Advances in Wavelet Algorithms and Applications

Mladen Victor Wickerhauser

Department of Mathematics,
Washington University,
St. Louis, MO 63130, USA
victor@math.wustl.edu

March 4, 2002

Abstract

Wavelet and wavelet packet transforms are presently used for image compression and denoising. There has been recent progress on three fronts: implementing multiplication operations in wavelet bases, estimating compressibility by wavelet packet transform coding, and designing wavelet packets to control frequency spreading and pointwise convergence. Some open problems are mentioned.

Keywords: Wavelet transforms, edge detection, singularity analysis, wavelet packets, nonlinear approximation, rate-distortion, frequency spreading, connection coefficients

1 Introduction

Wavelets and *wavelet packets* are special functions having three useful properties:

- They are almost as well localized in both time and frequency as the Heisenberg uncertainty inequality allows;
- They form orthogonal bases;
- They come equipped with fast, well-conditioned transforms: to compute N expansion coefficients of a function costs only $O(N \log N)$ operations.

A simple almost-example of wavelets, which lacks only adequate frequency localization, is the Haar basis [13] generated by the compactly-supported "mother" function $\psi = \psi(x) = \mathbf{1}(2x) - \mathbf{1}(2x-1)$, where $\mathbf{1}(x)$ is the characteristic or indicator function of the interval $[0,1)$. The linear span of the following orthogonal unit vectors is dense in $L^2(\mathbf{R})$:

$$\{\psi_{sn}(x) \stackrel{\text{def}}{=} 2^{-s/2}\psi(2^{-s}x - n) : s, n \in \mathbf{Z}\}. \tag{1}$$

This generalizes to Walsh functions, which are almost wavelet packets. Fix the initial functions $w_0 = \mathbf{1}$ and $w_1 = \psi$ in $L^2(\mathbf{R})$, and for each integer $n > 0$ define

$$w_{2n}(x) = w_n(2x) + w_n(2x-1); \qquad w_{2n+1}(x) = w_n(2x) - w_n(2x-1).$$

Putting $w_{n,k}(x) \stackrel{\text{def}}{=} w_n(x-k)$ gives an orthonormal basis $\{w_{n,k} : n \in \mathbf{N}, k \in \mathbf{Z}\}$ for $L^2(\mathbf{R})$.

Daubechies' famous smooth generalization [10] of Haar's basis is constructed using the *two-scale equations*:

$$\phi(x) = \sum_k h_k \sqrt{2}\phi(2x-k) \stackrel{\text{def}}{=} H\phi(x); \qquad \psi(x) = \sum_k g_k \sqrt{2}\phi(2x-k) \stackrel{\text{def}}{=} G\phi(x). \tag{2}$$

where $h = \{h_k : k \in \mathbf{Z}\}$ and $g_k = (-1)^{1-k} h_{1-k}$ are finitely-supported sequences satisfying the following conditions for all integers n, m:

$$\sum_k h_k \, h_{k+2n} = \delta(n); \qquad \sum_k [h_{n+2k} \, h_{m+2k} + g_{n+2k} \, g_{m+2k}] = \delta(n-m). \tag{3}$$

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*, or CQFs. One nontrivial example is the "Coifman 12" (C12) filters [28]. Here $h_k = 0$ if $k < 0$ or $k \geq 12$, and $\{h_k : 0 \leq k < 12\}$ is the following table of values:

$$\begin{array}{rrr} \{1.6387336463179785 \times 10^{-2}, & -4.1464936781966485 \times 10^{-2}, & -6.7372554722299874 \times 10^{-2}, \\ 3.8611006682309290 \times 10^{-1}, & 8.1272363544960613 \times 10^{-1}, & 4.1700518442377760 \times 10^{-1}, \\ -7.6488599078264594 \times 10^{-2}, & -5.9434418646471240 \times 10^{-2}, & 2.3680171946876750 \times 10^{-2}, \\ 5.6114348193659885 \times 10^{-3}, & -1.8232088709100992 \times 10^{-3}, & -7.2054944536811512 \times 10^{-4}\} \end{array} \tag{4}$$

The conjugate filter is $g_k = (-1)^{1-k} h_{11-k}$, shifted by 12, so that like h it is nonzero only at indices 0,...,11.

Similarly, wavelet packets are smooth generalized Walsh functions. Let $w_0 = \phi$ and $w_1 = \psi$ be the scaling function and mother wavelet, respectively, of an orthonormal wavelet basis, with operators H, G defined by CQFs h, g, and put

$$w_{2n}(x) = H w_n(x), \qquad w_{2n+1}(x) = G w_n(x), \tag{5}$$

for each $n = 1, 2, \ldots$. *Shannon wavelets and wavelet packets* can also be obtained by this recursion, if the condition that h and g be finitely supported is removed. Take

$$h_k = \frac{\sin\left[\frac{\pi}{2}(k-\frac{1}{2})\right]}{\frac{\pi}{2}(k-\frac{1}{2})}; \qquad g_k = (-1)^k h_{1-k} = (-1)^k \frac{\sin\left[\frac{\pi}{2}(k-\frac{1}{2})\right]}{\frac{\pi}{2}(k-\frac{1}{2})}, \tag{6}$$

to define H and G, and

$$\phi(x) = \frac{\sin\left[\pi(x-\frac{1}{2})\right]}{\pi(x-\frac{1}{2})}; \qquad \psi(x) = \frac{\sin\left[2\pi(x-\frac{1}{2})\right] - \sin\left[\pi(x-\frac{1}{2})\right]}{\pi(x-\frac{1}{2})}, \tag{7}$$

for the initial functions.

This overview paper will describe recent progress and open problems in four areas: detection of singularities in images, multiplication of functions given their wavelet aproximations; estimation of source coding efficiency, or compressibility, from wavelet packet coefficient distributions; and control of frequency spreading and convergence in wavelet packet bases.

2 Detecting Edge-like Singularities in Images

Many edge detection algorithms are known, broadly separated into *local* operations, such as discrete Laplacian or Sobel difference filtering, and *global* operations such as template matching that recognizes edge-like groups of pixels. A local operation based on approximate differentiation, either by finite differences as in the Sobel detector, or after transformation as in the Fourier and wavelet methods, assigns a recognizably large value at singular points of the image, and a small value at smooth points. It typically produces too many candidate edge points, which must then be screened for membership along some line or curve by a subsequent global operation. Candidate edge points are especially overabundant in noisy images.

This situation may be improved if we introduce a local operation that produces a large value at a point only if it and a few lined-up neighbors are singular points. Then both single-point singularites and nonsingular points of the image will produce small values. This will reduce the number of candidates to be checked by the global follow-up, especially in noisy images. The finer local step has complexity comparable to filtering or transformation, but the global part's complexity is reduced due to the smaller number of false edge points.

It is a classical observation that whenever a function is not smooth at some point, then the power in its Fourier transform localized near that point will be slowly decreasing at high frequencies. But then, if

the singularity has a direction, such as the normal direction to an edge discontinuity, the decrease will be particularly slow in that direction. This slow decrease creates a large variance in the slow direction, if we treat the local Fourier power spectrum as a probability density. By contrast, the variance in the other directions, in which the Fourier transform decreases rapidly, will be smaller. These variances are the two eigenvalues of the 2×2 autocovariance matrix, or equivalently the second-moment matrix, of the localized Fourier power spectrum.

One new technique, introduced in [7], is to recognize the edge-like nonsmooth points of a function by the differences between these two eigenvalues. The ratio or difference of the eigenvalues, in the limit as the localization shrinks to the point of interest in the continuum case, give the "edginess" of a function, with bigger differences or ratios giving more "edginess." Furthermore, the eigenvector of the larger eigenvalue will be normal to the edge, when such a normal exists.

The eigenvalues might be the same because they are small and equal, or large and equal. The first case arises at a point of smoothness, the second at a point singularity. Our technique assigns low "edginess" in both cases, and therefore differs from the differentiation-based edge detectors. Drawing a conclusion from two eigenvalues specializes our earlier work, in which we estimated the local rank of a complicated function from the number of relatively large eigenvalues of the autocovariance matrix.

To get localized information on the singularities of a function $f : \mathbf{R}^2 \to \mathbf{R}$, we will multiply it by a smooth cutoff function, or "bump," concentrated around the point of interest. This bump should be radial, to avoid introducing directional bias:

Definition 1 *Fix a nonzero radial function $g : \mathbf{R}^2 \to \mathbf{R}$ in the Schwartz class, centered at the origin, and fix $\epsilon > 0$. Then for each polynomially bounded function $f : \mathbf{R}^2 \to \mathbf{R}$, and each point $x^0 \in \mathbf{R}^2$, define the* dual local autocovariance matrix of f at x^0 *to be the 2×2 matrix*

$$E_{\epsilon,g}(f; x^0)_{ij} = \int_{B(0,1/\epsilon)} \xi_i \xi_j \left| \widehat{g_\epsilon f}(\xi) \right|^2 \, d\xi, \qquad i, j \in \{1, 2\},$$

where $g_\epsilon(\xi) = g\left(\dfrac{\xi - x^0}{\epsilon}\right)$.

By these assumptions, gf is integrable, so $\widehat{gf}$ is bounded and continuous and the matrix coefficients are well defined. It is the real, symmetric second moment matrix of the unnormalized probability density function $|\widehat{g_\epsilon f}|^2$. If f is nonzero in a neighborhood of x^0, then the matrix will be positive definite. Our technique is to use differences and ratios of its eigenvalues to define the "edginess" of the function f, at the point x^0.

The dual local autocovariance matrix can also be defined for certain singular measures and distributions. For example, let f be the Dirac delta measure supported at y^0, and let g_ϵ be centered at x^0 as in Definition 1. Then $|\widehat{g_\epsilon f}(\xi)|^2 = |g_\epsilon(y^0)|^2$, which tends to 0+ as $\epsilon \to 0$ if $x^0 \neq y^0$, but remains constantly 1 as $\epsilon \to 0$ if $x^0 = y^0$. In either case, $E_{\epsilon,g}(f; x^0)$ tends to a multiple of the identity as $\epsilon \to 0$, so both the ratio and the normalized difference of the eigenvalues is everywhere the same in the limit. Hence, either definition of edginess ignores point singularities.

The main result is that edge-like singularities are recognizable:

Theorem 1 *Suppose $D \subset \mathbf{R}^2$ is a domain with a smooth boundary, $x^0 \in \partial D$, and $g : \mathbf{R}^2 \to \mathbf{R}$ is a nonzero radial Schwartz function. Let H be either half-plane defined by the line tangent to ∂D at x^0. If we denote the eigenvalues of $E_{\epsilon,g}(\mathbf{1}_D; x^0)$ by $\lambda_1(\epsilon)$ and $\lambda_2(\epsilon)$, then $\liminf_{\epsilon \to 0+} |\lambda_1(\epsilon) - \lambda_2(\epsilon)| > 0$.*

The boundary smoothness assumption in Theorem 1 is not crucial. We can prove similar results for rougher ∂D, using the weaker Morrey-Campanato regularity assumption [17, 8].

On the other hand, at points where f is differentiable, the eigenvalues of its dual local autocovariance matrix are equal:

Theorem 2 *Suppose that $f : \mathbf{R}^2 \to \mathbf{R}$ is differentiable at x^0. Then for any smooth radial function $g : \mathbf{R}^2 \to \mathbf{R}$ of compact support, the matrix $E_{\epsilon,g}(f; x^0)$ converges to a multiple of the 2×2 identity matrix, as $\epsilon \to 0$.*

The converse to Theorem 2 is false: even if $\lim_{\epsilon\to 0+} |E_{\epsilon,g}(f;x^0)|$ is a multiple of the identity, we cannot conclude that f is differentiable at x^0, or even continuous. Sufficient symmetry can masquerade as smoothness, as the following example shows. Let $f(x) = \mathbf{1}_+(x_1)\mathbf{1}_+(x_2)$, where $\mathbf{1}_+$ is the characteristic function of $\mathbf{R}^+$, and fix $g(x) = \exp(-\pi|x|^2)$ as before. Then, after some straightforward calculations, we discover that $E_{\epsilon,g}(f;0)$ is a positive multiple of the identity, so $\lambda_1(\epsilon) = \lambda_2(\epsilon) = \lambda > 0$ for every ϵ, even though f is discontinuous at 0.

To calculate the autocovariance matrix in the discrete sampled case, we use the discrete Fourier transform on N real samples $\{f(n) : 0 \le n < N\}$, normalized as follows:

$$\hat{f}(k) = \sum_{n=0}^{N-1} \exp\left(-2\pi i \frac{kn}{N}\right) f(n), \qquad k \in B_N = \left[-\frac{N}{2}, \frac{N}{2}\right].$$

If only the first $q \ll N$ samples of f are nonzero, then the sum reduces to the smaller range $\{0, 1, \ldots, q-1\}$. Thus, the squared absolute value of $\hat{f}(k)$, when f is real-valued, is

$$|\hat{f}(k)|^2 = \sum_{n=0}^{q-1}\sum_{n'=0}^{q-1} \exp\left(-2\pi i \frac{k(n-n')}{N}\right) f(n)f(n'), \qquad k \in B_N.$$

The r-th moment of $|\hat{f}(k)|^2$ is therefore

$$\begin{aligned}\sum_{k\in B_N} k^r |\hat{f}(k)|^2 &= \sum_{k\in B_N}\sum_{n=0}^{q-1}\sum_{n'=0}^{q-1} k^r \exp\left(-2\pi i \frac{k(n-n')}{N}\right) f(n)f(n') \\ &= \sum_{n=0}^{q-1}\sum_{n'=0}^{q-1} f(n)f(n') \sum_{k\in B_N} k^r \exp\left(-2\pi i \frac{k(n-n')}{N}\right).\end{aligned}$$

The innermost sum in k is a function of the integer $n - n'$. Except for a factor of N^{r+1}, it is a Riemann approximation to the integral

$$\mu_r(n) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^r \exp(-2\pi i n x)\, dx,$$

evaluated at $n \leftarrow n - n'$, whose easily-computed values we shall use instead. Evidently $\mu_0(n) = 1$ if $n = 0$, but is zero otherwise, while

$$\mu_1(n) = \begin{cases} 0, & \text{if } n = 0, \\ i\frac{(-1)^n}{2\pi n}, & \text{otherwise;} \end{cases} \qquad \mu_2(n) = \begin{cases} 1/12, & \text{if } n = 0, \\ \frac{(-1)^n}{2\pi^2 n^2}, & \text{otherwise.} \end{cases}$$

We can apply the above results to analyze an image, which for our purposes will be a real-valued function supported on the rectangle $[0, M] \times [0, N] \subset \mathbf{R}^2$, sampled on a regular grid with grid point coordinates $\{(m, n) : 0 \le m < M;\ 0 \le n < N\}$. We will use a bump function supported on small subrectangles of size $p \times q$, rather than a dilated radial function, for g_ϵ. Translations of f have no effect on $|\hat{f}|^2$, so we may assume that the localized portion of the image has been translated to the subgrid $\{(m, n) : 0 \le m < p;\ 0 \le n < q\}$. The dual local autocovariance matrix may then be computed as follows:

$$\begin{aligned} E_{11} &= \sum_{m,m',n,n'} f(m,n)f(m',n')\mu_2(m-m')\mu_0(n-n') \qquad (8)\\ &= \sum_{m=0}^{p-1}\sum_{m'=0}^{p-1}\sum_{n=0}^{q-1} f(m,n)f(m',n)\mu_2(m-m'); \\ E_{22} &= \sum_{m=0}^{p-1}\sum_{n=0}^{q-1}\sum_{n'=0}^{q-1} f(m,n)f(m,n')\mu_2(n-n'); \qquad (9)\\ E_{12} = E_{21} &= \sum_{m=0}^{p-1}\sum_{m'=0}^{p-1}\sum_{n=0}^{q-1}\sum_{n'=0}^{q-1} f(m,n)f(m',n')\mu_1(m-m')\mu_1(n-n'). \qquad (10)\end{aligned}$$

Around each grid point x^0 of the image, we perform the following steps:

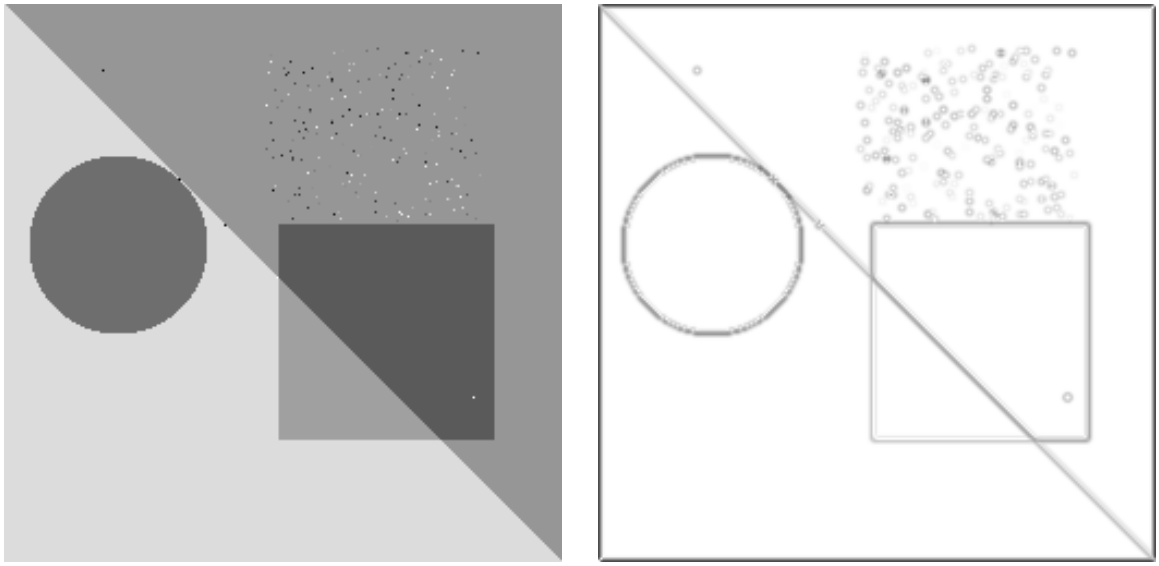

Figure 1: Geometrical figures: image, and edginess.

Localization. Extract the samples on the square subgrid $x^0 + [-\epsilon, \epsilon] \times [-\epsilon, \epsilon]$, where a small positive integer plays the role of ϵ in Definition 1. Then $p = q = 2\epsilon + 1$. The sample at x^0 becomes the sample at 0 in the $(2\epsilon + 1) \times (2\epsilon + 1)$ extracted subgrid $[-\epsilon, \epsilon]^2$. The sample at $x^0 + y$ is multiplied by the Gaussian bump function $\exp(-\pi|y|^2/\epsilon^2)$ and stored at y in the subgrid. This costs $O(\epsilon^2)$ operations per pixel.

If x^0 is within ϵ grid points of the boundary, then we simply pad any missing samples in the subgrid with zeros. For definiteness, we chose $\epsilon = 3$ to prepare our case studies.

Dual Autocovariance. Compute the 2×2 matrix $E = (E_{ij})$ using Equations 8,9, and 10. This always yields a real-valued, symmetric, positive semidefinite matrix. The computational complexity of the quadruple sum in Equation 10 dominates the triple sums of Equations 8 and 9, so this step costs $O(\epsilon^4)$ operations per pixel.

Eigenvalues. For symmetric 2×2 matrices E, the exact formula for eigenvalues is:

$$\lambda = \frac{1}{2}\left(E_{11} + E_{22} \pm \sqrt{(E_{11} - E_{22})^2 + 4E_{12}^2}\right), \tag{11}$$

where we take $+$ for λ_1 and $-$ for λ_2. These will satisfy $\lambda_1 \geq \lambda_2 \geq 0$, so in particular, if $\lambda_2 > 0$ we always have $\lambda_1/\lambda_2 \geq 1$. The greater the relative difference, the greater the edginess.

Edginess. We define this to be the ratio λ_1/λ_2. We actually compute the bounded reciprocal $\lambda_2/\lambda_1 \in [0, 1]$, amplified to fill the grayscale range of a write-black display device. That way, the darkest marks indicate the greatest edginess.

The Dual Autocovariance step dominates the computational complexity. It is therefore $O(p^2q^2)$ operations per pixel, if we localize to subgrids of $p \times q$ points.

The algorithm described above, localizing with Gaussian bumps restricted to 7×7 subgrids centered at x^0, is illustrated by its action on the various types of singularities in Figure 1. Software to produce these images is available from the author's web site, `http://www.math.wustl.edu/~victor`. There are piecewise constant functions with jump discontinuities along various rectifiable, mostly smooth curves, as well as point singularities of various magnitudes and a patch of independent, uniformly distributed noisy pixels.

3 Computation with Adapted Wavelet Bases

Speedups in numerical simulations have been obtained by representing solutions to complicated problems as superpositions of relatively few wavelet packets. This approximation scheme is nonlinear, keeping only a short series of those component functions with significant amplitudes; the others are discarded. Making this choice to minimize a description length or information cost criterion is called a *best basis* algorithm [6].

Computed simulations of fully-developed turbulence in the two-dimensional Navier–Stokes equation (2D-NSE) provide an example [11, 31]. 2D-NSE simulations on 10^4 to 10^6 grid points indicate that 10% of the components suffice for deterministic predictability for short, and 1% suffice statistical predictability such as estimates of the vorticity power spectrum. Turbulence simulations are thus an example "compressible" high-dimensional problem.

After reduction, all of the numerical computations are done in wavelet packet coordinates. Some of the known algorithms are matrix-vector and matrix-matrix multiplication [2, 27], numerical differentiation [20, 1, 30], multiplication [9, 23], and certain integral operators [12, 29]. The multiplication algorithm is key to solving nonlinear equations and is sketched here.

Given two functions approximable with short wavelet packet series, their sum is evidently approximable by another short wavelet series. If the wavelets are smooth and have vanishing moments, then the same is true for products. The short wavelet series representing their product may be found by pre-calculating the *connection coefficients* which express the product of two wavelets or scaling functions as a wavelet series. A method suggested by Daubechies and also used by Dahmen, *et al.*, allows rapid computation of these coefficients by matrix fixed-point iteration. The complexity of the multiplication algorithm is bounded by the number of nonnegligible connection coefficients.

Only compactly-supported wavelets of Daubechies and Mallat [10, 21] are considered here, rather than the more general wavelet packets. These are *refinable functions*, expressible as short linear combinations of dilated and translated versions of themselves. Refinable functions have a cross-scale self-similarity that can be used to compute integrals of their products, and thus to find connection coefficients.

The algebraic properties of refinable functions are well known, and have been heavily exploited in recent papers on wavelets and numerical analysis. Dahmen and Micchelli [9] considered the problem of evaluating integrals of products of refinable functions and their derivatives. Kunoth [18] later implemented the algorithms described in that paper. Latto, Resnikoff, and Tenenbaum [19] also derived a linear system of equations for connection coefficients involving two-scale equations for refinable functions.

Suppose that $\{e_k : k \in \mathbf{Z}\}$ is an orthonormal basis for $L^2(\mathbf{R})$ consisting of bounded functions. Then the triple product $e_j e_k e_l$ is defined and integrable over $\mathbf{R}$, so the abstract *connection coefficients* of this basis may be defined as the following integrals:

$$\Gamma_{jkl} = \langle e_j, e_k e_l \rangle \stackrel{\text{def}}{=} \int_{\mathbf{R}} \bar{e}_j(t)\, e_k(t) e_l(t)\, dt. \tag{12}$$

These coefficients are used to find the expansion of a product. If $u(t) = \sum_k u_k e_k(t)$ and $v(t) = \sum_l v_l e_l(t)$, then

$$u(t)v(t) = \sum_j \left(\sum_{k,l} \Gamma_{jkl} u_k v_l \right) e_j(t). \tag{13}$$

An example simpler than $L^2(\mathbf{R})$ is the space of sequences ℓ^2 with the *Kronecker basis* $e_k(n) = \delta(n-k)$, where $\delta(x)$ is the Kronecker symbol which is 1 if $x = 0$ and 0 otherwise. The inner product in Equation 12 is a sum rather than an integral, and we see that $\Gamma_{jkl} = \delta(j-k)\delta(j-l)$. The inner summation of Equation 13 simplifies into the pointwise multiplication formula $\sum_{k,l} \Gamma_{jkl} u_k v_l = u_j v_j$.

Another simple example is the Fourier basis $e_k(t) = e^{2\pi k t}$, $k \in \mathbf{Z}$, for $L^2([0,1])$. This basis is both orthonormal and closed under multiplication, so $\Gamma_{jkl} = \delta(k+l-j)$. A change of variables in the inner summation of Equation 13 gives the usual convolution formula $\sum_{k,l} \Gamma_{jkl} u_k v_l = \sum_k u_k v_{j-k}$.

A less simple example is the Haar basis of Equation 1. Its basis functions are indexed by a pair of integers, so their connection coefficients require six integer indices:

$$\Gamma^{str}_{nmk} \stackrel{\text{def}}{=} \int_{\mathbf{R}} \psi_{sn}(x) \psi_{tm}(x) \psi_{rk}(x)\, dx \tag{14}$$

The integral may be evaluated explicitly, since the Haar functions are almost closed under multiplication. The product of two Haar functions ψ_{sn} and ψ_{tm} is either zero (if their support intervals are disjoint), or $2^{-s}\mathbf{1}(2^{-s}x - n)$ (if they are equal, as when $s = t$ and $n = m$), or $\pm 2^{-\frac{s+t}{2}}\psi_{tm}$ (if their support intervals intersect, and $s > t$). Reordering so that $s \geq t \geq r$ gives

$$\Gamma^{str}_{nmk} = \begin{cases} 2^{-s/2}, & \text{if } s > t = r \text{ and } m = k \in 2^{s-t}[n, n+\frac{1}{2}) \stackrel{\text{def}}{=} \mathcal{L}(s-t, n); \\ -2^{-s/2}, & \text{if } s > t = r \text{ and } m = k \in 2^{s-t}[n+\frac{1}{2}, n+1) \stackrel{\text{def}}{=} \mathcal{R}(s-t, n); \\ 0, & \text{otherwise, with } s \geq t \geq r. \end{cases} \tag{15}$$

Now $n \in \mathcal{L}(a,j) \iff 2^{-a}n - \frac{1}{2} < j \leq 2^{-a}n$ and $n \in \mathcal{R}(a,j) \iff 2^{-a}n - 1 < j \leq 2^{-a}n - \frac{1}{2}$. For fixed n and $a > 0$ exactly one of these inequalities will have a solution $j \stackrel{\text{def}}{=} \mathcal{M}(a,n)$, and that solution will be unique. Putting $\mathcal{S}(a,n) = +1$ if $n \in \mathcal{L}(a, \mathcal{M}(a,n))$ and $\mathcal{S}(a,n) = -1$ if $n \in \mathcal{R}(a, \mathcal{M}(a,n))$ yields the following multiplication formula:

$$\begin{aligned} \sum_{t,m,r,k} \Gamma^{str}_{nmk} u_{tm} v_{rk} \quad &= \quad 2^{-s/2} \sum_{t=-\infty}^{s-1} \left[\sum_{m \in \mathcal{L}(s-t,n)} u_{tm} v_{tm} \; - \sum_{m \in \mathcal{R}(s-t,n)} u_{tm} v_{tm} \right] \\ &\quad + \sum_{t=s+1}^{\infty} 2^{-t/2} \mathcal{S}(t-s,n) \left[u_{t,\mathcal{M}(t-s,n)} v_{sn} \; + \; u_{sn} v_{t,\mathcal{M}(t-s,n)} \right]. \end{aligned} \tag{16}$$

Finally, consider orthonormal wavelets[10]. The two-scale equations produce the connection coefficients through iteration and filtering. As in the Haar case, there are six indices: Γ^{str}_{nmk}. One starts with the coefficients derived from the scaling functions:

$$A^{str}_{nmk} \stackrel{\text{def}}{=} \int_{\mathbf{R}} \phi_{sn}(x) \phi_{tm}(x) \phi_{rk}(x) \, dx \tag{17}$$

One may suppose without loss that $s \geq t \geq r$. Also, changing variables gives

$$A^{str}_{nmk} = 2^{-\frac{r}{2}} A^{s-r,t-r,0}_{nmk} = 2^{-\frac{r}{2}} A^{s-r,t-r,0}_{0,m-2^{s-t}n,k-2^{s-r}n}, \qquad \text{for } n,m,k \in \mathbf{Z} \text{ and } s \geq t \geq r \in \mathbf{Z}. \tag{18}$$

Thus, to obtain A^{str}_{nmk} it suffices to compute $A^{ij}_{mk} \stackrel{\text{def}}{=} A^{ij0}_{0mk}$ for $m,k \in \mathbf{Z}$ and $i \geq j \in \mathbf{Z}$. These values are themselves computed by fixed point iteration. For all triplets (n,m,k) of integers, define

$$A_{nmk} \stackrel{\text{def}}{=} A^{000}_{nmk} = \int_{\mathbf{R}} \phi(x-n) \phi(x-m) \phi(x-k) \, dx, \tag{19}$$

This quantity will vanish whenever the triplet is so large that the scaling function factors have disjoint support. It also satisfies $A_{nmk} = A_{n-k,m-k,0}$, so it suffices to compute the simpler quantity:

$$A(n,m) \stackrel{\text{def}}{=} A_{nm0} = \int_{\mathbf{R}} \phi(x-n) \phi(x-m) \phi(x) \, dx \tag{20}$$

But this matrix satisfies its own two-scale equation [23]:

Theorem 3 *Suppose that $h_k = 0$ unless $0 \leq k < L$. Then $A(n,m) = 0$ unless $-L < n < L$ and $-L < m < L$. Also, A satisfies the homogeneous fixed-point equation*

$$A(n,m) = \sum_{p,q} \alpha(p,q) A(2n-p, 2m-q), \qquad -L < n,m < L,$$

where

$$\alpha(p,q) \stackrel{\text{def}}{=} \sqrt{2} \sum_{k=0}^{L} h_{k-p} h_{k-q} h_k, \qquad -L < p,q < L.$$

The complete set of connection coefficients Γ^{str}_{nmk} may be obtained from the following numbers, which must be computed for all $m, k \in \mathbf{Z}$ and for all $i \geq j \in \mathbf{Z}$:

$$\Gamma^{ij}_{mk} \stackrel{\text{def}}{=} 2^{-\frac{i+j}{2}} \int_{\mathbf{R}} \psi(2^{-i}x)\psi(2^{-j}x - m)\psi(x - k)\, dx; \qquad \Gamma^{str}_{nmk} = 2^{-\frac{r}{2}} \Gamma^{s-r,t-r}_{m-2^{s-t}n,k-2^{s-r}n}. \tag{21}$$

$\Gamma^{ij}_{\cdots}$ comes from $A^{ij}_{\cdots}$ by filtering:

$$\Gamma^{ij}_{nmk} = \sqrt{2} \sum_{n',m',k'} g_{n'} g_{m'} g_{k'} A^{ij}_{2n+n',2m+m',2k+k'} \stackrel{\text{def}}{=} \sqrt{2} G_1 G_2 G_3 A^{ij}_{n,m,k}, \tag{22}$$

written as a separable filtering operation.

Likewise, $A^{ij}_{\cdots}$ comes from $A_{\cdots}$ by filtering. The two-scale equation for ϕ gives one step:

$$A^{ij}_{nmk} = \sum_{n'} h_{n'} A^{i-1,j}_{2n+n',m,k} \stackrel{\text{def}}{=} H_1 A^{i-1,j}_{n,m,k}. \tag{23}$$

Iterating i times in the first scale index and j times in the second gives $A^{ij}_{nmk} = H_1^i H_2^j A_{n,m,k}$, using the commuting operators

$$H_1 B(n,m,k) \stackrel{\text{def}}{=} \sum_{n'} h_{n'} B(2n + n', m, k); \qquad H_2 B(n,m,k) \stackrel{\text{def}}{=} \sum_{m'} h_{m'} B(n, 2m + m', k). \tag{24}$$

Combining the H and G operations yields Γ from A by filtering:

$$\Gamma^{ij}_{nmk} = \sqrt{2} G_1 G_2 G_3 H_1^i H_2^j A_{n,m,k}. \tag{25}$$

The C12 filters of Equation 4 provide a good example. They define the following α, whose entries are multiplied by 1000 and truncated to integers for display purposes:

0	0	−1	−1	0	0	−1	−1	−1	0	0	−1	0	0	0	0	0	0	0	0	0	0
0	0	−1	−1	0	0	−1	−1	0	0	−1	−1	0	0	0	0	0	0	0	0	0	0
−1	−1	0	0	−1	−1	0	0	0	0	0	−1	−1	0	0	0	0	0	0	0	0	0
−1	−1	0	0	−1	−1	0	0	0	−1	−1	−1	0	0	−1	0	0	0	0	0	0	0
0	0	−1	−1	0	0	−1	−2	−2	−2	−2	−1	−1	0	0	−1	0	0	0	0	0	0
0	0	−1	−1	0	−1	−1	−1	1	1	0	−1	−2	−1	0	−1	−1	0	0	0	0	0
−1	−1	0	0	−1	−1	3	8	8	10	14	9	0	−2	0	0	−1	0	0	0	0	0
−1	−1	0	0	−2	−1	8	7	−9	−21	−12	11	14	1	−2	0	0	0	0	0	0	0
−1	0	0	0	−2	1	8	−9	−45	−65	−76	−61	−12	10	1	−2	0	−1	0	−1	0	0
0	0	0	−1	−2	1	10	−21	−65	32	122	8	−76	−21	8	−1	−1	−1	−1	−1	0	0
0	−1	0	−1	−2	0	14	−12	−76	122	549	543	122	−65	−9	8	−1	0	−1	0	−1	0	0
−1	−1	−1	−1	−1	−1	9	11	−61	8	543	941	549	32	−45	7	3	−1	0	0	0	0	0
0	0	−1	0	−1	−2	0	14	−12	−76	122	549	543	122	−65	−9	8	−1	0	−1	0	−1	0
0	0	0	0	0	−1	−2	1	10	−21	−65	32	122	8	−76	−21	8	−1	−1	−1	−1	−1	−1
0	0	0	−1	0	0	0	−2	1	8	−9	−45	−65	−76	−61	−12	10	1	−2	0	−1	0	−1
0	0	0	0	−1	−1	0	0	−2	−1	8	7	−9	−21	−12	11	14	1	−2	0	0	0	0
0	0	0	0	0	−1	−1	0	0	−1	−1	3	8	8	10	14	9	0	−2	0	0	−1	0
0	0	0	0	0	0	0	0	−1	−1	0	−1	−1	−1	1	1	0	−1	−2	−1	0	−1	−1
0	0	0	0	0	0	0	0	0	−1	−1	0	0	−1	−2	−2	−2	−2	−1	−1	0	0	−1
0	0	0	0	0	0	0	0	−1	−1	0	0	−1	−1	0	0	0	−1	−1	−1	0	0	−1
0	0	0	0	0	0	0	0	0	−1	−1	0	0	−1	−1	0	0	0	0	0	−1	−1	0
0	0	0	0	0	0	0	0	0	0	0	0	−1	−1	0	0	−1	−1	0	0	−1	−1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	−1	−1	0	0	−1	−1	−1	0	0	−1

The origin $m = k = 0$ of $\alpha(m, k)$ is at the center, m increases downwards, and k increases to the right as in the convention for matrices. The corresponding fixed point A is plotted below in Figure 2.

After computing the kernel α of the fixed-point problem, iterate from the elementary double sequence $A(m, k) = 1 \iff m = k = 0$ until the maximum change per iteration in a coefficient of A falls below 10^{-6}. To get the other scaling and connection coefficients, apply the filter operators G_1, G_2, G_3, H_1, and H_2 as needed.

Each application of an operator G_1, G_2, G_3, H_1, or H_2 costs L operations per output coefficient. Unfortunately, the number of output coefficients grows with each application. For filters supported on $\{0, 1, \ldots, L-1\}$ and fixed $i, j \geq 0$, the matrices $A^{ij}(m, k)$ and $\Gamma^{ij}(m, k)$ will vanish outside $-L < m < 2^{i-j}L$ and $-L < k < 2^i L$. However, as seen in the graphs, many of the coefficients with indices in this range are negligible. The graphs show level lines of the logarithm of the absolute value of scaling and connection coefficients. The origin $m = k = 0$ is always at the center of the square. The graphs are oriented such that m increases to the right and k increases upwards as in the convention for xy plots in the first quadrant.

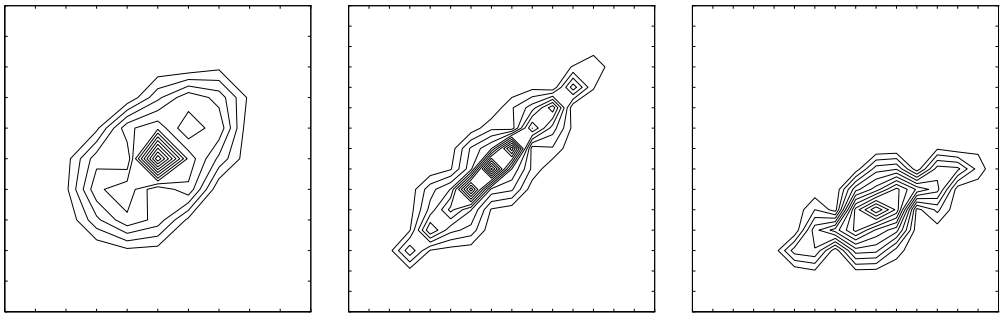

Figure 2: A, $A^{1,0}$ and $A^{1,1}$ for C12 filters.

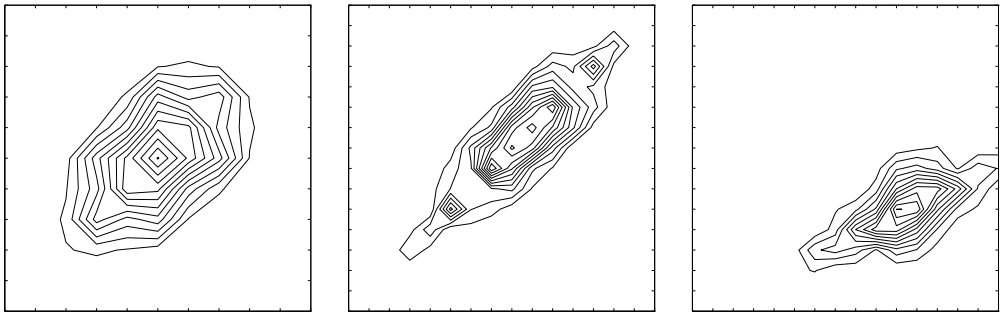

Figure 3: Γ, $\Gamma^{1,0}$ and $\Gamma^{1,1}$ for C12 filters.

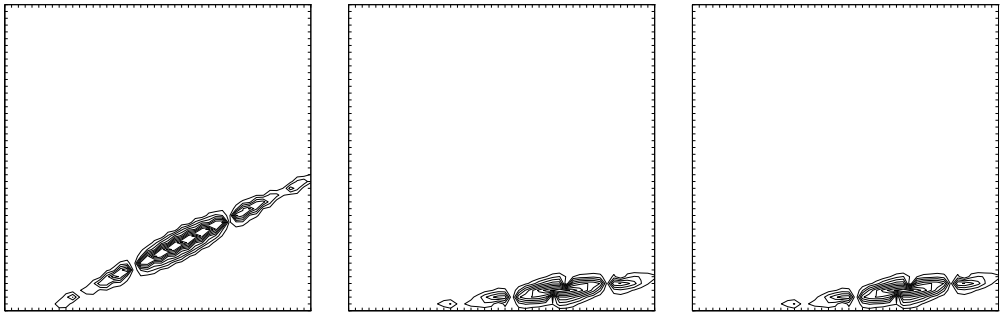

Figure 4: $A^{3,1}$, $A^{3,2}$, and $A^{3,3}$ for C12 filters.

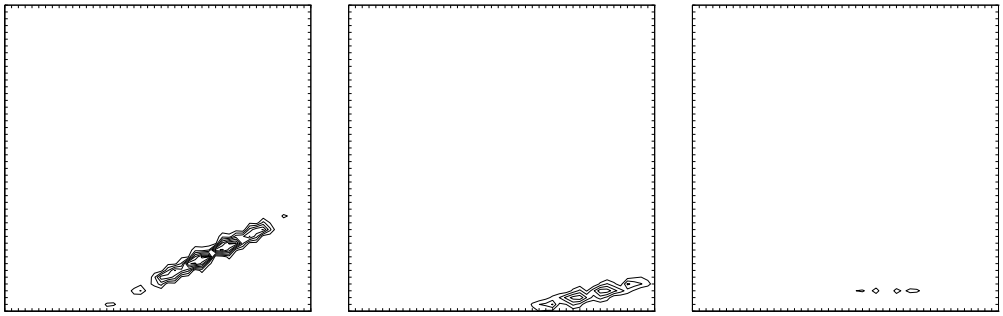

Figure 5: $\Gamma^{3,1}$, $\Gamma^{3,2}$, and $\Gamma^{3,3}$ for C12 filters.

4 Entropy of Wavelet Packet Coefficients

Having a choice of transforms for data compression suggests choosing one that optimizes coding efficiency. If the transform coefficients are independent random variables, then the Shannon–Weaver entropy[25] of the sequence determines the minimum bit rate needed to transmit them. This entropy cannot be calculated before the transform is chosen and all the coefficients are known, but if large coefficient values are very sparse, as is commonly observed in practice, then the entropy is equal to the logarithm of the *theoretical dimension* of the coefficient sequence, which can be computed on pieces of the signal to guide the transform choice. The result is a *best-basis* algorithm [6] that minimizes theoretical dimension over a library of transformations, choosing the transformation that yields best compression and also giving an estimate of the compression rate.

Transform output coefficients $\{x_m : m = 1, 2, \ldots\}$ are modeled by independent Bernoulli trials of a random variable with a fixed probability density function $\rho = \rho(t)$. For technical reasons, assume that ρ is continuous and strictly positive on $(0, 1)$. It defnes a *probability* $P\{E\} \stackrel{\text{def}}{=} \int_E \rho(t)\,dt$ for each measurable $E \subset [0, 1]$.

For fixed $1 \ll N < \infty$, *(uniform) quantization* to N values is defined by the formula $Q_N(x) \stackrel{\text{def}}{=} \lfloor Nx \rfloor / N$. If $x \in [0, 1)$ then $Q_N(x) \in \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}\}$. For coding or transmission, $\{x_n\}$ is replaced by a quantized version of itself, namely $\{Q_N(x_m) : m = 1, 2, \ldots\}$.

After quantization, the root-mean-square error, or *distortion*, per sequence element will have the following expected value:

$$D_N \stackrel{\text{def}}{=} \left(E\{|x_m - Q_N(x_m)|^2\}\right)^{1/2} = \left(\int_0^1 |t - Q_N(t)|^2 \rho(t)\,dt\right)^{1/2}. \tag{26}$$

Since the terms in the sequence are independent and identically distributed random variables, the distortion is independent of m. Each quantized value x_m will have the following discrete probability density function, independent of m:

$$p_n \stackrel{\text{def}}{=} P\left\{Q_N(x_m) = \frac{n-1}{N}\right\} = \int_{\frac{n-1}{N}}^{\frac{n}{N}} \rho(t)\,dt; \qquad n = 1, 2, \ldots, N. \tag{27}$$

Shannon's theorem [25] states that the expected number of bits per element required to encode this quantized sequence cannot be less than, but can be made arbitrarily close to, the entropy of the distribution, defined below:

$$H_N \stackrel{\text{def}}{=} -\sum_{n=1}^{N} p_n \log_2 p_n \tag{28}$$

Huffman coding [16] is one way to achieve this efficiency.

A simple rate-distortion curve for this combination of quantization and coding is obtained by plotting $10 \log_{10} D_N$ against H_N, so that the units are decibels of distortion versus bits per coefficient. The number of quantization intervals N parameterizes the curve. H_N and D_N can be estimated from ρ.

For H_N, write $p_n = \frac{1}{N}\rho(\xi_n)$ for some $\xi_n \in (\frac{n-1}{N}, \frac{n}{N})$. Then

$$H_N = -\sum_{n=1}^{N} \frac{1}{N}\rho(\xi_n) \log_2 \left[\frac{1}{N}\rho(\xi_n)\right] = \log_2 N - \sum_{n=1}^{N} \frac{1}{N}\rho(\xi_n) \log_2 \rho(\xi_n) \tag{29}$$

The second term is a Riemann sum approximating $-\int_0^1 \rho(t) \log_2 \rho(t)\,dt$, which may be called the *source entropy* $\mathcal{H}(\rho)$. The $\log_2 N$ term is present because at super fine quantizations the less significant digits contain most of the information even though they have almost no connection with ρ.

Now D_N, if ρ is continuous, has the following asymptotic behavior as $N \to \infty$:

$$\lim_{N\to\infty} N^2 D_N^2 = N^3 \int_0^{1/N} t^2\,dt = \frac{1}{3} \quad \Rightarrow \quad D_N \sim \frac{1}{\sqrt{3}\,N}. \tag{30}$$

Combining Equations 29 and 30 shows that

$$10\log_{10} D_N \sim -10\log_{10} N - 5\log_{10} 3 \sim -AH_N + B\mathcal{H}(\rho) - C, \qquad \text{as } N \to \infty, \tag{31}$$

where A, B, C are positive constants. Thus the rate-distortion curve is asymptotic to a line of negative slope with an intercept at $B\mathcal{H}(\rho) - C$. Shifting the curve to the left improves the rate-distortion relationship in the sense that the same transmission quality is obtained at a lower bit rate. Such a shift is accomplished by reducing $\mathcal{H}(\rho)$, or equivalently by transforming the sequence $\{x_m\}$ so that it appears to come from a lower-entropy source.

Fix $1 \ll M < \infty$, let $\{x_1, \ldots, x_M\}$ be a sequence of M Bernoulli trials of the random variable with density ρ, and let $\{x_1^*, \ldots, x_M^*\}$ be the decreasing rearrangement: $x_0^* \stackrel{\text{def}}{=} 1 \geq x_1^* \geq x_2^* \geq \cdots x_M^* \geq 0 \stackrel{\text{def}}{=} x_{M+1}^*$. Let $x^* = x^*(t)$ be defined on the interval $[0, 1]$ as follows:

$$x^*(t) = x_m^*, \quad \text{if } \tfrac{m-1}{M} < t \leq \tfrac{m}{M}; \qquad x^*(0) = 1. \tag{32}$$

This is a decreasing step function. The same sequence determines another step function as follows:

$$y(s) = \frac{m}{M}, \quad \text{if } x_{m+1}^* < s \leq x_m^*; \qquad y(1) = 0. \tag{33}$$

These two step functions y and x^* are approximate inverses: $y\left(x^*(t)\right) = Q_M(t)$, while $x^*\left(y(s)\right) = \max\{x_m^* : x_m^* \leq s\} \stackrel{\text{def}}{=} Q_x(s)$. Thus y inverts x^* up to the precision of the M-bin uniform quantization, while x^* inverts y up to the precision of the generally nonuniform quantization defined by the monotonic sequence $\{x_m^*\}$.

Now $y(s)$ is the fraction of values of $m \in \{1, 2, \ldots, M\}$ for which $x_m \geq s$. Since the x_m's are independent, this expectation depends only on ρ, not on M:

$$Ey(s) = \int_s^1 \rho(t)\, dt. \tag{34}$$

Thus $\frac{d}{ds} Ey(s) = -\rho(s)$, so $Ey(s)$ is strictly decreasing and continuously differentiable. Hence Ey has a differentiable inverse function $z = z(t)$, and the source entropy may be written in terms of z':

$$\mathcal{H}(\rho) = -\int_0^1 \rho(s)\log\rho(s)\, ds = \int_0^1 \left[\frac{1}{z'(Ey(s))}\right] \log\left[\frac{-1}{z'(Ey(s))}\right] ds = \int_0^1 \log\left[-z'(t)\right]\, dt. \tag{35}$$

But since y is the approximate inverse of x^*, while z is the inverse of Ey, one may approximate $z \approx x^*$ and

$$\int_0^1 \log\left[-z'(t)\right]\, dt \approx \sum_{m=1}^{M} \log\left[-\Delta x_m^*\right], \tag{36}$$

where $\Delta x_m^* \stackrel{\text{def}}{=} x_m^* - x_{m-1}^*$ for $m = 1, 2, \ldots, M$. If $x_1, \ldots, x_m$ are so concentrated near 0 that x^* decreases exponentially, then $-\Delta x_m^* \approx b x_m^*$ for some constant $b > 0$, and

$$\mathcal{H}(\rho) \approx M\log b + \sum_{m=1}^{M} \log|x_m^*| = M\log b + \sum_{m=1}^{M} \log|x_m|, \tag{37}$$

since the right-hand sum is independent of the order of summation.

Finally, note that $I(x) = \sum_{m=1}^{M} \log\left[x_m\right]$ is an additive information cost function [26]. In the best-basis method with I, a transform T is chosen for a signal u such that $I(Tu)$ is minimized. If the minimizing coefficient sequence $x = Tu$ has an exponentially decreasing rearrangement, then it will appear to come from a source whose entropy is approximately $I(x)$.

5 Wavelet Packet Spreading

Wavelet packets 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. One can obtain Schauder bases of uniformly bounded, uniformly compactly supported wavelet packets. By controlling size and support, one can apply the Carleson–Hunt theorem to show that certain wavelet packet Fourier series of a continuous function converges almost everywhere.

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:

$$w_n = F_{n_1} F_{n_2} \cdots F_{n_J} w_0, \tag{38}$$

where $2^{J-1} \le n < 2^J$ is written in binary as $n = \sum_{j=1}^{J} n_j 2^{j-1}$, $n_j \in \{0, 1\}$. 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 $2^{J-1} \le n < 2^J$ implies that $n_J = 1$.

But operators H and G also act as Fourier multipliers:

$$\hat{w}_{2n}(\xi) = \frac{1}{2} m_0\left(\frac{\xi}{2}\right) \hat{w}_n\left(\frac{\xi}{2}\right); \qquad \hat{w}_{2n+1}(\xi) = \frac{1}{2} m_1\left(\frac{\xi}{2}\right) \hat{w}_n\left(\frac{\xi}{2}\right), \tag{39}$$

where $m_0(\xi) = \sum_k h_k\, e^{-2\pi i k\xi}$ and $m_1(\xi) = \sum_k g_k\, e^{-2\pi i k\xi}$ are 1-periodic functions. They are trigonometric polynomials whenever h and g are finitely supported, as in the Walsh example where $m_0(\xi) = 1 + e^{2\pi i\xi} = 2e^{\pi i\xi} \cos \pi\xi$, and $m_1(\xi) = 1 - e^{2\pi i\xi} = -2ie^{\pi i\xi} \sin \pi\xi$. Hence, there is also a *multiplier formulation* of wavelet packets:

$$\hat{w}_n(\xi) = \frac{1}{2^J} \hat{w}_0(\tfrac{\xi}{2^J})\, m_{n_J}\left(\frac{\xi}{2^J}\right) m_{n_{J-1}}\left(\frac{\xi}{2^{J-1}}\right) \cdots m_{n_2}\left(\frac{\xi}{2^2}\right) m_{n_1}\left(\frac{\xi}{2}\right). \tag{40}$$

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 $d \approx N/5$ [10]. Since Daubechies' wavelets form an orthonormal basis, the associated wavelet packets $\{w_{n,k} : n \in \mathbf{N}, k \in \mathbf{Z}\}$ 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 [5]:

Theorem 4 *For any orthogonal CQFs (h, g) for which $m_0(\xi) \neq 0$ on $-\frac{\pi}{2} \le \xi \le \frac{\pi}{2}$, the wavelet packets $\{w_n\}$ satisfy*

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

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. This result was refined by M. Nielsen [22]:

Theorem 5 *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 $p = \infty$. In the $L = 4$ case, $p_{\min} = 2$ is the smallest possible value, and the same result holds for some other well-known CQFs, but the sharp lower bound is not known in general. 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.

N. Hess-Nielsen [14, 15] originally introduced the idea of building wavelet packet bases with more than one CQF pair in order 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. The algorithm is based on two generalizations of Equations 38 and 40. 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

$$w_n(x) = F^1_{n_1} F^2_{n_2} \cdots F^J_{n_J} w_0(x), \tag{41}$$

or alternatively, in the multiplier formulation, by

$$\hat{w}_n(x) = \frac{1}{2^J} w_0\left(\frac{\xi}{2^J}\right) m^J_{n_j}\left(\frac{\xi}{2^J}\right) \cdots m^2_{n_2}\left(\frac{\xi}{2^2}\right) m^1_{n_1}\left(\frac{\xi}{2}\right). \tag{42}$$

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. The associated transforms are as fast as ordinary wavelet packet transforms, but the new functions are better behaved. They may be designed to have the almost everywhere convergence property, or else to have uniform size, independent of frequency.

For example, (h^J, g^J) might be the Walsh CQF pair for all sufficiently large $J \geq J_0$. The resulting wavelet packets are called *Walsh-type*. Likewise, if (h^J, g^J) is the Shannon CQF pair of Equation 6 for all sufficiently large $J \geq J_0$, then the resulting wavelet packets are called *Shannon-type*. M. Nielsen [22] proved:

Theorem 6 *Both Walsh-type and Shannon-type wavelet packet series converge pointwise almost everywhere.*

These theorems are direct consequences of the Carleson–Hunt theorem [3] for Walsh series and Shannon series. Generalizing a result of Y. Meyer, M. Nielsen showed 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.

One can also use nonstationary wavelet packets to control the growth of $\|w_n\|_p$ for large p, as $n \to \infty$, using lengthening filters. the idea is to get a uniform bound from $\|w_n\|_\infty \leq \|\hat{w}_n\|_1$ by controlling frequency spreading $\|\hat{w}_n\|_1$. A. Cohen and E. Séré [4] showed the following:

Theorem 7 *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 [22] refined this result in the special case where h^J, g^J are the Daubechies orthogonal CQFs of length $L(J)$. It is necessary to redo the entire recursion for each new level. Let

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

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

$$w_n(x) = F^{J,1}_{n_1} F^{J,2}_{n_2} \cdots F^{J,J}_{n_J} w_0(x), \tag{43}$$

or alternatively, in the multiplier formulation, by

$$\hat{w}_n(x) = \frac{1}{2^J} w_0\left(\frac{\xi}{2^J}\right) m^{J,J}_{n_J}\left(\frac{\xi}{2^J}\right) \cdots m^{J,2}_{n_2}\left(\frac{\xi}{2^2}\right) m^{J,1}_{n_1}\left(\frac{\xi}{2}\right). \tag{44}$$

Here the superscripts indicate which pair of which sequence of CQFs defines the filter operator or multiplier. In fact, $h^{J,j} \stackrel{\text{def}}{=} h^J$ 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 basis, 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 [4]:

Theorem 8 *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 9 *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 [22]:

Theorem 10 *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.*

Acknowledgments

The author was supported in part by NSF grant DMS-0072234, "Adapted Wavelet Algorithms."

References

[1] Gregory Beylkin. On the representation of operators in bases of compactly supported wavelets. *SIAM Journal of Numerical Analysis*, 6-6:1716–1740, 1992.

[2] Gregory Beylkin, Ronald R. Coifman, and Vladimir Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.

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

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

[5] Ronald R. Coifman, Yves Meyer, and Mladen Victor Wickerhauser. Size properties of wavelet packets. In Ruskai et al. [24], pages 453–470.

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

[7] Wojciech Kladiusz Czaja and Mladen Victor Wickerhauser. Singularity detection in images using dual local autocovariance. *Applied and Computational Harmonic Analysis*, page 10, 2002. To appear.

[8] Wojciech Klaudiusz Czaja. *Applications of Local Autocovariance Matrices.* PhD thesis, Washington University, Saint Louis, Missouri, 2000.

[9] Wolfgang Dahmen and Charles A. Micchelli. Using the refinement equation for evaluating integrals of wavelets. *SIAM Journal of Numerical Analysis*, 30(2):507–537, April 1993.

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

[11] Marie Farge, Eric Goirand, Yves Meyer, Frédéric Pascal, and Mladen Victor Wickerhauser. Improved predictability of two-dimensional turbulent flows using wavelet packet compression. *Fluid Dynamics Research*, 10:229–250, 1992.

[12] Eric Goirand, Mladen Victor Wickerhauser, and Marie Farge. A parallel two dimensional wavelet packet transform and its application to matrix-vector multiplication. In Rodolphe L. Motard and Babu Joseph, editors, *Wavelet Applications in Chemical Engineering*, pages 275–319. Kluwer Academic Publishers, Norwell, Massachusetts, 1994.

[13] A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910.

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

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

[16] D. A. Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40:1098–1101, 1952.

[17] Svante Janson, Mitchell Taibleson, and Guido Weiss. Elementary characterizations of the Morrey-Campanato spaces. In Guido Weiss, Giancarlo Mauceri, and Fulvio Ricci, editors, *Proceedings of the July, 1982 Harmonic Analysis Conference Held in Cortona, Italy*, volume 992 of *Lecture Notes in Mathematics*, pages 101–114, New York, 1983. Springer-Verlag.

[18] Angela Kunoth. Computing refinable integrals, version 1.1. Available by anonymous FTP from ftp.igpm.rwth-aachen.de, 1995. C++ source code and documentation.

[19] Andy Latto, Howard L. Resnikoff, and Eric Tenenbaum. The evaluation of connection coefficients of compactly support wavelets. Preprint, Aware, Inc., Cambridge, Massachusetts, 1991.

[20] J. Liandrat, Valerie Perrier, and Philippe Tchamitchian. Numerical resolution of nonlinear partial differential equations using the wavelet approach. In Ruskai et al. [24], pages 227–238.

[21] Stéphane G. Mallat. A theory for multiresolution signal decomposition: The wavelet decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

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

[23] Valerie Perrier and Mladen Victor Wickerhauser. Multiplication of short wavelet series using connection coefficients. In Ka-Sing Lau, editor, *Advances in Wavelets*, pages 77–101. Springer-Verlag, Singapore, 1999.

[24] Mary Beth Ruskai, Gregory Beylkin, Ronald Coifman, Ingrid Daubechies, Stéphane Mallat, Yves Meyer, and Louise Raphael, editors. *Wavelets and Their Applications*. Jones and Bartlett, Boston, 1992.

[25] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, 1964.

[26] Hrvoje Šikić and Mladen Victor Wickerhauser. Information cost functions. *Applied and Computational Harmonic Analysis*, 11(2):147–166, September 2001.

[27] Mladen Victor Wickerhauser. An adapted waveform functional calculus. In Moody Chu, Robert Plemmons, David Brown, and Donald Ellison, editors, *Proceedings of the Cornelius Lanczos Centenary, Raleigh, North Carolina, 12–17 December 1993*, pages 418–421, Philadelphia, 1994. SIAM, SIAM Press.

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

[29] Mladen Victor Wickerhauser. Large-rank approximate principal component analysis with wavelets for signal feature discrimination and the inversion of complicated maps. *Journal of Chemical Information and Computer Science*, 34(5):1036–1046, September/October 1994.

[30] Mladen Victor Wickerhauser. Time localization techniques for wavelet transforms. *Croatica Chemica Acta*, 68(1):1–27, April 1995. Proceedings of the Ninth Dubrovnik International Course and Math-Chem-Comp 1994.

[31] Mladen Victor Wickerhauser, Marie Farge, Eric Goirand, Eva Wesfreid, and Echeyde Cubillo. Efficiency comparison of wavelet packet and adapted local cosine bases for compression of a two-dimensional turbulent flow. In Charles K. Chui, Laura Montefusco, and Luigia Puccio, editors, *Wavelets: Theory, Algorithms, and Applications*, Proceedings of the International Conference in Taormina, Sicily, 14–20 October 1993, pages 509–531. Academic Press, San Diego, California, 1994.