A Survey of the Theory of Wavelets

Haikel Hichri
Department of Computer Science
University of Saskatchewan
57 Campus Drive
Saskatoon, SK S7N 5A9
CANADA
e-mail: hah134@cs.usask.ca

Supervisor: Carl McCrosky

Abstract

This paper provides an introduction to the theory of wavelets and its relation to Fourier analysis and subband filtering. It starts with a brief overview of some signal processing preliminaries such as discrete-time signals, discrete-time systems and Fourier analysis. It then explains the motivation behind wavelet analysis and compares it to Fourier analysis. This is followed by an overview of the theory of wavelets in which we concentrate on one type of wavelets, namely *orthogonal* wavelets developed by Daubechies (Daubechies 1992). We describe them through the theoretical framework of *MultiResolution Analysis (MRA)*(Mallat 1989). Finally, the paper describes the *Discrete Wavelet Transform (DWT)* algorithm and explains its connection to subband filtering in discrete signal processing.

Introduction.

In the past decade, wavelet analysis has grown from a mathematical curiosity into a major source of new signal processing algorithms. Wavelets are functions that are used as building blocks to represent other functions. The idea is not new, approximation of functions using superposition of a simpler set of functions has existed since the early 1800's, when Fourier discovered that he could superpose sines and cosines to represent other functions. Each sinusoid in Fourier analysis has a different frequency. Therefore, Fourier methods analyse functions according to frequency. However, in wavelet analysis we change the frequency of the basis functions as well as their interval of support [1]. That is achieved by changing the *scale* of the wavelet. Increasing the scale stretches the wavelet whereas decreasing the scale squeezes it. The basic idea behind this is to take advantage of the fact that small features (high frequency) of a function are localised in time whereas gross features (low frequency) are spread over time. Because of that wavelets are superior to sinusoids at representing functions with sharp *spikes*. These kinds of functions appear in many practical applications.

[1] The support is the interval over which the function is non zero

Real life images, for example, are composed of large areas of similar information but sharp changes at object edges. Psychophysicists and visual psychologists have performed many experiments on the Human Visual System (HVS) to determine how it processes images. They have discovered that the eye processes an image at different scales (Marr 1982). So, the eye is more sensitive to object edges rather than minor details inside. Therefore, wavelet analysis produces better results for image processing application such as computer vision, pattern recognition and image compression.

Wavelet theory has been developed fully only recently, although similar ideas and constructions could be traced back to Haar (1910). The idea of looking at a signal at various scales and analysing it with various resolutions has emerged independently in many different fields of mathematics, physics and engineering. In the mid-eighties, a mathematician named Meyer and others in the fields of geophysics built strong mathematical foundations around the subject and named their work *wavelets* (Meyer 1989). The attention of the signal processing community was soon caught when Mallat, in addition to his contribution to theory of wavelets, established connections to discrete signal processing results (Mallat 1989). One year later Daubechies used Mallat's work to build a set of orthogonal wavelet basis (Daubechies 1992). Since then, a number of theoretical, as well as practical contributions have been made on various aspects of wavelets, and the subject is growing rapidly. There are many variants of wavelets such as *orthogonal wavelets*, *biorthogonal wavelets* and *wavelet packets* (Kaiser 1994). There are also several types of the wavelet transforms. We have the *continuous wavelet transform* when scale and time parameters of the wavelet basis are continuous. When these parameters are discrete we get the *wavelet series expansion*. Finally, if the signal to be analysed is also discrete we get the *discrete wavelet transform (DWT)*. Notice that these three types are analogous to the corresponding Fourier types, *continuous Fourier transform*, *Fourier series expansion* and *discrete Fourier transform (DFT)*, respectively.

This paper is restricted to the orthogonal wavelet transform developed by Daubechies. In this set of wavelets all scales used are powers of 2. In the paper, we first, discuss some signal processing preliminaries that are fundamental to understanding wavelets. That includes discrete-time signals, discrete systems and Fourier analysis. We, then, give a general overview of *MultiResolution Analysis (MRA)* which provides a strong theoretical framework for understanding wavelets. Next we describe the DWT algorithm in 1D and 2D. Finally, we briefly explain the connection between the wavelet algorithm and what is known in signal processing as *subband filtering* algorithm.

Signal processing preliminaries.

Fourier analysis.

This section presents an overview of Fourier analysis. For more details see any introductory signal processing book such as (Van den Enden & Verhoeckx 1991) (Kaiser 1994) (Higgins 1990) (Alkin 1994). Some of these books treat the subject from an engineering angle whereas others from a mathematical point of view.

It was remarked by *Fourier* that an arbitrary periodic function f, with period T, could be approximated by a superposition of sine and cosine functions.

$$f(t) = \sum_{k=-\infty}^{+\infty} F(k)\; e^{j2\pi k\,\frac{t}{T}} \tag{1}$$

where $e^{jw} = \cos w + j \sin w$ (called Euler's equation where j is the complex number $\sqrt{-1}$). This is known as the *Fourier series* expansion of f and $F(k)$'s are called *Fourier coefficients*. They are computed via the formula

$$F(k) = \frac{1}{T}\int_0^T f(t)\; e^{-j2\pi k\,\frac{t}{T}}\; dt \tag{2}$$

The function f must meet certain conditions or else the infinite sum in (1) may diverge. Namely, the function should be *square integrable* i.e. $\int_0^T |f(t)|^2 dt < +\infty$ (has *finite energy*) and differentiable. The set of *square integrable* periodic functions with period T is denoted by $L^2(T)$. Fortunately, for most practical applications, it will not be necessary to deal with these issues.

The representation of f by (1) can also be viewed (in the context of functional analysis) as an expansion in a bases $\{\phi_k(t) = e^{j2\pi k\,\frac{t}{T}}, k \in Z\}$ of $L^2(T)$. The inner product between two functions f and g is defined as, [2]

$$\langle f, g\rangle \stackrel{def}{=} \int_0^T f(t)\,\overline{g(t)}\; dt \tag{3}$$

[2] $\overline{g(t)}$ indicates complex conjugate, if $a = x + iy$ then $\overline{a} = x - iy$. Generally, in the set of all finite energy functions L^2, the integral stretches to infinity.

therefore (2) is simply $F(k) = \langle f, \phi_k\rangle$ and (1) becomes our usual *basis expansion relation*

$$f(t) = \sum_{k=-\infty}^{+\infty} \langle f, \phi_k\rangle\; \phi_k(t) \tag{4}$$

Furthermore, we have

$$\langle \phi_k, \phi_{k'}\rangle = \begin{cases} 1 & \text{if } k = k' \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

thus $\{\phi_k(t) = e^{j2\pi k\,\frac{t}{T}}\}$ is an *Orthonormal Bases (OB)* of $L^2(T)$.

The *Fourier Transform (FT)* is a generalisation of the Fourier series to an infinite interval. It is defined by the following pair

$$F(w) = \int_{-\infty}^{+\infty} f(t)\; e^{-j2\pi w\,t}\; dt \tag{6}$$

$$f(t) = \int_{-\infty}^{+\infty} F(w)\; e^{j2\pi w\,t}\; dw \tag{7}$$

The Fourier transform, $F(w)$ can also be viewed as describing the *spectral (frequency) content* of f, while f itself describes the *data (time) content.* The FT pair provides the means of converting between these two equivalent representations of f.

In the discrete case, function values are known only at discrete intervals t_s. The computation of the Fourier transform becomes an infinite sum

$$F(w) = \sum_{n=-\infty}^{+\infty} f(n)\; e^{-j2\pi w\,n} \tag{8}$$

Where the sampling interval t_s is set to 1 for simplicity. Sampling has a special effect on the nature of the Fourier transform. Namely, it becomes periodic with period $f_s = 1/t_s = 1$. This is because we are sampling at a fixed frequency f_s. If the function contains frequencies higher than f_s, Sampling may also cause what is known as the *aliasing* effect, where higher frequencies interfere with lower ones. This leads to the reconstructed signal being in error. Again for more details about aliasing, see any introductory signal processing book cited earlier. Since $F(w)$ is periodic then it has a Fourier series expansion according to (1). Therefore (8) is the Fourier series expansion of $f(w)$ because Fourier series are unique. Hence $f(n)$ act like Fourier coefficients and are computed similarly via the formula below (recall that the period is $f_s = 1$)

$$f(n) = \int_0^1 F(w)\; e^{j2\pi w\,n}\; dw \tag{9}$$

The *Discrete Fourier Transform (DFT)* is similar to (8) and (9) but in this case we only use N samples of *both* $F(w)$ and $f(n)$. The forward and inverse DFT are given by the following

$$F(k) = \sum_{n=0}^{N-1} f(n)\; e^{-j2\pi\,\frac{k}{N}\,n} \tag{10}$$

$$f(n) = \frac{1}{N} \sum_{k=0}^{N-1} F(k)\ e^{j2\pi \frac{k}{N} n} \quad (11)$$

Discrete-time signals and impulses.

A discrete-time signal is a continuous signal $f(t)$ sampled at discrete times. It is represented by a sequence $x(n)$ such that $x(n) = f(nt_s)$ where t_s is the sampling period.

An *impulse* δ is a sequence such that

$$\delta(n) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

Note that it is an *infinite sequence* with only one non-zero value equal to 1. Thus we can write a discrete-time signal x as

$$x(n) = \sum_{k=-\infty}^{+\infty} x(k)\delta(n-k) \quad (13)$$

Each one of $x(k)\delta(n-k)$ is an infinite sequence which is non-zero only at $n = k$ where it has the value of x. Thus x is expressed as a linear combination of the δ sequence and its linear shifts. In other words, the set of δ sequences are used as a bases for expressing other arbitrary sequences.

Discrete-time systems.

A *system* transforms a signal $x(t)$ into an output signal $y(t)$. A *discrete-time system* operates on discrete-time signals (sequences) as defined earlier. Let T be a discrete-time system and let $x(n)$ and $y(n)$ be the input and output sequence respectively, then we write

$$y(n) = T[x(n)] \quad (14)$$

The goal is to develop a good mathematical model for physical systems. A system is *linear* if

$$\forall\, sequences\ x, y\ and\ \forall\, \alpha, \beta \in R$$
$$T[\alpha x(n) + \beta y(n)] = \alpha T[x(n)] + \beta T[y(n)] \quad (15)$$

and *time invariant* if

$$\forall\, sequences\ x, y\ s.t.\ y(n) = T[x(n)]$$
$$\forall\, k \in Z\ T[x(n-k)] = y(n-k) \quad (16)$$

These two assumptions simplify system analysis considerably, so physical systems are modelled with linear and time-invariant mathematical models whenever possible. Recall from (13) that $x(n)$ can be expressed as a sum of shifted impulse sequences. Further, assume that T is linear and shift-invariant. Thus we can rewrite (14) as

$$\begin{aligned} y(n) &= T[\sum_{k=-\infty}^{+\infty} x(k)\delta(n-k)] \\ &= \sum_{k=-\infty}^{+\infty} x(k)T[\delta(n-k)] \quad (17) \\ &= \sum_{k=-\infty}^{+\infty} x(k)h(n-k) \quad (18) \end{aligned}$$

Step (17) in the above derivation uses linearity of T whereas step (18) uses its shift-invariance. $h(n)$ is an infinite sequence such that $h(n) = T[\delta(n)]$ and it is called the *impulse response* of system T. Therefore, the knowledge of $h(n)$, the impulse response, is sufficient to determine its response to any arbitrary input sequence $x(n)$.[3] The operation, performed in (18) between x and h, is called *convolution*. It is denoted by *, thus we write $y(n) = x(n) * h(n)$.

Similarly, we can use the Fourier decomposition of $x(n)$ given in (10) in the analysis of system T.

$$\begin{aligned} y(n) &= T[\frac{1}{N} \sum_{k=0}^{N-1} X(k)\ e^{j2\pi \frac{k}{N} n}] \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X(k) T[e^{j2\pi \frac{k}{N} n}] \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X(k)\ \ h(n) * e^{j2\pi \frac{k}{N} n} \quad (19) \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X(k) \sum_{l=-\infty}^{+\infty} h(l) e^{j2\pi \frac{k}{N}(n-1)} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi \frac{k}{N} n} \sum_{l=-\infty}^{+\infty} h(l) e^{-j2\pi \frac{k}{N} l} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi \frac{k}{N} n} H(k) \\ &= \frac{1}{N} \sum_{k=0}^{N-1} X(k)H(k) e^{j2\pi \frac{k}{N} n} \quad (20) \end{aligned}$$

Notice that in (19), we have used the fact that the convolution operation is commutative. Thus we have $T[x(n)] = y(n) = x(n) * h(n) = h(n) * x(n)$. $H(k)$ is the FT of the inpulse response $h(n)$ of T. It is called the frequency response of T.

Observing the similarity of (20) to the inverse DFT in (11), it is obvious that $X(k)H(k) = Y(k)$ is then the DFT of $y(n)$. Thus we have proved an important property of the DFT, that is, the convolution operation between x and h is equivalent to multiplication of the corresponding DFT's X and H. Formally, we write

$$\begin{gathered} y(n) = x(n) * h(n) \\ \Leftrightarrow \\ Y(k) = X(k)H(k) \end{gathered} \quad (21)$$

We have also shown that a system can be described in two equivalent ways either by its impulse response $h(n)$ or frequency response $H(k)$. Suppose that $H(k)$ is a sequence that is 1 at small frequencies k (k near 0) and 0 at large frequencies (k near $\pm N/2$). Then the

[3] Notice that h is not necessarily a one-value sequence like the δ sequence. In fact for most *useful* systems it is not.

application of the corresponding system T to an input sequence $x(n)$ eliminates its high frequency content (according to (21)). In this case system T is known, in signal processing, as a *low pass filter*. A *high pass filter* is defined similarly. Filters are common in signal processing and a lot of work has been done related to filter design and applications.

Motivation for wavelets and the Haar example.

Recall that in functional analysis a signal is said to be represented as a linear combination of basis functions. Generally we have

$$f(t) = \sum_i c_i \psi_i(t) \tag{22}$$

where the $\psi_i(x)$ are basis functions and the c_i are coefficients. Since the basis functions ψ_i are fixed, it is the coefficients which contain the information about the signal.

The simplest such representation uses translates of the impulse function as its only basis. It reveals information only about the time behaviour of the signal, because the coefficients give the values of the signal at a particular time. An impulse bases function cannot, however, provide information about the frequency content of the signal, because it has infinitesimally small support[4]. On the other hand, Fourier analysis uses sinusoidal functions as bases. Sinusoids give detailed information about the frequency content of the function, because each one has a well defined frequency and the coefficients show the importance of that frequency in the function. But, it cannot tell us where in time that frequency is, because its support is infinite.

It is also important to note that if we look at their respective supports in the frequency domain, we find that impulses have infinite support whereas sinusoids have infinitesimally small support.[5] The goal is to find basis functions that provide us with detailed information about a function in the time domain as well as the frequency domain. Unfortunately that is not possible because the resolution in time (Δt) and the resolution in frequency (Δw) [6] are governed by the *Heisenberg inequality*, namely

$$\Delta t \Delta w \geq \frac{1}{4\pi} \tag{23}$$

This inequality says that the resolution in time and in frequency cannot be made arbitrarily small, both at the same time. Thus we can have very good resolution in time if we are willing to settle for low resolution in frequency and vice versa. What we seek then, is a compromise. Fortunately, we can take advantage of the fact that high frequency events are concentrated (or localised) in time, whereas low frequency events are spread out (or non-local) in time. Thus we can design basis function that have variable support which allows us to trade off time and frequency resolution in different ways; for example, a basis function with wide support can examine a large region of the signal and resolve low frequency details accurately, while a short basis function can examine a small region of the signal to resolve high frequency (time) details accurately.

The *wavelet analysis* procedure adopts this idea with a further simplification. We only use one prototype function ψ called the *mother wavelet*. The rest of the functions in the basis set are scaled and translated versions of ψ, which are defined as follows

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k) \ \forall j,k \in Z \tag{24}$$

where j is the scale and k is the translation. Notice, first, that when j is small the wavelet is squeezed but it is streched when j is large. Second, notice that the value of the translation depends on the scale j because $\psi(2^{-j}x - k) = \psi(2^{-j}(x - 2^j k))$, so that the smaller is j, the smaller is the wavelets support and the smaller is the translation. This set of wavelets are *orthonormal* and relate to an old technique in signal processing known as *octave subband filtering*. This relation were first discovered by Mallat (Mallat 1989)

The earliest wavelet known is the *Haar* wavelet. It is a step function taking values 1 and -1 on $[0, 1/2)$ and $[1/2, 1)$. Figure 1 shows the Haar wavelet ψ and the dilated version $\psi_{1,0}$. It is known that any continuous function can be approximated uniformly by Haar wavelets. As a matter of fact, dilations and translations of the function ψ,

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k) \tag{25}$$

define an orthonormal basis in L^2. The orthonormality of $\psi_{j,k}$'s is easy to check using the definition of inner product given in (3).[7]

Any function f in L^2 is decomposed into the wavelet basis as follows,

$$f(t) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \langle f, \psi_{j,k} \rangle \, \psi_{j,k}(t) \tag{26}$$

Multiresolution analysis.

MultiResolution Analysis (MRA) was developed by Mallat (Mallat 1989) who was working in digital signal processing. He observed the relationship between *Quadrature mirror filters* (QMF) known in subband coding techniques and orthonormal wavelet bases. His work inspired Meyer who constructed the first non trivial wavelet. Unlike the Haar wavelet, Meyer's wavelet

[4]support of a function is the interval over which it is non-zero

[5]Because each sinusoid have a well defined frequency

[6]Both are equal to the support lengths in the time domain and the frequency domain respectively

[7]The inner product of $\psi_{j,k}$ and $\psi_{j',k'}$ is 1 only when $j = j'$ and $k = k'$, otherwise it is 0.

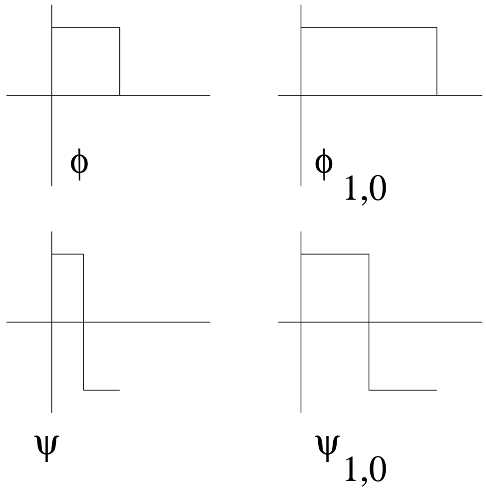

Figure 1: Haar example

is continuously differentiable. Daubechies used Mallat's work to construct a set of orthonormal wavelet basis functions that are perhaps the most elegant.

An MRA is defined as follows: let V_j be a set of subspaces of L^2 such that

$$... \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset ... \qquad (27)$$

$$\text{with} \quad \bigcup_{j \in Z} V_j = L^2 \quad and \quad \bigcap_{j \in Z} V_j = \emptyset$$

$$and$$

$$\forall f(t) \in V_0 \;\Leftrightarrow\; f(2^{-j}t) \in V_j \qquad (28)$$

The multiresolution aspect is a consequence of (28). An easy example is the *Haar* MRA system, where the subspaces are defined as follows

$$V_j = \{h \in L^2, \forall k \in Z : \\ h|_{[2^j k, 2^j(k+1)[} = Constant\} \qquad (29)$$

Figure 2 shows the projections $P_0 f$ and $P_1 f$ of a function f onto Haar subspaces V_0 and V_1 respectively. Notice that these projections are only an approximation of the original function f and that $P_0 f$ is a better approximation than $P_1 f$. In general we have the result

$$\lim_{n \to -\infty} P_n f = f \qquad (30)$$

In addition to (27) and (28), it is further required that there exists a function ϕ such that $\{\phi_{0,k}(t) = \phi(t-n); n \in Z\}$ is an in Orthonormal Basis (OB) in V_0. It follows (from (28)) that $\{\phi_{j,n}(t) = 2^{-j/2}\phi(2^{-j}t - n); n \in Z\}$ is an in OB in V_j. In the Haar example, one possible choice for ϕ is

$$\phi(t) = \begin{cases} 1 & \text{if } [0,1[\\ 0 & \text{otherwise} \end{cases} \qquad (31)$$

It is easy to check that these functions ϕ do form OB's for the Haar MRA subspaces V_j's defined in (29). Given all of the above requirements ((27) and (28) and

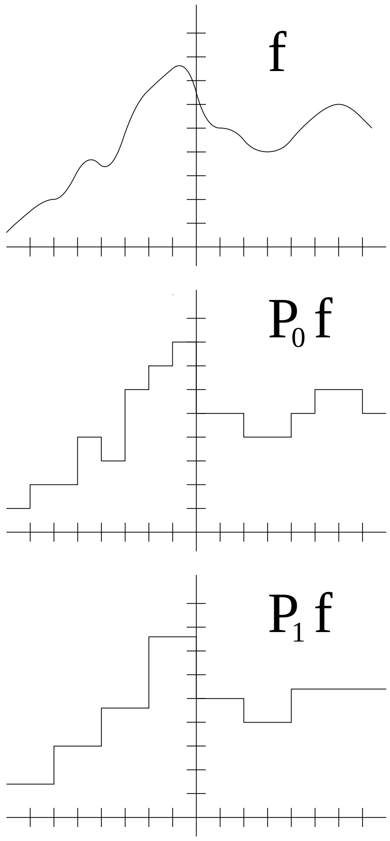

Figure 2: projections of f onto V_0 ($P_0 f$) and V_1 ($P_1 f$).

$$
\begin{pmatrix}
h(0) & h(1) & h(2) & h(3) & & & & & & & \\
h(3) & -h(2) & h(1) & -h(0) & & & & & & & \\
 & & h(0) & h(1) & h(2) & h(3) & & & & & \\
 & & h(3) & -h(2) & h(1) & -h(3) & & & & & \\
 & & & . & . & . & . & & & & \\
 & & & . & . & . & . & & & & \\
 & & & . & . & . & . & & & & \\
 & & & & & & & h(0) & h(1) & h(2) & h(3) \\
 & & & & & & & h(3) & -h(2) & h(1) & -h(0) \\
h(2) & h(3) & & & & & & & & h(0) & h(1) \\
h(1) & -h(0) & & & & & & & & h(3) & -h(2)
\end{pmatrix}
$$

Figure 3: Matrix of the wavelet filter coefficients

the existence of functions ϕ), it follows that there exists an OB of wavelets $\{\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k); j,k \in Z\}$ of L^2. It means that $\forall f \in L^2$ we have

$$f(t) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \langle f, \psi_{j,k} \rangle \psi_{j,k}(t) \tag{32}$$

This basis set could be explicitly constructed. Let us look at the general idea.

First we need to make a few more definitions. For all $j \in Z$, let W_j be the orthogonal complement of V_j in V_{j-1}, i.e. $W_j = V_{j-1} \cap V_j^{\perp}$. In other words W_j is the subspace that contains all elements that are orthogonal to all elements in V_j. Thus in a way W_j is like the orthogonal parts that when added to V_j results in V_{j-1}. We denote that by

$$V_{j-1} = V_j \oplus W_j \tag{33}$$
$$and$$
$$W_j \perp W_{j'} \quad if \;\; j \neq j'$$

By induction on j, it follows that

$$L^2 = \bigoplus_{j=-\infty}^{+\infty} W_j \tag{34}$$

In other words L^2 is decomposed into mutually orthogonal subspaces W_j. Furthermore, W_j subspaces inherit the scaling property (28) from V_j's

$$f(t) \in W_0 \leftrightarrow f(2^{-j}t) \in W_j \tag{35}$$

Therefore, if we have a function $\psi \in W_0$ such that $\{\psi_{0,k}(t) = \psi(t-k), k \in Z\}$ is an OB of W_0, then it follows from (35) that $\{\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k), k \in Z\}$ is an OB in V_j. So, all we need, is to construct this function $\psi \in W_0$ such that $\{\psi_{0,k}; k \in Z\}$ is an OB of W_0. Recall that $\phi \in V_0 \subset V_{-1}$ and $\{\phi_{-1,n}\}$ is an OB for V_{-1} thus

$$\phi(t) = \textstyle\sum_n h(n)\phi_{-1,n}(t) \tag{36}$$
$$with$$
$$h(n) = \langle \phi, \phi_{-1,n} \rangle \tag{37}$$
$$\tag{38}$$

We can rewrite (36) using the scaling property between V_{-1} and V_0 as follows

$$\phi(t) = \sum_n h(n) 2^{1/2} \phi(2t - n) \tag{39}$$

One possible set of ψ's that form and OB in W_0 is given bellow.

$$\psi(t) = \textstyle\sum_n (-1)^n h(-n+1)\phi_{-1,n}(t) \tag{40}$$
$$or$$
$$\psi(t) = \textstyle\sum_n (-1)^n h(-n+1) 2^{1/2}\phi(2t-n) \tag{41}$$

We omit the proof of the construction. It is easy to check that $\psi(t)$ as defined in (40) and $\phi(t)$ as defined in (36) are orthogonal[8]. We can also check that the set of functions $\{\psi_{j,k}\}$ alone are orthonormal. This is sufficient to show that the above choice in (40) is valid.

Let us see how all of this applies to the Haar example. Recall that ϕ in the Haar example was defined as

$$\phi(t) = \begin{cases} 1 & \text{if } t \in [0,1[\\ 0 & \text{otherwise} \end{cases} \tag{42}$$

Therefore we can compute $h(n)$ easily as [9]

$$\begin{aligned} h(n) &= 2^{1/2} \int \phi(t)\phi(2t-n)dt && (43) \\ &= \begin{cases} 2^{1/2}2^{-1} = 2^{-1/2} & \text{if } n = 0,1 \\ 0 & \text{otherwise} \end{cases} && (44) \end{aligned}$$

Consequently,

$$\begin{aligned} \phi(t) &= \phi_{0,0}(t) && (45) \\ &= 2^{-1/2}\phi_{-1,0}(t) + 2^{-1/2}\phi_{-1,1}(t) && (46) \\ & and && \\ \psi(t) &= \psi_{0,0}(t) && (47) \\ &= 2^{-1/2}\phi_{-1,0}(t) - 2^{-1/2}\phi_{-1,1}(t) && (48) \end{aligned}$$

Notice that ψ in (48) is just the Haar wavelet introduced in the previous section!

[8] Furthermore, $\psi_{j,k}(t)$ and $\phi_{j,k}(t)$ are orthogonal for all j and k. It is a good exercise to check.

[9] To compute the inner products just multiply the functions (ϕ and $\phi_{1,0}$ shown in Figure 1) and integrate the area under the curve.

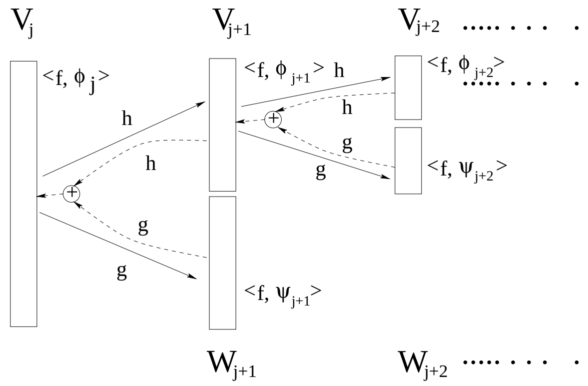

Figure 4: Block diagram of the forward and inverse wavelet algorithm

Wavelet algorithm.

Suppose we have computed, or are given the inner products of f with the $\phi_{j,k}$ at some given scale j. For example, in the Haar system, these are given by the projections of f onto subspace V_j i.e. $P_j f$. [10] Without loss of generality we can assume $j = 0$. Our goal is to compute the wavelet coefficients which are $\langle f, \psi_{j,k}\rangle$ in (32). Obviously, we can only do that for $j \geq 1$, because we cannot create extra detail about f that is not originally given to us. So, in (41) we have

$$\psi(t) = \sum_n (-1)^n h(-n+1)\, 2^{1/2}\phi(2t-n) \quad (49)$$

$$= \sum_n g_n 2^{1/2}\, \phi(2t-n) \quad (50)$$

where $g(n)$ is defined as $(-1)^n h(-n+1)$. Consequently,

$$\psi_{j,k}(t) = 2^{\frac{-j}{2}}\psi(2^{-j}t-k) \quad (51)$$

$$= 2^{\frac{-j}{2}}\sum_n g(n)2^{\frac{1}{2}}\phi(2(2^{-j}t-k)-n) \quad (52)$$

$$= \sum_n g(n)2^{\frac{-j+1}{2}}\phi(2^{-j+1}t-2k-n) \quad (53)$$

$$= \sum_n g(n)\phi_{j-1,2k+n}(t) \quad (54)$$

Taking the inner product with f of both sides of (54), we get

$$\langle f, \psi_{j,k}\rangle = \sum_n g(n)\langle f, \phi_{j-1,2k+n}\rangle \quad (55)$$

Similarly, from (39) we can derive the following relation

$$\langle f, \phi_{j,k}\rangle = \sum_n h(n)\langle f, \phi_{j-1,2k+n}\rangle \quad (56)$$

Now the algorithm is becoming clear. Since we have $\langle f, \phi_{0,k}\rangle$, then we can compute $\langle f, \psi_{1,k}\rangle$ and $\langle f, \phi_{1,k}\rangle$ using (55) and (56) respectively. Next we use $\langle f, \phi_{1,k}\rangle$ to compute $\langle f, \psi_{2,k}\rangle$ and $\langle f, \phi_{2,k}\rangle$, and so on.

In the discrete case, a finite number N, of inner product (or samples) are given at the start of the algorithm ($j = 0$). We can easily check that the number of inner products computed using (55) is $\frac{N}{2}$ (Notice the factor of $2k$ in the subscript of $\phi_{j-1,2k+n}$ which means that we ignore every other value of the input). This is similarly the case for (56). Hence, the total number of values is invariant from one step to the next, But half of these values represent inner products with the wavelet basis functions $\psi_{j,k}$ and the other half represent the inner products with the scaling basis functions $\phi_{j,k}$. Computing these inner products in each step is equivalent to applying a matrix operator on the input vector. For example supose h and g have 4 coefficients each, Figure 3 shows their arranged in the matrix (racall that $g(n) = (-1)^n h(-n+1)$). The result of the matrix operation is a vector of the same size which is then rearranged so that all wavelet coefficients are in the lower half, whereas the smooth version coefficients are in the upper half. Figure 4 depicts this algorithm graphically.

The inverse process is shown with dashed lines. To reconstruct the original signal we

- Upsample the vector of N values $\langle f, \psi_{j,k}\rangle$ by 2. i.e. consider them as even entries of a vector with odd entries equal to 0. Simillarly for the other vector of N values $\langle f, \phi_{j,k}\rangle$.
- Apply the sequences h and g to these vectors i.e. compute $\sum_n h(n)\langle f, \phi_{j,k+n}\rangle$ and $\sum_n g(n)\langle f, \psi_{j,k+n}\rangle$ for all $0 \leq k \leq 2N$.
- Add the resultant vectors, so for all $0 \leq k \leq 2N$

$$\langle f, \phi_{j-1,k}\rangle = \sum_n h(n)\langle f, \phi_{j,k+n}\rangle + \sum_n g(n)\langle f, \psi_{j,k+n}\rangle \quad (57)$$

The algorithm is generalised to 2 dimensions as shown in Figure 5. The 1D version is applied to the rows and columns separately. Next the same step is applied recursively to the smooth version of the image

[10] Note that, in the Haar example, this is the same as sampling the function!

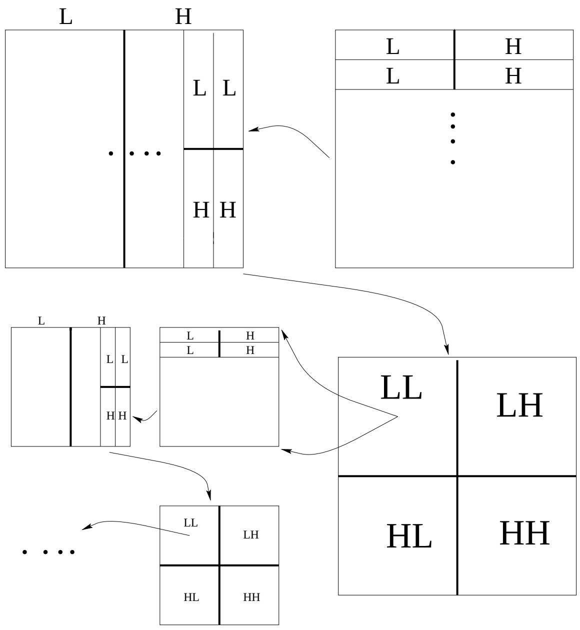

Figure 5: Block diagram of the 2D forward wavelet transform.

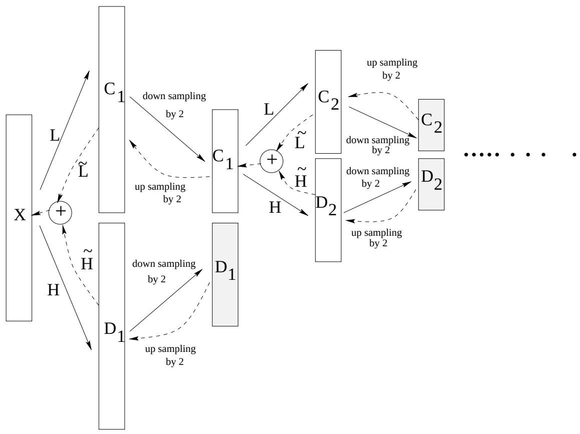

Figure 6: Block digram of subband filtering algorithm

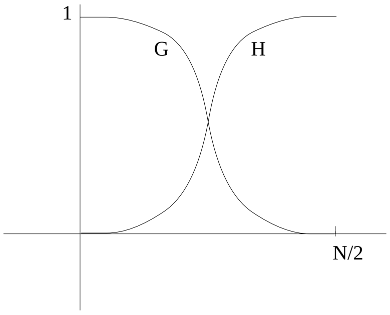

Figure 7: Fourier transform of wavelet low pass and high pass filters

(the top left quarter). This algorithm and its application to image compression are discussed in (Antonini, Mathieu, & Daubechies 1992). DWT is generalised to 3D in a similar fashion. Lewis and Knowles describe a video codec using a 3D DWT (Lewis 1990). However, this algorithm is not practical because of its large memory requirements.

Connection to subband filtering.

Subband filtering is a technique developed in signal processing. A block digram of the algorithm (also known as the *pyramidal algorithm*) is shown in Figure 6.

H and L are respectively high pass filter and low pass filter (introduced in Section Discrete-time systems). The signal X is filtered with both filters, producing two other signals C and D. C has the low frequency content of X while D has its high frequency content. Because the bandwidth of X is effectively divided in half, C and D need to be represented with only half as many points as X [11]. The process is then repeated for signal C recursively until the number of points is less than 4. Intuitively, in each step, low pass filtering produces a *smooth version* of X while high pass filtering produces a *detailed version* of X. To reconstruct the original signal we start from the right (top of pyramid) and work our way to the left. At each step, we upsample the vectors by two, apply the adjoint filters $\tilde{\mathrm{H}}$ and $\tilde{\mathrm{L}}$ and add the reasults. These steps are shown in the Figure 6 with dashed lines.

Clearly, there is a striking resemblance to the wavelet algorithm. In fact, the wavelet algorithm is a form of subband filtering. Look back at (55) and (56) and observe that they are similar to the convolution operation defined in (18) (except for the factor $2k$ instead of k). Thus we can consider $h(n)$ and $g(n)$ as impulse responses of some systems. But are h and g filters? To check, we need to look at their Fourier transforms. A plot of the Fourier transform H and G of two examples of h and g are shown in Figure 7.

[11] This is based on Nyquist theorem on sampling

Notice that the two curves are symmetric with respect to the middle line. A pair of filters that have this property are known as *Quadrature Mirror Filters* (QMF). Wavelet filters have the extra property of orthonormality which allows for the exact reconstruction of the signal from its wavelet coefficients.

Conclusions.

We have seen how wavelets relate to signal processing and filtering. We have also discussed the advantages of wavelet analysis versus Fourier analysis. Wavelets have been studied thoroughly during the last decade (Graps 1995). What remains is applying it to various fields in mathematics and engineering. Wavelet techniques has shown a great potential in image compression. It consistently outperforms old Fourier techniques in quality and compression ratios. For example the FBI‘s Criminal Justice Information Services Division developed standards for fingerprint digitization and compression in cooperation with the National Institute of Standards and Technology, Los Alamos National Laboratory, commercial vendors, and criminal justice communities (Bradley & Hopper 1993). The standard used wavelet transform to achieve high quality and high compression ratios of fingerprints. Many different image (Video) compression algorithms that use wavelets appear every year, with different strong and weak points. Examples include (Shapiro 1993) (Ohta 1993) (Yao 1992) (Katto, Nogaki, & Ohta 1994).

As for hardware implementations, there still needs to be more research in this area. In general, however, the wavelet transform requires much less hardware to implement than Fourier methods (Lewis 1992). Lewis and Knowles have also devised a clever idea of quantising the wavelet coefficients in order to build a simple VLSI architecture without multipliers (Lewis 1991).

References

Alkin, O. 1994. *Digital Signal Processing. A Laboratory Approach Using PC-DSP.* Prentice-Hall, Englewood Cliffs, NJ.

Antonini, Mark; Barlaud, M.; Mathieu, P.; and Daubechies, I. 1992. Image coding using the wavelet transform. *IEEE Transaction on Image Processing* 1(2).

Bradley, J.; Brislawn, C., and Hopper, T. 1993. The fbi wavelet/scaler quantization standard for grayscale fingerprint image compression. Technical Report LA-UR-93-1659, Los Alamos National Laboratory, Los Alamos, N.M.

Daubechies, I. 1992. *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series.

Graps, A. 1995. An introduction to wavelets. *IEEE Computational Science and Engineering* 2(2).

Higgins, R. J. 1990. *Digital Signal Processing in VLSI.* Prentice-Hall, Englewood Cliffs, NJ.

Kaiser, G. 1994. *A Freindly Guide To Wavelets.* Birkhauser, Boston.

Katto, Jiro; Ohki, J.-i.; Nogaki, S.; and Ohta, M. 1994. A wavelet codec with overlapped motion compression for very low bit-rate environment. *IEEE transactions on circuits and systems for video technology* 4(3):328–338.

Lewis, A. S.; Knowles, G. 1990. Video compression using 3d wavelet transforms. *Electronics Letters* 26(6):396–397.

Lewis, A. S.; Knowles, G. 1991. A vlsi architecture for the discrete wavelet transform without multipliers. *Electronics Letters* 27(2):171–173.

Lewis, A. S.; Knoweles, G. 1992. Image compression using the 2-d wavelet transform. *IEEE Transaction on Image Processing* 1(2):244–250.

Mallat, S. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7):674–693.

Marr, D. 1982. *Vision.* Freeman, New York.

Meyer, Y. 1989. Orthonormal wavelets. In *Wavelets, Time-Frequency Methods and Phase Space Proceedings of International conference.*

Ohta, Mutsumi; Nogaki, S. 1993. Hybrid picture coding with wavelet transform and overlapped motion-compensated interframe prediction coding. *IEEE Transactions on Signal Processing* 41(12):3416–3423.

Shapiro, J. M. 1993. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing* 41(12).

Van den Enden, A. W. M., and Verhoeckx, N. A. M. 1991. *Discrete-time signal processing An introduction.* Prentice Hall international(UK) Ltd.

Yao, S.; Clarke, R. J. 1992. Image sequence coding using adaptive vector quantisaion in wavelet transform domain. *Electronics Letters* 28(17):1566–1568.