Some Problems Related to Wavelet Packet Bases and Convergence

Mladen Victor Wickerhauser*

Department of Mathematics
Washington University
St. Louis, Missouri 63130 USA

February 20, 2003

Abstract

Wavelet packets are subsets of a multiresolution analysis and derive many of their properties therefrom. Those defined by a single filter pair have uncontrolled size and basis properties, in general. By substituting different filters at different scales according to a rule, these can be controlled. The number of orthonormal bases available in an MRA satisfies a recursion equation depending on the basis selection method, and some of these recursions have closed form solutions. Some of these orthonormal bases consist of uniformly bounded, uniformly compactly supported wavelet packets and are Schauder bases for many Banach spaces. With controlled size and support, the Carleson–Hunt theorem applies to show that a wavelet packet Fourier series of a continuous function converges pointwise almost everywhere.

1 Approximation With Refinable Functions

A sequence of samples $s = \{s(n) : n \in \mathbf{Z}\}$ may be regarded as a piecewise constant function:

$$f_s(x) \stackrel{\text{def}}{=} \sum_{n\in\mathbf{Z}} s(n)\mathbf{1}(x-n), \qquad t \in \mathbf{R},$$

where $\mathbf{1}$ is the indicator function of the interval $[0,1) \subset \mathbf{R}$. If s is square-summable, namely if $s \in \ell^2$, then $f_s \in L^2(\mathbf{R})$ is square-integrable by the linearity of the integral. The correspondence $s \mapsto f_s$ is injective but not surjective.

*Research supported by NSF Grant DMS-0072234.

Conversely, by the comparison test, any $f \in L^2(\mathbf{R})$ corresponds to a square-summable sequence

$$s_f(n) \stackrel{\text{def}}{=} \int_{\mathbf{R}} f(x)\mathbf{1}(x-n)\,dx, \qquad n \in \mathbf{Z}.$$

This correspondence $f \mapsto s_f$ is surjective but not injective.

The integral may also be written $\langle f, \mathbf{1}_n\rangle$, putting $\mathbf{1}_n(x) \stackrel{\text{def}}{=} \mathbf{1}(x-n)$ and using the ordinary inner product in $L^2(\mathbf{R})$:

$$\langle f, g\rangle \stackrel{\text{def}}{=} \int_{\mathbf{R}} f(x)\bar{g}(x)\,dx.$$

Note that $\langle \mathbf{1}_n, \mathbf{1}_m\rangle = 0$ if $n \neq m$, and $\|\mathbf{1}_n\| = 1$ for all $n \in \mathbf{Z}$. Thus $\{\mathbf{1}_n : n \in \mathbf{Z}\}$ is an orthonormal subset of $L^2(\mathbf{R})$. We may define an *approximation space* $V_0 = \text{span}\{\mathbf{1}_n : n \in \mathbf{Z}\} \subset L^2(\mathbf{R})$. Those shifted indicator functions $\mathbf{1}_n$ supply an orthonormal basis for V_0, so the orthogonal projection $P_0 : L^2(\mathbf{R}) \to V_0$ is given by

$$P_0 f(x) \stackrel{\text{def}}{=} \sum_{n\in\mathbf{Z}} \langle f, \mathbf{1}_n\rangle \mathbf{1}_n(x), \qquad f \in L^2(\mathbf{R}).$$

If we replace $L^2(\mathbf{R})$ with V_0, then $f \mapsto s_f$ and $s \mapsto f_s$ are bijections. In fact, they are inverses of each other. In addition, $f \in V_0$ (respectively $s \in \ell^2$) is positive, monotone, or compactly supported if and only if s_f (respectively f_s) has the same property. The proofs are elementary, and show that approximations to functions are interchangeable with sequences of samples.

We may refine V_0 into $V_J \stackrel{\text{def}}{=} \{f(2^J x) : f \in V_0\} \subset L^2(\mathbf{R})$. It is immediate that $\cdots V_1 \subset V_0 \subset V_{-1} \subset \cdots$, as V_J consists of piecewise constant functions with jumps at $k2^J$, $k \in \mathbf{Z}$. Therefore, $\cup_{J\in\mathbf{Z}} V_J$ is dense in $L^2(\mathbf{R})$ since it contains all step functions with jumps at dyadic rationals.

$\{V_J : J \in \mathbf{Z}\}$ is the simplest example of a *multiresolution analysis*, or MRA [1, 2]. Given any $f \in L^2(\mathbf{R})$, we may approximate it arbitrarily well in V_J for some adequate J, then replace the approximate function with the equivalent sequence. Since dilations of f by 2^J move it from V_J to V_0, we may assume without loss that $P_0 f \in V_0$ provides an adequate degree of approximation.

The function $\mathbf{1} \in V_0$ is *refinable*: for all $x \in \mathbf{R}$,

$$\mathbf{1}(x) = \mathbf{1}(2x) + \mathbf{1}(2x-1) = \frac{1}{\sqrt{2}}[\sqrt{2}\,\mathbf{1}(2x)] + \frac{1}{\sqrt{2}}[\sqrt{2}\,\mathbf{1}(2x-1)].$$

This expression may be deduced from the inclusion $V_0 \subset V_{-1}$. The second form uses the orthonormal basis vectors of V_{-1}. In general, a refinable function is a compactly-supported function $\phi \in L^2(\mathbf{R})$ satisfying a *two-scale refinement equation* for all $x \in \mathbf{R}$:

$$\phi(x) = \sum_{n\in\mathbf{Z}} \sqrt{2}\,h(n)\,\phi(2x-n) \stackrel{\text{def}}{=} H\phi(x), \qquad (1)$$

where h is a finite sequence of *filter coefficients*. The Haar-Walsh case has $h(0) = h(1) = 1/\sqrt{2}$, with $h(n) = 0$ for $n \notin \{0, 1\}$.

Integration of both sides of Equation 1 shows that the filter must satisfy

$$\sum_{k \in \mathbf{Z}} h(2k) = \sum_{k \in \mathbf{Z}} h(2k+1) = \frac{1}{\sqrt{2}}. \tag{2}$$

If we want $\{\phi_n : n \in \mathbf{Z}\}$ to be an orthonormal set, where $\phi_n(x) = \phi(x - n)$, we must have

$$\sum_{k \in \mathbf{Z}} \overline{h(k)} h(k + 2m) = \begin{cases} 1, & \text{if } m = 0, \\ 0, & \text{if } m \neq 0, \end{cases} \tag{3}$$

for all $m \in \mathbf{Z}$. This is the key *orthogonality condition* on filters [1].

Taking a different point of view, suppose a finite filter h satisfies Equations 2 and 3. We may seek a solution to Equation 1 by iteration. We let $\phi^0 = \mathbf{1}$ and for $j > 0$ define $\phi^j = H\phi^{j-1}$ to get the following:

Theorem 1.1 *If ϕ^j converges uniformly to a limit function ϕ as $j \to \infty$, then*

1. *ϕ has compact support;*
2. *ϕ satisfies the two-scale refinement equation $H\phi = \phi$;*
3. *ϕ is nonzero, integrable, and satisfies $\int_{\mathbf{R}} \phi(x)\, dx = 1$;*
4. *$\{\phi(x - k) : k \in \mathbf{Z}\}$ is an orthonormal set in $L^2(\mathbf{R})$.* □

The proof is straightforward and may be found, for example, in Reference [3], page 165. Unfortunately, the conditions on h that guarantee uniform convergence are complicated [4]. The prize, however, is a function ϕ that may be substituted for $\mathbf{1}$ in the definition of the approximation spaces V_J of the MRA. Many examples are known [5, 1, 6] for which ϕ is smooth and thus gives superior compact coding of smooth functions in L^2.

2 Exponential Fourier Series

The celebrated theorem of L. Carleson [7], on the convergence of Fourier series, can be stated as follows:

Theorem 2.1 *If $f = f(x)$ is continuous on the interval $[0, 1]$, and*

$$c_n = \int_0^1 f(x) e^{-2\pi i n x}\, dx,$$

then $\sum_n c_n e^{2\pi i n x}$ converges to $f(x)$ at almost every x in $[0, 1]$.

The same conclusion holds for f belonging to the class $L^2 = L^2([0, 1])$ of square-integrable functions defined on the interval $[0, 1]$. No stronger conclusion is possible, since members of L^2 are actually equivalence classes of functions that agree almost everywhere.

Now L^2 is a complete inner product space, or Hilbert space, with Hermitean inner product $\langle f, g\rangle \stackrel{\text{def}}{=} \int_0^1 f(x)\bar{g}(x)\,dx$ and norm $\|f\|_2 \stackrel{\text{def}}{=} (\int_0^1 |f(x)|^2\,dx)^{1/2}$, which will be written $\|f\|$ when there is no risk of confusion. A countable subset $\{b_n\} \subset L^2$ is an orthonormal basis, or Hilbert basis, for L^2 if it satisfies the following three conditions:

Normalization: $\|b_n\| = 1$ for all n;

Orthogonality: $\langle b_n, b_m\rangle = 0$ if $n \neq m$;

Density: $\operatorname{span}\{b_n\}$ is dense in L^2.

The generalized Fourier coefficient c_n of f, with respect to an orthonormal basis $\{b_n\}$, may be written $c_n = \langle f, b_n\rangle$, and the generalized Fourier series written as $\sum_n \langle f, b_n\rangle b_n = \sum_n c_n b_n$. For any Hilbert basis, there is at least one kind of convergence:

Theorem 2.2 *If f belongs to L^2 and $\{b_n : n \in \mathbf{Z}\}$ is any orthonormal basis for L^2, then $\|f - \sum_{n=-M}^{N} \langle f, b_n\rangle b_n\|$ tends to 0 as $M, N \to \infty$. Equivalently, the norms of the series tails $\|\sum_{n>N} \langle f, b_n\rangle b_n\|$ and $\|\sum_{n<-N} \langle f, b_n\rangle b_n\|$ must tend to zero as $N \to \infty$.*

This is called L^2 norm convergence, and it follows from the Riesz–Fischer theorem and Parseval's theorem. The proof is elementary and may be found, for example, in [8], pages 309–311. It does not, however, imply pointwise convergence even at a single point.

With a bit of effort, one shows that the exponential functions $\{e_n : n \in \mathbf{Z}\}$ defined by $e_n(x) = e^{2\pi i n x}$ form an orthonormal basis of L^2. Orthonormality can be shown with the calculus, and density follows from an analysis of the symmetric partial sum $\sum_{|n|<N} e_n$. Thus, the exponential Fourier series converges in L^2 norm by Theorem 2.2.

Carleson's theorem implies that when $\{b_n\} = \{e_n\}$, convergence occurs not only in L^2 norm but also pointwise almost everywhere. Equivalently, the series tail functions $\sum_{n>N} c_n e_n(x)$ and $\sum_{n<-N} c_n e_n(x)$ must tend to zero at almost every point x, as $N \to \infty$. The purpose of this survey is to describe some other orthonormal bases of L^2 that have this *pointwise almost everywhere convergence property.*

R. Hunt [9] extended Carleson's argument to show pointwise almost everywhere convergence for the Fourier series of any $f \in L^p = L^p([0,1])$, namely any f whose norm $\|f\|_p \stackrel{\text{def}}{=} \left(\int_0^1 |f(x)|^p\,dx\right)^{1/p}$ is finite for some $1 < p < \infty$. Like for L^2, no stronger conclusion is possible in any of these classes, since their members are only defined almost everywhere. Also, since L^p with $p \neq 2$ is not a Hilbert space, the notion of orthonormal basis must be replaced. A natural candidate is the Schauder basis, defined as a countable subset $\{b_n\}$ of the space whose span is dense and for which $\sum_n a_n b_n = 0$ implies $a_n = 0$ for all n. The exponentials $\{e_n : n \in \mathbf{Z}\}$ form a Schauder basis for all L^p with $1 < p < \infty$, but some of the generalizations considered in this article do not.

3 Walsh Functions

Walsh functions are analogues of $\{e_n\}$, in the sense that they form an orthonormal basis for L^2 and are indexed by a frequency parameter. They may be defined recursively, if they are considered to be functions defined on all of $\mathbf{R}$ but supported in $[0, 1]$. Namely, put $W_0 = \mathbf{1}$, the characteristic function of $[0, 1)$, and for $n = 0, 1, 2, \ldots$, define

$$W_{2n}(x) = W_n(2x) + W_n(2x + 1); \qquad W_{2n+1}(x) = W_n(2x) - W_n(2x + 1). \quad (4)$$

It is elementary to show that the functions in $\{W_n : n = 0, 1, 2, \ldots\}$ are uniformly bounded and uniformly compactly supported in $[0, 1]$, and that they are orthonormal with respect to the L^2 inner product. To show that $\text{span}\{W_n : n = 0, 1, 2, \ldots\}$ is dense in L^2, observe that this span contains the characteristic function of every dyadic interval $[2^{-N}k, 2^{-N}(k + 1))$ for every $N = 0, 1, 2, \ldots$ and every $0 \leq k < 2^N$. Clearly such characteristic functions are dense in L^2. It follows that $\{W_n\}$ is an orthonormal basis for L^2. Consequently, Parseval's theorem applies when $\{b_n\} = \{W_n\}$. But Carleson's result applies, too:

Theorem 3.1 *If $f = f(x)$ is continuous on the interval $[0, 1]$, then the Walsh series $\sum_n \langle f, W_n\rangle W_n$ converges to $f(x)$ at almost every x in $[0, 1]$.*

Furthermore, the result applies to $f \in L^p$ as well [10, 11, 12].

4 Shannon Functions

For integer $n \geq 0$, let $S_n = S_n(x)$ be defined by

$$S_n(x) = \frac{\sin\left[\pi(n + 1)(x - \frac{1}{2})\right] - \sin\left[\pi n(x - \frac{1}{2})\right]}{\pi(x - \frac{1}{2})}. \quad (5)$$

The Shannon functions are the doubly-indexed set $\{S_{nk} : n \in \mathbf{N}; k \in \mathbf{Z}\}$ defined by

$$S_{nk}(x) = S_n(x - k). \quad (6)$$

It may not be obvious, but is nonetheless true, that $\{S_{nk} : n \in \mathbf{N}, k \in \mathbf{Z}\}$ is an orthonormal basis for $L^2(\mathbf{R})$. A short proof is available, using some of the elementary tools of harmonic analysis. First note that the Fourier integral transforms $v_n = \hat{S}_n$ of Shannon functions have simple formulas. It can be verified by the calculus and the Fourier inversion theorem that $v_n(\xi) = \rho(\xi)\mathbf{1}(2|\xi| - n)$, where $\rho(\xi) = e^{-\pi i\xi}$. Since the absolute value of v_n is one on its support, it follows that $\int |v_n(\xi)|\, d\xi = \int |v_n(\xi)|^2\, d\xi = 1$ for all n. Plancherel's theorem then implies that $\|S_n\| = 1$ for all n. Furthermore, writing $v_{nk} = \hat{s}_{nk}$, one directly computes that $v_{nk}(\xi) = e^{2\pi i k\xi} v_n(\xi)$. Then, because of their nonoverlapping supports, it is clear that $\langle v_{nk}, v_{mj}\rangle = 0$ for $n \neq m$ and any $k, j \in \mathbf{Z}$. The orthogonality of exponential functions implies that $\langle v_{nk}, v_{nj}\rangle = 0$ if $k \neq j$, and the orthonormality of $\{S_{nk}\}$ follows by Plancherel's theorem.

The Riemann-Lebesgue Lemma implies that Shannon functions are uniformly bounded and continuous. But also, each Shannon function has compactly supported Fourier integral transform, and so it is band-limited. Such functions, by the Paley-Wiener theorem, are entire real analytic: they have derivatives of every order at every point, and are represented everywhere by their Taylor series.

5 The Carleson Operator

The difficult part of Carleson's theorem is obtaining a bound on the Carleson operator, which for any orthonormal basis $\{b_n : n \in \mathbf{N}\}$ of $L^2(\mathbf{R})$ is

$$Lf(x) \stackrel{\text{def}}{=} \sup_{N\geq 0} \sum_{0\leq n<N} \langle f, b_n\rangle b_n(x). \tag{7}$$

The rest of the proof is straightforward: suppose it has been shown that $\|Lf\| \leq c\|f\|$ for some fixed $c > 0$ and all $f \in L^2$. Consider the remainder term after subtracting a partial sum from f:

$$f_N = f - \sum_{0\leq n<N} \langle f, b_n\rangle b_n = \sum_{n\geq N} \langle f, b_n\rangle b_n.$$

Then Parseval's theorem implies $\|f_N\| \to 0$ as $N \to \infty$, so $\|Lf_N\| \to 0$ as $N \to \infty$. But for each fixed x, the sequence $\{Lf_N(x)\}$ decreases as $N \to \infty$. Thus $Lf_N(x) \to 0$ for almost every x, as $N \to \infty$, by the monotone convergence theorem ([8], page 265). But $f_N(x) \leq Lf_N(x)$ at every x, so $f_N(x) \to 0$ as $N \to \infty$ for almost every x as well, finishing the proof.

M. Lacey and C. Thiele [11] recently gave an interesting alternative proof of Carleson's theorem, in which they focused on the integer-valued function $N_f(x)$ defined by $Lf(x) = \sum_{0\leq n<N_f(x)} \langle f, b_n\rangle b_n(x)$. For Walsh series, they were able to estimate $\|Lf\|$ with a geometric argument.

6 Wavelet Packets

Equation 4 may be generalized as follows: Let $h = \{h(k) : k \in \mathbf{Z}\}$ and $g = \{g(k) : k \in \mathbf{Z}\}$ be two finitely-supported sequences, fix the initial functions w_0 and w_1 in $L^2(\mathbf{R})$, and for each integer $n > 0$ define

$$w_{2n}(x) = \sum_k h(k) w_n(2x - k) \stackrel{\text{def}}{=} Hw_n(x); \tag{8}$$

$$w_{2n+1}(x) = \sum_k g(k) w_n(2x - k) \stackrel{\text{def}}{=} Gw_n(x). \tag{9}$$

As before, define $w_{nk}(x) = w_n(x-k)$. The collection of functions $\{w_n : n \in \mathbf{N}\}$ will be an orthonormal basis for $L^2(\mathbf{R})$ if $\phi = w_0$ and $\psi = w_1$ are the scaling function and mother wavelet, respectively, of an orthonormal multiresolution analysis of $L^2(\mathbf{R})$, or MRA, and operators H, G are defined by sequences h, g satisfying the following conditions for all integers n, m:

- $\sum_k h(k)h(k+2n) = 2\delta(n)$;
- $\sum_k g(k)g(k+2n) = 2\delta(n)$;
- $\sum_k g(k)h(k+2n) = 0$;
- $\sum_k [h(n+2k)h(m+2k) + g(n+2k)g(m+2k)] = 2\delta(n-m)$.

Here δ is the Kronecker symbol; $\delta(0) = 1$, but $\delta(n) = 0$ if $n \neq 0$. Sequences h, g satisfying these conditions are called *orthogonal conjugate quadrature filters* (orthogonal CQF).

Walsh functions are obtained by taking $h(0) = h(-1) = g(0) = -g(-1) = 1$, with $h(k) = g(k) = 0$ for $k \notin \{0, -1\}$, to define H and G, and functions $\phi = \mathbf{1}$, and $\psi = G\mathbf{1}$.

Shannon functions can also be obtained by this recursion, if the condition that h and g be finitely supported is removed. Take

$$h(k) = \frac{\sin\left[\frac{\pi}{2}(k-\frac{1}{2})\right]}{\frac{\pi}{2}(k-\frac{1}{2})}; \qquad g(k) = (-1)^k h(1-k) = (-1)^k \frac{\sin\left[\frac{\pi}{2}(k-\frac{1}{2})\right]}{\frac{\pi}{2}(k-\frac{1}{2})}, \tag{10}$$

to define H and G, and

$$\phi(x) = \frac{\sin\left[\pi(x-\frac{1}{2})\right]}{\pi(x-\frac{1}{2})}; \qquad \psi(x) = \frac{\sin\left[2\pi(x-\frac{1}{2})\right] - \sin\left[\pi(x-\frac{1}{2})\right]}{\pi(x-\frac{1}{2})}, \tag{11}$$

for the initial functions.

Operators H and G act as Fourier multipliers:

$$\hat{w}_{2n}(\xi) = \frac{1}{2} m_0(\frac{\xi}{2}) \hat{w}_n(\frac{\xi}{2}); \qquad \hat{w}_{2n+1}(\xi) = \frac{1}{2} m_1(\frac{\xi}{2}) \hat{w}_n(\frac{\xi}{2}), \tag{12}$$

where $m_0(\xi) = \sum_k h(k) e^{-2\pi i k\xi}$ and $m_1(\xi) = \sum_k g(k) e^{-2\pi i k\xi}$. Functions m_0 and m_1 are 1-periodic, and are trigonometric polynomials whenever h and g are finitely supported.

In the Walsh case, $m_0(\xi) = 1 + e^{2\pi i\xi} = 2e^{\pi i\xi} \cos \pi\xi$, and $m_1(\xi) = 1 - e^{2\pi i\xi} = -2ie^{\pi i\xi} \sin \pi\xi$. In the Shannon case, one can take

$$\begin{aligned} m_0(\xi) &= \begin{cases} 2, & \text{if } k - \frac{1}{4} \leq \xi < k + \frac{1}{4} \text{ for some integer } k; \\ 0, & \text{otherwise;} \end{cases} \\ m_1(\xi) &= \begin{cases} 2, & \text{if } k + \frac{1}{4} \leq \xi < k + \frac{3}{4} \text{ for some integer } k; \\ 0, & \text{otherwise,} \end{cases} \quad = \; 2 - m_0(\xi). \end{aligned}$$

7 Daubechies' Wavelet Packets

The filters h and g that define the compactly-supported orthonormal wavelets of I. Daubechies [5] can be used here. For example, the Daubechies filter of length 4, which produces a scaling function supported in $[0,3]$ that satisfies $\phi = H\phi$, and a mother wavelet also supported in $[0,3]$ that satisfies $\psi = G\phi$, uses

$$h(k) = \begin{cases} \frac{1+\sqrt{3}}{4}, & \text{if } k = 0; \\ \frac{3+\sqrt{3}}{4}, & \text{if } k = -1; \\ \frac{3-\sqrt{3}}{4}, & \text{if } k = -2; \\ \frac{1-\sqrt{3}}{4}, & \text{if } k = -3; \\ 0, & \text{otherwise;} \end{cases} \qquad g(k) = \begin{cases} \frac{1-\sqrt{3}}{4}, & \text{if } k = 0; \\ -\frac{3-\sqrt{3}}{4}, & \text{if } k = -1; \\ \frac{3+\sqrt{3}}{4}, & \text{if } k = -2; \\ -\frac{1+\sqrt{3}}{4}, & \text{if } k = -3; \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

Note that $g(k) = (-1)^k h(-3-k)$.

For every positive integer $N > 1$ there is a Daubechies wavelet supported in $[0, 2N-1]$ which belongs to the smoothness class C^d for $d \approx N/5$ [5]. Since Daubechies' wavelets form an orthonormal MRA, the associated wavelet packets $\{w_{nk} : n \in \mathbf{N}, k \in \mathbf{Z}\}$ form an orthonormal basis for $L^2(\mathbf{R})$, and they are just as smooth as the mother wavelet and scaling function, because the filters are finitely supported. Unfortunately, though they are smooth, these wavelet packets are not uniformly bounded. The following is proved in [13]:

Theorem 7.1 *For any orthogonal CQFs (h, g) for which $m_0(\xi) \neq 0$ on $-\frac{\pi}{2} \leq \xi \leq \frac{\pi}{2}$, the wavelet packets $\{w_n\}$ satisfy*

$$\limsup_{n\to\infty} \frac{1}{n} \left(\|\hat{w}_0\|_1 + \cdots + \|\hat{w}_n\|_1\right) = \infty.$$

In particular, the nonvanishing condition on m_0 is satisfied by Daubechies' filters. If in addition m_0 is nonnegative, then $\|\hat{w}_n\|_1$ and $\|w_n\|_\infty$ will be equivalent, so

$$\limsup_{n\to\infty} \frac{1}{n} \left(\|w_0\|_\infty + \cdots + \|w_n\|_\infty\right) = \infty.$$

Thus, such wavelet packets are not bounded on average, as the frequency index increases.

A refined special case of this result is shown in [14]:

Theorem 7.2 *For Daubechies' filters of length $L = 4$ through $L = 20$, there exist $p_{\min} < \infty$, $C > 0$, and $r > 1$, all depending on L, such that*

$$\|w_{2^n-1}\|_p > Cr^n,$$

for all $p > p_{\min}$.

In particular, the theorem holds for $p = \infty$. The result depends on a calculation, and holds for some other well-known CQFs as well. In the $L = 4$ case, $p_{\min} = 2$.

There is numerical evidence that the wavelet packets with frequency index $2^n - 1$ have the fastest growth as $n \to \infty$, while those with frequency index 2^n seem to be uniformly bounded.

It is not known whether Daubechies' wavelet packets have the almost everywhere convergence property.

8 Counting Wavelet Packet Bases

Suppose that (h, g) is an orthogonal CQF pair derived from an MRA, and let $\{w_n : n = 0, 1, \ldots\} \subset L^2(\mathbf{R})$ be the resulting wavelet packets. We may rescale and shift each of them to define

$$w_{nsp}(x) \stackrel{\text{def}}{=} 2^{-s/2} w_n(2^{-s}t - p),$$

a function with nominal scale 2^s, $s \in \mathbf{Z}$, and nominal position $p \in \mathbf{Z}$. The collection $\{w_{nsp} : n \in \mathbf{N}, s \in \mathbf{Z}, p \in \mathbf{Z}\} \subset L^2(\mathbf{R})$ is an overcomplete set with many subsets constituting orthonormal bases of $L^2(\mathbf{R})$ and its subspaces [15].

Among these subsets, a *graph basis* of wavelet packets corresponds to $\{w_{nsp} : 0 \le n \le 2^J; s = s(n), 0 \le s \le J; p \in \mathbf{Z}\}$, where the function s is a constant on intervals of the form $[k2^j, (k+1)2^j)$, with $0 \le j \le J$ and appropriate k. These functions form an orthonormal basis of $V_0 = \text{span}\{w_0(x-p) : p \in \mathbf{Z}\}$, with the parameter J giving the depth of decomposition.

The number of graph bases with depth J or less is the number of ways we may decompose the interval $[0, 2^J)$ into subintervals of the form $[k2^j, (k+1)2^j)$ with $0 \le j \le J$ and appropriate $k \in \mathbf{N}$. Equivalently, it is the number of ways to partition $[0,1) \subset \mathbf{R}$ using dyadic rationals with denominator at most 2^J. We divide the points $k2^j$ by 2^J to obtain the partition of $[0, 1)$.

Denote the number of graph bases to depth J by B_J. Then $B_0 = 1$, $B_1 = 2$, and for all $J > 1$ we have

$$B_J = B_{J-1}^2 + 1, \tag{14}$$

since a partition of $[0,1)$ either has a single element $[0,1)$, or else consists of two independent partitions of $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$ to depth $J-1$ each. The recursion in Equation 14 has been studied in [16] and is known to have the solution

$$B_J = \left\lfloor C^{2^J} \right\rfloor, \qquad J = 0, 1, 2, \ldots, \tag{15}$$

where $C \approx 1.502837\ldots$ is a constant given by

$$C = \exp\left(\sum_{k=1}^{\infty} \frac{1}{2^k} \log \frac{B_k}{B_k - 1} \right). \tag{16}$$

The easy estimate $B_J > 2^{2^{J-1}} = (\sqrt{2})^{2^J}$, which we get from observing that $B_J > B_{J-1}^2$, shows that B_J increases rapidly with J. The sum in the expression for C thus converges very rapidly, and just a few terms give an accurate approximation.

For b-band wavelet packet decompositions to depth J, as described in Reference [17], the number of graph bases is governed by the recursion

$$D_J = D_{J-1}^b + 1, \qquad J = 1, 2, \ldots, \tag{17}$$

with initial condition $D_0 = 1$. Another easy estimate like the one above shows that $D_J > 2^{b^{J-1}}$, but there are better estimates and a formula analogous to Equations 15 and 16.

For Haar-Walsh filters, there are also double tree decompositions ([18], pages 256 and 342). These give bases that are more general and thus more numerous than graph bases. The count of bases satisfies a rapidly increasing recursion [19] similar to Equation 17.

Separable isotropic d-dimensional wavelet packet decompositions to depth J are likewise counted by Equation 17, using $b = 2^d$. There are also anisotropic and mixed isotropic/anisotropic multidimensional wavelet packet decompositions ([18], page 308) whose enumerations give rise to other recursions [20].

9 Nonstationary Wavelet Packets

An integer n in the range $0 \leq n < 2^J$, for integer $J \geq 0$, may be written in binary as

$$n = \sum_{j=1}^{J} n_j 2^{j-1},$$

where $n_j \in \{0, 1\}$. The numbering is chosen so that n_1 is the least significant bit and n_J is the most significant bit of the J-bit expansion of n. The restriction $2^{J-1} \leq n < 2^J$ implies that $n_J = 1$.

With the definitions $F_0 \stackrel{\text{def}}{=} H$ and $F_1 \stackrel{\text{def}}{=} G$, it is possible to write the *filter formulation* of wavelet packets:

$$w_n = F_{n_1} F_{n_2} \cdots F_{n_J} w_0, \tag{18}$$

where $2^{J-1} \leq n < 2^J$. Alternatively, there is also a *multiplier formulation*:

$$\hat{w}_n(\xi) = \frac{1}{2^J} \hat{w}_0(\frac{\xi}{2^J})\, m_{n_J}(\frac{\xi}{2^J}) m_{n_{J-1}}(\frac{\xi}{2^{J-1}}) \cdots m_{n_2}(\frac{\xi}{2^2}) m_{n_1}(\frac{\xi}{2}). \tag{19}$$

M. Nielsen [14] studied two generalizations of this recursive definition.

Let $\{(h^J, g^J : J = 1, 2, \ldots\}$ be a family of orthogonal CQF pairs. Fix w_0, and for $J \geq 2$ and $2^{J-1} \leq n < 2^J$ define *nonstationary* wavelet packets by

$$w_n(x) = F^1_{n_1} F^2_{n_2} \cdots F^J_{n_J} w_0(x), \tag{20}$$

or alternatively, in the multiplier formulation, define their Fourier integral transforms by

$$\hat{w}_n(x) = \frac{1}{2^J} w_0(\frac{\xi}{2^J})\, m^J_{n_j}(\frac{\xi}{2^J}) \cdots m^2_{n_2}(\frac{\xi}{2^2}) m^1_{n_1}(\frac{\xi}{2}). \tag{21}$$

The superscript indicates which pair of CQFs defines the filter operator or multiplier. The idea is to change the filters used to generate wavelet packets as their frequency increases, for example, to control their growth in L^∞.

But one can also redo the entire recursion for each new level. Let

$$\{\big((h^{J,J}, g^{J,J}), \ldots, (h^{J,1}, g^{J,1})\big) : J = 1, 2, \ldots\},$$

be a family of sequences of orthogonal CQF pairs. Fix w_0, and for $J \geq 2$ and $2^{J-1} \leq n < 2^J$ define *highly nonstationary* wavelet packets by

$$w_n(x) = F^{J,1}_{n_1} F^{J,2}_{n_2} \cdots F^{J,J}_{n_J} w_0(x), \tag{22}$$

or alternatively, in the multiplier formulation, define their Fourier integral transforms by

$$\hat{w}_n(x) = \frac{1}{2^J} w_0(\frac{\xi}{2^J})\, m^{J,J}_{n_J}(\frac{\xi}{2^J}) \cdots m^{J,2}_{n_2}(\frac{\xi}{2^2}) m^{J,1}_{n_1}(\frac{\xi}{2}). \tag{23}$$

Here the superscripts indicate which pair of which sequence of CQFs defines the filter operator or multiplier.

10 Walsh and Shannon Type Wavelet Packets

Suppose that (h^J, g^J) is the Walsh CQF pair for all sufficiently large $J \geq J_0$. The resulting wavelet packets are called *Walsh-type*, and we have the following theorem due to M. Nielsen [14]:

Theorem 10.1 *Walsh-type wavelet packet series converge pointwise almost everywhere.*

Likewise, if (h^J, g^J) is the Shannon CQF pair for all sufficiently large $J \geq J_0$, then the resulting wavelet packets are called *Shannon-type*, and we have another theorem by M. Nielsen:

Theorem 10.2 *Shannon-type wavelet packet series converge pointwise almost everywhere.*

These theorems are direct consequences of the Carleson–Hunt theorem for Walsh series and Shannon series. Generalizing a result of Y. Meyer [21], one shows that for each Walsh-type wavelet packet basis and each $1 < p < \infty$, there is an isomorphism of L^p that maps the basis onto Walsh functions. The L^p boundedness of the Carleson operator follows. Similarly, each Shannon-type wavelet packet basis is an L^p isomorphic image of the Shannon basis functions.

11 Growth Control for Wavelet Packets

One way to control the growth of $\|w_n\|_p$ for large p, as $n \to \infty$, is to use nonstationary or highly nonstationary wavelet packets with lengthening filters. One obtains a uniform bound on $\|w_n\|_\infty$, for example, from a uniform bound on $\|\hat{w}_n\|_1$, using the Riemann–Lebesgue lemma and the Fourier inversion theorem: $\|w_n\|_\infty \leq \|\hat{w}_n\|_1$.

For values $2 \leq p < \infty$, the bound for $\|w_n\|_p$ follows from the Hausdorff–Young inequality:

$$\|w_n\|_p \leq C\|\hat{w}_n\|_q,$$

where $q = p/(p-1)$ and the sharp constant [22] is $C = [q^{\frac{1}{q}}/p^{\frac{1}{p}}]^{\frac{1}{2}}$.

N. Hess-Nielsen [23, 24] originally introduced the idea of building wavelet packet bases with more than one CQF pair. An original application was to design a single short CQF pair with the same frequency localization as longer CQFs, given a desired depth J of wavelet packet decomposition. This resulted in a savings of approximately half the arithmetic operations in subband decompositions.

A. Cohen and E. Séré [25] showed the following:

Theorem 11.1 *Suppose (h^J, g^J) is a family of orthogonal CQFs whose length function $L = L(J)$ satisfies $L(J) \geq cJ^{3+\epsilon}$ for some $c > 0$ and $\epsilon > 0$. Then the associated nonstationary wavelet packets $\{w_n\}$ satisfy*

$$2^{-J}\left(\|\hat{w}_0\|_1 + \cdots + \|\hat{w}_{2^J-1}\|_1\right) \leq B,$$

for some $B < \infty$ and all $J \geq 0$. Thus,

$$2^{-J}\left(\|w_0\|_\infty + \cdots + \|w_{2^J-1}\|_\infty\right) \leq B,$$

as well.

M. Nielsen [14] refined this result in the special case where h^J, g^J are the Daubechies orthogonal CQFs of length $L = L(J)$, where the length function will be specified later. When highly nonstationary wavelet packets are called for, use $h^{J,j} \stackrel{\text{def}}{=} h^J$ and $g^{J,j} \stackrel{\text{def}}{=} g^J$ for all $j = 1, 2, \ldots, J$. One may suppose that w_0 is any scaling function that generates an orthonormal MRA, not necessarily a Daubechies scaling function. One must suppose, however, that w_0 is smooth enough so that $|\hat{w}_0(\xi)| = O(1/|\xi|^{1+\epsilon})$ for some $\epsilon > 0$. One first obtains a basic result, part of which was also shown in [25]:

Theorem 11.2 *For any length function $L = L(J)$, the nonstationary wavelet packets derived from $\{h^J, g^J\}$ and the highly nonstationary wavelet packets derived from $\{h^{J,j}, g^{J,j}\}$ form an orthonormal basis for $L^2(\mathbf{R})$.*

The additional properties of Daubechies' CQFs give a better growth result:

Theorem 11.3 *If the length function satisfies*

$$L(J) \geq cJ^{2+\epsilon}$$

for some $c > 0$ and $\epsilon > 0$, then the nonstationary wavelet packets derived from Daubechies' filters $\{h^J, g^J\}$ are uniformly bounded functions.

The support diameter of the nonstationary wavelet packet w_n grows without bound as $n \to \infty$, if $L(J) \to \infty$ as $J \to \infty$. This is overcome, strangely enough, by backing up and introducing longer filters earlier in the highly nonstationary wavelet packet algorithm [14]:

Theorem 11.4 *If the length function satisfies*

$$cJ^{2+\epsilon} \leq L(J) \leq \frac{2^J}{cJ^{1+\epsilon}}$$

for some $c > 0$ and $\epsilon > 0$, and w_1 has compact support, then the highly nonstationary wavelet packets $\{w_n\}$ derived from Daubechies' filters $\{h^{J,j}, g^{J,j}\}$ are uniformly bounded and have uniform compact support in a fixed interval independent of n.

12 Wavelet Packets as Schauder Bases

A countable set $B = \{b_n\} \subset L^p(\mathbf{R})$, $1 < p < \infty$, that has a dense span is called a *Schauder basis* if $\sum_n a_n b_n = 0$ implies that the coefficient $a_n = 0$ for all n. In particular, it means that finite subsets of B must be linearly independent. Every $f \in L^p$ has a unique representation $f = \sum_n a_n(f) b_n$ that converges in norm, and it is well known that the coefficient functionals $\{a_n\}$, which are given by functions in L^q, $q = p/(p-1)$, must satisfy

$$\sup_n \|a_n\|_q \|b_n\|_p < \infty. \tag{24}$$

An orthonormal basis for L^2 consisting of compactly-supported wavelet packets $\{w_n\}$, derived from an MRA with somewhat smooth scaling function w_0, has a dense span in L^p and satisfies $a_n = b_n = w_n$. Using Theorem 7.2, M. Nielsen showed [14] that

Theorem 12.1 *For Daubechies' filters of length $L = 4$ through $L = 20$, there exist $p_{\min} < \infty$, $C > 0$, and $r > 1$, all depending on L, such that*

$$\|w_{2^n-1}\|_p \|w_{2^n-1}\|_q \to \infty$$

as $n \to \infty$, for all $p > p_{\min}$. Thus these wavelet packets fail to be a Schauder basis for those L^p spaces.

It is known that Walsh functions are a Schauder basis for L^p, all $1 < p < \infty$, as are Shannon functions. Using that fact, M. Nielsen showed [14] that Walsh-type and Shannon-type wavelet packets likewise constitute a Schauder basis for L^p.

It is not known whether the nonstationary and highly nonstationary wavelet packet bases of Theorems 11.3 and 11.4, which are uniformly bounded, give Schauder bases, as Equation 24 only gives a necessary condition. However, a perturbation argument applied to periodized Shannon wavelet packets [26] shows that certain highly nonstationary periodic wavelet packets, which are very close to exponentials, do constitute a Schauder basis for $L^p([0, 1])$.

References

[1] Ingrid Daubechies. *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM Press, Philadelphia, 1992.

[2] Stéphane G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, New York, second edition, 1999.

[3] Mladen Victor Wickerhauser. *Mathematics for Multimedia*. Academic Press, San Diego, California, 2003. To appear.

[4] Alfred S. Cavaretta, Wolfgang Dahmen, and Charles A. Micchelli. *Stationary Subdivision*, volume 93:453 of *Memoirs of the American Mathematical Society*. American Mathematical Society, Providence, Rhode Island, 1991.

[5] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.

[6] Ingrid Daubechies and Jeffrey C. Lagarias. Two-scale difference equations I. global regularity of solutions. *SIAM Journal on Mathematical Analysis*, 22(5):1388–1410, 1991.

[7] Lennart Carleson. On convergence and growth of partial sums of Fourier series. *Acta Mathematica*, 116:135–157, 1966.

[8] Tom M. Apostol. *Mathematical Analysis*. Addison–Wesley, Reading, Massachusetts, second edition, 1974.

[9] Richard A. Hunt. On the convergence of Fourier series. In *Orthogonal Expansions and Their Continuous Analogues*, pages 235–255. Southern Illinois University Press, 1968.

[10] John Gosselin. On the convergence of Walsh–Fourier series for $L^2(0,1)$. *Advances in Mathematics Supplementary Studies*, 4:223–232, 1979.

[11] Michael Lacey and Christoph Thiele. A proof of the boundedness of the Carleson operator. *Mathematics Research Letters*, 7:361–370, 2000.

[12] Mitchell H. Taibleson. Fourier series on the ring of integers in a p-series field. *Bulletin of the American Mathematical Society*, 73(5):623–629, September 1971.

[13] Ronald R. Coifman, Yves Meyer, and Mladen Victor Wickerhauser. Size properties of wavelet packets. In Mary Beth Ruskai, Gregory Beylkin, Ronald Coifman, Ingrid Daubechies, Stéphane Mallat, Yves Meyer, and Louise Raphael, editors, *Wavelets and Their Applications*, pages 453–470. Jones and Bartlett, Boston, 1992.

[14] Morten Nielsen. *Size Properties of Wavelet Packets*. PhD thesis, Washington University, Saint Louis, Missouri, 1999.

[15] Ronald R. Coifman and Mladen Victor Wickerhauser. Entropy based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 32:712–718, March 1992.

[16] Alfred V. Aho and Neil J. A. Sloane. Some doubly exponential sequences. *Fibonacci Quarterly*, 11:429–437, 1970.

[17] Mladen Victor Wickerhauser. Best-adapted wavelet packet bases. In Ingrid Daubechies, editor, *Different Perspectives on Wavelets*, number 47 in Proceedings of Symposia in Applied Mathematics, pages 155–171. American Mathematical Society, Providence, Rhode Island, January 1993.

[18] Mladen Victor Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*. A K Peters, Ltd., Natick, Massachusetts, 1994.

[19] Christoph M. Thiele and Lars F. Villemoes. A fast algorithm for adapted time frequency tilings. *Applied and Computational Harmonic Analysis*, 3(2):91–99, April 1996.

[20] Nicolas N. Bennett. Fast algorithms for best anisotropic Walsh bases and relatives. *Applied and Computational Harmonic Analysis*, 8(1):86–103, January 2000.

[21] Yves Meyer. *Wavelets and Operators*, volume 37 of *Cambridge studies in advanced mathematics*. Cambridge University Press, Cambridge, 1992.

[22] Elliott H. Lieb and Michael Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, first edition, 1997.

[23] Nikolaj Hess-Nielsen. *Time-Frequency Analysis of Signals Using Generalized Wavelet Packets*. PhD thesis, University of Aalborg, 1992.

[24] Nikolaj Hess-Nielsen. Control of frequency spreading of wavelet packets. *Applied and Computational Harmonic Analysis*, 1(2):157–168, March 1994.

[25] Albert Cohen and Eric Séré. Time-frequency localization by nonstationary wavelet packets. In Ali N. Akansu and Mark J. T. Smith, editors, *Subband and Wavelet Transforms — Theory and Design*, pages 189–211. Kluwer Academic Publishers, 1996.

[26] Morten Nielsen. Highly nonstationary wavelet packets. *Applied and Computational Harmonic Analysis*, 12(2):209–229, March 2002.