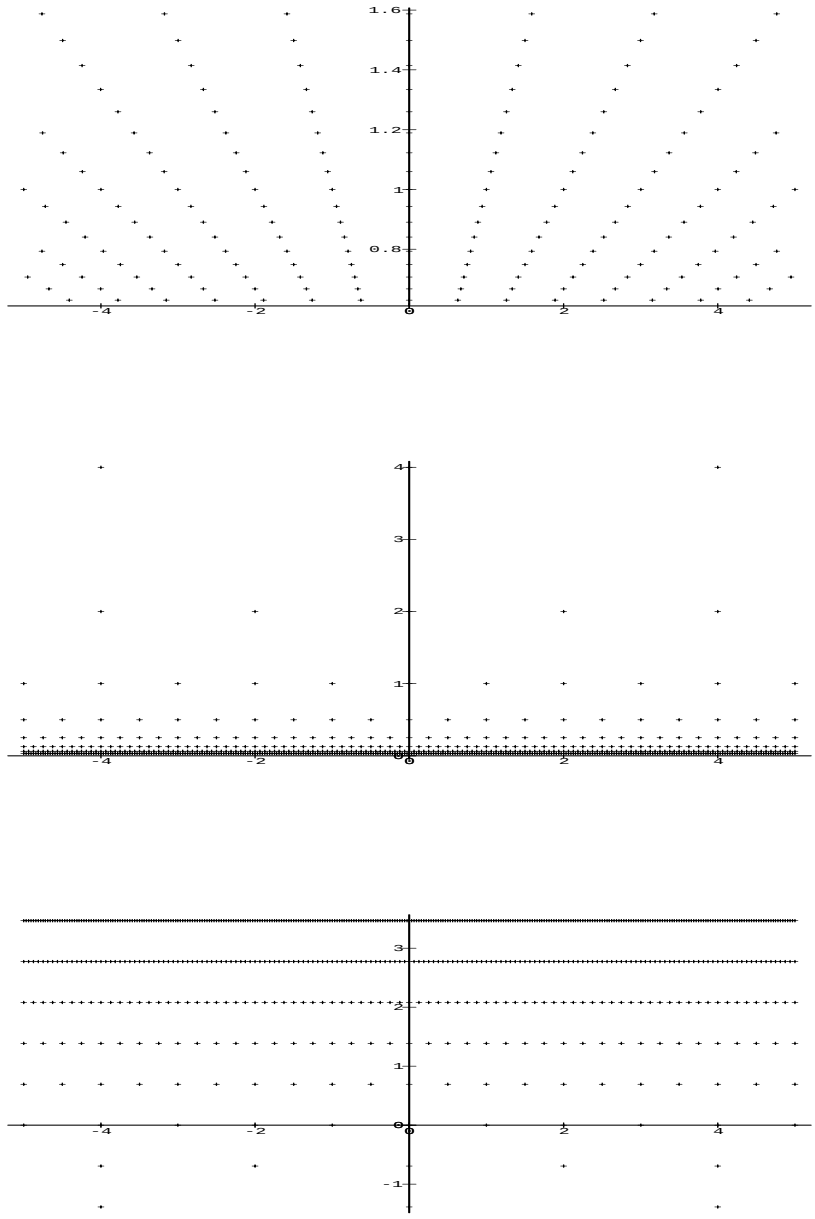

Fig. 6.6 Top: The sampling grid in the (b, a)-plane corresponding to $a_0 = 2^{1/12}, b_0 = 1$. Middle: The dyadic sampling grid with $a_0 = 2, b_0 = 1$. Bottom: The logarithmic dyadic sampling grid $(b, -\ln a)$.

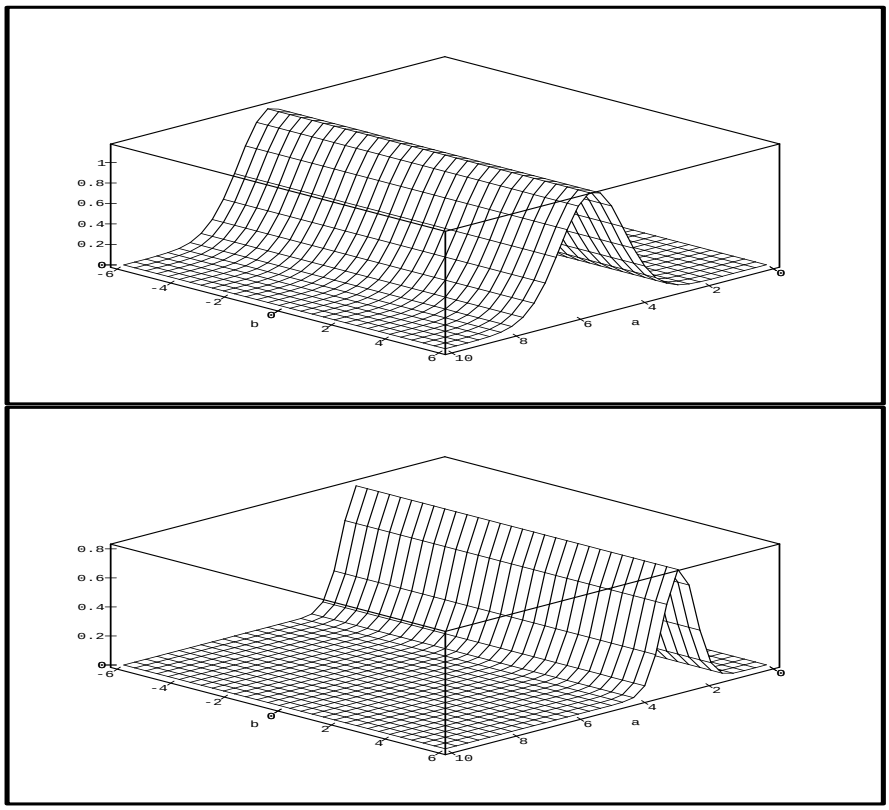

Fig. 6.7 Top: The scalogram of the *progessive* Morlet wavelet transform of the monochromatic signal $x(t) = \sin(t)$. Bottom: The scalogram of the *progessive* Morlet wavelet transform of the monochromatic signal $x(t) = \sin(2t)$.

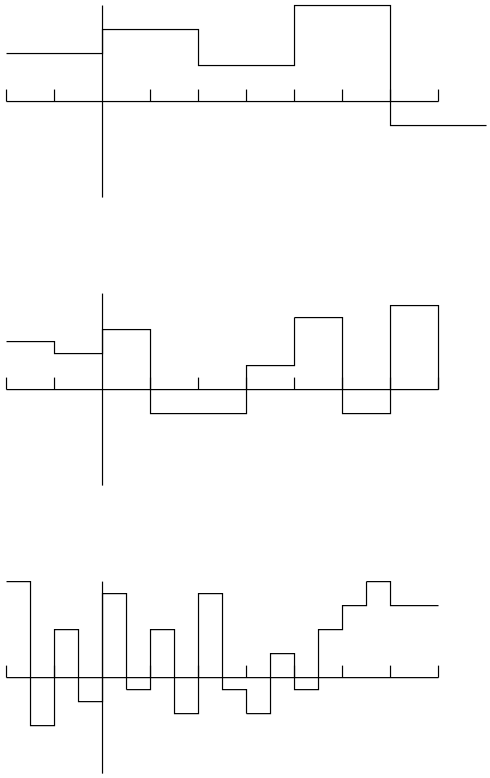

Fig. 6.8 Elements in different Haar subspaces.

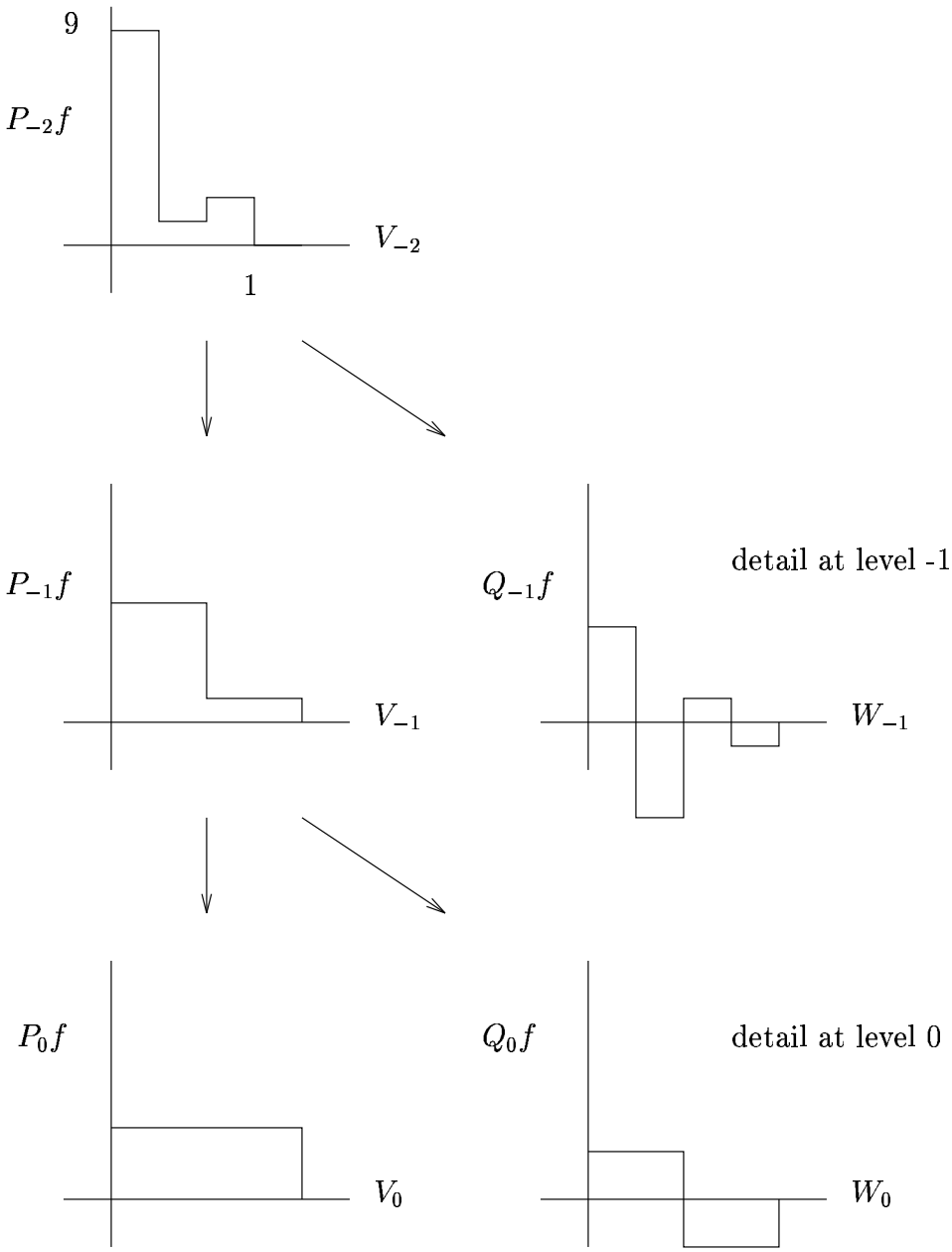

Fig. 6.9 The wavelet decomposition.

Part IV

Adaptive Nonlinear Mappings

7

Radial Basis Functions

8

Neural Networks

9

Nonlinear Reduction Architectures

Appendix J Mathemtical Preliminaries

References

1. R.V. Abadi, D. S. Broomhead, R.A. Clement, J.P. Whittle, and R. Worfolk. Dynamical systems analysis: A new method of analysing congenital nystagmus waveforms. Technical Report 97-3, UMIST Applied Mathematics, 1997.

2. Tom. M. Apsotol. *Mathematical Analysis*. Addison-Wesley, Reading, MA, 1974.

3. N. Aubry, R. Guyonnet, and R. Lima. Spatio-temporal symmetries and bifurcations via bi-orthogonal decomposition. *Journal of Nonlinear Science*, 2:183–215, 1992.

4. B. Mueller and J.Reinhardt. *Neural Networks: An Introduction*. Springer-Verlag, Berlin, 1990.

5. Andrew R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Neural Networks*, 39(3):930–945, 1993.

6. Andrew R. Barron. Approximation and estimation bounds for artificial neural networks. *Machine Learning*, 14:115–133, 1994.

7. Gal Berkooz, Philip Holmes, and John L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flow. *Annu. Rev. Fluid Mech.*, 25:539–575, 1993.

8. Christopher M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, Oxford, U.K., 1995.

9. D. S. Broomhead, R. Jones, and G. P. King. Topological dimension and local coordinates from time series data. *J. Phys. A: Math. Gen*, 20:L563–L569, 1987.

10. David Broomhead and M. Kirby. New approach for dimensionality reduction: Theory and algorithms. *accepted, SIAM J. of Applied Mathematics*, 1998.

11. D.S. Broomhead, R. Indik, A.C. Newell, and D.A. Rand. Local adaptive Galerkin bases for large-dimensional dynamical systems. *Nonlinearity*, 4:159–197, 1991.

12. D.S. Broomhead and David Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

13. C. Canuto, M.Y. Hussaini, A. Quarteroni, and T.A. Zang. *Spectral Methods in Fluid Dynamics.* Springer Series in Computational Physics. Springer-Verlag, 1988.

14. P. Constantin, C. Foias, B. Nicolaenko, and R. Temam. *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations.* Springer-Verlag, 1989.

15. G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems*, 2:303–314, 1989.

16. Ingrid Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, PA, 1992.

17. P.A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach.* Prentice Hall, 1982.

18. Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, New York, 1973.

19. R. M. Everson and L. Sirovich. The karhunen-loeve transform for incomplete data. *Journal of the Optical Society of America, A*, 12(8):1657–1664, 1995.

20. Kenneth Falconer. *Fractal Geometry: Mathematical Foundations and Applications.* John Wiley and Sons, 1990.

21. K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, Boston, MA, 1990.

22. K. Fukunaga and D.R. Olsen. An algorithm for finding intrinsic dimensionality of data. *IEEE Transactions on Computers*, C-20(2):176–183, 1971.

23. Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* Johns Hopkins, Baltimore, third edition, 1996.

24. Andrew A. Green, Mark Berman, Paul Switzer, and Maurice D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65–74, January 1988.

25. A. Grossmann, R. Kronland-Martinet, and J. Morlet. Reading and understanding the continuous wavelet transforms. In J.M. Combes, A. Grossmann, and Ph. Tchamitchian, editors, *Wavelets*, pages 2–20, Berlin, 1989. Springer Verlag.

26. V. Guillemin and A. Pollack. *Differential Topology.* Prentice Hall, Englewood Cliffs, NJ, 1974.

27. Simon Haykin. *Neural Networks A compreshensive foundation.* IEEE Press, New York, 1994.

28. Morris W. Hirsch. *Differential Topology.* Graduate Texts in Mathematics 33. Springer-Verlag, 1976.

29. Harry Hochstadt. *Integral Equations.* Wiley, New York, 1973.

30. K. Hoffman and R. Kunze. *Linear Algebra.* Prentice Hall, New Jersey, 1971.

31. R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, England, 1985.

32. R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, England, 1991.

33. H. Hotelling. Analysis of a complex of statistical variables into pricipal components. *Journal of Educational Psychology*, September, 1933.

34. D. Hundley and M. Kirby. An automated algorithm for dimensionality estimation. *submitted to Trans. on PAMI*, 1999.

35. D. Hundley, M. Kirby, and R. Miranda. Empirical dynamical system reduction II: Neural charts. In K. Coughlin, editor, *Semi-Analytic Methods for the Navier-Stokes Equations (Montreal, 1995)*, volume 20 of *CRM Proc. Lecture Notes*, pages 65–83, Providence, RI, 1999. Amer. Math. Soc.

36. D. R. Hundley. *Local Nonlinear Modeling via Neural Charts.* PhD dissertation, Colorado State University, Department of Mathematics, 1998.

37. Witold Hurewicz and Henry Wallman. *Dimension Theory.* Princeton University Press, Princeton, 1948.

38. J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation.* Addison Wesley, 1991.

39. I.T. Jolliffe. *Principal Component Analysis.* Springer, New York, 1986.

40. D.S. Jones. *Elementary Information Theory.* Oxford University Press, Oxford, U.K., 1979.

41. K. Karhunen. Uber lineare methoden in der wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae*, 37, 1946.

42. M. Kirby. Minimal dynamical systems from partial differential equations using sobolev eigenfunctions. *Physica D*, 57:466–475, 1992.

43. M. Kirby and D. Armbruster. Reconstructing phase-space from PDE simulations. *Z. angew. Math. Phys.*, 43:999–1022, 1992.

44. M. Kirby and R. Miranda. Empirical dynamical system reduction I: Global nonlinear transformations. In K. Coughlin, editor, *Semi-Analytic Methods for the Navier-Stokes Equations (Montreal, 1995)*, volume 20 of *CRM Proc. Lecture Notes*, pages 41–64, Providence, RI, 1999. Amer. Math. Soc.

45. M. Kirby and L. Sirovich. Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE trans. PAMI*, 12(1):103–108, 1990.

46. M. Kirby, F. Weisser, and G. Dangelmayr. A problem in facial animation: Analysis and synthesis of lip motion. In *Proc. of the 7th Scandinavian Conf. on Image Analysis*, pages 529–536, Aalborg, Denmark, 1991.

47. M. Kirby, Frank Weisser, and G. Dangelmayr. Speaking with images: a model problem in the representation of still and moving images. *Pattern Recognition*, 26(1):63–73, 1993.

48. T. Kohonen. Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, 43:59, 1982.

49. T. Kohonen. *Self-organization and Associative Memory.* Springer-Verlag, Berlin, 1984.

50. Mark A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2):233–243, 1991.

51. P. Lancaster and M. Tismenetsky. *The Theory of Matrices.* Academic Press, New York, 1985.

52. L.D. Landau and E.M. Lifshitz. *Fluid Mechanics.* Pergamon Press, New York, 1959.

53. M. Loève. *Probability Theory.* von Nostrand, Princeton, N.J., 1955.

54. E. Lorenz. Empirical orthogonal eigenfunctions and statistical weather prediction. Science Report No. 1, Statistical Forecasting Project 1, M.I.T., Cambridge, MA, 1956.

55. Marvin L. Minsky and Seymour A. Papert. *Perceptrons.* MIT Press, Cambridge, MA, 1990 (expanded edition).

56. John Moody and Christian Darken. Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1:281–294, 1989.

57. E. Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273, 1982.

58. E. Oja. Data compression, feature extraction, and autoassociation in feedforward neural networks. In T. Kohonen, K. Mäkisara., O. Simula, and J. Kangas, editors, *"Artificial Neural Networks"*, pages 737–745, NY, 1991. Elsevier Science Publishers.

59. E. Oja and Juha Karhunen. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *Journal of Mathematical Analysis and Applications*, 106:69–84, 1985.

60. N.H. Packard, J.P Crutchfield, J.D. Farmer, and R.S. Shaw. Geometry from a time series. *Physical Review Letters*, 45:712–716, 1980.

61. Athanasios Papoulis. *The Fourier Integral and Its Applications.* McGraw Hill Electronic Science Series. McGraw Hill, New York, 1962.

62. Karl Pearson. On lines and planes of closest fit to systems of points in space. *Phil. Mag. S.*, 2(11):559–572, 1901.

63. M.J.D. Powell. *Approximation Theory and Methods.* Cambridge University Press, Cambridge, England, 1981.

64. M.J.D. Powell. The theory of radial basis functions in 1990. In W. Light, editor, *Advances in Numerical Analysis II: Wavelets Subdivision and Radial Basis Functions*, pages 105–210. Oxford Univeristy Press, 1992.

65. William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, Cambridge, England, 1989.

66. Riesz and Sz.-Nagy. *Functional Analysis.* Dover, reprinted 1990.

67. Terence David Sanger. Optimal unsupervised learning in a single layer linear feedforward neural network. *Neural Networks*, 2(6):459–473, 1989.

68. Tim Sauer, James A. Yorke, and Martin Casdagli. Embedology. *Journal of Statisical Physics*, 65(3/4):579–616, 1991.

69. William D. Sellers. A statistical-dynamic approach to numerical weather prediction. Science Report No. 2, Statistical Forecasting Project 2, M.I.T., Cambridge, MA, 1957.

70. L. Sirovich. Turbulence and the dynamics of coherent structures, Part I: Coherent structures. *Quarterly of Applied Mathematics*, XLV(3):561–571, 1987.

71. L. Sirovich. Turbulence and the dynamics of coherent structures, Part II: Symmetries and transformations. *Quarterly of Applied Mathematics*, XLV(3):573–582, 1987.

72. L. Sirovich. Turbulence and the dynamics of coherent structures, Part III: Dynamics and scaling. *Quarterly of Applied Mathematics*, XLV(3):583–590, 1987.

73. L. Sirovich and M. Kirby. A low-dimensional procedure for the characterization of human faces. *J. of the Optical Society of America A*, 4:529–524, 1987.

74. F. Smithies. *Integral Equations.* Combridge University Press, 1970.

75. G.W. Stewart. On the early history of the singular value decomposition. *SIAM Review*, 35:551–566, 1993.

76. G. Strang. *Linear Algebra and its Applications.* Academic Press, 1980.

77. F. Takens. Detecting strange attractors in turbulence. In D.A. Rand and L.-S. Young, editors, *Dynamical Systems and Turbulenc*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381, Warwick 1980, 1981. Springer-Verlag, Berlin.

78. Lloyd N. Trefethen and III David Bau. *Numerical Linear Algebra.* SIAM, Philadelphia, PA, 1997.

79. S. Watanabe. Karhunen-loève expansion and factor analysis. In *Trans. 4th. Prague Conf. on Inf. Theory, Stat. Decision Functions, and Random Proc.*, pages 635–660, Prague, 1965.

80. B. Widrow and S.D. Stearns. *Adaptive Signal Processing.* Prentice Hall, Englewood Cliffs, N.J., 1985.