Catching up on Wavelets:

Recent Advances, Future Directions

Don Percival
Applied Physics Laboratory
Box 355640
University of Washington
Seattle, WA 98195–5640

206–543–1368

dbp@apl.washington.edu

http://faculty.washington.edu/dbp

A Little Background

- wavelets are analysis tools for time series and images
- as a subject, wavelets are
  - relatively new (1983 to present)
  - a synthesis of old/new ideas
  - keyword in 29,826+ articles and books since 1989 (4032 more since 2005: an inundation of material!!!)
- broadly speaking, there have been two waves of wavelets
  - continuous wavelet transform (1983 and on)
  - discrete wavelet transform (1988 and on)
- statistical community has focused largely on DWT

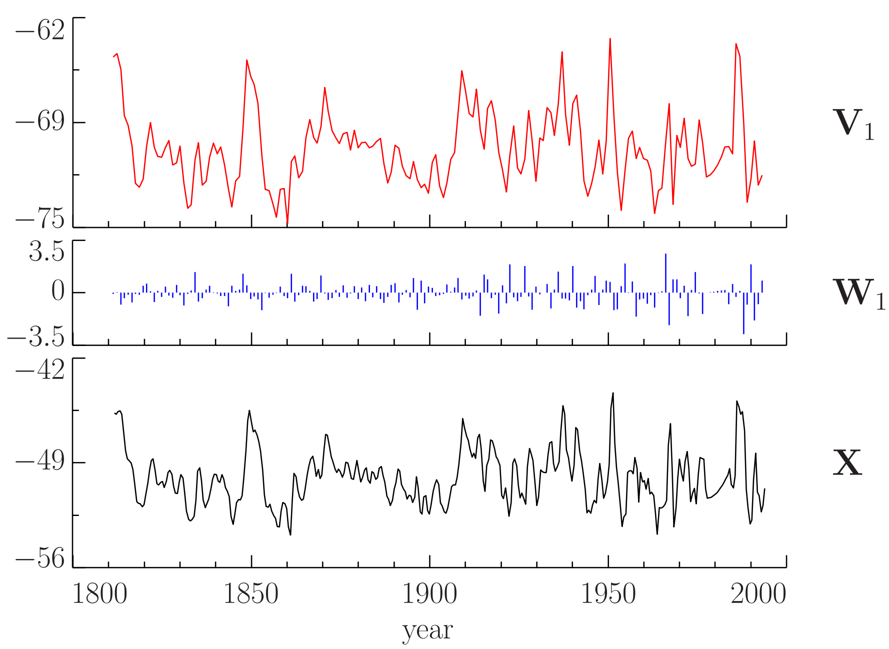
Why Focus on the DWT?

- let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ represent a time series
- can formulate DWT as an orthonormal transform $\mathcal{W}$ yielding N DWT coefficients $\mathbf{W} = \mathcal{W}\mathbf{X}$, where $\mathcal{W}^T\mathcal{W} = I_N$
- each wavelet coefficient can be interpreted as a difference between localized weighted averages over certain scales
- three basic properties of the DWT

1. yields additive decomposition (multiresolution analysis)
   - reexpresses $\mathbf{X}$ as sum of several new time series, each of which is associated with a particular scale (or scales)
   - starting point for wavelet-based signal extraction
2. yields a scale-based analysis of variance (wavelet variance)
3. tends to decorrelate certain time series

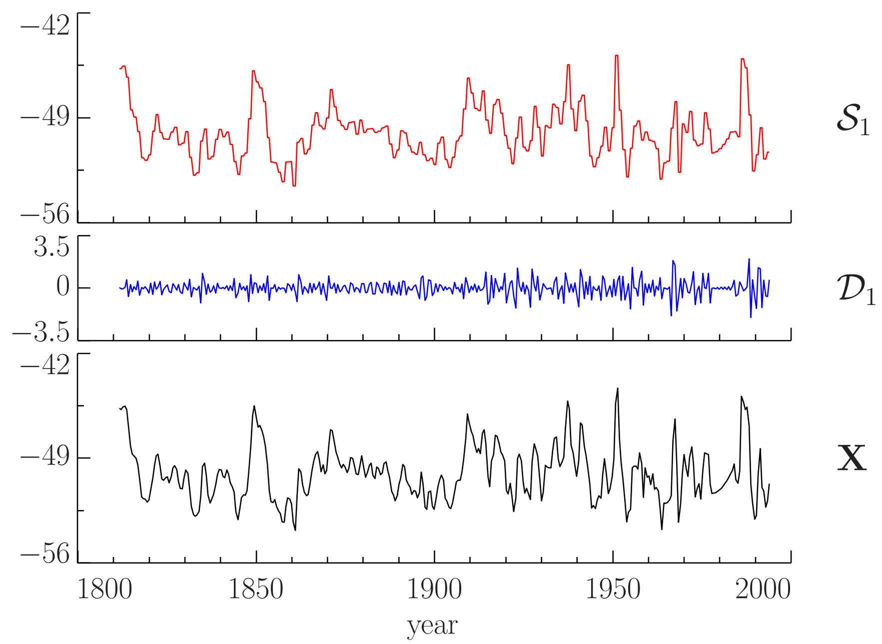
Example of DWT

- oxygen isotope records $\mathbf{X}$ from Antarctic ice core



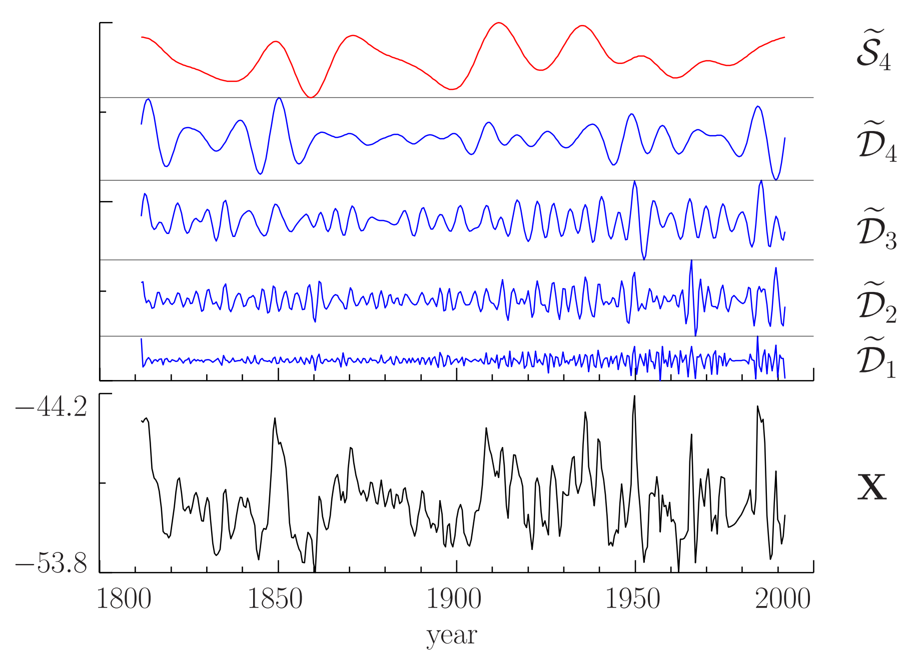
Example of Two Level Multiresolution Analysis

- oxygen isotope records **X** from Antarctic ice core



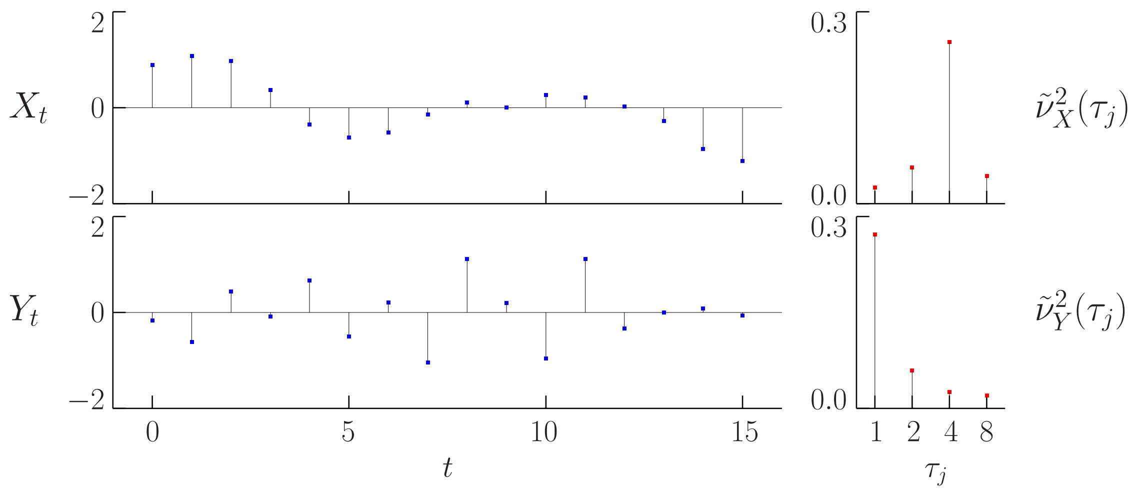
Example of Five Level Multiresolution Analysis

- oxygen isotope records **X** from Antarctic ice core



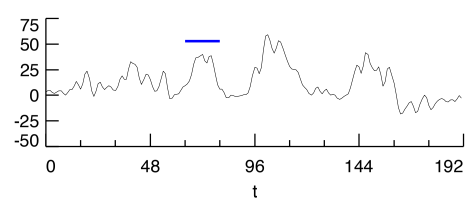
First Example of Wavelet Variance

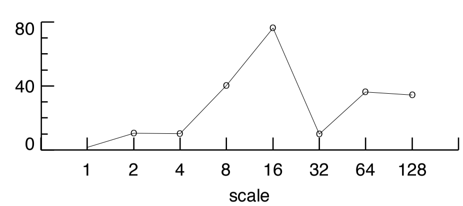
- wavelet variances for time series X_t and Y_t of length $N = 16$, each with zero sample mean and same sample variance



Second Example of Wavelet Variance

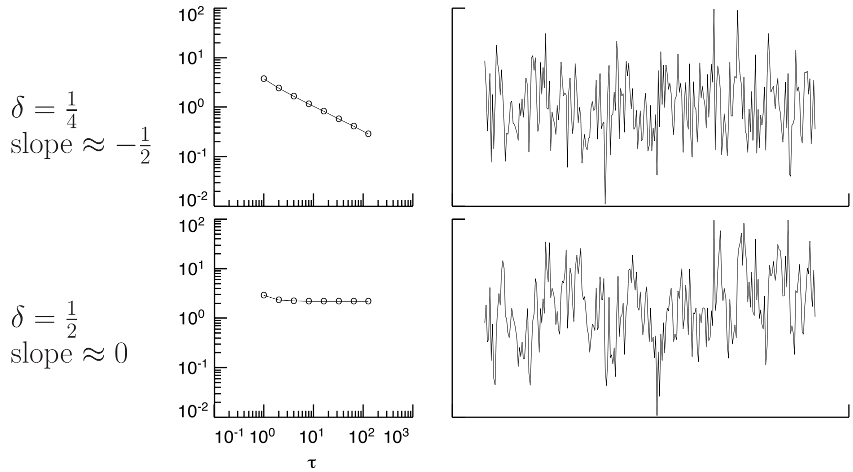
- top: part of subtidal sea level data (blue line shows scale of 16)





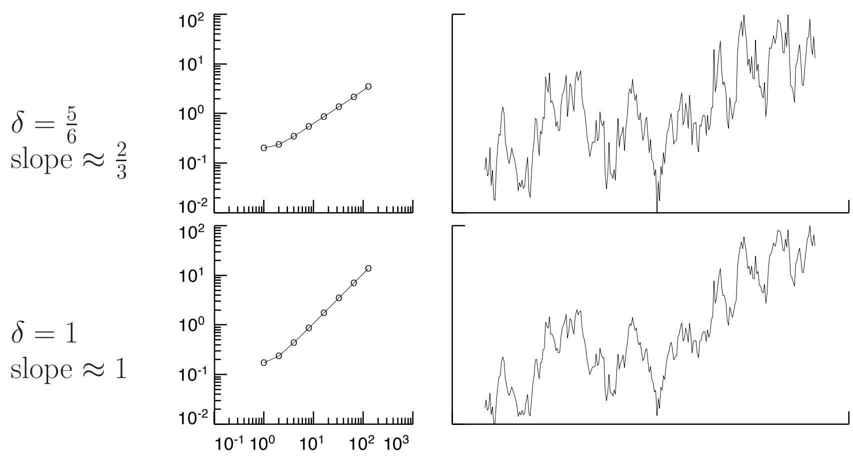
- bottom: wavelet variances $\tilde{\nu}^2_X(\tau_j)$

WV for Fractionally Differenced Processes



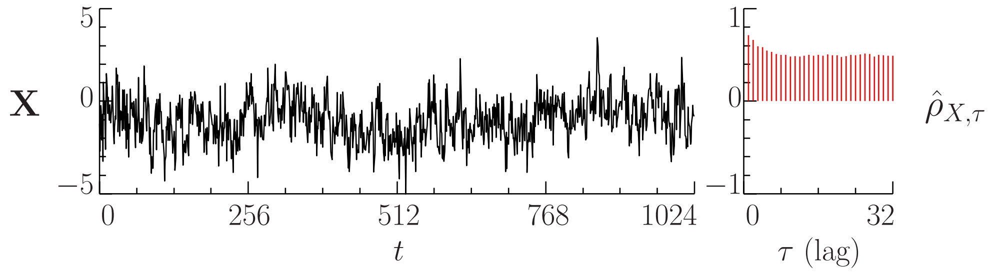
- left-hand column: $\nu_X^2(\tau_j)$ versus τ_j based upon LA(8) wavelet
- right-hand: realization of length $N = 256$ from each FD process

WV for 2 More FD Processes



- $\delta = \frac{5}{6}$ is Kolmogorov turbulence; $\delta = 1$ is random walk
- note: positive slope indicates nonstationarity, while negative slope indicates stationarity

DWT of a Long Memory Process: I

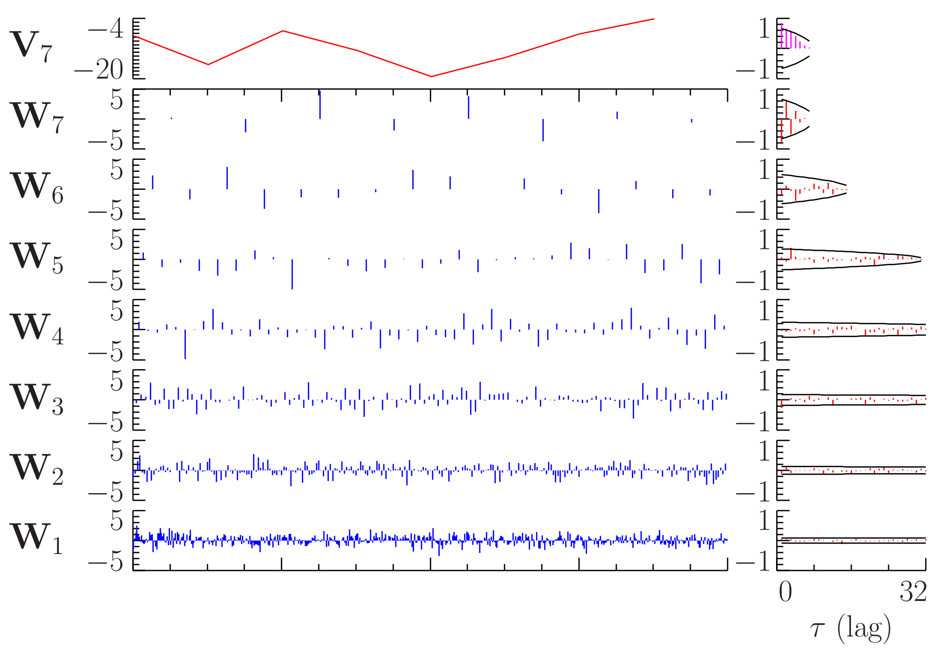


- realization of an FD(0.4) time series $\mathbf{X}$ along with its sample autocorrelation sequence (ACS): for $\tau \geq 0$,

$$\hat{\rho}_{X,\tau} \equiv \frac{\sum_{t=0}^{N-1-\tau} X_t X_{t+\tau}}{\sum_{t=0}^{N-1} X_t^2}$$

- note that ACS dies down slowly

DWT of a Long Memory Process: II



- LA(8) DWT of FD(0.4) series and sample ACSs for each $\mathbf{W}_j$ & $\mathbf{V}_7$, along with 95% confidence intervals for white noise

Wavelet-Based Signal Estimation

- key ideas
  - certain signals can be efficiently described by the DWT using
    * all of the scaling coefficients
    * a small number of 'large' wavelet coefficients
  - noise is manifested in a large number of 'small' wavelet coefficients
  - can either 'threshold' or 'shrink' wavelet coefficients to eliminate noise in the wavelet domain
- key ideas led to wavelet thresholding and shrinkage proposed by Donoho, Johnstone and coworkers in 1990s

Signal Estimation via Thresholding

- thresholding schemes involve
  1. computing $\mathbf{W} \equiv \mathcal{W}\mathbf{X}$
  2. defining $\mathbf{W}^{(t)}$ as vector with lth element

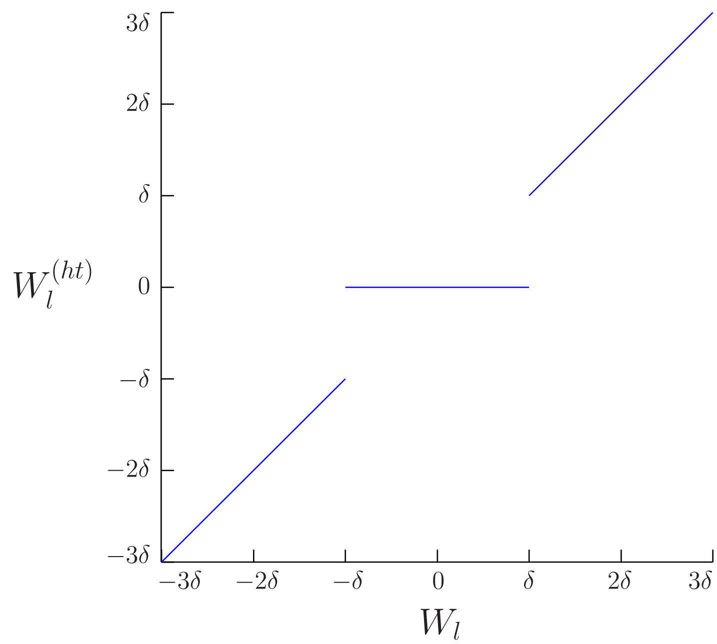
$$W_l^{(t)} = \begin{cases} 0, & \text{if } |W_l| \leq \delta; \\ \text{some nonzero value}, & \text{otherwise}, \end{cases}$$

     where nonzero values are yet to be defined
  3. estimating signal $\mathbf{D}$ via $\widehat{\mathbf{D}}^{(t)} \equiv \mathcal{W}^T\mathbf{W}^{(t)}$
- simplest scheme is 'hard thresholding' ('kill/keep' strategy):

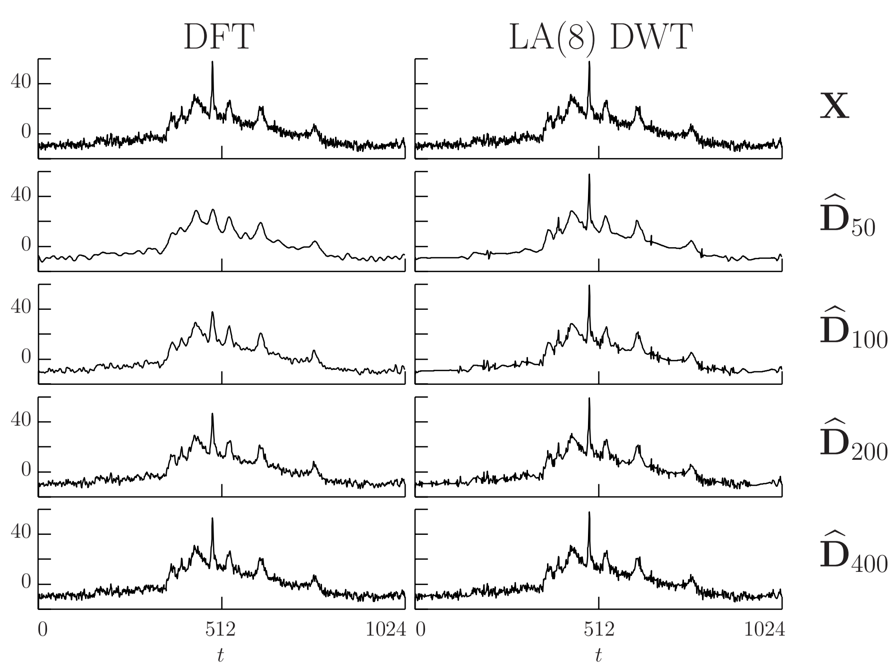
$$W_l^{(ht)} = \begin{cases} 0, & \text{if } |W_l| \leq \delta; \\ W_l, & \text{otherwise.} \end{cases}$$

Hard Thresholding Function

- plot shows mapping from W_l to $W_l^{(ht)}$



Fourier- and Wavelet-Based Thresholding



- discrete Fourier transform denoising (left-hand column) and DWT denoising (right) for NMR series $\mathbf{X}$ (A. Maudsley, UCSF)

Soft Thresholding

- alterative scheme is 'soft thresholding:'

$$W_l^{(st)} = \text{sign}\{W_l\}\,(|W_l| - \delta)_+\,,$$

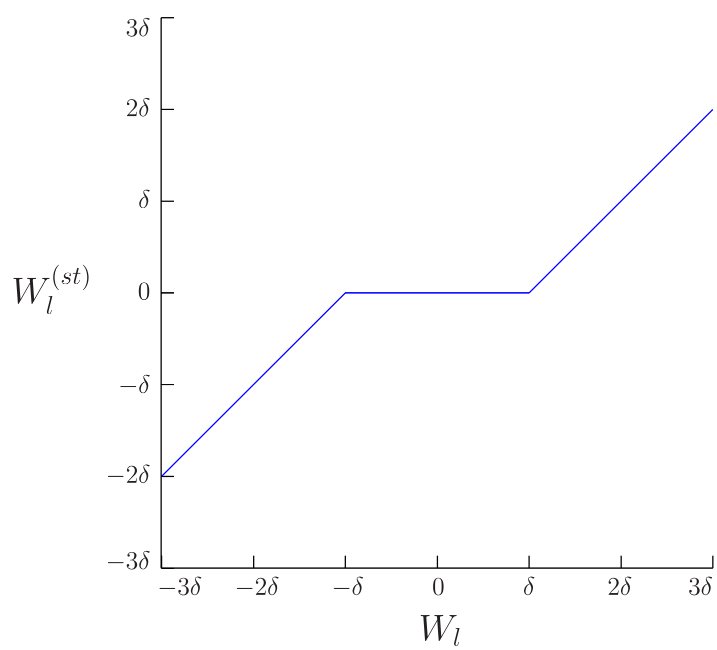
where

$$\text{sign}\{W_l\} \equiv \begin{cases} +1, & \text{if } W_l > 0; \\ 0, & \text{if } W_l = 0; \\ -1, & \text{if } W_l < 0. \end{cases} \quad \text{and} \quad (x)_+ \equiv \begin{cases} x, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

- one rationale for soft thresholding is that it fits into Stein's class of estimators

Soft Thresholding Function

- here is the mapping from W_l to $W_l^{(st)}$



Mid Thresholding

- third scheme is 'mid thresholding:'

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

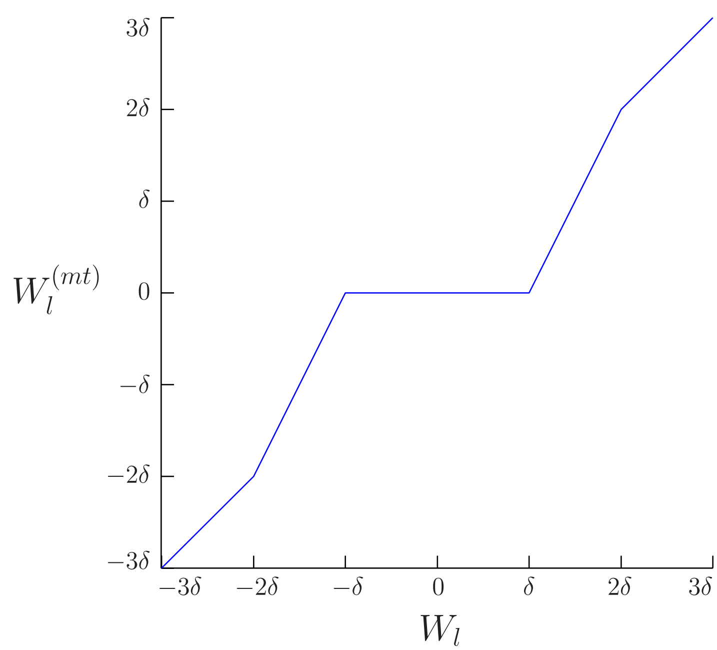
where

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

- provides compromise between hard and soft thresholding

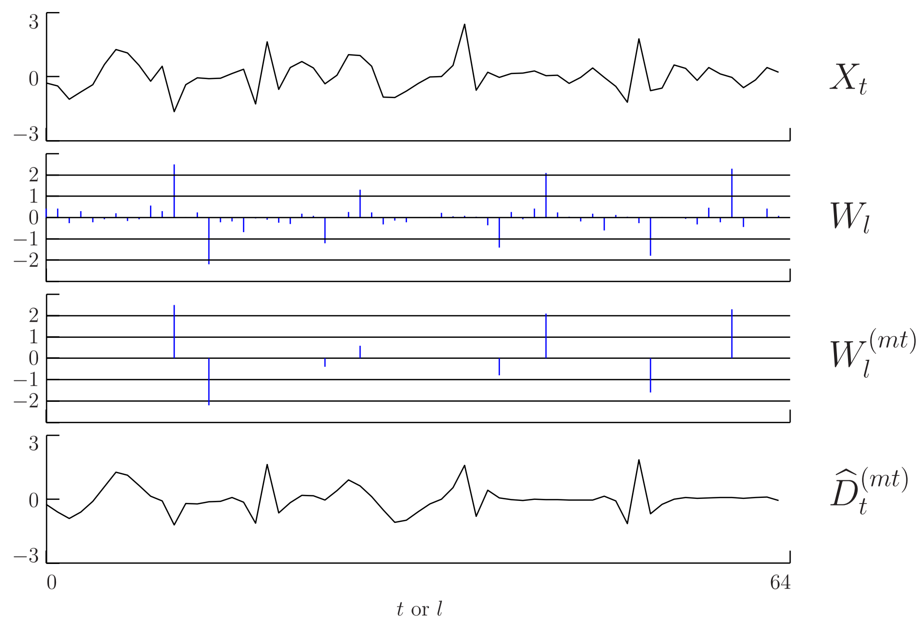
Mid Thresholding Function

- here is the mapping from W_l to $W_l^{(mt)}$

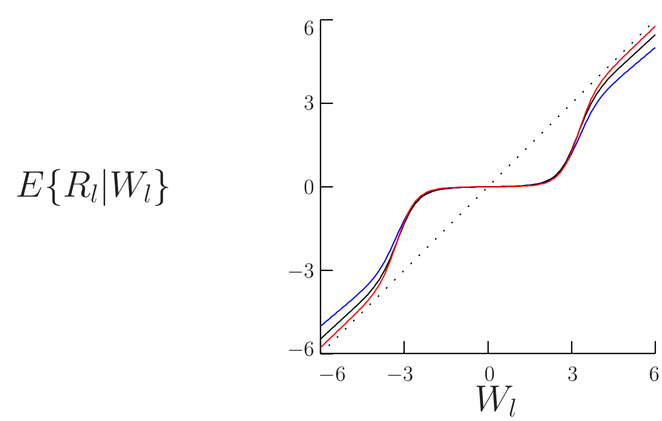


Example of Mid Thresholding

- example of mid thresholding with $\delta = 1$

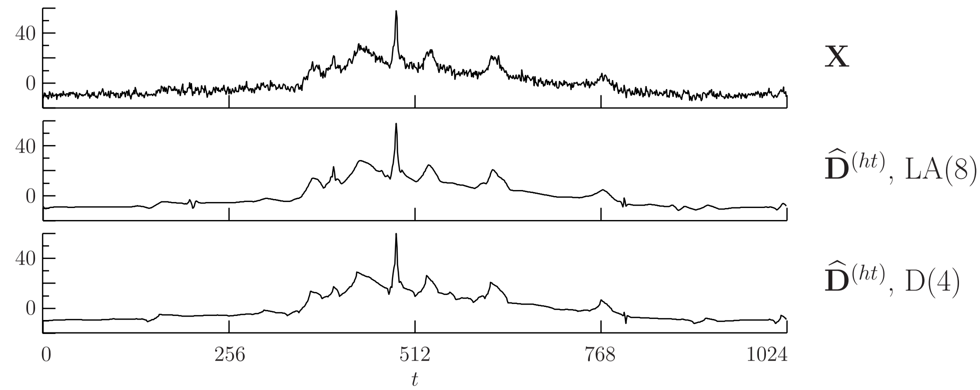


Shrinkage Functions



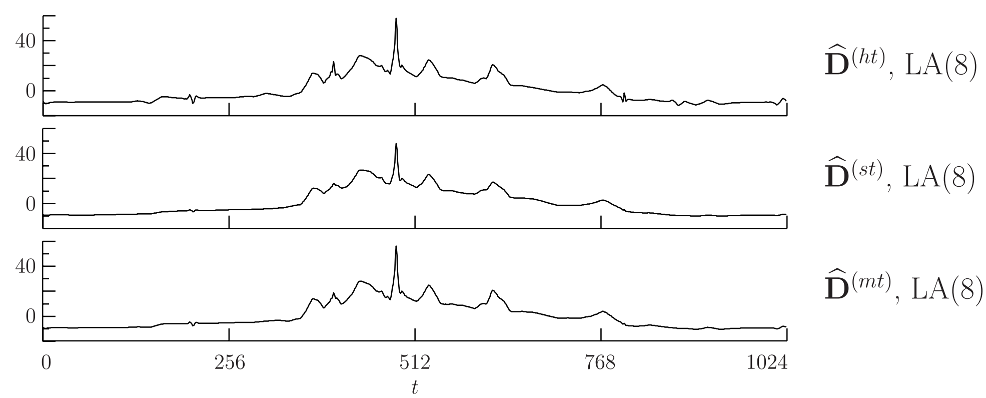
- conditional mean shrinkage rule for $p_l = 0.95$ (i.e., $\approx 95\%$ of signal coefficients are 0); $\sigma^2_{n_l} = 1$; and $\sigma^2_{G_l} = 5$ (curve furthest from dotted diagonal), 10 and 25 (curve nearest to diagonal)
- as $\sigma^2_{G_l}$ gets large (i.e., large signal coefficients increase in size), shrinkage rule starts to resemble mid thresholding rule

Examples of DWT-Based Thresholding: I



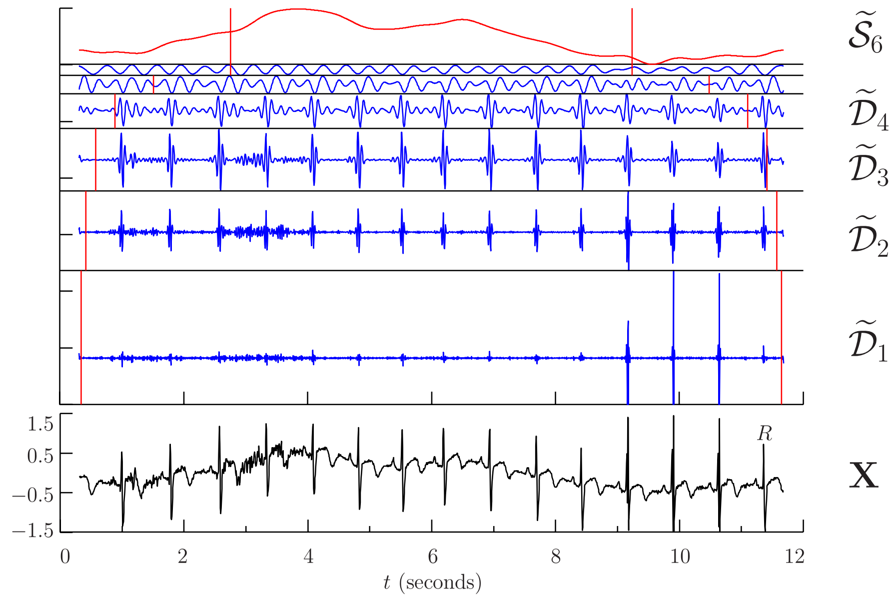
- top plot: NMR spectrum $\mathbf{X}$
- middle: signal estimate using $J_0 = 6$ partial LA(8) DWT with hard thresholding and universal thresholding
- bottom: same, but now using D(4) DWT

Examples of DWT-Based Thresholding: II



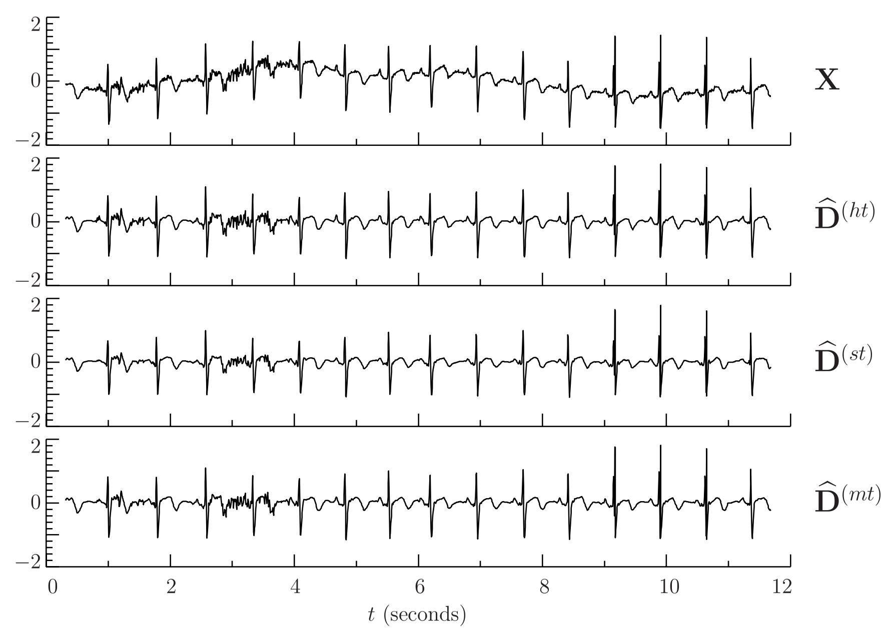
- top: signal estimate using $J_0 = 6$ partial LA(8) DWT with hard thresholding (repeat of middle plot of previous overhead)
- middle: same, but now with soft thresholding
- bottom: same, but now with mid thresholding

MRA of Electrocardiogram Data



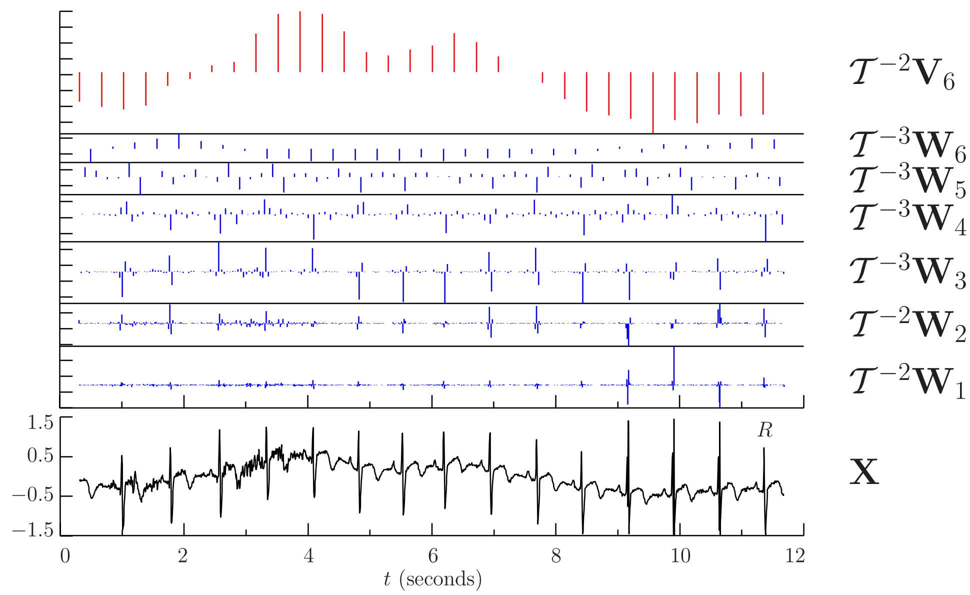
- LA(8) MODWT multiresolution analysis of ECG data

Denoising ECG Time Series



- hard/soft/mid threshold estimates with $J_0 = 6$ partial LA(8) DWT, MAD & scaling coefficients to 0 (zaps baseline drift)

‘2nd Generation’ Denoising: I



- ‘1st generation’ denoising looks at each W_l alone; for ‘real world’ signals, coefficients often cluster within a given level and persist across adjacent levels (ECG series offers an example)

‘2nd Generation’ Denoising: II

- here are some ‘2nd generation’ approaches that exploit these ‘real world’ properties:
  - Crouse *et al.* (1998) use hidden Markov models for stochastic signal DWT coefficients to handle clustering, persistence and non-Gaussianity
  - Huang and Cressie (2000) consider scale-dependent multi-scale graphical models to handle clustering and persistence
  - Cai and Silverman (2001) consider ‘block’ thesholding in which coefficients are thresholded in blocks rather than individually (handles clustering)
  - Dragotti and Vetterli (2003) introduce the notion of ‘wavelet footprints’ to track discontinuities in a signal across different scales (handles persistence)

'2nd Generation' Denoising: III

  – Huerta (2005) proposes a Bayesian approach involving a multivariate normal prior with a covariance matrix that accounts for clustering

- '1st generation' denoising also suffers from problem of overall significance of multiple hypothesis tests
- '2nd generation' work integrates idea of 'false discovery rate' (Benjamini and Hochberg, 1995) into denoising (see Wink and Roerdink, 2004, for a recent applications-oriented discussion)

A Sampling of Recent Advances: I

- Aykroyd and Mardia (2003) give a way to describe shape change and shape differences in curves, by constructing a deformation function in terms of a wavelet decomposition
- Johnstone *et al.* (2004) propose a deconvolution method involving a combination of fast Fourier and fast wavelet transforms that can recover a blurred function observed in white noise, with application to simulated light detection and ranging data suggested by underwater remote sensing
- Oh and Li (2004) use a spherical wavelet approach developed for multiscale representation and analysis of scattered data to estimate the entire temperature field for every location on the globe from scattered surface air temperatures observed by a network of weather-stations

A Sampling of Recent Advances: II

- Gupta *et al.* (2005) consider the problem of constructing wavelets that are specifically adapted to extract certain types of signals
- Ray and Mallick (2006) discuss clustering functional data using a Bayesian wavelet-based approach
- Morris and Carroll (2006) consider wavelet-based functional mixed models
- Fryzlewicz and Nason (2006) discuss estimation of wavelet variance in a time-varying context (Priestley's evolutionary spectra)
- ... and just for fun, see the 1 August 2006 edition of the *New York Times* for an article by Gretchen Cuda about kaleidoscopic art created by Mark Fischer based upon a wavelet transform of whale songs!

Future Directions

- wavelet-based bootstrapping (start provided by Sabatini, 1999; Percival *et al.*, 2000; Bullmore *et al.*, 2001; Breakspear *et al.*, 2004; and Whitcher, 2006)
- prediction (start provided by Wong *et al.*, 2003, and Fryzlewicz *et al.*, 2003)
- handling unequally and sparsely sampled data (particularly spatial)
- random fields (texture analysis – forthcoming work by Mondal)
- statistical significance of transient events

References: I

- R. G. Aykroyd and K. V. Mardia (2003), ‘A Wavelet Approach to Shape Analysis for Spinal Curves,’ *Journal of Applied Statistics*, **30**(6), pp. 605–23
- Y. Benjamini and Y. Hochberg (1995), ‘Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing,’ *Journal of the Royal Statistical Society, Series B*, **57**, pp. 289300
- M. Breakspear, M. J. Brammer, E. T. Bullmore, P. Das and L. M. Williams (2004), ‘Spatiotemporal Wavelet Resampling for Functional Neuroimaging Data,’ *Human Brain Mapping*, **23**, pp. 1–25
- E. Bullmore, Chris Long, John Suckling, Jalal Fadili, Gemma Calvert, Fernando Zelaya, T. Adrian Carpenter and Mick Brammer (2001), ‘Colored Noise and Computational Inference in Neurophysiological (fMRI) Time Series Analysis: Resampling Methods in Time and Wavelet Domains,’ *Human Brain Mapping*, **12**, pp. 61–78
- T. Cai and B. W. Silverman (2001), ‘Incorporating Information on Neighboring Coefficients into Wavelet Estimation,’ *Sankhya Series B*, **63**, pp. 127–48
- M. S. Crouse, R. D. Nowak and R. G. Baraniuk (1998), ‘Wavelet-Based Statistical Signal Processing Using Hidden Markov Models,’ *IEEE Transactions on Signal Processing*, **46**, pp. 886–902

References: II

- P. L. Dragotti and M. Vetterli (2003), ‘Wavelet Footprints: Theory, Algorithms, and Applications,’ *IEEE Transactions on Signal Processing*, **51**, pp. 1306–23
- P. Fryzlewicz and G. P. Nason (2006), ‘Haar-Fisz Estimation of Evolutionary Wavelet Spectra,’ *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**(4), pp. 611–34
- P. Fryzlewicz, S. Van Bellegem and R. Von Sachs (2003), ‘Forecasting Non-Stationary Time Series by Wavelet Process Modelling,’ *Annals of the Institute of Statistical Mathematics*, **55**(4), pp. 737–64
- A. Gupta, S. D. Joshi and S. Prasad (2005), ‘A New Approach for Estimation of Statistically Matched Wavelet,’ *IEEE Transactions on Signal Processing*, **53**(5), pp. 1778–93
- H.–C. Huang and N. Cressie (2000), ‘Deterministic/Stochastic Wavelet Decomposition for Recovery of Signal from Noisy Data,’ *Technometrics*, **42**, pp. 262–76
- G. Huerta (2005), ‘Multivariate Bayes Wavelet Shrinkage and Applications,’ *Journal of Applied Statistics*, **32**(5), pp. 529–42
- I. M. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo (2004), ‘Wavelet Deconvolution in a Periodic Setting,’ *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**(3), pp. 547–73

References: III

- J. S. Morris and R. J. Carroll (2006), 'Wavelet-Based Functional Mixed Models,' *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**(2), pp. 179–99
- H.–S. Oh and T.–H. Li (2004), 'Estimation of Global Temperature Fields from Scattered Observations by a Spherical-Wavelet-Based Spatially Adaptive Method,' *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**(1), pp. 221–38
- D. B. Percival, S. Sardy and A. C. Davison (2001), 'Wavestrapping Time Series: Adaptive Wavelet-Based Bootstrapping,' in *Nonlinear and Nonstationary Signal Processing*, edited by W. J. Fitzgerald, R. L. Smith, A. T. Walden and P. C. Young. Cambridge, England: Cambridge University Press, pp. 442-70
- S. Ray and B. Mallick (2006), 'Functional Clustering by Bayesian Wavelet Methods,' *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**(2), pp. 305–32
- A. M. Sabatini (1999), 'Wavelet-Based Estimation of $1/f$-Type Signal Parameters: Confidence Intervals Using the Bootstrap,' *IEEE Transactions on Signal Processing*, **47**(12), pp. 3406–9
- B. J Whitcher, 'Wavelet-Based Bootstrapping of Spatial Patterns on a Finite Lattice,' *Computational Statistics & Data Analysis* **50**(9), pp. 2399–421

References: IV

- A. M. Wink and J. B. T. M. Roerdink (2004), 'Denoising Functional MR Images: A Comparison of Wavelet Denoising and Gaussian Smoothing,' *IEEE Transactions on Medical Imaging*, **23**(3), pp. 374–87
- H. Wong, W.–C. Ip, Z. Xie, X. Lui (2003), 'Modelling and Forecasting by Wavelets, and the Application to Exchange Rates,' *Journal of Applied Statistics*, **30**(5), pp. 537–53