Wavelet Methods for Time Series Analysis

Part II: Introduction to the Discrete Wavelet Transform

- will give precise definition of DWT in Part III
- 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)
- need to assume $N = 2^J$ for some positive integer J (restrictive!)
- 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 (jth component is W_j)
  - $\mathcal{W}$ is $N \times N$ orthonormal transform matrix; i.e., $\mathcal{W}^T\mathcal{W} = I_N$, where I_N is $N \times N$ identity matrix
- inverse of $\mathcal{W}$ is just its transpose, so $\mathcal{W}\mathcal{W}^T = I_N$ also

Implications of Orthonormality: I

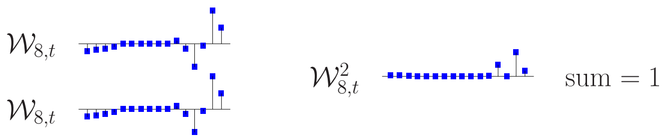
- let $\mathcal{W}_{j\bullet}^T$ denote the jth row of $\mathcal{W}$, where $j = 0, 1, \ldots, N-1$
- note that $\mathcal{W}_{j\bullet}$ itself is a column vector
- let $\mathcal{W}_{j,l}$ denote element of $\mathcal{W}$ in row j and column l
- note that $\mathcal{W}_{j,l}$ is also lth element of $\mathcal{W}_{j\bullet}$
- let's consider two vectors, say, $\mathcal{W}_{j\bullet}$ and $\mathcal{W}_{k\bullet}$
- orthonormality says

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

  - $\langle \mathcal{W}_{j\bullet}, \mathcal{W}_{k\bullet} \rangle$ is inner product of jth & kth rows
  - $\|\mathcal{W}_{j\bullet}\|^2 \equiv \langle \mathcal{W}_{j\bullet}, \mathcal{W}_{j\bullet} \rangle$ is squared norm (energy) for $\mathcal{W}_{j\bullet}$

Implications of Orthonormality: II

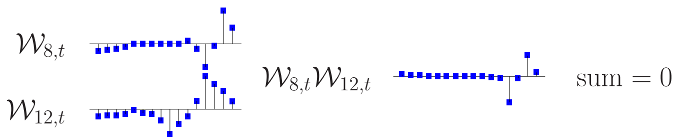
- example from $\mathcal{W}$ of dimension 16×16 we'll see later on
  - inner product of row 8 with itself (i.e., squared norm):



  - row 8 said to have 'unit energy' since squared norm is 1

Implications of Orthonormality: III

- another example from same $\mathcal{W}$
  - inner product of rows 8 and 12:

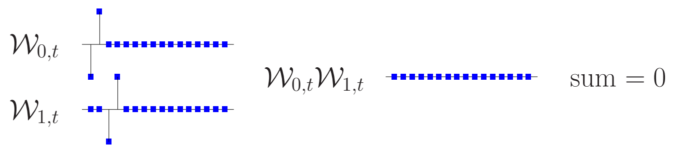


  - rows 8 & 12 said to be orthogonal since inner product is 0

The Haar DWT: I

- like CWT, DWT tell us about variations in local averages
- to see this, let's look inside $\mathcal{W}$ for the Haar DWT for $N = 2^J$
- row $j = 0$: $[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \underbrace{0, \ldots, 0}_{N-2 \text{ zeros}}] \equiv \mathcal{W}_{0\bullet}^T$

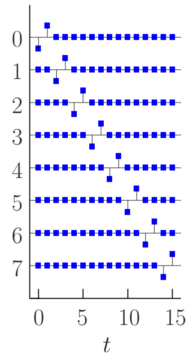
  note: $\|\mathcal{W}_{0\bullet}\|^2 = \frac{1}{2} + \frac{1}{2} = 1$ & hence has required unit energy
- row $j = 1$: $[0, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \underbrace{0, \ldots, 0}_{N-4 \text{ zeros}}] \equiv \mathcal{W}_{1\bullet}^T$
- $\mathcal{W}_{0\bullet}$ and $\mathcal{W}_{1\bullet}$ are orthogonal



The Haar DWT: II

- keep shifting by two to form rows until we come to …
- row $j = \frac{N}{2} - 1$: $[\underbrace{0, \ldots, 0}_{N-2 \text{ zeros}}, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] \equiv \mathcal{W}_{\frac{N}{2}-1\bullet}^T$
- first $N/2$ rows form orthonormal set of $N/2$ vectors

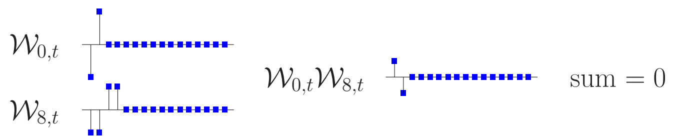
$N = 16$ example



The Haar DWT: III

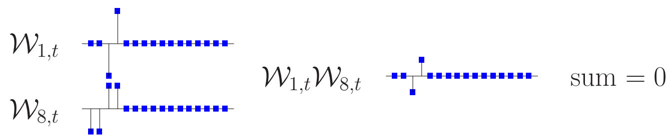
- to form next row, stretch $[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \ldots, 0]$ out by a factor of two and renormalize to preserve unit energy
- $j = \frac{N}{2}$: $[-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \underbrace{0, \ldots, 0}_{N-4 \text{ zeros}}] \equiv \mathcal{W}_{\frac{N}{2}\bullet}^T$

  note: $\|\mathcal{W}_{\frac{N}{2}\bullet}\|^2 = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$, as required
- $\mathcal{W}_{0\bullet}$ and $\mathcal{W}_{\frac{N}{2}\bullet}$ are orthogonal ($\frac{N}{2} = 8$ in example)

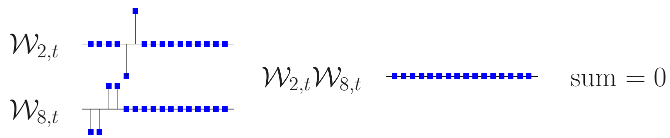


The Haar DWT: IV

- $\mathcal{W}_{1\bullet}$ and $\mathcal{W}_{\frac{N}{2}\bullet}$ are orthogonal

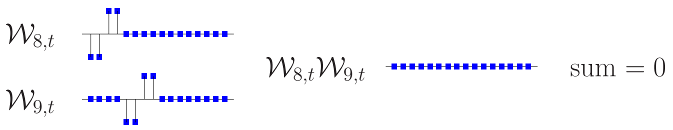


- $\mathcal{W}_{2\bullet}$ and $\mathcal{W}_{\frac{N}{2}\bullet}$ are orthogonal



The Haar DWT: V

- form next row by shifting $\mathcal{W}_{\frac{N}{2}\bullet}$ to right by 4 units
- $j = \frac{N}{2} + 1$: $[0, 0, 0, 0, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \underbrace{0, \ldots, 0}_{N-8 \text{ zeros}}] \equiv \mathcal{W}^T_{\frac{N}{2}+1\bullet}$
- $\mathcal{W}_{\frac{N}{2}+1\bullet}$ orthogonal to first $N/2$ rows and also to $\mathcal{W}_{\frac{N}{2}\bullet}$



- continue shifting by 4 units to form more rows, ending with ...
- row $j = \frac{3N}{4} - 1$: $[\underbrace{0, \ldots, 0}_{N-4 \text{ zeros}}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}] \equiv \mathcal{W}^T_{\frac{3N}{4}-1\bullet}$

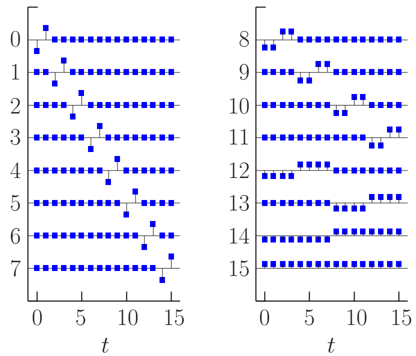
The Haar DWT: VI

- to form next row, stretch $[-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, \ldots, 0]$ out by a factor of two and renormalize to preserve unit energy
- $j = \frac{3N}{4}$: $[\underbrace{-\frac{1}{\sqrt{8}}, \ldots, -\frac{1}{\sqrt{8}}}_{4 \text{ of these}}, \underbrace{\frac{1}{\sqrt{8}}, \ldots, \frac{1}{\sqrt{8}}}_{4 \text{ of these}}, \underbrace{0, \ldots, 0}_{N-8 \text{ zeros}}] \equiv \mathcal{W}^T_{\frac{3N}{4}\bullet}$

  note: $\|\mathcal{W}_{\frac{3N}{4}\bullet}\|^2 = 8 \cdot \frac{1}{8} = 1$, as required
- $j = \frac{3N}{4} + 1$: shift row $\frac{3N}{4}$ to right by 8 units
- continue shifting and stretching until finally we come to ...
- $j = N - 2$: $[\underbrace{-\frac{1}{\sqrt{N}}, \ldots, -\frac{1}{\sqrt{N}}}_{\frac{N}{2} \text{ of these}}, \underbrace{\frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}}}_{\frac{N}{2} \text{ of these}}] \equiv \mathcal{W}^T_{N-2\bullet}$
- $j = N - 1$: $[\underbrace{\frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}}}_{N \text{ of these}}] \equiv \mathcal{W}^T_{N-1\bullet}$

The Haar DWT: VII

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

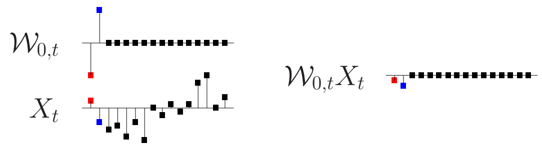


Haar DWT Coefficients: I

- obtain Haar DWT coefficients $\mathbf{W}$ by premultiplying $\mathbf{X}$ by $\mathcal{W}$:
$$\mathbf{W} = \mathcal{W}\mathbf{X}$$
- jth coefficient W_j is inner product of jth row $\mathcal{W}_{j\bullet}$ and $\mathbf{X}$:
$$W_j = \langle \mathcal{W}_{j\bullet}, \mathbf{X} \rangle$$
- can interpret coefficients as difference of averages
- to see this, let
$$\overline{X}_t(\lambda) \equiv \frac{1}{\lambda} \sum_{l=0}^{\lambda-1} X_{t-l} = \text{'scale } \lambda\text{' average}$$
  - note: $\overline{X}_t(1) = X_t$ = scale 1 'average'
  - note: $\overline{X}_{N-1}(N) = \overline{X}$ = sample average

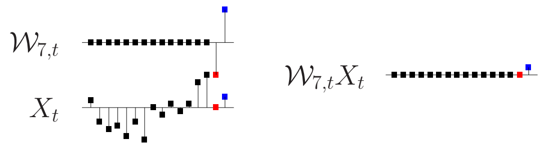
Haar DWT Coefficients: II

- consider form $W_0 = \langle \mathcal{W}_{0\bullet}, \mathbf{X} \rangle$ takes in $N = 16$ example:



sum $\propto \overline{X}_1(1) - \overline{X}_0(1)$

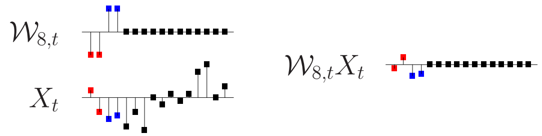
- similar interpretation for $W_1, \ldots, W_{\frac{N}{2}-1} = W_7 = \langle \mathcal{W}_{7\bullet}, \mathbf{X} \rangle$:



sum $\propto \overline{X}_{15}(1) - \overline{X}_{14}(1)$

Haar DWT Coefficients: III

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

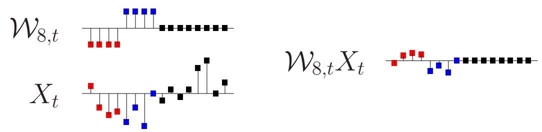


sum $\propto \overline{X}_3(2) - \overline{X}_1(2)$

- similar interpretation for $W_{\frac{N}{2}+1}, \ldots, W_{\frac{3N}{4}-1}$

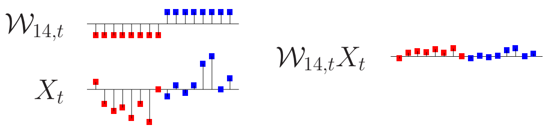
Haar DWT Coefficients: IV

- $W_{\frac{3N}{4}} = W_{12} = \langle \mathcal{W}_{8\bullet}, \mathbf{X} \rangle$ takes the following form:



sum $\propto \overline{X}_7(4) - \overline{X}_3(4)$

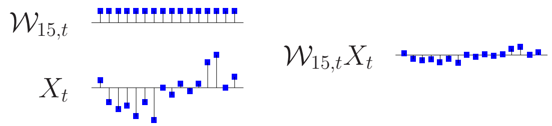
- continuing in this manner, come to $W_{N-1} = \langle \mathcal{W}_{14\bullet}, \mathbf{X} \rangle$:



sum $\propto \overline{X}_{15}(8) - \overline{X}_7(8)$

Haar DWT Coefficients: V

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



sum $\propto \overline{X}_{15}(16)$

- 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
  - $\tau_j \, \Delta t$ is physical scale, where Δt is sampling interval
- each W_j localized in time: as scale $\uparrow$, localization $\downarrow$
- rows of $\mathcal{W}$ for given scale τ_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 $\mathbf{X}$ as the sum of $J+1$ new time series, each of which is associated with a particular scale τ_j
  - called multiresolution analysis (MRA)
  - related to first 'scary-looking' CWT equation
- energy decomposition
  - yields analysis of variance across J scales
  - called wavelet spectrum or wavelet variance
  - related to second 'scary-looking' CWT equation

Partitioning of DWT Coefficient Vector W

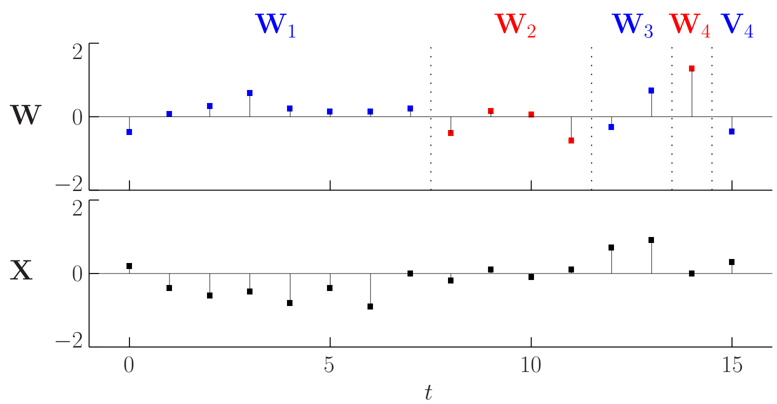
- decompositions are based on partitioning of $\mathbf{W}$ and $\mathcal{W}$
- partition $\mathbf{W}$ into subvectors associated with scale:

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

- $\mathbf{W}_j$ has $N/2^j$ elements (scale $\tau_j = 2^{j-1}$ changes)
  note: $\sum_{j=1}^{J} \frac{N}{2^j} = \frac{N}{2} + \frac{N}{4} + \cdots + 2 + 1 = 2^J - 1 = N - 1$
- $\mathbf{V}_J$ has 1 element, which is equal to $\sqrt{N} \cdot \overline{X}$ (scale N average)

Example of Partitioning of W

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



Partitioning of DWT Matrix $\mathcal{W}$

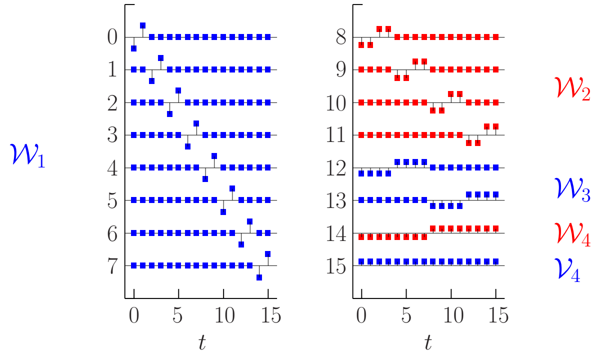
- partition $\mathcal{W}$ commensurate with partitioning of $\mathbf{W}$:

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

- $\mathcal{W}_j$ 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) $\mathcal{W}_j \mathcal{W}_j^T = I_{\frac{N}{2^j}}$

DWT Analysis and Synthesis Equations

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

$$\mathbf{X} = \mathcal{W}^T\mathbf{W} = \left[\mathcal{W}_1^T, \mathcal{W}_2^T, \ldots, \mathcal{W}_J^T, \mathcal{V}_J^T\right] \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_J \\ \mathbf{V}_J \end{bmatrix}$$

$$= \sum_{j=1}^{J} \mathcal{W}_j^T \mathbf{W}_j + \mathcal{V}_J^T \mathbf{V}_J$$

Multiresolution Analysis: I

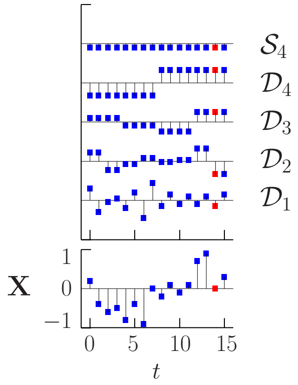
- synthesis equation leads to additive decomposition:

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

- $\mathcal{D}_j \equiv \mathcal{W}_j^T \mathbf{W}_j$ is portion of synthesis due to scale τ_j
- $\mathcal{D}_j$ is vector of length N and is called jth ‘detail’
- $\mathcal{S}_J \equiv \mathcal{V}_J^T \mathbf{V}_J = \overline{X}\mathbf{1}$, where $\mathbf{1}$ is a vector containing N ones (later on we will call this the ‘smooth’ of Jth order)
- additive decomposition called multiresolution analysis (MRA)

Multiresolution Analysis: II

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



- 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 $\mathbf{X}$ as its squared norm:

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

(usually not really energy, but will use term as shorthand)

- energy of $\mathbf{X}$ is preserved in its DWT coefficients $\mathbf{W}$ because

$$\begin{aligned}\|\mathbf{W}\|^2 = \mathbf{W}^T\mathbf{W} &= (\mathcal{W}\mathbf{X})^T\mathcal{W}\mathbf{X} \\ &= \mathbf{X}^T\mathcal{W}^T\mathcal{W}\mathbf{X} \\ &= \mathbf{X}^T I_N \mathbf{X} = \mathbf{X}^T\mathbf{X} = \|\mathbf{X}\|^2\end{aligned}$$

- note: same argument holds for any orthonormal transform

Wavelet Spectrum (Variance Decomposition): I

- let $\overline{X}$ denote sample mean of X_t's: $\overline{X} \equiv \frac{1}{N}\sum_{t=0}^{N-1} X_t$
- let $\hat{\sigma}_X^2$ denote sample variance of X_t's:

$$\begin{aligned}\hat{\sigma}_X^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1}\left(X_t - \overline{X}\right)^2 &= \frac{1}{N}\sum_{t=0}^{N-1} X_t^2 - \overline{X}^2 \\ &= \frac{1}{N}\|\mathbf{X}\|^2 - \overline{X}^2 \\ &= \frac{1}{N}\|\mathbf{W}\|^2 - \overline{X}^2\end{aligned}$$

- since $\|\mathbf{W}\|^2 = \sum_{j=1}^{J} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_J\|^2$ and $\frac{1}{N}\|\mathbf{V}_J\|^2 = \overline{X}^2$,

$$\hat{\sigma}_X^2 = \frac{1}{N}\sum_{j=1}^{J} \|\mathbf{W}_j\|^2$$

Wavelet Spectrum (Variance Decomposition): II

- define discrete wavelet power spectrum:

$$P_X(\tau_j) \equiv \frac{1}{N}\|\mathbf{W}_j\|^2, \text{ where } \tau_j = 2^{j-1}$$

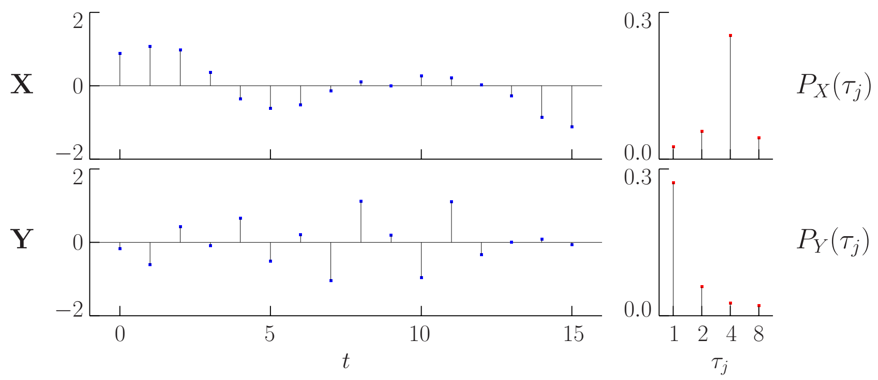
- gives us a scale-based decomposition of the sample variance:

$$\hat{\sigma}_X^2 = \sum_{j=1}^{J} P_X(\tau_j)$$

- in addition, each $W_{j,t}$ in $\mathbf{W}_j$ associated with a portion of $\mathbf{X}$; 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 $\mathbf{X}$ and $\mathbf{Y}$ of length $N = 16$, each with zero sample mean and same sample variance



Summary of Qualitative Description of DWT

- DWT is expressed by an $N \times N$ orthonormal matrix $\mathcal{W}$
- transforms time series $\mathbf{X}$ into DWT coefficients $\mathbf{W} = \mathcal{W}\mathbf{X}$
- each coefficient in $\mathbf{W}$ associated with a scale and location
  - $\mathbf{W}_j$ is subvector of $\mathbf{W}$ with coefficients for scale $\tau_j = 2^{j-1}$
  - coefficients in $\mathbf{W}_j$ related to differences of averages over τ_j
  - last coefficient in $\mathbf{W}$ related to average over scale N
- orthonormality leads to basic scale-based decompositions
  - multiresolution analysis (additive decomposition)
  - discrete wavelet power spectrum (analysis of variance)
- stayed tuned for precise definition of DWT!