Wavelet-Based Simulation of Stochastic Processes

Don Percival

Applied Physics Lab, University of Washington

overheads available via

`http://www.staff.washington.edu/dbp`

under "Recent Talks"

Motivation: Simulation of Gaussian Stationary Processes

- let $\{X_t\}$ be a Gaussian stationary process
  - assume t to be discrete: $t = \ldots, -1, 0, 1, \ldots$
  - any finite subset of $\{X_t\}$ is multivariate normal
  - $\text{cov}\{X_t, X_{t+\tau}\} = s_{X,\tau}$ for all t and τ (defines autocovariance sequence (ACVS))
  - $s_{X,\tau}$ determines spectrum $S_X(f) = \Sigma_\tau \, s_{X,\tau} e^{-i2\pi f \tau}$
- Q: given supply of Gaussian white noise deviates, how can we generate realizations of $X_0, \ldots, X_{N-1}$?
- common approaches fall in two categories
  - time domain (e.g., Cholesky decomposition)
  - frequency domain (e.g., circulant embedding)
- disadvantages to common approaches
  - can be slow for use in 'real-time' or with large N
  - not easy to adapt for non-Gaussian processes
  - not easy to adapt for nonstationary processes
- claim: wavelet-based method attractive alternative

Outline of Remainder of Talk

- overview of discrete wavelet transform (DWT)
  - orthonormal transform
  - localized in time and frequency
  - DWT as decorrelator of 'power law' or '$1/f$ type' processes
- wavelet-based simulation of time series
  - previously proposed wavelet-based scheme (Wornell, 1995; McCoy & Walden, 1996)
  - 'circularization' produces non-Gaussian stationary processes
  - can adapt for time-varying power-law processes
  - can adapt for 'real-time' implementations
- comments on extending scheme to other processes via discrete wavelet packet transforms (DWPTs)

Overview of DWT

- let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ denote a time series (for convenience, assume $N = 2^J$)
- let $\mathcal{W}$ be $N \times N$ orthonormal DWT matrix; i.e., $\mathcal{W}^{-1} = \mathcal{W}^T$ so $\mathcal{W}^T\mathcal{W} = \mathcal{W}\mathcal{W}^T = I_N$
- $\mathbf{W} = \mathcal{W}\mathbf{X}$ is vector of DWT coefficients
- can partition $\mathbf{W}$ as follows:

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

- $\mathbf{W}_j$ 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
- $\mathbf{V}_J$ contains a single scaling coefficient
  - proportion to sample mean of time series

DWT in Terms of Filters: I

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

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

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

  - L is length of unit level filter (must be even)
  - jth level filter has length $L_j = (2^j - 1)(L - 1) + 1$
  - note use of circular filtering

- subsample to obtain wavelet coefficients:

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

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

- Fig. 2: Haar wavelet filters of levels $j = 1, \ldots, 7$
- Fig. 3: D(4), C(6) & LA(8) wavelet filters
  - ‘D’ = Daubechies’ extremal phase filters
  - ‘C’ & ‘LA’ =‘coiflets’ & ‘least asymmetric’ (have approximately linear phase)

Examples of DWTs: I

- formation of Haar DWT coefficients for $N = 16$:

$$\mathbf{W} = \mathcal{W}\mathbf{X} = \begin{bmatrix}
\bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet
\end{bmatrix}
\begin{bmatrix}
\bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet
\end{bmatrix}$$

Examples of DWTs: II

- formation of D(4) DWT coefficients for $N = 16$:

$$
\mathbf{W} = \mathcal{W}\mathbf{X} = \begin{bmatrix}
\bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet \\
0 & 0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet
\end{bmatrix}
\begin{bmatrix}
\bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet
\end{bmatrix}
$$

- localization in time decreases with increasing L
- there are $N/2^j$ coefficients at level j, of which no more than $L - 2$ are 'boundary' coefficients

DWT in Terms of Filters: II

- 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
- Fig. 4: squared gain functions for $j = 1, 2, 3$ and $L = 2, 4, \ldots, 14$
- band-pass approximation improves as L increases

DWT of FD Processes

- X_t called fractionally differenced (FD) process
  if its spectrum is given by
$$S_X(f) = \frac{\sigma^2}{|2\sin(\pi f)|^{2\delta}},$$
  where $\sigma^2 > 0$
- note: for small f, have $S_X(f) \approx C/|f|^{2\delta}$;
  i.e., power law or '$1/f$ type' process
- if $\delta = 0$, FD process is white noise
- if $0 < \delta < \frac{1}{2}$, process 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 5: DWT of FDP, $\delta = 0.4$

DWT as a Whitening Transform

- since FD process is stationary, $\mathbf{W}_j$ is also (ignoring terms influenced by circularity)
- Fig. 6: SDFs $S_j(\cdot)$ for $\mathbf{W}_j$, $j = 1, 2, 3, 4$
- DWT acts as whitening filter for FD series because SDFs for $\mathbf{W}_j$ are $\approx$ flat over pass-bands $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$
- Figs. 7 & 8: auto- and cross-correlations
- can regard $\mathbf{W}_1, \ldots, \mathbf{W}_J$ as $\approx$ uncorrelated
  - cross-correlation decreases as L increases
  - refinement: model remaining autocorrelation using autoregressive model of low order
- if Gaussian, close to independently distributed
  - $\approx$ IID within given $\mathbf{W}_j$, but not between ($\mathbf{W}_j$ & $\mathbf{W}_{j'}$ can have different variances)

DWT-Based Simulation: I

- basic DWT-based simulation scheme
  - approximate $\mathbf{W}_j$ as Gaussian white noise with zero mean and variance

    $$\int_{-1/2}^{1/2} \mathcal{H}_j(f) S_X(f)\, df \approx \frac{1}{\frac{1}{2^j} - \frac{1}{2^{j+1}}} \int_{1/2^{j+1}}^{1/2^j} S_X(f)\, df \equiv C_j,$$

    where $\mathcal{H}_j(\cdot)$ is squared gain function for $h_{j,l}$
  - approximate $\mathbf{X}$ via

    $$\mathbf{Y}_N \equiv \mathcal{W}_N^T \Lambda_N^{1/2} \mathbf{Z}_N,$$

    where
    * Λ is $N \times N$ diagonal matrix with diagonal

      $$\underbrace{C_1, \ldots, C_1}_{\frac{N}{2} \text{ of these}}, \underbrace{C_2, \ldots, C_2}_{\frac{N}{4} \text{ of these}}, \ldots, \underbrace{C_{J-1}, C_{J-1}}_{2 \text{ of these}}, C_J, C_{J+1}$$

      ($C_{J+1} \propto$ variance of scaling coefficient in $\mathbf{V}_J$)
    * $\mathbf{Z}_N$ is vector of Gaussian white noise (zero mean and unit variance)

- Q: do covariance matrices of $\mathbf{X}$ & $\mathbf{Y}_N$ match up?
  - $\Sigma_{\mathbf{X}}$ is Toeplitz with diagonals $s_{X,0}, \ldots, s_{X,N-1}$
  - Fig. 9: $\Sigma_{\mathbf{Y}}$ is not Toeplitz!

DWT-Based Simulation: II

- can force Toeplitz structure via following 'trick'
  - recall $\mathcal{W}$ treats $\mathbf{X}$ as if it were circular
  - let $\mathcal{T}$ be $N \times N$ 'circular rotation' matrix:

$$\mathcal{T}\begin{bmatrix} Y_0 \\ Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_0 \end{bmatrix}; \quad \mathcal{T}^2\begin{bmatrix} Y_0 \\ Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} Y_2 \\ Y_3 \\ Y_0 \\ Y_1 \end{bmatrix}; \quad \text{etc.}$$

  - let κ be uniformily distributed over $0, \ldots, N-1$
  - define $\widetilde{\mathbf{Y}}_N \equiv \mathcal{T}^\kappa \mathbf{Y}_N$
- can show that $\widetilde{\mathbf{Y}}_N$ has Toeplitz covariance matrix; i.e., $\widetilde{\mathbf{Y}}_N$ is stationary with ACVS $s_{\widetilde{Y},\tau}$
- can argue $s_{\widetilde{Y},N-\tau} = s_{\widetilde{Y},\tau}$ for $\tau = 1, \ldots, N/2$
  - thus $s_{\widetilde{Y},\tau} \not\approx s_{X,\tau}$ for τ close to N
  - can patch up by simulating $\widetilde{\mathbf{Y}}_M$ with $M > N$ and then extracting first N deviates
  - Fig. 10: approximate and target ACVSs
  - Fig. 11: example of simulated FD series

Simulating Non-Gaussian Series

- marginal distributions of $\mathbf{Y}_M$ are Gaussian, but $\widetilde{\mathbf{Y}}_M$ obeys Gaussian mixture model
- let $\mathcal{I}$ be a binomial random variable (RV):

$$\mathbf{P}[\mathcal{I} = 1] = p \text{ and } \mathbf{P}[\mathcal{I} = 0] = 1 - p$$

- let $\mathcal{N}(0, \sigma^2)$ be Gaussian RV, mean 0 & variance σ^2
- example of RV with Gaussian mixture:

$$R = \mathcal{I}\mathcal{N}(0, \sigma_1^2) + (1 - \mathcal{I})\mathcal{N}(0, \sigma_2^2),$$

  where $\sigma_1^2 \neq \sigma_2^2$
- Fig. 12: example of Gaussian mixture density
- generalizes readily to multinomial
- consider replacing $\frac{M}{2^j}$ occurrences of C_j in Λ_M with values whose average is C_j
- Q: how does this affect statistical properties of $\widetilde{\mathbf{Y}}_M$?
  - spectral properties over $[\frac{1}{2^{j+1}}, \frac{1}{2^j}] \approx$ invariant (due to circularization)
  - univariate distribution changes significantly
- Fig. 13: FD processes with different marginals

Simulating Non-Stationary Series

- recall that $W_{j,t} = \overline{W}_{j,2^j(t+1)-1}$, where

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

- if $h_{j,l}$ is coiflet or least asymmetric filter, can associate $W_{j,t}$ with position in time series (apply shifts to turn linear phase into zero phase)
- can thus generate time-dependent C_j to simulate process with time-evolving spectrum (e.g., evolving from $S(f) \propto |f|^{-0.2}$ to $S(f) \propto |f|^{-0.8}$)
- Fig. 14: example of time series with evolving spectrum

‘Real-Time’ Simulation: I

- form of simulated series via inverse Haar DWT:

$$\mathbf{Y}_M = \mathcal{W}_M^T \Lambda_M^{1/2} \mathbf{Z}_M = \begin{bmatrix}
\bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & \bullet & 0 & \bullet & \bullet \\
\bullet & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & \bullet & 0 & \bullet & \bullet \\
0 & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & \bullet & 0 & \bullet & \bullet \\
0 & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & \bullet & 0 & \bullet & \bullet \\
0 & 0 & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & \bullet & 0 & \bullet & \bullet \\
0 & 0 & \bullet & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & \bullet & 0 & \bullet & \bullet \\
0 & 0 & 0 & \bullet & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & \bullet & 0 & \bullet & \bullet \\
0 & 0 & 0 & \bullet & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & \bullet & 0 & \bullet & \bullet \\
0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & \bullet & \bullet & \bullet \\
0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & \bullet & \bullet & \bullet \\
0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & \bullet & \bullet & \bullet \\
0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & \bullet & \bullet & \bullet \\
0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & 0 & \bullet & 0 & \bullet & \bullet & \bullet \\
0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & 0 & \bullet & 0 & \bullet & \bullet & \bullet \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & \bullet & 0 & \bullet & \bullet & \bullet \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \bullet & 0 & 0 & 0 & \bullet & 0 & \bullet & \bullet & \bullet
\end{bmatrix}
\begin{bmatrix}
\bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet
\end{bmatrix}$$

- need $\log_2(M) + 1$ multiplications to compute any Y_t (even less if elements computed recursively)
- only $\log_2(M) + 1$ random deviates required at once

‘Real-Time’ Simulation: II

- can generate series with large N since, e.g., $\log_2(1,073,741,824) = 30$
- for other wavelets (LA(8) etc.), required storage and computational burden increase linearly ($h_{j,l}$ can be generated using approximation scheme)

Beyond Power Law Processes

- DWT ideally suited for power-law processes, but problematic for other processes
- Fig. 15: DWT applied to two autoregressive processes
- key to scheme is finding decorrelating transform
- possible to do so for given spectrum using DWPT (discrete wavelet packet transform)
- Figs. 16, 17 & 18: key ideas behind DWPT

Concluding Remarks

- wavelet-based simulations quite promising, but still some theoretical work to be done
- would welcome opportunities for collaborations involving implementing scheme for applications