Progress in Wavelet Algorithms and Applications

M. Victor Wickerhauser

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

victor@math.wustl.edu
www.math.wustl.edu/~victor

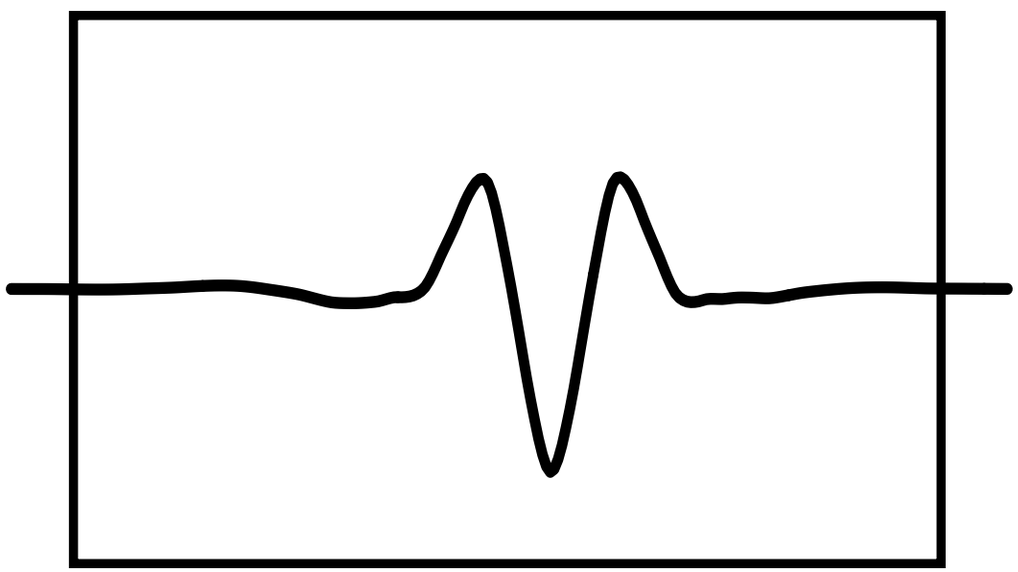
SPIE Orlando
April 4, 2002

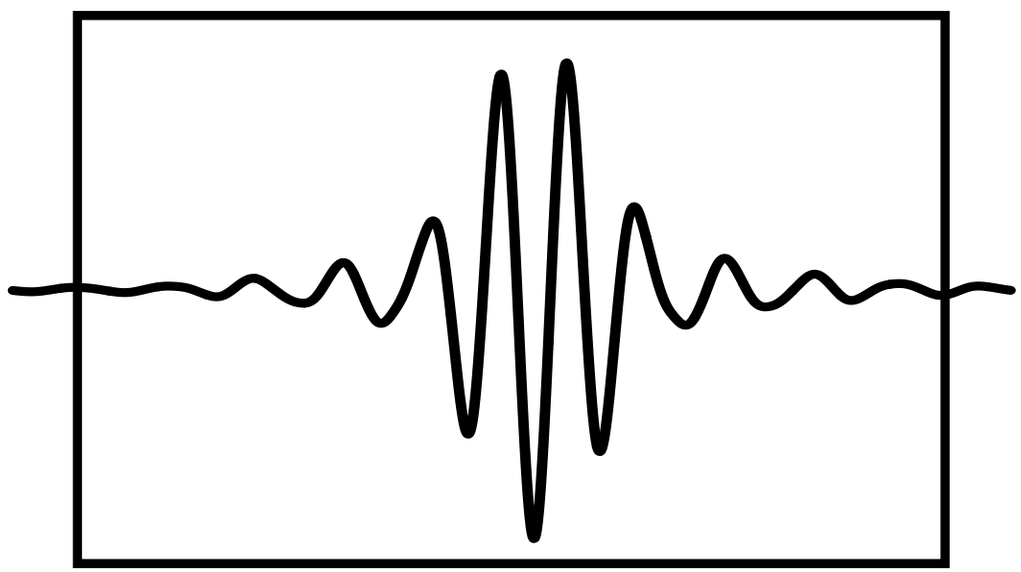
0. Definitions

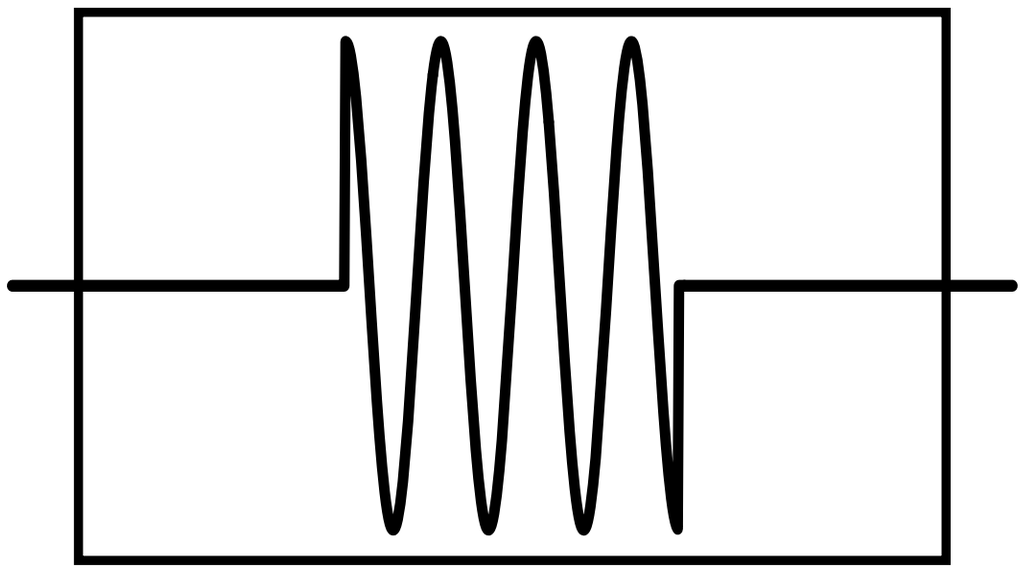
Wavelets and *wavelet packets* are special functions having three useful properties:

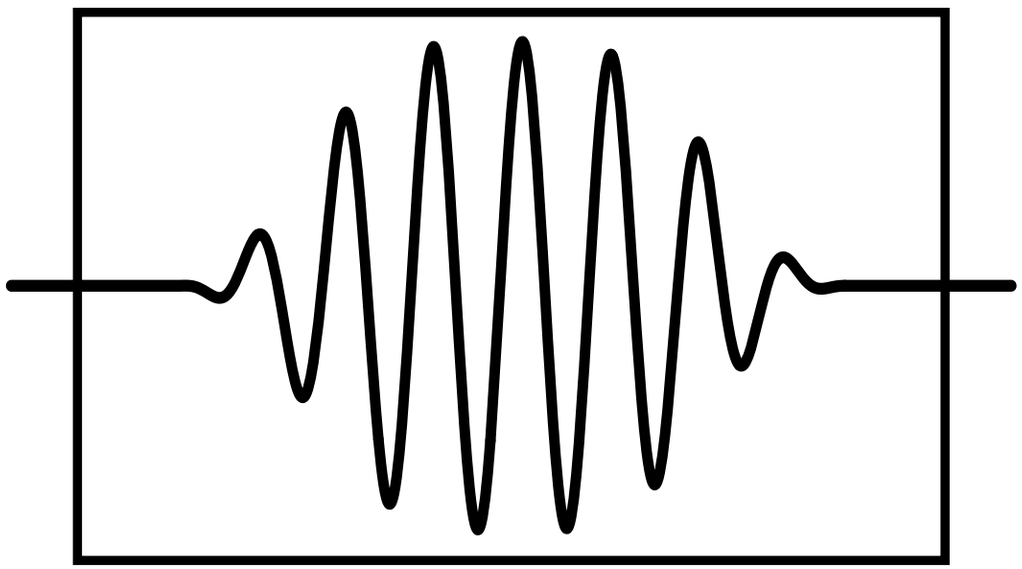
- They are almost as well localized in both time and frequency as the Heisenberg uncertainty inequality allows;
- They form orthogonal bases;
- They come equipped with fast, well-conditioned transforms: to compute N expansion coefficients of a function costs only $O(N \log N)$ operations.

Example waveforms









Almost-Examples

Almost-wavelets: the "mother" function $\psi = \psi(x) = \mathbf{1}(2x) - \mathbf{1}(2x-1)$ generates the Haar orthonormal basis for $L^2(\mathbf{R})$:

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

Almost-wavelet-packets: let $w_0 = \mathbf{1}$ 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 the Walsh orthonormal basis $\{w_{n,k} : n \in \mathbf{N}, k \in \mathbf{Z}\}$ for $L^2(\mathbf{R})$.

Actual Wavelets

Daubechies' and Mallat's generalization of Haar's basis uses the *two-scale equations*:

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

where $h = \{h_k : k \in \mathbf{Z}\}$ and $g_k = (-1)^{1-k} h_{1-k}$ are FIR filters satisfying:

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

for all integers n, m.

Sequences h, g satisfying these conditions are called *(orthogonal) conjugate quadrature filters*, or CQFs. Nontrivial example: "Coifman 12" (C12) filters. Here $h_k = 0$ if $k < 0$ or $k \geq 12$, and $\{h_k : 0 \leq k < 12\}$ is the following table of values:

$$\begin{array}{rrr}
\{1.6387336463179785 \times 10^{-2}, & -4.146493678196648 5 \times 10^{-2}, & -6.737255472229987 4 \times 10^{-2}, \\
3.8611006682309290 \times 10^{-1}, & 8.1272363544960613 \times 10^{-1}, & 4.1700518442377760 \times 10^{-1}, \\
-7.6488599078264594 \times 10^{-2}, & -5.9434418646471240 \times 10^{-2}, & 2.3680171946876750 \times 10^{-2}, \\
5.6114348193659885 \times 10^{-3}, & -1.8232088709100992 \times 10^{-3}, & -7.2054944536811512 \times 10^{-4}\}
\end{array}$$

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

Another Almost-Example

Shannon wavelets and wavelet packets use the same recursion, but with h and g having infinite support:

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

are used to define H and G, and

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

are the initial functions.

Actual Wavelet Packets

Smooth generalizations of Walsh functions have scaling function $w_0 = \phi$ and mother wavelet $w_1 = \psi$, respectively, of an orthonormal wavelet basis, with operators H, G defined by CQFs h, g. Put

$$w_{2n}(x) = Hw_n(x), \qquad w_{2n+1}(x) = Gw_n(x),$$

for each $n = 1, 2, \ldots$.

Relax the FIR condition on CQFs h, g and use the Shannon filters to get *Shannon wavelet packets*.

1. Computation

Addition: Short expansions give low-complexity componentwise addition.

Multiplication: Find connection coefficients that combine short expansions.

- Orthonormal basis $\{e_k : k \in \mathbf{Z}\}$, of bounded functions, for $L^2(\mathbf{R})$
- Triple product $e_j e_k e_l$ is defined and integrable over $\mathbf{R}$
- *connection coefficients* of the basis are integrals:

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

- Expansion of the product $u(t) = \sum_k u_k e_k(t)$ times $v(t) = \sum_l v_l e_l(t)$:

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

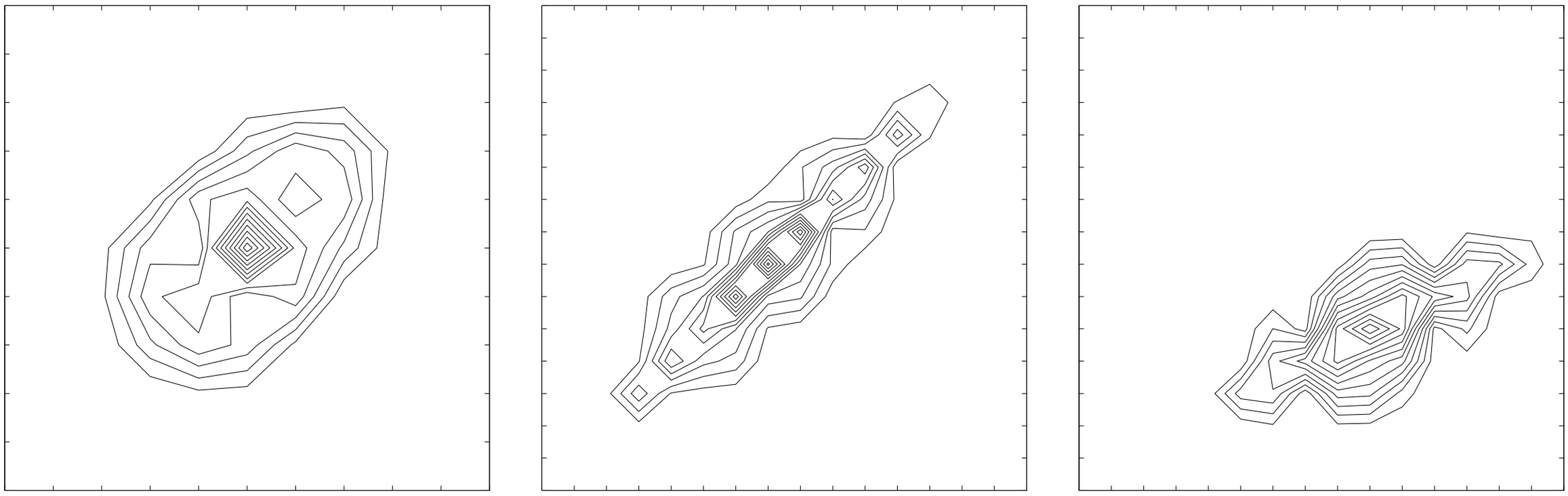
Low complexity requires rapid decrease of Γ.

Actual Connection Coefficients for C12 Filters

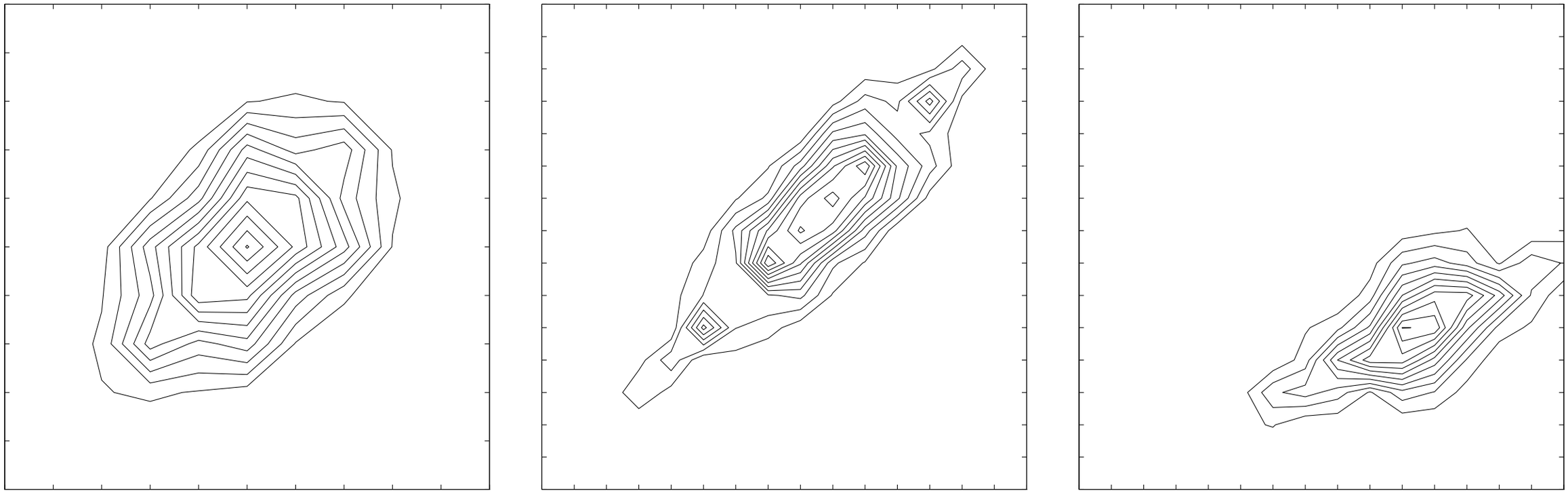
0	0	−1	−1	0	0	−1	−1	−1	0	0	−1	0	0	0	0	0	0	0	0	0	0	0
0	0	−1	−1	0	0	−1	−1	0	0	−1	−1	0	0	0	0	0	0	0	0	0	0	0
−1	−1	0	0	−1	−1	0	0	0	0	0	−1	−1	0	0	0	0	0	0	0	0	0	0
−1	−1	0	0	−1	−1	0	0	0	−1	−1	−1	0	0	−1	0	0	0	0	0	0	0	0
0	0	−1	−1	0	0	−1	−2	−2	−2	−2	−1	−1	0	0	−1	0	0	0	0	0	0	0
0	0	−1	−1	0	−1	−1	−1	1	1	0	−1	−2	−1	0	−1	−1	0	0	0	0	0	0
−1	−1	0	0	−1	−1	3	8	8	10	14	9	0	−2	0	0	−1	0	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	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	0	−1	−2	1	10	−21	−65	32	122	8	−76	−21	8	−1	−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

Mask $\alpha(i, j)$ entries are multiplied by 1000 and truncated to integers for display.

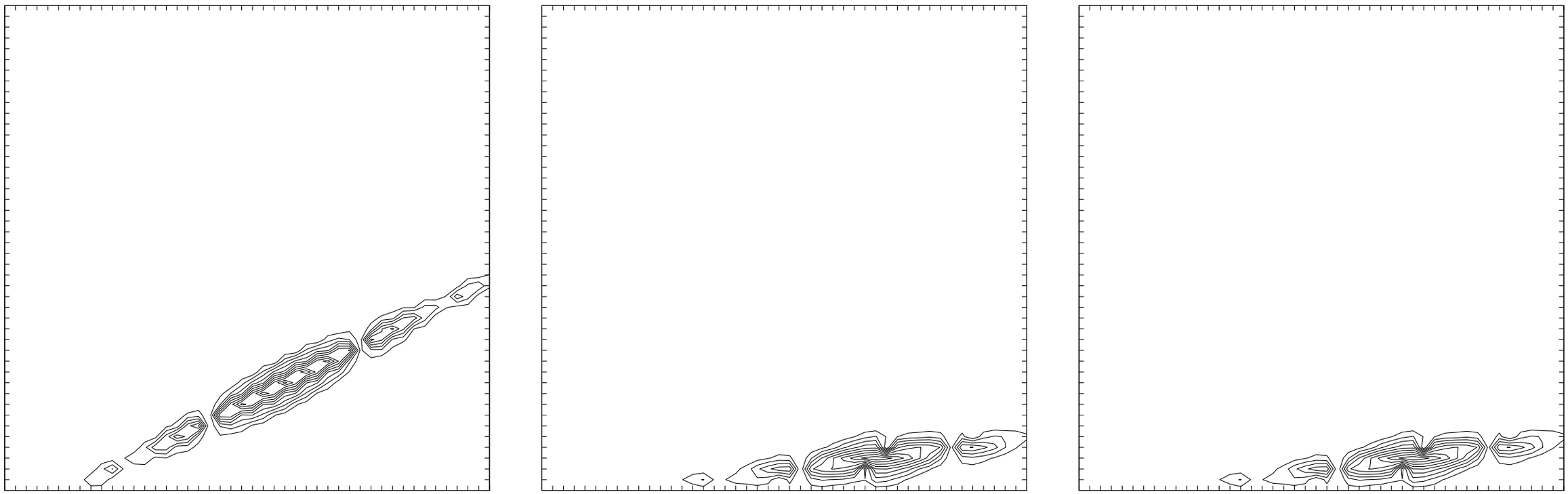
Δ, $\Delta^{1,0}$ and $\Delta^{1,1}$ for C12 filters



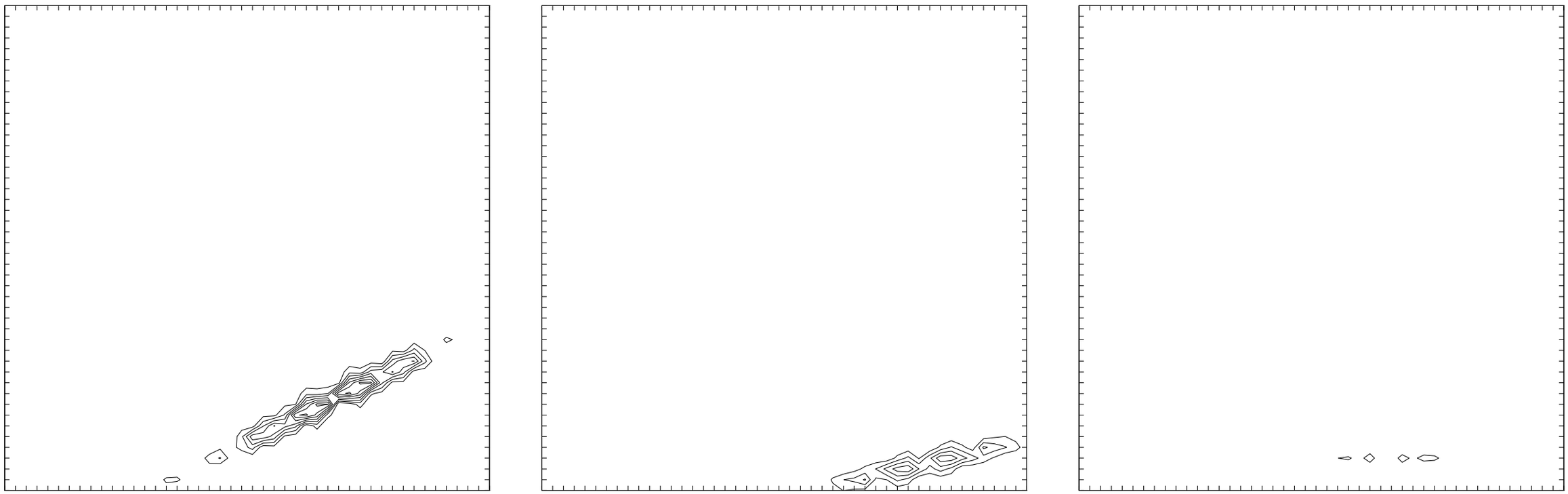
Γ, $\Gamma^{1,0}$ and $\Gamma^{1,1}$ for C12 filters



$\Delta^{3,1}$ $\Delta^{3,2}$ and $\Delta^{3,3}$ for C12 filters



$\Gamma^{3,1}$, $\Gamma^{3,2}$, and $\Gamma^{3,3}$ for C12 filters



2. Entropy of Coefficient Sequences

GOAL: choose a transform to optimize coding efficiency for data compression.

MODEL: transform coefficients are independent RVs, so Shannon–Weaver entropy determines minimum bit rate.

This S-W entropy cannot be calculated before the transform is chosen and all coefficients are known, so no divide-and-conquer method is available.

OBSERVATION: if large coefficients are very rare, then S-W entropy equals the logarithm of *theoretical dimension*, which can be computed on pieces.

BEST-BASIS ALGORITHM: minimize theoretical dimension over a library of transformations organized for divide-and-conquer.

Independent Components; Quantization

Transform coefficients $\{x_m : m = 1, 2, \ldots\}$ are i.i.d. RVs with pdf $\rho = \rho(t)$.

Technical assumption: ρ is continuous, $\rho > 0$ on $(0, 1)$.

Probability $P\{E\} \stackrel{\text{def}}{=} \int_E \rho(t)\, dt$ for each measurable $E \subset [0, 1]$.

Uniform quantization to N values, $1 \ll N < \infty$, is

$$Q_N(x) \stackrel{\text{def}}{=} \lfloor Nx \rfloor / N.$$

If $x \in [0, 1)$ then $Q_N(x) \in \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}\}$.

Rate versus Distortion

RMS *distortion* per sequence element has the expected value:

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

RVs $\{x_m\}$ i.i.d. implies that distortion is independent of m, so each x_m will have the following discrete pdf:

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

Bit rate, from Shannon's theorem:

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

Huffman coding is one way to achieve this efficiency.

Rate-distortion curve, for this combination of quantization and coding, is the plot of $10 \log_{10} D_N$ against H_N. Number of quantization intervals N parameterizes the curve.

Estimating Rate-Distortion

H_N and D_N can be estimated from ρ.

Continuous ρ implies that D_N has asymptotic behavior

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

Thus $\log D_N \sim -AH_N + B\mathcal{H}(\rho) - C$, as $N \to \infty$, where A, B, C are positive constants, for

$$\mathcal{H}(\rho) = -\int_0^1 \rho(s) \log \rho(s)\, ds.$$

Rate-distortion curve is asymptotically a line with intercept $B\mathcal{H}(\rho) - C$.

IDEA: Shifting the curve to the left by reducing $\mathcal{H}(\rho)$ gives the same transmission quality at a lower bit rate.

Estimating $\mathcal{H}(\rho)$ from Theoretical Dimension

Suppose $\{x_m : m = 1, \ldots, M\}$ is concentrated near 0.

- Let $\{1 = x_0^* \geq x_1^* \geq \cdots \geq x_M^*, x_{M+1}^* = 0\}$ be the decreasing rearrangement of $\{x_m\}$.
- Define a decreasing step function $x^* = x^*(t)$ on the interval $[0, 1]$ by

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

- Let $\Delta x_m^* \stackrel{\text{def}}{=} x_m^* - x_{m-1}^*$ for $m = 1, 2, \ldots, M$ be the decrements of x^*.

Then

$$\mathcal{H}(\rho) \approx \sum_{m=1}^{M} \log\left[-\Delta x_m^*\right],$$

If x^* decreases exponentially, then $-\Delta x_m^* \approx b x_m^*$ for some $b > 0$, so

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

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

$I(x) \stackrel{\text{def}}{=} \sum_{m=1}^{M} \log[x_m]$ is an *additive information cost function*.

3. Wavelet Packet Spreading

PROBLEM: In general, wavelet packets defined by a single filter pair have uncontrolled size and basis properties. For example, putting $m_0(\xi) = \sum_k h_k\, e^{-2\pi i k\xi}$, we have

Theorem 1 (Coifman-Meyer-MVW, 1992) *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*

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

The nonvanishing condition on m_0 is satisfied by Daubechies' filters. Also,

Theorem 2 (M. Nielsen,1998) *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* $p > p_{\min}$.

IDEA: Substitute different filters at different scales to control size.

RESULTS: Obtain Schauder bases of uniformly bounded, uniformly compactly supported wavelet packets. With controlled size and support, the Carleson–Hunt theorem applies to show that certain wavelet packet Fourier series of a continuous function converge almost everywhere.

Nonstationary Wavelet Packets

Define

$$w_n(x) = F_{n_1}^1 F_{n_2}^2 \cdots F_{n_J}^J w_0(x),$$

where the superscript indicates which CQF pair defines the filter operator. We change filters as frequency increases. The associated wavelet packets can be better behaved, with no additional computational complexity.

Example 1: (h^J, g^J) are Walsh CQFs for all sufficiently large $J \geq J_0$. Result: *Walsh-type* wavelet packets.

Example 2: (h^J, g^J) are Shannon CQFs for all sufficiently large $J \geq J_0$. Result: *Shannon-type* wavelet packets.

Theorem 3 (M. Nielsen,1999) *Both Walsh-type and Shannon-type wavelet packet series converge pointwise almost everywhere.*

Highly Nonstationary Wavelet Packets

We may 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 CQFs, fix w_0, and for $J \geq 2$ and $2^{J-1} \leq n < 2^J$ define *highly nonstationary* wavelet packets by

$$w_n(x) = F_{n_1}^{J,1} F_{n_2}^{J,2} \cdots F_{n_J}^{J,J} w_0(x).$$

For example, use Daubechies orthogonal CQFs h^J, g^J, of length $L(J)$, and put $h^{J,j} \stackrel{\text{def}}{=} h^J$ and $g^{J,j} \stackrel{\text{def}}{=} g^J$ for all $j = 1, 2, \ldots, J$.

Here w_0 is any scaling function that generates an orthonormal basis, not necessarily a Daubechies scaling function.

Uniformly Bounded and Supported Wavelet Packets

We suppose that w_0 is smooth enough so that $|\widehat{w}_0(\xi)| = O(1/|\xi|^{1+\epsilon})$ for some $\epsilon > 0$. Then

Theorem 4 (M. Nielsen, 2000) *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 $L^2(\mathbf{R})$.*

Theorem 5 (M. Nielsen, 2000) *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 nonstationary wavelet packet w_n grows without bound as $n \to \infty$, if $L(J) \to \infty$ as $J \to \infty$. Overcome this by introducing longer filters earlier:

Theorem 6 (M. Nielsen, 2001) *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.*