Discrete-Time Wavelet Analysis - Theorems and Definitions

1 Discrete Wavelet Transform (DWT)

Discrete Wavelet Transform (DWT): The discrete wavelet transform can be written in matrix form as $\mathbf{W} = \mathcal{W}\mathbf{X}$, where $\mathbf{X} \in \mathbb{R}^N$ is the time series, $\mathbf{W} \in \mathbb{R}^N$ are the DWT coefficients, and $\mathcal{W} \in \mathbb{R}^{N \times N}$ is the orthonormal wavelet transform matrix. We require that $N = 2^J$ with J an integer. Note that since $\mathcal{W}$ is an orthonormal matrix, we have $\mathcal{W}^T\mathcal{W} = \mathcal{W}\mathcal{W}^T = I_N$ and thus $\mathbf{X} = \mathcal{W}^T\mathbf{W}$.

Notation for the Wavelet Transform Matrix: Let $\mathcal{W}_{j\bullet}$ denote the jth row of $\mathcal{W}$ and $\mathcal{W}_{j,l}$ denote the element in row j, column l. Note that $< \mathcal{W}_{j\bullet}, \mathcal{W}_{k\bullet} >= \delta_{jk}$ and thus $\|\mathbf{X}\|^2 = \|\mathbf{W}\|^2$.

DWT Coefficients: We denote the jth DWT coefficient by $W_j =< \mathcal{W}_{j\bullet}, \mathbf{X} >$. We decompose $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_J, \mathbf{V}_J)^T$, where $\mathbf{W}_j \in \mathbb{R}^{N/2^j}$ and $\mathcal{W} = (\mathcal{W}_1, \ldots, \mathcal{W}_J, \mathcal{V}_J)^T$, where $\mathcal{W}_j \in \mathbb{R}^{N/2^j \times N}$, $\mathcal{V}_J = (1/\sqrt{N}, \ldots, 1/\sqrt{N})^T \in \mathbb{R}^{1\times N}$. Hence $\mathbf{X} = \sum_{j=1}^J \mathcal{W}_j^T\mathbf{W}_j + \mathcal{V}_J\mathbf{V}_J$.

Multiresolution Analysis (MRA): We may express $\mathbf{X}$ in an additive decomposition known as multiresolution analysis by $\mathbf{X} = \sum_{j=1}^J \mathcal{D}_j + \mathcal{S}_J$, where $\mathcal{D}_j = \mathcal{W}_j^T\mathbf{W}_j$ and $\mathcal{S}_J = \mathcal{V}_J^T\mathbf{V}_J = \mathbf{X}\mathbf{1}$. The vectors $\mathcal{D}_j$ are refered to as the jth level detail and relate to changes of $\mathbf{X}$ on a scale of $\tau_j = 2^{j-1}$.

Wavelet Spectrum (Variance Decomposition): Let $\overline{X} = \frac{1}{N}\sum_{t=0}^{N-1} X_t$ denote the sample mean of $\mathbf{X}$ and $\widehat{\sigma}_X^2$ denote the sample variance of $\mathbf{X}$. Then

$$\begin{aligned}
\widehat{\sigma}_X^2 &= \frac{1}{N}\sum_{t=0}^{N-1}(X_t - \overline{X})^2 = \frac{1}{N}\|\mathbf{X}\|^2 - \overline{X} = \frac{1}{N}\|\mathbf{W}\|^2 - \overline{X} \\
&= \frac{1}{N}\left(\sum_{j=1}^J \|\mathbf{W}_j\|^2 + \|\mathbf{V}_J\|^2\right) - \overline{X} = \frac{1}{N}\sum_{j=1}^J \|\mathbf{W}_j\|^2 = \sum_{j=1}^J P_X(\tau_j).
\end{aligned}$$

Variable Level Decomposition: One can decompose $\mathbf{X}$ by $\mathbf{X} = \sum_{j=1}^{J_0} \mathcal{D}_j + \mathcal{S}_{J_0}$ for any $J_0 \in \mathbb{Z}$, where $\mathcal{D}_j$ and $\mathcal{S}_j$ are defined as above.

Convolution and Correlation: The convolution of a_t and b_t is defined as

$$a * b_t = \sum_u a_u b_{t-u}.$$

The correlation of a_t and b_t is defined as

$$a \star b_t = \sum_u a_u b_{t+u}.$$

Note that

$$\mathcal{F}\{a * b_t\}(f) = \mathcal{F}\{a\}(f)\mathcal{F}\{b\}(f), \quad \mathcal{F}\{a \star b_t\}(f) = \mathcal{F}\{a\}^*(f)\mathcal{F}\{b\}(f).$$

Periodized Filters: Let $\{a_t\}_{t=-\infty}^{\infty}$. We defined the periodized sequence of length N by

$$a_u^\circ = \sum_{n=-\infty}^{\infty} a_{u+nN} \leftrightarrow^{FT} A(k/N).$$

Note that if the support of a_u is contained in $[0, L-1] \cap \mathbb{N}$, where $L \leq N$, then

$$a_l^\circ = \begin{cases} a_l, & 0 \leq l \leq L-1, \\ 0, & \text{otherwise.} \end{cases}$$

Wavelet (Mother) Filter: The sequence $\{h_l\}_{l=0}^{L-1}$ is called a wavelet filter if

$$\sum_{l=0}^{L-1} h_l = 0, \quad \sum_{l=0}^{L-1} h_l^2 = 1, \quad \sum_{l=0}^{L-1} h_l h_{l+2n} = 0, \qquad \text{for } n \neq 0.$$

We define

$$H(f) \equiv \mathcal{F}\{h\}(f), \quad \mathcal{H}(f) \equiv |H(f)|^2.$$

From the orthonormality condition, we have that

$$\mathcal{H}(f) + \mathcal{H}(f + 1/2) = 2.$$

Haar Wavelet Filter: The Haar wavelet filter is given by $h_0 = 1/\sqrt{2}$, $h_1 = -1/\sqrt{2}$.

D(4) Wavelet Filter: The D(4) wavelet filter is given by

$$h_0 = \frac{1-\sqrt{3}}{4\sqrt{2}}, \quad h_1 = \frac{-3+\sqrt{3}}{4\sqrt{2}}, \quad h_2 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \quad h_3 = \frac{-1-\sqrt{3}}{4\sqrt{2}}.$$

First Level Wavelet Coefficients: For $N = 2^J$, let $\mathbf{W}_1 = \{W_{1,t}\}_{t=0}^{N/2-1}$, where

$$W_{1,t} = \sum_{l=0}^{N-1} h_l^\circ X_{2t+1-l \bmod N}, \quad t = 0, 1, \ldots, \frac{N}{2} - 1$$

and h_l° is periodized to length N. Note that these coefficients are obtained by filtering the time series with h_l° and then downsampling by a factor of 2.

Upper Half of the DWT Matrix: From the above, we see that $\mathcal{W}_{0\bullet}^T = (h_1^\circ, h_0^\circ, h_{N-1}^\circ, h_{N-2}^\circ, \ldots, h_3^\circ, h_2^\circ)$, $\mathcal{W}_{1\bullet}^T = (h_3^\circ, h_2^\circ, h_1^\circ, h_0^\circ, h_{N-1}^\circ, \ldots, h_5^\circ, h_4^\circ)$. Hence $\mathcal{W}_{j\bullet} = \mathcal{T}^{2j}\mathcal{W}_{0\bullet}$ for $j = 0, 1, \ldots, N/2-1$. One can show that $\mathcal{W}_1\mathcal{W}_1^T = I_{N/2}$, where $\mathbf{W}_1 = \mathcal{W}_1\mathbf{X}$.

Scaling (Father) Filter: The scaling filter is defined as

$$g_l \equiv (-1)^{l+1} h_{L-1-l} \quad \Rightarrow \quad h_l = (-1)^l g_{L-1-l}.$$

Then we have that

$$\sum_{l=0}^{L-1} g_l = \pm\sqrt{2}, \quad \sum_{l=0}^{L-1} g_l^2 = 1, \quad \sum_{l=0}^{L-1} g_l g_{l+2n} = 0, \quad \text{for } n \neq 0.$$

Also define

$$G(f) \equiv \mathcal{F}\{g\}(f), \quad \mathcal{G}(f) \equiv |G(f)|^2.$$

From the orthonormality condition, we have that

$$\mathcal{G}(f) + \mathcal{G}(f + 1/2) = 2.$$

We also have that $\mathcal{G}(f) = \mathcal{H}(f + 1/2)$, so $\mathcal{H}(f) + \mathcal{G}(f) = 2$.

First Level Scaling Coefficients: Let g_l° be the periodized version of g_l to length N. Then the first level scaling coefficients are given by

$$V_{1,t} \equiv \sum_{l=0}^{N-1} g_l^\circ X_{2t+1-l \bmod N}, \quad t = 0, 1, \ldots, \frac{N}{2} - 1.$$

Let $\mathbf{V}_1 = \{V_{1,t}\}_{t=0}^{N/2-1}$. Note that these coefficients are obtained by filtering the time series with g_l° and then downsampling by a factor of 2. One can show that $\mathcal{V}_1 \mathcal{V}_1^T = I_{N/2}$ and $\mathcal{W}_1 \mathcal{V}_1^T = 0_{N/2}$, where $\mathbf{V}_1 = \mathcal{V}_1 \mathbf{X}$.

Synthesizing X from $\mathcal{D}_1$ and $\mathcal{S}_1$: Define the upsampled version of $W_{1,t}$ by a factor of 2

$$W_{1,t}^{\uparrow} \equiv \begin{cases} 0, & t = 0, 2, 4, \ldots, N-2, \\ W_{1,(t-1)/2} = W_{(t-1)/2}, & t = 1, 3, 5, \ldots, N-1 \end{cases}$$

and similarly define $V_{1,t}^{\uparrow}$. Then

$$X_t = \mathcal{D}_{1,t} + \mathcal{S}_{1,t} = \sum_{l=0}^{N-1} h_l^\circ W_{1,t+l \bmod N}^{\uparrow} + \sum_{l=0}^{N-1} g_l^\circ V_{1,t+l \bmod N}^{\uparrow}.$$

Second Stage of the Pyramid Algorithm: We have

$$\begin{array}{llll}
\mathbf{V}_2 \overrightarrow{\uparrow 2} \boxed{G^*\left(\frac{k}{N/2}\right)} \searrow & & & \\
& + \rightarrow \mathbf{V}_1 \overrightarrow{\uparrow 2} \boxed{G^*\left(\frac{k}{N}\right)} \searrow & & \\
\mathbf{W}_2 \overrightarrow{\uparrow 2} \boxed{H^*\left(\frac{k}{N/2}\right)} \nearrow & & + \longrightarrow \mathbf{X} & \\
& \mathbf{W}_1 \overrightarrow{\uparrow 2} \boxed{H^*\left(\frac{k}{N}\right)} \nearrow & &
\end{array}$$

Let $\{h_{2,l}\} \equiv \{g * h_l^{\uparrow}\}$ and $\{g_{2,l}\} \equiv \{g * g_l^{\uparrow}\}$, where $\{h_l^{\uparrow}\} = \{h_0, 0, h_1, \ldots\}$. Then we have that $\{h_{2,l}\} \leftrightarrow^{FT} H_2(f) = H(2f)G(f)$, $\{g_{2,l}\} \leftrightarrow^{FT} H_2(f) = G(2f)G(f)$, and

$$\begin{array}{ll}
 & \nearrow \boxed{G_2(k/N)} \overrightarrow{\downarrow 4} \mathbf{V}_2 \\
\mathbf{X} & \\
 & \searrow \boxed{H_2(k/N)} \overrightarrow{\downarrow 4} \mathbf{W}_2
\end{array}$$

Note that $\sum h_{2,l} = 0$, $\|h_{2,\cdot}\|^2 = 1$, but $\sum h_{2,l}h_{2,l+2n} \neq 0$. The filter $\{h_{2,l}\}$ has length $L_2 \equiv 3L-2$. Because the elements of $\mathbf{W}_2$ are obtained by circularly filtering $\{X_t\}$ with $\{h_{2,l} : 0, 1, \ldots, L_2-1\}$ they can also be obtained by circularly filtering with the periodized version of $\{h_{2,l}\}$, which we denote by $\{h^{\circ}_{2,l} : l = 0, 1, \ldots, N-1\}$, i.e., we have

$$W_{2,t} = \sum_{l=0}^{N-1} h^{\circ}_{2,l} X_{4(t+1)-1-l \bmod N}, \quad t = 0, 1, \ldots, N/4-1.$$

The first row of $\mathcal{W}_2$ (or row $N/2$ of $\mathcal{W}$) is given by

$$\left[h^{\circ}_{2,3}, h^{\circ}_{2,2}, h^{\circ}_{2,1}, h^{\circ}_{2,0}, h^{\circ}_{2,N-1}, h^{\circ}_{2,N-2}, \ldots, h^{\circ}_{2,5}, h^{\circ}_{2,4}\right] = \mathcal{W}^T_{N/2\bullet}$$

the remaining $N/4-1$ rows are given by $\mathcal{T}^{4k}\mathcal{W}^T_{N/2\bullet}$, $k = 1, \ldots, N/4-1$.

General Stage of the Pyramid Algorithm: We have

$$\begin{array}{ccccccccc} & \nearrow & \boxed{G\left(\frac{k}{N/2^{j-1}}\right)} & \overrightarrow{\downarrow 2} & \mathbf{V}_j & \overrightarrow{\uparrow 2} & \boxed{G^*\left(\frac{k}{N/2^{j-1}}\right)} & \searrow & \\ \mathbf{V}_{j-1} & & & & & & & & + \mathbf{V}_{j-1} \\ & \searrow & \boxed{H\left(\frac{k}{N/2^{j-1}}\right)} & \overrightarrow{\downarrow 2} & \mathbf{W}_j & \overrightarrow{\uparrow 2} & \boxed{H^*\left(\frac{k}{N/2^{j-1}}\right)} & \nearrow & \end{array}$$

$$\begin{array}{cccc} & \nearrow & \boxed{G_j(k/N)} & \overrightarrow{\downarrow 2^j} \ \mathbf{V}_j \\ \mathbf{X} & & & \\ & \searrow & \boxed{H_j(k/N)} & \overrightarrow{\downarrow 2^j} \ \mathbf{W}_j \end{array}$$

The equivalent filter $\{h_{j,l}\}$ relating $\{W_{j,t}\}$ with $\{X_t\}$ is formed by convolving together the following j filters:

$$\begin{array}{rl} \text{filter 1:} & g_0, g_1, \ldots, g_{L-1} \\ \text{filter 2:} & g_0, 0, g_1, 0, \ldots, g_{L-2}, 0, g_{L-1} \\ \text{filter 3:} & g_0, 0, 0, 0, g_1, 0, 0, 0, \ldots, g_{L-2}, 0, 0, 0, g_{L-1} \\ & \vdots \\ \text{filter j-1:} & g_0, 0, \ldots, 0, g_1, 0, \ldots, 0, \ldots, g_{L-2}, 0, \ldots, 0, g_{L-1} \\ \text{filter j:} & h_0, 0, \ldots, 0, h_1, 0, \ldots, 0, \ldots, h_{L-2}, 0, \ldots, 0, h_{L-1}. \end{array}$$

Note that the ith filter ($1 \leq i \leq j$) has $2^{i-1}-1$ zeros between samples and that $\{h_{j,l}\}$ has length $L_j \equiv (2^j-1)(L-1)+1$ and transfer function

$$H_j(f) \equiv H(2^{j-1}f) \prod_{l=0}^{j-2} G(2^l f).$$

We similarly define $\{g_{j,l}\}$ relating $\{V_{j,t}\}$ with $\{X_t\}$ and thus it has transfer function

$$G_j(f) \equiv \prod_{l=0}^{j-1} G(2^l f).$$

Then we have that

$$\begin{aligned} W_{j,t} &= \sum_{l=0}^{L_j-1} h_{j,l} X_{2^j(t+1)-1-l \bmod N}, \quad t = 0,1,\ldots,N_j-1 \\ V_{j,t} &= \sum_{l=0}^{L_j-1} g_{j,l} X_{2^j(t+1)-1-l \bmod N}, \quad t = 0,1,\ldots,N_j-1. \end{aligned}$$

Construction of Remaining DWT Coefficients: Treat $\mathbf{V}_{j-1} \in \mathbb{R}^{N/2^{j-1}}$ in the same manner as $\mathbf{X} \in \mathbb{R}^N$ to find $\mathbf{W}_j \in \mathbb{R}^{N/2^j}$ and $\mathbf{V}_j \in \mathbb{R}^{N/2^j}$, i.e., to find $\mathbf{W}_j$ we periodize h_l to length $N/2^{j-1}$ to obtain h_l°, filter h_l° with $\mathbf{V}_{j-1}$, and then downsample by a factor of 2.

Matrix Description of Pyramid Algorithm: Form $\mathcal{B}_j \in \mathbb{R}^{N/2^j \times N/2^{j-1}}$ the same way as $\mathcal{W}_1 \in \mathbb{R}^{N/2 \times N}$, i.e., the rows contain $\{h_l\}$ periodized to length $N/2^{j-1}$. Similarly form $\mathcal{A}_j \in \mathbb{R}^{N/2^j \times N/2^{j-1}}$ the same way as $\mathcal{V}_1 \in \mathbb{R}^{N/2 \times N}$, i.e., the rows contain $\{g_l\}$ periodized to length $N/2^{j-1}$. Then we have that $\mathbf{W}_j = \mathcal{B}_j \mathbf{V}_{j-1}$ and $\mathbf{V}_j = \mathcal{A}_j \mathbf{V}_{j-1}$, so

$$\begin{aligned} \mathbf{W}_j &= \mathcal{B}_j \mathbf{V}_{j-1} = \mathcal{B}_j \mathcal{A}_{j-1} \mathbf{V}_{j-2} = \cdots = \mathcal{B}_j \mathcal{A}_{j-1} \mathcal{A}_{j-2} \cdots \mathcal{A}_1 \mathbf{X} \equiv \mathcal{W}_j \mathbf{X} \\ \mathbf{V}_J &= \mathcal{A}_J \mathbf{V}_{J-1} = \mathcal{A}_J \mathcal{A}_{J-1} \mathbf{V}_{J-2} = \cdots = \mathcal{A}_J \mathcal{A}_{J-1} \cdots \mathcal{A}_1 \mathbf{X} \equiv \mathcal{V}_J \mathbf{X} \end{aligned}$$

which leads to the following decomposition of $\mathcal{W}$

$$\mathcal{W} = \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \mathcal{W}_3 \\ \vdots \\ \mathcal{W}_J \\ \mathcal{V}_J \end{bmatrix} = \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \mathcal{A}_1 \\ \mathcal{B}_3 \mathcal{A}_2 \mathcal{A}_1 \\ \vdots \\ \mathcal{B}_J \mathcal{A}_{J-1} \cdots \mathcal{A}_1 \\ \mathcal{A}_J \mathcal{A}_{J-1} \cdots \mathcal{A}_1 \end{bmatrix}.$$

Pyramid Synthesis Equation: One can show that $\mathbf{V}_{j-1} = \mathcal{B}_j^T \mathbf{W}_j + \mathcal{A}_j^T \mathbf{V}_j$, so

$$\begin{aligned} \mathbf{X} \equiv \mathbf{V}_0 &= \mathcal{B}_1^T \mathbf{W}_1 + \mathcal{A}_1^T \mathbf{V}_1 \\ &= \mathcal{B}_1^T \mathbf{W}_1 + \mathcal{A}_1^T (\mathcal{B}_2^T \mathbf{W}_2 + \mathcal{A}_2^T \mathbf{V}_2) \\ &\vdots \\ &= \mathcal{B}_1^T \mathbf{W}_1 + \mathcal{A}_1^T \mathcal{B}_2^T \mathbf{W}_2 + \mathcal{A}_1^T \mathcal{A}_2^T \mathcal{B}_3^T \mathbf{W}_3 + \mathcal{A}_1^T \mathcal{A}_2^T \mathcal{A}_3^T \mathbf{V}_3 \\ &= \mathcal{D}_1 + \mathcal{D}_2 + \mathcal{D}_3 + \mathcal{S}_3. \end{aligned}$$

Now we describe in terms of filtering operations. Define the upsampled version of $W_{j,t}$ by a factor of 2 by

$$W_{j,t}^{\uparrow} \equiv \begin{cases} 0, & t = 0,2,4,\ldots,N_{j-1}-2, \\ W_{j,(t-1)/2}, & t = 1,3,5,\ldots,N_{j-1}-1 \end{cases}$$

and similarly define $V_{j,t}^{\uparrow}$. Then

$$V_{j-1,t} = \mathcal{D}_{j,t} + \mathcal{S}_{j,t} = \sum_{l=0}^{L-1} h_l W_{j,t+l \bmod N_{j-1}}^{\uparrow} + \sum_{l=0}^{L-1} g_l V_{j,t+l \bmod N_{j-1}}^{\uparrow}.$$

2 Maximal Overlap Discrete Wavelet Transform (MODWT)

MODWT Wavelet and Scaling Filters: We define the MODWT wavelet and scaling coefficients by $\{\widetilde{h}_l\} \equiv \{h_l/\sqrt{2}\}$ and $\{\widetilde{g}_l\} \equiv \{g_l/\sqrt{2}\}$, respectively. Similarly to the DWT case, we have

$$\begin{aligned}
\sum_{l=0}^{L-1} \widetilde{h}_l &= 0, \quad \sum_{l=0}^{L-1} \widetilde{h}_l^2 = \frac{1}{2}, \quad \sum_{l=-\infty}^{\infty} \widetilde{h}_l \widetilde{h}_{l+2n} = 0, \quad \text{for } n \in \mathbb{Z}\backslash\{0\} \\
\sum_{l=0}^{L-1} \widetilde{g}_l &= 1, \quad \sum_{l=0}^{L-1} \widetilde{g}_l^2 = \frac{1}{2}, \quad \sum_{l=-\infty}^{\infty} \widetilde{g}_l \widetilde{g}_{l+2n} = 0, \quad \text{for } n \in \mathbb{Z}\backslash\{0\}.
\end{aligned}$$

Now define $\widetilde{h}_l \leftrightarrow^{FT} \widetilde{H}(f)$ and $|\widetilde{H}(f)|^2 = \widetilde{\mathcal{H}}(f)$. Then $\widetilde{H}(f) = 2^{-1/2}H(f)$ and $\widetilde{\mathcal{H}}(f) + \widetilde{\mathcal{H}}(f + 1/2) = 1$ for all f. We also have

$$\sum_{l=-\infty}^{\infty} \widetilde{g}_l \widetilde{h}_{l+2n} = 0, \quad \text{for } n \in \mathbb{Z}\backslash\{0\}, \quad \widetilde{\mathcal{G}}(f) + \widetilde{\mathcal{H}}(f) = 1 \;\text{ for all } f.$$

Then the MODWT coefficients are given by

$$\begin{aligned}
\widetilde{W}_{1,t} &\equiv \sum_{l=0}^{L-1} \widetilde{h}_l X_{t-l \bmod N} \\
\widetilde{V}_{1,t} &\equiv \sum_{l=0}^{L-1} \widetilde{g}_l X_{t-l \bmod N}.
\end{aligned}$$

Note that we did not down-sample X, as we did in the DWT. Similar relations that apply to $\widetilde{\mathcal{H}}(f)$ also of course apply to $\widetilde{\mathcal{G}}(f)$.

Basic Concepts of the MODWT: We have

$$\begin{aligned}
\mathbf{W}_1 &= [\sqrt{2}\widetilde{W}_{1,1}, \sqrt{2}\widetilde{W}_{1,3}, \ldots, \sqrt{2}\widetilde{W}_{1,N-1}]^T \\
\mathbf{V}_1 &= [\sqrt{2}\widetilde{V}_{1,1}, \sqrt{2}\widetilde{V}_{1,3}, \ldots, \sqrt{2}\widetilde{V}_{1,N-1}]^T.
\end{aligned}$$

Define

$$\mathbf{W}_{\mathcal{T},1} \equiv \mathcal{B}_{\mathcal{T},1}\mathbf{X} \equiv \mathcal{B}_1\mathcal{T}\mathbf{X}.$$

Then

$$\mathbf{W}_{\mathcal{T},1} = [\sqrt{2}\widetilde{W}_{1,0}, \sqrt{2}\widetilde{W}_{1,2}, \ldots, \sqrt{2}\widetilde{W}_{1,N-2}]^T.$$

From $\widetilde{\mathcal{B}}_1$ by interleaving the rows of $\mathcal{B}_{\mathcal{T},1}$ and $\mathcal{B}_1$ and replacing h_l by $\widetilde{h}_l$. We then have $\widetilde{\mathbf{W}}_1 = \widetilde{\mathcal{B}}_1\mathbf{X}$. Explicitly stated, we have

$$\begin{aligned}
\widetilde{\mathcal{B}}_1^T\widetilde{\mathbf{W}}_1 &= \frac{1}{2}\left(\mathcal{B}_1^T\mathbf{W}_1 + \mathcal{B}_{\mathcal{T},1}^T\mathbf{W}_{\mathcal{T},1}\right) \\
\widetilde{\mathcal{A}}_1^T\widetilde{\mathbf{V}}_1 &= \frac{1}{2}\left(\mathcal{A}_1^T\mathbf{V}_1 + \mathcal{A}_{\mathcal{T},1}^T\mathbf{V}_{\mathcal{T},1}\right).
\end{aligned}$$

First Level Detail and Scaling: We have

$$\mathbf{X} = \widetilde{\mathcal{B}}_1^T\widetilde{\mathbf{W}}_1 + \widetilde{\mathcal{A}}_1^T\widetilde{\mathbf{V}}_1 \equiv \widetilde{\mathcal{D}}_1 + \widetilde{\mathcal{S}}_1,$$

where

$$\begin{aligned}\widetilde{\mathcal{D}}_1 &\equiv \sum_{l=0}^{L-1} \widetilde{h}_l \widetilde{W}_{1,t+l \bmod N} = \sum_{l=0}^{N-1} \widetilde{h}_l^{\circ} \widetilde{W}_{1,t+l \bmod N}, \quad t = 0, 1, \ldots, N-1 \\ \widetilde{\mathcal{S}}_1 &\equiv \sum_{l=0}^{L-1} \widetilde{g}_l \widetilde{V}_{1,t+l \bmod N} = \sum_{l=0}^{N-1} \widetilde{g}_l^{\circ} \widetilde{V}_{1,t+l \bmod N}, \quad t = 0, 1, \ldots, N-1.\end{aligned}$$

Analysis of Variation: We have that

$$\|\mathbf{X}\|^2 = \|\widetilde{\mathbf{W}}_1\|^2 + \|\widetilde{\mathbf{V}}_1\|^2.$$

$$\begin{array}{ccccccccc} & \nearrow & \boxed{\widetilde{G}\left(\frac{k}{N}\right)} & \longrightarrow & \widetilde{\mathbf{V}}_1 & \longrightarrow & \boxed{\widetilde{G}^*\left(\frac{k}{N}\right)} & \searrow & \\ \mathbf{X} & & & & & & & & + \mathbf{X} \\ & \searrow & \boxed{\widetilde{H}\left(\frac{k}{N}\right)} & \longrightarrow & \widetilde{\mathbf{W}}_1 & \longrightarrow & \boxed{\widetilde{H}^*\left(\frac{k}{N}\right)} & \nearrow & \end{array}$$

Definition of the jth Level MODWT Coefficients: We have

$$\begin{aligned}\widetilde{W}_{j,t} &\equiv \sum_{l=0}^{L_j-1} \widetilde{h}_{j,l} X_{t-l \bmod N} \\ \widetilde{V}_{j,t} &\equiv \sum_{l=0}^{L_j-1} \widetilde{g}_{j,l} X_{t-l \bmod N}, \ t = 0, 1, \ldots, N-1,\end{aligned}$$

where $\widetilde{h}_{j,l} \equiv h_{j,l}/2^{j/2}$ and $\widetilde{g}_{j,l} \equiv g_{j,l}/2^{j/2}$. Then clearly

$$\begin{aligned}\widetilde{H}_j(f) &= \widetilde{H}(2^{j-1}f) \prod_{l=0}^{j-2} \widetilde{G}(2^l f) \\ \widetilde{G}_j(f) &= \prod_{l=0}^{j-1} \widetilde{G}(2^l f)\end{aligned}$$

and

$$\begin{aligned}\|\mathbf{X}\|^2 &= \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2 \\ \mathbf{X} &= \sum_{j=1}^{J_0} \widetilde{\mathcal{D}}_j + \widetilde{\mathcal{S}}_{J_0}.\end{aligned}$$

General Stage of the Pyramid Algorithm: We have

$$\begin{array}{ccccccccc} & \nearrow & \boxed{\widetilde{G}\left(2^{j-1}\frac{k}{N}\right)} & \longrightarrow & \widetilde{\mathbf{V}}_j & \longrightarrow & \boxed{\widetilde{G}^*\left(2^{j-1}\frac{k}{N}\right)} & \searrow & \\ \widetilde{\mathbf{V}}_{j-1} & & & & & & & & + \widetilde{\mathbf{V}}_{j-1}. \\ & \searrow & \boxed{\widetilde{H}\left(2^{j-1}\frac{k}{N}\right)} & \longrightarrow & \widetilde{\mathbf{W}}_j & \longrightarrow & \boxed{\widetilde{H}^*\left(2^{j-1}\frac{k}{N}\right)} & \nearrow & \end{array}$$

3 Discrete Wavelet Packet Transform (DWPT)

Wavelet Packet Table: Below we have the wavelet packet table for $J_0 = 3$, which shows the levels $j = 0, 1, 2, 3$. Note that for level j the wavelet coefficients $\{W_{j,n}\}_{n=0}^{2^j-1}$ are nominally associated with frequencies in the interval $\mathcal{I}_{j,n} \equiv \left[\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}\right]$ and time width $\lambda_j = 2^j$. Note that $\mathbf{V}_j = \mathbf{W}_{j,0}$.

<table>
<tr><td colspan="8">$\mathbf{W}_{0,0} = \mathbf{X}$</td></tr>
<tr><td colspan="4">$\downarrow$
$G(k/N)$
$\downarrow 2$</td><td colspan="4">$\downarrow$
$H(k/N)$
$\downarrow 2$</td></tr>
<tr><td colspan="4">$\mathbf{W}_{1,0}$</td><td colspan="4">$\mathbf{W}_{1,1}$</td></tr>
<tr><td colspan="2">$\downarrow$
$G(k/N_1)$
$\downarrow 2$</td><td colspan="2">$\downarrow$
$H(k/N_1)$
$\downarrow 2$</td><td colspan="2">$\downarrow$
$H(k/N_1)$
$\downarrow 2$</td><td colspan="2">$\downarrow$
$G(k/N_1)$
$\downarrow 2$</td></tr>
<tr><td colspan="2">$\mathbf{W}_{2,0}$</td><td colspan="2">$\mathbf{W}_{2,1}$</td><td colspan="2">$\mathbf{W}_{2,2}$</td><td colspan="2">$\mathbf{W}_{2,3}$</td></tr>
<tr><td>$\downarrow$
$G(k/N_2)$
$\downarrow 2$</td><td>$\downarrow$
$H(k/N_2)$
$\downarrow 2$</td><td>$\downarrow$
$H(k/N_2)$
$\downarrow 2$</td><td>$\downarrow$
$G(k/N_2)$
$\downarrow 2$</td><td>$\downarrow$
$G(k/N_2)$
$\downarrow 2$</td><td>$\downarrow$
$H(k/N_2)$
$\downarrow 2$</td><td>$\downarrow$
$H(k/N_2)$
$\downarrow 2$</td><td>$\downarrow$
$G(k/N_2)$
$\downarrow 2$</td></tr>
<tr><td>$\mathbf{W}_{3,0}$</td><td>$\mathbf{W}_{3,1}$</td><td>$\mathbf{W}_{3,2}$</td><td>$\mathbf{W}_{3,3}$</td><td>$\mathbf{W}_{3,4}$</td><td>$\mathbf{W}_{3,5}$</td><td>$\mathbf{W}_{3,6}$</td><td>$\mathbf{W}_{3,7}$</td></tr>
</table>

Disjoint Dyadic Decomposition: From the above table, we see that one can decompose $\mathbf{X}$ in many ways. For instance,

$$\begin{bmatrix} \mathbf{W}_{1,1} \\ \mathbf{W}_{3,3} \\ \mathbf{W}_{3,2} \\ \mathbf{W}_{2,0} \end{bmatrix} = \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{A}_3\mathcal{B}_2\mathcal{A}_1 \\ \mathcal{B}_3\mathcal{B}_2\mathcal{A}_1 \\ \mathcal{A}_2\mathcal{A}_1 \end{bmatrix} \mathbf{X}$$

and we have a decomposition of energy (ANOVA)

$$\|\mathbf{X}\|^2 = \|\mathbf{W}_{1,1}\|^2 + \|\mathbf{W}_{3,3}\|^2 + \|\mathbf{W}_{3,2}\|^2 + \|\mathbf{W}_{2,0}\|^2.$$

The collection of doublets (j, n) forming the indices of the table nodes will be denoted by $\mathcal{N} \equiv \{(j, n) : j = 0, 1, \ldots, J_0;\ n = 0, 1, \ldots, 2^j - 1\}$, where we are free to pick $J_0 \leq J$. The doublets (j, n) that form the indices of the WP coefficients corresponding to an orthonormal transform will be denoted by $\mathcal{C}$; for example, above we have $\mathcal{C} = \{(1, 1), (3, 3), (3, 2), (2, 0)\}$.

Best Basis Algorithm: (1) Given a WP table to level J_0, then for every $(j, n) \in \mathcal{N}$ we associate with $\mathbf{W}_{j,n}$ a cost $M(\mathbf{W}_{j,n})$, where $M(\cdot)$ is an additive cost functional of the form

$$M(\mathbf{W}_{j,n}) \equiv \sum_{t=0}^{N_j-1} m(|\mathbf{W}_{j,n,t}|), \quad (N_j = N/2^j),$$

and $m(\cdot)$ is a real-valued function defined on $[0, \infty)$ with $m(0) = 0$. (2) The 'optimal' orthonormal transform that can be extracted from the WP table is the solution of

$$\min_{\mathcal{C}} \sum_{(j,n)\in\mathcal{C}} M(\mathbf{W}_{j,n}),$$

i.e., we seek the orthonormal transform specified by $\mathcal{C} \subseteq \mathcal{N}$ such that the cost summed over all the doublets $(j, n) \in \mathcal{C}$ is minimized. Explicitly stated, the algorithm is as follows:

1. We mark all the costs in the nodes at the bottom of the (wavelet packet) table in some way. We start by examining this bottom row of nodes.
2. We compare the costs of the sum of each pair of children nodes with their parent node and then do one of the following:
   (a) if the parent node has a lower cost than the sum of the costs of the children nodes, we mark the parent node, while
   (b) if the sum of the costs of the children nodes is lower than the cost of the parent node, we replace the cost of the parent node by the sum of the costs of the children nodes (no new marking is done).
3. We then repeat step 2 for each level as we move up the table.
4. Once we have reached the top of the table, we look back down the table at the marked nodes. The top-most marked nodes that correspond to a disjoint dyadic decomposition define the best basis transform.

A Few Cost Functionals for the Best Basis Algorithm: Here are three examples.

1. The $-l^2 \log(l^2)$ norm, also called the 'entropy' information cost functional, where

   $$m(|\overline{W}_{j,n,t}|) = \begin{cases} -\overline{W}^2_{j,n,t} \log(\overline{W}^2_{j,n,t}), & W_{j,n,t} \neq 0, \\ 0, & W_{j,n,t} = 0, \end{cases}$$

   and $\overline{W}_{j,n,t} \equiv W_{j,n,t}/\|\mathbf{X}\|$. This quantity has a monotonic relationship with the entropy of a sequence.
2. The threshold functional, i.e., the number of terms exceeding a specified threshold δ, with

   $$M(|W_{j,n,t}|) = \begin{cases} 1, & |W_{j,n,t}| > \delta, \\ 0, \text{otherwise}. \end{cases}$$

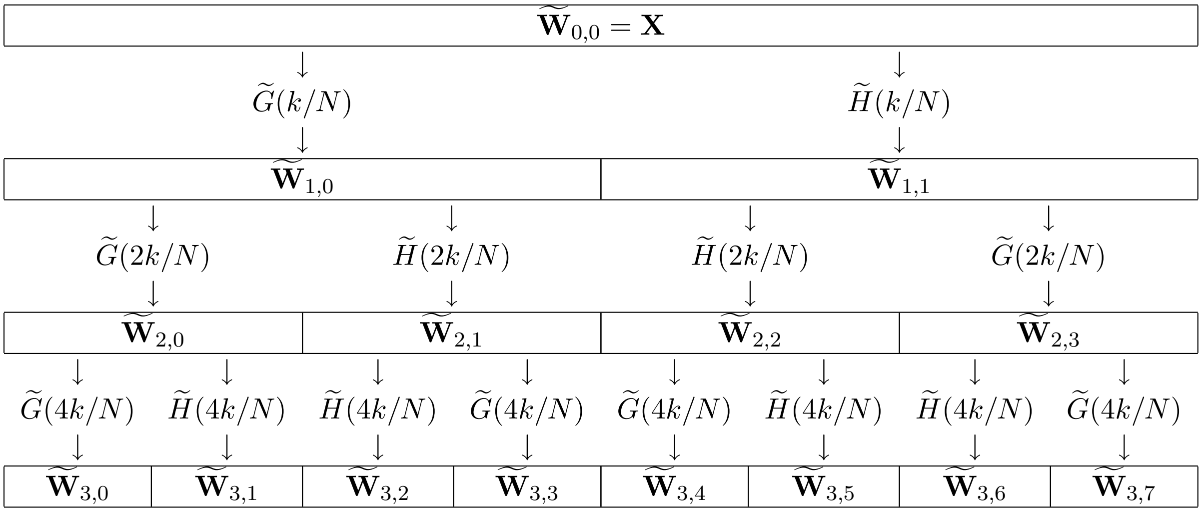
3. The l_p information cost functional, defined by

   $$m(|W_{j,n,t}|) = |W_{j,n,t}|^p,$$

   for $0 < p < 2$. If this cost functional is used, then $M^{1/p}(\mathbf{W}_{j,n})$ gives the l_p norm of the sequence.

These cost functionals give a lower cost if energy is concentrated in a few $|W_{j,n,t}|$'s.

Maximal Overlap Wavelet Packet Table: Just as we had a packet table for the DWT, we have one for the MODWT.



Dictionary: A dictionary is a set of vectors $\mathbb{D} \equiv \{\mathbf{d}_\gamma : \gamma \in \Gamma\}$, where $\mathbf{d}_\gamma \in \mathbb{R}^N$, $\|\mathbf{d}_\gamma\|^2 = 1$ and Γ is a finite set that contains a subset that is a basis for $\mathbb{R}^N$.

Matching Pursuit: In matching pursuit, we try to approximate a signal $\mathbf{X}$ using a small number of 'time/frequency' vectors from a large set of such vectors. We find these vectors from successive orthogonal projections of $\mathbf{X}$. Let $R^{(0)} = X$ and

$$\mathbf{X} = \mathbf{X}^{(m)} + \mathbf{R}^{(m)}, \quad \text{where} \qquad \mathbf{X}^{(m)} = \sum_{n=0}^{m-1} < \mathbf{R}^{(n)}, \mathbf{d}_{\gamma_n} > \mathbf{d}_{\gamma_n}$$

$$\mathbf{d}_{\gamma_n} = \text{argmin}_{\gamma\in\Gamma} | < \mathbf{R}^{(n)}, \mathbf{d}_\gamma > |.$$

Then we have that

$$\|\mathbf{X}\|^2 = \sum_{n=0}^{m-1} | < \mathbf{R}^{(n)}, \mathbf{d}_{\gamma_n} > |^2 + \|\mathbf{R}^{(m)}\|^2.$$

Note that $< \mathbf{R}^{(m)}, \mathbf{R}^{(n)} >= 0$ for $m \neq n$.

4 The Wavelet Variance

Wavelet Variance: Let $\{X_t\}_{t=-\infty}^{\infty}$ be a stationary stochastic process and let

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

represent the stochastic process obtained by filtering $\{X_t\}$ with the MODWT wavelet filter $\{\widetilde{h}_{j,l}\}$. Then the time-independent wavelet variance is given by

$$\nu_X^2(\tau_j) \equiv \text{Var}(\overline{W}_{j,t}).$$

Wavelet Variance Estimators: We may estimate the wavelet variance at scale τ_j with the MODWT coefficients by

$$\widetilde{\nu}_X^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{N}\left(\sum_{t=0}^{L_j-2} \widetilde{W}_{t,j}^2 + \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2\right)$$

and also by

$$\widehat{\nu}_X^2(\tau_j) \equiv \frac{1}{M_j}\sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{M_j}\sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2, \quad M_j \equiv N - L_j + 1.$$

Since $E[\widehat{\nu}_X^2(\tau_j)] = \nu_X^2(\tau_j)$, we refer to $\widehat{\nu}_X^2(\tau_j)$ as the unbiased MODWT estimator of the wavelet variance. We refer to $\widetilde{\nu}_X^2(\tau_j)$ as the biased MODWT estimator of the wavelet variance.

Decomposition of Variance: Suppose that $\{Y_t\}$ is a stationary process with spectral density function (SDF) $S_Y(f)$ defined for $f \in [-1/2, 1/2]$. Then we have that

$$\int_{-1/2}^{1/2} S_Y(f)\, df = \sigma_Y^2 \equiv \mathrm{Var}(Y_t).$$

The analog of this fundamental result for the wavelet variance is

$$\sum_{j=1}^{\infty} \nu_Y^2(\tau_j) = \sigma_Y^2,$$

where the decomposition is now across various scales. Note that $\nu_Y^2(\tau_j) \approx 2\int_{1/2^{j+1}}^{1/2^j} S_Y(f)\, df$.

Accuracy of the Unbiased MODWT estimator of the Wavelet Variance: Because $\widehat{\nu}_X^2(\tau_j)$ is a random variable, it is of interest to know how close it is likely to be to $\nu_X^2(\tau_j)$. To address this question, let us assume that $\{\overline{W}_{j,t}\}$ is a Gaussian stationary process with mean zero and SDF $S_j(\cdot)$. Then if $S_j(f) > 0$ a.e. and if

$$A_j \equiv \int_{-1/2}^{1/2} S_j(f)\, df < \infty$$

then the estimator $\widehat{\nu}_X^2(\tau_j)$ is asymptotically normally distributed with mean $\nu_X^2(\tau_j)$ and large sample variance $2A_j/M_j$. Thus, to a good approximation for large $M_j = N_j - L_j + 1$, we have

$$\frac{M_j^{1/2}(\widehat{\nu}_X^2(\tau_j) - \nu_X^2(\tau_j))}{(2A_j)^{1/2}} =^d \mathcal{N}(0,1).$$

Periodogram: Let Y_t be a stationary process. The discrete Fourier transform produces a basic estimator of $S_Y(\cdot)$ called the periodogram, which is given by

$$\widehat{S}_Y^{(p)}(f_k) \equiv \frac{1}{N}\left|\sum_{t=0}^{N-1}(Y_t - \overline{Y})e^{-2\pi i f_k t}\right|^2, \quad f_k = k/N.$$

The periodogram is limited in its usefulness because it is an inconsistent estimator of $S_Y(\cdot)$ and because it can be badly biased even for sample sizes that would be considered "large" by statisticians.

Variance Decomposition for a White Noise Process: Let X_t be a white noise process. Then $\nu_X^2(\tau_j) \propto \tau_j^{-1}$ and thus $\log(\nu_X^2(\tau_j)) \propto -\log(\tau_j)$. We see that the log-log plot of $\nu_X^2(\tau_j)$ is a line of slope -1.

Backward Difference Process: The backwards difference of orders one and two are given by

$$X_t^{(1)} = X_t - X_{t-1}, \quad X_t^{(2)} = X_t^{(1)} - X_{t-1}^{(1)} = X_t - 2X_{t-1} + X_{t-2},$$

respectively. Then X_t is said to have dth order stationary backward differences (SBD) if

$$Y_t \equiv \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k}$$

forms a stationary process.

Fractional Differenced Process: A popular time series model $\{X_t\}$ for a stationary long memory process is the fractionally differenced (FD) process. Here $\{X_t\}$ is related to a (typically Gaussian) white noise process $\{\varepsilon_t\}$ with mean zero and variance σ_ε^2 through $(1-B)^\delta X_t = \varepsilon_t$, where B is the backward shift operator (e.g. $(1-B)X_t = X_t - X_{t-1}$); $-\frac{1}{2} < \delta < \frac{1}{2}$; and $(1-B)^\delta$ is interpreted as

$$(1-B)^\delta = \sum_{k=0}^{\infty} \binom{\delta}{k} (-1)^k B^k, \text{ so } \sum_{k=0}^{\infty} \binom{\delta}{k} (-1)^k X_{t-k} = \varepsilon_t,$$

where

$$\binom{\delta}{k} \equiv \frac{\delta!}{k!(\delta-k)!} = \frac{\Gamma(\delta+1)}{\Gamma(k+1)\Gamma(\delta-k+1)}.$$

The SDF for an FD process is given by

$$S_X(f) = \frac{\sigma_\varepsilon^2}{(4\sin^2(\pi f))^\delta}, \quad -\frac{1}{2} \le f \le \frac{1}{2}.$$

For small f we have $S_X(f) \propto |f|^{2\delta}$ approximately, so an FD process is a stationary long memory process when $0 < \delta < \frac{1}{2}$. We denote an FD process with parameter δ by $FD(\delta)$.

Daubechies Wavelets and SBD: Let X_t be a nonstationary process with d-th order stationary differences. Suppose we use a Daubechies wavelet filter of width $L \ge 2d$. Then $\sum_{j=1}^{\infty} \nu_X^2(\tau_j) = \infty$. Let $X_t = \sum_{u=1}^{t} \varepsilon_u$ be a random walk process. Thus with $L = 2$ $(2 \ge 2 \cdot 1)$

$$\nu_X^2(\tau_j) = \frac{\mathrm{Var}(\varepsilon_t)}{6}\left(\tau_j + \frac{1}{2\tau_j}\right) \approx \frac{\mathrm{Var}(\varepsilon_t)}{6}\tau_j$$

and $\log(\nu_X^2(\tau_j)) \propto \log(\tau_j)$, so the log-log plot has slope $+1$. Note that a fractionally differenced process with parameter $\delta = 1$ is the same as a random walk process.