An Overview of Wavelet Based Multiresolution Analyses

Björn Jawerth* and Wim Sweldens†

February 5, 1993

Abstract

In this paper we give an overview of some wavelet based multiresolution analyses. First, we briefly discuss the continuous wavelet transform in its simplest form. Then we give the definition of multiresolution analysis and show how wavelets fit into it. We take a closer look at orthogonal, biorthogonal and semiorthogonal wavelets. The fast wavelet transform, wavelets on closed sets (boundary wavelets), multidimensional wavelets and wavelet packets are discussed. Several examples of wavelet families are given and compared. Finally, the essentials of two major applications are outlined: data compression and compression of linear operators.

Keywords: wavelets, multiresolution analysis, compression

AMS(MOS) classification: 42-02, 42C10

1 Introduction

Wavelets and wavelet techniques have recently generated much interest, both in applied areas as well as in more theoretical ones. The class of wavelet techniques is not really precisely defined and it keeps changing. Hence, it is virtually impossible to give a precise definition of "wavelet" that incorporates all different aspects. It is equally hard to write a comprehensive overview of wavelets. In this paper we shall focus on wavelets in connection with multiresolution analysis. We shall also briefly discuss existing wavelet functions and, whenever appropriate, their advantages and disadvantages. It goes without saying (almost) that this short overview is still highly incomplete and does not cover many important and interesting developments in this area. Many of the results we do not mention are more significant than the ones we include, and we apologize to the people whose work is not discussed. For example, we hardly mention the significant volume of work done more in the direction of approximation theory, and the efforts in the field of fractal functions and the more applied areas are left out almost entirely.

Although wavelets are a relatively recent phenomenon, there are already several books on the subject, for example [14, 15, 33, 40, 60, 79, 88, 97].

*Partially supported by DARPA Grant AFOSR 89-0455 and ONR Grant N00014-90-J-1343.

†Research Assistant of the National Fund of Scientific Research Belgium, partially supported by ONR Grant N00014-90-J-1343.

2 Notation and definitions

Much of the notation will be presented as we go along. Here we just note that the inner product of two square integrable functions $f, g \in L^2(\mathbb{R})$ is defined as

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

and that the Fourier transform of a function $f \in L^2(\mathbb{R})$ is defined as

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x)\,e^{-i\omega x}\,dx.$$

We shall also use the Poisson summation formula in the following two forms,

$$\sum_l f(x-l) = \sum_k \hat{f}(2k\pi)\,e^{i2k\pi x},$$

and

$$\sum_l \langle f, g(\cdot - l) \rangle\, e^{-i\omega l} = \sum_k \hat{f}(\omega + k2\pi)\,\overline{\hat{g}(\omega + k2\pi)}.$$

If no bounds are indicated under a summation sign, $\in \mathbb{Z}$ is understood.

A countable set $\{f_n\}$ of a Hilbert space is a *Riesz basis* if every element f of the space can be written uniquely as $f = \sum_n c_n f_n$, and positive constants A and B exist such that

$$A\,\|f\|^2 \leq \sum_n |c_n|^2 \leq B\,\|f\|^2.$$

3 A short history of wavelets

Wavelet theory involves representing general functions in terms of simpler, fixed building blocks at different scales and positions. This has been found to be a useful approach in several different areas. For example, we have subband coding techniques, quadrature mirror filters, pyramid schemes, etc., in signal and image processing, while in mathematical physics similar ideas are studied as part of the theory of Coherent States. Wavelet theory represents a useful synthesis of these different approaches.

In abstract mathematics it has been known for quite some time that techniques based on Fourier series and Fourier transforms are not quite adequate for many problems and Littlewood-Paley techniques are often effective substitutes. These techniques were initially developed in the 30's to understand, for example, summability properties of Fourier series and boundary behavior of analytic functions. However, in the 50's and 60's they developed into powerful tools for understanding other things such as solutions of partial differential equations and integral equations. It was realized that they fit into so called Calderón–Zygmund theory, an area of harmonic analysis which is still very heavily researched.

One of the standard approaches, not only in Calderón-Zygmund theory but in analysis in general, is to break up a complicated phenomenon into many simple pieces and study each of the pieces separately.

In the 70's, sums of simple functions, called atomic decompositions [28], were widely used, especially in Hardy space theory. One method used to establish that a general function f has such a decomposition is to start with the "Calderón formula": for a function f, holds that

$$f(x) \;=\; \int_0^{+\infty}\int_{-\infty}^{+\infty} (\psi_t * f)(y)\, \tilde{\psi}_t(x-y)\, dy\, \frac{dt}{t}.$$

The $*$ denotes convolution. Here $\psi_t(x) = t^{-1}\psi(x/t)$, and similarly for $\tilde{\psi}_t(x)$, for appropriate fixed functions ψ and $\tilde{\psi}$. In fact, as we shall see below, this representation is an example of a continuous wavelet transform. In the context of trying to further understand Hardy spaces, as well as other spaces used to measure the size and smoothness of functions, and showing very deep, but also very abstract, functional analytic properties, the first orthogonal wavelets were discovered by Strömberg [104]. A discrete version of the Calderón formula had also been used for similar purposes in [74] and long before this there were results by Haar [68], Franklin [56], Ciesielski [20], Peetre [93], and others.

Independently from these developments in harmonic analysis, Alex Grossmann, Jean Morlet, and their coworkers studied the wavelet transform in its continuous form [65, 66, 67]. The theory of "frames" [41] provided a suitable general framework for these investigations.

In the early to mid 80's there were several groups, perhaps most notably the one associated with Yves Meyer and his collaborators, that independently realized, with some excitement, that some of the tools that had been so effective in Calderón-Zygmund theory, in particular the Littlewood-Paley representations, had discrete analogs and could be used both to give a unified view of many of the results in harmonic analysis and also, at least potentially, could be effective substitutes for Fourier series in numerical applications. (The first named author of this paper came to this understanding through the joint work with Mike Frazier [57, 58, 59].) As the emphasis shifted more towards the representations themselves, and the building blocks involved, the name also shifted: Yves Meyer and Jean Morlet suggested the word wavelet for the building blocks, and what earlier had been referred to as Littlewood-Paley theory now started to be called wavelet theory.

Pierre-Gilles Lemarié and Yves Meyer [80], independently of Strömberg, constructed new orthogonal wavelet expansions. With the notion of multiresolution analysis, introduced by Stéphane Mallat and Yves Meyer, a systematic framework for understanding these orthogonal expansions was developed [85, 86, 87]. It also provided the connection with quadrature mirror filtering. Soon Ingrid Daubechies [37] gave a construction of wavelets, non-zero only on a finite interval and with arbitrarily high, but fixed, regularity. This takes us up to a fairly recent time in the history of wavelet theory. Several people have made substantial contributions to the field over the past few years. Some of their work and the appropriate references will be discussed in the body of the paper.

4 The continuous wavelet transform

Since we are going to be brief, let us start by pointing out that more detailed treatments of the continuous wavelet transform can be found in [14, 65, 64, 70]. As mentioned above, a wavelet expansion consists of translations and dilations of one fixed function, the wavelet $\psi \in L^2(\mathbb{R})$. In the continuous wavelet transform the translation and dilation parameter vary continuously. This means that we use the functions

$$\psi_{a;b}(x) \;=\; \frac{1}{\sqrt{|a|}}\,\psi\left(\frac{x-b}{a}\right) \quad \text{with} \quad a, b \in \mathbb{R},\; a \neq 0.$$

These functions are scaled so that their $L^2(\mathbb{R})$ norms are independent of a. The *continuous wavelet transform* of a function $f \in L^2(\mathbb{R})$ is now defined as

$$\mathcal{W}(a,b) \;=\; \langle\, f, \psi_{a;b}\,\rangle\,. \tag{1}$$

Using the Parseval identity we can also write this as

$$2\pi\, \mathcal{W}(a,b) \;=\; \langle\, \hat{f}, \hat{\psi}_{a;b}\,\rangle\,. \tag{2}$$

where

$$\hat{\psi}_{a;b}(\omega) \;=\; \frac{a}{\sqrt{|a|}}\, e^{-i\omega b}\, \hat{\psi}(a\omega).$$

Note that the continuous wavelet transform takes a one-dimensional function into a two-dimensional one. The representation of a function by its continuous wavelet transform is redundant and the inverse transform is possibly not unique. Furthermore, not every function $\mathcal{W}(a,b)$ is the continuous wavelet transform of a function f.

We assume that the wavelet ψ and its Fourier transform $\hat{\psi}$ are functions with finite centers $\bar{x}$ and $\bar{\omega}$ and finite radii Δ_x and Δ_ω. The latter quantities are defined as

$$\begin{aligned} \bar{x} \;&=\; \frac{1}{\|\psi\|_{L_2}^2} \int_{-\infty}^{+\infty} x\, |\psi(x)|^2\, dx, \\ \Delta_x^2 \;&=\; \frac{1}{\|\psi\|_{L_2}^2} \int_{-\infty}^{+\infty} (x-\bar{x})^2\, |\psi(x)|^2\, dx, \end{aligned}$$

and similarly for $\bar{\omega}$ and Δ_ω. Although the variable x typically represents either time or space, we shall refer to it as time. From (1) and (2) we see that the continuous wavelet transform at (a,b) essentially contains information from the time interval $[b + a\bar{x} - a\Delta_x, b + a\bar{x} + a\Delta_x]$ and the frequency interval $[(\bar{\omega} - \Delta_\omega)/a, (\bar{\omega} + \Delta_\omega)/a]$. These two intervals determine a *time-frequency window*. Its width, height and position are governed by a and b. Its area is constant and given by $4\Delta_x\Delta_\omega$. Due to the Heisenberg uncertainty principle the area has to be greater then 2. These time-frequency windows are also called *Heisenberg boxes*.

Suppose that the wavelet ψ satisfies the *admissibility condition*

$$C_\psi \;=\; \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\, d\omega < \infty.$$

Then the continuous wavelet transform $\mathcal{W}(a,b)$ has an inverse given by the relation

$$f(x) \;=\; \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}(a,b)\, \psi_{a;b}(x)\, \frac{da\, db}{a^2}. \tag{3}$$

From the admissibility condition we see that $\hat{\psi}(0)$ has to be 0, and, hence, ψ has to oscillate. This together with the decay property gave ψ the name *wavelet* or "small wave" (French: *ondelette*). Other, more efficient inverse transforms exist that only use $\mathcal{W}(a,b)$ for positive values of a in the reconstruction, or even only use $\mathcal{W}(a,b)$ at discrete values of a [38].

This transform can be used to analyze signals and this was done successfully e.g. in geophysics. The transform is often graphically represented as two two-dimensional images with color or grey value corresponding to the modulus and phase of $\mathcal{W}(a,b)$.

The continuous wavelet transform is also used in singularity detection and characterization [57, 82]. A typical result in this direction is that if a function f is Hölder (Lipschitz) continuous of order $0 < \alpha < 1$, so that $|f(x+h) - f(x)| = \mathcal{O}(h^\alpha)$, then the continuous wavelet transform has an asymptotic behavior like

$$\mathcal{W}(a,b) \;=\; \mathcal{O}(a^{\alpha+1/2}) \quad \text{for} \quad a \to 0.$$

In fact, the converse is true as well. The advantage of this characterization compared to the Fourier transform is that it does not only provide information about the kind of singularity, but also about its location in time. There is also a corresponding characterization of Hölder (Lipschitz) continuous functions of higher order $\alpha \geq 1$; the wavelet must then have a number of vanishing moments greater than α, i.e.

$$\int_{-\infty}^{+\infty} \psi(x)\, x^p \, dx \;=\; 0 \quad \text{for} \quad 0 \leq p \leq \alpha \quad \text{and} \quad p \in \mathbb{Z}.$$

So the number of vanishing wavelet moments limits the order of smoothness that can be characterized.

Example: A classical example of a wavelet is the so-called *Mexican hat*,

$$\psi(x) \;=\; (1 - 2x^2)e^{-x^2}.$$

This is the second derivative of a Gaussian and it has thus two vanishing moments.

5 Multiresolution analysis

5.1 The scaling function and the subspaces V_j

There are at least two ways to introduce wavelets: one is through the continuous wavelet transform as in the previous section, another is through multiresolution analysis. Here we shall start by introducing the concept of multiresolution analysis and then point out the connections with the continuous wavelet transform.

A *multiresolution analysis* of $\mathrm{L}^2(\mathbb{R})$ is defined as a sequence of closed subspaces V_j of $\mathrm{L}^2(\mathbb{R})$, $j \in \mathbb{Z}$, with the following properties [37, 85]:

1. $V_j \subset V_{j+1}$,
2. $v(x) \in V_j \Leftrightarrow v(2x) \in V_{j+1}$,

3. $v(x) \in V_0 \Leftrightarrow v(x+1) \in V_0$,

4. $\bigcup_{j=-\infty}^{+\infty} V_j$ is dense in $L^2(\mathbb{R})$ and $\bigcap_{j=-\infty}^{+\infty} V_j = \{\mathbf{0}\}$,

5. A *scaling function* $\phi \in V_0$ with a nonvanishing integral exists such that the collection $\{\phi(x-l) \mid l \in \mathbb{Z}\}$ is a Riesz basis of V_0.

Let us make a couple of simple observations related to this definition. Since $\phi \in V_0 \subset V_1$, a sequence $(h_k) \in l^2(\mathbb{Z})$ exists such that the scaling function satisfies

$$\phi(x) = 2 \sum_k h_k \, \phi(2x-k). \tag{4}$$

This functional equation is also called the *refinement equation*, *dilation equation* or *two-scale difference equation*. Here we will use the first choice. The collection of functions $\{\phi_{j,l} \mid l \in \mathbb{Z}\}$, with $\phi_{j,l}(x) = \sqrt{2^j}\, \phi(2^j x - l)$, is now a Riesz basis of V_j.

We also note that a multiresolution analysis allows us to approximate a given function f by a function f_j in each of the spaces V_j. We call f_j the approximation at resolution j and a possible way to find it is projection. Since the union of all the V_j is dense in $L^2(\mathbb{R})$, we are guaranteed that there are such approximations converging to the original function, or $f = \lim_{j \to +\infty} f_j$.

By integrating both sides of (4) and using the fact that the integral of ϕ does not vanish, we see that

$$\sum_k h_k \; = \; 1. \tag{5}$$

The scaling function is, under very general conditions, uniquely defined by its refinement equation and the normalization [42],

$$\int_{-\infty}^{+\infty} \phi(x)\, dx \; = \; 1.$$

In many cases, no explicit expression for ϕ is available. However, there are fast algorithms that use the refinement equation to evaluate the scaling function ϕ at dyadic points ($x = 2^{-j}k,\ j, k \in \mathbb{Z}$) (see, for example, [8, 12, 37, 42, 43, 102]). In many applications, we never need the scaling function itself; instead we may often work directly with the h_k.

To be able to use the collection $\{\phi(x-l) \mid l \in \mathbb{Z}\}$ to approximate even the simplest functions (such as constants), it is natural to assume that the scaling function and its integer translates form a *partition of unity*, or, in other words,

$$\forall\, x \in \mathbb{R} \; : \; \sum_k \phi(x-k) \; = \; 1.$$

This is also used to prove that a given ϕ generates a multiresolution analysis. By Poisson's summation formula, the partition of unity relation is (essentially) equivalent with

$$\hat{\phi}(2\pi k) \; = \; \delta_k \quad \text{for} \quad k \in \mathbb{Z}. \tag{6}$$

By (4), the Fourier transform of the scaling function must satisfy

$$\hat{\phi}(\omega) \;=\; H(w/2)\,\hat{\phi}(\omega/2), \tag{7}$$

where H is a 2π-periodic function defined by

$$H(\omega) \;=\; \sum_k h_k\, e^{-ik\omega}.$$

Since $\hat{\phi}(0) = 1$, we can apply (7) recursively. This yields, at least formally,

$$\hat{\phi}(\omega) \;=\; \prod_{j=1}^{\infty} H(2^{-j}\omega).$$

The convergence of this product is examined in [21, 37]. The product formula for $\hat{\phi}$ is nice to have in many situations. For example, it can be used to construct $\phi(x)$ from the h_k. Using (6) and (7), we see that we obtain a partition of unity if

$$H(\pi) = 0 \quad \text{or} \quad \sum_k (-1)^k\, h_k \;=\; 0.$$

We also see that (5) can be written as

$$H(0) \;=\; 1.$$

Examples of scaling functions:

- A well known family of scaling functions is the set of cardinal B–splines. The cardinal B–spline of order 1 is the box function $N_1(x) = \chi_{[0,1]}(x)$. For $m > 1$ the cardinal B–spline N_m is defined recursively as a convolution,

  $$N_m \;=\; N_{m-1} * N_1.$$

  These functions satisfy

  $$N_m(x) \;=\; 2^{m-1} \sum_k \binom{m}{k} N_m(2x-k),$$

  and

  $$\hat{N}_m(\omega) \;=\; \left(\frac{1-e^{-i\omega}}{i\omega}\right)^m.$$

- Another classical example is the Shannon sampling function,

  $$\phi(x) = \frac{\sin(\pi x)}{\pi x} \quad \text{with} \quad \hat{\phi}(\omega) \;=\; \chi_{[-\pi,\pi]}(\omega).$$

  We may take

  $$H(\omega) \;=\; \chi_{[-\pi/2,\pi/2]}(\omega) \quad \text{for} \quad \omega \in [-\pi,\pi],$$

  and, consequently,

  $$h_{2k} \;=\; 1/2\,\delta_k \quad \text{and} \quad h_{2k+1} \;=\; \frac{(-1)^k}{(2k+1)\pi} \quad \text{for} \quad k \in \mathbb{Z}.$$

For the remainder of this paper, it will be useful to define the following 2π-periodic function,

$$F(\omega) \;=\; \sum_k |\hat{\phi}(\omega + k2\pi)|^2.$$

The fact that ϕ and its translates form a Riesz basis corresponds to the existence of positive constants A and B such that

$$0 < A \leq F(\omega) \leq B < \infty.$$

Using (7) and rearranging the even and odd terms, we have

$$\begin{aligned} F(2\omega) &= \sum_k |\hat{\phi}(2\omega + k2\pi)|^2 \\ &= \sum_k |H(\omega + k\pi)|^2 \, |\hat{\phi}(\omega + k\pi)|^2 \\ &= \sum_k |H(\omega + k2\pi)|^2 \, |\hat{\phi}(\omega + k2\pi)|^2 + |H(\omega + \pi + k2\pi)|^2 \, |\hat{\phi}(\omega + \pi + k2\pi)|^2 \\ &= |H(\omega)|^2 \, F(\omega) + |H(\omega + \pi)|^2 \, F(\omega + \pi). \end{aligned} \tag{8}$$

5.2 The wavelet function and the detail spaces W_j

We will use W_j to denote a space complementing V_j in V_{j+1}, i.e. a space that satisfies

$$V_{j+1} \;=\; V_j \oplus W_j,$$

where the symbol $\oplus$ stands for direct sum. This means that the space W_j contains the "detail" information needed to go from an approximation at resolution j to an approximation at resolution $j+1$. Consequently,

$$\bigoplus_j W_j \;=\; \mathrm{L}^2(\mathrm{I\!R}).$$

Note that defined this way, the space W_j is not unique.

A function ψ is a *wavelet* if the collection of functions $\{\psi(x - l) \mid l \in \mathbb{Z}\}$ is a Riesz basis of W_0. The collection of wavelet functions $\{\psi_{j,l} \mid l, j \in \mathbb{Z}\}$ is then a Riesz basis of $\mathrm{L}^2(\mathrm{I\!R})$, where $\psi_{j,l}$ is defined similarly to $\phi_{j,l}$. Since the wavelet ψ is an element of V_1, a sequence $(g_k) \in \mathrm{l}^2(\mathbb{Z})$ exists such that

$$\psi(x) = 2 \sum_k g_k \, \phi(2x - k). \tag{9}$$

Also here we require the wavelet to have a vanishing integral, or

$$\int_{-\infty}^{+\infty} \psi(x) dx \;=\; 0. \tag{10}$$

The Fourier transform of the wavelet is given by

$$\hat{\psi}(\omega) \;=\; G(\omega/2) \, \hat{\phi}(\omega/2), \tag{11}$$

where G is the 2π-periodic function

$$G(\omega) \;=\; \sum_k g_k \, e^{-ik\omega}.$$

From (9) and (10) we have

$$\sum_k g_k \;=\; 0 \quad \text{or} \quad G(0) \;=\; 0.$$

Each space V_j and W_j has an $\mathrm{L}^2(\mathbb{R})$ complement denoted by V_j^c and W_j^c, respectively. We have:

$$V_j^c \;=\; \bigoplus_{i=j}^{\infty} W_i \quad \text{and} \quad W_j^c \;=\; \bigoplus_{i \neq j} W_i.$$

We define P_j and Q_j as the projection operators onto V_j and W_j and parallel to V_j^c or W_j^c, respectively. A function f can now be written as

$$f(x) \;=\; \sum_j Q_j f(x) \;=\; \sum_{j,l} \mu_{j,l} \, \psi_{j,l}(x).$$

This can be seen as a discrete version of the inverse continuous wavelet transform (3). The mapping from the function f to the coefficients $\mu_{j,l}$ is usually referred to as the *discrete wavelet transform*. How the coefficients $\mu_{j,l}$ are found will become clear in the following sections.

6 Orthogonal wavelets

A particularly interesting class of wavelets are the *orthogonal wavelets*. We start their construction by introducing an *orthogonal multiresolution analysis*. This is a multiresolution analysis where the wavelet spaces W_j are defined as the *orthogonal* complement of V_j in V_{j+1}. Consequently the spaces W_j with $j \in \mathbb{Z}$ are all mutually orthogonal, the projections P_j and Q_j are orthogonal, and the expansion

$$f(x) \;=\; \sum_j Q_j f(x)$$

is an orthogonal expansion. A sufficient condition for a general multiresolution analysis to be an orthogonal multiresolution analysis is

$$W_0 \;-\; V_0,$$

or

$$\langle \, \psi, \phi(\cdot - l) \, \rangle \;=\; 0 \qquad l \in \mathbb{Z},$$

since the other conditions simply follow from scaling. Using the Poisson summation formula, we see that this condition is (essentially) equivalent with

$$\forall \, \omega \in \mathbb{R} \;:\; \sum_k \hat{\psi}(\omega + k2\pi) \, \overline{\hat{\phi}(\omega + k2\pi)} \;=\; 0. \tag{12}$$

An *orthogonal scaling function* is a function ϕ such that the set $\{\phi(x-l) \mid l \in \mathbb{Z}\}$ is an *orthonormal* basis, or

$$\langle\, \phi, \phi(\cdot - l)\, \rangle \;=\; \delta_l \qquad\qquad l \in \mathbb{Z}. \tag{13}$$

With such a ϕ, the collection of functions $\{\phi(x-l) \mid l \in \mathbb{Z}\}$ is an orthonormal basis of V_0 and the collection of functions $\{\phi_{j,l} \mid l \in \mathbb{Z}\}$ is an orthonormal basis of V_j. Using Poisson's formula, equation (13) is (essentially) equivalent with

$$\forall\, \omega \in \mathbb{R} \;:\; \sum_k |\hat{\phi}(\omega + k2\pi)|^2 \;=\; F(\omega) \;=\; 1. \tag{14}$$

From equation (8) we now see that,

$$\forall\, \omega \in \mathbb{R} \;:\; |H(\omega)|^2 + |H(\omega+\pi)|^2 \;=\; 1, \tag{15}$$

or

$$\sum_k h_k\, h_{k-2l} \;=\; \delta_l/2 \quad \text{for} \quad l \in \mathbb{Z}.$$

The last two equations are equivalent but they provide only a necessary condition for the orthogonality of the scaling function and its translates. This relationship is investigated in detail in [77].

Now, an *orthogonal wavelet* is a function ψ such that the collection of functions $\{\psi(x-l) \mid l \in \mathbb{Z}\}$ is an orthonormal basis of W_0. This is the case if

$$\langle\, \psi, \psi(\cdot - l)\, \rangle \;=\; \delta_l$$

Again these conditions are (essentially) equivalent with

$$\forall\, \omega \in \mathbb{R} \;:\; \sum_k |\hat{\psi}(\omega + k2\pi)|^2 \;=\; 1,$$

and, using a similar argument as above, a necessary condition is given by

$$\forall\, \omega \in \mathbb{R} \;:\; |G(\omega)|^2 + |G(\omega+\pi)|^2 \;=\; 1.$$

Since the spaces W_j are mutually orthogonal, the collection of functions $\{\psi_{j,l} \mid j, l \in \mathbb{Z}\}$ is an orthonormal basis of $\mathrm{L}^2(\mathbb{R})$.

The projection operators P_j and Q_j can now be written as

$$P_j f(x) \;=\; \sum_l \langle\, f, \phi_{j,l}\, \rangle\, \phi_{j,l}(x) \quad \text{and} \quad Q_j f(x) \;=\; \sum_l \langle\, f, \psi_{j,l}\, \rangle\, \psi_{j,l}(x).$$

They are the best $\mathrm{L}^2(\mathbb{R})$ approximations of the function f in V_j and W_j respectively. For a function $f \in \mathrm{L}^2(\mathbb{R})$ we now have the orthogonal expansion

$$f(x) \;=\; \sum_{j,l} \mu_{j,l}\, \psi_{j,l}(x) \quad \text{with} \quad \mu_{j,l} \;=\; \langle\, f, \psi_{j,l}\, \rangle\,.$$

Again, this can be viewed as a discrete version of the continuous wavelet transform. Examples of orthogonal wavelets are given in section 10.

Using equation (14) we can write condition (12) as

$$\forall\, \omega \in \mathbb{R} \;:\;\; G(\omega)\,\overline{H(\omega)} + G(\omega+\pi)\,\overline{H(\omega+\pi)} \;=\; 0.$$

From this last equation we see that a possible choice for the function $G(\omega)$ is

$$G(\omega) \;=\; -e^{-i\omega}\,\overline{H(\omega+\pi)}.$$

For this choice the orthogonality of the wavelet immediately follows from the orthogonality of the scaling function. This means we can derive an orthogonal wavelet from an orthogonal scaling function by choosing

$$g_k \;=\; (-1)^k\,\overline{h_{1-k}}. \tag{16}$$

In [78] an orthonormalization procedure to find orthonormal wavelets is proposed. It states that if a function ϕ and its integer translates form a Riesz basis of V_0, then an orthonormal basis of V_0 is given by ϕ_{orth} and its integer translates with

$$\hat{\phi}_{orth}(\omega) \;=\; \frac{\hat{\phi}(\omega)}{\sqrt{F(\omega)}}. \tag{17}$$

The fact that we started from a Riesz basis guarantees that $F(\omega)$ is strictly positive. We see that ϕ now indeed satisfies the orthogonality condition (14). Note that if ϕ is compactly supported, ϕ_{orth} will in general not be compactly supported.

Now, from condition (15) and the fact that $H(0) = G(\pi) = 1$ and $G(0) = H(\pi) = 0$, we see that $H(\omega)$ essentially represents a low pass filter for the interval $[0, \pi/2]$ and $G(\omega)$ represents a band pass filter for the interval $[\pi/2, \pi]$. Then, from (7) and (11) we conclude that the main part of the energy of $\hat{\phi}(\omega)$ and $\hat{\psi}(\omega)$ is concentrated in the intervals $[0, \pi]$ and $[\pi, 2\pi]$, respectively. This means that the wavelet expansion essentially splits the frequency space into dyadic blocks $[2^j\pi, 2^{j+1}\pi]$ with $j \in \mathbb{Z}$.

7 Biorthogonal wavelets

The orthogonality property puts a strong limitation on the construction of wavelets. For instance, there are hardly any wavelets that are compactly supported, symmetric and orthogonal. Hence, the generalization to *biorthogonal wavelets* has been introduced. Here, a dual scaling function $\tilde{\phi}$ and a dual wavelet $\tilde{\psi}$ exist who generate a dual multiresolution analysis with subspaces $\tilde{V}_j$ and $\tilde{W}_j$ such that

$$\tilde{V}_j - W_j \quad \text{and} \quad V_j - \tilde{W}_j, \tag{18}$$

and consequently,

$$\tilde{W}_j - W_{j'} \quad \text{for} \quad j \neq j'.$$

The dual multiresolution analysis is not necessarily the same as the one generated by the primary functions. An equivalent condition for (18) is

$$\langle\, \tilde{\phi}, \psi(\cdot - l)\,\rangle \;=\; \langle\, \tilde{\psi}, \phi(\cdot - l)\,\rangle \;=\; 0.$$

Moreover, the dual functions also have to satisfy

$$\langle\, \tilde{\phi}, \phi(\cdot - l)\,\rangle \;=\; \delta_l \quad \text{and} \quad \langle\, \tilde{\psi}, \psi(\cdot - l)\,\rangle \;=\; \delta_l.$$

Again using a scaling argument we have now that

$$\langle\, \tilde{\phi}_{j,l}, \phi_{j,l'}\,\rangle \;=\; \delta_{l-l'} \qquad l, l', j \in \mathbb{Z} \tag{19}$$

and

$$\langle\, \tilde{\psi}_{j,l}, \psi_{j',l'}\,\rangle \;=\; \delta_{j-j'}\delta_{l-l'} \qquad l, l', j, j' \in \mathbb{Z}. \tag{20}$$

where $\tilde{\phi}_{j,l}$ and $\tilde{\psi}_{j,l}$ are defined similarly to $\phi_{j,l}$ and $\psi_{j,l}$. Note that the role of the primary (i.e. the ϕ and ψ) and dual functions can be interchanged. Using the same Fourier techniques as in the previous section, the biorthogonality conditions are (essentially) equivalent with

$$\forall \omega \in \mathbb{R} \;:\; \begin{cases} \sum_k \hat{\tilde{\phi}}(\omega + k2\pi)\,\overline{\hat{\phi}(\omega + k2\pi)} &=\; 1 \\ \sum_k \hat{\tilde{\psi}}(\omega + k2\pi)\,\overline{\hat{\psi}(\omega + k2\pi)} &=\; 1 \\ \sum_k \hat{\tilde{\psi}}(\omega + k2\pi)\,\overline{\hat{\phi}(\omega + k2\pi)} &=\; 0 \\ \sum_k \hat{\tilde{\phi}}(\omega + k2\pi)\,\overline{\hat{\psi}(\omega + k2\pi)} &=\; 0. \end{cases} \tag{21}$$

As they define a multiresolution analysis, the dual functions satisfy

$$\tilde{\phi}(x) = 2\sum_k \tilde{h}_k\,\tilde{\phi}(2x-k) \quad \text{and} \quad \tilde{\psi}(x) = 2\sum_k \tilde{g}_k\,\tilde{\phi}(2x-k). \tag{22}$$

If we define the functions $\tilde{H}$ and $\tilde{G}$ similar to H and G, then necessary conditions are again given by,

$$\forall \omega \in \mathbb{R} \;:\; \begin{cases} \tilde{H}(\omega)\,\overline{H(\omega)} &+& \tilde{H}(\omega+\pi)\,\overline{H(\omega+\pi)} &=& 1 \\ \tilde{G}(\omega)\,\overline{G(\omega)} &+& \tilde{G}(\omega+\pi)\,\overline{G(\omega+\pi)} &=& 1 \\ \tilde{G}(\omega)\,\overline{H(\omega)} &+& \tilde{G}(\omega+\pi)\,\overline{H(\omega+\pi)} &=& 0 \\ \tilde{H}(\omega)\,\overline{G(\omega)} &+& \tilde{H}(\omega+\pi)\,\overline{G(\omega+\pi)} &=& 0, \end{cases}$$

or

$$\forall \omega \in \mathbb{R} \;:\; \begin{bmatrix} \tilde{H}(\omega) & \tilde{H}(\omega+\pi) \\ \tilde{G}(\omega) & \tilde{G}(\omega+\pi) \end{bmatrix} \overline{\begin{bmatrix} H(\omega) & G(\omega) \\ H(\omega+\pi) & G(\omega+\pi) \end{bmatrix}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

or

$$\tilde{M}(\omega)\,\overline{M^t(\omega)} \;=\; \mathbf{1}, \tag{23}$$

where

$$M(\omega) \;=\; \begin{bmatrix} H(\omega) & H(\omega+\pi) \\ G(\omega) & G(\omega+\pi) \end{bmatrix},$$

and similarly for $\tilde{M}$. By interchanging the matrices on the left hand side of (23), we get

$$\forall \omega \in \mathbb{R} \;:\; \begin{cases} \overline{H(\omega)}\,\tilde{H}(\omega) &+& \overline{G(\omega)}\,\tilde{G}(\omega) &=& 1 \\ \overline{H(\omega)}\,\tilde{H}(\omega+\pi) &+& \overline{G(\omega)}\,\tilde{G}(\omega+\pi) &=& 0. \end{cases} \tag{24}$$

Note that the orthogonal case corresponds to M being a unitary matrix. Cramer's rule now states that

$$\tilde{H}(\omega) \;=\; \frac{\overline{G(\omega+\pi)}}{\overline{\Delta(\omega)}} \tag{25}$$

and

$$\tilde{G}(\omega) \;=\; -\frac{\overline{H(\omega+\pi)}}{\overline{\Delta(\omega)}}, \tag{26}$$

where

$$\Delta(\omega) \;=\; \det M(\omega).$$

The projection operators take the form

$$P_j f(x) \;=\; \sum_l \langle\, f, \tilde{\phi}_{j,l} \,\rangle\, \phi_{j,l}(x) \quad \text{and} \quad Q_j f(x) \;=\; \sum_l \langle\, f, \tilde{\psi}_{j,l} \,\rangle\, \psi_{j,l}(x).$$

From the equations (19), (20), and (22) we see that

$$\tilde{h}_{k-2l} \;=\; \langle\, \tilde{\phi}(x-l), \phi(2x-k) \,\rangle \quad \text{and} \quad \tilde{g}_{k-2l} \;=\; \langle\, \tilde{\psi}(x-l), \phi(2x-k) \,\rangle\,.$$

In particular by writing $\phi(2x-k) \in V_1$ in the bases of V_0 and W_0 we obtain

$$\phi(2x-k) = \sum_l \tilde{h}_{k-2l}\, \phi(x-l) + \sum_l \tilde{g}_{k-2l}\, \psi(x-l). \tag{27}$$

Since primary and dual functions are interchangeable, we also have

$$\tilde{\phi}(2x-k) = \sum_l h_{k-2l}\, \tilde{\phi}(x-l) + \sum_l g_{k-2l}\, \tilde{\psi}(x-l). \tag{28}$$

The fact that the scaling function and wavelet are not orthogonal does not necessarily mean that the multiresolution analysis is not orthogonal. In fact, a biorthogonal scaling function and wavelet are *semiorthogonal* if they generate an orthogonal multiresolution analysis [14]. Also the name *pre-wavelet* is used in this context. Since the W_j subspaces are mutually orthogonal we have that

$$W_j - \tilde{W}_{j'} \quad \text{and} \quad W_j - W_{j'} \quad \text{for} \quad j \neq j'.$$

Consequently, $W_j = \tilde{W}_j$ and thus $V_j = \tilde{V}_j$. Hence primary and dual functions generate the same (orthogonal) multiresolution analysis. A dual scaling function can now be found by letting

$$\hat{\tilde{\phi}}(\omega) \;=\; \frac{\hat{\phi}(\omega)}{F(\omega)}.$$

We see that the first equation of (21) is now satisfied and, since F is a bounded, 2π-periodic function that does not vanish, the translates of ϕ and $\tilde{\phi}$ will generate the same space. This corresponds to

$$\tilde{H}(\omega) \;=\; \frac{H(\omega)\, F(\omega)}{F(2\omega)}.$$

Since $\Delta(\omega+\pi) = -\Delta(\omega)$, we can choose Δ as

$$\Delta(\omega) \;=\; e^{-i\omega}\, F(2\omega),$$

such that

$$G(\omega) \;=\; -e^{-i\omega}\, \overline{\tilde{H}(\omega+\pi)}\, F(2\omega) \;=\; -e^{-i\omega}\, \overline{H(\omega+\pi)}\, F(\omega+\pi),$$

and

$$\tilde{G}(\omega) \;=\; -e^{-i\omega}\, \frac{\overline{H(\omega+\pi)}}{F(2\omega)}.$$

If ϕ is a compactly supported function, this construction guarantees that ψ is compactly supported too. However, in general the dual functions will not be compactly supported.

8 Wavelets and polynomials

The moments of the scaling function and wavelet are defined as:

$$\mathcal{M}_p \;=\; \int_{-\infty}^{+\infty} x^p\, \phi(x)\, dx \quad \text{and} \quad \mathcal{N}_p \;=\; \int_{-\infty}^{+\infty} x^p\, \psi(x)\, dx \quad \text{with} \quad p \geq 0.$$

Of course, these integrals only make sense if ϕ and ψ have sufficient decay. The scaling function has $\mathcal{M}_0 = 1$. Recursion formulae to calculate these moments are derived in [10, 105]. The number of vanishing wavelet moments is denoted by $\tilde{N}$ where $\tilde{N}$ is at least 1:

$$\mathcal{N}_p \;=\; 0 \quad \text{for} \quad 0 \leq p < \tilde{N} \quad \text{and} \quad \mathcal{N}_{\tilde{N}} \neq 0.$$

This is equivalent with

$$\hat{\psi}^{(p)}(0) \;=\; 0 \quad \text{for} \quad 0 \leq p < \tilde{N},$$

and, since $\hat{\phi}(0) = \mathcal{M}_0 \neq 0$, also with

$$G^{(p)}(0) \;=\; 0 \quad \text{for} \quad 0 \leq p < \tilde{N}.$$

The sequence (g_k) thus has also $\tilde{N}$ vanishing discrete moments. The number of vanishing moments of the dual wavelet is denoted by N and similar statements can be made for the dual functions by adding or omitting the tilde. Since this is true for other statements in this section as well we will not mention their dual equivalents explicitly. At this point it might seem more logical to switch the notations $\tilde{N}$ and N around, but our choice will become clear in a moment. Using equation (26) we see that

$$\tilde{G}^{(p)}(0) \;=\; 0 \quad \text{for} \quad 0 \leq p < N.$$

is equivalent to

$$H^{(p)}(\pi) \;=\; 0 \quad \text{for} \quad 0 \leq p < N.$$

This means we can factor H as

$$H(\omega) \;=\; \left(\frac{1+e^{-i\omega}}{2}\right)^N K(\omega),$$

with $K(0) = 1$ and $K(\pi) \neq 0$. This factorization together with the (bi)orthogonality conditions is used as a starting point for construction of compactly supported wavelets [24, 37]. We also have that

$$i^p \hat{\phi}^{(p)}(2k\pi) \;=\; \delta_k \, \mathcal{M}_p \quad \text{for} \quad 0 \leq p < N, \tag{29}$$

and, by the Poisson summation formula, that

$$\sum_l (x-l)^p \, \phi(x-l) \;=\; \mathcal{M}_p \quad \text{for} \quad 0 \leq p < N.$$

By rearranging the last expression we see that any polynomial with degree smaller than N can be written as a linear combination of the functions $\phi(x-l)$ with $l \in \mathbb{Z}$. The coefficients in the linear combination themselves are polynomials in l. Or, stated in another way, if Π^p denotes the set of polynomials of degree p,

$$\forall A \in \Pi^{N-1}, \, \exists B \in \Pi^{N-1} : A(x) \;=\; \sum_l B(l) \, \phi(x-l). \tag{30}$$

The fact that B is indeed a polynomial can easily be seen from

$$B(l) \;=\; \int A(x) \, \tilde{\phi}(x-l) \, dx \;=\; \int A(x+l) \, \tilde{\phi}(x) \, dx.$$

Also

$$A(x) \;=\; \sum_l B(x-l) \, \phi(l),$$

which is true because left and right hand side are polynomials that match at every integer.

We already saw that the number of vanishing wavelet moments is important for the characterization of singularities. It also defines the convergence rate of the wavelet approximation for smooth functions [55, 102, 103], since if $f \in \mathcal{C}^N$, then

$$\| P_j f(x) - f(x) \| \;=\; \mathcal{O}(h^N) \quad \text{with} \quad h = 2^{-n}.$$

In fact, the conditions (29) are usually referred to as the Strang–Fix conditions, and these conditions were established long before the development of wavelet theory.

An asymptotic error expansion in powers of h, which can be used in numerical extrapolation, is derived in [106]. There it is also proved that the wavelet approximation of a smooth function interpolates the function in almost twice the number of points as compared to the number of basis functions.

The exponent N in the factorization of H also plays a role in the regularity of ϕ. The regularity is $N-1$ at most, but in many cases it is smaller due to the influence of K. The regularity of solutions of refinement equations is studied in detail in [32, 42, 43, 54, 95].

9 The fast wavelet transform

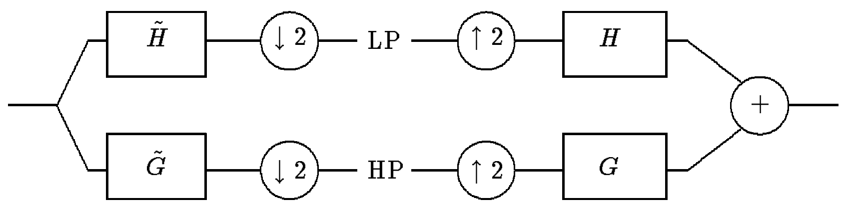

Figure 1: The subband coding scheme.

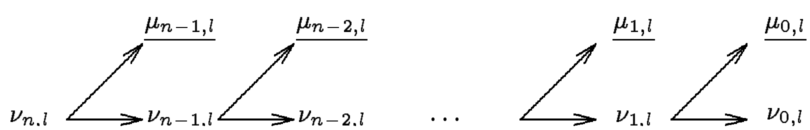

Figure 2: The decomposition scheme.

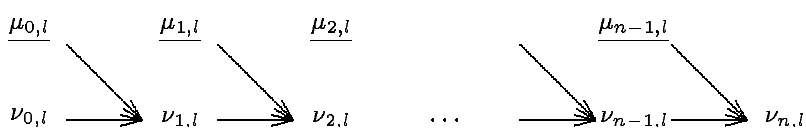

Figure 3: The reconstruction scheme.

Since V_j is equal to $V_{j-1} \oplus W_{j-1}$, a function $v_j \in V_j$ can be written uniquely as the sum of a function $v_{j-1} \in V_{j-1}$ and a function $w_{j-1} \in W_{j-1}$:

$$\begin{aligned} v_j(x) &= \sum_k \nu_{j,k}\, \phi_{j,k}(x) \;=\; v_{j-1}(x) + w_{j-1}(x) \\ &= \sum_l \nu_{j-1,l}\, \phi_{j-1,l}(x) + \sum_l \mu_{j-1,l}\, \psi_{j-1,l}(x). \end{aligned}$$

There is a one-to-one relationship between the coefficients of these functions. The decomposition formulae can be found using (22):

$$\begin{aligned} \nu_{j-1,l} &= \langle\, v_j, \tilde{\phi}_{j-1,l} \,\rangle \;=\; \sqrt{2}\, \langle\, v_j, \sum_k \tilde{h}_{k-2l}\, \tilde{\phi}_{j,k} \,\rangle \\ &= \sqrt{2} \sum_k \tilde{h}_{k-2l}\, \nu_{j,k}, \end{aligned} \tag{31}$$

and, similarly,

$$\mu_{j-1,l} = \sqrt{2} \sum_k \tilde{g}_{k-2l}\, \nu_{j,k}. \tag{32}$$

The reconstruction step involves calculating the $\nu_{j,k}$ from the $\nu_{j-1,l}$ and the $\mu_{j-1,l}$. Using (27) we have

$$\nu_{j,k} = \sqrt{2} \sum_l h_{k-2l}\, \nu_{j-1,l} + \sqrt{2} \sum_l g_{k-2l}\, \mu_{j-1,l}. \tag{33}$$

When applied recursively, these formulae define a transformation, the *fast wavelet transform* [85, 86].

In signal processing this technique is known as subband coding or more specifically as quadrature mirror filtering. Quadrature mirror filters were originally studied before wavelet theory. The decomposition step consists of applying a low-pass ($\tilde{H}$) and a band-pass ($\tilde{G}$) filter followed by downsampling ($\downarrow 2$) (i. e. retaining only the even index samples), see figure 1. The reconstruction consists of upsampling ($\uparrow 2$) (i.e. adding a zero between every two samples) followed by filtering and addition. One can show that the conditions (24) correspond to the exact reconstruction of a subband coding scheme. More details can be found in [96, 108, 109, 110].

An interesting problem is: given a function f, determine, with a certain accuracy and in a computationally favorable way, the coefficients $\nu_{n,l}$ of a function in the space V_n which are needed to start the fast wavelet transform. A trivial solution could be

$$\nu_{n,l} \;=\; f(l/2^n).$$

Other sampling procedures, such as (quasi-)interpolation and quadrature formulae were proposed in [73, 100, 105, 111]

An implementation of a fast wavelet transform in pseudo code is given in the appendix.

10 Examples of wavelets

Now that we have discussed the essentials of wavelet multiresolution analysis, we shall take a look at which properties of wavelets are important.

Orthogonality: If the wavelets are orthogonal, the fast wavelet transform has perfect numerical condition and stable numerical computation is ensured. If the multiresolution analysis is orthogonal (remember that this includes semiorthogonal wavelets), the projection operators onto the different subspaces yield optimal approximations in the $L^2(\mathbb{R})$ sense.

Compact support: If the scaling function and wavelet are compactly supported, the filters H and G are finite impulse response filters, which is of use in implementations. If they are not compactly supported, a fast decay is desirable so the filters can be approximated reasonably well by finite impulse response filters.

Rational coefficients: For computer implementations it is of use if the filter coefficients h_k and g_k are rationals or, even better, dyadic rationals. Dividing by a power of two on a computer just corresponds to shifting bits.

Symmetry: If the scaling function and wavelet are (anti-)symmetric, then the filters have generalized linear phase. If they don't have this property, this can lead to phase distortion.

Smoothness: Smoothness is of importance in compression applications. Compression is usually achieved by setting coefficients $\mu_{j,l}$ to zero and this corresponds to leaving out a component $\mu_{j,l}\,\psi_{j,l}(x)$ from the original function. If the original function represents e.g. an image and the wavelet is not smooth, the error can easily be caught by the human eye. From this simple argument we also conclude that the smoothness of the primary functions is more important to this aspect than that of the dual. Also more smoothness corresponds to better frequency localization of the filters. Finally, smooth basis functions are desired in applications to numerical analysis where derivatives are involved.

Number of vanishing moments: As we saw this can be important in singularity detection and characterization of smoothness spaces and it determines the convergence rate of wavelet approximations of smooth functions. We also mentioned that the number of vanishing moments is connected to the smoothness of the wavelet.

Analytic expressions: As already mentioned, there is in general no analytic expression for a scaling function or wavelet. In some cases an analytic expression is available and is nice to have. In harmonic analysis, analytic expressions of the Fourier transform are particularly useful.

Interpolation: If the scaling function satisfies

$$\phi(k) \;=\; \delta_k \quad \text{for} \quad k \in \mathbb{Z},$$

then it is trivial to find the function of V_j that interpolates data sampled on a grid with spacing 2^{-j}, as the coefficients are just the sample values.

As could be expected, it will not be possible to construct wavelets that have all these properties and there is a trade-off between them. So we will have to settle for a compromise. We will take a look at several solutions.

Examples of orthogonal wavelets:

- Two simple examples of orthogonal scaling functions are the box function $\chi_{[0,1]}(x)$ and the Shannon sampling function $\text{sinc}(\pi x)$. The orthogonality conditions are trivial to verify here either in time or frequency space. The corresponding wavelet for the box function is the Haar wavelet

$$\psi_{Haar}(x) = \chi_{[0,1/2]}(x) - \chi_{[1/2,1]}(x),$$

  and the Shannon wavelet is

$$\psi_{Shannon}(x) = \frac{\sin(2\pi x) - \sin(\pi x)}{\pi x}.$$

  These two, however, are not very useful in practice, since the first has very low regularity and the second has very slow decay.

- A more interesting example is the Meyer wavelet and scaling function [88]. These functions belong to $\mathcal{C}^\infty$ and have faster than polynomial decay. Their Fourier transform is compactly supported. The scaling function and wavelet are symmetric around 0 and 1/2 respectively and the wavelet has an infinite number of vanishing moments.

- The Battle-Lemarié wavelets are constructed by orthogonalizing B–spline functions using (17) and have exponential decay [7, 78]. The wavelet with N vanishing moments is a piecewise polynomial of degree $N-1$ that belongs to $\mathcal{C}^{N-2}$.

- Probably the most commonly used orthogonal wavelets are the original Daubechies wavelets [37, 40]. It is a family of orthogonal wavelets indexed by $N \in \mathbb{N}$, where N is the number of vanishing wavelet moments. They are supported on an interval of length $2N-1$. A disadvantage is that, except for the Haar wavelet (which has $N = 1$), they cannot be symmetric or antisymmetric. Their regularity increases linearly with N and is approximately equal to $0.3N$. In [39] three variations of this family, all three with orthogonal and compactly supported functions, are constructed:

  1. The previous construction does not lead to a unique solution for fixed N and support length $2N-1$, and there is a family where for each N the solution with closest to linear phase (or closest to symmetry) is chosen. In fact it turns out that the original family corresponds to choosing the extremal phase.
  2. Another family has more regularity at the price of a slightly larger support length $(2N+1)$.
  3. In a third family also the scaling function has vanishing moments ($\mathcal{M}_p = 0$ for $0 < p < N$). This is of use in numerical analysis applications where inner products of arbitrary functions with scaling functions have to be calculated very fast [10]. Their construction was asked by Ronald Coifman and Ingrid Daubechies therefore named them *coiflets*. They are supported on an interval with length $3N-1$.

wavelet family	compact support		analytic expression		symmetry	orthogonality		compact support $\tilde{\psi}$
	primary	dual	primary	dual		semi	full	
a	x	x	o	o	o	x	x	o
b	x	x	x	o	x	o	o	o
c	x	o	x	x	x	x	o	o
d	o	o	o	o	x	x	x	x
e	o	o	x	x	x	x	x	o

a: Daubechies' orthogonal wavelets
b: biorthogonal spline-wavelets
c: semiorthogonal spline-wavelets
d: Meyer wavelet
e: orthogonal spline-wavelets

Table 1: A quick comparison of wavelet families.

Examples of biorthogonal wavelets:

- Biorthogonal wavelets were constructed by Albert Cohen, Ingrid Daubechies and Jean-Christophe Feauveau in [21, 24]. Here $\Delta(\omega)$ is chosen equal to $e^{-i\omega}$, and thus

$$G(\omega) \;=\; -e^{-i\omega}\,\overline{\tilde{H}(\omega+\pi)} \quad \text{and} \quad \tilde{G}(\omega) \;=\; -e^{-i\omega}\,\overline{H(\omega+\pi)}.$$

  The scaling functions are the cardinal B-splines and the wavelets too are spline functions. All functions including the dual ones have compact support and linear phase. Moreover, all filter coefficients are dyadic rationals. A disadvantage is that for small filter lengths, the dual functions have very low regularity.

- Examples of semiorthogonal wavelets are the ones constructed by Charles Chui and Jianzhong Wang in [17, 18, 19]. The scaling functions are cardinal B–splines of order m and the wavelet functions are splines with compact support $[0, 2m-1]$. All primary and dual functions still have generalized linear phase and all scaling and wavelet parameters are rationals. A powerful feature here is that analytic expressions for the wavelet, scaling function, and dual functions are available. A disadvantage is that the dual functions do not have compact support but instead have exponential decay. The same wavelets, but in a different setting, were also derived by Akram Aldroubi, Murray Eden and Michael Unser in [107].

- Other semiorthogonal wavelets can be found in [75, 90, 91, 94].

Some of these families and properties are summerized in table 1.

Examples of interpolating scaling functions:

- The Shannon sampling function

$$\phi_{Shannon} \;=\; \frac{\sin(\pi x)}{\pi x},$$

is an interpolating scaling function. It is bandlimited but has very slow decay.

- An interpolating scaling function, whose translates also generate V_0, can be found by letting

$$\hat{\phi}_{interpol}(\omega) \;=\; \frac{\hat{\phi}(\omega)}{\sum_l \phi(l) e^{-i\omega l}},$$

  provided that the denominator does not vanish [111]. Even if ϕ is compactly supported, $\phi_{interpol}$ is in general not compactly supported. The cardinal spline interpolation functions of even order are constructed this way [99].

- An interpolation scaling function can also be constructed from a pair of biorthogonal scaling functions as

$$\phi_{interpol}(x) \;=\; \int_{-\infty}^{+\infty} \phi(y+x)\, \overline{\tilde{\phi}(y)}\, dy.$$

  The interpolation property immediately follows from the biorthogonality condition. In the case of an orthogonal scaling function this is just its autocorrelation function. The interpolating function and its translates do not generate the same space as ϕ and its translates. This construction, started from the Daubechies orthogonal or biorthogonal wavelets, yields a family of interpolating functions which were studied by Gilles Deslauriers and Serge Dubuc in [45, 46]. These functions are smooth and compactly supported. More information can also be found in [50, 98].

11 Wavelets on closed sets

So far we have been discussing wavelet theory on the real line (and its higher dimensional analogs). For many applications the functions involved are only defined on a compact set, such as an interval or a square, and to apply wavelets then requires some modifications.

11.1 Simple solutions

To be specific, let us discuss the case of the unit interval $[0,1]$. Given a function f on $[0,1]$, the most obvious approach is to set $f(x) = 0$ outside $[0,1]$, and then use wavelet theory on the line. However, for a general function f this "padding with 0's" introduces discontinuities at the endpoints 0 and 1; consider for example the simple function $f(x) = 1$, $x \in [0,1]$. Now, as we have said earlier, wavelets are effective for detecting singularities, so artificial ones are likely to introduce significant errors.

Another approach, which is often better, is to consider the function to be periodic with period 1, $f(x+1) = f(x)$. Expressed in another way, we assume that the function is defined on the torus and identify the torus with $[0,1]$. Wavelet theory on the torus parallels that on the line. In fact, note that if f has period 1, then the the wavelet coefficients on a given scale satisfy $\langle f, \psi_{j,k} \rangle = \langle f, \psi_{j,k+2^j} \rangle$, $k \in \mathbb{Z}$, $j \geq 0$. This simple observation readily allows us to rewrite wavelet expansions on the line as analogous ones on the torus, with wavelets defined on $[0,1]$. A periodic multiresolution analysis on the interval $[0,1]$ can be constructed by periodizing the basis functions as follows,

$$\phi^*_{j,l}(x) = \chi_{[0,1]}(x) \sum_m \phi_{j,l}(x+m) \quad \text{for} \quad 0 \le l < 2^j \quad \text{and} \quad j \ge 0. \tag{34}$$

If the support of $\phi_{j,l}(x)$, is a subset of $[0,1]$, then $\phi^*_{j,l}(x) = \phi_{j,l}(x)$. Otherwise $\phi_{j,l}(x)$ is chopped into pieces of length 1 which are shifted onto $[0,1]$ and added up, yielding $\phi^*_{j,l}(x)$. Similar definitions hold for $\psi^*_{j,l}$, $\tilde{\phi}^*_{j,l}$ and $\tilde{\psi}^*_{j,l}$. The algorithm in the appendix uses this periodic fast wavelet transform . This "wrap around" procedure is satisfactory in many situations (and certainly takes care of functions like $f(x) = 1$, $x \in [0,1]$, for example). However, unless the behavior of the function f at 0 matches that at 1, then the periodic version of f will have singularities there. A simple function like $f(x) = x$, $x \in [0,1]$, gives a good illustration of this.

A third method, which works if the basis functions are symmetric, is to use reflection across the edges. This preserves continuity, but introduces discontinuities in the first derivative. This solution is sometimes satisfactory in image processing applications.

11.2 Meyer's boundary wavelets

What really is needed then are wavelets intrinsically defined on $[0,1]$. Such wavelets were recently given by Yves Meyer [89], and we shall sketch his construction next. We start from the Daubechies wavelets and a scaling function with $2N$ non-zero coefficients:

$$\phi(x) \;=\; 2 \sum_{k=0}^{2N-1} h_k\, \phi(2x-k). \tag{35}$$

It easy to see that $\text{clos}\{x : \phi(x) \neq 0\} = [0, 2N-1]$, and, as a consequence,

$$B_{j,k} \;=\; \text{clos}\{x : \phi_{j,k}(x) \neq 0\} \;=\; [2^{-j}k, 2^{-j}(k+2N-1)]. \tag{36}$$

This implies that for sufficiently small scales 2^{-j}, $j \ge j_0$, say, a function $\phi_{j,k}$ can only intersect at most one of the endpoints 0 or 1. Let us restate this in a different way. Define the set of indices

$$S_j \;=\; \{k : B_{j,k} \cap (0,1) \neq \emptyset\}.$$

We define three subsets of this set containing the indices of the basis functions at the left boundary, in the interior, and at the right boundary:

$$\begin{aligned} S_j^{(1)} &= \{k : 0 \in B^\circ_{j,k}\} \\ S_j^{(2)} &= \{k : (0,1) \subset B^\circ_{j,k}\} \\ S_j^{(3)} &= \{k : 1 \in B^\circ_{j,k}\}. \end{aligned}$$

Here E° denotes the interior of the set E. The sets $S_j^{(1)}$ and $S_j^{(3)}$ are disjoint for sufficiently large j. We also have that

$$S_j \;=\; S_j^{(1)} \cup S_j^{(2)} \cup S_j^{(3)},$$

and the sets on the right are all pairwise disjoint. It is easy to write down what these sets are more explicitly:

$$
\begin{aligned}
S_j^{(1)} &= \{k : -2N+2 \le k \le -1\} \\
S_j^{(2)} &= \{k : 0 \le k \le 2^j - 2N + 1\} \\
S_j^{(3)} &= \{k : 2^j - 2N + 2 \le k \le 2^j - 1\}.
\end{aligned}
$$

Note, in particular, that the sets $S_j^{(1)}$ and $S_j^{(3)}$ contain $2N-2$ functions, independently of j. We now let $V_j^{[0,1]}$ denote the restriction of functions in V_j:

$$V_j^{[0,1]} \;=\; \{f : f(x) = g(x),\, x \in [0,1], \text{ for some function } g \in V_j\}.$$

Clearly, since the V_j's form an increasing sequence of spaces,

$$V_j^{[0,1]} \;\subset\; V_{j+1}^{[0,1]},$$

and $V_j^{[0,1]}$, $j \ge j_0$, form a multiresolution analysis of $L^2([0,1])$. It is also obvious that the functions in $\{\phi(x-l)|_{[0,1]} : l \in S_j\}$ span $V_j^{[0,1]}$. Here $g(x)\,|_{[0,1]}$ denotes the restriction of $g(x)$ to $[0,1]$. Not quite as obvious, but still easy, is the fact that the functions in this collection are linearly independent and, hence, form a basis for $V_j^{[0,1]}$. In order to obtain an orthonormal basis, we may argue as follows. As long as the function $\phi_{j,k}$ lives entirely inside $[0,1]$, restricting it to $[0,1]$ has no effect. In particular, the functions $\phi_{j,k}$, $k \in S_j^{(2)}$ are still pairwise orthogonal. A key observation now is that for $k \in S_j^{(1)}$, $l \in S_j^{(2)} \cup S_j^{(3)}$,

$$\langle\, \phi_{j,k}, \phi_{j,l} \,\rangle_{[0,1]} \;=\; \int_0^1 \phi_{j,k}(x)\, \phi_{j,l}(x)\, dx \;=\; \int_{-\infty}^{+\infty} \phi_{j,k}(x)\, \phi_{j,l}(x)\, dx \;=\; 0, \tag{37}$$

and similarly when $k \in S_j^{(3)}$, $l \in S_j^{(2)} \cup S_j^{(1)}$. Hence, we see that the three collections $\{\phi(x-l)|_{[0,1]} : l \in S_j^{(1)}\}$, $\{\phi(x-l)|_{[0,1]} : l \in S_j^{(2)}\}$, and $\{\phi(x-l)|_{[0,1]} : l \in S_j^{(3)}\}$ are mutually orthogonal. So, since the functions in $\{\phi(x-l)|_{[0,1]} : l \in S_j^{(2)}\}$ already form an orthonormal set, there only remains to separately orthonormalize the functions in $\{\phi(x-l)|_{[0,1]} : l \in S_j^{(1)}\}$ and in $\{\phi(x-l)|_{[0,1]} : l \in S_j^{(3)}\}$. This is easily accomplished with a Gram-Schmidt procedure.

Now, if we let $W_j^{[0,1]}$ denote the restriction of functions in W_j to $[0,1]$, then we have that

$$V_{j+1}^{[0,1]} \;=\; V_j^{[0,1]} + W_j^{[0,1]}. \tag{38}$$

So, the basis elements in $V_j^{[0,1]}$ together with the restriction of the wavelets $\psi_{j,k}$ to $[0,1]$ span $V_{j+1}^{[0,1]}$. However there are $2^j + 2N - 2$ wavelets that intersect $[0,1]$, and since $\dim V_{j+1}^{[0,1]} - \dim V_j^{[0,1]} = 2^j$ we have too many functions. The restrictions of the wavelets in W_j that live entirely inside $[0,1]$ are still mutually orthogonal and, by an observation similar to (37), they are also orthogonal to $V_j^{[0,1]}$. Among the $2N-2$ that intersect the endpoints, we use (27) to find the redundant ones and remove them. After that we just apply a Gram-Schmidt argument again, and we have an orthonormal basis for $W_j^{[0,1]}$.

This elegant construction of Yves Meyer has a couple of disadvantages. Among the functions $\phi_{j,k}$ that intersect $[0,1]$ there are some that are almost zero there. Hence, the set $\{\phi_{j,k}\}_{k\in S_j}$ is almost linearly dependent, and, as a consequence, the condition number of the matrix, corresponding to the change of basis from $\{\phi_{j,k}\}_{k\in S_j}$ to the orthonormal one, becomes quite large. Furthermore, we have $\dim V_j^{[0,1]} \neq \dim W_j^{[0,1]}$ which means that there is an inherent imbalance between the spaces $V_j^{[0,1]}$ and $W_j^{[0,1]}$, which is not present in the case of the whole real line.

11.3 Dyadic boundary wavelets

As we noted earlier (30) all polynomials of degree $\leq N-1$ are in V_j. Hence, the restriction of such polynomials to $[0,1]$ are in $V_j^{[0,1]}$. Since this fact is directly linked to many of the approximation properties of wavelets, any construction of a multiresolution analysis on $[0,1]$ should preserve this. The construction in [25] uses this as a starting point and is slightly different than the one by Yves Meyer. Let us briefly describe this construction as well. Again we start with the scaling function ϕ from the Daubechies construction with $2N$ non-zero scaling parameters, and assume that we have picked the scale fine enough so that the endpoints are independent as before. By (30) and since the $\{\phi_{j,k}\}$ is an orthonormal basis for V_j, each monomial x^α, $\alpha \leq N-1$, has the representation $x^\alpha = \sum_k \langle\, x^\alpha, \phi_{j,k}\,\rangle\, \phi_{j,k}(x)$. The restriction to $[0,1]$ can then be written

$$x^\alpha|_{[0,1]} \;=\; \Big(\sum_{k=-2N+2}^{0} + \sum_{k=1}^{2^j-2N} + \sum_{k=2^j-2N+1}^{2^j-1}\Big)\,\langle\, x^\alpha, \phi_{j,k}\,\rangle\, \phi_{j,,k}(x)|_{[0,1]}.$$

If we let

$$x_{j,L}^\alpha \;=\; 2^{j(\alpha+1/2)} \sum_{k=-2N+2}^{0} \langle\, x^\alpha, \phi_{j,k}\,\rangle\, \phi_{j,k}(x)|_{[0,1]}$$

and, similarly,

$$x_{j,R}^\alpha \;=\; 2^{j(\alpha+1/2)} \sum_{k=2^j-2N+1}^{2^j-1} \langle\, x^\alpha, \phi_{j,k}\,\rangle\, \phi_{j,k}(x)|_{[0,1]},$$

then

$$2^{j/2}(2^j x)^\alpha|_{[0,1]} \;=\; x_{j,L}^\alpha + 2^{j(\alpha+1/2)} \sum_{k=1}^{2^j-2N} \langle\, x^\alpha, \phi_{j,k}\,\rangle\, \phi_{j,k}(x)|_{[0,1]} + x_{j,R}^\alpha.$$

The spaces $\bar{V}_j$, $j \geq j_0$, that will form our multiresolution analysis of $L^2([0,1])$, we take to be the linear span of the functions $\{x_{j,L}^\alpha\}_{\alpha\leq N-1}$, $\{x_{j,R}^\alpha\}_{\alpha\leq N-1}$, and $\{\phi_{j,k}|_{[0,1]}\}_{k=1}^{2^j-2N}$:

$$\bar{V}_j \;=\; \overline{\{x_{j,L}^\alpha\}_{\alpha\leq N-1}} \,\cup\, \overline{\{\phi_{j,k}\}_{k=1}^{2^j-2N}} \,\cup\, \overline{\{x_{j,R}^\alpha\}_{\alpha\leq N-1}}$$

Finding an orthonormal basis for $\bar{V}_j$ is easy; in fact, the collections $\{x_{j,L}^\alpha\}_{\alpha\leq N-1}$, $\{\phi_{j,k}\}_{k=1}^{2^j-2N}$, and $\{x_{j,R}^\alpha\}_{\alpha\leq N-1}$ are mutually orthogonal, and all of the functions in these are linearly independent. We thus only have to orthonormalize the functions $x^{\alpha}_{j,L}$ and $x^{\alpha}_{j,R}$ to get our orthonormal basis. Note that, by construction, $\dim \bar{V}_j = 2^j$ and all polynomials of degree $\leq N-1$ are in $\bar{V}_j$. It is also easy to see that

$$\bar{V}_j \subset \bar{V}_{j+1}.$$

To get to the corresponding wavelets we let $\bar{W}_j$ be the orthogonal complement of $\bar{V}_j$ in $\bar{V}_{j+1}$. The wavelets $\psi_{j,k}$ with $1 \leq k \leq 2^j - 2N$ are all in $\bar{V}_{j+1}$ and live entirely inside $[0,1]$. The remaining $2N$ functions required for an orthonormal basis of $\bar{W}_j$, can be found, for example by using (27) again.

This last construction carries over to more general situations [71]; for example, we can also use biorthogonal wavelets and much more general closed sets than $[0,1]$.

There are also other constructions of wavelets on $[0,1]$. In fact, for historical perspective it is interesting to notice that Franklin's original construction [56] was given for $[0,1]$. Another interesting one, in the case of semiorthogonal spline-wavelets, has been given by Charles Chui and Ewald Quak [13]; we refer to the original paper for details.

12 Wavelet packets

A simple, but most powerful extension of wavelets and multiresolution analysis are wavelet packets [29, 31]. In this section it will be useful to switch to the following notation,

$$m_e(\omega) = H^e(\omega)\, G^{1-e}(\omega) \quad \text{for} \quad e = 0, 1.$$

The fundamental observation is the following fact, called the *splitting trick* [16, 22, 88]:

Suppose that the set of functions $\{f(x-k) \mid k \in \mathbb{Z}\}$ is a Riesz basis for its closed linear span S, then the functions

$$f^0_k = \frac{1}{\sqrt{2}} f^0(x/2 - k) \quad \text{and} \quad f^1_k = \frac{1}{\sqrt{2}} f^1(x/2-k) \quad \text{for} \quad k \in \mathbb{Z}.$$

also constitute a Riesz basis for S, where

$$\hat{f}^e(\omega) = m_e(\omega/2)\, \hat{f}(\omega/2).$$

We see that the classical multiresolution analysis is obtained by splitting V_j with this trick into V_{j-1} and W_{j-1} and then doing the same for V_{j-1} recursively. The wavelet packets are the basis functions that we obtain if we also use the splitting trick on the W_j spaces. So starting from a space V_j, we obtain, after applying the splitting trick L times, the basis functions

$$\psi^L_{e_1,\ldots,e_L;j,k}(x) = 2^{(j-L)/2} \psi^L_{e_1,\ldots,e_L}(2^{j-L}x - k),$$

with

$$\hat{\psi}^L_{e_1,\ldots,e_L}(\omega) = \prod_{i=1}^{L} m_{e_i}(2^{-i}\omega)\, \hat{\phi}(2^{-L}\omega).$$

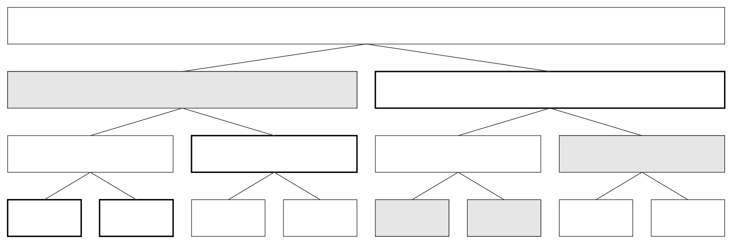

Figure 4: Wavelet packets scheme.

So, after L splittings, we have 2^L basis functions and their translates over integer multiples of 2^{L-j} as a basis of V_j. The connection between the wavelet packets and the wavelet and scaling functions is

$$\phi = \psi^L_{0,\ldots,0} \quad \text{and} \quad \psi = \psi^L_{1,0,\ldots,0}.$$

However, we do not necessarily have to split each subspace at every stage. In figure 4 we give a schematical representation of a space and its subspaces after using the splitting on 4 levels. The top rectangle could e.g. represent the space V_4 and each other rectangle then corresponds to a certain subspace of V_4 generated by wavelet packets. The slanted lines between the rectangles indicate the splitting, the left referring to the filter m_0 and the right to m_1. The bold rectangles then correspond to the wavelet multiresolution analysis $V_4 = V_0 \oplus W_0 \oplus W_1 \oplus W_2$. The shaded rectangles correspond to a possible wavelet packet splitting and a basis with functions

$$\left\{\psi^1_1(4x-k),\, \psi^2_{1,1}(2x-k),\, \psi^3_{0,0,1}(x-k),\, \psi^3_{1,0,1}(x-k) \mid k \in \mathbb{Z}\right\}.$$

For the dual functions, a similar procedure has to be followed.

In the Fourier domain the splitting trick corresponds to dividing the frequency interval essentially represented by the original space into a lower and an upper part. So the wavelet packets allow more flexibility in adapting the basis to the frequency contents of a signal.

It is easy to develop a fast wavelet packet transform. It just involves applying the same low and band pass filters also to the coefficient of functions of W_j again in a iterative manner. This means that starting from M samples we construct a full binary tree with $(M \log_2 M)$ entries. The power of this construction lies in the fact that we have much more freedom in deciding which basis functions we will use to represent the given function. We can choose to use the set of M coefficients of the tree to represent the function that is optimal with respect to a certain criterion. This idea is called *best basis selection*, and, using the tree structure, it can be implemented in $\mathcal{O}(M)$ operations. The criterion is determined by the application, and which basis functions will end up in the basis depends on the data.

Entropy based criteria for applications in image compression, were proposed in [27]. Applications in signal processing can be found in [30, 112].

This wavelet packets construction can also be combined with wavelets on closed sets and wavelets in higher dimensions.

13 Multidimensional wavelets

Up till now we have focused on the one-dimensional situation. However, there are also wavelets in higher dimensions. A simple way to obtain these is to use tensor products. To fix ideas, let us consider the case of the plane. Let

$$\Phi(x, y) \;=\; \phi(x)\,\phi(y) \;=\; \phi \otimes \phi(x, y),$$

and define

$$V_0 \;=\; \{f : f(x, y) = \sum_{k_1, k_2} \lambda_{k_1, k_2}\,\Phi(x - k_1, y - k_2), \lambda \in l^2(\mathbb{Z}^2)\}.$$

Of course, if $\{\phi(x - l) \mid l \in \mathbb{Z}\}$ is an orthonormal set, then $\{\Phi(x - k_1, y - k_2)\}$ form an orthonormal basis for V_0. By dyadic scaling we obtain a multiresolution analysis of $L^2(\mathbb{R}^2)$. The complement W_0 of V_0 in V_1 is similarly generated by the translates of the three functions

$$\Psi^{(1)} \;=\; \phi \otimes \psi, \quad \Psi^{(2)} \;=\; \psi \otimes \phi, \quad \text{and} \quad \Psi^{(3)} \;=\; \psi \otimes \psi. \tag{39}$$

There is another, perhaps even more straightforward, wavelet decomposition in higher dimensions. By carrying out a one-dimensional wavelet decomposition for each variable separately, we obtain

$$f(x, y) \;=\; \sum_{i,l} \sum_{j,k} \langle\, f, \psi_{i,l} \otimes \psi_{j,k} \,\rangle\; \psi_{i,l} \otimes \psi_{j,k}(x, y). \tag{40}$$

Note that the functions $\psi_{i,l} \otimes \psi_{j,k}$ involve two scales, 2^{-i} and 2^{-j}, and each of these functions are (essentially) supported on a rectangle. The decomposition (40) is therefore called the *rectangular wavelet decomposition* of f while the functions in (39) are the basis functions of the *square wavelet decomposition*. For both decompositions, the corresponding fast wavelet transform consists of applying the one-dimensional fast wavelet transform to the rows and columns of a matrix.

There are also several other extensions to higher dimensions. We mention nonseparable basis functions [23, 44, 94, 101], other lattices corresponding to different symmetries [26], Clifford valued wavelets [3], etc. However we leave these topics for now.

14 Applications

14.1 Data compression

One of the applications of wavelet theory is data compression. There are two basic kinds of compression schemes: lossless and lossy. In the case of lossless compression one is interested in reconstructing the data exactly, without any loss of information. We shall consider here lossy compression. Here we are ready to accept an error as long as the quality after compression is acceptable. With lossy compression schemes we potentially can achieve much higher compression ratios than with lossless compression.

To be specific, let us assume that we are given a digitized image. The compression ratio is defined as the number of bits the initial image takes to store on the computer divided by the number of bits required to store the compressed image. The interest in compression in general has grown as the amount

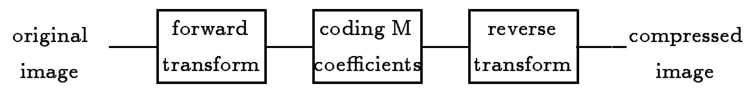

Figure 5: Image transform coding.

of information we pass around has increased. This is easy to understand when we consider the fact that to store a moderately large image, say a 512×512 pixels, 24 bit color image, takes about 0.75 MBytes. This is only for still images; in the case of video, the situation becomes even worse. Then we need this kind of storage for each frame and we have something like 30 frames per second. There are several other reasons than just the storage requirement for the interest in compression techniques. However, instead of going into this, let us now look at the connection with wavelet theory.

First, let us define, somewhat mathematically, what we mean by an image. Let us for simplicity discuss an $L \times L$ grayscale image with 256 grayscales (i.e. 8 bit). This can be considered to be a piecewise constant function f defined on a square

$$f(x,y) \;=\; p_{ij}, \quad \text{for} \quad i \leq x < i+1 \quad \text{and} \quad j \leq y < j+1 \quad \text{and} \quad 0 \leq i,j < L,$$

where $0 \leq p_{ij} \leq 255$ are integers. Now, one of the standard procedures for lossy compression is through transform coding, see figure 5. The most common transform used in this context is the "Discrete Cosine Transform" which uses a Fourier transform of the image f. However, we are more interested in the case when the transform is the fast wavelet transform.

There are in fact several ways to use the wavelet transform for compression purposes [83, 84]. One way is to consider compression to be an approximation problem [47, 48]. More specifically, let us fix an orthogonal wavelet ψ. Given an integer $M \geq 1$ we try to find the "best" approximation of f by using a representation

$$f_M(x) \;=\; \sum_{kl} b_{jk}\,\psi_{jk}(x) \quad \text{with } M \text{ non-zero coefficients } b_{jk}. \tag{41}$$

The basic reason why this potentially might be useful is that each wavelet picks up information about the image f essentially at a given location and at a given scale. Where the image has more interesting features, we can spend more coefficients, and where the image is nice and smooth we can use fewer and still get good quality of approximation. In other words, the wavelet transform allows us to focus on the most relevant parts of f. Now, to give this mathematical meaning we need to agree on an error measure. Ideally, for image compression we should use a norm that corresponds as closely as possible to the human eye. However, let us make it simple and discuss the case of L^2.

So we are interested in finding an optimal approximation minimizing the error $\|f - f_M\|_{L^2}$. Because of the orthogonality of the wavelets this equals

$$\Big(\sum_{jk} |\,\langle\, f, \psi_{jk}\,\rangle \,-\, b_{jk}|^2\Big)^{1/2}. \tag{42}$$

A moments thought, reveals that the best way to pick M non-zero coefficients b_{jk}, making the error as small as possible, is by simply picking the M coefficients with largest absolute value, and setting $b_{j,k} = \langle f, \psi_{jk} \rangle$ for these numbers. This then yields the optimal approximation f_M^{opt}.

Another fundamental question is which images can be approximated well by using the procedure just sketched. Let us take this to mean that the error satisfies

$$\|f - f_M^{opt}\|_{L^2} \;=\; \mathcal{O}(M^{-\beta}), \tag{43}$$

for some $\beta > 0$. The larger β, the faster the error decays as M increases and the fewer coefficients are generally needed to obtain a given error. The exponent β can be found easily, in fact it can be shown that

$$\left(\sum_{M \geq 1} (M^{\beta} \|f - f_M^{opt}\|_{L^2})^p \frac{1}{M} \right)^{1/p} \;\approx\; (\sum_{jk} | \langle f, \psi_{jk} \rangle |^p)^{1/p} \tag{44}$$

with $1/p = 1/2 + \beta$. The maximal β for which (43) is valid can be estimated by finding the smallest p for which the right hand side of (44) is finite. The expression on the right is one of many equivalent norms on the Besov space $\dot{B}_p^{2\beta,p}$ (Besov spaces are smoothness spaces generalizing the Lipschitz continuous functions). The β in the left hand side of (44) is actually not exactly the same as in (43). However, for practical purposes, the difference is of no consequence.

14.2 Numerical analysis

As mentioned earlier, interest in wavelets historically grew from the fact that they are effective tools for studying problems in partial differential equations and operator theory. More specifically, they are useful for understanding properties of so-called Calderón-Zygmund operators.

Let us first make a general observation about the representation of a linear operator T and wavelets. Suppose that f has the representation

$$f(x) \;=\; \sum_{jk} \langle f, \psi_{jk} \rangle \, \psi_{jk}(x).$$

Then

$$Tf(x) \;=\; \sum_{jk} \langle f, \psi_{jk} \rangle \, T\psi_{jk}(x),$$

and, using the wavelet representation of the function $T\psi_{jk}(x)$, this equals

$$\sum_{jk} \langle f, \psi_{jk} \rangle \sum_{il} \langle T\psi_{jk}, \psi_{il} \rangle \, \psi_{il}(x) \;=\; \sum_{il} \left(\sum_{jk} \langle T\psi_{jk}, \psi_{il} \rangle \; \langle f, \psi_{jk} \rangle \right) \psi_{il}(x).$$

In other words, the action of the operator T on the function f is directly translated into the action of the infinite matrix $A_T = \{ \langle T\psi_{jk}, \psi_{il} \rangle \}_{il,jk}$ on the sequence $\{ \langle f, \psi_{jk} \rangle \}_{jk}$. This representation of T as the matrix A_T is often referred to as the "standard representation" of T [10]. There is also a "nonstandard representation". For virtually all linear operators there is a function (or, more generally, a distribution) K such that

$$Tf(x) = \int K(x,y) f(y)\, dy.$$

The nonstandard representation of T is now simply the (two-dimensional) wavelet coefficients of the kernel K, using the square decomposition $\{ \langle K, \Psi^{(j)}_{k_1,k_2} \rangle \}$ (again, we have more than one wavelet function in two dimensions), while the standard representation corresponds to the rectangular decomposition.

Let us then briefly discuss the connection with Calderón-Zygmund operators. Consider a typical example. Let H be the Hilbert transform,

$$Hf(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(s)}{x-s}\, ds.$$

The basic idea is now that the wavelets ψ_{jk} are approximate eigenfunctions for this, as well as for many other related (Calderón-Zygmund) operators. We note that if ψ_{jk} were exact eigenfunctions, then we would have $H\psi_{jk}(x) = \lambda_{jk}\psi_{jk}(x)$, for some number λ_{jk} and the standard representation would be a diagonal "matrix":

$$A_H = \{ \langle H\psi_{il}, \psi_{jk} \rangle \} = \{ \lambda_{il} \langle \psi_{il}, \psi_{jk} \rangle \} = \{ \lambda_{il} \delta_{il,jk} \}$$

This is unfortunately not the case. However, it turns out that A_T is in fact an almost diagonal operator, in the appropriate, technical sense, with the off diagonal elements quickly becoming small. To get some idea why this is the case, note that for large $|x|$, we have, at least heuristically,

$$H\psi_{jk}(x) \approx \frac{1}{x} \int \psi_{jk}(y)\, dy.$$

A priori, the decay of the right hand side would thus be $\mathcal{O}(1/x)$, which of course is far from the rapid decay of a wavelet ψ_{jk} (some wavelets are even zero outside a finite set). Recall, however, that ψ_{jk} has at least one vanishing moment so the decay is in fact much faster than just $\mathcal{O}(1/x)$, and the shape of $H\psi_{jk}(x)$ closely resembles that of $\psi_{jk}(x)$.

So, for a large class of operators, the matrix representation, either the standard or the nonstandard, has a rather precise structure with many small elements. In this representation, we then expect to be able to compress the operator by simply omitting small elements. In fact, note that this is essentially the same situation, especially in the case of the nonstandard representation, as in the case of image compression, the "image" now being the kernel $K(x,y)$. Hence, if we could do basic operations such as inversion, and multiplication, with compressed matrices, rather than with the discretized versions of T, then we may significant speed up of the numerical treatment. This program of using the wavelet representations for the efficient numerical treatment of operators was initiated in [10]. We also refer to [1, 2] for related material and many more details.

In a different direction, because of the close similarities between the scaling function and finite elements, it seems natural to try wavelets where traditionally finite element methods are used, e.g. for solving boundary value problems [72]. There are interesting results showing that this might be fruitful; for example, it has been shown [11, 36, 92, 113]. that for many problems the condition number of the $N \times N$ stiffness matrix remains bounded as the dimension N goes to infinity. This is in contrast with the situation for regular finite elements where the condition number in general tends to infinity.

One of the first problems we have to address when discussing boundary problems on domains is how to take care of the boundary values and the fact that the problem is closely associated with a finite set rather than with the entire Euclidean plane. This is similar to the problem we discussed with wavelets on closed sets, and, indeed, the techniques discussed there can be often used to handle these two problems [4, 5].

Wavelets have also been used in the solution of evolution equations [6, 63, 76, 81]. A typical test problem here is Burgers' equation:

$$\frac{\partial u}{\partial t} + u\,\frac{\partial u}{\partial x} \;=\; \nu\,\frac{\partial^2 u}{\partial x^2}.$$

The time discretization is obtained here using standard schemes such as Crank-Nicholson or Adams-Moulton. Wavelets are used in the space discretization. Adaptivity can be used both in time and space [6].

One of the nice features of wavelets and finite elements is that they allow us to treat a large class of operators or partial differential equations in a unified way, allowing for example general pde solvers to be designed. In specific instances, though, it is sometimes possible to find particular wavelets, adapted to the operator or problem at hand. For example, Stefan Dahlke and Ilona Weinrich constructed wavelets adapted to a pseudo differential operator [35, 34]. In [9] Gregory Beylkin develops fast wavelet-based algorithms for the solution of differential equations.

Note: Applications in statistics such as the smoothing of data were investigated by David Donoho and Iain Johnstone in [51, 52, 53]

Acknowledgement

We would like to thank Gilbert Strang for useful comments and suggestions.

References

[1] B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin. Wavelet-like bases for the fast solution of second-kind integral equations. *SIAM J. Scient. Comp.*, 14, 1.

[2] B. K. Alpert. Wavelets and other bases for fast numerical linear algebra. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 181–216. Academic Press, 1992.

[3] L. Andersson, B. Jawerth, and M. Mitrea. The Cauchy singular integral operator and Clifford wavelets. In J. Benedetto and M. Frazier, editors, *Wavelets and Applications*. CRC Press, 1993.

[4] L. Andersson, B. Jawerth, and G. Peters. Wavelet theory on closed sets. In preparation.

[5] P. Auscher. Wavelets with boundary conditions on the interval. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 217–236. Academic Press, 1992.

[6] E. Bacry, S. Mallat, and G. Papanicolaou. A wavelet based space-time adaptive numerical method for partial differential equations. Technical Report 591, Courant Institute of Mathematical Sciences, 1991.

[7] G. Battle. A block spin construction of ondelettes. *Comm. Math. Phys.*, 110:601–615, 1987.

[8] M. A. Berger. Random affine iterated function systems: Curve generation and wavelets. *SIAM Review*, 31(4):614–627, 1989.

[9] G. Beylkin. On wavelet-based algorithms for solving differential equations. Preprint University of Colorado at Boulder, ftp from newton.colorado.edu.

[10] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Comm. Pure and Appl. Math.*, 44:141–183, 1991.

[11] G. Beylkin, R. Coifman, and V. Rokhlin. Wavelets in numerical analysis. In M. B. Ruskai et. al., editor, *Wavelets and Their Applications*, pages 181–210. Jones and Bartlett Publishers, 1992.

[12] A. Cavaretta, W. Dahmen, and C. Micchelli. Subdivision for Computer Aided Geometric Design. *Memoirs Amer. Math. Soc.*, 93, 1991.

[13] C. Chui and E. Quak. Wavelets on a bounded interval. In D. Braess and L. L. Schumaker, editors, *Numerical Methods of Approximation Theory*, pages 1–24. Birkhäuser Verlag, Basel, 1992.

[14] C. K. Chui. *An Introduction to Wavelets.* Academic Press, 1992.

[15] C. K. Chui, editor. *Wavelets: A Tutorial in Theory and Applications.* Academic Press, 1992.

[16] C. K. Chui and C. Li. Non-orthogonal wavelet packets. Technical Report CAT Report #261, Center for Approximation Theory, Texas A&M University, 1991.

[17] C. K. Chui and J. Z. Wang. On compactly supported spline wavelets and a duality principle. To appear in Trans. Amer. Math. Soc.

[18] C. K. Chui and J. Z. Wang. A cardinal spline approach to wavelets. *Proc. Amer. Math. Soc.*, 113:785–793, 1991.

[19] C. K. Chui and J. Z. Wang. A general framework of compactly supported splines and wavelets. *J. Approx. Th.*, 71(3):263–304, 1992.

[20] Z. Ciesielski. Constructive function theory and spline systems. *Studia Math.*, 52:277–302, 1973.

[21] A. Cohen. Biorthogonal wavelets. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 123–152. Academic Press, 1992.

[22] A. Cohen and I. Daubechies. On the instability of arbitrary biorthogonal wavelet packets. Preprint.

[23] A. Cohen and I. Daubechies. Non-separable bidimensional wavelet bases. Preprint AT&T Bell Laboratories, New Jersey, 1991.

[24] A. Cohen, I. Daubechies, and J. Feauveau. Bi-orthogonal bases of compactly supported wavelets. To appear in Comm. Pure and Appl. Math.

[25] A. Cohen, I. Daubechies, B. Jawerth, and P. Vial. Multiresolution analysis, wavelets and fast algorithms on an interval. To appear in C. R. Acad. Sci. Paris.

[26] A. Cohen and J.-M. Schlenker. Compactly supported bidimensional wavelet bases with hexagonal symmetry. *Constructive approximation*, 9(2):209–236, 1993.

[27] R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. IEEE Trans. on Inf. Theory, special issue on wavelets, to appear in 1992.

[28] R. R. Coifman. A real variable characterization of H^p. *Studia Math*, 51, 1974.

[29] R. R. Coifman and Y. Meyer. Orthonormal wave packet bases. Preprint.

[30] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser. Signal processing and compression with wave packets. in Proceedings of the conference on Wavelets, Marseille, 1989.

[31] R. R. Coifman, Y. Meyer, and V. Wickerhauser. Size properties of wavelet packets. In M. B. Ruskai et. al., editor, *Wavelets and Their Applications*, pages 453–470. Jones and Bartlett Publishers, 1992.

[32] D. Colella and C. E. Heil. Characterizations of scaling functions. SIAM J. Matrix Anal. and Appl., To be published.

[33] J. M. Combes, A. Grossmann, and Ph. Tchamitchian, editors. *Wavelets: Time-Frequency Methods and Phase Space.* Inverse problems and theoretical imaging. Springer-Verlag, 1989.

[34] S. Dahlke and I. Weinrich. Wavelet bases adapted to pseudo-differential operators. Technical report, RWTH Aachen, 1992.

[35] S. Dahlke and I. Weinrich. Wavelet-Galerkin-methods: An adapted biorthogonal wavelet basis. *Constructive approximation*, 9(2):237–262, 1993.

[36] W. Dahmen and A. Kunoth. Multilevel preconditioning. Preprint Freie Universität Berlin.

[37] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.*, 41:909–996, 1988.

[38] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. on Inf. Theory*, 36(5):961–1005, 1990.

[39] I. Daubechies. Orthonormal bases of compactly supported wavelets II. Variations on a theme. Preprint AT&T Bell Laboratories, New Jersey, Submitted SIAM J. Math. Anal., 1991.

[40] I. Daubechies. *Ten Lectures on Wavelets.* Number 61 in CBMS-NSF Series in Applied Mathematics. SIAM Publications, Philadelphia, 1992.

[41] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27(5):1271–1283, 1986.

[42] I. Daubechies and J. C. Lagarias. Two-scale difference equations I. Existence and global regularity of solutions. *SIAM J. Math. Anal.*, 22(5):1388–1410, 1991.

[43] I. Daubechies and J. C. Lagarias. Two-scale difference equations II. Local regularity, infinite products of matrices and fractals. *SIAM J. Math. Anal.*, 23(4):1031–1079, 1992.

[44] C. de Boor, R. A. DeVore, and A. Ron. On the construction of multivariate (pre)wavelets. *Constructive approximation*, 9(2):123–166, 1993.

[45] G. Deslauriers and S. Dubuc. Interpolation dyadique. In *Fractals, Dimensions non-entières et applications*, Paris, 1987. Masson.

[46] G. Deslauriers and S. Dubuc. Symmetric iterative interpolation processes. *Constructive approximation*, 5(1):49–68, 1989.

[47] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. on Inf. Theory*, 38(2):719–746, 1992.

[48] R. A. DeVore, B. Jawerth, and B. J. Lucier. Surface compression. *Comp. Aid. Geom. Des.*, 9(3):219–239, 1992.

[49] R. A. DeVore and B. J. Lucier. Wavelets. In *Acta Numerica 1*, pages 1–56, 1991.

[50] D. L. Donoho. Interpolating wavelet transforms. Preprint Department of Statistics, Stanford University, 1992.

[51] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness by wavelet shrinkage. Preprint Department of Statistics, Stanford University, 1992.

[52] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. Preprint Department of Statistics, Stanford University, 1992.

[53] D. L. Donoho and I. M. Johnstone. New minimax theorems, thresholding, and adaptation. Preprint Department of Statistics, Stanford University, 1992.

[54] T. Eirola. Sobolev characterization of solutions of dilation equations. *SIAM J. Math. Anal.*, 23(4):1015–1030, 1992.

[55] G. Fix and G. Strang. Fourier analysis of the finite element method in Ritz-Galerkin theory. *Stud. Appl. Math*, 48:265–273, 1969.

[56] P. Franklin. A set of continuous orthogonal functions. *Math. Ann*, 100:522–529, 1928.

[57] M. Frazier and B. Jawerth. Decomposition of Besov spaces. *Indiana Univ. Math. J.*, 34(4):777–799, 1985.

[58] M. Frazier and B. Jawerth. The ϕ-transform and applications to distribution spaces. In M. Cwikel et al., editor, *Function Spaces and Applications*, number 1302 in Springer Lecture Notes in Math, pages 223–246, 1988.

[59] M. Frazier and B. Jawerth. A discrete transform and decompositions of distribution spaces. *J. Func. Anal*, 93:34–170, 1990.

[60] M. Frazier, B. Jawerth, and G. Weiss. *Littlewood-Paley theory and the study of function spaces.* Number 79 in Regional Conference Series in Mathematics. American Mathematical Society, Providence, 1991.

[61] J. Geronimo, D. Hardin, and P. R. Massopust. Fractal functions and wavelet expansions based on several scaling functions. Preprint, Department of Mathematics, Georgia Tech.

[62] J. Geronimo, D. Hardin, and P. R. Massopust. Fractal surfaces, multiresolution analyses and wavelet transforms. To appear in proc. of the Nato Advanced Research Workshop "Shape in Picture", 7–11 september 1992.

[63] R. Glowinski, W. Lawton, M. Ravechol, and E. Tenenbaum. Wavelet solution of linear and nonlinear elliptic parabolic and hyperbolic problems in one space dimension. Technical Report AD 890527.1.1, Aware Inc., 1989.

[64] A. Grossmann, R. Kronland-Martinet, and J. Morlet. Reading and understanding continuous wavelet transforms. In J. M. Combes, A. Grossmann, and P. Tchamitchian, editors, *Wavelets: Time-Frequency Methods and Phase Space*, pages 2–20. Springer-Verlag, 1989.

[65] A. Grossmann and J. Morlet. Decompostion of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.*, 15(4):723–736, 1984.

[66] A. Grossmann and J. Morlet. Decomposition of functions into wavelets of constant shape, and related transforms. In L. Streit, editor, *Mathematics and Physics, Lectures on Recent Results*, Singapore, 1985. World Scientific Publishing.

[67] A. Grossmann, J. Morlet, and T. Paul. Transforms associated to square integrable group representations I. General results. *J. Math. Phys.*, 26(10):2473–2479, 1985.

[68] A. Haar. Zur Theorie der orthogonalen Funktionen-Systeme. *Math. Ann.*, 69:331–371, 1910.

[69] D. P. Hardin, B. Kessler, and P. R. Massopust. Multiresolution analyses based on fractal functions. *J. Approx. Th.*, 71(1):104–120, 1992.

[70] C. E. Heil and D. F. Walnut. Continuous and discrete wavelet transforms. *SIAM Review*, 31(4):628–666, 1989.

[71] T. Huntsberger, B. Jawerth, S. Lopresto, G. Peters, and A. Tirumalai. Wavelets on closed sets and image processing. In preparation.

[72] S. Jaffard and Ph. Laurencot. Orthonormal wavelets, analysis of operators, and applications to numerical analysis. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 543–602. Academic Press, 1992.

[73] A. J. E. M. Janssen. The Zak transform: a signal transform for sampled time-continuous signals. *Philips J. Res.*, 43:23–69, 1988.

[74] B. Jawerth. On Besov spaces. Technical Report 1, Lund, 1977.

[75] R.-Q. Jia and C. A. Micchelli. Using the refinement equations for the construction of pre-wavelets II: Powers of two. In P. J. Laurent, A. Le Méhauté, and L. L. Schumaker, editors, *Curves and Surfaces*. Academic Press, New York, 1991.

[76] A. Latto and E. Tenenbaum. Compactly supported wavelets and the numerical solution of Burgers' equation. Technical Report AD 900307, Aware Inc., 1990.

[77] W. M. Lawton. Necessary and sufficient conditions for constructing orthonormal wavelets bases. *J. Math. Phys.*, 32(1):57–61, 1991.

[78] P.-G. Lemarié. Ondelettes a localisation exponentielle. *J. de Math. Pures et Appl.*, 67(3):227–236, 1988.

[79] P.-G. Lemarié, editor. *Les Ondelettes en 1989.* Number 1438 in Lecture Notes in Mathematics. Springer-Verlag, 1990.

[80] P.-G. Lemarié and Y. Meyer. Ondelettes et bases hilbertiennes. *Rev. Mat. Iberoamericana*, 2:1–18, 1986.

[81] Y. Maday, V. Perrier, and J.-C. Ravel. Adaptivité dynamique sur bases d'ondelettes pour l'approximation d'équitations aux dérivées partielles. *C. R. Acad. Sci. Paris*, I(312):405–410, 1991.

[82] S. Mallat and W. L. Hwang. Singularity detection and processing with wavelets. Preprint Courant Institute of Mathematical Sciences, New York University, 1991.

[83] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 14:710–732, 1992.

[84] S. Mallat and S. Zhong. Wavelet transform maxima and multiscale edges. In M. B. Ruskai et. al., editor, *Wavelets and Their Applications*, pages 67–104. Jones and Bartlett Publishers, 1992.

[85] S. G. Mallat. Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. on Acoust. Signal Speech Process.*, 37(12):2091–2110, 1989.

[86] S. G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, 315(1):69–87, 1989.

[87] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 11(7):674–693, 1989.

[88] Y. Meyer. *Ondelettes et Opérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III: (*with R. Coifman*), Opérateurs multilinéaires.* Hermann, Paris, English translation is being prepared by Cambridge University Press, 1990.

[89] Y. Meyer. Ondelettes sur l'intervalle. Technical Report 9020, CEREMADE, Université Paris-Dauphine, To appear in Rev. Mat. Iberoamer., 1991.

[90] C. A. Micchelli. Using the refinement equations for the construction of pre-wavelets. *Numerical Algorithms*, 1(1):75–116, 1991.

[91] C. A. Micchelli, C. Rabut, and F. I. Utretas. Using the refinement equations for the construction of pre-wavelets III: Elliptic splines. *Numerical Algorithms*, 1(1):331–352, 1991.

[92] P. Oswald. On discrete norm estimates related to multilevel preconditioners in the finite element method. Preprint, 1991.

[93] J. Peetre. *New Thoughts on Besov Spaces.* Duke Univ. Math. Series, Durham, NC, 1976.

[94] S. D. Riemenschneider and Z. Shen. Wavelets and pre-wavelets in low dimensions. *J. Approx. Th.*, 71(1):18–38, 1992.

[95] O. Rioul. Simple regularity cirteria for subdivision schemes. *SIAM J. Math. Anal.*, 23(6):1544–1576, 1992.

[96] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Proc. Mag.*, pages 14–38, October 1991.

[97] M. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael, editors. *Wavelets and their Applications.* Jones and Bartlett, 1992.

[98] N. Saito and G. Beylkin. Multiresolution representations using the autocorrelation functions of compactly supported wavelets. Preprint University of Colorado at Boulder.

[99] I. J. Schoenberg. Cardinal interpolation and spline functions. *J. Approx. Th.*, 2:167–627, 1969.

[100] M. J. Shensa. Wedding the à trous and Mallat algorithms. *IEEE Trans. on Signal Process.*, 40(10):2464–2482, 1992.

[101] J. Stöckler. Multivariate wavelets. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 325–356. Academic Press, 1992.

[102] G. Strang. Wavelets and dilation equations: A brief introduction. *SIAM Review*, 31(4):614–627, 1989.

[103] G. Strang and G. Fix. A Fourier analysis of the finite element variational method. In *Constructive aspects of Functional Analysis*, Rome, 1973. Edizione Cremonese.

[104] J. 0. Strömberg. A modified Franklin system and higher order spline systems on $\mathbb{R}^n$ as unconditional bases for Hardy spaces. In Beckner et al., editor, *Conference on Harmonic Analysis in Honor of Antoni Zygmund*, volume II, pages 475–494, Chicago, 1981.

[105] W. Sweldens and R. Piessens. Quadrature formulae for the calculation of the wavelet decomposition. Preprint Department of Computer Science, K.U.Leuven, Belgium.

[106] W. Sweldens and R. Piessens. Asymptotic error expansions of wavelet approximations of smooth functions. Technical Report TW164, Department of Computer Science, K.U.Leuven, Belgium, 1992.

[107] M. Unser, A. Aldroubi, and M. Eden. A family of polynomial spline wavelet transforms. NCRR report 153/90, To be published in Signal Processing.

[108] P. P. Vaidyanathan. Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having perfect reconstruction property. *IEEE Trans. on Acoust. Signal Speech Process.*, 36:476–492, 1987.

[109] M. Vetterli. Filter banks allowing perfect reconstruction. *Signal Process.*, 10:219–244, 1986.

[110] M. Vetterli and C. Herley. Wavelets and filter banks. IEEE Trans. on Acoust. Signal Speech Process., To be published.

[111] G. G. Walter. A sampling theorem for wavelet subspaces. *IEEE Trans. on Inf. Theory*, 38:881–884, 1992.

[112] M. V. Wickerhauser. Acoustic signal compression with wavelet packets. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 679–700. Academic Press, 1992.

[113] J.-C. Xu and W.-C. Shann. Galerkin-wavelet methods for two-point boundary value problems. Numer. Math., To be published.

Appendix: The periodic fast wavelet transform algorithm

We will give here a pseudo code implementation of the periodic fast wavelet transform. We assume that len_hp coefficients h_k are non-zero starting with the one with index $k = min\_hp$. Similar assumptions hold for the g_k, $\tilde{h}_k$, and $\tilde{g}_k$ with lengths len_gp, len_hd and len_gp and starting indices min_gp, min_hd and min_gp respectively. These coefficients are stored in 4 vectors such that

$$hp[k] \;=\; a\,h_{k+min\_hp}, \;\; gp[k] \;=\; a\,g_{k+min\_gp}, \;\; hd[k] \;=\; b\,\tilde{h}_{k+min\_hd}, \;\text{ and }\; gd[k] \;=\; b\,\tilde{g}_{k+min\_gd},$$

where $a\,b = 2$. We start with 2^n coefficients $\nu_{n,l}$ of a function of V_n and can thus apply n steps of the algorithm. These are initially stored in a vector $v[l]$. The computed wavelet coefficients are stored in a vector w such that

$$w \;=\; [\; \nu_{0,0} \quad \mu_{0,0} \quad \mu_{1,0} \quad \mu_{1,1} \quad \mu_{2,0} \quad \cdots \quad \mu_{2,3} \quad \cdots \quad \mu_{n-1,0} \quad \cdots \quad \mu_{n-1,2^{n-1}-1} \;]\,.$$

The algorithms are written in such a way to reduce operations in the inner loops. They are however not highly optimized not to affect readability too much. The index notation $a\,(b)\,c$ stands for a, $a+c$, $\ldots$, b and the operator floor(a) rounds a to the nearest integer towards minus infinity.

```
for  j ← n − 1 (−1) 0
    w[0 (1) 2^{j+1} − 1] ← 0
    for  l ← 0 (1) 2^j − 1
        i ← (2 * l + min_hd) mod 2^{j+1}
        for  k ← 0 (1) len_hd
            w[l]  ←  w[l] + hd[k] * v[i]
               i  ←  (i + 1) mod 2^{j+1}
        end for
        i ← (2 * l + min_gd) mod 2^{j+1}
        ls ← l + 2^j
        for  k ← 0 (1) len_gd
            w[ls]  ←  w[ls] + gd[k] * v[i]
                i  ←  (i + 1) mod 2^{j+1}
        end for
    end for
    v ← w[0 (1) 2^j − 1]
end for
```

```
for  j ← 1 (1) n
    v[0 (1) 2^j − 1] ← 0
    for  k ← 0 (1) 2^j − 1
        i ← (floor((k − min_hp)/2)) mod 2^(j−1)
        lb ← (k − min_hp) mod 2
        for  l ← lb (2) len_hp
            v[k]  ←  v[k] + hp[l] ∗ w[i]
               i  ←  (i − 1) mod 2^(j−1)
        end for
        i ← (floor((k − min_gp)/2)) mod 2^(j−1)
        lb ← (k − min_gp) mod 2
        for  l ← lb (2) len_gp
            v[k]  ←  v[k] + gp[l] ∗ w[i + 2^(j+1)]
               i  ←  (i − 1) mod 2^(j−1)
        end for

    end for
    w[0 (1) 2^j − 1]  =  v
end for
```

Björn Jawerth
University of South Carolina
Department of Mathematics
Columbia, SC 29208
USA

Wim Sweldens
Katholieke Universteit Leuven
Department of Computer Science
Celestijnenlaan 200A
B 3001 Leuven
Belgium
and
University of South Carolina
Department of Mathematics