AN INTRODUCTION TO WAVELETS

or:

THE WAVELET TRANSFORM: WHAT'S IN IT FOR YOU?

Andrew E. Yagle and Byung-Jae Kwak
Dept. of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor MI

Presentation to Ford Motor Co.
May 21, 1996

OUTLINE OF PRESENTATION

1. Signal Representation using Orthonormal Bases
   - 1.1 Definitions and Properties
   - 1.2 Example: Fourier Series
   - 1.3 Example: Bandlimited Signals
   - 1.4 Example: Wavelet Transform
2. Multiresolution Analysis
   - 2.1 Multiresolution Subspaces
   - 2.2 Wavelet Scaling Functions
   - 2.3 Wavelet Basis Functions
   - 2.4 Summary of Wavelet Design
3. Wavelet Transforms
   - 3.1 Haar Function→Haar Transform
   - 3.2 Sinc Function→LP Wavelet
   - 3.3 Splines→Battle-Lemarie
   - 3.4 General Properties
   - 3.5 2-D Wavelet Transform
4. Applications of Wavelets
   - 4.1 Sparsification of Operators
   - 4.2 Compression of Signals
   - 4.3 Localized Denoising
   - 4.4 Tomography under Wavelet Constraints

1. SIGNAL REP. USING ORTHONORMAL BASES

1.1 Definitions

DEF: The (L^2) *inner product* of $x_1(t)$ and $x_2(t)$ is

$$< x_1(t), x_2(t) >= \int_{-\infty}^{\infty} x_1(t) x_2^*(t) dt$$

DEF: The (L^2) *norm* of $x(t)$ is

$$||x(t)|| = \sqrt{< x(t), x(t) >} = \sqrt{\int_{-\infty}^{\infty} x(t) x^*(t) dt}$$

DEF: $x(t) \in L^2$ if $||x(t)|| < \infty$

DEF: $x_i(t)$ and $x_j(t)$ are *orthonormal* if

$$< x_i(t), x_j(t) >= \delta_{i-j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}$$

DEF: The space *spanned* by basis funcs $\{\phi_i(t)\}$

$SPAN\{\phi_i(t)\} = \{\sum_i c_i \phi_i(t)\}$ for any constants c_i

DEF: A set of basis functions is *complete* if any L^2 function can be represented as a linear combination of the basis functions.

1.1 Properties

Let $\{\phi_i(t)\}$ be a complete orthonormal basis set. Then for any L^2 signal $x(t)$ we have:

$$x(t) = \sum_{i=-\infty}^{\infty} x_i \phi_i(t) \quad \textit{expansion}$$

$$x_i = \int_{-\infty}^{\infty} x(t)\phi_i^*(t)dt \quad \textit{orthonormal}$$

$$x_N(t) = \sum_{i=-N}^{N} x_i \phi_i(t) \rightarrow$$

$$\lim_{N\rightarrow\infty} ||x(t) - x_N(t)|| = 0 \quad \textit{complete}$$

1. Instead of processing the uncountably infinite $x(t)$, we may process the countably infinite x_i;
2. We may process the FINITE $\{x_i, |i| \leq N\}$ with arbitrarily small error if N sufficiently large;
3. $x_N(t)$ is the best approximation to $x(t)$ using only $2N + 1$ x_i;
4. Each x_i carries different information about $x(t)$– no redundancy.
5. Easy to compute x_i and update $x_N(t)$.

1.2 Example: Fourier Series

Let $x(t)$ be defined for $0 < t < P$ (periodic?)

Then $x(t)$ has the orthonormal expansion

$$x(t) = \sum_{n=-\infty}^{\infty} x_n e^{j2\pi nt/P}$$

$$x_n = \frac{1}{P} \int_0^P x(t) e^{-j2\pi nt/P}$$

1. Convergence in L^2 norm; Gibbs phenomenon at discontinuities;
2. Note basis functions orthogonal in SCALE.

Approximation of square wave using $N = 15$: (uncountably infinite $\rightarrow$ 15)

1.3 Example: Bandlimited Signals

Let $x(t)$ be bandlimited to $|\omega| < B$.
Then $x(t)$ has the orthonormal expansion

$$x(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin B(t-n\Delta)}{B(t-n\Delta)}, \quad \Delta = \frac{\pi}{B}$$

$$x_n = \int_{-\infty}^{\infty} x(t) \frac{\sin B(t-n\Delta)}{B(t-n\Delta)} \frac{1}{\Delta} dt = x(t = n\Delta)$$

1. Since $x(t)$ and $sinc(t)$ are both bandlimited $x(t) * sinc(t) = x(t); \quad sinc(t) * sinc(t) = sinc(t)$ Then sample;

2. Here x_n happen to be sampled values of $x(t)$;
3. Basis functions orthogonal in TRANSLATION:

$$\int_{-\infty}^{\infty} \frac{\sin B(t-i\Delta)}{B(t-i\Delta)} \frac{\sin B(t-j\Delta)}{B(t-j\Delta)} dt = \Delta \delta_{i-j}$$

1.4 Example: Wavelet Transform

Let $x(t) \in L^2$.

Then $x(t)$ has the orthonormal expansion

$$x(t) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} x_j^i 2^{-i/2} \psi(2^{-i}t - j)$$

$$x_j^i = \int_{-\infty}^{\infty} x(t) 2^{-i/2} \psi(2^{-i}t - j)$$

1. Note double-indexed x_j^i;
2. Basis functions areorthogonal in both SCALE AND TRANSLATION;
3. Therefore localized in time and frequency;
4. Note scale-dependent shift in t: $2^i j$
5. Need $2^{-i/2}$ to make $||2^{-i/2}\psi(2^{-i}t - j)|| = 1$.

How do we find such basis functions?

2. MULTIRESOLUTION ANALYSIS

2.1 Multiresolution Subspaces

DEF: A *multiresolution analysis* is a sequence of closed subspaces V_i such that:

1. $\{0\} \subset \ldots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset \ldots \subset L^2$
2. $x(t) \in V_i \Leftrightarrow x(2^i t) \in V_0$
3. $x(t) \in V_0 \Leftrightarrow x(t-j) \in V_0$
4. $\exists$orthonormal basis so $V_0 = SPAN\{\phi(t-j)\}$
   Latter implies $V_i = SPAN\{2^{-i/2}\phi(2^{-i}t-j)\}$

Example: Piecewise constant functions

1. Approximate function by piecewise constant steps
2. Smaller steps $\rightarrow$ more accurate approximation
3. Step size 2^i $\rightarrow$ project onto V_i (want small i)

2.2 Wavelet Scaling Functions

$$V_i = SPAN\{2^{-i/2}\phi(2^{-i}t - j)\}$$

$\phi(t)$ is the wavelet *scaling* function.

$$\phi(t) \in V_0 \subset V_{-1} \rightarrow \phi(t) \in V_{-1}$$

$$\rightarrow \phi(t) = \sum_{n=-\infty}^{\infty} g_n 2^{1/2} \phi(2t - n)$$

Taking Fourier transform and $G(e^{j\omega}) = DTFT[g_n] \rightarrow$ (for MEs: set backshift $q = e^{j\omega}$)

$$\Phi(\omega) = \frac{1}{\sqrt{2}} G(e^{j\omega/2}) \Phi(\omega/2)$$

Orthonormality of $\{\phi(t - j)\} \rightarrow$

$$||G(e^{j\omega})||^2 + ||G(e^{j(\omega+\pi)})||^2 = 2$$

Suggests g_n should be lowpass filter.

2.3 Wavelet Basis Functions

Define $\{W_i\}$ as orthogonal complements of $\{V_i\}$:

1. W_i=orthogonal complement of V_i in V_{i-1}
2. $V_{i-1} = V_i \bigoplus W_i$, $\bigoplus$=direct sum

$$\ldots \bigoplus W_1 \bigoplus W_0 \bigoplus W_{-1} \bigoplus \ldots = L^2$$

1. Still have $x(t) \in W_i \Leftrightarrow x(2^i t) \in W_0$
2. Need basis $\psi(t)$ so $W_0 = SPAN\{\psi(t-j)\}$
3. Compare the above to $V_0 = SPAN\{\phi(t-j)\}$

$$\psi(t) \in W_0 \subset V_{-1} \rightarrow \psi(t) \in V_{-1}$$

$$\rightarrow \psi(t) = \sum_{n=-\infty}^{\infty} h_n 2^{1/2} \phi(2t-n)$$

Taking Fourier transform and $H(e^{j\omega}) = DTFT[h_n] \rightarrow$

$$\Psi(\omega) = \frac{1}{\sqrt{2}} H(e^{j\omega/2}) \Phi(\omega/2)$$

$\psi(t) \in W_0 \perp V_0 \rightarrow \psi(t) \perp \{\phi(t-j)\} \rightarrow$

$$H(e^{j\omega})G^*(e^{j\omega}) + H(e^{j(\omega+\pi)})G^*(e^{j(\omega+\pi)}) = 0$$

2.3 Wavelet Basis Functions, continued

This leads to $h_n = g_{1-n}(-1)^n$

1. Now have $\psi(t)$ in terms of $\phi(t)$
2. $W_i = SPAN\{2^{-i/2}\psi(2^{-i}t - j)\}$
3. But $\{W_i\}$ (unlike $\{V_i\}$) are orthogonal spaces
4. So $\{2^{-i/2}\psi(2^{-i}t - j)\}$ is a complete orthonormal basis for L^2.

Note that g_n is a *lowpass filter* while

$$h_n = g_{1-n}(-1)^n \rightarrow H(e^{j\omega}) = -e^{-j\omega}G^*(e^{j(\omega+\pi)})$$

is a *bandpass filter*.

This shows the distinction between

1. wavelet scaling ($\phi(t)$) and
2. wavelet basis ($\psi(t)$) functions:

Scaled $\psi(t)$ sweep out different frequency bands.

2.4 Summary of Wavelet Design

1. Find scaling function $\phi(t)$, often by iterating 2-scale eqn

$$\phi(t) = \sum_{n=-\infty}^{\infty} g_n 2^{1/2} \phi(2t - n)$$

   a. Start with $\phi(t)$=LPF
   b. Repeatedly convolve and upsample by 2
   c. Converges if g_n *regular* filter: $\prod_{k=1}^{\infty} G(e^{j\omega/2^k})$ converges
   d. $\phi(t)$=cont. limit discrete filter g_n: upsampling spreads spectrum so that it is periodic with $2^k\pi$, k=iteration #

2. Find wavelet function $\psi(t)$:

$$\psi(t) = \sum_{n=-\infty}^{\infty} h_n 2^{1/2} \phi(2t - n)$$

$$\Psi(\omega) = \frac{1}{\sqrt{2}} H(e^{j\omega/2}) \Phi(\omega/2)$$

$$H(e^{j\omega}) = -e^{-j\omega} G^*(e^{j(\omega+\pi)})$$

3. WAVELET TRANSFORMS

3.1 Haar Function→Haar Transform

1. $V_i = \{\text{piecewise constant functions}\}$
2. $x(t)$ changes levels at $t = j2^i$
3. SMALLER $i \rightarrow$ better resolution

Consider DECREASE in resolution:

In going from V_i to coarser $V_{i+1} \subset V_i$:

1. Replace $x(t)$ over the two intervals $[2^i j, 2^i(j+1)]$ and $[2^i(j+1), 2^i(j+2)]$
2. with $x(t)$ over interval $[2^{i+1}(j/2), 2^{i+1}(j/2+1)]$
3. Note that $[2^{i+1}(j/2), 2^{i+1}(j/2+1)]$ $= [2^i j, 2^i(j+1)] \bigcup [2^i(j+1), 2^i(j+2)]$

How do we do this?

Replace $x(t)$ with its *average*
(which is its projection onto V_{i+1})
and with its *difference*
(which is its projection onto W_{i+1})
over the longer interval.

3.1 Haar Function→Haar Transform: Example

$$x(t) = \sin(t)$$

Projection onto V_i:

$$x_i(t) = \begin{cases} \sin(2^i j), & \text{if } 2^i j < t < 2^i(j+1); \\ \sin(2^i(j+1)), & \text{if } 2^i(j+1) < t < 2^i(j+2); \end{cases}$$

Projection onto V_{i+1}:
$x_{i+1}(t) = [\sin(2^i j) + \sin(2^i(j+1))]/2$,
if $2^{i+1}(j/2) < t < 2^{i+1}(j/2+1)$
Same interval; replace 2 values with average.

Projection onto W_{i+1}:
$x'_{i+1}(t) = [\sin(2^i j) + \sin(2^i(j+1))]/2$,
if $2^{i+1}(j/2) < t < 2^{i+1}(j/2+1)$
Replace 2 values with difference.

3.1 Haar Function→Haar Transform: Bases

$$\phi(t) = \begin{cases} 1, & \text{if } 0 < t < 1; \\ 0, & \text{otherwise.} \end{cases}$$

$$\phi(t) = \phi(2t) + \phi(2t-1) \rightarrow g_0 = g_1 = 1$$

$$h_n = g_{1-n}(-1)^n \rightarrow h_0 = 1, h_1 = -1 \rightarrow$$

$$\psi(t) = \phi(2t) - \phi(2t-1) = \begin{cases} 1, & \text{if } 0 < t < 1/2; \\ -1, & \text{if } 1/2 < t < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Scaling function $\phi(t)$ **Basis function** $\psi(t)$

3.2 Sinc Function→LP Wavelet

1. $V_0 = \{$functions bandlimited to $[-\pi, \pi]$
2. $V_{-i} = \{$functions bandlimited to $[-2^i\pi, 2^i\pi]$
3. $W_{-i} = \{$functions bandlimited to $[-2^{i+1}\pi, -2^i\pi] \bigcup [2^i\pi, 2^{i+1}\pi]$

Note V_i=space of *lowpass* signals
while W_i=space of *bandpass* signals

$\phi(t) = \frac{\sin(\pi t)}{\pi t}$ (lowpass)
$\psi(t) = \frac{\sin(\pi t/2)}{\pi t/2} \cos(3\pi t/2)$ (bandpass)

$$G(e^{j\omega}) = \begin{cases} \sqrt{2}, & \text{if } |\omega| < \frac{\pi}{2}; \\ 0, & \text{if otherwise.} \end{cases}$$

$$H(e^{j\omega}) = \begin{cases} -\sqrt{2}e^{-j\omega}, & \text{if } \frac{\pi}{2} < |\omega| < \pi; \\ 0, & \text{if otherwise.} \end{cases}$$

Scaling function $\phi(t)$ Basis function $\psi(t)$

3.2 Sinc Function→LP Wavelet, continued

1. LP=Littlewood-Paley wavelet
2. Octave-band decomposition
3. Constant-Q filtering

3.3 Splines→Battle-Lemarie

DEF: *Splines* are polynomials over finite intervals with continuous derivatives at interval ends.

DEF: V_i^k = {piecewise polynomial functions of degree k with boundaries at $t = 2^i j$ and $k-1$ continuous derivatives at those boundaries}.

DEF: *B-Splines* are convolutions of Haar scaling function with itself. They and their translations are (non-orthogonal) bases for V_0^k.

1. Since B-splines not orthogonal, Gram-Schmidt
2. Go through 2-scale equation, get:
3. Battle-Lemarie wavelet basis function
4. $\lim_{k\to\infty} \phi(t) = sinc$ →L-P wavelet
   $\lim_{k\to 0} \phi(t) = Haar$ →Haar wavelet

Thus Battle-Lemarie wavelet becomes Haar and L-P wavelets in extreme cases.

Scaling function $\phi(t)$ **Basis function** $\psi(t)$

3.4 General Properties of Wavelets

1. $\phi(t)$ is lowpass filter
   $\psi(t)$ is bandpass filter
   Sometimes need BOTH (e.g., DC for L-P)
2. Localized in both time and frequency:
   a. x_j^i has information about $x(t)$ for $t \approx 2^i j$ or $j \approx 2^{-i} t$
   b. Since $\psi(t)$ is bandpass, x_j^i has information about $X(\omega)$ for $\omega \approx 2^{-i}$center frequency
3. Sampling rate is scale-dependent:

4. $\Psi(0) = 0$ always; often also zero *moments*:

$$\int_{-\infty}^{\infty} t^k \psi(t) dt = \frac{d^k \Psi}{d\omega^k}|_{\omega=0} = 0$$

Battle-Lemarie: based on k^{th}-order spline.
→ first $k+1$ moments zero.

3.5 2-D Wavelet Transform

Just perform wavelet transform in both variables:

$$f(x,y) = \sum\sum f_{j,n}^{i,m} 2^{-i/2}\psi(2^{-i}x - j) 2^{-m/2}\psi(2^{-m}y - n)$$

$$f_{j,n}^{i,m} = \int\int f(x,y) 2^{-i/2}\psi(2^{-i}x - j) 2^{-m/2}\psi(2^{-m}y - n) dx\, dy$$

There is another decomposition that is more useful:

$$x(t) = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} x_j^i 2^{-i/2}\psi(2^{-i}t - j) =$$

$$\sum_{j=-\infty}^{\infty} c_j^N 2^{-N/2}\phi(2^{-N}t - j) + \sum_{i=-\infty}^{N}\sum_{j=-\infty}^{\infty} x_j^i 2^{-i/2}\psi(2^{-i}t - j)$$

where

$$x_j^i = \int_{-\infty}^{\infty} x(t) 2^{-i/2}\psi(2^{-i}t - j)$$

$$c_j^N = \int_{-\infty}^{\infty} x(t) 2^{-N/2}\phi(2^{-N}t - j)$$

The scaling function at scale N replaces the effects of all basis functions at scales coarser (larger) than N.

3.5 2-D Wavelet Transform, continued

In 2-D this becomes

$$f(x,y) = \sum_{j=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f_{j,n}^{N} 2^{-N} \phi(2^{-N}x - j)\phi(2^{-N}y - n)$$

$$+ \sum_{i=-\infty}^{N} \sum_{j=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f_{j,n}^{i,(1)} 2^{-i} \psi(2^{-i}x - j)\phi(2^{-i}y - n)$$

$$+ f_{j,n}^{i,(2)} \phi(2^{-i}x - j)\psi(2^{-i}y - n) + f_{j,n}^{i,(3)} \psi(2^{-i}x - j)\psi(2^{-i}y - n)$$

where

$$f_{j,n}^{N} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) 2^{-N} \phi(2^{-N}x - j)\phi(2^{-N}y - n)$$

$$f_{j,n}^{i,(1)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) 2^{-i} \psi(2^{-i}x - j)\phi(2^{-i}y - n) dx\, dy$$

$$f_{j,n}^{i,(2)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) 2^{-i} \phi(2^{-i}x - j)\psi(2^{-i}y - n) dx\, dy$$

$$f_{j,n}^{i,(3)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) 2^{-i} \psi(2^{-i}x - j)\psi(2^{-i}y - n) dx\, dy$$

Note products of scaling and basis functions are

1. bandpass×lowpass
2. lowpass×bandpass
3. bandpass×bandpass

4. APPLICATIONS OF WAVELETS

4.1 Sparsification of Operators

Problem: Solve the integral equation

$$g(x) = \int_a^b h(x,y)u(y)dy, \quad a < x < b$$

Examples of Applications:

1. Deconvolution
2. kernel=electromagnetic Green's function
3. Laplace equation in free space:
   $h(x,y) = h(|x-y|) = h(r)$
   a. 2-D: $= -\frac{1}{2\pi}\log r$
   b. 3-D: $= \frac{1}{4\pi r}$

Galerkin's method (project onto bases)$\rightarrow$

$$g_j^i = \sum_m \sum_n h_{j,n}^{i,m} u_n^m$$

FACT: Using Battle-Lemarie wavelets makes the block matrix $h_{j,n}^{i,m}$ *sparse*:

4.1 Sparsification of Operators, continued

Why does this happen?
$h(x,y)$ is *Calderon-Zygmund* operator if

$$|\frac{\partial^k}{\partial x^k}h(x,y)|+|\frac{\partial^k}{\partial y^k}h(x,y)|<\frac{C_k}{|x-y|^{k+1}}$$

Many Green's functions are Calderon-Zygmund.

1. If use wavelets as Galerkin basis function
2. and if first k moments of basis function=0 (e.g., Battle-Lemarie wavelets)
3. Then for $|j-n|>2k$ we have

$$|h_{j,n}^{i,(1)}|+|h_{j,n}^{i,(2)}|+|h_{j,n}^{i,(3)}|<\frac{C_k}{1+|j-n|^{k+1}}$$

1. Dropoff from main diagonal as $|j-n|^{-(k+1)}$
2. Faster than discretization$\rightarrow |j-n|^{-1}$
3. Matrix sparse $\rightarrow$ faster iterative algorithms.

Why does this happen?

1. First k moments zero$\rightarrow \Psi(\omega)\approx\omega^k$
2. So wavelet transform$\approx k^{th}$ derivative.
3. Wavelet basis function≈differentiatiors.

$h(x,y)$ Calderon-Zygmund$\rightarrow$
k^{th} derivative drops off as $|x-y|^{k+1}$.

4.2 Compression of Signals

IDEA: Represent signals using fewer numbers

1. Simplifies processing
   (e.g., operator sparsification)
2. Greatly reduces storage
   (especially important for images)

What advantages to using wavelets?

1. Signals have slow-varying parts (regions) and fast-varying parts (edges)
2. Fast-varying part ANYWHERE in signal
   $\rightarrow$high Fourier frequencies EVERYWHERE
   $\rightarrow$sample finely everywhere$\rightarrow$many numbers
3. Fast-varying parts using local bases and slow-varying parts using global bases
4. Use many numbers for fast-varying parts, BUT ONLY LOCALLY.
   Use few numbers for slow-varying parts.

4.3 Localized Denoising

1. Can perform "localized lowpass filtering":
   a. Where signal is LOCALLY smooth, filter→reduce noise
   b. Where signal is LOCALLY rapidly changing (edges), accept noise
   c. Threshold wavelet coefficients
   d. Very attractive for images
2. Contrast to Wiener filtering:
   a. Fourier→spatially invariant
   b. Must do same filtering everywhere
   c. Trade off denoising & smoothing edges
3. Localized lowpass filtering:
   a. ≈Wiener filtering in space and scale
   b. If edge present, high-resolution coeff$\neq 0$
   c. If edge absent, high-resolution coeff$= 0$
4. Use threshold (1-bit Wiener filter):
   a. If above threshold, keep coefficient
   b. If below threshold, set coefficient to zero.
   c. Eliminates noise.

4.4 Tomography under Wavelet Constraints

B. Sahiner and A.E. Yagle

Since wavelets are basis functions,
can incorporate localized LPF
INTO the reconstruction procedure:

Problem: Reconstruct image $\mu(x, y)$
from projections $p(r, \theta)$

$$p(r,\theta) = \int\int \mu(x,y)\delta(r - x\cos\theta - y\sin\theta)dx\,dy$$

Interested in high-resolution features (such as edges, calcifications) which are localized.

Don't want them smoothed.

Procedure:

1. Reconstruct noisy $WT\{\mu(x, y)\}$
2. Threshold $WT\{\mu(x, y)\}$
3. Impose constraints: some $WT\{\mu(x, y)\} = 0$
4. Solve constrained problem

Advantage: Improves reconstruction OUTSIDE constrained region!

Reason: $\mathcal{R}$ not unitary operator.