An Introduction to the Wavelet Analysis of Time Series

Don Percival

Applied Physics Lab, University of Washington, Seattle

Dept. of Statistics, University of Washington, Seattle

MathSoft, Inc., Seattle

Overview

- wavelets are analysis tools mainly for
  - time series analysis (focus of this tutorial)
  - image analysis (will not cover)
- as a subject, wavelets are
  - relatively new (1983 to present)
  - synthesis of many new/old ideas
  - keyword in $10,558+$ articles & books since 1989 (2000+ in the last year alone)
- broadly speaking, have been two waves of wavelets
  - continuous wavelet transform (1983 and on)
  - discrete wavelet transform (1988 and on)

Game Plan

- introduce subject via CWT
- describe DWT and its main 'products'
  - multiresolution analysis (additive decomposition)
  - analysis of variance ('power' decomposition)
- describe selected uses for DWT
  - wavelet variance (related to Allan variance)
  - decorrelation of fractionally differenced processes (closely related to power law processes)
  - signal extraction (denoising)

What is a Wavelet?

- wavelet is a 'small wave' (sinusoids are 'big waves')
- real-valued $\psi(t)$ is a wavelet if
  1. integral of $\psi(t)$ is zero: $\int_{-\infty}^{\infty} \psi(t)\, dt = 0$
  2. integral of $\psi^2(t)$ is unity: $\int_{-\infty}^{\infty} \psi^2(t)\, dt = 1$ (called 'unit energy' property)
- wavelets so defined deserve their name because
  - #2 says we have, for every small $\epsilon > 0$,
  $$\int_{-T}^{T} \psi^2(t)\, dt < 1 - \epsilon,$$
  for some finite T (might be quite large!)
  - length of $[-T, T]$ small compare to $[-\infty, \infty]$
  - #2 says $\psi(t)$ must be nonzero somewhere
  - #1 says $\psi(t)$ balances itself above/below 0
- Fig. 1: three wavelets
- Fig. 2: examples of complex-valued wavelets

Basics of Wavelet Analysis: I

- wavelets tell us about variations in local averages
- to quantify this description, let $x(t)$ be a 'signal'
  - real-valued function of t
  - will refer to t as time (but can be, e.g., depth)
- consider average value of $x(t)$ over $[a, b]$:

$$\frac{1}{b-a}\int_a^b x(u)\,du \equiv \alpha(a, b)$$

- reparameterize in terms of λ & t

$$A(\lambda, t) \equiv \alpha(t - \tfrac{\lambda}{2}, t + \tfrac{\lambda}{2}) = \frac{1}{\lambda}\int_{t-\frac{\lambda}{2}}^{t+\frac{\lambda}{2}} x(u)\,du$$

  - $\lambda \equiv b - a$ is called scale
  - $t = (a + b)/2$ is center time of interval
- $A(\lambda, t)$ is average value of $x(t)$ over scale λ at t

Basics of Wavelet Analysis: II

- average values of signals are of wide-spread interest
  - hourly rainfall rates
  - monthly mean sea surface temperatures
  - yearly average temperatures over central England
  - etc., etc., etc. (Rogers & Hammerstein, 1951)
- Fig. 3: fractional frequency deviates in clock 571
  - can regard as averages of form $[t - \frac{1}{2}, t + \frac{1}{2}]$
  - t is measured in days (one measurment per day)
  - plot shows $A(1, t)$ versus integer t
  - $A(1, t) = 0 \Rightarrow$ master clock & 571 agree perfectly
  - $A(1, t) < 0 \Rightarrow$ clock 571 is losing time
  - can easily correct if $A(1, t)$ constant
  - quality of clock related to changes in $A(1, t)$

Basics of Wavelet Analysis: III

- can quantify changes in $A(1,t)$ via
$$\begin{aligned} D(1,t-\tfrac{1}{2}) &\equiv A(1,t) - A(1,t-1) \\ &= \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} x(u)\,du - \int_{t-\frac{3}{2}}^{t-\frac{1}{2}} x(u)\,du, \end{aligned}$$
or, equivalently,
$$\begin{aligned} D(1,t) &= A(1,t+\tfrac{1}{2}) - A(1,t-\tfrac{1}{2}) \\ &= \int_{t}^{t+1} x(u)\,du - \int_{t-1}^{t} x(u)\,du \end{aligned}$$
- generalizing to scales other than unity yields
$$\begin{aligned} D(\lambda,t) &\equiv A(\lambda,t+\tfrac{\lambda}{2}) - A(\lambda,t-\tfrac{\lambda}{2}) \\ &= \frac{1}{\lambda}\int_{t}^{t+\lambda} x(u)\,du - \frac{1}{\lambda}\int_{t-\lambda}^{t} x(u)\,du \end{aligned}$$
- $D(\lambda,t)$ often of more interest than $A(\lambda,t)$
- can connect to Haar wavelet: write
$$D(\lambda,t) = \int_{-\infty}^{\infty} \tilde{\psi}_{\lambda,t}(u)x(u)\,du$$
with
$$\tilde{\psi}_{\lambda,t}(u) \equiv \begin{cases} -1/\lambda, & t-\lambda \le u < t; \\ 1/\lambda, & t \le u < t+\lambda; \\ 0, & \text{otherwise.} \end{cases}$$

Basics of Wavelet Analysis: IV

- specialize to case $\lambda = 1$ and $t = 0$:

$$\tilde{\psi}_{1,0}(u) \equiv \begin{cases} -1, & -1 \le u < 0; \\ 1, & 0 \le u < 1; \\ 0, & \text{otherwise.} \end{cases}$$

  comparison to $\psi^{\mathrm{H}}(u)$ yields $\tilde{\psi}_{1,0}(u) = \sqrt{2}\psi^{\mathrm{H}}(u)$

- Haar wavelet mines out info on difference between unit scale averages at $t = 0$ via

$$\int_{-\infty}^{\infty} \psi^{\mathrm{H}}(u)x(u)\,du \equiv W^{\mathrm{H}}(1,0)$$

- to mine out info at other t's, just shift $\psi^{\mathrm{H}}(u)$:

$$\psi^{\mathrm{H}}_{1,t}(u) \equiv \psi^{\mathrm{H}}(u-t); \text{ i.e., } \psi^{\mathrm{H}}_{1,t}(u) = \begin{cases} -\frac{1}{\sqrt{2}}, & t-1 \le u < t; \\ \frac{1}{\sqrt{2}}, & t \le u < t+1; \\ 0, & \text{otherwise} \end{cases}$$

  Fig. 4: top row of plots

- to mine out info about other λ's, form

$$\psi^{\mathrm{H}}_{\lambda,t}(u) \equiv \frac{1}{\sqrt{\lambda}}\psi^{\mathrm{H}}\left(\frac{u-t}{\lambda}\right) = \begin{cases} -\frac{1}{\sqrt{2\lambda}}, & t-\lambda \le u < t; \\ \frac{1}{\sqrt{2\lambda}}, & t \le u < t+\lambda; \\ 0, & \text{otherwise.} \end{cases}$$

  Fig. 4: bottom row of plots

Basics of Wavelet Analysis: V

- can check that $\psi^{\mathrm{H}}_{\lambda,t}(u)$ is a wavelet for all λ & t
- use $\psi^{\mathrm{H}}_{\lambda,t}(u)$ to obtain

  $$W^{\mathrm{H}}(\lambda,t) \equiv \int_{-\infty}^{\infty} \psi^{\mathrm{H}}_{\lambda,t}(u)x(u)\,du \propto D(\lambda,t)$$

  left-hand side is Haar CWT
- can do the same with other wavelets:

  $$W(\lambda,t) \equiv \int_{-\infty}^{\infty} \psi_{\lambda,t}(u)x(u)\,du, \text{ where } \psi_{\lambda,t}(u) \equiv \frac{1}{\sqrt{\lambda}}\psi\left(\frac{u-t}{\lambda}\right)$$

  left-hand side is CWT based on $\psi(u)$
- interpretation for $\psi^{\mathrm{fdG}}(u)$ and $\psi^{\mathrm{Mh}}(u)$ (Fig. 1): differences of adjacent weighted averages

Basics of Wavelet Analysis: VI

- basic CWT result: if $\psi(u)$ satisfies admissibility condition, can recover $x(t)$ from its CWT:

$$x(t) = \frac{1}{C_\psi} \int_0^\infty \left[\int_{-\infty}^\infty W(\lambda, t) \frac{1}{\sqrt{\lambda}} \psi\left(\frac{t-u}{\lambda}\right) du \right] \frac{d\lambda}{\lambda^2},$$

  where C_ψ is constant depending just on ψ

- conclusion: $W(\lambda, t)$ equivalent to $x(t)$
- can also show that

$$\int_{-\infty}^\infty x^2(t)\, dt = \frac{1}{C_\psi} \left[\int_0^\infty \int_{-\infty}^\infty W^2(\lambda, t)\, dt \right] \frac{d\lambda}{\lambda^2}$$

  - LHS called energy in $x(t)$
  - RHS integrand is energy density over λ & t
- Fig. 3: Mexican hat CWT of clock 571 data

Beyond the CWT: the DWT

- critique: have transformed signal into an image
- can often get by with subsamples of $W(\lambda, t)$
- leads to notion of discrete wavelet transform (DWT)
  - can regard as dyadic 'slices' through CWT
  - can further subsample slices at various t's
- DWT has appeal in its own right
  - most time series are sampled as discrete values (can be tricky to implement CWT)
  - can formulate as orthonormal transform (facilitates statistical analysis)
  - approximately decorrelates certain time series (including power law processes)
  - standardization to dyadic scales often adequate
  - can be faster than the fast Fourier transform!
- will concentrate on DWT for remainder of tutorial

Overview of DWT

- let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ be observed time series (for convenience, assume N integer multiple of 2^{J_0})
- let $\mathcal{W}$ be $N \times N$ orthonormal DWT matrix
- $\mathbf{W} = \mathcal{W}\mathbf{X}$ is vector of DWT coefficients
- orthonormality says $\mathbf{X} = \mathcal{W}^T\mathbf{W}$, so $\mathbf{X} \Leftrightarrow \mathbf{W}$
- can partition $\mathbf{W}$ as follows:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_{J_0} \\ \mathbf{V}_{J_0} \end{bmatrix}$$

- $\mathbf{W}_j$ contains $N_j = N/2^j$ wavelet coefficients
  - related to changes of averages at scale $\tau_j = 2^{j-1}$ (τ_j is jth 'dyadic' scale)
  - related to times spaced 2^j units apart
- $\mathbf{V}_{J_0}$ contains $N_{J_0} = N/2^{J_0}$ scaling coefficients
  - related to averages at scale $\lambda_{J_0} = 2^{J_0}$
  - related to times spaced 2^{J_0} units apart

Example: Haar DWT

- Fig. 5: $\mathcal{W}$ for Haar DWT with $N = 16$
  - first 8 rows yield $\mathbf{W}_1 \propto$ *changes* on scale 1
  - next 4 rows yield $\mathbf{W}_2 \propto$ *changes* on scale 2
  - next 2 rows yield $\mathbf{W}_3 \propto$ *changes* on scale 4
  - next to last row yields $\mathbf{W}_4 \propto$ *change* on scale 8
  - last row yields $\mathbf{V}_4 \propto$ *average* on scale 16
- Fig. 6: Haar DWT coefficients for clock 571

DWT in Terms of Filters

- filter $X_0, X_1, \ldots, X_{N-1}$ to obtain

  $$2^{j/2}\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1$$

  where $h_{j,l}$ is jth level wavelet filter

  – note: circular filtering

- subsample to obtain wavelet coefficients:

  $$W_{j,t} = 2^{j/2}\widetilde{W}_{j,2^j(t+1)-1}, \quad t = 0, 1, \ldots, N_j - 1,$$

  where $W_{j,t}$ is tth element of $\mathbf{W}_j$

- Figs. 7 & 8: Haar, D(4), C(6) & LA(8) wavelet filters
- jth wavelet filter is band-pass with pass-band $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$
- note: jth scale related to interval of frequencies
- similarly, scaling filters yield $\mathbf{V}_{J_0}$
- Figs. 9 & 10: Haar, D(4), C(6) & LA(8) scaling filters
- J_0th scaling filter is low-pass with pass-band $[0, \frac{1}{2^{J_0+1}}]$

Pyramid Algorithm: I

- can formulate DWT via 'pyramid algorithm'
  - elegant iterative algorithm for computing DWT
  - implicitly *defines* $\mathcal{W}$
  - computes $\mathbf{W} = \mathcal{W}\mathbf{X}$ using $O(N)$ multiplications
    * 'brute force' method uses $O(N^2)$
    * FFT algorithm uses $O(N \log_2 N)$
- algorithm makes use of two basic filters
  - wavelet filter h_l of unit scale $h_l \equiv h_{1,l}$
  - associated scaling filter g_l

The Wavelet Filter: I

- let $h_l, l = 0, \ldots, L-1$, be a real-valued filter
  - L is filter width so $h_0 \neq 0$ & $h_{L-1} \neq 0$
  - L must be even
  - assume $h_l = 0$ for $l < 0$ & $l \geq L$
- h_l called a wavelet filter if it has these 3 properties
  1. summation to zero:
  $$\sum_{l=0}^{L-1} h_l = 0$$
  2. unit energy:
  $$\sum_{l=0}^{L-1} h_l^2 = 1$$
  3. orthogonality to even shifts:
  $$\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0$$

  for all nonzero integers n
- 2 & 3 together called orthonormality property

The Wavelet Filter: II

- transfer & squared gain functions for h_l:

$$H(f) \equiv \sum_{l=0}^{L-1} h_l e^{-i2\pi f l} \;\;\&\;\; \mathcal{H}(f) \equiv |H(f)|^2$$

- can argue that orthonormality property equivalent to

$$\mathcal{H}(f) + \mathcal{H}(f + \tfrac{1}{2}) = 2 \quad \text{for all } f$$

- Fig. 11: $\mathcal{H}(f)$ for Daubechies wavelet filters
  - $L = 2$ case is Haar wavelet filter
  - filter cascade with averaging & differencing filters
  - high-pass filter with pass-band $[\frac{1}{4}, \frac{1}{2}]$
  - can regard as half-band filter

The Scaling Filter: I

- scaling filter: $g_l \equiv (-1)^{l+1} h_{L-1-l}$
  - reverse h_l & flip sign of every other coefficient
  - e.g.: $h_0 = \frac{1}{\sqrt{2}}$ & $h_1 = -\frac{1}{\sqrt{2}} \Rightarrow g_0 = g_1 = \frac{1}{\sqrt{2}}$
  - g_l is 'quadrature mirror' filter for h_l
- properties of h_l imply g_l has these properties:
  1. summation to $\pm\sqrt{2}$, so will assume
  $$\sum_{l=0}^{L-1} g_l = \sqrt{2}$$
  2. unit energy:
  $$\sum_{l=0}^{L-1} g_l^2 = 1$$
  3. orthogonality to even shifts:
  $$\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0$$
  for all nonzero integers n
  4. orthogonality to wavelet filter at even shifts:
  $$\sum_{l=0}^{L-1} g_l h_{l+2n} = \sum_{l=-\infty}^{\infty} g_l h_{l+2n} = 0$$
  for all integers n

The Scaling Filter: II

- transfer & squared gain functions for g_l:

$$G(f) \equiv \sum_{l=0}^{L-1} g_l e^{-i2\pi fl} \ \& \ \mathcal{G}(f) \equiv |G(f)|^2$$

- can argue that $\mathcal{G}(f) = \mathcal{H}(f - \frac{1}{2})$
  - have $\mathcal{G}(0) = \mathcal{H}(-\frac{1}{2}) = \mathcal{H}(\frac{1}{2})$ & $\mathcal{G}(\frac{1}{2}) = \mathcal{H}(0)$
  - since h_l is high-pass, g_l must be low-pass
  - low-pass filter with pass-band $[0, \frac{1}{4}]$
  - can also regard as half-band filter
- orthonormality property equivalent to

$$\mathcal{G}(f) + \mathcal{G}(f + \tfrac{1}{2}) = 2 \text{ or } \mathcal{H}(f) + \mathcal{G}(f) = 2 \quad \text{for all } f$$

Pyramid Algorithm: II

- define $\mathbf{V}_0 \equiv \mathbf{X}$ and set $j = 1$
- input to jth stage of pyramid algorithm is $\mathbf{V}_{j-1}$
  - $\mathbf{V}_{j-1}$ is full-band
  - related to frequencies $[0, \frac{1}{2^j}]$ in $\mathbf{X}$
- filter with half-band filters and downsample:

$$\begin{aligned} W_{j,t} &\equiv \sum_{l=0}^{L-1} h_l V_{j-1,2t+1-l \bmod N_{j-1}} \\ V_{j,t} &\equiv \sum_{l=0}^{L-1} g_l V_{j-1,2t+1-l \bmod N_{j-1}}, \end{aligned}$$

  $t = 0, \ldots, N_j - 1$
- place these in vectors $\mathbf{W}_j$ & $\mathbf{V}_j$
  - $\mathbf{W}_j$ are wavelet coefficients for scale $\tau_j = 2^{j-1}$
  - $\mathbf{V}_j$ are scaling coefficients for scale $\lambda_j = 2^j$
- increment j and repeat above until $j = J_0$
- yields DWT coefficients $\mathbf{W}_1, \ldots, \mathbf{W}_{J_0}, \mathbf{V}_{J_0}$

Pyramid Algorithm: III

- can formulate inverse pyramid algorithm (recovers $\mathbf{V}_{j-1}$ from $\mathbf{W}_j$ and $\mathbf{V}_j$)
- algorithm implicitly defines transform matrix $\mathcal{W}$
- partition $\mathcal{W}$ commensurate with $\mathbf{W}_j$:

$$\mathcal{W} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \vdots \\ \mathcal{W}_{J_0} \\ \mathcal{V}_{J_0} \end{bmatrix} \text{ parallels } \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_{J_0} \\ \mathbf{V}_{J_0} \end{bmatrix}$$

- rows of $\mathcal{W}_j$ use jth level filter $h_{j,l}$ with DFT

$$H(2^{j-1}f)\prod_{l=0}^{j-2} G(2^l f)$$

  ($h_{j,l}$ has $L_j = (2^j - 1)(L-1) + 1$ nonzero elements)

- $\mathcal{W}_j$ is $N_j \times N$ matrix such that

$$\mathbf{W}_j = \mathcal{W}_j \mathbf{X} \text{ and } \mathcal{W}_j \mathcal{W}_j^T = I_{N_j}$$

Two Consequences of Orthonormality

- multiresolution analysis (MRA)

$$\mathbf{X} = \mathcal{W}^T \mathbf{W} = \sum_{j=1}^{J_0} \mathcal{W}_j^T \mathbf{W}_j + \mathcal{V}_{J_0}^T \mathbf{V}_{J_0} \equiv \sum_{j=1}^{J_0} \mathcal{D}_j + \mathcal{S}_{J_0}$$

  – scale-based additive decomposition

  – $\mathcal{D}_j$'s & $\mathcal{S}_{J_0}$ called details & smooth

- analysis of variance

  – consider 'energy' in time series:

$$\|\mathbf{X}\|^2 = \mathbf{X}^T \mathbf{X} = \sum_{t=0}^{N-1} X_t^2$$

  – energy preserved in DWT coefficients:

$$\|\mathbf{W}\|^2 = \|\mathcal{W}\mathbf{X}\|^2 = \mathbf{X}^T \mathcal{W}^T \mathcal{W} \mathbf{X} = \mathbf{X}^T \mathbf{X} = \|\mathbf{X}\|^2$$

  – since $\mathbf{W}_1, \ldots, \mathbf{W}_{J_0}, \mathbf{V}_{J_0}$ partitions $\mathbf{W}$, have

$$\|\mathbf{W}\|^2 = \sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_{J_0}\|^2,$$

  leading to analysis of sample variance:

$$\hat{\sigma}^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} \left(X_t - \overline{X}\right)^2 = \frac{1}{N} \sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \left(\frac{1}{N} \|\mathbf{V}_{J_0}\|^2 - \overline{X}^2\right)$$

  – scale-based decomposition (cf. frequency-based)

Variation: Maximal Overlap DWT

- can eliminate downsampling and use

  $$\widetilde{W}_{j,t} \equiv \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1$$

  to define MODWT coefficients $\widetilde{\mathbf{W}}_j$ (& also $\widetilde{\mathbf{V}}_j$)

- unlike DWT, MODWT is not orthonormal (in fact MODWT is highly redundant)

- like DWT, can do MRA & analysis of variance:

  $$\|\mathbf{X}\|^2 = \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2$$

- unlike DWT, MODWT works for all samples sizes N (i.e., power of 2 assumption is not required)

  - if N is power of 2, can compute MODWT using $O(N \log_2 N)$ operations (i.e., same as FFT algorithm)
  - contrast to DWT, which uses $O(N)$ operations

- Fig. 12: Haar MODWT coefficients for clock 571 (cf. Fig. 6 with DWT coefficients)

Definition of Wavelet Variance

- let X_t, $t = \ldots, -1, 0, 1, \ldots,$ be a stochastic process
- run X_t through jth level wavelet filter:

$$\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t = \ldots, -1, 0, 1, \ldots,$$

  which should be contrasted with

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1$$

- definition of time dependent wavelet variance (also called wavelet spectrum):

$$\nu^2_{X,t}(\tau_j) \equiv \operatorname{var}\{\overline{W}_{j,t}\},$$

  assuming $\operatorname{var}\{\overline{W}_{j,t}\}$ exists and is finite
- $\nu^2_{X,t}(\tau_j)$ depends on τ_j and t
- will consider time independent wavelet variance:

$$\nu^2_X(\tau_j) \equiv \operatorname{var}\{\overline{W}_{j,t}\}$$

  (can be easily adapted to time varying situation)

Rationale for Wavelet Variance

- decomposes variance on scale by scale basis
- useful substitute/complement for spectrum
- useful substitute for process/sample variance

Variance Decomposition

- suppose X_t has power spectrum $S_X(f)$:

$$\int_{-1/2}^{1/2} S_X(f)\, df = \mathrm{var}\,\{X_t\};$$

  i.e., decomposes $\mathrm{var}\,\{X_t\}$ across frequencies f

  - involves uncountably infinite number of f's
  - $S_X(f)\,\Delta f \approx$ contribution to $\mathrm{var}\,\{X_t\}$ due to f's in interval of length Δf centered at f

- wavelet variance analog to fundamental result:

$$\sum_{j=1}^{\infty} \nu_X^2(\tau_j) = \mathrm{var}\,\{X_t\}$$

  i.e., decomposes $\mathrm{var}\,\{X_t\}$ across scales τ_j

  - recall DWT/MODWT and sample variance
  - involves countably infinite number of τ_j's
  - $\nu_X^2(\tau_j)$ contribution to $\mathrm{var}\,\{X_t\}$ due to scale τ_j
  - $\nu_X(\tau_j)$ has same units as X_t (easier to interpret)

Spectrum Substitute/Complement

- because $\tilde{h}_{j,l} \approx$ bandpass over $[1/2^{j+1}, 1/2^j]$,

$$\nu_X^2(\tau_j) \approx 2 \int_{1/2^{j+1}}^{1/2^j} S_X(f)\, df$$

- if $S_X(f)$ 'featureless', info in $\nu_X^2(\tau_j) \Leftrightarrow$ info in $S_X(f)$
- $\nu_X^2(\tau_j)$ more succinct: only 1 value per octave band
- example: $S_X(f) \propto |f|^\alpha$, i.e., power law process
  - can deduce α from slope of $\log S_X(f)$ vs. $\log f$
  - implies $\nu_X^2(\tau_j) \propto \tau_j^{-\alpha-1}$ approximately
  - can deduce α from slope of $\log \nu_X^2(\tau_j)$ vs. $\log \tau_j$
  - no loss of 'info' using $\nu_X^2(\tau_j)$ rather than $S_X(f)$
- with Haar wavelet, obtain pilot spectrum estimate proposed in Blackman & Tukey (1958)

Substitute for Variance: I

- can be difficult to estimate process variance
- $\nu_X^2(\tau_j)$ useful substitute: easy to estimate & finite
- let $\mu = E\{X_t\}$ be known, $\sigma^2 = \text{var}\,\{X_t\}$ unknown
- can estimate σ^2 using

$$\tilde{\sigma}^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1}(X_t - \mu)^2$$

- estimator above is unbiased: $E\{\tilde{\sigma}^2\} = \sigma^2$
- if μ is unknown, can estimate σ^2 using

$$\hat{\sigma}^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1}(X_t - \overline{X})^2$$

- there is some (non-pathological!) X_t such that

$$\frac{E\{\hat{\sigma}^2\}}{\sigma^2} < \epsilon$$

  for any gvien $\epsilon > 0$ & $N \geq 1$
- $\hat{\sigma}^2$ can badly underestimate σ^2!
- example: power law process with $-1 < \alpha < 0$

Substitute for Variance: II

- Q: why is wavelet variance useful when σ^2 is not?
- replaces 'global' variability with variability over scales
- if X_t stationary with mean μ, then

$$E\{\overline{W}_{j,t}\} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} E\{X_{t-l}\} = \mu \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} = 0$$

  because $\sum_l \tilde{h}_{j,l} = 0$
- $E\{\overline{W}_{j,t}\}$ known, so can get unbiased estimator of $\operatorname{var}\{\overline{W}_{j,t}\} = \nu_X^2(\tau_j)$
- certain nonstationary X_t have well-defined $\nu_X^2(\tau_j)$
- example: power law processes with $\alpha \le -1$ (example of process with stationary increments)

Estimation of Wavelet Variance: I

- can base estimator on MODWT of $X_0, X_1, \ldots, X_{N-1}$:

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1$$

  (DWT-based estimator possible, but less efficient)

- recall that

$$\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t = 0, \pm 1, \pm 2, \ldots$$

  so $\widetilde{W}_{j,t} = \overline{W}_{j,t}$ if mod not needed: $L_j - 1 \le t < N$

- if $N - L_j \ge 0$, unbiased estimator of $\nu_X^2(\tau_j)$ is

$$\hat{\nu}_X^2(\tau_j) \equiv \frac{1}{N - L_j + 1} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2,$$

  where $M_j \equiv N - L_j + 1$

- can also construct biased estimator of $\nu_X^2(\tau_j)$:

$$\tilde{\nu}_X^2(\tau_j) \equiv \frac{1}{N} \sum_{t=0}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{N} \Big(\sum_{t=0}^{L_j-2} \widetilde{W}_{j,t}^2 + \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2 \Big)$$

  1st sum in parentheses influenced by circularity

Estimation of Wavelet Variance: II

- biased estimator unbiased if $\{X_t\}$ white noise
- biased estimator offers exact analysis of $\hat{\sigma}^2$; unbiased estimator need not
- biased estimator can have better mean square error (Greenhall *et al.*, 1999; need to 'reflect' X_t)

Statistical Properties of $\hat{\nu}_X^2(\tau_j)$

- suppose $\{\overline{W}_{j,t}\}$ Gaussian, mean 0 & spectrum $S_j(f)$
- suppose square integrability condition holds:

$$A_j \equiv \int_{-1/2}^{1/2} S_j^2(f)\,df < \infty \;\;\&\;\; S_j(f) > 0$$

  (holds for power law processes if L large enough)

- can show $\hat{\nu}_X^2(\tau_j)$ asymptotically normal with mean $\nu_X^2(\tau_j)$ & large sample variance $2A_j/M_j$
- can estimate A_j and use with $\hat{\nu}_X^2(\tau_j)$ to construct confidence interval for $\nu_X^2(\tau_j)$
- example
  - Fig. 13: clock errors $X_t \equiv X_t^{(0)}$ along with differences $X_t^{(i)} \equiv X_t^{(i-1)} - X_{t-1}^{(i-1)}$ for $i = 1, 2$
  - Fig. 14: $\hat{\nu}_X^2(\tau_j)$ for clock errors
  - Fig. 15: $\hat{\nu}_{\overline{Y}}^2(\tau_j)$ for $\overline{Y}_t \propto X_t^{(1)}$
  - Haar $\hat{\nu}_{\overline{Y}}^2(\tau_j)$ related to Allan variance $\sigma_{\overline{Y}}^2(2, \tau_j)$:

$$\nu_{\overline{Y}}^2(\tau_j) = \tfrac{1}{2}\sigma_{\overline{Y}}^2(2, \tau_j)$$

Decorrelation of FD Processes

- X_t 'fractionally differenced' if its spectrum is
$$S_X(f) = \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}},$$
where $\sigma_\epsilon^2 > 0$ and $-\frac{1}{2} < \delta < \frac{1}{2}$
- note: for small f, have $S_X(f) \approx C/|f|^{2\delta}$; i.e., power law with $\alpha = -2\delta$
- if $\delta = 0$, FD process is white noise
- if $0 < \delta < \frac{1}{2}$, FD stationary with 'long memory'
- can extend definition to $\delta \geq \frac{1}{2}$
  - nonstationary $1/f$ type process
  - also called ARFIMA(0,δ,0) process
- Fig. 16: DWT of simulated FD process, $\delta = 0.4$ (sample autocorrelation sequences (ACSs) on right)

DWT as Whitening Transform

- sample ACSs suggest $\mathbf{W}_j \approx$ uncorrelated
- since FD process is stationary, so are $\mathbf{W}_j$ (ignoring terms influenced by circularity)
- Fig. 17: spectra for $\mathbf{W}_j$, $j = 1, 2, 3, 4$
- $\mathbf{W}_j$ & $\mathbf{W}_{j'}$, $j \neq j'$, approximately uncorrelated (approximation improves as L increases)
- DWT thus acts as a whitening transform
- lots of uses for whitening property, including:
  1. testing for variance changes
  2. bootstrapping time series statistics
  3. estimating δ for stationary/nonstationary fractional difference processes with trend

Estimation for FD Processes: I

- extension of work by Wornell; McCoy & Walden
- problem: estimate δ from time series U_t such that

$$U_t = T_t + X_t$$

  where

  - $T_t \equiv \sum_{j=0}^{r} a_j t^j$ is polynomial trend
  - X_t is FD process, but can have $\delta \geq \frac{1}{2}$

- DWT wavelet filter of width L has embedded differencing operation of order $L/2$
- if $\frac{L}{2} \geq r + 1$, reduces polynomial trend to 0
- can partition DWT coefficients as

$$\mathbf{W} = \mathbf{W}_s + \mathbf{W}_b + \mathbf{W}_w$$

  where

  - $\mathbf{W}_s$ has scaling coefficients and 0s elsewhere
  - $\mathbf{W}_s$ has boundary-dependent wavelet coefficients
  - $\mathbf{W}_w$ has boundary-independent wavelet coefficients

Estimation for FD Processes: II

- since $\mathbf{U} = \mathcal{W}^T \mathbf{W}$, can write

$$\mathbf{U} = \mathcal{W}^T (\mathbf{W}_s + \mathbf{W}_b) + \mathcal{W}^T \mathbf{W}_w \equiv \widehat{\mathbf{T}} + \widehat{\mathbf{X}}$$

- Fig. 18: example with fractional frequency deviates
- can use values in $\mathbf{W}_w$ to form likelihood:

$$L(\delta, \sigma_\epsilon^2) \equiv \prod_{j=1}^{J_0} \prod_{t=1}^{N'_j} \frac{1}{\left(2\pi\sigma_j^2\right)^{1/2}} e^{-W^2_{j,t+L'_j-1}/(2\sigma_j^2)}$$

  where

$$\sigma_j^2 \equiv \int_{-1/2}^{1/2} \mathcal{H}_j(f) \frac{\sigma_\epsilon^2}{|2\sin(\pi f)|^{2\delta}}\, df;$$

  and $\mathcal{H}_j(f)$ is squared gain for $h_{j,l}$

- leads to maximum likelihood estimator $\hat{\delta}$ for δ
- works well in Monte Carlo simulations
- get $\hat{\delta} \doteq 0.39 \pm 0.03$ for fractional frequency deviates

DWT-based Signal Extraction: I

- DWT analysis of $\mathbf{X}$ yields $\mathbf{W} = \mathcal{W}\mathbf{X}$
- DWT synthesis $\mathbf{X} = \mathcal{W}^T\mathbf{W}$ yields
  - multiresolution analysis (MRA)
  - estimator of 'signal' $\mathbf{D}$ hidden in $\mathbf{X}$:
    * modify $\mathbf{W}$ to get $\mathbf{W}'$
    * use $\mathbf{W}'$ to form signal estimate:

$$\widehat{\mathbf{D}} \equiv \mathcal{W}^T\mathbf{W}'$$

- key ideas behind wavelet-based signal estimation
  - DWT can isolate signals in small number of W_n's
  - can 'threshold' or 'shrink' W_n's
- key ideas lead to 'waveshrink' (Donoho and Johnstone, 1995)

DWT-based Signal Extraction: II

- thresholding schemes involve
  1. computing $\mathbf{W} \equiv \mathcal{W}\mathbf{X}$
  2. defining $\mathbf{W}^{(t)}$ as vector with nth element
     $$W_n^{(t)} = \begin{cases} 0, & \text{if } |W_n| \leq \delta; \\ \text{some nonzero value}, & \text{otherwise}, \end{cases}$$
     where nonzero values are yet to be defined
  3. estimating $\mathbf{D}$ via $\widehat{\mathbf{D}}^{(t)} \equiv \mathcal{W}^T\mathbf{W}^{(t)}$
- simplest scheme is 'hard thresholding:'
  $$W_n^{(ht)} = \begin{cases} 0, & \text{if } |W_n| \leq \delta; \\ W_n, & \text{otherwise.} \end{cases}$$
  Fig. 19: solid line ('kill/keep' strategy)
- alterative scheme is 'soft thresholding:'
  $$W_n^{(st)} = \operatorname{sign}\{W_n\}\left(|W_n| - \delta\right)_+,$$
  where
  $$\operatorname{sign}\{W_n\} \equiv \begin{cases} +1, & \text{if } W_n > 0; \\ 0, & \text{if } W_n = 0; \\ -1, & \text{if } W_n < 0. \end{cases} \quad \text{and} \quad (x)_+ \equiv \begin{cases} x, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$
  Fig. 19: dashed line

DWT-based Signal Extraction: III

- third scheme is 'mid thresholding:'

  $$W_n^{(mt)} = \operatorname{sign}\{W_n\}\,(|W_n| - \delta)_{++}\,,$$

  where

  $$(|W_n| - \delta)_{++} \equiv \begin{cases} 2(|W_n| - \delta)_{+}, & \text{if } |W_n| < 2\delta; \\ |W_n|, & \text{otherwise} \end{cases}$$

  Fig. 19: dotted line

- Q: how should δ be set?

- A: universal' threshold (Donoho & Johnstone, 1995) (lots of other answers have been proposed)

  - specialize to model $\mathbf{X} = \mathbf{D} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is Gaussian white noise with variance σ_ϵ^2
  - 'universal' threshold: $\delta_{\mathbf{U}} \equiv \sqrt{[2\sigma_\epsilon^2 \log(N)]}$
  - rationale for $\delta_{\mathbf{U}}$:
    * suppose $\mathbf{D} = \mathbf{0}$ & hence $\mathbf{W}$ is white noise also
    * as $N \to \infty$, have

      $$\mathbf{P}\left[\max_n |W_n| \leq \delta_{\mathbf{U}}\right] \to 1$$

      so all $\mathbf{W}^{(ht)} = 0$ with high probability
    * will estimate correct $\mathbf{D}$ with high probability

DWT-based Signal Extraction: IV

- can estimate σ_ϵ^2 using median absolute deviation (MAD):

$$\hat{\sigma}_{(\text{MAD})} \equiv \frac{\text{median}\,\{|W_{1,0}|, |W_{1,1}|, \ldots, |W_{1,\frac{N}{2}-1}|\}}{0.6745},$$

  where $W_{1,t}$'s are elements of $\mathbf{W}_1$
- Fig. 20: application to NMR series
- has potential application in dejamming GPS signals (with roles of 'signal' and 'noise' swapped!)

Web Material and Books

- Wavelet Digest

  `http://www.wavelet.org/`

- MathSoft's wavelet resource page

  `http://www.mathsoft.com/wavelets.html`

- books

  - R. Carmona, W.–L. Hwang & B. Torrésani (1998), *Practical Time-Frequency Analysis*, Academic Press
  - S. G. Mallat (1999), *A Wavelet Tour of Signal Processing* (Second Edition), Academic Press
  - R. T. Ogden (1997), *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser
  - D. B. Percival & A. T. Walden (2000), *Wavelet Methods for Time Series Analysis*, Cambridge University Press (will appear in July/August) `http://www.staff.washington.edu/dbp/wmtsa.html`
  - B. Vidakovic (1999), *Statistical Modeling by Wavelets*, John Wiley & Sons.

Software

- `Matlab`
  - WaveLab (free):

    `http://www-stat.stanford.edu/~wavelab`
  - WaveBox (commercial):

    `http://www.toolsmiths.com/`
- `Mathcad` Wavelets Extension Pack (commercial):

  `http://www.mathsoft.com/mathcad/ebooks/wavelets.asp`
- `S-Plus` software
  - WaveThresh (free):

    `http://lib.stat.cmu.edu/S/wavethresh`
  - S+Wavelets (commercial):

    `http://www.mathsoft.com/splsprod/wavelets.html`