DWT: Qualitative Description

Like ODFT, Discrete Wavelet Transform, (DWT) is orthonormal transform

- **analysis of variance: wavelet variance (discrete wavelet power spectrum)**
- **additive decomposition: multiresolution analysis**

DWT: Different from ODFT?

- Real-valued (complex-valued DWTs *do* exist!)
- Basis vectors associated with scale & location (time)
- Requires $N = 2^J$ for some positive integer J (a restrictive assumption)

$\mathbf{DWT: W = \mathcal{W}X.}$

$\mathbf{W}$ is $N \times 1$ vector of DWT coefficients

(jth component denoted as W_j)

$\mathcal{W}$ is $N \times N$ transform matrix: $\mathcal{W}^T\mathcal{W} = I_N$.

$$\mathcal{E}_{\mathbf{X}} = \|\mathbf{X}\|^2 = \mathcal{E}_{\mathbf{W}} = \|\mathbf{W}\|^2 = \sum_{j=0}^{N-1} W_j^2$$

Key: W_j^2 is 'scale/location' contribution to $\mathcal{E}_{\mathbf{X}}$

The Haar DWT: Row 0 to $\frac{N}{2}$

Row $j = 0$: $\left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \underbrace{0, \ldots, 0}_{N-2 \text{ zeros}}\right] \equiv \mathcal{W}_{0\bullet}^T$

Row $j = 1$: $\left[0, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \underbrace{0, \ldots, 0}_{N-4 \text{ zeros}}\right] \equiv \mathcal{W}_{1\bullet}^T$

Transpose of jth row as

$$\mathcal{W}_{j\bullet} = \mathcal{T}^{2j}\mathcal{W}_{0\bullet}, \;\; j = 0, \ldots \frac{N}{2} - 1$$

First $\frac{N}{2}$ rows form orthonormal set of $\frac{N}{2}$ vectors yields $\frac{N}{2}$ wavelet coefficients of 'scale 1,' location $2j$

The Haar DWT: Row $\frac{N}{2}$ to $\frac{3N}{4}$

Row $j = \frac{N}{2}$:

$$\left[-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \underbrace{0, \ldots, 0}_{N-4 \text{ zeros}} \right] \equiv \mathcal{W}^T_{\frac{N}{2}\bullet}$$

Transpose of row $j = \frac{N}{2} + k$ as

$$\mathcal{W}_{\frac{N}{2}+k\bullet} = \mathcal{T}^{4k}\mathcal{W}_{\frac{N}{2}\bullet}, \quad k = 0, \ldots \frac{N}{4} - 1$$

First $\frac{3N}{4}$ rows form orthonormal set of $\frac{3N}{4}$ vectors yield $\frac{N}{4}$ wavelet coefficients of ‘scale 2,’ location $4j$

The Haar DWT: Row $\frac{3N}{4}$ to $\frac{7N}{8}$

Row $j = \frac{3N}{4}$:

$$\Big[\underbrace{-\frac{1}{\sqrt{8}}, \ldots, -\frac{1}{\sqrt{8}}}_{4\text{ of these}}, \underbrace{\frac{1}{\sqrt{8}}, \ldots, \frac{1}{\sqrt{8}}}_{4\text{ of these}}, \underbrace{0, \ldots, 0}_{N-8\text{ zeros}}\Big] \equiv \mathcal{W}^T_{\frac{3N}{4}\bullet}$$

Transpose of row $j = \frac{3N}{4} + k$ as

$$\mathcal{W}_{\frac{3N}{4}+k\bullet} = \mathcal{T}^{8k}\mathcal{W}_{\frac{3N}{4}\bullet}, \quad k = 0, \ldots \frac{N}{8} - 1$$

$\frac{N}{8}$ rows starting with $j = \frac{3N}{4}$ yield $\frac{N}{8}$ wavelet coefficients of ‘scale 4,’ location $8k$

The Haar DWT: Row ... to $N-2$

Row $j = N-2$:

$$\Big[\underbrace{-\frac{1}{\sqrt{N}},\ldots,-\frac{1}{\sqrt{N}}}_{\frac{N}{2}\text{ of these}},\underbrace{\frac{1}{\sqrt{N}},\ldots,\frac{1}{\sqrt{N}}}_{\frac{N}{2}\text{ of these}}\Big] \equiv \mathcal{W}^T_{N-2\bullet}$$

associated with wavelet coefficient of scale $\frac{N}{2}$

Row $j = N-1$: $\Big[\underbrace{\frac{1}{\sqrt{N}},\ldots,\frac{1}{\sqrt{N}}}_{N\text{ of these}}\Big] \equiv \mathcal{W}^T_{N-1\bullet}$

associated with coefficient of scale N

We have created a set of N orthonormal vectors in all

Interpretation of Haar DWT

- Define

$$\overline{\boldsymbol{X}}_t(\boldsymbol{\lambda}) \equiv \frac{1}{\boldsymbol{\lambda}} \sum_{l=0}^{\lambda-1} \boldsymbol{X}_{t-l}$$

‘scale $\boldsymbol{\lambda}$’ average

Note:

$\overline{\boldsymbol{X}}_t(1) = \boldsymbol{X}_t =$ scale 1 ‘average’

$\overline{\boldsymbol{X}}_{N-1}(N) = \overline{\boldsymbol{X}} =$ sample average

Interpretation of Haar DWT : $\mathbf{W} = \mathcal{W}\mathbf{X}$

$$
\begin{aligned}
W_0 &= (X_1 - X_0)/\sqrt{2} = (\overline{X}_1(1) - \overline{X}_0(1))/\sqrt{2} \\
W_1 &= (X_3 - X_2)/\sqrt{2} = (\overline{X}_3(1) - \overline{X}_2(1))/\sqrt{2} \\
&\vdots \\
W_{\frac{N}{2}-1} &= (X_{N-1} - X_{N-2})/\sqrt{2} = (\overline{X}_{N-1}(1) - \overline{X}_{N-2}(1))/\sqrt{2}
\end{aligned}
$$

First $\frac{N}{2}$ rows yield W_j's $\propto$ *changes* on scale 1

Interpretation of Haar DWT : $\mathbf{W} = \mathcal{W}\mathbf{X}$

$$
\begin{aligned}
W_{\frac{N}{2}} &= (X_3 + X_2 - X_1 - X_0)/2 = \overline{X}_3(2) - \overline{X}_1(2) \\
&\vdots \\
W_{\frac{3N}{4}-1} &= (X_{N-1} + X_{N-2} - X_{N-3} - X_{N-4})/2 \\
&= \overline{X}_{N-1}(2) - \overline{X}_{N-3}(2)
\end{aligned}
$$

Next $\frac{N}{4}$ rows yield W_j's $\propto$ *changes* on scale 2

Interpretation of Haar DWT : $\mathbf{W} = \mathcal{W}\mathbf{X}$

$$
\begin{aligned}
W_{\frac{3N}{4}} &= (X_7 + \cdots + X_4 - X_3 - \cdots - X_0)/\sqrt{8} \\
&= \sqrt{2}(\overline{X}_7(4) - \overline{X}_3(4)) \\
&\;\vdots \\
W_{\frac{7N}{8}-1} &= (X_{N-1} + \cdots + X_{N-4} - X_{N-5} - \cdots - X_{N-8})/\sqrt{8} \\
&= \sqrt{2}(\overline{X}_{N-1}(4) - \overline{X}_{N-5}(4))
\end{aligned}
$$

Next $\frac{N}{8}$ rows yield W_j's $\propto$ *changes* on scale 4

Interpretation of Haar DWT : $\mathbf{W} = \mathcal{W}\mathbf{X}$

$$\begin{aligned}
&\vdots \\
W_{N-2} &= (X_{N-1} + \cdots + X_{\frac{N}{2}} - X_{\frac{N}{2}-1} - \cdots - X_0 \\
&= \sqrt{N}(\overline{X}_{N-1}(\tfrac{N}{2}) - \overline{X}_{\frac{N}{2}-1}(\tfrac{N}{2}))/2 \\
W_{N-1} &= (X_{N-1} + \cdots + X_0)/\sqrt{N} = \sqrt{N}\overline{X}
\end{aligned}$$

Next to last row yields $W_j \propto$ *change* on scale $\frac{N}{2}$

Last row yields $W_j \propto$ *average* on scale $N = 2^J$

Structure of DWT Matrix $\mathcal{W}$

- 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 $N = 2^J$
- $\frac{N}{2\tau_j}$ wavelet coeff.'s for scale $\tau_j \equiv 2^{j-1}$, $j = 1, \ldots, J$

(τ_j is standardized scale; $\tau_j \, \Delta t$ is physical scale)

Structure of DWT Matrix $\mathcal{W}$

- Each $\boldsymbol{W}_j$ localized in time: as scale $\uparrow$, localization $\downarrow$
- Rows of $\boldsymbol{\mathcal{W}}$ for given scale $\boldsymbol{\tau_j}$:
  - circularly shifted with respect to each other
  - shift between adjacent rows is $\mathbf{2}\boldsymbol{\tau_j} = \mathbf{2}^{\boldsymbol{j}}$
- Differences of averages common theme for DWTs

DWT-Based Analysis of Variance

$$\mathcal{E}_{\mathbf{W}} = \|\mathbf{W}\|^2 = \|\mathbf{X}\|^2 = \mathcal{E}_{\mathbf{X}}$$

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

DWT-Based Analysis of Variance

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 $\frac{N}{2^j}$ elements (scale $\tau_j = 2^{j-1}$ changes)

DWT-Based Analysis of Variance

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

$\mathbf{V}_J$ has 1 element, namely, $\sqrt{N}\,\overline{X}$ (scale N average)

DWT-Based Analysis of Variance

Define discrete wavelet power spectrum:

$$P_{\mathcal{W}}(\tau_j) \equiv \frac{1}{N}\|\mathbf{W}_j\|^2, \quad \tau_j = 1, 2, 4, \ldots, \frac{N}{2},$$

so $\sum_{j=1}^{J} P_{\mathcal{W}}(\tau_j) = \hat{\sigma}_X^2.$

$P_{\mathcal{W}}(\tau_j)$ *not* invariant as $\mathbf{X}$ circularly shifts

DWT-Based Additive Decomposition

Synthesis: $\mathbf{X} = \mathcal{W}^T \mathbf{W}$

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

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

DWT-Based Additive Decomposition

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

$\mathcal{W}_j$ is $\frac{N}{2^j} \times N$ matrix (scale $\tau_j = 2^{j-1}$ changes)
Two properties: (a) $\mathbf{W}_j = \mathcal{W}_j \mathbf{X}$ &
(b) $\mathcal{W}_j \mathcal{W}_j^T = I_{\frac{N}{2^j}}$

DWT-Based Additive Decomposition

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

$\mathcal{V}_J$ is $1 \times N$ row vector (each element is $\frac{1}{\sqrt{N}}$)

DWT-Multi Resolution Analysis

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

where $\mathcal{D}_j \equiv \mathcal{W}_j^T \mathrm{W}_j$ (synthesis-scale τ_j)

$$\mathcal{S}_J \equiv \mathcal{V}_J^T \mathrm{V}_J = \overline{X}1$$

$$\mathrm{X} = \sum_{j=1}^{J} \mathcal{D}_j + \mathcal{S}_J \text{ (Multiresolution analysis)}$$

Analysis of Variance

$$\|\mathcal{D}_j\|^2 = \|\mathcal{W}_j^T \mathbf{W}_j\|^2 = \mathbf{W}_j^T \underbrace{\mathcal{W}_j \mathcal{W}_j^T}_{I_{\frac{N}{2^j}}} \mathbf{W}_j = \mathbf{W}_j^T \mathbf{W}_j = \|\mathbf{W}_j\|^2$$

Analysis of variance using details:

$$\hat{\sigma}_X^2 = \sum_{j=1}^{J} P_{\mathcal{W}}(\tau_j) = \frac{1}{N} \sum_{j=1}^{J} \|\mathbf{W}_j\|^2 = \frac{1}{N} \sum_{j=1}^{J} \|\mathcal{D}_j\|^2$$

Note: $\frac{1}{N}\|\mathcal{D}_j\|^2$ is sample variance of detail series
(Argue that sample mean of $\mathcal{D}_j$ is 0)

Wavelet Smooths

Define jth level wavelet smooth for $0 \leq j \leq J - 1$:

$$\mathcal{S}_j \equiv \sum_{k=j+1}^{J} \mathcal{D}_k + \mathcal{S}_J$$

'smooth' since small τ_j variations removed from X:

$$\mathcal{S}_j = \mathrm{X} - \sum_{k=1}^{j} \mathcal{D}_k$$

Wavelet Roughs

Define jth level wavelet rough:

$$\mathcal{R}_j \equiv \begin{cases} 0, & j = 0; \\ \sum_{k=1}^{j} \mathcal{D}_k, & 1 \leq j \leq J, \end{cases}$$

Three interpretations of details, roughs and smooths: $\mathcal{S}_j + \mathcal{R}_j = \mathbf{X}$,

$$\mathcal{D}_j = \mathcal{S}_{j-1} - \mathcal{S}_j,$$
$$\mathcal{D}_j = \mathcal{R}_j - \mathcal{R}_{j-1}$$

Defining the DWT

- can formulate DWT via 'pyramid algorithm'
  - *defines* $\mathcal{W}$ for non-Haar wavelets
  - leads to same definition for Haar $\mathcal{W}$
  - computes $\mathbf{W} = \mathcal{W}\mathbf{X}$ using $O(N)$ multiplications
    - 'brute force' method uses $O(N^2)$
    - faster than the fast Fourier transform!