AN OVERVIEW OF WAVELET BASED MULTIRESOLUTION ANALYSES*

BJÖRN JAWERTH[†‡] AND WIM SWELDENS[†§]

Abstract. In this paper we present an overview of wavelet based multiresolution analyses. First, we briefly discuss the continuous wavelet transform in its simplest form. Then, we give the definition of a 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 an interval, multidimensional wavelets and wavelet packets are discussed. Several examples of wavelet families are introduced and compared. Finally, the essentials of two major applications are outlined: data compression and compression of linear operators.

Key words. wavelet, multiresolution analysis, compression

AMS subject classifications. 42-02, 42C10

Contents.

1. Introduction. Wavelets have generated a tremendous interest in both theoretical and applied areas, especially over the past few years. The number of researchers, already large, continues to grow, so progress is being made at a rapid pace. In fact, advancements in the area are occurring at such a rate that the very meaning of "wavelet analysis" keeps changing to incorporate new ideas.

In a rapidly developing field, overview papers are particularly useful, and several good ones concerning wavelets are already available, such as [60, 83, 115, 122, 123, 125]. Of these, [122] contains a brief introduction to multiresolution analysis, [60] describes wavelets from an approximation theory point of view, [83] discusses continuous and discrete wavelets, [125] focuses on the construction of wavelets, [115] looks at wavelets from a signal processing point of view and [123] compares wavelets with Fourier techniques.

* Manuscript recieved by the editors ??, accepted for publication ??

† Department of Mathematics, University of South Carolina, Columbia SC 29208

‡ This author is partially supported by DARPA Grant AFOSR F49620-93-1-0083 and ONR Grant N00014-90-J-1343.

§ Departement Computerwetenschappen, Katholieke Universiteit Leuven, Celestijnenlaan 200 A, B 3001 Leuven, Belgium. This author is Research Assistant of the National Fund of Scientific Research Belgium, and is partially supported by ONR Grant N00014-90-J-1343 and the Experimental Program to Stimulate Competitive Research (EPSCOR NSF Grant EHR-9108772).

Our paper differs from these in that it contains some more recent developments and that it focuses on the "multiresolution analysis" aspect of wavelets. The emphasis on multiresolution analysis allows us to look at a number of different constructions of wavelets such as orthogonal, semiorthogonal, and biorthogonal wavelets. By examining these constructions in a unified setting, we are ideally positioned to make comparisons between them. The recent developments contain wavelets on an interval, multidimensional wavelets, and wavelet packets.

We have selected an expository style and a level of rigor that we hope will present the ideas without obscuring them in too much detail. Instead of giving exact, detailed statements in a "theorem-structured" way, we have opted for a more informal style. References are given throughout, pointing to more details when needed.

For example, this paper occasionally contains statements of the form "A is (essentially) equivalent with B". The interpretation that we have in mind is that, for "all practical purposes", A is equivalent to B. Strictly speaking, the equivalence may only hold under some extra technical conditions. Good examples are when a formula is guaranteed to be true only "almost everywhere" or in a "weak sense".

The style we have chosen is motivated by the intended audience: people with a more theoretical interest as well as those working in various applied areas. For a reader in the first category this paper might provide some of the theory, and point to some of the right references for further study. A reader in the second category could use the paper to make comparisons, and find connections to related material.

The paper is organized as follows. After a brief sketch of the history of wavelets we introduce the "continuous wavelet transform." The discussion of the continuous wavelet transform is mainly included for historical purposes and for comparison with the multiresolution analysis wavelets. Next, we give the definitions of "multiresolution analysis" and "scaling function" (Section 5), derive some basic properties and illustrate these with some examples. In this section we also give the basic definition of "wavelet." Wavelets are then studied in more detail in the next sections. Section 6 discusses orthogonal wavelets, while Section 7 treats biorthogonal wavelets, a generalization of the orthogonal ones, and semiorthogonal wavelets, a compromise between the previous two. In the following section we study the connection between wavelets and polynomials, and show how this relates to the approximation properties of wavelet expansions. In Section 9 we show how a "fast wavelet transform" can be derived from the multiresolution analysis properties. In the appendix, the reader can find a pseudocode implementation of this algorithm. At this point, i.e. after the study of the basic properties of multiresolution analysis, we are ready to single out some desirable properties of wavelets. This is done in Section 10. We also give several examples of wavelet families, such as Daubechies' and spline wavelets, and compare their properties. The next three sections focus on more recent developments such as wavelets on an interval, wavelet packets and multidimensional wavelets. These sections can be read independently. Finally, in the last section (Section 14) we consider the basic ideas associated with two important applications: data compression and analysis of linear operators.

It goes without saying (almost) that this short overview is still highly incomplete. It is unfortunate that we were unable to cover many other important and interesting developments in the area, some of which are more significant than the ones we have included. For example, we hardly mention the significant volume of work done 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. We apologize to the people whose

results we were unable to discuss due to the constraints imposed by the overview format.

Finally, let us point out that, although wavelets are a relatively recent phenomenon, there are a number of useful sources of information about them. First of all, there are three new journals with an emphasis on wavelets: Applied Computational Harmonic Analysis, Journal of Fourier Analysis, and Advances in Computational Mathematics. Secondly, several journals have had special issues on wavelets, such as Constructive Approximation, IEEE Transactions on Signal Processing, IEEE Transactions on Information Theory, International Journal of Optical Computing, Journal of Mathematical Imaging and Vision, and Optical Engineering. Also, an electronic information service exists on the Internet, the Wavelet Digest, with the address wavelet@math.scarolina.edu. Last but not least, several books on the subject exist, monographs as well as edited volumes. The list includes [13, 20, 21, 43, 49, 74, 92, 96, 106, 108, 116, 119].

2. Notation. Most of the notation will be presented as we go along. The space of square integrable functions, $\mathrm{L}^2(\mathbf{R})$, is defined as the space of Lebesgue measurable functions for which

$$\|f\|^2 = \int_{-\infty}^{+\infty} |f(x)|^2 \, dx < \infty.$$

The inner product of two functions $f, g \in \mathrm{L}^2(\mathbf{R})$ is given by

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

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

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

The Poisson summation formula is used in the following two forms,

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

and

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

If no bounds are indicated under a summation sign, $\in \mathbf{Z}$ 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

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

3. A short history of wavelets. The history of wavelets could be the topic of a separate paper. Let us give a short, subjective account.

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 filtering 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 so-called *Littlewood-Paley techniques* often are effective substitutes. These techniques were initially developed in the 30's to understand, among other things, summability properties of Fourier series and boundary behavior of analytic functions. In the 50's and 60's, these developed into powerful tools for studying other things, such as solutions of partial differential equations and integral equations. It was realized that they fit into *Calderón–Zygmund theory*, an area of harmonic analysis that 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 [35], 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, one has that

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

The $*$ denotes convolution. Here $\psi_t(x) = t^{-1}\psi(x/t)$, and $\widetilde{\psi}_t(x)$ is defined similarly, for appropriate fixed functions ψ and $\widetilde{\psi}$. As we shall see below, this representation is an example of a continuous wavelet transform. In mathematical physics the Aslaksen-Klauder construction of the $(ax+b)$-coherent states can be seen as another independent derivation of the Calderón formula [7, 91].

In the early 80's, Strömberg discovered the first orthogonal wavelets [126]. This was done in the context of trying to further understand Hardy spaces, as well as other spaces used to measure the size and smoothness of functions. A discrete version of the Calderón formula had also been used for similar purposes in [86] and long before this there were results by Haar [81], Franklin [70], Ciesielski [26], Peetre [112], and others.

Independent from these developments in harmonic analysis, Alex Grossmann, Jean Morlet, and their coworkers studied the wavelet transform in its continuous form [78, 79, 80]. The theory of "frames" [51] provided a suitable general framework for these investigations.

In the early to mid 80's, several groups, perhaps most notably the one associated with Yves Meyer and his collaborators, independently realized, with some excitement, that tools from Calderón-Zygmund theory, in particular the Littlewood-Paley representations, had discrete analogs and could give a unified view of many of the results in harmonic analysis. Also, one started to understand that these techniques 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 [71, 72, 73].) As the emphasis shifted more towards the representations themselves, and the building blocks involved, the name of the theory also shifted. Alex Grossmann 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 [97], independent 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 [103, 104, 105]. It also provided the connection with quadrature mirror filtering. Soon, Ingrid Daubechies [47] 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 [20, 77, 78, 83]. As mentioned above, a wavelet expansion uses translations and dilations of one fixed function, the wavelet $\psi \in \mathrm{L}^2(\mathbf{R})$. In the case of the continuous wavelet transform, the translation and dilation parameters vary continuously. In other words, the transform makes use of the functions

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

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

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

Using Parseval's identity, we can also write this as

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

where

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

We assume now that the wavelet ψ and its Fourier transform $\widehat{\psi}$ are functions with finite centers $\bar{x}$ and $\bar{\omega}$ and finite radii Δ_x and Δ_ω. These quantities are defined by

$$\bar{x} = \frac{1}{\|\psi\|^2} \int_{-\infty}^{+\infty} x\, |\psi(x)|^2\, dx,$$
$$\Delta_x^2 = \frac{1}{\|\psi\|^2} \int_{-\infty}^{+\infty} (x \Leftrightarrow \bar{x})^2\, |\psi(x)|^2\, dx,$$

and similarly for $\bar{\omega}$ and Δ_ω. The variable x usually represents either time or space; we shall settle for the first and refer to x as time. From (1) and (2), we see that the continuous wavelet transform at (a,b) picks up information about f, mostly from the time interval $[b + a\bar{x} \Leftrightarrow a\Delta_x, b + a\bar{x} + a\Delta_x]$ and from the frequency interval $[(\bar{\omega} \Leftrightarrow \Delta_\omega)/a, (\bar{\omega} + \Delta_\omega)/a]$. 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 $4\Delta_x\Delta_\omega$. The Heisenberg uncertainty principle says that this area has to be greater than 2. These time-frequency windows are also called *Heisenberg boxes*.

Suppose that the wavelet ψ satisfies the *admissibility condition*

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

Then, the continuous wavelet transform $\mathcal{W}(a,b)$ is invertible on its range, and an inverse transform is 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 $\widehat{\psi}(0)$ has to be 0, and, in particular, ψ has to oscillate. This, together with the decay property, has given ψ the name *wavelet* or "small wave" (French: *ondelette*). This shows that the frequency localization of the wavelets is much better than pointed out above. In most cases $\bar{\omega}$ is zero and the frequency localization is really in a band $[\Leftrightarrow\omega_2/a, \Leftrightarrow\omega_1/a] \cup [\omega_1/a, \omega_2/a]$, because $\widehat{\psi}(0)$ vanishes. This can help to understand why a reconstruction formula of type (3) is possible.

The transform is often represented graphically and plotted as two two-dimensional images with color or grey-scale value corresponding to the modulus and phase of $\mathcal{W}(a,b)$. This representation has been used extensively in areas such as geophysics.

In applications, it is of interest to find inverse transforms that do not make use of $\mathcal{W}$ over the whole range of a and b. Transforms exist that only use positive values of a or even only discrete values for a. Furthermore, using the theory of frames it is possible to study the case where only discrete values for a *and* b are used, see [83] for an excellent overview. The most common choice is to use a dyadic grid, i.e. to let $a = 2^{-j}$ and $b/a = l$ with $j, l \in \mathbf{Z}$ [48, 73]. In general, the fewer values of a and b one wants to use, the more restrictive the condition on the wavelet becomes. The continuous wavelet transform allows us to use a very general wavelet. At the other extreme, we shall see that much more restrictive conditions hold for a wavelet used in multiresolution analysis. This allows us, on the other hand, to prove powerful results such as the construction of orthogonal bases.

The transform that only uses the dyadic values of a and b was originally called the *discrete wavelet transform*. At this moment, however, this term is ambiguous, since it is also used to denote the transform from the sequence of scaling function coefficients of a function to its wavelet coefficients (see Section 9).

The case when a, and b belong to more irregular sets have also been covered. Such irregular sampling results can be found in [14, 39, 68, 67].

The continuous wavelet transform is also used in *singularity detection* and characterization [71, 100]. A typical result in this direction is that if a function f is Hölder (Lipschitz) continuous of order $0 < \alpha < 1$, so that $|f(x+h) \Leftrightarrow f(x)| = \mathcal{O}(h^\alpha)$, then the continuous wavelet transform has an asymptotic behavior like

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

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 a corresponding characterization of Hölder (Lipschitz) continuous functions of higher order $\alpha \geqslant 1$; the wavelet must then have a number of vanishing moments greater than α, i.e.

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

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

Example. A classical example of a wavelet is the *Mexican hat* function,

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

Being the second derivative of a Gaussian, it has 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, and another is through multiresolution analysis. Here we start by defining multiresolution analysis, and then point out some of the connections with the continuous wavelet transform.

A *multiresolution analysis* of $\mathrm{L}^2(\mathbf{R})$ is defined as a sequence of closed subspaces V_j of $\mathrm{L}^2(\mathbf{R})$, $j \in \mathbf{Z}$, with the following properties [47, 103]:

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 $\mathrm{L}^2(\mathbf{R})$ and $\bigcap_{j=-\infty}^{+\infty} V_j = \{\mathbf{0}\}$,
5. A *scaling function* $\varphi \in V_0$, with a non-vanishing integral, exists such that the collection $\{\varphi(x \Leftrightarrow l) \mid l \in \mathbf{Z}\}$ is a Riesz basis of V_0.

The references [122, 123] contain an introduction to the concept of multiresolution analysis.

Let us make a couple of simple observations concerning this definition. Since $\varphi \in V_0 \subset V_1$, a sequence $(h_k) \in \ell^2(\mathbf{Z})$ exists such that the scaling function satisfies

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

This functional equation goes by several different names: the *refinement equation*, the *dilation equation* or the *two-scale difference equation*. We shall use the first.

It is immediate that the collection of functions $\{\varphi_{j,l} \mid l \in \mathbf{Z}\}$, with $\varphi_{j,l}(x) = \sqrt{2^j}\,\varphi(2^j x \Leftrightarrow l)$, is a Riesz basis of V_j.

By integrating both sides of (4), and dividing by the (non-vanishing) integral of φ, we see that

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

If the scaling function belongs to L^1, it is, under very general conditions, uniquely defined by its refinement equation and the normalization [52],

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

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 ($x = 2^{-j}k$, $j, k \in \mathbf{Z}$) [15, 18, 47, 52, 53, 122]. In many applications, we never need the scaling function itself; instead we may often work directly with the h_k.

The spaces V_j will be used to approximate general functions. This will be done by defining appropriate projections onto these spaces. Since the union of all the V_j is dense in $L^2(\mathbf{R})$, we are guaranteed that any given function can be approximated arbitrarily close by such projections.

To be able to use the collection $\{\varphi(x \Leftrightarrow l) \mid l \in \mathbf{Z}\}$ 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,

$$\forall\, x \in \mathbf{R} \;:\; \sum_k \varphi(x \Leftrightarrow k) = 1. \tag{6}$$

Note that by Poisson's summation formula, the partition of unity is (essentially) equivalent with

$$\widehat{\varphi}(2\pi k) = \delta_k \quad \text{for} \quad k \in \mathbf{Z}. \tag{7}$$

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

$$\widehat{\varphi}(\omega) = H(\omega/2)\, \widehat{\varphi}(\omega/2), \tag{8}$$

where H is a 2π-periodic function defined by

$$H(\omega) = \sum_k h_k\, e^{-ik\omega}. \tag{9}$$

Since $\widehat{\varphi}(0) = 1$, we can apply (8) recursively. This yields, at least formally, the product formula

$$\widehat{\varphi}(\omega) = \prod_{j=1}^{\infty} H(2^{-j}\omega).$$

The convergence of this product is examined in [27, 47]. The representation of $\widehat{\varphi}$ is nice to have in many situations. For example, it can be used to construct $\varphi(x)$ from the h_k. Using (7) and (8), we see that we obtain a partition of unity if

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

Also note that (5) can be written as

$$H(0) = 1.$$

Examples of scaling functions.

(i) 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

$$N_m(x) = 2^{m-1} \sum_{k=0}^{m} \binom{m}{k} N_m(2x \Leftrightarrow k),$$

and

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

(ii) Another classical example is the Shannon sampling function,

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

We may take

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

and, consequently,

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

Now, for later reference, let us introduce the following 2π-periodic function:

$$F(\omega) = \sum_k |\widehat{\varphi}(\omega + k2\pi)|^2.$$

The fact that φ and its translates form a Riesz basis, corresponds to the fact that there are positive constants A and B such that

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

Using (8) and rearranging the even and odd terms, we have

$$\begin{aligned} F(2\omega) &= \sum_k |\widehat{\varphi}(2\omega + k2\pi)|^2 \\ &= \sum_k |H(\omega + k\pi)|^2 \, |\widehat{\varphi}(\omega + k\pi)|^2 \\ &= \sum_k |H(\omega + k2\pi)|^2 \, |\widehat{\varphi}(\omega + k2\pi)|^2 + |H(\omega + \pi + k2\pi)|^2 \, |\widehat{\varphi}(\omega + \pi + k2\pi)|^2 \\ &= |H(\omega)|^2 \, F(\omega) + |H(\omega + \pi)|^2 \, F(\omega + \pi). \end{aligned} \tag{10}$$

This shows that F is actually π-periodic.

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. In other words, each element of V_{j+1} can be written, in a unique way, as the sum of an element of W_j and an element of V_j. We note that the spaces W_j themselves are not necessarily unique; there may be several ways to complement V_j in V_{j+1}.

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,

$$\bigoplus_j W_j = \mathrm{L}^2(\mathbf{R}).$$

A function ψ is a *wavelet* if the collection of functions $\{\psi(x - l) \mid l \in \mathbf{Z}\}$ is a Riesz basis of W_0. The collection of wavelet functions $\{\psi_{j,l} \mid l, j \in \mathbf{Z}\}$ is then a Riesz basis

of $L^2(\mathbf{R})$. The definition of $\psi_{j,l}$ is similar to the one of $\varphi_{j,l}$ in the previous section. Note that a union of Riesz bases does not necessarily give a Riesz basis for the total span. Even though we did not impose any orthogonality, spaces W_j and $W_{j'}$ are "almost" diagonal for $|j \Leftrightarrow j'|$ large, and this allows the collection of all $\psi_{j,l}$ to form a Riesz basis for L^2. Since the wavelet ψ is an element of V_1, a sequence $(g_k) \in \ell^2(\mathbf{Z})$ exists such that

$$\psi(x) = 2 \sum_k g_k \, \varphi(2x \Leftrightarrow k). \tag{11}$$

The Fourier transform of the wavelet is given by

$$\widehat{\psi}(\omega) = G(\omega/2) \, \widehat{\psi}(\omega/2), \tag{12}$$

where G is a 2π-periodic function given by

$$G(\omega) = \sum_k g_k \, e^{-ik\omega}. \tag{13}$$

Each space V_j and W_j has a complement in $L^2(\mathbf{R})$ 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 $\mathcal{P}_j$ as the projection operator onto V_j and parallel to V_j^c, and $\mathcal{Q}_j$ as the projection operator onto W_j and parallel to W_j^c. A function f can now be written as

$$f(x) = \sum_j \mathcal{Q}_j f(x) = \sum_{j,l} \gamma_{j,l} \, \psi_{j,l}(x).$$

Recalling the discussion in Section 4, we see that this last equation is in fact an inverse "discrete" wavelet transform. At this moment the exact conditions on the wavelet are still unclear. They will made more precise in the next sections. There it will also become clear how to find the coefficients $\gamma_{j,l}$. We first turn to the case where the $\psi_{j,l}$ form an orthonormal basis for $L^2(\mathbf{R})$.

6. Orthogonal wavelets. The class of *orthogonal wavelets* is particularly interesting. We start by introducing the concept of 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 $j \in \mathbf{Z}$ are all mutually orthogonal, the projections $\mathcal{P}_j$ and $\mathcal{Q}_j$ are orthogonal, and the expansion

$$f(x) = \sum_j \mathcal{Q}_j f(x)$$

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

$$W_0 \perp V_0,$$

or

$$\langle \psi, \varphi(\cdot \Leftrightarrow l) \rangle = 0 \qquad l \in \mathbf{Z},$$

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

$$\forall\, \omega \in \mathbf{R} \;:\; \sum_k \widehat{\psi}(\omega + k2\pi)\, \overline{\widehat{\varphi}(\omega + k2\pi)} = 0. \tag{14}$$

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

$$\langle\, \varphi, \varphi(\cdot \Leftrightarrow l)\,\rangle = \delta_l \qquad l \in \mathbf{Z}. \tag{15}$$

With such a φ, the collection of functions $\{\varphi(x \Leftrightarrow l) \mid l \in \mathbf{Z}\}$ is an orthonormal basis of V_0 and the collection of functions $\{\varphi_{j,l} \mid l \in \mathbf{Z}\}$ is an orthonormal basis of V_j. Using Poisson's formula, (15) is (essentially) equivalent to

$$\forall\, \omega \in \mathbf{R} \;:\; \sum_k |\widehat{\varphi}(\omega + k2\pi)|^2 = F(\omega) = 1. \tag{16}$$

From (10) we now see that,

$$\forall\, \omega \in \mathbf{R} \;:\; |H(\omega)|^2 + |H(\omega + \pi)|^2 = 1, \tag{17}$$

or

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

The last two equations are equivalent, but they only provide a necessary condition for the orthogonality of the scaling function and its translates. This relationship is investigated in detail in [28, 94].

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

$$\langle\, \psi, \psi(\cdot \Leftrightarrow l)\,\rangle = \delta_l.$$

Again these conditions are (essentially) equivalent to

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

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

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

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

The projection operators $\mathcal{P}_j$ and $\mathcal{Q}_j$ can now be written as

$$\mathcal{P}_j f(x) = \sum_l \langle\, f, \varphi_{j,l}\,\rangle\, \varphi_{j,l}(x) \quad \text{and} \quad \mathcal{Q}_j f(x) = \sum_l \langle\, f, \psi_{j,l}\,\rangle\, \psi_{j,l}(x).$$

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

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

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

Using (16) we can write the condition (14) as

$$\forall\, \omega \in \mathbf{R} \;:\; G(\omega)\,\overline{H(\omega)} + G(\omega+\pi)\,\overline{H(\omega+\pi)} = 0. \tag{18}$$

From this last equation it follows that the function $G(\omega)$ needs to be of the form

$$G(\omega) = A(\omega)\,\overline{H(\omega+\pi)},$$

where A is a 2π-periodic function such that

$$A(\omega+\pi) = \Leftrightarrow A(\omega).$$

The orthogonality of the wavelet immediately follows from the orthogonality of the scaling function if

$$|A(\omega)| = 1.$$

As we will see later on, it is important for the scaling function and wavelet to have compact support. The compact support of the wavelet and scaling function is equivalent with the fact that H and G are trigonometric polynomials (i.e. the sums in (9) and (13) are finite). In the above case, we see that if the scaling function is compactly supported, so is the wavelet, provided that A is a trigonometric polynomial. The only trigonometric polynomials that satisfy the conditions for A are monomials of the form,

$$C\, e^{-(2k+1)\omega} \quad \text{with} \quad |C| = 1 \quad \text{and} \quad k \in \mathbf{Z}.$$

Up to the constant C and an integer translation, the different A all give rise to the same wavelet. Any other choice for A will lead to a wavelet without compact support. If the coefficients h_k are real, so are the g_k if $C = \pm 1$. The standard choice is $A(\omega) = \Leftrightarrow e^{-i\omega}$. This means that we derive an orthogonal wavelet from an orthogonal scaling function by choosing

$$g_k = (\Leftrightarrow 1)^k\, \overline{h_{1-k}}. \tag{19}$$

This still leaves us with the problem of constructing a compactly supported scaling function. We will comment on this in Section 8.

In [95] an orthogonalization 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

$$\widehat{\varphi}_{orth}(\omega) = \frac{\widehat{\varphi}(\omega)}{\sqrt{F(\omega)}}. \tag{20}$$

The fact that we started from a Riesz basis guarantees that $F(\omega)$ is strictly positive. We see that φ indeed satisfies the orthogonality condition (16). Note that if φ is compactly supported, φ_{orth} will, in general, not be compactly supported.

7. Biorthogonal wavelets. The orthogonality property puts a strong limitation on the construction of wavelets. For example, it is known that the Haar wavelet is the only real-valued wavelet that is compactly supported, symmetric and orthogonal [47]. The generalization to *biorthogonal wavelets* has been considered to gain more flexibility. Here, a dual scaling function $\widetilde{\varphi}$ and a dual wavelet $\widetilde{\psi}$ exist that generate a dual multiresolution analysis with subspaces $\widetilde{V}_j$ and $\widetilde{W}_j$, such that

$$\widetilde{V}_j \perp W_j \quad \text{and} \quad V_j \perp \widetilde{W}_j, \tag{21}$$

and, consequently,

$$\widetilde{W}_j \perp W_{j'} \quad \text{for} \quad j \neq j'.$$

The dual multiresolution analysis is not necessarily the same as the one generated by the original basis functions. An equivalent condition to (21) is

$$\langle\, \widetilde{\varphi}, \psi(\cdot \Leftrightarrow l) \,\rangle \;=\; \langle\, \widetilde{\psi}, \varphi(\cdot \Leftrightarrow l) \,\rangle \;= 0.$$

Moreover, the dual functions also have to satisfy

$$\langle\, \widetilde{\varphi}, \varphi(\cdot \Leftrightarrow l) \,\rangle \;= \delta_l \quad \text{and} \quad \langle\, \widetilde{\psi}, \psi(\cdot \Leftrightarrow l) \,\rangle \;= \delta_l.$$

By using a scaling argument, we have the seemingly more general properties that

$$\langle\, \widetilde{\varphi}_{j,l}, \varphi_{j,l'} \,\rangle \;= \delta_{l-l'} \qquad l, l', j \in \mathbf{Z} \tag{22}$$

and

$$\langle\, \widetilde{\psi}_{j,l}, \psi_{j',l'} \,\rangle \;= \delta_{j-j'}\delta_{l-l'} \qquad l, l', j, j' \in \mathbf{Z}. \tag{23}$$

Here the definitions of $\widetilde{\varphi}_{j,l}$ and $\widetilde{\psi}_{j,l}$ are similar to the ones for $\varphi_{j,l}$ and $\psi_{j,l}$. Note that the role of the basis (i.e. the φ and ψ) and the dual basis can be interchanged. Using the same Fourier techniques as in the previous section, the biorthogonality conditions are (essentially) equivalent with

$$\forall \omega \in \mathbf{R} \;:\; \begin{cases} \displaystyle\sum_k \widehat{\widetilde{\varphi}}(\omega + k2\pi)\,\overline{\widehat{\varphi}(\omega + k2\pi)} &=& 1 \\ \displaystyle\sum_k \widehat{\widetilde{\psi}}(\omega + k2\pi)\,\overline{\widehat{\psi}(\omega + k2\pi)} &=& 1 \\ \displaystyle\sum_k \widehat{\widetilde{\psi}}(\omega + k2\pi)\,\overline{\widehat{\varphi}(\omega + k2\pi)} &=& 0 \\ \displaystyle\sum_k \widehat{\widetilde{\varphi}}(\omega + k2\pi)\,\overline{\widehat{\psi}(\omega + k2\pi)} &=& 0. \end{cases} \tag{24}$$

Since they define a multiresolution analysis, the dual functions must satisfy

$$\widetilde{\varphi}(x) = 2\sum_k \widetilde{h}_k\, \widetilde{\varphi}(2x \Leftrightarrow k) \quad \text{and} \quad \widetilde{\psi}(x) = 2\sum_k \widetilde{g}_k\, \widetilde{\varphi}(2x \Leftrightarrow k). \tag{25}$$

If we define the functions $\widetilde{H}$ and $\widetilde{G}$ in the same fashion as we did for H and G, then necessary conditions are again given by

$$\forall \omega \in \mathbf{R} \;:\; \begin{cases} \widetilde{H}(\omega)\,\overline{H(\omega)} &+& \widetilde{H}(\omega+\pi)\,\overline{H(\omega+\pi)} &=& 1 \\ \widetilde{G}(\omega)\,\overline{G(\omega)} &+& \widetilde{G}(\omega+\pi)\,\overline{G(\omega+\pi)} &=& 1 \\ \widetilde{G}(\omega)\,\overline{H(\omega)} &+& \widetilde{G}(\omega+\pi)\,\overline{H(\omega+\pi)} &=& 0 \\ \widetilde{H}(\omega)\,\overline{G(\omega)} &+& \widetilde{H}(\omega+\pi)\,\overline{G(\omega+\pi)} &=& 0, \end{cases} \tag{26}$$

or

$$\forall \omega \in \mathbf{R} \ : \ \begin{bmatrix} \widetilde{H}(\omega) & \widetilde{H}(\omega+\pi) \\ \widetilde{G}(\omega) & \widetilde{G}(\omega+\pi) \end{bmatrix} \overline{\begin{bmatrix} H(\omega) & G(\omega) \\ H(\omega+\pi) & G(\omega+\pi) \end{bmatrix}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hence, if we let

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

and similarly for $\widetilde{M}$, then

$$\widetilde{M}(\omega)\,\overline{M^t(\omega)} = \mathbf{1}.$$

By interchanging the matrices on the left-hand side, we get

$$\forall \omega \in \mathbf{R} \ : \ \begin{cases} \overline{H(\omega)}\,\widetilde{H}(\omega) & + & \overline{G(\omega)}\,\widetilde{G}(\omega) & = & 1 \\ \overline{H(\omega)}\,\widetilde{H}(\omega+\pi) & + & \overline{G(\omega)}\,\widetilde{G}(\omega+\pi) & = & 0. \end{cases} \tag{27}$$

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

$$\widetilde{H}(\omega) = \frac{\overline{G(\omega+\pi)}}{\overline{\Delta(\omega)}} \tag{28}$$

and

$$\widetilde{G}(\omega) = \Leftrightarrow \frac{\overline{H(\omega+\pi)}}{\overline{\Delta(\omega)}}, \tag{29}$$

where

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

The fact that the wavelets form a basis for the complementary spaces ensures that Δ does not vanish.

The projection operators take the form

$$\mathcal{P}_j f(x) = \sum_l \langle\, f, \widetilde{\varphi}_{j,l} \,\rangle\, \varphi_{j,l}(x) \quad \text{and} \quad \mathcal{Q}_j f(x) = \sum_l \langle\, f, \widetilde{\psi}_{j,l} \,\rangle\, \psi_{j,l}(x),$$

and

$$f = \sum_{j,l} \langle\, f, \widetilde{\psi}_{j,l} \,\rangle\, \psi_{j,l}.$$

Note that this can be viewed as a "discrete" wavelet transform and that the conditions on ψ are less restrictive than in the orthogonal case. From the equations (22), (23), and (25) we see that

$$\widetilde{h}_{k-2l} = \langle\, \widetilde{\varphi}(x \Leftrightarrow l), \varphi(2x \Leftrightarrow k) \,\rangle \quad \text{and} \quad \widetilde{g}_{k-2l} = \langle\, \widetilde{\psi}(x \Leftrightarrow l), \varphi(2x \Leftrightarrow k) \,\rangle.$$

In particular, by writing $\varphi(2x \Leftrightarrow k) \in V_1$ in the bases of V_0 and W_0, we obtain that

$$\varphi(2x \Leftrightarrow k) = \sum_l \widetilde{h}_{k-2l}\, \varphi(x \Leftrightarrow l) + \sum_l \widetilde{g}_{k-2l}\, \psi(x \Leftrightarrow l). \tag{30}$$

Even if the scaling function and the wavelet are not orthogonal, the multiresolution analysis may still be orthogonal. Let us study this in a little more detail

A biorthogonal scaling function and wavelet are *semiorthogonal* if they generate an orthogonal multiresolution analysis [1, 2, 20]. The name *pre-wavelet* is also used for such a wavelet. Since the W_j subspaces are mutually orthogonal we have that

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

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

$$\widehat{\widetilde{\varphi}}(\omega) = \frac{\widehat{\varphi}(\omega)}{F(\omega)}.$$

We see that the first equation of (24) is satisfied, and, since F is a bounded, 2π-periodic function that does not vanish, the translates of φ and $\widetilde{\varphi}$ generate the same space. This corresponds to:

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

In order to have an orthogonal multiresolution analysis, (18) must also be satisfied. As before, this means that we need to pick G so that

$$G(\omega) = A(\omega)\, \overline{H(\omega + \pi)},$$

where A is a 2π-periodic function with

$$A(\omega + \pi) = \Leftrightarrow A(\omega).$$

If A is a trigonometric polynomial, then the scaling function is compactly supported. By looking at the last equation of (26) it is clear that a simple choice is

$$A(\omega) = \Leftrightarrow e^{-i\omega}\, F(\omega + \pi),$$

so that

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

and, consequently,

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

If φ is a compactly supported function, this construction guarantees that ψ is compactly supported too, since H and F, and hence also G, are trigonometric polynomials. However, the dual functions are, in general, not compactly supported.

8. Wavelets and polynomials. The moments of the scaling function and wavelet are defined by:

$$\mathcal{M}_p = \int_{-\infty}^{+\infty} x^p\, \varphi(x)\, dx \quad \text{and} \quad \mathcal{N}_p = \int_{-\infty}^{+\infty} x^p\, \psi(x)\, dx \quad \text{with} \quad p \in \mathbf{N},$$

and similarly for the dual functions. The scaling functions are normalized with $\mathcal{M}_0 = \widetilde{\mathcal{M}}_0 = 1$.

Recall that we want the scaling function to satisfy a "partition of unity" property and, furthermore, that this corresponds to $H(\pi) = 0$. From (29) we see that this implies that $\widetilde{G}(0) = 0$ and, hence, that $\widetilde{\mathcal{N}}_0 = 0$. So the dual wavelet needs to have a vanishing integral. This is reminiscent of the case of the continuous wavelet transform where we needed the wavelet to have a vanishing integral.

As we pointed out before, the fact that the wavelet has a vanishing integral allows us to give a precise characterization of the functions with a certain smoothness (when the order of smoothness α is less than 1), in terms of the decay of the continuous wavelet transform. The analogous fact is true here: the wavelet coefficients are given by inner products with the dual wavelets and the fact that these have a vanishing integral allows us to characterize exactly which functions will be of a certain smoothness by looking at the decay of the coefficients.

As in the case of the continuous wavelet transform, to obtain similar characterizations of classes of functions of smoothness $\alpha > 1$, the dual wavelet needs to have more vanishing moments. This is in fact closely related to the property that the scaling function and its translates can be used to represent polynomials. We make this statement more precise.

Let N denote the number of vanishing moments of the dual wavelet,

$$\widetilde{\mathcal{N}}_p = 0 \quad \text{for} \quad 0 \leqslant p < N \quad \text{and} \quad \widetilde{\mathcal{N}}_N \neq 0.$$

This is the same as saying that $\widehat{\widetilde{\psi}}(\omega)$ has a root of multiplicity N at $\omega = 0$. Since $\widehat{\widetilde{\varphi}}(0) \neq 0$, it is also equivalent to the fact that $\widetilde{G}(\omega)$ has a root of multiplicity N at $\omega = 0$. Thus, the sequence $\{\widetilde{g}_k\}$ also has N vanishing discrete moments,

$$\sum_k \widetilde{g}_k \, k^p = 0, \quad \text{for} \quad 0 \leqslant p < N.$$

From (29), we see that this is equivalent to $H(\omega)$ having a root of multiplicity N at $\omega = \pi$, which, by using (8), implies that

$$i^p \widehat{\varphi}^{(p)}(2k\pi) = \delta_k \, \mathcal{M}_p \quad \text{for} \quad 0 \leqslant p < N. \tag{31}$$

By Poisson's summation formula, it follows that

$$\sum_l (x \Leftrightarrow l)^p \, \varphi(x \Leftrightarrow l) = \mathcal{M}_p \quad \text{for} \quad 0 \leqslant p < N. \tag{32}$$

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 $\varphi(x \Leftrightarrow l)$ with $l \in \mathbf{Z}$.

At this point we digress a little and make two small remarks.

1. The fact that $H(\omega)$ has a root of multiplicity N at $\omega = \pi$ means that we can factor $H(\omega)$ 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 and the fact that K is a trigonometric polynomial is used as a starting point for the construction of compactly supported wavelets [31, 47].

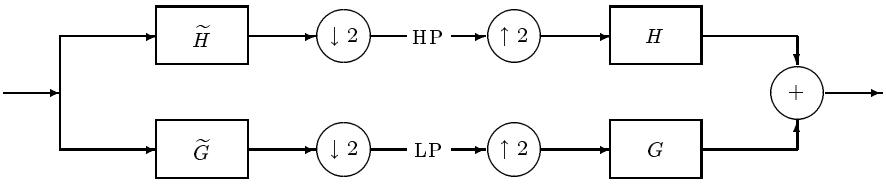

FIG. 1. *The subband filtering scheme.*

2. When writing a polynomial as a linear combination of the $\varphi(x \Leftrightarrow l)$, the coefficients in the linear combination themselves are polynomials of the same degree in l. More precisely, if A is a polynomial of degree $p \leqslant N \Leftrightarrow 1$, then a polynomial B, of the same degree, exists such that

$$A(x) = \sum_l B(l)\, \varphi(x \Leftrightarrow l). \tag{33}$$

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

$$B(l) = \int A(x)\, \widetilde{\varphi}(x \Leftrightarrow l)\, dx = \int A(x + l)\, \widetilde{\varphi}(x)\, dx.$$

Furthermore,

$$A(x) = \sum_l B(x \Leftrightarrow l)\, \varphi(l),$$

since the polynomials on the left and right-hand sides match at each integer.

With the extra vanishing moment conditions on the dual wavelet, we can characterize smoothness up to order $\alpha < N$. Another consequence is that the convergence rate of the wavelet approximation for smooth functions now immediately follows: if $f \in \mathcal{C}^N$, then

$$\|\mathcal{P}_j f(x) \Leftrightarrow f(x)\| = \mathcal{O}(h^N) \quad \text{with} \quad h = 2^{-j}. \tag{34}$$

The conditions (31) are referred to as the Strang–Fix conditions, and these were established long before the development of wavelet theory [69, 122, 124].

An asymptotic error expansion in powers of h, which can be used in numerical extrapolation, is derived in [127, 128]. For results on the pointwise convergence properties of wavelet series, see [90].

The exponent N in the factorization of H also plays a role in the regularity of φ. The Hölder regularity is $N \Leftrightarrow 1$ at most, but in many cases it is lower due to the influence of K. The regularity of solutions of refinement equations is studied in detail in [42, 41, 52, 53, 66, 114, 135, 136].

Note that we never required the dual scaling function to satisfy a partition of unity property, nor the wavelet to have a vanishing moment. In fact, it is possible to have a wavelet with a non-vanishing integral. In that case the regularity of the dual functions is very low. It may even be that they are distributions instead of functions, but this is not necessarily a problem in applications.

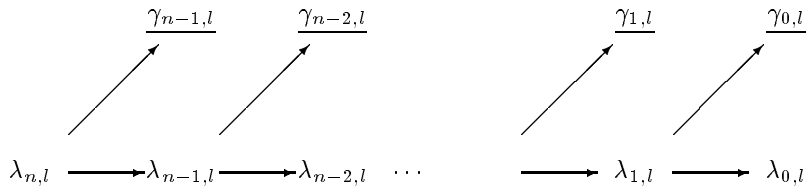

FIG. 2. *The decomposition scheme.*

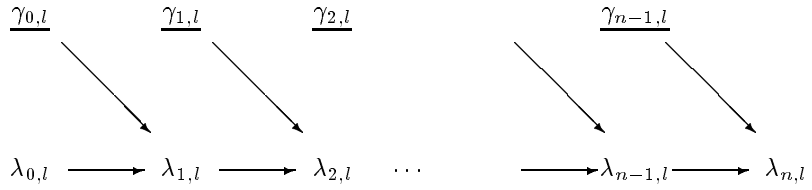

FIG. 3. *The reconstruction scheme.*

9. The fast wavelet transform. 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 $w_{j-1} \in W_{j-1}$:

$$\begin{aligned} v_j(x) &= \sum_k \lambda_{j,k}\, \varphi_{j,k}(x) = v_{j-1}(x) + w_{j-1}(x) \\ &= \sum_l \lambda_{j-1,l}\, \varphi_{j-1,l}(x) + \sum_l \gamma_{j-1,l}\, \psi_{j-1,l}(x). \end{aligned}$$

In other words, we have two representations of the function v_j, one as an element in V_j and associated with the sequence $\{\lambda_{j,k}\}$, and another as a sum of elements in V_{j-1} and W_{j-1} and associated with the sequences $\{\lambda_{j-1,k}\}$ and $\{\gamma_{j-1,k}\}$. The following relations show how to pass between these representations. By (25),

$$\begin{aligned} \lambda_{j-1,l} &= \langle\, v_j, \widetilde{\varphi}_{j-1,l} \,\rangle = \sqrt{2}\, \langle\, v_j, \sum_k \widetilde{h}_{k-2l}\, \widetilde{\varphi}_{j,k} \,\rangle \\ &= \sqrt{2} \sum_k \widetilde{h}_{k-2l}\, \lambda_{j,k}, \end{aligned} \tag{35}$$

and, similarly,

$$\gamma_{j-1,l} = \sqrt{2} \sum_k \widetilde{g}_{k-2l}\, \lambda_{j,k}. \tag{36}$$

The opposite direction, from the $\lambda_{j-1,l}$ and the $\gamma_{j-1,l}$ to the $\lambda_{j,k}$, is equally easy. Using (30) we have

$$\lambda_{j,k} = \sqrt{2} \sum_l h_{k-2l}\, \lambda_{j-1,l} + \sqrt{2} \sum_l g_{k-2l}\, \gamma_{j-1,l}. \tag{37}$$

When applied recursively, these formulae define the *fast wavelet transform*; the relations (35) and (36) define the forward transform, while (37) defines the inverse transform.

Now, from the fact that $H(0) = G(\pi) = 1$ and $G(0) = H(\pi) = 0$, we see that $H(\omega)$ acts like a low pass filter for the interval $[0, \pi/2]$ and $G(\omega)$ similarly behaves like a band pass filter for the interval $[\pi/2, \pi]$. Equation (8) (respectively (12)) then implies that the major part of the energy of the functions in V_0 (respectively W_0) is concentrated in the intervals $[0, \pi]$ (respectively $[\pi, 2\pi]$). The basic behavior of the dual functions is the same. In an approximate sense, this means that the wavelet expansion splits the frequency space into dyadic blocks $[2^j\pi, 2^{j+1}\pi]$ with $j \in \mathbf{Z}$ [103, 104].

In signal processing this idea is known as subband filtering, or, more specifically, as quadrature mirror filtering. Quadrature mirror filters were studied before wavelet theory. The decomposition step consists of applying a low-pass ($\widetilde{H}$) and a band-pass ($\widetilde{G}$) 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. putting a zero between every two samples) followed by filtering and addition. One can show that the conditions (27) correspond to the exact reconstruction of a subband filtering scheme. More details about this can be found in [115, 132, 133, 134].

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

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

Other sampling procedures, such as (quasi-)interpolation and quadrature formulae were proposed in [1, 2, 85, 120, 128, 138].

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 take a look at some important properties of wavelets.

Orthogonality. Orthogonality is convenient to have in many situations, e.g. it directly links the L^2 norm of a function to the norm of its wavelet coefficients by

$$\|f\| = \sqrt{\sum_{j,l} \gamma_{j,l}^2}.$$

In the biorthogonal case these two quantities are only equivalent. Another advantage of orthogonal wavelets is that the fast wavelet transform is a unitary transformation (i.e. its adjoint is its inverse). Consequently, its condition number is equal to 1, which is the optimal case. (Recall that the condition number of a linear transformation A is defined as $\|A\|.\|A^{-1}\|$). This is of importance in numerical calculations. It means that an error present in the initial data will not grow under the transformation, and that stable numerical computations are possible.

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 sense.

Compact support. If the scaling function and wavelet are compactly supported, the filters H and G are finite impulse response filters, so that the summations in the fast wavelet transform are finite. This obviously is of use in implementations. If they are not compactly supported, a fast decay is desirable so that the filters can be approximated reasonably 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. Multiplication by a power of two on a computer corresponds to shifting bits, which is a very fast operation.

Symmetry. If the scaling function and wavelet are symmetric, then the filters have generalized linear phase. The absence of this property can lead to phase distortion. This is important in signal processing applications.

Smoothness. The smoothness of wavelets plays an important role in compression applications. Compression is usually achieved by setting small coefficients $\gamma_{j,l}$ to zero, and thus leaving out a component $\gamma_{j,l}\,\psi_{j,l}(x)$ from the original function. If the original function represents an image and the wavelet is not smooth, the error can easily be detected visually. Note that the smoothness of the primary functions is more important to this aspect than that of the dual. Also, a higher degree of smoothness corresponds to better frequency localization of the filters. Finally, smooth basis functions are desired in numerical analysis applications where derivatives are involved.

Number of vanishing moments of the dual wavelet. We saw earlier that this is important in singularity detection and characterization of smoothness spaces. Also, it determines the convergence rate of wavelet approximations of smooth functions. Finally, the number of vanishing moments of the dual wavelet is connected to the smoothness of the wavelet (and vice versa).

Analytic expressions. As previously noted, an analytic expression for a scaling function or wavelet does not always exists but in some cases it is available and nice to have. In harmonic analysis, analytic expressions of the Fourier transform are particularly useful.

Interpolation. If the scaling function satisfies

$$\varphi(k) = \delta_k \quad \text{for} \quad k \in \mathbf{Z},$$

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

As could be expected, it is not possible to construct wavelets that have all these properties and there is a trade-off between them. We now take a look at several compromises.

Examples of orthogonal wavelets.

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

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

and the Shannon wavelet is

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

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

(ii) A more interesting example is the *Meyer wavelet* and scaling function [106]. 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.

(iii) The *Battle-Lemarié wavelets* are constructed by orthogonalizing B–spline functions using (20) and have exponential decay [12, 95]. The wavelet with N vanishing moments is a piecewise polynomial of degree $N \Leftrightarrow 1$ that belongs to $\mathcal{C}^{N-2}$.

(iv) Probably the most frequently used orthogonal wavelets are the original Daubechies wavelets [47, 49]. They are a family of orthogonal wavelets indexed by $N \in \mathbf{N}$, where N is the number of vanishing wavelet moments. They are supported on an interval of length $2N \Leftrightarrow 1$. 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 $0.2075N$ for large N. In [137] a different family with regularity asymptotically equal to $0.3N$ was presented. In [50] three variations of the original family, all with orthogonal and compactly supported functions, are constructed:

1. The previous construction does not lead to a unique solution if N and the support length are fixed. One family is constructed by choosing, for each N, the solution with closest to linear phase (or closest to symmetry). In fact, 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, the scaling function also 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 [17]. Their construction was asked by Ronald Coifman and Ingrid Daubechies therefore named them *coiflets*. They are supported on an interval with length $3N \Leftrightarrow 1$.

Examples of biorthogonal wavelets.

(i) Biorthogonal wavelets were constructed by Albert Cohen, Ingrid Daubechies and Jean-Christophe Feauveau in [31]. Here $\Delta(\omega)$ is chosen equal to $e^{-i\omega}$, and thus

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

In one of the families constructed in [31], 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.

(ii) Semiorthogonal spline wavelets were constructed by Charles Chui and Jianzhong Wang in [23, 24, 25]. The scaling functions are cardinal B–splines of order m and the wavelet functions are splines with support $[0, 2m \Leftrightarrow 1]$. All primary and dual functions still have generalized linear phase and all coefficients used in the fast

TABLE 1
A quick comparison of wavelet families.

wavelet	compact support		analytic expression		symmetry	orthogonality		compact
family	primary	dual	primary	dual		semi	full	support $\widehat{\psi}$
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

wavelet transform 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 have exponential decay instead. The same wavelets, but in a different setting, were also derived by Akram Aldroubi, Murray Eden and Michael Unser in [129, 131]. They also showed that for N going to infinity, the spline wavelets converge to Gabor functions [130].

(iii) Other semiorthogonal wavelets can be found in [89, 109, 110, 113]. A characterization of all semiorthogonal wavelets is given in [1, 2].

The properties of some of the orthogonal, biorthogonal and semiorthogonal wavelet families are summarized in Table 1.

Examples of interpolating scaling functions.

(i) The Shannon sampling function

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

is an interpolating scaling function. It is band limited, but it has very slow decay.

(ii) An interpolating scaling function, whose translates also generate V_0, can be found by letting

$$\widehat{\varphi}_{interpol}(\omega) = \frac{\widehat{\varphi}(\omega)}{\sum_l \varphi(l) e^{-i\omega l}},$$

provided that the denominator does not vanish [1, 2, 129, 138]. Even if φ is compactly supported, $\varphi_{interpol}$ is in general not compactly supported. The cardinal spline interpolants of even order are constructed this way [118].

(iii) An interpolating scaling function can also be constructed from a pair of biorthogonal scaling functions as

$$\varphi_{interpol}(x) = \int_{-\infty}^{+\infty} \varphi(y+x)\, \overline{\widetilde{\varphi}(y)}\, dy.$$

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, starting from the Daubechies orthogonal or biorthogonal wavelets, yields a family of interpolating functions which had been studied earlier by Gilles Deslauriers and Serge Dubuc in [56, 57]. These functions are smooth and compactly supported. More information can also be found in [61, 117]. A natural choice for the wavelet here is $\psi(x) = \varphi(2x \Leftrightarrow 1)$ and this is a typical example of a wavelet with a non-vanishing integral. The dual scaling function is a Dirac impulse and the dual wavelet is a linear combination of Dirac impulses (and has several vanishing moments). We still have a fast wavelet transform with finite impulse response filters.

(iv) Also wavelets can be interpolating. In [2] wavelets that are both symmetrical and interpolating were constructed.

11. Wavelets on an interval. 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 0s" 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 wavelet coefficients on a given scale satisfy $\langle f, \psi_{j,k} \rangle = \langle f, \psi_{j,k+2^j} \rangle$, $k \in \mathbf{Z}$, $j \geq 0$. 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,

$$\varphi^*_{j,l}(x) = \chi_{[0,1]}(x) \sum_m \varphi_{j,l}(x+m) \quad \text{for} \quad 0 \leqslant l < 2^j \quad \text{and} \quad j \geqslant 0. \tag{38}$$

If the support of $\varphi_{j,l}(x)$, is a subset of $[0,1]$, then $\varphi^*_{j,l}(x) = \varphi_{j,l}(x)$. Otherwise $\varphi_{j,l}(x)$ is chopped into pieces of length 1, which are shifted onto $[0,1]$ and added up, yielding $\varphi^*_{j,l}(x)$. Similar definitions hold for $\psi^*_{j,l}$, $\widetilde{\varphi}^*_{j,l}$ and $\widetilde{\psi}^*_{j,l}$. The algorithm in the appendix describes the periodic fast wavelet transform. This "wrap around" procedure is satisfactory in many situations (and certainly takes care of functions like $f(x) = 1$, $x \in [0,1]$). However, unless the behavior of the function f at 0 matches that at 1, the periodic version of f has 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, are wavelets intrinsically defined on $[0,1]$. We sketch a construction of orthogonal wavelets on $[0,1]$,

recently presented by Yves Meyer [107]. We start from an orthogonal Daubechies scaling function with $2N$ non-zero coefficients:

$$\varphi(x) = 2 \sum_{k=0}^{2N-1} h_k \, \varphi(2x \Leftrightarrow k). \tag{39}$$

It is easy to see that $\mathrm{clos}\{x : \varphi(x) \neq 0\} = [0, 2N \Leftrightarrow 1]$, and, as a consequence,

$$B_{j,k} = \mathrm{clos}\{x : \varphi_{j,k}(x) \neq 0\} = [2^{-j}k, 2^{-j}(k + 2N \Leftrightarrow 1)]. \tag{40}$$

This implies that for sufficiently small scales 2^{-j}, $j \geq j_0$, a function $\varphi_{j,k}$ 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:

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

Here E° denotes the interior of the set E. For sufficiently large j the sets $S_j^{(1)}$ and $S_j^{(3)}$ are disjoint and

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

It is easy to write down what these sets are more explicitly:

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

Note, in particular, that the sets $S_j^{(1)}$ and $S_j^{(3)}$ contain the indices of $2N \Leftrightarrow 2$ 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 form an increasing sequence of spaces,

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

and $V_j^{[0,1]}$, $j \geq j_0$, form a multiresolution analysis of $\mathrm{L}^2([0,1])$. It is also obvious that the functions in $\{\varphi(x \Leftrightarrow l)|_{[0,1]} : l \in S_j\}$ span $V_j^{[0,1]}$. Here $g(x)\ |_{[0,1]}$ denotes the restriction of $g(x)$ to $[0,1]$. Not quite as obvious 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 $\varphi_{j,k}$ lives entirely inside $[0,1]$, restricting it to $[0,1]$ has no effect. In particular, the functions $\varphi_{j,k}$, $k \in S_j^{(2)}$ are still pairwise orthogonal. A key observation now is that for $k \in S_j^{(1)}$, $l \in S_j^{(2)} \cup S_j^{(3)}$,

$$\langle \varphi_{j,k}, \varphi_{j,l} \rangle_{[0,1]} = \int_0^1 \varphi_{j,k}(x)\, \varphi_{j,l}(x)\, dx = \int_{-\infty}^{+\infty} \varphi_{j,k}(x)\, \varphi_{j,l}(x)\, dx = 0, \tag{41}$$

and similarly when $k \in S_j^{(3)}$, $l \in S_j^{(2)} \cup S_j^{(1)}$. We see that the three collections $\{\varphi(x \Leftrightarrow l)|_{[0,1]} : l \in S_j^{(1)}\}$, $\{\varphi(x \Leftrightarrow l)|_{[0,1]} : l \in S_j^{(2)}\}$, and $\{\varphi(x \Leftrightarrow l)|_{[0,1]} : l \in S_j^{(3)}\}$ are mutually orthogonal. So, since the functions in $\{\varphi(x \Leftrightarrow l)|_{[0,1]} : l \in S_j^{(2)}\}$ already form an orthonormal set, there only remains to separately orthogonalize the functions in $\{\varphi(x \Leftrightarrow l)|_{[0,1]} : l \in S_j^{(1)}\}$ and in $\{\varphi(x \Leftrightarrow l)|_{[0,1]} : l \in S_j^{(3)}\}$. 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

$$V_{j+1}^{[0,1]} = V_j^{[0,1]} + W_j^{[0,1]}. \tag{42}$$

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 $2^j + 2N \Leftrightarrow 2$ wavelets that intersect $[0,1]$, and, since $\dim V_{j+1}^{[0,1]} \Leftrightarrow \dim V_j^{[0,1]} = 2^j$ 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 (41), they are also orthogonal to $V_j^{[0,1]}$. There are $2N \Leftrightarrow 2$ wavelets whose support intersects an endpoint. However, we only need $N \Leftrightarrow 1$ basis functions at each endpoint. One can now use (30) to write out the dependencies, and construct $N \Leftrightarrow 1$ basis functions at each endpoint. After that we just apply a Gram-Schmidt procedure 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 $\varphi_{j,k}$ that intersect $[0,1]$ there are some that are almost zero there. Hence, the set $\{\varphi_{j,k}\}_{k \in S_j}$ is almost linearly dependent, and, as a consequence, the condition number of the matrix, corresponding to the change of basis from $\{\varphi_{j,k}\}_{k \in S_j}$ 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 (33) all polynomials of degree $\leqslant N \Leftrightarrow 1$ can be written as linear combinations of the $\varphi_{j,l}$ for $l \in \mathbf{Z}$. 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 [5, 32, 33] uses this as a starting point and is slightly different from the one by Yves Meyer. Let us briefly describe this construction as well. Again we start with an orthogonal Daubechies scaling function φ with $2N$ non-zero coefficients, and assume that we have picked the scale fine enough so that the endpoints are independent as before. By (33) and, since the $\{\varphi_{j,k}\}$ is an orthonormal basis for V_j, each monomial x^α, $\alpha \leqslant N \Leftrightarrow 1$, has the representation $x^\alpha = \sum_k \langle x^\alpha, \varphi_{j,k} \rangle\, \varphi_{j,k}(x)$. The restriction to $[0,1]$ can then

be written

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

If we let

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

and, similarly,

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

then

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

We let the spaces $\bar{V}_j$, $j \geq j_0$, that form a multiresolution analysis of $\mathrm{L}^2([0,1])$, be the linear span of the functions $\{x^\alpha_{j,L}\}_{\alpha \leqslant N-1}$, $\{x^\alpha_{j,R}\}_{\alpha \leqslant N-1}$, and $\{\varphi_{j,k}|_{[0,1]}\}_{k=1}^{2^j-2N}$:

$$\bar{V}_j = \overline{\{x^\alpha_{j,L}\}_{\alpha \leqslant N-1}} \cup \overline{\{\varphi_{j,k}\}_{k=1}^{2^j-2N}} \cup \overline{\{x^\alpha_{j,R}\}_{\alpha \leqslant N-1}}$$

Finding an orthonormal basis for $\bar{V}_j$ is easy; in fact, the collections $\{x^\alpha_{j,L}\}_{\alpha \leqslant N-1}$, $\{\varphi_{j,k}\}_{k=1}^{2^j-2N}$, and $\{x^\alpha_{j,R}\}_{\alpha \leqslant N-1}$ are mutually orthogonal, and all of the functions in these are linearly independent. We thus only have to orthogonalize 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 $\leq N \Leftrightarrow 1$ 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 $1 \leqslant k \leqslant 2^j \Leftrightarrow 2N$ 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 (30) again.

This last construction carries over to more general situations. For example, we can also use biorthogonal wavelets and much more general closed sets than $[0,1]$ [5, 33, 87].

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 [70] was given for $[0,1]$. Another interesting one, in the case of semiorthogonal spline wavelets, has been given by Charles Chui and Ewald Quak [19]; 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 [37, 38]. In this section it will be useful to switch to the following notation:

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

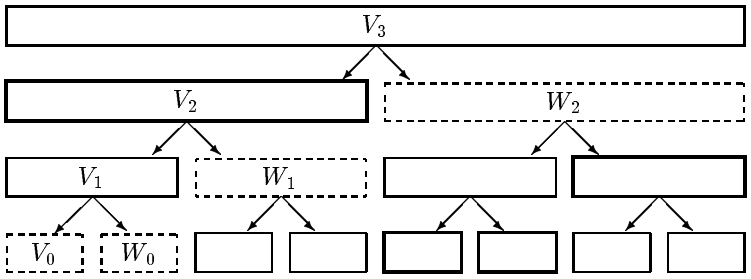

FIG. 4. *Wavelet packets scheme.*

The fundamental observation is the following fact, called the *splitting trick* [22, 30, 106]:
Suppose that the set of functions $\{f(x \Leftrightarrow k) \mid k \in \mathbf{Z}\}$ *is a Riesz basis for its closed linear span* S. *Then the functions*

$$f_k^0 = \frac{1}{\sqrt{2}} f^0(x/2 \Leftrightarrow k) \quad \textit{and} \quad f_k^1 = \frac{1}{\sqrt{2}} f^1(x/2 \Leftrightarrow k) \quad \textit{for} \quad k \in \mathbf{Z},$$

also constitute a Riesz basis for S, *where*

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

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

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

with

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

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

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

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 3 levels. The top rectangle represents the space V_3 and each other rectangle corresponds to a certain subspace of V_3 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 dashed rectangles then correspond to the wavelet

multiresolution analysis $V_3 = V_0 \oplus W_0 \oplus W_1 \oplus W_2$. The bold rectangles correspond to a possible wavelet packet splitting and a basis with functions

$$\left\{\psi_0^1(4x \Leftrightarrow k),\ \psi_{1,1}^2(2x \Leftrightarrow k),\ \psi_{0,0,1}^3(x \Leftrightarrow k),\ \psi_{1,0,1}^3(x \Leftrightarrow k) \mid k \in \mathbf{Z}\right\}.$$

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 two parts. 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 an 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 procedure is called *best basis selection*, and one can design fast algorithms that make use of the tree structure. The particular criterion is determined by the application, and which basis functions that will end up in the basis depends on the data.

For applications in image processing, entropy-based criteria were proposed in [40]. The best basis selection in that case has a numerical complexity of $\mathcal{O}(M)$. Applications in signal processing can be found in [36, 139].

This wavelet packets construction can also be combined with wavelets on an interval and wavelets in higher dimensions [55].

13. Multidimensional wavelets. Up till now we have focused on functions of one variable and 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

$$\Phi(x,y) = \varphi(x)\,\varphi(y) = \varphi \otimes \varphi(x,y),$$

and define

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

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

$$\Psi^{(1)} = \varphi \otimes \psi,\quad \Psi^{(2)} = \psi \otimes \varphi,\quad \text{and}\quad \Psi^{(3)} = \psi \otimes \psi. \tag{43}$$

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

$$f(x,y) = \sum_{i,l}\sum_{j,k} \langle\, f, \psi_{i,l} \otimes \psi_{j,k} \,\rangle\ \psi_{i,l} \otimes \psi_{j,k}(x,y). \tag{44}$$

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 (44) is therefore called the *rectangular wavelet decomposition* of f while the functions in (43) 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.

These simple constructions are insufficient in many cases. What we need sometimes are wavelets intrinsically constructed for higher dimensions. One of the interesting problems here is how to split a space into complementary subspaces. In the univariate case we split into two spaces, each with essentially the same "size." If we use the square tensor product basis in d dimensions, we split into 2^d subspaces, $2^d \Leftrightarrow 1$ of which are spanned by wavelets. There are several constructions of nonseparable wavelets that use this kind of splitting. One of the problems here is, given the scaling function, is there an easy way, cf. (19), to find the wavelets? This was studied in [54, 113, 121]. Another idea is to still try to split into just two subspaces. This involves the use of different lattices [99]. In the bivariate case, Ingrid Daubechies and Albert Cohen constructed smooth, compactly supported, biorthogonal wavelets, using ideas from the univariate construction [29].

By now, there is a lot of material about multivariate wavelets. However, we shall leave this topic for now and just mention some other possibilities such as hexagonal lattices, and Clifford valued wavelets [6, 9, 34].

14. Applications.

14.1. Data compression. One of the most common 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 consider here lossy compression. This means 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 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 reasons other 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

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

where $0 \leqslant p_{ij} \leqslant 255$. 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

FIG. 5. *Image transform coding.*

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 [101, 102]. One way is to consider compression to be an approximation problem [58, 59]. More specifically, let us fix an orthogonal wavelet ψ. Given an integer $M \geqslant 1$, we try to find the "best" approximation of f by using a representation

$$f_M(x) = \sum_{kl} b_{jk}\, \psi_{jk}(x) \quad \text{with } M \text{ non-zero coefficients } b_{jk}. \tag{45}$$

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 [58]. 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 $\|f \Leftrightarrow f_M\|_{L^2}$. Because of the orthogonality of the wavelets this equals

$$\left(\sum_{jk} |\, \langle\, f, \psi_{jk} \,\rangle \Leftrightarrow b_{jk}\,|^2 \right)^{1/2}. \tag{46}$$

A moment's 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

$$\|f \Leftrightarrow f_M^{opt}\|_{L^2} = \mathcal{O}(M^{-\beta}), \tag{47}$$

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

$$\left(\sum_{M \geq 1} (M^{\beta} \|f \Leftrightarrow f_M^{opt}\|_{L^2})^p \frac{1}{M} \right)^{1/p} \approx \left(\sum_{jk} |\, \langle\, f, \psi_{jk} \,\rangle\, |^p \right)^{1/p} \tag{48}$$

with $1/p = 1/2 + \beta$. The maximal β for which (47) is valid can be estimated by finding the smallest p for which the right-hand side of (48) 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 (48) is actually not exactly the same as in (47). However, for practical purposes, the difference is of no consequence.

14.2. Operator 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 [17]. 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*,

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

The basic idea now is 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,

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

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} (remember that 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 $H\psi_{jk}(x)$ resembles that of $\psi_{jk}(x)$. By expanding the kernel as a Taylor series,

$$\frac{1}{x \Leftrightarrow s} = \frac{1}{x}\left(1 + \frac{s}{x} + \frac{s^2}{x^2}\,\cdots\right),$$

we see that the more vanishing moments ψ has, the faster the decay of $H\psi_{j,k}$.

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, 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 significantly speed up the numerical treatment. This program of using the wavelet representations for the efficient numerical treatment of operators was initiated in [17]. We also refer to [4, 3] 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 [84]. There are interesting results showing that this might be fruitful; for example, it has been shown [17, 46, 111, 140] 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 defined on a finite set rather than on the entire Euclidean plane. This is similar to the problem we discussed with wavelets on an interval, and, indeed, the techniques discussed there can be often used to handle these two problems [5, 8].

Wavelets have also been used in the solution of evolution equations [11, 76, 93, 98]. 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 [11].

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 possible to find particular wavelets, adapted to the operator or problem at hand [10, 44, 45, 88]. In [16], 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 [62, 63, 64, 65].

Acknowledgement. We would like to thank Gilbert Strang and the referees for useful comments and suggestions, and Maurits Malfait and Nathan Hall for helping us with the presentation.

REFERENCES

[1] A. Aldroubi and M. Unser, *Families of wavelet transforms in connection with Shannon's sampling theory and the Gabor transform*, in [21], pp. 509–528.

[2] ———, *Families of multiresolution and wavelet spaces with optimal properties*, Numer. Funct. Anal. and Optimiz., (1993), pp. 417–446.

[3] B. Alpert, *A class of bases in* L^2 *for the sparse representation of integral operators*, SIAM J. Math. Anal., 24 (1993), pp. 246–262.

[4] B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin, *Wavelet-like bases for the fast solution of second-kind integral equations*, SIAM J. Sci. Comp., 14 (1993), pp. 159–184.

[5] L. Andersson, N. Hall, B. Jawerth, and G. Peters, *Wavelets on closed subsets of the real line*, in [119], pp. 1–61.

[6] L. Andersson, B. Jawerth, and M. Mitrea, *The Cauchy singular integral operator and Clifford wavelets*, in [13], pp. 525–546.

[7] E. W. Aslaksen and J. R. Klauder, *Unitary representations of the affine group*, J. Math. Phys., 9 (1968), pp. 2267–2275.

[8] P. Auscher, *Wavelets with boundary conditions on the interval*, in [21], pp. 217–236.

[9] ———, *Ondelettes fractales et Applications*, PhD thesis, Université de Paris-Dauphine, 1989.

[10] P. Auscher and P. Tchamitchian, *Conjecture de Kato sur les ouverts de* $\mathbf{R}$, Rev. Mat. Iberoamericana, 8 (1992), pp. 149–199.

[11] E. Bacry, S. Mallat, and G. Papanicolaou, *A wavelet based space-time adaptive numerical method for partial differential equations*, RAIRO Mathematical Modelling and Numerical Analysis, 26.

[12] G. Battle, *A block spin construction of ondelettes*, Comm. Math. Phys., 110 (1987), pp. 601–615.

[13] J. Benedetto and M. Frazier, eds., *Wavelets: Mathematics and Applications*, CRC Press, Boca Raton, 1993.

[14] J. J. Benedetto, *Irregular sampling and frames*, in [21], pp. 445–507.

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

[16] G. Beylkin, *On wavelet-based algorithms for solving differential equations*, in [13], pp. 449–466.

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

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

[19] C. Chui and E. Quak, *Wavelets on a bounded interval*, in Numerical Methods of Approximation Theory, D. Braess and L. L. Schumaker, eds., Birkhäuser Verlag, Basel, 1992, pp. 1–24.

[20] C. K. Chui, *An Introduction to Wavelets*, Academic Press, San Diego, 1992.

[21] ———, ed., *Wavelets: A Tutorial in Theory and Applications*, Academic Press, San Diego, 1992.

[22] C. K. Chui and C. Li, *Non-orthogonal wavelet packets*, SIAM J. Math. Anal., 24 (1993), pp. 712–738.

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

[24] ———, *A general framework of compactly supported splines and wavelets*, J. Approx. Theory, 71 (1992), pp. 263–304.

[25] ———, *On compactly supported spline wavelets and a duality principle*, Trans. Amer. Math. Soc., 330 (1992), pp. 903–915.
[26] Z. Ciesielski, *Constructive function theory and spline systems*, Studia Math., 52 (1973), pp. 277–302.
[27] A. Cohen, *Biorthogonal wavelets*, in [21], pp. 123–152.
[28] ———, *Ondelettes, analyses multiresolutions et filtres miroirs en quadrature*, Ann. Inst. H. Poincaré, Anal. Non Linéaire, 7 (1990), pp. 439–459.
[29] A. Cohen and I. Daubechies, *Non-separable bidimensional wavelet bases*. Preprint AT&T Bell Laboratories, New Jersey, 1991.
[30] ———, *On the instability of arbitrary biorthogonal wavelet packets*, SIAM J. Math. Anal., 24 (1993), pp. 1340–1354.
[31] A. Cohen, I. Daubechies, and J. Feauveau, *Bi-orthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 45 (1992), pp. 485–560.
[32] A. Cohen, I. Daubechies, B. Jawerth, and P. Vial, *Multiresolution analysis, wavelets and fast algorithms on an interval*, C. R. Acad. Sci. Paris Sér. I Math., I (1993), pp. 417–421.
[33] A. Cohen, I. Daubechies, and P. Vial, *Multiresolution analysis, wavelets and fast algorithms on an interval*, Appl. Comput. Harmon. Anal., 1 (1993), pp. 54–81.
[34] A. Cohen and J.-M. Schlenker, *Compactly supported bidimensional wavelet bases with hexagonal symmetry*, Constr. Approx., 9 (1993), pp. 209–236.
[35] R. R. Coifman, *A real variable characterization of H^p*, Studia Math, 51 (1974).
[36] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, *Signal processing and compression with wave packets*, in Proceedings of the International Conference on Wavelets, Marseille, 1989, Y. Meyer, ed., Masson, Paris.
[37] R. R. Coifman, Y. Meyer, and V. Wickerhauser, *Size properties of wavelet packets*, in [116], pp. 453–470.
[38] ———, *Wavelet analysis and signal processing*, in [116], pp. 453–470.
[39] R. R. Coifman and R. Rochberg, *Representation theorems for holomorphic and harmonic functions in L^p*, Astérisque, 77 (1980), pp. 11–66.
[40] R. R. Coifman and M. L. Wickerhauser, *Entropy based algorithms for best basis selection*, IEEE Trans. Inform. Theory, 38 (1992), pp. 713–718.
[41] D. Colella and C. Heil, *Characterizations of scaling functions: I. Continuous solutions*, SIAM J. Math. Anal., To appear.
[42] ———, *The characterization of continuous four-coefficient scaling functions and wavelets*, IEEE Trans. Inform. Theory, 38 (1992).
[43] J. M. Combes, A. Grossmann, and P. Tchamitchian, eds., *Wavelets: Time-Frequency Methods and Phase Space*, Inverse problems and theoretical imaging, Springer-Verlag, 1989.
[44] S. Dahlke and I. Weinreich, *Wavelet bases adapted to pseudo-differential operators*, Appl. Comput. Harmon. Anal., To appear.
[45] ———, *Wavelet-Galerkin-methods: An adapted biorthogonal wavelet basis*, Constr. Approx., 9 (1993), pp. 237–262.
[46] W. Dahmen and A. Kunoth, *Multilevel preconditioning*, Numer. Math., 63 (1992), pp. 315–344.
[47] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 41 (1988), pp. 909–996.
[48] ———, *The wavelet transform, time-frequency localization and signal analysis*, IEEE Trans. Inform. Theory, 36 (1990), pp. 961–1005.
[49] ———, *Ten Lectures on Wavelets*, no. 61 in CBMS-NSF Series in Applied Mathematics, SIAM, Philadelphia, 1992.
[50] ———, *Orthonormal bases of compactly supported wavelets II. Variations on a theme*, SIAM J. Math. Anal., 24 (1993), pp. 499–519.
[51] I. Daubechies, A. Grossmann, and Y. Meyer, *Painless nonorthogonal expansions*, J. Math. Phys., 27 (1986), pp. 1271–1283.
[52] I. Daubechies and J. C. Lagarias, *Two-scale difference equations I. Existence and global regularity of solutions*, SIAM J. Math. Anal., 22 (1991), pp. 1388–1410.
[53] ———, *Two-scale difference equations II. Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal., 23 (1992), pp. 1031–1079.
[54] C. de Boor, R. A. DeVore, and A. Ron, *On the construction of multivariate (pre)wavelets*, Constr. Approx., 9 (1993), pp. 123–166.
[55] B. Deng and B. Jawerth, *Biorthogonal wavelet packets on closed intervals*. Preprint.
[56] G. Deslauriers and S. Dubuc, *Interpolation dyadique*, in Fractals, Dimensions non entières

et applications, Masson, Paris, Paris, 1987, pp. 44–55.
[57] ———, *Symmetric iterative interpolation processes*, Constr. Approx., 5 (1989), pp. 49–68.
[58] R. A. DeVore, B. Jawerth, and B. J. Lucier, *Image compression through wavelet transform coding*, IEEE Trans. Inform. Theory, 38 (1992), pp. 719–746.
[59] ———, *Surface compression*, Comput. Aided Geom. Des., 9 (1992), pp. 219–239.
[60] R. A. DeVore and B. J. Lucier, *Wavelets*, in Acta Numerica 1, Cambridge University Press, 1991, pp. 1–56.
[61] D. L. Donoho, *Interpolating wavelet transforms*. Preprint Department of Statistics, Stanford University, 1992.
[62] ———, *Unconditional bases are optimal bases for data compression and for statistical estimation*, Appl. Comput. Harmon. Anal., 1 (1993), pp. 100–115.
[63] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness by wavelet shrinkage*. Preprint Department of Statistics, Stanford University, 1992.
[64] ———, *Ideal spatial adaptation via wavelet shrinkage*. Preprint Department of Statistics, Stanford University, 1992.
[65] ———, *New minimax theorems, thresholding, and adaptation*. Preprint Department of Statistics, Stanford University, 1992.
[66] T. Eirola, *Sobolev characterization of solutions of dilation equations*, SIAM J. Math. Anal., 23 (1992), pp. 1015–1030.
[67] H. G. Feichtinger and K. Gröchenig, *Theory and practise of irregular sampling*, in [13], pp. 305–363.
[68] ———, *Irregular sampling theorems and series expansions of band-limited functions*, SIAM J. Math. Anal., 23 (1992), pp. 530–556.
[69] G. Fix and G. Strang, *Fourier analysis of the finite element method in Ritz-Galerkin theory*, Stud. Appl. Math, 48 (1969), pp. 265–273.
[70] P. Franklin, *A set of continuous orthogonal functions*, Math. Ann, 100 (1928), pp. 522–529.
[71] M. Frazier and B. Jawerth, *Decomposition of Besov spaces*, Indiana Univ. Math. J., 34 (1985), pp. 777–799.
[72] ———, *The φ-transform and applications to distribution spaces*, in Function Spaces and Applications, M. C. et al., ed., no. 1302 in Lecture Notes in Math., 1988, pp. 223–246.
[73] ———, *A discrete transform and decompositions of distribution spaces*, J. Func. Anal, 93 (1990), pp. 34–170.
[74] M. Frazier, B. Jawerth, and G. Weiss, *Littlewood-Paley theory and the study of function spaces*, no. 79 in Regional Conference Series in Mathematics, American Mathematical Society, Providence, 1991.
[75] J. Geronimo, D. Hardin, and P. R. Massopust, *Fractal functions and wavelet expansions based on several scaling functions*, J. Approx. Theory, To appear.
[76] R. Glowinski, W. M. Lawton, M. Ravechol, and E. Tenenbaum, *Wavelet solution of linear and nonlinear elliptic parabolic and hyperbolic problems in one space dimension*, in Proceedings of the 9th International Conference on Numerical Methods in Applied Sciences and Engineering, SIAM, Philadelphia, 1990.
[77] A. Grossmann, R. Kronland-Martinet, and J. Morlet, *Reading and understanding continuous wavelet transforms*, in [43], pp. 2–20.
[78] A. Grossmann and J. Morlet, *Decompostion of Hardy functions into square integrable wavelets of constant shape*, SIAM J. Math. Anal., 15 (1984), pp. 723–736.
[79] ———, *Decomposition of functions into wavelets of constant shape, and related transforms*, in Mathematics and Physics, Lectures on Recent Results, L. Streit, ed., World Scientific Publishing, Signapore, 1985.
[80] A. Grossmann, J. Morlet, and T. Paul, *Transforms associated to square integrable group representations I. General results*, J. Math. Phys., 26 (1985), pp. 2473–2479.
[81] A. Haar, *Zur Theorie der orthogonalen Funktionen-Systeme*, Math. Ann., 69 (1910), pp. 331–371.
[82] D. P. Hardin, B. Kessler, and P. R. Massopust, *Multiresolution analyses based on fractal functions*, J. Approx. Theory, 71 (1992), pp. 104–120.
[83] C. E. Heil and D. F. Walnut, *Continuous and discrete wavelet transforms*, SIAM Rev., 31 (1989), pp. 628–666.
[84] S. Jaffard, *Wavelet methods for fast resolution of elliptic problems*, SIAM J. Numer. Anal., 29 (1992), pp. 965–986.
[85] A. J. E. M. Janssen, *The Zak transform: a signal transform for sampled time-continuous signals*, Philips J. Res., 43 (1988), pp. 23–69.
[86] B. Jawerth, *On Besov spaces*, Tech. Report 1, Lund, 1977.
[87] B. Jawerth and G. Peters, *Wavelets on non smooth sets of $\mathbf{R}^n$*. Preprint, 1993.

[88] B. Jawerth and W. Sweldens, *Wavelet multiresolution analyses adapted for the fast solution of boundary value ordinary differential equations*, in Sixth Copper Mountain Conference on Multigrid Methods, N. D. Melson, T. A. Manteuffel, and S. F. McCormick, eds., NASA Conference Publication 3224, 1993, pp. 259–273.

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

[90] S. Kelly, M. Kon, and L. Raphael, *Pointwise convergence of wavelet expansions*, Bull. Amer. Math. Soc. (N.S.), 29 (1994).

[91] J. R. Klauder and B.-S. Skagerstam, *Coherent States*, World Scientific, Signapore, 1985.

[92] T. H. Koornwinder, ed., *Wavelets: an elementary treatment of theory and applications*, no. 1 in Series in Approximations and Decompositions, World Scientific, Signapore, 1993.

[93] A. Latto and E. Tenenbaum, *Les ondelettes à support compact et la solution numérique de l'équation de burgers*, C. R. Acad. Sci. Paris Sér. I Math., 311 (1990), pp. 903–909.

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

[95] P.-G. Lemarié, *Ondelettes a localisation exponentielle*, J. Math. Pures Appl., 67 (1988), pp. 227–236.

[96] ———, ed., *Les Ondelettes en 1989*, no. 1438 in Lecture Notes in Math., Springer-Verlag, 1990.

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

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

[99] W. Madych, *Some elementary properties of multiresolution analysis of* $\mathrm{L}^2(\mathbf{R})$, in [21], pp. 259–294.

[100] S. Mallat and W. L. Hwang, *Singularity detection and processing with wavelets*, IEEE Trans. Inform. Theory, (1992), pp. 617–643.

[101] S. Mallat and S. Zhong, *Wavelet transform maxima and multiscale edges*, in [116], pp. 67–104.

[102] ———, *Characterization of signals from multiscale edges*, IEEE Trans. Patt. Anal. Mach. Intell., 14 (1992), pp. 710–732.

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

[104] ———, *Multiresolution approximations and wavelet orthonormal bases of* $L^2(\mathbf{R})$, Trans. Amer. Math. Soc., 315 (1989), pp. 69–87.

[105] ———, *A theory for multiresolution signal decomposition: The wavelet representation*, IEEE Trans. Patt. Anal. Mach. Intell., 11 (1989), pp. 674–693.

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

[107] ———, *Ondelettes sur l'intervalle*, Rev. Mat. Iberoamericana, 7 (1992), pp. 115–133.

[108] ———, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, 1993.

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

[110] 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 (1991), pp. 331–352.

[111] P. Oswald, *On discrete norm estimates related to multilevel preconditioners in the finite element method*, in Constructive Theory of Functions, Proc. Int. Conf. Varna 1991, Bulg. Acad. Sci., Sofia, 1992, pp. 203–214.

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

[113] S. D. Riemenschneider and Z. Shen, *Wavelets and pre-wavelets in low dimensions*, J. Approx. Theory, 71 (1992), pp. 18–38.

[114] O. Rioul, *Simple regularity criteria for subdivision schemes*, SIAM J. Math. Anal., 23 (1992), pp. 1544–1576.

[115] O. Rioul and M. Vetterli, *Wavelets and signal processing*, IEEE Signal Proc. Mag., (1991), pp. 14–38.

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

[117] N. Saito and G. Beylkin, *Multiresolution representations using the autocorrelation func-*

tions of compactly supported wavelets. Preprint University of Colorado at Boulder.
[118] I. J. Schoenberg, *Cardinal interpolation and spline functions*, J. Approx. Theory, 2 (1969), pp. 167–627.
[119] L. L. Schumaker and G. Webb, eds., *Recent Advances in Wavelet Analysis*, Academic Press, 1993.
[120] M. J. Shensa, *Wedding the à trous and Mallat algorithms*, IEEE Trans. Signal Process., 40 (1992), pp. 2464–2482.
[121] J. Stöckler, *Multivariate wavelets*, in [21], pp. 325–356.
[122] G. Strang, *Wavelets and dilation equations: A brief introduction*, SIAM Rev., 31 (1989), pp. 614–627.
[123] ———, *Wavelet transforms versus Fourier transforms*, Bull. Amer. Math. Soc. (N.S.), 28 (1993), pp. 288–305.
[124] G. Strang and G. Fix, *A Fourier analysis of the finite element variational method*, in Constructive aspects of Functional Analysis, Edizione Cremonese, Rome, 1973.
[125] R. S. Strichartz, *How to make wavelets*, Amer. Math. Monthly, 100 (1993), pp. 539–556.
[126] J. O. Strömberg, *A modified Franklin system and higher order spline systems on* $\mathbf{R}^n$ *as unconditional bases for Hardy spaces*, in Conference on Harmonic Analysis in Honor of Antoni Zygmund, B. et al., ed., vol. II, Univ. of Chicago Press, 1981, pp. 475–494.
[127] W. Sweldens and R. Piessens, *Asymptotic error expansions of wavelet approximations of smooth functions II*, Tech. Report 1993:4, Industrial Mathematics Initiative, Department of Mathematics, University of South Carolina.
[128] ———, *Quadrature formulae and asymptotic error expansions for wavelet approximations of smooth functions*, SIAM J. Numer. Anal., To appear.
[129] M. Unser and A. Aldroubi, *Polynomial splines and wavelets — a signal processing perspective*, in [21], pp. 543–601.
[130] M. Unser, A. Aldroubi, and M. Eden, *On the asymptotic convergence of b-spline wavelets to Gabor functions*, IEEE Trans. Inform. Theory, 38 (1992), pp. 864–872.
[131] ———, *A family of polynomial spline wavelet transforms*, Signal Process., 30 (1993), pp. 141–162.
[132] P. P. Vaidyanathan, *Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having perfect reconstruction property*, IEEE Trans. Acoust. Speech Signal Process., 36 (1987), pp. 476–492.
[133] M. Vetterli, *Filter banks allowing perfect reconstruction*, Signal Process., 10 (1986), pp. 219–244.
[134] M. Vetterli and C. Herley, *Wavelets and filter banks: theory and design*, IEEE Trans. Acoust. Speech Signal Process., 40 (1992), pp. 2207–??
[135] L. F. Villemoes, *Wavelet analysis of refinable functions*, in Curves and Surfaces II, P. J. Laurent, A. L. Méhauté, and L. L. Schumaker, eds., A. K. Peters, Boston.
[136] ———, *Energy moments in time and frequency for two-scale difference equation solutions and wavelets*, SIAM J. Math. Anal., 23 (1992), pp. 1119–1543.
[137] H. Volkmer, *On the regularity of wavelets*, IEEE Trans. Inform. Theory, 38 (1992), pp. 872–876.
[138] G. G. Walter, *A sampling theorem for wavelet subspaces*, IEEE Trans. Inform. Theory, 38 (1992), pp. 881–884.
[139] M. V. Wickerhauser, *Acoustic signal compression with wavelet packets*, in [21], pp. 679–700.
[140] J.-C. Xu and W.-C. Shann, *Galerkin-wavelet methods for two-point boundary value problems*, Numer. Math., 63 (1992), pp. 123–142.

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_gd, and starting indices min_gp, min_hd and min_gd respectively. These coefficients are stored in 4 vectors such that

$$hp[k] = a\, h_{k+min\_hp}, \quad gp[k] = a\, g_{k+min\_gp},$$

$$hd[k] = b\, \tilde{h}_{k+min\_hd}, \quad \text{and} \quad gd[k] = b\, \tilde{g}_{k+min\_gd},$$

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

$$w = \left[\begin{array}{ccccccccccc} \lambda_{0,0} & \gamma_{0,0} & \gamma_{1,0} & \gamma_{1,1} & \gamma_{2,0} & \cdots & \gamma_{2,3} & \cdots & \gamma_{n-1,0} & \cdots & \gamma_{n-1,2^{n-1}-1} \end{array} \right].$$

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, $a+b$, $\ldots$, c 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
```