Wavelet Methods for Time Series Analysis

Part I: Introduction to Wavelets and Wavelet Transforms

- wavelets are analysis tools for time series and images (mostly)
- following work on continuous wavelet transform by Morlet and co-workers in 1983, Daubechies, Mallat and others introduced discrete wavelet transform (DWT) in 1988
- begin with qualitative description of the DWT
- discuss two key descriptive capabilities of the DWT:
  - multiresolution analysis (an additive decomposition)
  - wavelet variance or spectrum (decomposition of sum of squares)
- look at how DWT is formed based on a wavelet filter
- discuss maximal overlap DWT (MODWT)
Qualitative Description of DWT: I

- let \( \mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T \) be a vector of \( N \) time series values (note: ‘\( T \)’ denotes transpose; i.e., \( \mathbf{X} \) is a column vector)
- assume initially \( N = 2^J \) for some positive integer \( J \) (will relax this restriction later on)
- example of time series with \( N = 16 = 2^4 \):

\[
\mathbf{X} = \begin{bmatrix}
0.2, -0.4, -0.6, -0.5, -0.8, -0.4, -0.9, 0.0, \\
-0.2, 0.1, -0.1, 0.1, 0.7, 0.9, 0.0, 0.3
\end{bmatrix}^T
\]

\[X_t\]
Qualitative Description of DWT: II

• DWT is a linear transform of \( \mathbf{X} \) yielding \( N \) DWT coefficients

• notation: \( \mathbf{W} = \mathcal{W} \mathbf{X} \)
  
  – \( \mathbf{W} \) is vector of DWT coefficients (\( j \)th component is \( W_j \))
  
  – \( \mathcal{W} \) is \( N \times N \) orthonormal transform matrix

• orthonormality says \( \mathcal{W}^T \mathcal{W} = I_N \) (\( N \times N \) identity matrix)

• inverse of \( \mathcal{W} \) is just its transpose, so \( \mathcal{W} \mathcal{W}^T = I_N \) also
Implications of Orthonormality

• let $\mathbf{W}_j^T$ denote the $j$th row of $\mathbf{W}$, where $j = 0, 1, \ldots, N - 1$
• let $\mathbf{W}_{j,l}$ denote $l$th element of $\mathbf{W}_j$
• consider two rows, say, $\mathbf{W}_j^T$ and $\mathbf{W}_k^T$
• orthonormality says

$$\langle \mathbf{W}_j, \mathbf{W}_k \rangle = \sum_{l=0}^{N-1} \mathbf{W}_{j,l} \mathbf{W}_{k,l} = \begin{cases} 1, & \text{when } j = k, \\ 0, & \text{when } j \neq k \end{cases}$$

- $\langle \mathbf{W}_j, \mathbf{W}_k \rangle$ is inner product of $j$th & $k$th rows
- $\langle \mathbf{W}_j, \mathbf{W}_j \rangle = \| \mathbf{W}_j \|^2$ is squared norm (energy) for $\mathbf{W}_j$
Example: the Haar DWT

• $N = 16$ example of Haar DWT matrix $\mathcal{W}$

• note that rows are orthogonal to each other (i.e., inner products are zero)
Haar DWT Coefficients: I

• obtain Haar DWT coefficients $\mathbf{W}$ by premultiplying $\mathbf{X}$ by $\mathbf{W}$:
  \[ \mathbf{W} = \mathcal{W} \mathbf{X} \]

• $j$th coefficient $W_j$ is inner product of $j$th row $\mathcal{W}_j^T$ and $\mathbf{X}$:
  \[ W_j = \langle \mathcal{W}_j \cdot, \mathbf{X} \rangle \]

• can interpret coefficients as difference of averages

• to see this, let
  \[ \bar{X}_t(\lambda) \equiv \frac{1}{\lambda} \sum_{l=0}^{\lambda-1} X_{t-l} = \text{‘scale } \lambda \text{’ average} \]
  
  – note: \( \bar{X}_t(1) = X_t = \text{scale 1 ‘average’} \)
  
  – note: \( \bar{X}_{N-1}(N) = \bar{X} = \text{sample average} \)
Haar DWT Coefficients: II

• consider form $W_0 = \langle \mathcal{W}_0, X \rangle$ takes in $N = 16$ example:

\[
\mathcal{W}_{0,t} X_t \quad \text{sum } \propto \overline{X}_{1}(1) - \overline{X}_{0}(1)
\]

• similar interpretation for $W_1, \ldots, W_{N/2} - 1 = W_7 = \langle \mathcal{W}_7, X \rangle$:

\[
\mathcal{W}_{7,t} X_t \quad \text{sum } \propto \overline{X}_{15}(1) - \overline{X}_{14}(1)
\]
Haar DWT Coefficients: III

• now consider form of $W_{N/2} = W_8 = \langle \mathcal{W}_{8\bullet}, X \rangle$:

\[ \mathcal{W}_{8,t} X_t \quad \text{sum} \propto \bar{X}_3(2) - \bar{X}_1(2) \]

• similar interpretation for $W_{N/2+1}, \ldots, W_{3N/4-1}$
Haar DWT Coefficients: IV

\[ W_{3N/4} = W_{12} = \langle \mathcal{W}_{12}, X \rangle \] takes the following form:

\[ \mathcal{W}_{8,t} X_t \quad \text{sum } \propto X_7(4) - X_3(4) \]

continuing in this manner, come to \( W_{N-2} = \langle \mathcal{W}_{14}, X \rangle \):

\[ \mathcal{W}_{14,t} X_t \quad \text{sum } \propto X_{15}(8) - X_7(8) \]
Haar DWT Coefficients: V

- final coefficient $W_{N-1} = W_{15}$ has a different interpretation:

$$W_{15,t}$$  

$W_{15,t}X_t$  

sum $\propto X_{15}(16)$

- structure of rows in $\mathcal{W}$
  - first $\frac{N}{2}$ rows yield $W_j$’s $\propto$ changes on scale 1
  - next $\frac{N}{4}$ rows yield $W_j$’s $\propto$ changes on scale 2
  - next $\frac{N}{8}$ rows yield $W_j$’s $\propto$ changes on scale 4
  - next to last row yields $W_j \propto$ change on scale $\frac{N}{2}$
  - last row yields $W_j \propto$ average on scale $N$
Structure of DWT Matrices

- \( \frac{N}{2\tau_j} \) wavelet coefficients for scale \( \tau_j \equiv 2^{j-1}, j = 1, \ldots, J \)
  - \( \tau_j \equiv 2^{j-1} \) is standardized scale
  - \( \tau_j \Delta \) is physical scale, where \( \Delta \) is sampling interval
- each \( W_j \) localized in time: as scale ↑, localization ↓
- rows of \( \mathcal{W} \) for given scale \( \tau_j \):
  - circularly shifted with respect to each other
  - shift between adjacent rows is \( 2\tau_j = 2^j \)
- similar structure for DWTs other than the Haar
- differences of averages common theme for DWTs
  - simple differencing replaced by higher order differences
  - simple averages replaced by weighted averages
Two Basic Decompositions Derivable from DWT

• additive decomposition
  – reexpresses $X$ as the sum of $J + 1$ new time series, each of which is associated with a particular scale $\tau_j$
  – called multiresolution analysis (MRA)

• energy decomposition
  – yields analysis of variance across $J$ scales
  – called wavelet spectrum or wavelet variance
Partitioning of DWT Coefficient Vector $W$

- decompositions are based on partitioning of $W$ and $\mathcal{W}$
- partition $W$ into subvectors associated with scale:

$$
W = \begin{bmatrix}
W_1 \\
W_2 \\
\vdots \\
W_j \\
\vdots \\
W_J \\
V_J
\end{bmatrix}
$$

- $W_j$ has $N/2^j$ elements (scale $\tau_j = 2^{j-1}$ changes)
  note: $\sum_{j=1}^{J} \frac{N}{2^j} = \frac{N}{2} + \frac{N}{4} + \cdots + 2 + 1 = 2^J - 1 = N - 1$
- $V_J$ has 1 element, which is equal to $\sqrt{N \cdot X}$ (scale $N$ average)
Example of Partitioning of $W$

- consider time series $X$ of length $N = 16$ & its Haar DWT $W$
Partitioning of DWT Matrix $\mathcal{W}$

- partition $\mathcal{W}$ commensurate with partitioning of $\mathbf{W}$:

$$
\mathcal{W} = \begin{bmatrix}
\mathcal{W}_1 \\
\mathcal{W}_2 \\
\vdots \\
\mathcal{W}_j \\
\vdots \\
\mathcal{W}_J \\
\mathcal{V}_J
\end{bmatrix}
$$

- $\mathcal{W}_j$ is $\frac{N}{2^j} \times N$ matrix (related to scale $\tau_j = 2^{j-1}$ changes)

- $\mathcal{V}_J$ is $1 \times N$ row vector (each element is $\frac{1}{\sqrt{N}}$)
Example of Partitioning of $\mathcal{W}$

- $N = 16$ example of Haar DWT matrix $\mathcal{W}$

- two properties: (a) $\mathbf{W}_j = \mathcal{W}_j \mathbf{X}$ and (b) $\mathcal{W}_j \mathcal{W}_j^T = I_{N/2^j}$

$\mathcal{W}_1$

$\mathcal{W}_2$

$\mathcal{W}_3$

$\mathcal{W}_4$

$\mathcal{V}_4$
DWT Analysis and Synthesis Equations

- recall the DWT analysis equation $\mathbf{W} = \mathbf{W} \mathbf{X}$
- $\mathbf{W}^T \mathbf{W} = \mathbf{I}_N$ because $\mathbf{W}$ is an orthonormal transform
- implies that $\mathbf{W}^T \mathbf{W} = \mathbf{W}^T \mathbf{W} \mathbf{X} = \mathbf{X}$
- yields DWT synthesis equation:

\[
\mathbf{X} = \mathbf{W}^T \mathbf{W} = \begin{bmatrix}
\mathbf{W}_1^T, \mathbf{W}_2^T, \ldots, \mathbf{W}_J^T, \mathbf{V}_J^T
\end{bmatrix}
\begin{bmatrix}
\mathbf{W}_1 \\
\mathbf{W}_2 \\
\vdots \\
\mathbf{W}_J \\
\mathbf{V}_J
\end{bmatrix}
= \sum_{j=1}^{J} \mathbf{W}_j^T \mathbf{W}_j + \mathbf{V}_j^T \mathbf{V}_J
\]
Multiresolution Analysis: I

- synthesis equation leads to additive decomposition:

\[
X = \sum_{j=1}^{J} W_j^T W_j + \nu_j^T \nu_j \equiv \sum_{j=1}^{J} D_j + S_J
\]

- \( D_j \equiv W_j^T W_j \) is portion of synthesis due to scale \( \tau_j \)
- \( D_j \) is vector of length \( N \) and is called \( j \)th ‘detail’
- \( S_J \equiv \nu_j^T \nu_j = X \mathbf{1} \), where \( \mathbf{1} \) is a vector containing \( N \) ones
  (later on we will call this the ‘smooth’ of \( J \)th order)
- additive decomposition called multiresolution analysis (MRA)
**Multiresolution Analysis: II**

- example of MRA for time series of length $N = 16$

![Graph showing multiresolution analysis]

- adding values for, e.g., $t = 14$ in $\mathcal{D}_1, \ldots, \mathcal{D}_4$ & $\mathcal{S}_4$ yields $X_{14}$
Energy Preservation Property of DWT Coefficients

- define ‘energy’ in $X$ as its squared norm:

$$\|X\|^2 = \langle X, X \rangle = X^TX = \sum_{t=0}^{N-1} X_t^2$$

- energy of $X$ is preserved in its DWT coefficients $W$ because

$$\|W\|^2 = W^TW = (WX)^TWX = X^TWTWX = X^TINX = X^TX = \|X\|^2$$

- note: same argument holds for any orthonormal transform
Wavelet Spectrum (Variance Decomposition): I

- let $\bar{X}$ denote sample mean of $X_t$’s: $\bar{X} \equiv \frac{1}{N} \sum_{t=0}^{N-1} X_t$
- let $\hat{\sigma}_X^2$ denote sample variance of $X_t$’s:
  \[
  \hat{\sigma}_X^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \bar{X})^2 = \frac{1}{N} \sum_{t=0}^{N-1} X_t^2 - \bar{X}^2
  \]
  \[
  = \frac{1}{N} \|X\|^2 - \bar{X}^2
  \]
  \[
  = \frac{1}{N} \|W\|^2 - \bar{X}^2
  \]
- since $\|W\|^2 = \sum_{j=1}^{J} \|W_j\|^2 + \|V_J\|^2$ and $\frac{1}{N} \|V_J\|^2 = \bar{X}^2$,
  \[
  \hat{\sigma}_X^2 = \frac{1}{N} \sum_{j=1}^{J} \|W_j\|^2
  \]
Wavelet Spectrum (Variance Decomposition): II

- define discrete wavelet power spectrum:
  \[ P_X(\tau_j) \equiv \frac{1}{N}\|W_j\|^2, \text{ where } \tau_j = 2^{j-1} \]
- gives us a scale-based decomposition of the sample variance:
  \[ \hat{\sigma}_X^2 = \sum_{j=1}^{J} P_X(\tau_j) \]
- in addition, each \( W_{j,t} \) in \( W_j \) associated with a portion of \( X \); i.e., \( W_{j,t}^2 \) offers scale- \& time-based decomposition of \( \hat{\sigma}_X^2 \)
Wavelet Spectrum (Variance Decomposition): III

- wavelet spectra for time series $X$ and $Y$ of length $N = 16$, each with zero sample mean and same sample variance.
Defining the Discrete Wavelet Transform (DWT)

• can formulate DWT via elegant ‘pyramid’ algorithm
• *defines* \( \mathcal{W} \) for non-Haar wavelets (consistent with Haar)
• computes \( \mathbf{W} = \mathcal{W} \mathbf{X} \) using \( O(N) \) multiplications
  – ‘brute force’ method uses \( O(N^2) \) multiplications
  – faster than celebrated algorithm for fast Fourier transform!
    (this uses \( O(N \cdot \log_2(N)) \) multiplications)
• can formulate algorithm using linear filters or matrices
  (two approaches are complementary)
• need to review ideas from theory of linear (time-invariant) filters, which requires some Fourier theory
Fourier Theory for Sequences: I

- let \( \{a_t\} \) denote a real-valued sequence such that \( \sum_t a_t^2 < \infty \)
- discrete Fourier transform (DFT) of \( \{a_t\} \):
  \[
  A(f) \equiv \sum_t a_t e^{-i2\pi ft}
  \]
- \( f \) called frequency: \( e^{-i2\pi ft} = \cos(2\pi ft) - i \sin(2\pi ft) \)
- \( A(f) \) defined for all \( f \), but \( 0 \leq f \leq 1/2 \) is of main interest:
  - \( A(\cdot) \) periodic with unit period, i.e., \( A(f + 1) = A(f) \), all \( f \)
  - \( A(-f) = A^*(f) \), complex conjugate of \( A(f) \)
  - need only know \( A(f) \) for \( 0 \leq f \leq 1/2 \) to know it for all \( f \)
- ‘low frequencies’ are those in lower range of \( [0, 1/2] \)
- ‘high frequencies’ are those in upper range of \( [0, 1/2] \)
Fourier Theory for Sequences: II

- can recover (synthesize) \( \{a_t\} \) from its DFT:

\[
\int_{-1/2}^{1/2} A(f)e^{i2\pi ft} df = a_t;
\]

left-hand side called inverse DFT of \( A(\cdot) \)

- \( \{a_t\} \) and \( A(\cdot) \) are two representations for one ‘thingy’

- large \( |A(f)| \) says \( e^{i2\pi ft} \) important in synthesizing \( \{a_t\} \); i.e.,

\( \{a_t\} \) resembles some combination of \( \cos(2\pi ft) \) and \( \sin(2\pi ft) \)
Convolution of Sequences

• given two sequences \( \{a_t\} \) and \( \{b_t\} \), define their convolution by

\[
c_t \equiv \sum_{u=-\infty}^{\infty} a_u b_{t-u}
\]

• DFT of \( \{c_t\} \) has a simple form, namely,

\[
\sum_{t=-\infty}^{\infty} c_t e^{-i2\pi ft} = A(f)B(f),
\]

where \( A(\cdot) \) is the DFT of \( \{a_t\} \), and \( B(\cdot) \) is the DFT of \( \{b_t\} \); i.e., just multiply two DFTs together!!!
Basic Concepts of Filtering

• convolution & linear time-invariant filtering are same concepts:
  – \{b_t\} is input to filter
  – \{a_t\} represents the filter
  – \{c_t\} is filter output

• flow diagram for filtering: \{b_t\} \rightarrow \boxed{\{a_t\}} \rightarrow \{c_t\}

• \{a_t\} is called impulse response sequence for filter

• its DFT \(A(\cdot)\) is called transfer function

• in general \(A(\cdot)\) is complex-valued, so write \(A(f) = |A(f)|e^{i\theta(f)}\)
  – \(|A(f)|\) defines gain function
  – \(A(f) \equiv |A(f)|^2\) defines squared gain function
  – \(\theta(\cdot)\) called phase function (well-defined at \(f\) if \(|A(f)| > 0\))
Example of a Low-Pass Filter

- consider $b_t = \frac{3}{16} \left( \frac{4}{5} \right) |t| + \frac{1}{20} \left( -\frac{4}{5} \right) |t|$ & $a_t = \begin{cases} \frac{1}{2}, & t = 0 \\ \frac{1}{4}, & t = -1 \text{ or } 1 \\ 0, & \text{otherwise} \end{cases}$

- note: $A(\cdot)$ & $B(\cdot)$ both real-valued ($A(\cdot) =$ its gain function)
Example of a High-Pass Filter

- consider same \( \{b_t\} \), but now let \( a_t = \begin{cases} \frac{1}{2}, & t = 0 \\ -\frac{1}{4}, & t = -1 \text{ or } 1 \\ 0, & \text{otherwise} \end{cases} \)

- note: \( \{a_t\} \) resembles some wavelet filters we’ll see later
The Wavelet Filter: I

• precise definition of DWT begins with notion of wavelet filter
  • let \( \{h_l : l = 0, \ldots, L - 1\} \) be a real-valued filter of width \( L \)
    – both \( h_0 \) and \( h_{L-1} \) must be nonzero
    – for convenience, will define \( h_l = 0 \) for \( l < 0 \) and \( l \geq L \)
    – \( L \) must be even \((2, 4, 6, 8, \ldots)\) for technical reasons (hence ruling out \( \{a_t\} \) on the previous overhead)
The Wavelet Filter: II

\[ \{h_l\} \text{ called a wavelet filter if it has these 3 properties} \]

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: for all nonzero integers \( n \), have
\[ \sum_{l=0}^{L-1} h_l h_{l+2n} = 0 \]

\[ \text{\bullet 2 and 3 together are called the orthonormality property} \]
The Wavelet Filter: III

• summation to zero and unit energy relatively easy to achieve
• orthogonality to even shifts is key property & hardest to satisfy
• define transfer and squared gain functions for wavelet filter:

\[ H(f) \equiv \sum_{l=0}^{L-1} h_l e^{-i2\pi fl} \quad \text{and} \quad \mathcal{H}(f) \equiv |H(f)|^2 \]

• orthonormality property is equivalent to

\[ \mathcal{H}(f) + \mathcal{H}(f + \frac{1}{2}) = 2 \quad \text{for all } f \]

(an elegant – but not obvious! – result)
Haar Wavelet Filter

• simplest wavelet filter is Haar ($L = 2$): $h_0 = \frac{1}{\sqrt{2}}$ & $h_1 = -\frac{1}{\sqrt{2}}$

• note that $h_0 + h_1 = 0$ and $h_0^2 + h_1^2 = 1$, as required

• orthogonality to even shifts also readily apparent

\[ h_l h_{l-2} \quad \text{sum} = 0 \]
D(4) Wavelet Filter: I

- next simplest wavelet filter is D(4), for which $L = 4$:
  
  \[
  h_0 = \frac{1-\sqrt{3}}{4\sqrt{2}} , \quad h_1 = \frac{-3+\sqrt{3}}{4\sqrt{2}} , \quad h_2 = \frac{3+\sqrt{3}}{4\sqrt{2}} , \quad h_3 = \frac{-1-\sqrt{3}}{4\sqrt{2}}
  \]

- ‘D’ stands for Daubechies
- $L = 4$ width member of her ‘extremal phase’ wavelets
- computations show $\sum_l h_l = 0$ & $\sum_l h_l^2 = 1$, as required
- orthogonality to even shifts apparent except for $\pm 2$ case:
D(4) Wavelet Filter: II

- Q: what is rationale for D(4) filter?
- consider $X_t^{(1)} \equiv X_t - X_{t-1} = a_0X_t + a_1X_{t-1}$, where $\{a_0 = 1, a_1 = -1\}$ defines 1st difference filter:
  \[
  \{X_t\} \longrightarrow \{1, -1\} \longrightarrow \{X_t^{(1)}\}
  \]
  - Haar wavelet filter is normalized 1st difference filter
  - $X_t^{(1)}$ is difference between two ‘1 point averages’
- consider filter ‘cascade’ with two 1st difference filters:
  \[
  \{X_t\} \longrightarrow \{1, -1\} \longrightarrow \{1, -1\} \longrightarrow \{X_t^{(2)}\}
  \]
- by considering convolution of $\{1, -1\}$ with itself, can reexpress the above using a single ‘equivalent’ (2nd difference) filter:
  \[
  \{X_t\} \longrightarrow \{1, -2, 1\} \longrightarrow \{X_t^{(2)}\}
  \]

WMTSA: 60–61
D(4) Wavelet Filter: III

- renormalizing and shifting 2nd difference filter yields high-pass filter considered earlier:

\[ a_t = \begin{cases} 
\frac{1}{2}, & t = 0 \\
-\frac{1}{4}, & t = -1 \text{ or } 1 \\
0, & \text{otherwise}
\end{cases} \]

- consider ‘2 point weighted average’ followed by 2nd difference:

\[
\{X_t\} \rightarrow \begin{bmatrix} a, b \end{bmatrix} \rightarrow \begin{bmatrix} 1, -2, 1 \end{bmatrix} \rightarrow \{Y_t\}
\]

- convolution of \{a, b\} and \{1, -2, 1\} yields an equivalent filter, which is how the D(4) wavelet filter arises:

\[
\{X_t\} \rightarrow \begin{bmatrix} h_0, h_1, h_2, h_3 \end{bmatrix} \rightarrow \{Y_t\}
\]
D(4) Wavelet Filter: IV

- using conditions
  1. summation to zero: \( h_0 + h_1 + h_2 + h_3 = 0 \)
  2. unit energy: \( h_0^2 + h_1^2 + h_2^2 + h_3^2 = 1 \)
  3. orthogonality to even shifts: \( h_0 h_2 + h_1 h_3 = 0 \)

  can solve for feasible values of \( a \) and \( b \)

- one solution is \( a = \frac{1+\sqrt{3}}{4\sqrt{2}} \approx 0.48 \) and \( b = \frac{-1+\sqrt{3}}{4\sqrt{2}} \approx 0.13 \)

  (other solutions yield essentially the same filter)

- interpret D(4) filtered output as changes in weighted averages
  - ‘change’ now measured by 2nd difference (1st for Haar)
  - average is now 2 point weighted average (1 point for Haar)
  - can argue that effective scale of weighted average is one
Another Popular Daubechies Wavelet Filter

- LA(8) wavelet filter (‘LA’ stands for ‘least asymmetric’)

  - resembles three-point high-pass filter \([-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}]\) (somewhat)
  - can interpret this filter as cascade consisting of
    - 4th difference filter
    - weighted average filter of width 4, but effective width 1
  - filter output can be interpreted as changes in weighted averages
First Level Wavelet Coefficients: I

- given wavelet filter \( \{h_l\} \) of width \( L \) & time series of length \( N = 2^J \), obtain first level wavelet coefficients as follows

- *circularly* filter \( X \) with wavelet filter to yield output
  \[
  \sum_{l=0}^{L-1} h_l X_{t-l} = \sum_{l=0}^{L-1} h_l X_{t-l \mod N}, \quad t = 0, \ldots, N - 1;
  \]
  i.e., if \( t - l \) does not satisfy \( 0 \leq t - l \leq N - 1 \), interpret \( X_{t-l} \) as \( X_{t-l \mod N} \); e.g., \( X_{-1} = X_{N-1} \) and \( X_{-2} = X_{N-2} \)

- take every other value of filter output to define
  \[
  W_{1,t} \equiv \sum_{l=0}^{L-1} h_l X_{2t+1-l \mod N}, \quad t = 0, \ldots, \frac{N}{2} - 1;
  \]
  \( \{W_{1,t}\} \) formed by *downsampling* filter output by a factor of 2
First Level Wavelet Coefficients: II

- example of formation of \( \{W_{1,t}\} \)

\[
h_i^\circ X_{-l \mod 16} \sum = \]

\[
h_i^\circ \]

\[
X_{-l \mod 16} \]
First Level Wavelet Coefficients: II

• example of formation of \( \{W_{1,t}\} \)
First Level Wavelet Coefficients: II

- example of formation of \( \{W_{1,t}\} \)

\[
h_l \quad \sum =
\]

\[
h_l X_{2-l \mod 16}
\]
First Level Wavelet Coefficients: II

- example of formation of \{W_{1,t}\}

\[ h_{l}^{\circ} X_{3-l \mod 16} \]

\[ \sum = \]
First Level Wavelet Coefficients: II

- example of formation of \( \{ W_{1,t} \} \)

\[
h_l \circ X_{4-l \mod 16} \quad \sum = \]

WMTSA: 70  
I-41
First Level Wavelet Coefficients: II

- example of formation of \( \{W_{1,t}\} \)

\[ h_l^o X_{5-l \mod 16} \]

\[ \sum = \]
First Level Wavelet Coefficients: II

- example of formation of \( \{W_{1,t}\} \)
First Level Wavelet Coefficients: II

- example of formation of \( \{ W_{1,t} \} \)

\[
h_l \odot X_{7-l \mod 16} = \sum \]

WMTSA: 70  I–41
First Level Wavelet Coefficients: II

- example of formation of \( \{W_{1,t}\} \)

\[
h_l^\circ \quad h_l^\circ X_{8-l \mod 16} \quad \sum =
\]

\[X_{8-l \mod 16}\]
First Level Wavelet Coefficients: II

- example of formation of \( \{W_{1,t}\} \)

\[
h_l \circ X_{9-l \mod 16} \quad \sum = \quad \text{graph}
\]
First Level Wavelet Coefficients: II

- example of formation of \{W_{1,t}\}
First Level Wavelet Coefficients: II

- example of formation of \( \{W_{1,t}\} \)

\[
h_l^\circ \quad h_l^\circ X_{11-l \mod 16} \quad \sum = \]

WMTSA: 70

I–41
First Level Wavelet Coefficients: II

• example of formation of \{W_{1,t}\}

\[ h_l \odot X_{12 - l \mod 16} \]
First Level Wavelet Coefficients: II

- example of formation of \{W_{1,t}\}
First Level Wavelet Coefficients: II

- example of formation of \( \{W_{1,t}\} \)

![Diagram showing wavelet coefficients](image-url)
First Level Wavelet Coefficients: II

- Example of formation of \( \{W_{1,t}\} \)

\[
\begin{align*}
&h_l^\circ \\
&X_{15-l \mod 16} \\
&h_l^\circ X_{15-l \mod 16} \\
&\sum = 
\end{align*}
\]
First Level Wavelet Coefficients: II

- example of formation of \( \{W_{1,t}\} \)

\[
\begin{align*}
    h_l & \quad h_l X_{15-l \mod 16} \\
    X_{15-l \mod 16} & \quad \sum = \downarrow 2 \\
    W_{1,t} &
\end{align*}
\]

- \( \{W_{1,t}\} \) are unit scale wavelet coefficients – these are the elements of \( \mathbf{W}_1 \) and first \( N/2 \) elements of \( \mathbf{W} = \mathbf{W_1X} \)

- also have \( \mathbf{W}_1 = \mathbf{W}_1 \mathbf{X} \), with \( \mathbf{W}_1 \) being first \( N/2 \) rows of \( \mathbf{W} \)

- hence elements of \( \mathbf{W}_1 \) dictated by wavelet filter
Upper Half $\mathcal{W}_1$ of Haar DWT Matrix $\mathcal{W}$

- consider Haar wavelet filter ($L = 2$): $h_0 = \frac{1}{\sqrt{2}}$ & $h_1 = -\frac{1}{\sqrt{2}}$
- when $N = 16$, $\mathcal{W}_1$ looks like

$$
\begin{bmatrix}
h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_1 & h_0 & 0 & 0 \\
\end{bmatrix}
$$

- rows obviously orthogonal to each other
Upper Half $\mathcal{W}_1$ of D(4) DWT Matrix $\mathcal{W}$

- when $L = 4 \& N = 16$, $\mathcal{W}_1$ looks like

$$
\begin{bmatrix}
  h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_3 & h_2 \\
  h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
$$

- rows orthogonal because $h_0h_2 + h_1h_3 = 0$

- note: $\langle \mathcal{W}_0\bullet, X \rangle$ yields $W_0 = h_1X_0 + h_0X_1 + h_3X_{14} + h_2X_{15}$

- unlike other coefficients from above, this ‘boundary’ coefficient depends on circular treatment of $X$ (a curse, not a feature!)
Orthonormality of Upper Half of DWT Matrix: I

• can show that, for all $L$ and even $N$,

$$W_{1,t} = \sum_{l=0}^{L-1} h_l X_{2t+1-l \mod N},$$

or, equivalently, $W_1 = W_1 X$

forms half an orthonormal transform; i.e.,

$$W_1 W_1^T = I_{N/2}$$

• Q: how can we construct the other half of $W$?
The Scaling Filter: I

- create scaling (or ‘father wavelet’) filter \( \{g_l\} \) by reversing \( \{h_l\} \) and then changing sign of coefficients with even indices

\[
\{h_l\} \quad \{h_l\} \text{ reversed} \quad \{g_l\}
\]

- 2 filters related by \( g_l \equiv (-1)^{l+1} h_{L-1-l} \) & \( h_l = (-1)^l g_{L-1-l} \)
The Scaling Filter: II

- \( \{g_l\} \) is ‘quadrature mirror’ filter corresponding to \( \{h_l\} \)

- properties 2 and 3 of \( \{h_l\} \) are shared by \( \{g_l\} \):
  2. unit energy:
  \[
  \sum_{l=0}^{L-1} g_l^2 = 1
  \]

  3. orthogonality to even shifts: for all nonzero integers \( n \), have
  \[
  \sum_{l=0}^{L-1} g_l g_{l+2n} = 0
  \]

- scaling & wavelet filters both satisfy orthonormality property
First Level Scaling Coefficients: I

- orthonormality property of \( \{h_l\} \) is all that is needed to prove \( \mathcal{W}_1 \) is half of an orthonormal transform (never used \( \sum_l h_l = 0 \))
- going back and replacing \( h_l \) with \( g_l \) everywhere yields another half of an orthonormal transform
- circularly filter \( \mathbf{X} \) using \( \{g_l\} \) and downsample to define

\[
V_{1,t} \equiv \sum_{l=0}^{L-1} g_l X_{2t+1-l \text{ mod } N}, \quad t = 0, \ldots, \frac{N}{2} - 1
\]

- \( \{V_{1,t}\} \) called scaling coefficients for level \( j = 1 \)
- place these \( N/2 \) coefficients in vector called \( \mathbf{V}_1 \)
First Level Scaling Coefficients: III

- define $\mathbf{v}_1$ in a manner analogous to $\mathbf{w}_1$ so that $\mathbf{v}_1 = \mathbf{v}_1 \mathbf{x}$
- when $L = 4$ and $N = 16$, $\mathbf{v}_1$ looks like

$$
\begin{bmatrix}
g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 \\
g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
$$

- $\mathbf{v}_1$ obeys same orthonormality property as $\mathbf{w}_1$:

$$
\text{similar to } \mathbf{w}_1 \mathbf{w}_1^T = \mathbf{I}_{\frac{N}{2}}, \text{ have } \mathbf{v}_1 \mathbf{v}_1^T = \mathbf{I}_{\frac{N}{2}}
$$
Orthonormality of $\mathcal{V}_1$ and $\mathcal{W}_1$: I

- **Q:** how does $\mathcal{V}_1$ help us?
- **A:** rows of $\mathcal{V}_1$ and $\mathcal{W}_1$ are pairwise orthogonal!
- readily apparent in Haar case:

$$
\begin{align*}
  g_l & \quad g_l h_l \quad \text{sum } = 0 \\
  h_l & \quad
\end{align*}
$$
Orthonormality of $\mathcal{V}_1$ and $\mathcal{W}_1$: II

- let’s check that orthogonality holds for $D(4)$ case also:

\[
\begin{align*}
    g_l & \quad \text{sum} = 0 \\
    h_l & \quad \text{sum} = 0 \\
    h_{l-2} & \quad \text{sum} = 0
\end{align*}
\]
Orthonormality of $\mathcal{V}_1$ and $\mathcal{W}_1$: III

- implies that

$$\mathcal{P}_1 \equiv \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{V}_1 \end{bmatrix}$$

is an $N \times N$ orthonormal matrix since

$$\mathcal{P}_1 \mathcal{P}_1^T = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{V}_1 \end{bmatrix} \begin{bmatrix} \mathcal{W}_1^T, \mathcal{V}_1^T \end{bmatrix} = \begin{bmatrix} \mathcal{W}_1 \mathcal{W}_1^T & \mathcal{W}_1 \mathcal{V}_1^T \\ \mathcal{V}_1 \mathcal{W}_1^T & \mathcal{V}_1 \mathcal{V}_1^T \end{bmatrix} = \begin{bmatrix} I_N & 0_N \\ 0_N & I_N \end{bmatrix} = I_N$$

- if $N = 2$ (not of too much interest!), in fact $\mathcal{P}_1 = \mathcal{W}$

- if $N > 2$, $\mathcal{P}_1$ is an intermediate step: $\mathcal{V}_1$ spans same subspace as lower half of $\mathcal{W}$ and will be further manipulated
Interpretation of Scaling Coefficients: I

- consider Haar scaling filter \((L = 2)\): \(g_0 = g_1 = \frac{1}{\sqrt{2}}\)

- when \(N = 16\), matrix \(\mathcal{V}_1\) looks like

\[
\begin{bmatrix}
g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

- since \(V_1 = \mathcal{V}_1X\), each \(V_{1,t}\) is proportional to a 2 point average:

\[
V_{1,0} = g_1X_0 + g_0X_1 = \frac{1}{\sqrt{2}}X_0 + \frac{1}{\sqrt{2}}X_1 \propto X_1(2) \text{ and so forth}
\]
Interpretation of Scaling Coefficients: II

- reconsider shapes of \( \{g_l\} \) seen so far:

  \begin{align*}
  \text{Haar} & \quad \begin{array}{c}
  \text{Haar Shape}
  \end{array} \\
  \text{D(4)} & \quad \begin{array}{c}
  \text{D(4) Shape}
  \end{array} \\
  \text{LA(8)} & \quad \begin{array}{c}
  \text{LA(8) Shape}
  \end{array}
  \end{align*}

- for \( L > 2 \), can regard \( V_{1,t} \) as proportional to weighted average

- can argue that effective width of \( \{g_l\} \) is 2 in each case; thus scale associated with \( V_{1,t} \) is 2, whereas scale is 1 for \( W_{1,t} \)
Frequency Domain Properties of Scaling Filter

• define transfer and squared gain functions for \( \{g_l\} \)

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

• can argue that \( G(f) = \mathcal{H}(f + \frac{1}{2}) \), which, combined with

\[
\mathcal{H}(f) + \mathcal{H}(f + \frac{1}{2}) = 2,
\]

yields

\[
\mathcal{H}(f) + G(f) = 2
\]
Frequency Domain Properties of \( \{h_l\} \) and \( \{g_l\} \)

- since \( W_1 \) & \( V_1 \) contain output from filters, consider their squared gain functions, recalling that \( H(f) + G(f) = 2 \)

- example: \( H(\cdot) \) and \( G(\cdot) \) for Haar & D(4) filters

- \( \{h_l\} \) is high-pass filter with nominal pass-band \( [1/4, 1/2] \)
- \( \{g_l\} \) is low-pass filter with nominal pass-band \( [0, 1/4] \)
Frequency Domain Properties of \( \{ h_l \} \) and \( \{ g_l \} \)

- since \( W_1 \) & \( V_1 \) contain output from filters, consider their squared gain functions, recalling that \( \mathcal{H}(f) + \mathcal{G}(f) = 2 \)
- example: \( \mathcal{H}(\cdot) \) and \( \mathcal{G}(\cdot) \) for Haar & LA(8) filters

\( \{ h_l \} \) is high-pass filter with nominal pass-band \([1/4, 1/2]\)

\( \{ g_l \} \) is low-pass filter with nominal pass-band \([0, 1/4]\)
Example of Decomposing $X$ into $W_1$ and $V_1$: I

- oxygen isotope records $X$ from Antarctic ice core
Example of Decomposing $X$ into $W_1$ and $V_1$: II

- oxygen isotope record series $X$ has $N = 352$ observations
- spacing between observations is $\Delta \doteq 0.5$ years
- used Haar DWT, obtaining 176 scaling and wavelet coefficients
- scaling coefficients $V_1$ related to averages on scale of $2\Delta$
- wavelet coefficients $W_1$ related to changes on scale of $\Delta$
- coefficients $V_{1,t}$ and $W_{1,t}$ plotted against mid-point of years associated with $X_{2t}$ and $X_{2t+1}$
- note: variability in wavelet coefficients increasing with time (thought to be due to diffusion)
- data courtesy of Lars Karlöf, Norwegian Polar Institute, Polar Environmental Centre, Tromsø, Norway
Reconstructing $X$ from $W_1$ and $V_1$

- in matrix notation, form wavelet & scaling coefficients via

$$\begin{bmatrix} \mathbf{W}_1 \\ \mathbf{V}_1 \end{bmatrix} = \begin{bmatrix} \mathcal{W}_1 X \\ \mathcal{V}_1 X \end{bmatrix} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{V}_1 \end{bmatrix} \mathbf{X} = \mathcal{P}_1 \mathbf{X}$$

- recall that $\mathcal{P}^T_1 \mathcal{P}_1 = I_N$ because $\mathcal{P}_1$ is orthonormal

- since $\mathcal{P}^T_1 \mathcal{P}_1 \mathbf{X} = \mathbf{X}$, premultiplying both sides by $\mathcal{P}^T_1$ yields

$$\mathcal{P}^T_1 \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{V}_1 \end{bmatrix} = \begin{bmatrix} \mathcal{W}_1^T & \mathcal{V}_1^T \end{bmatrix} \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{V}_1 \end{bmatrix} = \mathcal{W}_1^T \mathbf{W}_1 + \mathcal{V}_1^T \mathbf{V}_1 = \mathbf{X}$$

- $\mathcal{D}_1 \equiv \mathcal{W}_1^T \mathbf{W}_1$ is the first level detail

- $\mathcal{S}_1 \equiv \mathcal{V}_1^T \mathbf{V}_1$ is the first level ‘smooth’

- $\mathbf{X} = \mathcal{D}_1 + \mathcal{S}_1$ in this notation
Example of Synthesizing $X$ from $D_1$ and $S_1$

- Haar-based decomposition for oxygen isotope records $X$
First Level Variance Decomposition: I

- recall that ‘energy’ in $\mathbf{X}$ is its squared norm $\|\mathbf{X}\|^2$
- because $\mathcal{P}_1$ is orthonormal, have $\mathcal{P}_1^T \mathcal{P}_1 = I_N$ and hence
  \[ \|\mathcal{P}_1 \mathbf{X}\|^2 = (\mathcal{P}_1 \mathbf{X})^T \mathcal{P}_1 \mathbf{X} = \mathbf{X}^T \mathcal{P}_1^T \mathcal{P}_1 \mathbf{X} = \mathbf{X}^T \mathbf{X} = \|\mathbf{X}\|^2 \]
- can conclude that $\|\mathbf{X}\|^2 = \|\mathbf{W}_1\|^2 + \|\mathbf{V}_1\|^2$ because
  \[ \mathcal{P}_1 \mathbf{X} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{V}_1 \end{bmatrix} \] and hence $\|\mathcal{P}_1 \mathbf{X}\|^2 = \|\mathbf{W}_1\|^2 + \|\mathbf{V}_1\|^2$
- leads to a decomposition of the sample variance for $\mathbf{X}$:
  \[ \hat{\sigma}_X^2 = \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \overline{X})^2 = \frac{1}{N} \|\mathbf{X}\|^2 - \overline{X}^2 = \frac{1}{N} \|\mathbf{W}_1\|^2 + \frac{1}{N} \|\mathbf{V}_1\|^2 - \overline{X}^2 \]
First Level Variance Decomposition: II

• breaks up $\hat{\sigma}_X^2$ into two pieces:
  1. $\frac{1}{N}||W_1||^2$, attributable to changes in averages over scale 1
  2. $\frac{1}{N}||V_1||^2 - \overline{X}^2$, attributable to averages over scale 2
• Haar-based example for oxygen isotope records
  – first piece: $\frac{1}{N}||W_1||^2 \approx 0.295$
  – second piece: $\frac{1}{N}||V_1||^2 - \overline{X}^2 \approx 2.909$
  – sample variance: $\hat{\sigma}_X^2 \approx 3.204$
  – changes on scale of $\Delta \approx 0.5$ years account for 9% of $\hat{\sigma}_X^2$
    (standardized scale 1 corresponds to physical scale $\Delta$)
Summary of First Level of Basic Algorithm

- transforms \( \{ X_t : t = 0, \ldots, N - 1 \} \) into 2 types of coefficients
- \( N/2 \) wavelet coefficients \( \{ W_{1,t} \} \) associated with:
  - \( W_1 \), a vector consisting of first \( N/2 \) elements of \( \mathbf{W} \)
  - changes on scale 1 and nominal frequencies \( \frac{1}{4} \leq |f| \leq \frac{1}{2} \)
  - first level detail \( D_1 \)
  - \( \mathbf{W}_1 \), an \( \frac{N}{2} \times N \) matrix consisting of first \( \frac{N}{2} \) rows of \( \mathbf{W} \)
- \( N/2 \) scaling coefficients \( \{ V_{1,t} \} \) associated with:
  - \( V_1 \), a vector of length \( N/2 \)
  - averages on scale 2 and nominal frequencies \( 0 \leq |f| \leq \frac{1}{4} \)
  - first level smooth \( S_1 \)
  - \( \mathbf{V}_1 \), an \( \frac{N}{2} \times N \) matrix spanning same subspace as last \( N/2 \) rows of \( \mathbf{W} \)
Constructing Remaining DWT Coefficients: I

- have regarded time series $X_t$ as ‘one point’ averages $\overline{X}_t(1)$ over scale of 1
- first level of basic algorithm transforms $X$ of length $N$ into
  - $N/2$ wavelet coefficients $W_1 \propto$ changes on a scale of 1
  - $N/2$ scaling coefficients $V_1 \propto$ averages of $X_t$ on a scale of 2
- in essence basic algorithm takes length $N$ series $X$ related to scale 1 averages and produces
  - length $N/2$ series $W_1$ associated with the same scale
  - length $N/2$ series $V_1$ related to averages on double the scale
Constructing Remaining DWT Coefficients: II

- Q: what if we now treat $V_1$ in the same manner as $X$?
- basic algorithm will transform length $N/2$ series $V_1$ into
  - length $N/4$ series $W_2$ associated with the same scale (2)
  - length $N/4$ series $V_2$ related to averages on twice the scale
- by definition, $W_2$ contains the level 2 wavelet coefficients
- Q: what if we treat $V_2$ in the same way?
- basic algorithm will transform length $N/4$ series $V_2$ into
  - length $N/8$ series $W_3$ associated with the same scale (4)
  - length $N/8$ series $V_3$ related to averages on twice the scale
- by definition, $W_3$ contains the level 3 wavelet coefficients
Constructing Remaining DWT Coefficients: III

- continuing in this manner defines remaining subvectors of $\mathbf{W}$ (recall that $\mathbf{W} = \mathbf{W} \mathbf{X}$ is the vector of DWT coefficients)
- at each level $j$, outputs $\mathbf{W}_j$ and $\mathbf{V}_j$ from the basic algorithm are each half the length of the input $\mathbf{V}_{j-1}$
- length of $\mathbf{V}_j$ given by $N/2^j$
- since $N = 2^J$, length of $\mathbf{V}_J$ is 1, at which point we must stop
- $J$ applications of the basic algorithm defines the remaining subvectors $\mathbf{W}_2, \ldots, \mathbf{W}_J, \mathbf{V}_J$ of DWT coefficient vector $\mathbf{W}$
- overall scheme is known as the ‘pyramid’ algorithm
Scales Associated with DWT Coefficients

• $j$th level of algorithm transforms scale $2^{j-1}$ averages into
  – differences of averages on scale $2^{j-1}$, i.e., wavelet coefficients $W_j$
  – averages on scale $2 \times 2^{j-1} = 2^j$, i.e., scaling coefficients $V_j$
• $\tau_j \equiv 2^{j-1}$ denotes scale associated with $W_j$
  – for $j = 1, \ldots, J$, takes on values $1, 2, 4, \ldots, N/4, N/2$
• $\lambda_j \equiv 2^j = 2^{\tau_j}$ denotes scale associated with $V_j$
  – takes on values $2, 4, 8, \ldots, N/2, N$
Matrix Description of Pyramid Algorithm: I

• form $\frac{N}{2^j} \times \frac{N}{2^{j-1}}$ matrix $\mathcal{B}_j$ in same way as $\frac{N}{2} \times N$ matrix $\mathcal{W}_1$

• when $L = 4$ and $N/2^{j-1} = 16$, have

$$
\mathcal{B}_j = \begin{bmatrix}
  h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_3 & h_2 \\
  h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
$$

• matrix gets us $j$th level wavelet coefficients via $\mathcal{W}_j = \mathcal{B}_j \mathcal{V}_{j-1}$
Matrix Description of Pyramid Algorithm: II

- form $\frac{N}{2^j} \times \frac{N}{2^{j-1}}$ matrix $A_j$ in same way as $\frac{N}{2} \times N$ matrix $V_1$
- when $L = 4$ and $N/2^{j-1} = 16$, have

$$A_j = \begin{bmatrix}
g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 \\
g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0
\end{bmatrix}$$

- matrix gets us $j$th level scaling coefficients via $V_j = A_j V_{j-1}$
Matrix Description of Pyramid Algorithm: III

• if we define $V_0 = X$ and let $j = 1$, then

$$W_j = B_j V_{j-1}$$

reduces to

$$W_1 = B_1 V_0 = B_1 X = \mathcal{W}_1 X$$

because $B_1$ has the same definition as $\mathcal{W}_1$

• likewise, when $j = 1$,

$$V_j = A_j V_{j-1}$$

reduces to

$$V_1 = A_1 V_0 = A_1 X = \mathcal{V}_1 X$$

because $A_1$ has the same definition as $\mathcal{V}_1$
Formation of Submatrices of $\mathcal{W}$: I

- using $V_j = A_j V_{j-1}$ repeatedly and $V_1 = A_1 X$, can write
  \[ W_j = B_j V_{j-1} \]
  \[ = B_j A_{j-1} V_{j-2} \]
  \[ = B_j A_{j-1} A_{j-2} V_{j-3} \]
  \[ = B_j A_{j-1} A_{j-2} \cdots A_1 X \equiv W_j X, \]
  where $W_j$ is $\frac{N}{2^j} \times N$ submatrix of $\mathcal{W}$ responsible for $W_j$

- likewise, can get $1 \times N$ submatrix $V_J$ responsible for $V_J$
  \[ V_J = A_J V_{J-1} \]
  \[ = A_J A_{J-1} V_{J-2} \]
  \[ = A_J A_{J-1} A_{J-2} V_{J-3} \]
  \[ = A_J A_{J-1} A_{J-2} \cdots A_1 X \equiv V_J X \]

- $V_J$ is the last row of $\mathcal{W}$, & all its elements are equal to $1/\sqrt{N}$
Formation of Submatrices of $\mathcal{W}$: II

- have now constructed all of DWT matrix:

\[
\mathcal{W} = \begin{bmatrix}
\mathcal{W}_1 \\
\mathcal{W}_2 \\
\mathcal{W}_3 \\
\mathcal{W}_4 \\
\vdots \\
\mathcal{W}_j \\
\vdots \\
\mathcal{W}_J \\
\mathcal{V}_J
\end{bmatrix} = \begin{bmatrix}
\mathcal{B}_1 \\
\mathcal{B}_2 \mathcal{A}_1 \\
\mathcal{B}_3 \mathcal{A}_2 \mathcal{A}_1 \\
\mathcal{B}_4 \mathcal{A}_3 \mathcal{A}_2 \mathcal{A}_1 \\
\vdots \\
\mathcal{B}_j \mathcal{A}_{j-1} \cdots \mathcal{A}_1 \\
\vdots \\
\mathcal{B}_J \mathcal{A}_{J-1} \cdots \mathcal{A}_1 \\
\mathcal{A}_J \mathcal{A}_{J-1} \cdots \mathcal{A}_1
\end{bmatrix}
\]
Examples of $\mathcal{W}$ and its Partitioning: I

- $N = 16$ case for Haar DWT matrix $\mathcal{W}$

- above agrees with qualitative description given previously
Examples of $\mathcal{W}$ and its Partitioning: II

- $N = 16$ case for D(4) DWT matrix $\mathcal{W}$

- note: elements of last row equal to $1/\sqrt{N} = 1/4$, as claimed
Partial DWT: I

- $J$ repetitions of pyramid algorithm for $X$ of length $N = 2^J$ yields ‘complete’ DWT, i.e., $W = \mathcal{W}X$
- can choose to stop at $J_0 < J$ repetitions, yielding a ‘partial’ DWT of level $J_0$:

$$
\begin{bmatrix}
\mathcal{W}_1 \\
\mathcal{W}_2 \\
\vdots \\
\mathcal{W}_j \\
\vdots \\
\mathcal{W}_{J_0} \\
\mathcal{V}_{J_0}
\end{bmatrix}
\begin{bmatrix}
B_1 \\
B_2A_1 \\
\vdots \\
B_jA_{j-1} \cdots A_1 \\
\vdots \\
B_{J_0}A_{J_0-1} \cdots A_1 \\
A_{J_0}A_{J_0-1} \cdots A_1
\end{bmatrix}
\begin{bmatrix}
W_1 \\
W_2 \\
\vdots \\
W_j \\
\vdots \\
W_{J_0} \\
V_{J_0}
\end{bmatrix}
$$

- $\mathcal{V}_{J_0}$ is $\frac{N}{2^{J_0}} \times N$, yielding $\frac{N}{2^{J_0}}$ coefficients for scale $\lambda_{J_0} = 2^{J_0}$
Partial DWT: II

- only requires $N$ to be integer multiple of $2^J_0$
- partial DWT more common than complete DWT
- choice of $J_0$ is application dependent
- multiresolution analysis for partial DWT:
  \[X = \sum_{j=1}^{J_0} D_j + S_{J_0}\]
  
  $S_{J_0}$ represents averages on scale $\lambda J_0 = 2^J_0$ (includes $\overline{X}$)
- analysis of variance for partial DWT:
  \[\hat{\sigma}_X^2 = \frac{1}{N} \sum_{j=1}^{J_0} \|W_j\|^2 + \frac{1}{N} \|V_{J_0}\|^2 - \overline{X}^2\]
Example of $J_0 = 4$ Partial Haar DWT

- oxygen isotope records $X$ from Antarctic ice core
Example of $J_0 = 4$ Partial Haar DWT

- oxygen isotope records $\mathbf{X}$ from Antarctic ice core
Example of MRA from $J_0 = 4$ Partial Haar DWT

- oxygen isotope records $X$ from Antarctic ice core
Example of Variance Decomposition

- decomposition of sample variance from \( J_0 = 4 \) partial DWT

\[
\hat{\sigma}_X^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \bar{X})^2 = \sum_{j=1}^{4} \frac{1}{N} \| W_j \|^2 + \frac{1}{N} \| V_4 \|^2 - \bar{X}^2
\]

- Haar-based example for oxygen isotope records
  - 0.5 year changes: \( \frac{1}{N} \| W_1 \|^2 \doteq 0.295 (\doteq 9.2\% \text{ of } \hat{\sigma}_X^2) \)
  - 1.0 years changes: \( \frac{1}{N} \| W_2 \|^2 \doteq 0.464 (\doteq 14.5\%) \)
  - 2.0 years changes: \( \frac{1}{N} \| W_3 \|^2 \doteq 0.652 (\doteq 20.4\%) \)
  - 4.0 years changes: \( \frac{1}{N} \| W_4 \|^2 \doteq 0.846 (\doteq 26.4\%) \)
  - 8.0 years averages: \( \frac{1}{N} \| V_4 \|^2 - \bar{X}^2 \doteq 0.947 (\doteq 29.5\%) \)
  - sample variance: \( \hat{\sigma}_X^2 \doteq 3.204 \)
Haar Equivalent Wavelet & Scaling Filters

\[
\{h_1\}, \quad L = 2 \\
\{h_{2,1}\}, \quad L_2 = 4 \\
\{h_{3,1}\}, \quad L_3 = 8 \\
\{h_{4,1}\}, \quad L_4 = 16 \\
\{g_1\}, \quad L = 2 \\
\{g_{2,1}\}, \quad L_2 = 4 \\
\{g_{3,1}\}, \quad L_3 = 8 \\
\{g_{4,1}\}, \quad L_4 = 16
\]

- \(L_j = 2^j\) is width of \(\{h_{j,1}\}\) and \(\{g_{j,1}\}\)
- note: convenient to define \(\{h_{1,1}\}\) to be same as \(\{h_1\}\)
D(4) Equivalent Wavelet & Scaling Filters

- $L_j$ dictated by general formula $L_j = (2^j - 1)(L - 1) + 1$,
  but can argue that effective width is $2^j$ (same as Haar $L_j$)
LA(8) Equivalent Wavelet & Scaling Filters

\{h_1\} \quad L = 8
\{h_{2,1}\} \quad L_2 = 22
\{h_{3,1}\} \quad L_3 = 50
\{h_{4,1}\} \quad L_4 = 106
\{g_1\} \quad L = 8
\{g_{2,1}\} \quad L_2 = 22
\{g_{3,1}\} \quad L_3 = 50
\{g_{4,1}\} \quad L_4 = 106
Maximal Overlap Discrete Wavelet Transform

• abbreviation is MODWT (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 invariant DWT
  – 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 samples 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 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 $\mathcal{D}_j$, then $\mathcal{T}^mX$ has detail $\mathcal{T}^m\mathcal{D}_j$, where $\mathcal{T}^m$ circularly shifts $X$ by $m$ units)
- 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 $\mathcal{T}^mX$
Definition of MODWT Coefficients: I

- define MODWT filters \( \{ \tilde{h}_{j,l} \} \) and \( \{ \tilde{g}_{j,l} \} \) by renormalizing the DWT filters:
  \[
  \tilde{h}_{j,l} = h_{j,l}/2^j/2 \quad \text{and} \quad \tilde{g}_{j,l} = g_{j,l}/2^j/2
  \]

- level \( j \) MODWT wavelet and scaling coefficients are defined to be output obtaining by filtering \( X \) with \( \{ \tilde{h}_{j,l} \} \) and \( \{ \tilde{g}_{j,l} \} \):
  \[
  X \rightarrow \{ \tilde{h}_{j,l} \} \rightarrow \tilde{W}_j \quad \text{and} \quad X \rightarrow \{ \tilde{g}_{j,l} \} \rightarrow \tilde{V}_j
  \]

- compare the above to its DWT equivalent:
  \[
  X \rightarrow \{ h_{j,l} \} \downarrow 2^j \rightarrow W_j \quad \text{and} \quad X \rightarrow \{ g_{j,l} \} \downarrow 2^j \rightarrow V_j
  \]

- level \( J_0 \) MODWT consists of \( J_0 + 1 \) vectors, namely,
  \[
  \tilde{W}_1, \tilde{W}_2, \ldots, \tilde{W}_{J_0} \quad \text{and} \quad \tilde{V}_{J_0},
  \]
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

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{2^j(t+1)-1-l \mod N} \quad \text{and} \quad \tilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \mod N},$$

and we have the relationship

$$W_{j,t} = 2^{j/2} \tilde{W}_{j,2^j(t+1)-1} \quad \text{and, likewise,} \quad V_{J_0,t} = 2^{J_0/2} \tilde{V}_{J_0,2^{J_0}(t+1)-1}$$

(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):
    \[ X = \sum_{j=1}^{J_0} \tilde{D}_j + \tilde{S}_{J_0} \]
  – a scale-based energy decomposition (basis for ANOVA):
    \[ \|X\|^2 = \sum_{j=1}^{J_0} \|\tilde{W}_j\|^2 + \|\tilde{V}_{J_0}\|^2 \]
• in addition, the MODWT can be computed efficiently via a pyramid algorithm
Example of $J_0 = 4$ LA(8) MODWT

- oxygen isotope records $X$ from Antarctic ice core

\[ T^{-45} \tilde{V}_4 \]
\[ T^{-53} \tilde{W}_4 \]
\[ T^{-25} \tilde{W}_3 \]
\[ T^{-11} \tilde{W}_2 \]
\[ T^{-4} \tilde{W}_1 \]

$X$
Relationship Between MODWT and DWT

- bottom plot shows $W_4$ from DWT after circular shift $\mathcal{T}^{-3}$ 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 $J_0 = 4$ LA(8) MODWT MRA

- oxygen isotope records $X$ from Antarctic ice core
Example of Variance Decomposition

- decomposition of sample variance from MODWT

\[
\hat{\sigma}_X^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \bar{X})^2 = \sum_{j=1}^{4} \frac{1}{N} \|\tilde{W}_j\|^2 + \frac{1}{N} \|\tilde{V}_4\|^2 - \bar{X}^2
\]

- LA(8)-based example for oxygen isotope records
  - 0.5 year changes: \( \frac{1}{N} \|\tilde{W}_1\|^2 \doteq 0.145 \ (\doteq 4.5\% \text{ of } \hat{\sigma}_X^2) \)
  - 1.0 years changes: \( \frac{1}{N} \|\tilde{W}_2\|^2 \doteq 0.500 \ (\doteq 15.6\%) \)
  - 2.0 years changes: \( \frac{1}{N} \|\tilde{W}_3\|^2 \doteq 0.751 \ (\doteq 23.4\%) \)
  - 4.0 years changes: \( \frac{1}{N} \|\tilde{W}_4\|^2 \doteq 0.839 \ (\doteq 26.2\%) \)
  - 8.0 years averages: \( \frac{1}{N} \|\tilde{V}_4\|^2 - \bar{X}^2 \doteq 0.969 \ (\doteq 30.2\%) \)
  - sample variance: \( \hat{\sigma}_X^2 \doteq 3.204 \)
Summary of Key Points about the DWT: I

- The DWT $\mathcal{W}$ is orthonormal, i.e., satisfies $\mathcal{W}^T \mathcal{W} = I_N$

- Construction of $\mathcal{W}$ starts with a wavelet filter $\{h_l\}$ of even length $L$ that by definition
  1. Sums to zero; i.e., $\sum_l h_l = 0$;
  2. Has unit energy; i.e., $\sum_l h_l^2 = 1$; and
  3. Is orthogonal to its even shifts; i.e., $\sum_l h_l h_{l+2n} = 0$

- 2 and 3 together called orthonormality property

- Wavelet filter defines a scaling filter via $g_l = (-1)^{l+1} h_{L-1-l}$

- Scaling filter satisfies the orthonormality property, but sums to $\sqrt{2}$ and is also orthogonal to $\{h_l\}$; i.e., $\sum_l g_l h_{l+2n} = 0$

- While $\{h_l\}$ is a high-pass filter, $\{g_l\}$ is a low-pass filter
Summary of Key Points about the DWT: II

• \{h_l\} and \{g_l\} work in tandem to split time series \( \mathbf{X} \) into
  – wavelet coefficients \( \mathbf{W}_1 \) (related to changes in averages on a unit scale) and
  – scaling coefficients \( \mathbf{V}_1 \) (related to averages on a scale of 2)
• \{h_l\} and \{g_l\} are then applied to \( \mathbf{V}_1 \), yielding
  – wavelet coefficients \( \mathbf{W}_2 \) (related to changes in averages on a scale of 2) and
  – scaling coefficients \( \mathbf{V}_2 \) (related to averages on a scale of 4)
• continuing beyond these first 2 levels, scaling coefficients \( \mathbf{V}_{j-1} \) at level \( j - 1 \) are transformed into wavelet and scaling coefficients \( \mathbf{W}_j \) and \( \mathbf{V}_j \) of scales \( \tau_j = 2^{j-1} \) and \( \lambda_j = 2^j \)
Summary of Key Points about the DWT: III

- after $J_0$ repetitions, this ‘pyramid’ algorithm transforms time series $X$ whose length $N$ is an integer multiple of $2^{J_0}$ into DWT coefficients $W_1, W_2, \ldots, W_{J_0}$ and $V_{J_0}$ (sizes of vectors are $N/2$, $N/4$, $\ldots$, $N/2^{J_0}$ and $N/2^{J_0}$, for a total of $N$ coefficients in all)
- DWT coefficients lead to two basic decompositions
- first decomposition is additive and is known as a multiresolution analysis (MRA), in which $X$ is reexpressed as
  \[
  X = \sum_{j=1}^{J_0} D_j + S_{J_0},
  \]
  where $D_j$ is a time series reflecting variations in $X$ on scale $\tau_j$, while $S_{J_0}$ is a series reflecting its $\lambda_{J_0}$ averages
Summary of Key Points about the DWT: IV

- second decomposition reexpresses the energy (squared norm) of $\mathbf{X}$ on a scale by scale basis, i.e.,

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

leading to an analysis of the sample variance of $\mathbf{X}$:

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

$$
= \frac{1}{N} \sum_{j=1}^{J_0} \| \mathbf{W}_j \|^2 + \frac{1}{N} \| \mathbf{V}_{J_0} \|^2 - \bar{X}^2
$$
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 \( X \)
  – highly redundant in that a level \( J_0 \) transform consists of \( (J_0 + 1)N \) values rather than just \( N \)

• MODWT can eliminate ‘alignment’ artifacts, but its redundancies are problematic for some uses