Wavelet Methods for Time Series Analysis

Part IV: MODWT and Examples of DWT/MODWT Analysis

- MODWT stands for ‘maximal overlap discrete wavelet transform’ (pronounced ‘mod WT’)
- transforms very similar to the MODWT have been studied in the literature under the following names:
  - undecimated DWT (or nondecimated DWT)
  - stationary DWT
  - translation (or time) invariant DWT
  - redundant DWT
- also related to notions of ‘wavelet frames’ and ‘cycle spinning’
- basic idea: use values removed from DWT by downsampling

Quick Comparison of the MODWT to the DWT

- unlike the DWT, MODWT is not orthonormal (in fact MODWT is highly redundant)
- unlike the DWT, MODWT is defined naturally for all sample sizes (i.e., \(N\) need not be a multiple of a power of two)
- similar to the DWT, can form multiresolution analyses (MRAs) using MODWT, but with certain additional desirable features; e.g., unlike the DWT, MODWT-based MRA has details and smooths that shift along with \(X\) (if \(X\) has detail \(\tilde{D}_j\), then \(T^mX\) has detail \(T^m\tilde{D}_j\))
- similar to the DWT, an analysis of variance (ANOVA) can be based on MODWT wavelet coefficients
- unlike the DWT, MODWT discrete wavelet power spectrum same for \(X\) and its circular shifts \(T^mX\)

DWT Wavelet & Scaling Filters and Coefficients

- recall that we obtain level \(j = 1\) DWT wavelet and scaling coefficients from \(X\) by filtering and downsampling:
  \[
  X \rightarrow H\left(\frac{k}{N}\right) \downarrow 2 \rightarrow W_1 \quad \text{and} \quad X \rightarrow G\left(\frac{k}{N}\right) \downarrow 2 \rightarrow V_1
  \]
- transfer functions \(H(\cdot)\) and \(G(\cdot)\) are associated with impulse response sequences \(\{h_l\}\) and \(\{g_l\}\) via the usual relationships
  \[
  \{h_l\} \leftrightarrow H(\cdot) \quad \text{and} \quad \{g_l\} \leftrightarrow G(\cdot)
  \]

Level \(j\) Equivalent Wavelet & Scaling Filters

- for any level \(j\), rather than using the pyramid algorithm, we could get the DWT wavelet and scaling coefficients directly from \(X\) by filtering and downsampling:
  \[
  X \rightarrow H_j\left(\frac{k}{N}\right) \downarrow 2^j \rightarrow W_j \quad \text{and} \quad X \rightarrow G_j\left(\frac{k}{N}\right) \downarrow 2^j \rightarrow V_j
  \]
- transfer functions \(H_j(\cdot)\) & \(G_j(\cdot)\) depend just on \(H(\cdot)\) & \(G(\cdot)\)
  - actually can say ‘just on \(H(\cdot)\)’ since \(G(\cdot)\) depends on \(H(\cdot)\)
  - note that \(H_1(\cdot)\) & \(G_1(\cdot)\) are the same as \(H(\cdot)\) & \(G(\cdot)\)
- impulse response sequences \(\{h_{j,l}\}\) and \(\{g_{j,l}\}\) are associated with transfer functions via the usual relationships
  \[
  \{h_{j,l}\} \leftrightarrow H_j(\cdot) \quad \text{and} \quad \{g_{j,l}\} \leftrightarrow G_j(\cdot),
  \]
  and both filters have width \(L_j = (2^j - 1)(L - 1) + 1\)
Haar Equivalent Wavelet & Scaling Filters

- \{h_1\}
- \{h_{2,1}\}
- \{h_{3,1}\}
- \{h_{4,1}\}
- \{g_1\}
- \{g_{2,1}\}
- \{g_{3,1}\}
- \{g_{4,1}\}

- \(L_j = 2^j\) is width of \(\{h_{j,1}\}\) and \(\{g_{j,1}\}\)

D(4) Equivalent Wavelet & Scaling Filters

- \{h_1\}
- \{h_{2,1}\}
- \{h_{3,1}\}
- \{h_{4,1}\}
- \{g_1\}
- \{g_{2,1}\}
- \{g_{3,1}\}
- \{g_{4,1}\}

- \(L_j\) dictated by general formula \(L_j = (2^j - 1)(L - 1) + 1\),
  but can argue that effective width is \(2^j\) (same as Haar \(L_j\))

D(6) Equivalent Wavelet & Scaling Filters

- \{h_1\}
- \{h_{2,1}\}
- \{h_{3,1}\}
- \{h_{4,1}\}
- \{g_1\}
- \{g_{2,1}\}
- \{g_{3,1}\}
- \{g_{4,1}\}

- \(\{h_{4,1}\}\) resembles discretized version of Mexican hat wavelet

LA(8) Equivalent Wavelet & Scaling Filters

- \{h_1\}
- \{h_{2,1}\}
- \{h_{3,1}\}
- \{h_{4,1}\}
- \{g_1\}
- \{g_{2,1}\}
- \{g_{3,1}\}
- \{g_{4,1}\}

- \(\{h_{j,1}\}\) resembles discretized version of Mexican hat wavelet,
  again with an effective width of \(2^j\)
Squared Gain Functions for Equivalent Filters

- squared gain functions give us frequency domain properties:
  \[ H_j(f) \equiv |H_j(f)|^2 \quad \text{and} \quad G_j(f) \equiv |G_j(f)|^2 \]
- example: squared gain functions for LA(8) \( J_0 = 4 \) partial DWT

\[ H_4(f) \begin{bmatrix} \cdot \end{bmatrix} \]
\[ G_4(f) \begin{bmatrix} \cdot \end{bmatrix} \]
\[ H_3(f) \begin{bmatrix} \cdot \end{bmatrix} \]
\[ G_3(f) \begin{bmatrix} \cdot \end{bmatrix} \]
\[ H_2(f) \begin{bmatrix} \cdot \end{bmatrix} \]
\[ G_2(f) \begin{bmatrix} \cdot \end{bmatrix} \]
\[ H_1(f) \begin{bmatrix} \cdot \end{bmatrix} \]
\[ G_1(f) \begin{bmatrix} \cdot \end{bmatrix} \]

Definition of MODWT Wavelet & Scaling Filters

- define MODWT filters \( \{\tilde{h}_{j,l}\} \) and \( \{\tilde{g}_{j,l}\} \) by renormalizing the DWT filters (widths of MODWT & DWT filters are the same):
  \[ \tilde{h}_{j,l} = h_{j,l}/2^{j/2} \quad \text{and} \quad \tilde{g}_{j,l} = g_{j,l}/2^{j/2} \]
- whereas DWT filters have unit energy, MODWT filters satisfy
  \[ \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} \tilde{h}_{j,l}^2 = \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} \tilde{g}_{j,l}^2 = 1 \]
- let \( \tilde{H}_j(\cdot) \) and \( \tilde{G}_j(\cdot) \) be the corresponding transfer functions:
  \[ \tilde{H}_j(f) = \frac{1}{2^{j/2}} H_j(f) \quad \text{and} \quad \tilde{G}_j(f) = \frac{1}{2^{j/2}} G_j(f) \]
  so that
  \[ \{\tilde{h}_{j,l}\} \leftrightarrow \tilde{H}_j(\cdot) \quad \text{and} \quad \{\tilde{g}_{j,l}\} \leftrightarrow \tilde{G}_j(\cdot) \]

Definition of MODWT Coefficients: I

- level \( j \) MODWT wavelet and scaling coefficients are defined to be output obtained by filtering \( X \) with \( \{\tilde{h}_{j,l}\} \) and \( \{\tilde{g}_{j,l}\} \):
  \[ X \rightarrow \tilde{H}_j(\frac{k}{N}) \rightarrow \tilde{W}_j \quad \text{and} \quad X \rightarrow \tilde{G}_j(\frac{k}{N}) \rightarrow \tilde{V}_j \]
- compare the above to its DWT equivalent:
  \[ X \rightarrow H_j(\frac{k}{N}) \rightarrow W_j \quad \text{and} \quad X \rightarrow G_j(\frac{k}{N}) \rightarrow V_j \]
- DWT and MODWT have different normalizations for filters, and there is no downsampling by \( 2^j \) in the MODWT
- level \( J_0 \) MODWT consists of \( J_0 + 1 \) vectors, namely, \( \tilde{W}_1, \tilde{W}_2, \ldots, \tilde{W}_{J_0} \) and \( \tilde{V}_{J_0} \), each of which has length \( N \)

Definition of MODWT Coefficients: II

- MODWT of level \( J_0 \) has \((J_0 + 1)N\) coefficients, whereas DWT has \( N\) coefficients for any given \( J_0 \)
- whereas DWT of level \( J_0 \) requires \( N \) to be integer multiple of \( 2^{J_0} \), MODWT of level \( J_0 \) is well-defined for any sample size \( N \)
- when \( N \) is divisible by \( 2^{J_0} \), we can write
  \[ W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{2^j(t+1) - l \mod N} \quad \text{and} \quad \tilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \mod N} \]
  we have the relationship
  \[ W_{j,t} = 2^{j/2} \tilde{W}_{j,2^j(t+1)-1} \quad \text{and} \quad V_{J_0,t} = 2^{J_0/2} \tilde{V}_{J_0,2^{J_0}(t+1)-1} \]
  (here \( \tilde{W}_{j,t} \) & \( \tilde{V}_{J_0,t} \) denote the \( t \)th elements of \( \tilde{W}_j \) & \( \tilde{V}_{J_0} \))
Properties of the MODWT

- as was true with the DWT, we can use the MODWT to obtain
  - a scale-based additive decomposition (MRA) and
  - a scale-based energy decomposition (ANOVA)
- in addition, the MODWT can be computed efficiently via a pyramid algorithm

MODWT Multiresolution Analysis: I

- starting from the definition
  \[ \widetilde{W}_j = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \mod N} \]
  can write \( \widetilde{W}_j = \sum_{l=0}^{N-1} \tilde{h}_{j,l} X_{t-l \mod N} \)
  where \( \{ \tilde{h}_{j,l} \} \) is \( \{ h_{j,l} \} \) periodized to length \( N \)
- can express the above in matrix notation as \( \widetilde{W}_j = \tilde{W}_j X \), where \( \tilde{W}_j \) is the \( N \times N \) matrix given by
  \[
  \begin{bmatrix}
  \tilde{h}_{j,0} & \tilde{h}_{j,1} & \cdots & \tilde{h}_{j,N-1} \\
  \tilde{h}_{j,1} & \tilde{h}_{j,0} & \cdots & \tilde{h}_{j,N-2} \\
  \vdots & \vdots & \ddots & \vdots \\
  \tilde{h}_{j,N-1} & \tilde{h}_{j,N-2} & \cdots & \tilde{h}_{j,0}
  \end{bmatrix}
  \]

MODWT Multiresolution Analysis: II

- recalling the DWT relationship \( D_j = W_j^T W_j \), define \( j \)th level MODWT detail as \( \tilde{D}_j = \tilde{W}_j^T \tilde{W}_j \)
- similar development leads to definition for \( j \)th level MODWT smooth as \( \tilde{S}_j = \tilde{V}_j^T \tilde{V}_j \)
- can show that level \( J_0 \) MODWT-based MRA is given by
  \[ X = \sum_{j=1}^{J_0} \tilde{D}_j + \tilde{S}_{J_0} \]
  which is analogous to the DWT-based MRA

MODWT Multiresolution Analysis: III

- if we form DWT-based MRAs for \( X \) and its circular shifts \( T^m X, m = 1, \ldots, N - 1 \), we can obtain \( \tilde{D}_j \) by appropriately averaging all \( N \) DWT-based details (‘cycle spinning’)
MODWT Multiresolution Analysis: IV

- left-hand plots show $\tilde{D}_j$, while right-hand plots show average of $T^{-m}D_j$ in MRA for $T^mX$, $m = 0, 1, \ldots, 15$

MODWT Decomposition of Energy

- for any $J_0 \geq 1$ & $N \geq 1$, can show that
  \[ ||X||^2 = \sum_{j=1}^{J_0} ||\tilde{W}_j||^2 + ||\tilde{V}_{J_0}||^2, \]
  leading to an analysis of the sample variance of $X$:
  \[ \hat{\sigma}_X^2 = \frac{1}{N} \sum_{j=1}^{J_0} ||\tilde{W}_j||^2 + \frac{1}{N} ||\tilde{V}_{J_0}||^2 - X^2, \]
  which is analogous to the DWT-based analysis of variance

MODWT Pyramid Algorithm

- goal: compute $\tilde{W}_j$ & $\tilde{V}_j$ using $\tilde{V}_{j-1}$ rather than $X$
- letting $\tilde{V}_{0,t} = X_t$, can show that, for all $j \geq 1$,
  \[ \tilde{W}_{j,t} = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{V}_{j-1,t-2^j-l \mod N} \]
  and \[ \tilde{V}_{j,t} = \sum_{l=0}^{L-1} \tilde{g}_l \tilde{V}_{j-1,t-2^j-l \mod N} \]
- inverse pyramid algorithm is given by
  \[ \tilde{V}_{j-1,t} = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{W}_{j,t+2^j-l \mod N} + \sum_{l=0}^{L-1} \tilde{g}_l \tilde{V}_{j,t+2^j-l \mod N} \]
- algorithm requires $N \log_2(N)$ multiplications, which is the same as needed by fast Fourier transform algorithm

Example of $J_0 = 4$ LA(8) MODWT

- oxygen isotope records $X$ from Antarctic ice core
Relationship Between MODWT and DWT

- bottom plot shows $W_4$ from DWT after circular shift $T^{-3}$ to align coefficients properly in time
- top plot shows $\tilde{W}_4$ from MODWT and subsamples that, upon rescaling, yield $W_4$ via $W_4,t = 4\tilde{W}_{4,16(t+1)}-1$

Example of Variance Decomposition

- decomposition of sample variance from MODWT
  $$\hat{\sigma}^2_X \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \bar{X})^2 = \sum_{j=1}^{4} \frac{1}{N} \|\tilde{W}_j\|^2 + \frac{1}{N} \|\tilde{V}_4\|^2 - \bar{X}^2$$
- LA(8)-based example for oxygen isotope records $X$
  - 0.5 year changes: $\frac{1}{N} \|\tilde{W}_1\|^2 = 0.145 (\approx 4.5\%$ of $\hat{\sigma}^2_X$)
  - 1.0 years changes: $\frac{1}{N} \|\tilde{W}_2\|^2 = 0.500 (\approx 15.6\%$)
  - 2.0 years changes: $\frac{1}{N} \|\tilde{W}_3\|^2 = 0.751 (\approx 23.4\%$)
  - 4.0 years changes: $\frac{1}{N} \|\tilde{W}_4\|^2 = 0.839 (\approx 26.2\%$)
  - 8.0 years averages: $\frac{1}{N} \|\tilde{V}_4\|^2 - \bar{X}^2 = 0.969 (\approx 30.2\%$)
  - sample variance: $\hat{\sigma}^2_X \approx 3.204$

Example of $J_0 = 4$ LA(8) MODWT MRA

- oxygen isotope records $X$ from Antarctic ice core
- LA(8)-based example for oxygen isotope records
- $\tilde{S}_1$, $\tilde{D}_4$, $\tilde{D}_3$, $\tilde{D}_2$, $\tilde{D}_1$, $X$

Summary of Key Points about the MODWT

- similar to the DWT, the MODWT offers
  - a scale-based multiresolution analysis
  - a scale-based analysis of the sample variance
  - a pyramid algorithm for computing the transform efficiently
- unlike the DWT, the MODWT is
  - defined for all sample sizes (no ‘power of 2’ restrictions)
  - unaffected by circular shifts to $X$ in that coefficients, details and smooths shift along with $X$ (example coming later)
  - highly redundant in that a level $J_0$ transform consists of $(J_0 + 1)N$ values rather than just $N$
- as we shall see, the MODWT can eliminate ‘alignment’ artifacts, but its redundancies are problematic for some uses
Examples of DWT & MODWT Analysis: Overview

- look at DWT analysis of electrocardiogram (ECG) data
- discuss potential alignment problems with the DWT and how they are alleviated with the MODWT
- look at MODWT analysis of ECG data, subtidal sea level fluctuations, Nile River minima and ocean shear measurements
- discuss practical details
  - choice of wavelet filter and of level $J_0$
  - handling boundary conditions
  - handling sample sizes that are not multiples of a power of 2
  - definition of DWT not standardized

Electrocardiogram Data: I

- ECG measurements $X$ taken during normal sinus rhythm of a patient who occasionally experiences arrhythmia (data courtesy of Gust Bardy and Per Reinhall, University of Washington)
- $N = 2048$ samples collected at rate of 180 samples/second; i.e., $\Delta t = 1/180$ second
- 11.38 seconds of data in all
- time of $X_0$ taken to be $t_0 = 0.31$ merely for plotting purposes

Electrocardiogram Data: II

- features include
  - baseline drift (not directly related to heart)
  - intermittent high-frequency fluctuations (again, not directly related to heart)
  - ‘PQRST’ portion of normal heart rhythm
- provides useful illustration of wavelet analysis because there are identifiable features on several scales

Electrocardiogram Data: III

- partial DWT coefficients $W$ of level $J_0 = 6$ for ECG time series using the Haar, D(4) and LA(8) wavelets (top to bottom)
Electrocardiogram Data: IV

- elements $W_n$ of $W$ are plotted versus $n = 0, \ldots, N - 1 = 2047$
- vertical dotted lines delineate 7 subvectors $W_1, \ldots, W_6$ & $V_6$
- sum of squares of 2048 coefficients $W$ is equal to those of $X$
- gross pattern of coefficients similar for all three wavelets

Electrocardiogram Data: V

- LA(8) DWT coefficients stacked by scale and aligned with time
- spacing between major tick marks is the same in both plots

Electrocardiogram Data: VI

- R waves aligned with spikes in $W_2$ and $W_3$
- intermittent fluctuations appear mainly in $W_1$ and $W_2$
- setting $J_0 = 6$ results in $V_6$ capturing baseline drift

Electrocardiogram Data: VII

- to quantify how well various DWTs summarize $X$, can form normalized partial energy sequences (NPESs)
- given $\{U_t : t = 0, \ldots, N - 1\}$, square and order such that
  \[ U_2^2(0) \geq U_2^2(1) \geq \cdots \geq U_2^2(N-2) \geq U_2^2(N-1) \]
- $U_2^2(0)$ is largest of all the $U_t^2$ values while $U_2^2(N-1)$ is the smallest
- NPES for $\{U_t\}$ defined as
  \[ C_n \equiv \frac{\sum_{m=0}^{n} U^2_{(m)}}{\sum_{m=0}^{N-1} U^2_{(m)}}, \quad n = 0, 1, \ldots, N - 1 \]
Electrocardiogram Data: VIII

- plots show NPESs for
  - original time series (dashed curve, plot (a))
  - Haar DWT (solid curves, both plots)
  - D(4) DWT (dashed curve, plot (b)); LA(8) is virtually identical
  - DFT (dotted curve, plot (a)) with $|U_t|^2$ rather than $U_t^2$

Electrocardiogram Data: IX

- Haar DWT multiresolution analysis of ECG time series
- blocky nature of Haar basis vectors readily apparent

Electrocardiogram Data: X

- D(4) DWT multiresolution analysis
- ‘shark’s fin’ evident in $D_5$ and $D_6$

Electrocardiogram Data: XI

- LA(8) DWT MRA (shape of filter less prominent here)
- note where features end up (will find MODWT does better)
Effect of Circular Shifts on DWT

- Bottom row: bump X and bump shifted to right by 5 units
- \( J_0 = 4 \) LA(8) DWTs (first 2 columns) and MRAs (last 2)

Effect of Circular Shifts on MODWT

- Unlike the DWT, shifting a time series shifts the MODWT coefficients and components of MRA

Electrocardiogram Data: XII

- Level \( J_0 = 6 \) LA(8) MODWT, with \( \tilde{W}_j \)'s circularly shifted
- Vertical lines delineate 'boundary' coefficients (explained later)

Electrocardiogram Data: XIII

- Comparison of level 6 MODWT and DWT wavelet coefficients, after shifting for time alignment
- Boundary coefficients delineated by vertical red lines
- Subsampling & rescaling \( \tilde{W}_6 \) yields \( W_6 \) (note 'aliasing' effect)
Electrocardiogram Data: XIV

- LA(8) MODWT multiresolution analysis of ECG data

Electrocardiogram Data: XV

- MODWT details seem more consistent across time than DWT details; e.g., $\tilde{D}_6$ does not fade in and out as much as $D_6$
- ‘bumps’ in $D_6$ are slightly asymmetric, whereas those in $\tilde{D}_6$ aren’t

Electrocardiogram Data: XVI

- MODWT coefficients and MRA resemble each other, with latter being necessarily smoother due to second round of filtering
- in the above, $\tilde{S}_6$ is somewhat smoother than $\tilde{V}_6$ and is an intuitively reasonable estimate of the baseline drift

Subtidal Sea Level Fluctuations: I

- subtidal sea level fluctuations $X$ for Crescent City, CA, collected by National Ocean Service with permanent tidal gauge
- $N = 8746$ values from Jan 1980 to Dec 1991 (almost 12 years)
- one value every 12 hours, so $\Delta t = 1/2$ day
- ‘subtidal’ is what remains after diurnal & semidiurnal tides are removed by low-pass filter (filter seriously distorts frequency band corresponding to first physical scale $\tau_1 \Delta t = 1/2$ day)
Subtidal Sea Level Fluctuations: II

- level $J_0 = 7$ LA(8) MODWT multiresolution analysis

Subtidal Sea Level Fluctuations: III

- LA(8) picked in part to help with time alignment of wavelet coefficients, but MRAs for D(4) and C(6) are OK
- Haar MRA suffers from ‘leakage’
- with $J_0 = 7$, $\tilde{S}_7$ represents averages over scale $\lambda_7 \Delta t = 64$ days
- this choice of $J_0$ captures intra-annual variations in $\tilde{S}_7$ (not of interest to decompose these variations further)

Subtidal Sea Level Fluctuations: IV

- expanded view of 1985 and 1986 portion of MRA
- lull in $\tilde{D}_2$, $\tilde{D}_3$ and $\tilde{D}_4$ in December 1985 (associated with changes on scales of 1, 2 and 4 days)

Subtidal Sea Level Fluctuations: V

- MRA suggests seasonally dependent variability at some scales
- because MODWT-based MRA does not preserve energy, preferable to study variability via MODWT wavelet coefficients
- cumulative variance plots for $\tilde{W}_j$ useful tool for studying time dependent variance
- can create these plots for LA or coiflet-based $\tilde{W}_j$ as follows
- form $T^{-|\nu_j(W)|} \tilde{W}_j$, i.e., circularly shift $\tilde{W}_j$ to align with $X$
Subtidal Sea Level Fluctuations: VI

- form normalized cumulative sum of squares:
  \[ C_{j,t} \equiv \frac{1}{N} \sum_{u=0}^{t} \tilde{W}_j^2 \mod N, \quad t = 0, \ldots, N - 1; \]
  note that \( C_{j,N-1} = \|T^{-|\nu_j(H)|} \tilde{W}_j\|^2 / N = \|\tilde{W}_j\|^2 / N \)
- examples for \( j = 2 \) (left-hand plot) and \( j = 7 \) (right-hand)

Nile River Minima: I

- time series \( X \) of minimum yearly water level of the Nile River
- data from 622 to 1284, but actually extends up to 1921
- data after about 715 recorded at the Roda gauge near Cairo
- method(s) used to record data before 715 source of speculation
- oldest time series actually recorded by humans?!

Subtidal Sea Level Fluctuations: VII

- easier to see how variance is building up by subtracting uniform rate of accumulation \( tC_{j,N-1} / (N - 1) \) from \( C_{j,t} \):
  \[ C'_{j,t} \equiv C_{j,t} - tC_{j,N-1} / (N - 1) \]
  yields rotated cumulative variance plots
- \( C'_{2,t} \) and \( C'_{7,t} \) associated with physical scales of 1 and 32 days
- helps build up picture of how variability changes within a year

Nile River Minima: II

- level \( J_0 = 4 \) Haar MODWT MRA points out enhanced variability before 715 at scales \( \tau_1 \Delta t = 1 \) year and \( \tau_2 \Delta t = 2 \) year
- Haar wavelet adequate (minimizes # of boundary coefficients)
Ocean Shear Measurements: I

- $J_0 = 6$ MODWT multiresolution analysis using LA(8) wavelet of vertical shear measurements (in inverse seconds) versus depth (in meters; series collected & supplied by Mike Gregg, Applied Physics Laboratory, University of Washington)

Ocean Shear Measurements: II

- $\Delta t = 0.1$ meters and $N = 6875$
- LA(8) protects against leakage and permits coefficients to be aligned with depth
- $J_0 = 6$ yields smooth $\tilde{S}_6$ that is free of bursts (these are isolated in the details $\tilde{D}_j$)
- note small distortions at beginning/end of $\tilde{S}_6$ evidently due to assumption of circularity
- vertical blue lines delineate subseries of 4096 ‘burst free’ values (to be reconsidered later)
- since MRA is dominated by $\tilde{S}_6$, let’s focus on details alone

Ocean Shear Measurements: III

- $\tilde{D}_j$’s pick out bursts around 450 and 975 meters, but two bursts have somewhat different characteristics
- possible physical interpretation for first burst: turbulence in $\tilde{D}_4$ drives shorter scale turbulence at greater depths
- hints of increased variability in $\tilde{D}_5$ and $\tilde{D}_6$ prior to second burst

Choice of Wavelet Filter: I

- basic strategy: pick wavelet filter with smallest width $L$ that yields an acceptable analysis (smaller $L$ means fewer boundary coefficients)
- very much application dependent
  - LA(8) good choice for MRA of ECG data and for time/depth dependent analysis of variance (ANOVA) of subtidal sea levels and shear data
  - D(4) or LA(8) good choice for MRA of subtidal sea levels, but Haar isn’t (details ‘locked’ together, i.e., are not isolating different aspects of the data)
  - Haar good choice for MRA of Nile River minima
Choice of Wavelet Filter: II

- can often pick $L$ via simple procedure of comparing different MRAs or ANOVAs (this will sometimes rule out Haar if it differs too much from $D(4)$, $D(6)$ or $LA(8)$ analyses)
- for MRAs, might argue that we should pick \( \{h_l\} \) that is a good match to the ‘characteristic features’ in $X$
  - hard to quantify what this means, particularly for time series with different features over different times and scales
  - Haar and $D(4)$ are often a poor match, while the LA filters are usually better because of their symmetry properties
  - can use NPESs to quantify match between \( \{h_l\} \) and $X$
- use LA filters if time alignment of \( \{W_j,t\} \) with $X$ is important (LA filters with even $L/2$, i.e., 8, 12, 16 or 20, yield better alignment than those with odd $L/2$)

Choice of Level $J_0$: I

- again, very much application dependent, but often there is a clear choice
  - $J_0 = 6$ picked for ECG data because it isolated the baseline drift into $V_6$ and $\tilde{V}_6$, and decomposing this drift further is of no interest in studying heart rhythms
  - $J_0 = 7$ picked for subtidal sea levels because it trapped intra-annual variations in $\tilde{V}_7$ (not of interest to analyze these)
  - $J_0 = 6$ picked for shear data because $\tilde{V}_6$ is free of bursts; i.e., $\tilde{V}_{J_0}$ for $J_0 < 6$ would contain a portion of the bursts
  - $J_0 = 4$ picked for Nile River minima to demonstrate that its time-dependent variance is due to variations on the two smallest scales

IV–57

Choice of Level $J_0$: II

- as $J_0$ increases, there are more boundary coefficients to deal with, which suggests not making $J_0$ too big
- if application doesn’t naturally suggest what $J_0$ should be, an ad hoc (but reasonable) default is to pick $J_0$ such that circularity assumption influences $< 50\%$ of $W_{J_0}$ or $D_{J_0}$ (next topic of discussion)

IV–58

Handling Boundary Conditions: I

- DWT and MODWT treat time series $X$ as if it were circular
- circularity says $X_{N-1}$ is useful surrogate for $X_{-1}$ (sometimes this is OK, e.g., subtidal sea levels, but in general it is questionable)
- first step is to delineate which parts of $W_j$ and $D_j$ are influenced (at least to some degree) by circular boundary conditions
- by considering
  \[
  W_{j,t} = 2^j/2^{j(2j+1)-1} \text{ and } \tilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \mod N},
  \]
can determine that circularity affects
  \[
  W_{j,t}, \quad t = 0, \ldots, L_j' - 1 \text{ with } L_j' \equiv \left( (L-2) \left( 1 - \frac{1}{2^j} \right) \right)
  \]

IV–59

IV–60
Handling Boundary Conditions: II

- can argue that \( L'_1 = \frac{L}{2} - 1 \) and \( L'_j = L - 2 \) for large enough \( j \)
- circularity also affects the following elements of \( D_j \):
  \( t = 0, \ldots, 2^j L'_j - 1 \) and \( t = N - (L_j - 2^j), \ldots, N - 1 \),
  where \( L_j = (2^j - 1)(L - 1) + 1 \)
- for MODWT, circularity affects
  \( \tilde{W}_{j,t}, \ t = 0, \ldots, \min\{L_j - 2, N - 1\} \)
- circularity also affects the following elements of \( \tilde{D}_j \):
  \( t = 0, \ldots, L_j - 2 \) and \( t = N - L_j + 1, \ldots, N - 1 \)

Handling Boundary Conditions: III

- examples of delineating LA(8) DWT boundary coefficients for ECG data and of marking parts of MRA influenced by circularity

Handling Boundary Conditions: IV

- boundary regions increase as the filter width \( L \) increases
- for fixed \( L \), boundary regions in DWT MRAs are smaller than those for MODWT MRAs
- for fixed \( L \), MRA boundary regions increase as \( J_0 \) increases (an exception is the Haar DWT)
- these considerations might influence our choice of \( L \) and DWT versus MODWT

Handling Boundary Conditions: V

- comparison of DWT smooths \( S_6 \) (top 3 plots) and MODWT smooths \( \tilde{S}_6 \) (bottom 3) for ECG data using, from top to bottom within each group, the Haar, D(4) and LA(8) wavelets
Handling Boundary Conditions: VI

• just delineating parts of $W_j$ and $D_j$ that are influenced by circular boundary conditions can be misleading (too pessimistic)
• effective width $\lambda_j = 2\tau_j = 2^j$ of $j$th level equivalent filters can be much smaller than actual width $L_j = (2^j - 1)(L-1) + 1$
• arguably less pessimistic delineations would be to always mark boundaries appropriate for the Haar wavelet (its actual width is the effective width for other filters)

Handling Boundary Conditions: VII

$\{h_1\}$
$\{h_2,1\}$
$\{h_3,1\}$
$\{h_4,1\}$
$\{g_1\}$
$\{g_2,1\}$
$\{g_3,1\}$
$\{g_4,1\}$

• plots of LA(8) equivalent wavelet/scaling filters, with actual width $L_j$ compared to effective width of $2^j$

Handling Boundary Conditions: VIII

• to lessen the impact of boundary conditions, we can use ‘tricks’ from Fourier analysis, which also treats $X$ as if it were circular
  – extend series with $X$ (similar to zero padding)
  – polynomial extrapolations
  – use ‘reflection’ boundary conditions by pasting a reflected (time-reversed) version of $X$ to end of $X$

– note that series so constructed of length $2N$ has same sample mean and sample variance as original series $X$

Handling Boundary Conditions: IX

• comparison of effect of reflection (red/blue) and circular (black) boundary conditions on LA(8) DWT-based MRA for oxygen isotope data
Handling Non-Power of Two Sample Sizes

- not a problem with the MODWT, which is defined naturally for all sample sizes $N$
- partial DWT requires just $N = M2^{J_0}$ rather than $N = 2^J$
- can pad with sample mean $\overline{x}$ etc.
- can truncate down to multiple of $2^{J_0}$
  - truncate at beginning of series & do analysis
  - truncate at end of series & do analysis
  - combine two analyses together
- can use a specialized pyramid algorithm involving at most one special term at each level

Lack of Standard Definition for DWT: I

- our definition of DWT matrix $W$ based upon
  - convolutions rather than inner products
  - odd indexed downsampling rather than even indexed
  - using $(-1)^{l+1}h_{L-1-l}$ to define $g_l$ rather than $(-1)^{l-1}h_{1-l}$
  - ordering coefficients in resulting transform from small to large scale rather than large to small
- choices other than the above are used frequently elsewhere, resulting in DWTs that can differ from what we have presented

Lack of Standard Definition for DWT: II

- two left-hand columns: D(4) DWT matrix $W$ as defined here
- two right-hand columns: S-Plus Wavelets D(4) DWT matrix (after reordering of its row vectors)
- only the scaling coefficient is guaranteed to be the same!!!