MA495 Essay
Wavelet Bases of $L^2(\mathbb{R})$

Timothy John Sullivan
0006586

January 22, 2004

In this essay we shall briefly review the basic notions of Fourier analysis in the spaces $L^2[-\pi, \pi]$ and $L^2(\mathbb{R}^d)$, $d \in \mathbb{N}$. We shall see how the conventional Fourier basis and its associated series and transform have certain limitations and undesirable features. From this position, we shall introduce wavelets and multiresolution analysis and see the usefulness of these techniques in overcoming some of those limitations. We shall also follow the construction of some example wavelet bases of $L^2(\mathbb{R})$.

0 Introduction

Why are wavelets of interest? Wavelet analysis is a subfield of harmonic analysis, an outgrowth of Fourier analysis. In Fourier analysis, one attempts to describe a "signal" (in space or time) in terms of frequencies of oscillation. From a practical viewpoint, this scheme seems to offer the potential for great data compression, but in fact involves considerable loss of data because of fundamental limitations such as the Heisenberg Uncertainty Principle (Theorem 1.3). Wavelets offer ways to circumvent the problems associated with traditional Fourier analysis.

Yves Meyer, in the preface to the English edition of [12], gives an excellent summary of the status of wavelet theory:

> The "Theory of Wavelets" lies on the boundaries between (1) mathematics (2) scientific calculation (3) signal processing and

(4) image processing. The aim of the theory is to give a coherent set of concepts, methods and algorithms to deal with the difficulties met in each of these disciplines.

Wavelet analysis is a branch of mathematics with strong grounding on both the “pure” and “applied” sides of the subject. The mathematician will take interest in the area simply because of the formal qualities of the bases of $L^2(\mathbb{R})$ that are involved. The computer scientist or communications engineer will note instead the tremendous practical power of wavelet methods. The tools of wavelet analysis have, in many respects, developed in parallel in the “pure” and “applied” worlds; the synthesis of these two approaches lends much richness to the theory.

1 Basic Fourier Analysis

The notation for a function of a single variable shall be $f(\cdot)$. Many mathematicians would simply write f, and many physicists and engineers would write $f(x)$ (something the mathematians would consider heresy!). The author hopes that writing $f(\cdot)$ will combine the advantages of both notations, making it clear that $f(\cdot)$ is not the single quantity f, nor is it the function evaluated at a specific point, as in $f(x)$.

We shall work with the usual inner products and induced norms on $\mathbb{C}$-valued functions and sequences given by

$$\begin{aligned}
\langle f(\cdot), g(\cdot)\rangle_{L^2(\mathcal{D})} &:= \int_{\mathcal{D}} f(x)\overline{g(x)}\, dx; \\
\|f(\cdot)\|_{L^2(\mathcal{D})} &:= \langle f(\cdot), f(\cdot)\rangle_{L^2(\mathcal{D})}^{1/2} \\
&= \int_{\mathcal{D}} |f(x)|^2\, dx; \\
\langle (x_k), (y_k)\rangle_{\ell^2(K)} &:= \sum_{k\in K} x_k\overline{y_k}; \\
\|(x_k)\|_{\ell^2(K)} &:= \langle (x_k), (x_k)\rangle_{\ell^2(K)}^{1/2} \\
&= \sum_{k\in K} |x_k|^2;
\end{aligned}$$

where $\mathcal{D} \subseteq \mathbb{R}^d$ and K is some countable set. ($\mathcal{D}$ is assumed to be Lebesgue measurable, and all integrals are taken in the sense of Lebesgue.) Of course, $L^2(\mathcal{D})$ is the space of functions $\mathcal{D} \to \mathbb{C}$ of finite $L^2(\mathcal{D})$ norm (factored out by equality almost everywhere, if you wish!), and $\ell^2(K)$ the space of sequences $K \to \mathbb{C}$ of finite $\ell^2(K)$ norm.

At this point it may be useful to recall a few definitions and results from the theory of Hilbert spaces:

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. A subset $S \subseteq \mathcal{H}$ is called *orthonormal* if for all $x, y \in \mathcal{H}$,

$$\langle x, y \rangle = \begin{cases} 0 & \text{if } x \neq y; \\ 1 & \text{if } x = y. \end{cases}$$

Say $x \perp y$ if $\langle x, y \rangle = 0$ and define the *orthogonal complement* of $S \subseteq \mathcal{H}$ to be

$$S^\perp := \{x \in \mathcal{H} | \forall s \in S, x \perp s\}.$$

Naturally, we say $S \perp T$ for $S, T \subseteq \mathcal{H}$ if $s \perp t$ for all $s \in S$ and $t \in T$.

Theorem 1.1. (Orthonormal Bases of Hilbert Spaces) [13] *Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$, and let $\{e_k | k \in K\}$ be a countable orthonormal subset of $\mathcal{H}$. Then the following conditions are equivalent:*

(i) $\{e_i | k \in K\}^\perp = \{0\}$;

(ii) $\overline{\operatorname{span}\{e_k | k \in K\}} = \mathcal{H}$;

(iii) $\forall x \in \mathcal{H}$, $\|x\|^2 = \sum_{k \in K} |\langle x, e_k \rangle|^2 =: \|\langle x, e_k \rangle\|^2_{\ell^2(K)}$;

(iv) $\forall x \in \mathcal{H}$, $x = \sum_{k \in K} \langle x, e_k \rangle e_k$;

(v) *$\mathcal{H}$ is separable, i.e. has a countable subset that is dense in $\mathcal{H}$ with respect to the norm $\| \cdot \|$.*

If any (hence all) of these conditions holds then we say $\{e_k | k \in K\}$ is an *orthonormal basis* of $\mathcal{H}$.

The "next best" kind of basis after an orthonormal basis is a Riesz basis. A countable subset $\{e_k | k \in K\} \subseteq \mathcal{H}$ is a *Riesz basis* [2] for $\mathcal{H}$ if for all $x \in \mathcal{H}$,

(i) $\exists A, B > 0$ such that

$$A\|x\|^2 \leq \sum_{k\in K} |\langle x, e_k\rangle|^2 \leq B\|x\|^2;$$

(ii) $\overline{\operatorname{span}\{e_k | k \in K\}} = \mathcal{H}$;

(iii) $\forall (c_k) \in \ell^2(K)$, $\exists A, B > 0$ such that

$$A\|(c_k)\|^2_{\ell^2(K)} \leq \left\|\sum_{k\in K} c_k e_k\right\|^2 \leq B\|(c_k)\|^2_{\ell^2(K)}.$$

From now on, unless otherwise indicated, $\langle\cdot,\cdot\rangle := \langle\cdot,\cdot\rangle_{L^2(\mathcal{D})}$ and $\|\cdot\| := \|\cdot\|_{L^2(\mathcal{D})}$.

The family of functions $\{e_k(\cdot)|k \in \mathbb{Z}\}$ forms an orthonormal basis of the Hilbert space $L^2[-\pi,\pi]$, where

$$e_k(x) := \frac{1}{\sqrt{2\pi}} e^{-ikx}.$$

We define the *Fourier coefficients* $\widehat{f_k} := \langle f(\cdot), e_k(\cdot)\rangle$. By condition (iii) in Theorem 1.1 we see that $f(\cdot) \in L^2[-\pi,\pi] \Leftrightarrow (\widehat{f_k})_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ and have *Parseval's equality* for $f(\cdot) \in L^2[-\pi,\pi]$:

$$\|f(\cdot)\|_{L^2[-\pi,\pi]} = \left\|\left(\widehat{f_k}\right)\right\|_{\ell^2(\mathbb{Z})}. \tag{1.1}$$

By condition (iii) in Theorem 1.1 we have the reconstruction property that $\forall f(\cdot) \in L^2[-\pi,\pi]$,

$$f(\cdot) = \sum_{k\in\mathbb{Z}} \widehat{f_k} e_k(\cdot). \tag{1.2}$$

We call the expression $\sum_{k\in\mathbb{Z}} \widehat{f_k} e_k(\cdot)$ the *Fourier series* for f.

We have a similar notion of the *Fourier transform* $\widehat{f}$ for $f(\cdot) \in L^1(\mathbb{R}^d)$, $d \in \mathbb{N}$, defined by

$$\widehat{f}(\xi) = \frac{1}{\sqrt{2\pi}^d} \int_{\mathbb{R}^d} f(x) e^{i\xi\cdot x}\, dx.$$

This Fourier transform is then continuously extended so that it is defined on $L^2(\mathbb{R}^d)$ as well:

Theorem 1.2. [7] *There exists a unique map* $\mathcal{F} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ *such that*

(i) $\forall f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, $(\mathcal{F}f)(\cdot) = \widehat{f}(\cdot)$;

(ii) $\mathcal{F}$ *is sequentially continuous, i.e. if* $\lim_{j\to\infty} \|f_j(\cdot) - f(\cdot)\| = 0$ *then* $\lim_{j\to\infty} \|(\mathcal{F}f_j)(\cdot) - (\mathcal{F}f)(\cdot)\| = 0$.

If we suppress the $\mathcal{F}$ for the moment and again write $\widehat{f}(\cdot)$ for the Fourier transform of $f(\cdot)$, we have the reconstruction law that $\forall f(\cdot) \in L^2(\mathbb{R}^d)$,

$$f(x) = \frac{1}{\sqrt{2\pi}^d} \int_{\mathbb{R}^d} \widehat{f}(\xi) e^{-i\xi\cdot x}\, d\xi. \tag{1.3}$$

We have *Plancherel's equality* for $f(\cdot) \in L^2(\mathbb{R}^d)$:

$$\|f(\cdot)\| = \left\|\widehat{f}(\cdot)\right\|. \tag{1.4}$$

The uniqueness of the reconstruction formula (1.3) and Plancherel's equality (1.4) imply that the Fourier transform $\mathcal{F} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is an isometric isomorphism (modulo equality almost everywhere). It is customary to refer to the first copy of $\mathbb{R}^d$ as "space" and the second as "phase space". The two classic space-phase space pairs are time-frequency and position-momentum.

While the functions $x \mapsto e^{-i\xi\cdot x}$ are nice in many respects (all are C^∞, 2π-periodic, &c.), they have certain disadvantages, chief among these being that they do not admit simultaneous localization in both position x and frequency ξ, as illustrated by the Heisenberg Uncertainty Principle:

Theorem 1.3. (Heisenberg Uncertainty Principle.) [7] *Let* $f(\cdot) \in L^2(\mathbb{R}^d)$ *be such that* $\|f(\cdot)\| = \|\widehat{f}(\cdot)\| = 1$. *Define*

$$\begin{aligned} \operatorname{mean}\left(|f(\cdot)|^2\right) &:= \int_{\mathbb{R}^d} x|f(x)|^2\, dx \\ \operatorname{var}\left(|f(\cdot)|^2\right) &:= \int_{\mathbb{R}^d} \left|x - \operatorname{mean}\left(|f(\cdot)|^2\right)\right|^2 |f(x)|^2\, dx. \end{aligned}$$

Then

$$\operatorname{var}\left(|f(\cdot)|^2\right) \cdot \operatorname{var}\left(\left|\widehat{f}(\cdot)\right|^2\right) \geq \frac{d^2}{4} \tag{1.5}$$

The Heisenberg Uncertainty Principle is usually understood as applying to a quantum particle in $\mathbb{R}^3$: $|f(\cdot)|^2$ is the probability distribution of the particle's position in space; $|\widehat{f}(\cdot)|^2$ is the corresponding probability distribution for the particle's momentum, i.e. its position in phase space. The theorem tells us that we cannot simultaneously measure a quantum particle's position and momentum to arbitrary precision, and gives a lower bound for the product of the uncertainties in these two measurements. (The fact that position and momentum cannot be observed simultaneously can also be derived from the non-commutativity of their respective operators. This is, however, more of an algebraic viewpoint than an analytical one.)

2 Wavelets

What are wavelets? To the author, the word "wavelet" conjours up the image of a "small and lonely wave". Indeed, this is not too far from being a productive intuitive guide to the basic nature of wavelets. Roughly speaking, a wavelet is a function whose scaled translates span $L^2(\mathbb{R})$.

In [12], Meyer defines a wavelet $\psi(\cdot)$ of class $m \in \mathbb{N} \cup \{0\}$ to be a function of a real variable satisfying the four conditions of

(i) *regularity*: if $m = 0$, $\psi(\cdot) \in L^\infty(\mathbb{R})$; if $m \geq 1$, $\psi(\cdot), \psi'(\cdot), \ldots, \psi^{(m)}(\cdot) \in L^\infty(\mathbb{R})$;

(ii) *localization*: $|\psi(x)|, |\psi'(x)|, \ldots, |\psi^{(m)}(x)|$ decrease rapidly as $x \to \pm\infty$;

(iii) *zero moments*: $\int_{-\infty}^{\infty} x^k \psi(x)\, dx = 0$ for $0 \leq k \leq m$;

(iv) *orthonormal basis*: the collection of functions

$$\langle\langle \psi(\cdot) \rangle\rangle := \left\{ 2^{-j/2} \psi\left(2^{-j} \cdot -k\right) \middle|\, j, k \in \mathbb{Z} \right\}$$

is an orthonormal basis of $L^2(\mathbb{R})$.

The notation $\langle\langle \psi(\cdot) \rangle\rangle$ for the space of scaled translates of $\psi(\cdot)$ is not standard, but is intended to echo the traditional group-theoretic notation of $\langle S \rangle$ for the subgroup of a given group generated by elements of $S \subseteq G$. $\langle\langle \psi(\cdot) \rangle\rangle$ is a subspace of $L^2(\mathbb{R})$; Meyer's condition (iv) is that it has dense span, i.e. we are in the situation of Theorem 1.1.

Daubechies, [3], [5], gives the conditions that a wavelet $\psi(\cdot)$ should have

(i) *decay*: e.g. $|\psi(x)| \leq C_1 (1+|x|)^{-(1+\varepsilon_1)}$ for some $C_1, \varepsilon_1 \in \mathbb{R}$;

(ii) *smoothness*: as measured by the decay of the Fourier transform, e.g. $|\widehat{\psi}(\xi)| \leq C_2 (1+|\xi|)^{-(1+\varepsilon_2)}$ for some $C_2, \varepsilon_2 \in \mathbb{R}$;

(iii) *oscillation*: $\int_{-\infty}^{\infty} \psi(x)\, dx = 0$.

She does not require that $\langle\langle \psi(\cdot) \rangle\rangle$ should constitute an orthonormal basis for $L^2(\mathbb{R})$. While we shall encounter some wavelets that do not give rise to orthonormal bases of $L^2(\mathbb{R})$ we shall consider such wavelets to be somewhat "degenerate" and shall prefer those that do give rise to orthonormal bases.

Most definitions of wavelets have the common idea of simultaneous localization in both space and phase space. With such simultaneous localization, we can hope to circumvent Heisenberg's Uncertainty Principle (Theorem 1.3). This point merits a little further explanation to avoid confusion. We have the Fourier transform $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ and know that $\operatorname{var}|f(\cdot)|^2$ and $\operatorname{var}|(\mathcal{F}f)(\cdot)|^2$ cannot both simultaneously be small – this cannot be circumvented. The sense in which Heisenberg can be circumvented is that we can define a *(continuous) wavelet transform* $\mathcal{W}_\psi$ by

$$\psi^{a,b}(x) := |a|^{-1/2} \psi\left(\frac{x-b}{a}\right) \tag{2.1}$$

$$\begin{aligned}(\mathcal{W}_\psi f)(a,b) &:= \left\langle f(\cdot), \psi^{a,b}(\cdot) \right\rangle \\ &= |a|^{-1/2} \int_{-\infty}^{\infty} f(x) \overline{\psi\left(\frac{x-b}{a}\right)}\, dx\end{aligned}$$

where (a,b) takes values in $\mathbb{R}^* \times \mathbb{R}$, with $\mathbb{R}^* := \mathbb{R}\backslash\{0\}$. There is no *a priori* reason why we should not be able to make both $\operatorname{var}|f(\cdot)|^2$ and $\operatorname{var}|\left(\mathcal{W}_\psi f\right)(\cdot)|^2$ small. We do, of course, need to check that

$$\mathcal{W}_\psi : L^2(\mathbb{R}; dx) \to L^2(\mathbb{R}^* \times \mathbb{R}; a^{-2}\, da\, db)$$

is invertible. The reconstruction law is known as *Calderón's formula* [14]:

$$f(x) = C_\psi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\langle f(\cdot), \psi^{a,b}(\cdot) \right\rangle \psi^{a,b}(x) a^{-2}\, da\, db. \tag{2.2}$$

Rather than this continuous wavelet transform we are more often concerned with a discrete wavelet transform, which corresponds in some sense to the Fourier series. We consider wavelets

$$\psi_{j,k}(x) := a_0^{-j/2}\psi\left(a_0^{-j}x - kb_0\right) \tag{2.3}$$

for suitable fixed $a_0, b_0 \in \mathbb{R}$ and varying $j, k \in \mathbb{Z}$. In the terms used above in (2.1) to describe the continuous wavelet transform, $\psi_{j,k}(\cdot) = \psi^{a,b}(\cdot)$ with $a = a_0^j$, $b = kb_0a_0^j$. We seek an expansion

$$f(\cdot) = \sum_{j,k\in\mathbb{Z}} \langle f(\cdot), \psi_{j,k}(\cdot)\rangle\, \psi_{j,k}(\cdot). \tag{2.4}$$

Henceforth, we shall fix $a_0 = 2, b_0 = 1$ and use the notation of (2.3) for scaled translates of $\psi(\cdot)$ indexed by $j, k \in \mathbb{Z}$, as in Daubechies [3], [5]:

$$\psi_{j,k}(x) := 2^{-j/2}\psi\left(2^{-j}x - k\right). \tag{2.5}$$

By Theorem 1.1, the equality (2.4) will hold if and only if $\langle\langle\psi(\cdot)\rangle\rangle$ is an orthonomal set with $\overline{\operatorname{span}\langle\langle\psi(\cdot)\rangle\rangle} = L^2(\mathbb{R})$, i.e. $\langle\langle\psi(\cdot)\rangle\rangle$ is an orthonormal basis of $L^2(\mathbb{R})$.

Wavelet expansions do not give us everything that we could wish for. As we shall see, there is always some trading off of one or more attributes against one or more others. Such attributes include:

(i) the regularity of $\psi(\cdot)$ and $\widehat{\psi}(\cdot)$;

(ii) the size and compactness of $\operatorname{supp}\psi(\cdot)$ and $\operatorname{supp}\widehat{\psi}(\cdot)$;

(iii) whether or not the $\psi_{j,k}(\cdot)$ are orthonormal;

(iv) whether $\psi(\cdot)$ or $\widehat{\psi}(\cdot)$ have axes of symmetry or anti-symmetry;

(v) whether we can have an explicit formula for $\psi(\cdot)$.

However, there are a good many wavelet bases that give us most of what we want. From Daubechies [5]:

> Intuitively speaking, wavelet expansions do so well, in such a variety of frameworks, because their smoothness allows them to adjust well to smooth functions (or to smooth portions of functions), their scaling properties allow them to “zoom in” on singularities, and their good spatial concentration allows them to handle decay well.

3 Multiresolution Analysis

The theory of wavelet analysis is intimately connected with that of multiresolution analysis. We shall see in this section how a multiresolution analysis gives rise to a wavelet basis of $L^2(\mathbb{R})$. Later, as Theorem 6.1, we shall encounter Daubechies' converse construction.

The idea of multiresolution analysis (henceforth, MRA) is to build up $L^2(\mathbb{R})$ as a countable succession of subspaces, each allowing more "detail" than the previous one. Daubechies [3] defines a *multiresolution analysis of* $L^2(\mathbb{R})$ to be a family of closed nested subspaces $\{V_j \subset L^2(\mathbb{R}) | j \in \mathbb{Z}\}$, such that

(i) $\forall j \in \mathbb{Z}$, $V_j \subset V_{j-1}$;

(ii) $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$, $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$;

(iii) $f(\cdot) \in V_j \Leftrightarrow f(2^j \cdot) \in V_0$;

(iv) $\exists \phi(\cdot) \in V_0$ such that $V_j = \overline{\operatorname{span}\{\phi_{j,k}(\cdot) | k \in \mathbb{Z}\}}$;

(v) $\exists 0 < A \leq B < \infty$ such that $\forall (c_k)_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$,

$$A \left\| (c_k) \right\|^2_{\ell^2(\mathbb{Z})} \leq \left\| \sum_{k\in\mathbb{Z}} c_k \phi_{j,k}(\cdot) \right\|^2_{L^2(\mathbb{R})} \leq B \left\| (c_k) \right\|^2_{\ell^2(\mathbb{Z})}.$$

The function $\phi(\cdot)$ is called the *averaging function.* The last two requirements are sufficient for $\{\phi_{j,k}(\cdot) | k \in \mathbb{Z}\}$ to be a Riesz basis of $L^2(\mathbb{R})$. Note also that an immediate consequence of the scaling requirement (iii) of an MRA is that (iv) can be rewritten as

(iv)$'$ $\exists \phi(\cdot) \in V_0$ such that $V_0 = \overline{\operatorname{span}\{\phi_{0,k}(\cdot) | k \in \mathbb{Z}\}}$.

Just what are these spaces V_j? A brief consideration of the above axioms (i)–(v) leads to the formal (and not very helpful) result

$$V_j = \left\{ \sum_{s=j}^{\infty} \sum_{k=-\infty}^{\infty} c_{s,k} \phi_{s,k}(\cdot) \in L^2(\mathbb{R}) \,\middle|\, c_{s,k} \in \mathbb{R} \right\}.$$

A better way to think of V_j is that it is the subspace of $L^2(\mathbb{R})$ consisting of those functions that can be expressed using copies of ϕ that are at scale 2^j or bigger. So, for the Haar basis, for which $\phi(\cdot) = \chi_{[0,1)}(\cdot)$, V_j is the space of L^2 functions that are constant on intervals of length 2^j:

$$V_j = \left\{ f(\cdot) \in L^2(\mathbb{R}) \,\middle|\, \forall k \in \mathbb{Z}, f|_{[2^j k, 2^j(k+1))}(\cdot) \text{ is constant} \right\}.$$

Suppose we are given an MRA of $L^2(\mathbb{R})$; we wish to construct the associated orthonormal wavelet basis.

Proposition 3.1. *Given a multiresolution analysis* $\{V_j | j \in \mathbb{Z}\}$ *of* $L^2(\mathbb{R})$ *with associated averaging function* $\phi(\cdot)$*, there is an associated wavelet* $\psi(\cdot)$*, that we can construct from* $\phi(\cdot)$*, such that the set of scaled translates* $\langle\langle \psi(\cdot) \rangle\rangle$ *forms an orthonormal basis of* $L^2(\mathbb{R})$*.*

Proof. We first define a sequence of subspaces complementary to the V_j: let $W_j :=$ the orthogonal complement of V_j in V_{j-1}. So for each $j \in \mathbb{Z}$, $V_{j-1} = V_j \oplus W_j$ and $V_j \perp W_j$:

$$V_{j-1} = V_j \overset{\perp}{\oplus} W_j$$

From the scaling property of the V_j we get a scaling property for the W_j:

$$f(\cdot) \in W_j \Leftrightarrow f(2^j \cdot) \in W_0. \tag{3.1}$$

The W_j are thus orthogonal spaces summing to $L^2(\mathbb{R})$:

$$L^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} W_j \tag{3.2}$$

(Compare (3.2) with MRA axiom (ii).) Very roughly speaking, the $\phi_{j,k}(\cdot)$ "live in" the spaces V_j and the wavelets $\psi_{j,k}(\cdot)$ will "live in" the spaces W_j.

The MRA axioms imply that there is a $\psi(\cdot) \in W_0$ such that $W_0 = \overline{\operatorname{span}\{\psi_{0,k}(\cdot) | k \in \mathbb{Z}\}}$. (See Meyer [11] and Mallat [10].) So, by the scaling property for the W_j, (3.1), $W_j = \overline{\operatorname{span}\{\psi_{j,k}(\cdot) | k \in \mathbb{Z}\}}$.

At this point we need to resolve the technical difficulty that although the $\phi_{0,k}(\cdot)$ span V_0, they may not be linearly independent. If this is the case, define $\widetilde{\phi}(\cdot)$ by

$$\widehat{\left(\widetilde{\phi}\right)}(\xi) := C\widehat{\phi}(\xi) \left(\sum_{k \in \mathbb{Z}} \left| \widehat{\phi}(\xi + 2k\pi) \right|^2 \right)^{-1/2}. \tag{3.3}$$

Subject to the reasonable assumption that $\widehat{\phi}(\cdot)$ decays nicely, Meyer [11] shows:

Lemma 3.2. [11] *With $\widetilde{\phi}(\cdot)$ defined as above, the $\widetilde{\phi}_{0,k}$, $k \in \mathbb{Z}$, form an orthonormal basis of V_0, i.e.*

$$\overline{\operatorname{span}\{\,\phi_{0,k}(\cdot)|\,k \in \mathbb{Z}\}} = \overline{\operatorname{span}\left\{\,\widetilde{\phi}_{0,k}(\cdot)\middle|\,k \in \mathbb{Z}\right\}}$$

and

$$\left\langle \widetilde{\phi}_{0,k}(\cdot), \widetilde{\phi}_{0,k'}(\cdot) \right\rangle = \delta_{kk'}.$$

Armed with Lemma 3.2, we may assume that $\overline{\text{the } \phi_{0,k}(\cdot) \text{ form an or-}}$ thonormal basis for V_0. Since $\phi(\cdot) \in V_0 \subset V_{-1} = \overline{\operatorname{span}\{\phi(2\cdot -k)|k \in \mathbb{Z}\}}$, $\exists (c_k)_{k\in\mathbb{Z}}$ such that

$$\phi(x) = \sum_{k\in\mathbb{Z}} c_k \phi(2x-k). \tag{3.4}$$

Define

$$\psi(x) := \sum_{k\in\mathbb{Z}} (-1)^k c_{k+1} \phi(2x+k). \tag{3.5}$$

Then the collection $\{\psi_{0,k}(\cdot)|k \in \mathbb{Z}\}$ forms an orthonormal basis for W_0. The corresponding collection $\{\psi_{j,k}(\cdot)|k \in \mathbb{Z}\}$ forms an orthonormal basis for W_j, and so $\langle\langle\psi(\cdot)\rangle\rangle = \{\psi_{j,k}(\cdot)|j,k \in \mathbb{Z}\}$ forms an orthonormal basis of wavelets for $L^2(\mathbb{R})$. ☐

4 MRA, Wavelet Decomposition and The Laplacian Pyramid

We define the sequence of projection maps $P_j : L^2(\mathbb{R}) \to L^2(\mathbb{R})$

$$P_j(\cdot) := \operatorname{Proj}_{V_j}(\cdot) \tag{4.1}$$

We have seen how an MRA of $L^2(\mathbb{R})$ naturally gives rise to an orthonormal basis of wavelets; we shall see later that for suitable wavelets the converse holds. The layered structure of an MRA lends itself to a fast algorithm for the decomposition of a function into wavelets. We start with a function $f(\cdot)$ given to finite resolution – since we can change our unit length to whatever is convenient, we shall assume we have $P_0 f(\cdot)$. We suppose we know an initial expansion

$$P_0 f(\cdot) = \sum_{k \in \mathbb{Z}} s_{0,k} \phi_{0,k}(\cdot), \tag{4.2}$$

where $s_{0,k} := \langle f(\cdot), \phi_{0,k}(\cdot) \rangle$. Suppose we have N non-zero $s_{0,k}$.

Burt and Adelson [1] published an ingenious application of MRA in 1983, decomposing a signal and compressing it using a scheme known as a *Laplacian pyramid.* They described their method as

> ... a technique for image encoding in which local operators of many scales but identical shape serve as the basis functions.

The connection with MRA is obvious. (Since they were concerned with image encoding, Burt and Adelson were, of course, working in $L^2(\mathbb{R}^2)$ rather that $L^2(\mathbb{R})$.) We quickly recall Shannon's Sampling Theorem:

Theorem 4.1. (Shannon's Sampling Theorem.) *Let* $f(\cdot) \in L^2(\mathbb{R})$. *Let*

$$\omega_0 := \inf \left\{ \omega \in \mathbb{R} \,\middle|\, \overline{\operatorname{supp} \widehat{f}(\cdot)} \subseteq [-\omega, \omega] \right\},$$

and suppose that $\omega_0 < \infty$, *i.e.* $f(\cdot)$ *has limited bandwidth* ω_0. *Then*

$$f(x) = \sum_{n \in \mathbb{Z}} f\left(\frac{2\pi}{\omega_0} n\right) \frac{\sin(\omega_0 x/2 - \pi n)}{\omega_0 x/2 - \pi n}.$$

The important point to take from Shannon's Sampling Theorem is that a signal of lower bandwidth needs to be sampled less often to guarantee perfect reconstruction.

Burt and Adelson's idea is remarkably simple. Suppose we have an MRA of $L^2(\mathbb{R})$ and a limited bandwidth signal $f(\cdot) \in L^2(\mathbb{R})$. Since $f(\cdot)$ has limited bandwidth, there is a $j_0 \in \mathbb{Z}$ such that $f(\cdot) = P_{j_0} f(\cdot)$. From Daubechies [6]:

> In the Laplacian pyramid scheme [...] you compute a blurred version of the picture, and you subtract it from the original; the difference gives you the desired fine scale features which you break up into tiny "elementary" pieces. Successive layers of decreasing spatial resolution are obtained by repeating the procedure on the blurred picture.

We start this procedure by stripping away the layer containing detail coarser than scale 2^{j_0}. At each stage we have a "blurred layer" B and a "detail layer" D. Because B necessarily has a lower bandwidth than D, it need not be sampled as often to guarantee perfect reconstruction.

((Continue as understood.))

5 Splines and the Haar Basis

Polynomial functions are exceedingly nice objects to work with. Accordingly, a nice class of wavelets consists of suitable *spline functions.* Intuitively, a spline function is made of polynomial pieces that join up smoothly. Formally, a function $f(\cdot)$ defined on $\mathbb{R}$ is called a *spline of order* $r \in \mathbb{N} \cup \{0\}$ if

- (i) $\exists$ a finite collection of disjoint intervals[1] $\{I_j \subseteq \mathbb{R} | j = 1, \ldots, N\}$ such that
  - (a) $\mathbb{R} = \coprod_{j=1}^{N} I_j$;
  - (b) $\forall j \in \mathbb{N}$, $f|_{I_j}(\cdot)$ is a polynomial of order r;
- (ii) if $r \neq 0$, $f(\cdot) \in C^{r-1}(\mathbb{R})$.

A rather nice sequence of splines is the sequence of B-splines: the uniform B-spline blending function $\sigma_r(\cdot)$ of order $r \in \mathbb{N} \cup \{0\}$ is defined recursively by

$$\sigma_r(\cdot) := \begin{cases} \chi_{[0,1)}(\cdot) & \text{for } r = 0; \\ (\sigma_{r-1} * \sigma_0)(\cdot) & \text{for } r > 0, \end{cases}$$

where $*$ denotes the convolution operator defined by

$$(f * g)(t) := \int_{\mathbb{R}} f(\tau) g(t - \tau) \, d\tau.$$

One of the earliest known wavelet bases of $L^2(\mathbb{R})$ is that given by the *Haar wavelet* $\psi_{\text{Haar}}(\cdot)$:

$$\psi_{\text{Haar}}(x) := \begin{cases} 1 & \text{for } x \in [0, \frac{1}{2}); \\ -1 & \text{for } x \in [\frac{1}{2}, 1); \\ 0 & \text{otherwise.} \end{cases}$$

---

[1] We allow infinte intervals such as $(-\infty, a)$ or $[b, \infty)$.

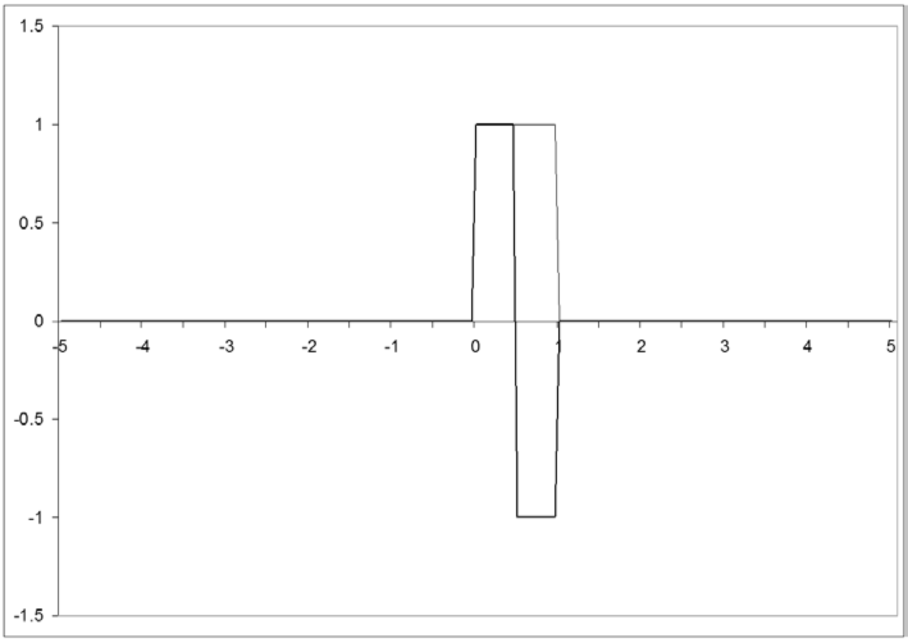

Figure 1: The Haar wavelet $\psi_{\text{Haar}}(\cdot)$ (heavy line) and averaging function $\sigma_0(\cdot) = \phi_{\text{Haar}}(\cdot)$ (light line).

It was known as early as 1910 (see Haar [8]) that $\langle\langle\psi_{\text{Haar}}(\cdot)\rangle\rangle$ forms an orthonormal basis of $L^2(\mathbb{R})$ – the *Haar basis.* The associated averaging function, $\phi_{\text{Haar}}(\cdot)$, is simply $\chi_{[0,1)}(\cdot)$. Observe that $\sigma_0(\cdot) = \phi_{\text{Haar}}(\cdot)$. See Figure 1 for graphs of $\psi_{\text{Haar}}(\cdot)$ and $\phi_{\text{Haar}}(\cdot)$.

In a later section we shall go through several worked examples using the Haar basis.

The linear spline function $\phi(\cdot) := \sigma_1(\cdot) \in C^0(\mathbb{R})$ is

$$\phi(x) := \begin{cases} x & \text{for } x \in [0,1); \\ 2-x & \text{for } x \in [1,2); \\ 0 & \text{otherwise.} \end{cases}$$

This averaging function $\phi(\cdot)$ satisfies

$$\phi(x) = \frac{1}{2}\phi(2x) + \phi(2x-1) + \frac{1}{2}\phi(2x-2)$$

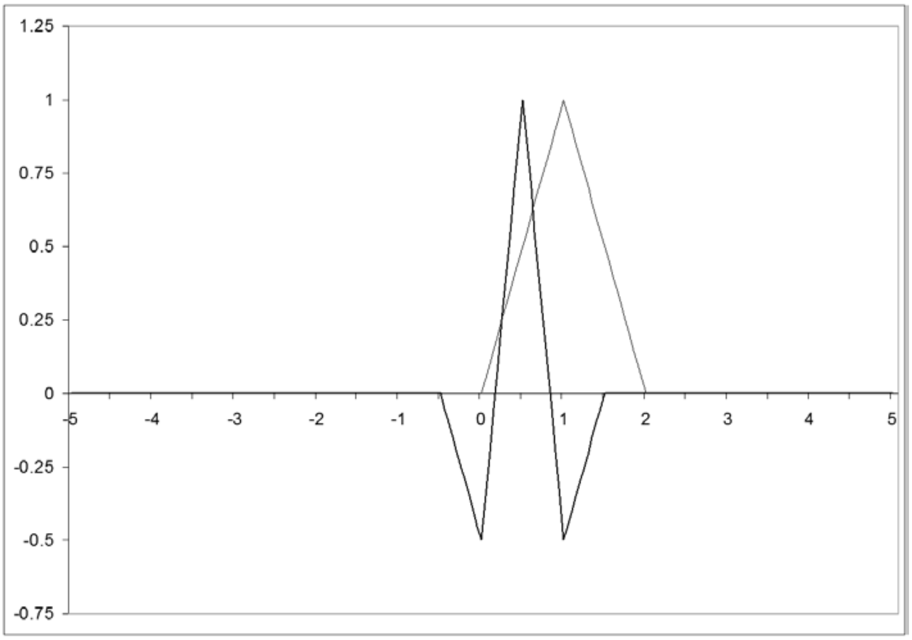

Figure 2: The linear spline wavelet (heavy line) and averaging function $\sigma_1(\cdot)$ (light line).

and so, by (3.5), gives rise to the wavelet

$$\begin{aligned}\psi(x) &:= -\frac{1}{2}\phi(2x-1) + \phi(2x) - \frac{1}{2}\phi(2x+2)\\ &= \begin{cases} -x-\frac{1}{2} & \text{for } x \in [-\frac{1}{2}, 0);\\ 3x-\frac{1}{2} & \text{for } x \in [0, \frac{1}{2});\\ -3x+\frac{5}{2} & \text{for } x \in [\frac{1}{2}, 1);\\ x-\frac{3}{2} & \text{for } x \in [1, \frac{3}{2});\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

The quadratic spline function $\phi(\cdot) := \sigma_2(\cdot) \in C^1(\mathbb{R})$ is given by

$$\phi(x) := \begin{cases} \frac{1}{2}x^2 & \text{for } x \in [0, 1);\\ \frac{1}{2}(-2x^2+6x-3) & \text{for } x \in [1, 2);\\ \frac{1}{2}(3-x)^2 & \text{for } x \in [2, 3);\\ 0 & \text{otherwise.}\end{cases}$$

This averaging function $\phi(\cdot)$ satisfies

$$\phi(x) = \frac{1}{4}\phi(2x) + \frac{3}{4}\phi(2x-1) + \frac{3}{4}\phi(2x-2) + \frac{1}{4}\phi(2x-3)$$

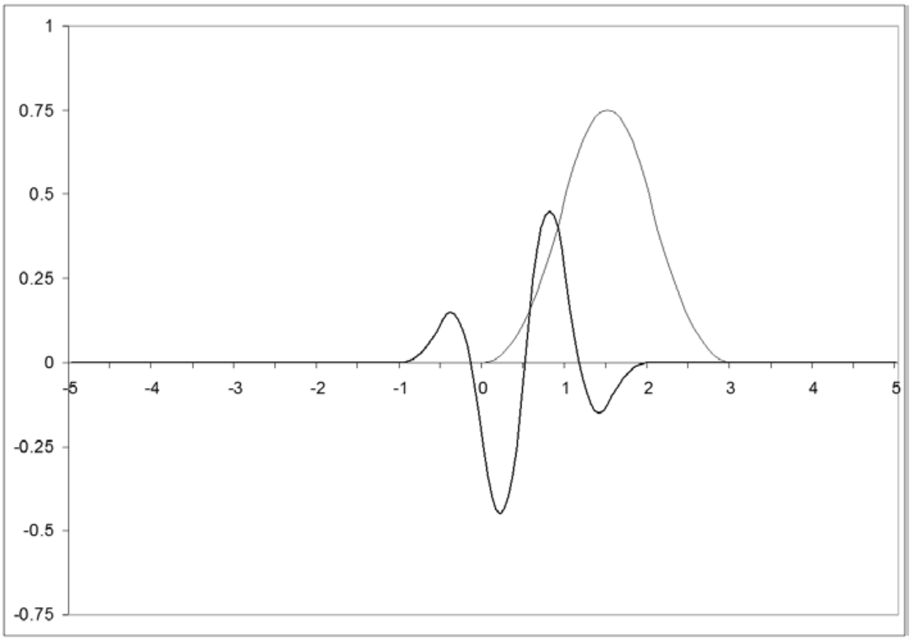

Figure 3: The quadratic spline wavelet (heavy line) and averaging function $\sigma_2(\cdot)$ (light line).

and so, by (3.5), gives rise to the wavelet

$$\psi(x) := -\frac{1}{4}\phi(2x-1) + \frac{3}{4}\phi(2x) - \frac{3}{4}\phi(2x+1) + \frac{1}{4}\phi(2x+2).$$

To the human eye, the graphs of the higher-order spline functions $\sigma_3(\cdot)$, $\sigma_4(\cdot), \ldots$ look just like that of $\sigma_2(\cdot)$, except with wider support and a lower peak value. The additional regularity is not readily visible in the graphs of the averaging functions $\phi(\cdot)$ or the associated wavelets $\psi(\cdot)$.

In the notation of (3.4) and (3.5), suppose that $\operatorname{supp}\phi(\cdot) \subseteq [a,b] \subset \mathbb{R}$ and that $c_k = 0$ for $k < K^-$ and $k > K^+$, with $K^+, K^- \in \mathbb{Z}$. Then since $\operatorname{supp}\phi(2\cdot -k) \subseteq \left[\frac{a-k}{2}, \frac{b-k}{2}\right]$,

$$\operatorname{supp}\psi(\cdot) \subseteq \left[\frac{a-K^+}{2}, \frac{b-K^-}{2}\right]. \tag{5.1}$$

Unfortunately, for $r > 1$, $\langle\langle\sigma_r(\cdot)\rangle\rangle$ is not an orthonormal wavelet basis of $L^2(\mathbb{R})$, and although one can apply an orthogonalization procedure (such as

Meyer's one, Lemma 3.2) to obtain such a basis, one loses the compactness of $\operatorname{supp}\psi(\cdot)$ in the process. We shall see later that the Haar basis is the only orthonormal basis of compactly supported wavelets for which the associated averaging function $\phi(\cdot)$ has an axis of symmetry.

6 Daubechies' Construction

We have already seen how we can pick an MRA of $L^2(\mathbb{R})$ as our starting point and construct an associated orthonormal wavelet basis (Proposition 3.1). Daubechies [3] provides the following theorem for the construction of an orthonormal wavelet basis and associated MRA of $L^2(\mathbb{R})$ from a sequence $(h_n)_{n\in\mathbb{Z}}$ with suitably rapid decay:

Theorem 6.1. *Let $(h_n)_{n\in\mathbb{Z}}$ be a sequence such that*

(i) for some $\varepsilon > 0$, $\sum_{n\in\mathbb{Z}} |h_n||n|^\varepsilon < \infty$;

(ii) $\sum_{n\in\mathbb{Z}} h_{n-2k}h_{n-2\ell} = \delta_{k\ell}$;

(iii) $\sum_{n\in\mathbb{Z}} h_n = \sqrt{2}$.

Also $m_0(\xi) := \frac{1}{\sqrt{2}}\sum_{n\in\mathbb{Z}} h_n e^{in\xi}$ can be written as

$$m_0(\xi) = \left(\frac{1}{2}\left(1 + e^{i\xi}\right)\right)^N \left(\sum_{n\in\mathbb{Z}} f_n e^{in\xi}\right)$$

where

(iv) for some $\varepsilon > 0$, $\sum_{n\in\mathbb{Z}} |f_n||n|^\varepsilon < \infty$;

(v) $\sup_{\xi\in\mathbb{R}} \left|\sum_{n\in\mathbb{Z}} f_n e^{in\xi}\right| < 2^{N-1}$.

Define

$$\begin{aligned} g_n &:= (-1)^n h_{-n+1} \\ \widehat{\phi}(\xi) &:= \frac{1}{\sqrt{2\pi}} \prod_{j=1}^{\infty} m_0\left(2^{-j}\xi\right) \\ \psi(x) &:= \sqrt{2}\sum_{n\in\mathbb{Z}} g_n \phi(2x - n) \end{aligned}$$

Then the $\phi_{j,k}(\cdot)$ define a multiresolution analysis of $L^2(\mathbb{R})$ and the $\psi_{j,k}(\cdot)$ are the associated orthonormal wavelet basis.

Proof. We shall sketch Daubechies' proof. Having made the above definitions of $(g_n)_{n\in\mathbb{Z}}$, $m_0(\cdot)$, $\widehat{\phi}(\cdot)$ and $\psi(\cdot)$, Daubechies first observes that $\phi(\cdot)$ is bounded and uniformly continuous, and that $\widehat{\phi}(\cdot) \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, therefore $\phi(\cdot) \in L^2(\mathbb{R})$. It is easy to see that

$$\sum_{n\in\mathbb{Z}} |g_n| = \sum_{n\in\mathbb{Z}} |h_n| < \infty,$$

and hence

$$\left|\widehat{\psi}(\xi)\right| \le \frac{1}{\sqrt{2}} \sum_{n\in\mathbb{Z}} |h_n| \left|\widehat{\phi}\left(\frac{\xi}{2}\right)\right|.$$

Thus $\psi(\cdot)$ is also bounded, uniformly continuous and L^2. The definitions of $\widehat{\phi}(\cdot)$ and $\psi(\cdot)$ in the statement of the theorem immediately give the relations

$$\phi_{j,k}(\cdot) = \sum_{n\in\mathbb{Z}} h_{n-2k}\phi_{j-1,n}(\cdot)$$

and

$$\psi_{j,k}(\cdot) = \sum_{n\in\mathbb{Z}} g_{n-2k}\phi_{j-1,n}(\cdot).$$

The next step is to prove the following orthogonality relations, which are restricted to one scale at a time – we shall need another lemma to show orthogonality between different scales.

Lemma 6.2. *Let (h_n), (g_n), $\phi(\cdot)$ and $\psi(\cdot)$ be as above. Then $\phi(\cdot), \psi(\cdot) \in L^2(\mathbb{R})$ and, for all $j, k, k' \in \mathbb{Z}$,*

$$\begin{aligned}\langle \psi_{j,k}(\cdot), \psi_{j,k'}(\cdot)\rangle &= \delta_{kk'},\\ \langle \psi_{j,k}(\cdot), \phi_{j,k'}(\cdot)\rangle &= 0,\\ \langle \phi_{j,k}(\cdot), \phi_{j,k'}(\cdot)\rangle &= \delta_{kk'}.\end{aligned}$$

Proof. Sketch proof? □

The next step is an analogue of Parseval's Equality (1.1), which also shows the orthogonality between different scales that Lemma 6.2 omitted:

Lemma 6.3. *With the same hypotheses as Lemma 6.2, for all $f(\cdot) \in L^2(\mathbb{R})$,*

$$\|(\langle \psi_{j,k}(\cdot), f(\cdot)\rangle)\|^2_{\ell^2(\mathbb{Z}^2)} := \sum_{j,k\in\mathbb{Z}} |\langle \psi_{j,k}(\cdot), f(\cdot)\rangle|^2 = \|f(\cdot)\|^2.$$

Proof. Sketch proof. ☐

This completes the proof of Theorem 6.1. ☐

Theorem 6.1 is the key to Daubechies' construction in [3] of an orthonormal wavelet basis of $L^2(\mathbb{R})$ where the mother wavelet $\psi(\cdot)$ has compact support. Suppose $(h_n)_{n\in\mathbb{Z}}$ has only finitely many non-zero terms. That is, if we define $N^+ := \max\{n|h_n \neq 0\}$ and $N^- := \min\{n|h_n \neq 0\}$, suppose $N^+ - N^- < \infty$. Then the corresponding wavelet has compact support. We construct $\phi(\cdot)$ as the limit of a sequence of recursively defined approximations:

$$\phi(x) := \lim_{\ell\to\infty} \eta_\ell(x)$$

where

$$\eta_\ell(x) := \begin{cases} \chi_{[-1/2,1/2)}(x) & \text{for } \ell = 0; \\ \sqrt{2}\sum_{n\in\mathbb{Z}} h_n \eta_{\ell-1}(2x-n) & \text{for } \ell > 0. \end{cases}$$

Then $\operatorname{supp} \eta_\ell(\cdot) \subset \left[N_\ell^-, N_\ell^+\right]$, where N_ℓ^+, N_ℓ^- satisfy the recursions

$$N_\ell^+ := \begin{cases} 1/2 & \text{for } \ell = 0; \\ (N_{\ell-1}^+ + N^+)/2 & \text{for} \ell > 0; \end{cases}$$
$$N_\ell^- := \begin{cases} -1/2 & \text{for } \ell = 0; \\ (N_{\ell-1}^- + N^-)/2 & \text{for} \ell > 0. \end{cases}$$

It is easy to see that as $\ell \to \infty$, $N_\ell^+ \to N^+$ and $N_\ell^- \to N^-$. Hence $\phi(\cdot)$ has compact support, contained within $[N^-, N^+]$. $\psi(\cdot)$ also has compact support with

$$\operatorname{supp} \psi(\cdot) \subseteq \left[\frac{1 - N^+ - N^-}{2}, \frac{1 + N^+ - N^-}{2}\right]$$

Daubechies' compactly supported wavelet of [3] is particularly well-known, and is given by

$$h_n := \begin{cases} \frac{1\mp\sqrt{3}}{4\sqrt{2}} & \text{for } n = 0; \\ \frac{3\mp\sqrt{3}}{4\sqrt{2}} & \text{for } n = 1; \\ \frac{3\pm\sqrt{3}}{4\sqrt{2}} & \text{for } n = 2; \\ \frac{1\pm\sqrt{3}}{4\sqrt{2}} & \text{for } n = 3; \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}$$

((Diagrams of Daubechies' wavelet and averaging function.))

In stark contrast to the high regularity (C^∞) and symmetry of the Fourier basis functions $e_k(x) := e^{ikx}$ the Daubechies wavelet appears to be very rough and non-symmetric. The irregularity problem can be overcome. Simply by increasing the number of non-zero h_n, the resulting wavelet can be made C^k for arbitrarily high (but still finite) k, at the cost of a linear increase in support width, while maintaining the compactness of the support. Daubechies [4] gives and proves the following result:

Proposition 6.4. *If a C^k function $\phi(\cdot)$ is such that*

$$\phi(x) = \sum_{n=0}^{N} c_n \phi(2x - n)$$

and $\operatorname{supp} \phi(\cdot) \subseteq [0, N]$ *then $k \leq N - 2$.*

The non-symmetry problem, however, is insurmountable. Daubechies [3] provides the following remarkable result:

Proposition 6.5. *The Haar basis is the only orthonormal basis of compactly supported wavelets for which the associated averaging function $\phi(\cdot)$ has a symmetry axis.*

Proof. The strategy of this proof is to take an orthonormal compactly-supported wavelet basis, assume the existence of a symmetry axis for the averaging function, and show that we must necessarily have been working with the Haar basis.

Recall our definitions from Daubechies wavelet construction, Theorem 6.1: $\phi(\cdot)$ is the function with Fourier transform $\widehat{\phi}(\cdot)$ such that

$$\widehat{\phi}(\xi) = \frac{1}{\sqrt{2\pi}} \prod_{j=1}^{\infty} m_0\left(2^{-j}\xi\right)$$

where

$$m_0(\xi) := \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} h_n e^{in\xi}.$$

We define

$$\begin{aligned} a_n &:= h_{2n}, & \alpha(\xi) &:= \textstyle\sum_{n\in\mathbb{Z}} a_n e^{in\xi}, \\ b_n &:= h_{2n+1}, & \beta(\xi) &:= \textstyle\sum_{n\in\mathbb{Z}} b_n e^{in\xi}, \\ c_n &:= g_{2n}, & \gamma(\xi) &:= \textstyle\sum_{n\in\mathbb{Z}} c_n e^{in\xi}, \\ d_n &:= g_{2n+1}, & \delta(\xi) &:= \textstyle\sum_{n\in\mathbb{Z}} d_n e^{in\xi}. \end{aligned}$$

For a trigonometric polynomial $P(\xi) = \sum_{n\in\mathbb{Z}} p_n e^{in\xi}$ define

$$\begin{aligned} N^+(P) &:= \max\{n | p_n \neq 0\}, \\ N^-(P) &:= \min\{n | p_n \neq 0\}. \end{aligned}$$

By inspection, we see that for any P,

$$N^+\left(|P|^2\right) = -N^-\left(|P|^2\right) = N^+(P) - N^-(P).$$

We know by 3.14 that $|\alpha(\xi)|^2 + |\beta(\xi)|^2 = 1$ and by 3.15 that $\alpha \neq 0$, $\beta \neq 0$. So we have

$$N^+(\alpha) - N^-(\alpha) = N^+(\beta) - N^-(\beta).$$

3.7 implies that

$$\begin{aligned} N^+(m_0) &= \max\{2N^+(\alpha), 2N^+(\beta) + 1\}, \\ N^-(m_0) &= \min\{2N^-(\alpha), 2N^-(\beta) + 1\}. \end{aligned}$$

So $N^+(m_0) - N^-(m_0) = \max\{2N^+(\alpha) - 2N^-(\beta) - 1, 2N^+(\beta) - 2N^-(\alpha) + 1\}$ is an odd number. $(*)$

Now suppose that $\phi(\cdot)$ were symmetric about zero, i.e. $\phi(x) = \phi(-x)$. It would follow that $h_n = h_{-n}$ and so that $N^+(m_0) = -N^-(m_0)$. So $N^+(m_0) - N^-(m_0) = 2N^+(m_0)$, an even number. This contradicts observation $(*)$, so $\phi(\cdot)$ cannot be symmetric about 0.

What about symmetry about $\lambda \in \mathbb{R}\backslash\{0\}$, i.e. $\phi(\lambda + x) = \phi(\lambda - x)$? Without loss of generality, shift by an integer so that λ lies in the interval $(0, 1)$. Then $\widehat{\phi}(\xi) = e^{2i\lambda\xi}\widehat{\phi}(-\xi)$, and so $m_0(\xi) = e^{2i\lambda\xi} m_0(-\xi)$. $m_0(\xi)$ and $m_0(-\xi)$ are both trigonometric polynomials, and so $\lambda = 1/2$. Thus $\phi(x+1) = \phi(-x)$, so $h_{2n+1} = h_{-2n}$, so $b_n = a_{-n}$, so $\beta(\xi) = \overline{\alpha(\xi)}$. We now have that $|\alpha(\xi)|^2 = 1/2$ and so $a_n = \pm\frac{1}{\sqrt{2}}\delta_{nk} = b_{-n}$ for some $k \in \mathbb{N}$. Since we may freely translate by integers, assume $k = 0$. We then obtain precisely the h_n that correspond to the Haar basis: $h_0 = h_1 = \frac{1}{\sqrt{2}}$, all others zero. So we have the Haar basis, with $\phi(\cdot) = \chi_{[0,1)}(\cdot)$. $\square$

7 A Sketch of 2-D MRA

Suppose we are given an MRA $\{V_j | j \in \mathbb{Z}\}$ of $L^2(\mathbb{R})$ with associated averaging function $\phi(\cdot)$. We can then define a multiresolution analysis of $L^2(\mathbb{R}^2)$ by setting $\mathbf{V}_j := V_j \oplus V_j$ and $\Phi(x_1, x_2) := \phi(x_1)\phi(x_2)$. Then

(i) $\forall j \in \mathbb{Z}$, $\mathbf{V}_j \subset \mathbf{V}_{j-1}$;

(ii) $\overline{\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j} = L^2(\mathbb{R}^2)$, $\bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$;

(iii) $f(\cdot) \in \mathbf{V}_j \Leftrightarrow f(2^j \cdot) \in \mathbf{V}_0$;

(iv) $\mathbf{V}_0 = \overline{\text{span}\{\Phi_{0,k}(\cdot,\cdot) | k \in \mathbb{Z}^2\}}$;

(v) $\exists 0 < A \leq B < \infty$ such that $\forall (c_k)_{k\in\mathbb{Z}^2} \in \ell^2(\mathbb{Z}^2)$,

$$A \left\|(c_k)\right\|^2_{\ell^2(\mathbb{Z}^2)} \leq \left\| \sum_{k\in\mathbb{Z}^2} c_k \Phi_{j,k}(\cdot,\cdot) \right\|^2_{L^2(\mathbb{R}^2)} \leq B \left\|(c_k)\right\|^2_{\ell^2(\mathbb{Z}^2)}.$$

The definition of the scaled translates is worthy of note, since we use the same scale j but different translations k_1, k_2:

$$\Phi_{j,k}(x_1, x_2) := 2^{-j}(2^{-j}x_1 - k_1, 2^{-j}x_2 - k_2) = \phi_{j,k_1}(x_1)\phi_{j,k_2}(x_2). \tag{7.1}$$

If we define, as before, W_j to be the orthogonal complement of V_j in V_{j-1} we can see that

$$\mathbf{V}_{j-1} = \mathbf{V}_j \oplus \left[(V_j \oplus W_j) \oplus (W_j \oplus V_j) \oplus (W_j \oplus W_j)\right].$$

Hence, an orthonormal basis for $\mathbf{W}_j$, the orthogonal complement of $\mathbf{V}_j$ in $\mathbf{V}_{j-1}$, is given by

$$\left\{ \phi_{j,k_1}(\cdot)\psi_{j,k_2}(\cdot), \psi_{j,k_1}(\cdot)\phi_{j,k_2}(\cdot), \psi_{j,k_1}(\cdot)\psi_{j,k_2}(\cdot) | \, (k_1, k_2) \in \mathbb{Z}^2 \right\}.$$

Equivalently, such a basis is given by the collection of 2-D wavelets $\Psi^\ell_{j,k}(\cdot,\cdot)$, where

$$\Psi^\ell_{j,k}(x_1, x_2) := 2^{-j}\Psi^{ell}(2^{-j}x_1 - k_1, 2^{-j}x_2 - k_2)$$

and

$$\begin{aligned}
\Psi^1(x_1, x_2) &:= \phi(x_1)\psi(x_2); \\
\Psi^2(x_1, x_2) &:= \psi(x_1)\phi(x_2); \\
\Psi^3(x_1, x_2) &:= \psi(x_1)\psi(x_2).
\end{aligned}$$

Thus, the collection

$$\left\{ \Psi^{\ell}_{j,k}(\cdot,\cdot) \middle| j \in \mathbb{Z}, k \in \mathbb{Z}^2, \ell = 1,2,3 \right\}$$

forms an orthonormal basis of 2-dimensional wavelets for $L^2(\mathbb{R}^2)$.

A similar construction for $L^2(\mathbb{R}^d)$ would yield an orthonormal basis of n-dimensional wavelets of the form

$$\left\{ \Psi^{\ell}_{j,k}(\cdot,\ldots,\cdot) \middle| j \in \mathbb{Z}, k \in \mathbb{Z}^d, 1 \le \ell < 2^d \right\}.$$

8 The Haar Basis Revisited

Recall the definition of the Haar wavelet:

$$\psi_{\text{Haar}}(x) = \psi(x) := \begin{cases} 1 & \text{for } x \in [0, \frac{1}{2}); \\ -1 & \text{for } x \in [\frac{1}{2}, 1); \\ 0 & \text{otherwise,} \end{cases}$$

with its associated averaging function $\phi(\cdot) = \chi_{[0,1)}(\cdot)$.

Let $f(\cdot) \in L^2(\mathbb{R})$. Let us compute its Haar coefficients:

$$\begin{aligned} \langle f(\cdot), \psi_{j,k}(\cdot) \rangle &:= \int_{\mathbb{R}} f(x) \overline{\psi_{j,k}(x)}\, dx \\ &= \int_{\mathbb{R}} f(x) 2^{-j/2} \psi(2^{-j}x - k)\, dx. \end{aligned}$$

At this point it helps to see that

$$\psi(2^{-j}x - k) = \begin{cases} 0 & \text{for } x < k2^j; \\ 1 & \text{for } k2^j \le x < (k+\frac{1}{2})2^j; \\ -1 & \text{for } (k+\frac{1}{2})2^j \le x < (k+1)2^j; \\ 0 & \text{for } x \ge (k+1)2^j. \end{cases}$$

Thus,

$$\langle f(\cdot), \psi_{j,k}(\cdot) \rangle = 2^{-j/2} \int_{k2^j}^{(k+\frac{1}{2})2^j} f(x)\, dx - 2^{-j/2} \int_{(k+\frac{1}{2})2^j}^{(k+1)2^j} f(x)\, dx$$

which, if $F(\cdot)$ is an anti-derivative for $f(\cdot)$, we can write as

$$= 2^{-j/2} \left(2F((k+1/2)2^j) - F((k+1)2^j) - F(k2^j) \right).$$

References

[1] Burt, P., & Adelson, E.H. The Laplacian Pyramid as a Compact Image Code. *IEEE Trans. Communications*, **31**, 482–540. 1983.

[2] Casazza, P., Christensen, O., Li, S., Lindner, A. On Riesz Fischer sequences and lower frame bounds. *Z. Anal. Anwend.*, **21** no. 2 (2002), 305-314.

[3] Daubechies, I. Orthonormal Bases of Compactly Suported Wavelets. *Comm. Pure & Appl. Math.*, **41**, 909–996. 1988.

[4] Daubechies, I., and Lagarias, J.C. Two-scale difference equations. I. Existence and global regularity of solutions. *SIAM J. Math. Anal.*, **22**, 1388–1410, 1991. —. II. Local regularity, infinite products of matrices, and fractals. *SIAM J. Math. Anal.*, **23**, pp. 1031–1079, 1992.

[5] Daubechies, I. Wavelets and Other Phase Space Localization Methods. *Proceedings of the International Congress of Mathematicians*, Zürich, Switzerland, 1994, Birkhäuser Verlag, Basel Switzerland, 1995.

[6] Daubechies, I. Where do Wavelets Come From? – A Personal Point of View. *Proceedings of the IEEE Special Issue on Wavelets*, **84** (no. 4), 510–513, April 1996.

[7] Friesecke, G. Lectures for *MA433 Fourier Analysis*, University of Warwick, 2003.

[8] Haar, A. Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.* **69**, 331–371, 1910.

[9] Kirillov, A.A. *Elementy Teorii Predstavleniĭ.* Nauka, Moscow. 1972. English translation: *Elements of the Theory of Representations.* Trans. Hewitt, E. Springer-Verlag. 1975.

[10] Mallat, S. Multiresolution Approximation and Wavelet Orthonormal Bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, **315**, 69–88.

[11] Meyer, Y. Ondelettes et Functions Splines, *Séminaire EDP*, Ecole Polytechnique, Paris, France, December, 1986.

[12] Meyer, Y. *Ondelettes et Opérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III: Opérateurs multilinéaires.* Hermann, Paris. 1990. English translation: Cambridge University Press. 1992.

[13] Pritchard, A.J. Lectures for *MA248 Applied Analysis*, University of Warwick, 2002.

[14] Weisstein, E.W. *http://mathworld.wolfram.com*. Wolfram Research / CRC Press LLC, 1999.