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 study algorithm using linear filters & matrix manipulations
- will look at both approaches since they are complementary

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
  - L called filter width
  - both h_0 and h_{L-1} must be nonzero
  - L must be even $(2, 4, 6, 8, \ldots)$ for technical reasons
  - will assume $h_l \equiv 0$ for $l < 0$ and $l \geq L$

The Wavelet Filter: II

- $\{h_l\}$ 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
  $$\sum_{l=0}^{L-1} h_l h_{l+2n} = 0$$

- 2 and 3 together are called the orthonormality property

The Wavelet Filter: III

- define transfer and squared gain functions for wavelet filter:

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

- claim: orthonormality property equivalent to

$$\mathcal{H}(f) + \mathcal{H}(f + \tfrac{1}{2}) = 2 \quad \text{for all } f$$

- to show equivalence, first assume above holds
- consider autocorrelation of $\{h_l\}$:

$$h \star h_j \equiv \sum_{l=-\infty}^{\infty} h_l h_{l+j} \quad j = \ldots, -1, 0, 1, \ldots$$

- $\{h_l\} \longleftrightarrow H(\cdot)$ implies that $\{h \star h_j\} \longleftrightarrow |H(\cdot)|^2 = \mathcal{H}(\cdot)$

The Wavelet Filter: IV

- inverse DFT says $h \star h_j = \int_{-1/2}^{1/2} \mathcal{H}(f')e^{i2\pi f'j}\, df'$
- Exer. [23b] says that, if $\{a_j\} \longleftrightarrow A(\cdot)$, then

$$\{a_{2n}\} \longleftrightarrow \tfrac{1}{2}\left[A(\tfrac{f}{2}) + A(\tfrac{f}{2} + \tfrac{1}{2})\right]$$

- application of this result here says that

$$\{h \star h_{2n}\} \longleftrightarrow \tfrac{1}{2}\left[\mathcal{H}(\tfrac{f}{2}) + \mathcal{H}(\tfrac{f}{2} + \tfrac{1}{2})\right]$$

- $\mathcal{H}(f) + \mathcal{H}(f + \frac{1}{2}) = 2$ for all f says that $\mathcal{H}(\frac{f}{2}) + \mathcal{H}(\frac{f}{2} + \frac{1}{2}) = 2$
- leads to orthonormality condition because

$$\sum_{l=-\infty}^{\infty} h_l h_{l+2n} = h \star h_{2n} = \int_{-1/2}^{1/2} e^{i2\pi fn}\, df = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

The Wavelet Filter: VI

- hence $\mathcal{H}(f) + \mathcal{H}(f + \frac{1}{2}) = 2$ implies orthonormality
- Exer. [70]: orthonormality implies

$$\mathcal{H}(f) + \mathcal{H}(f + \tfrac{1}{2}) = 2 \text{ for all } f$$

- this establishes the equivalence between above and

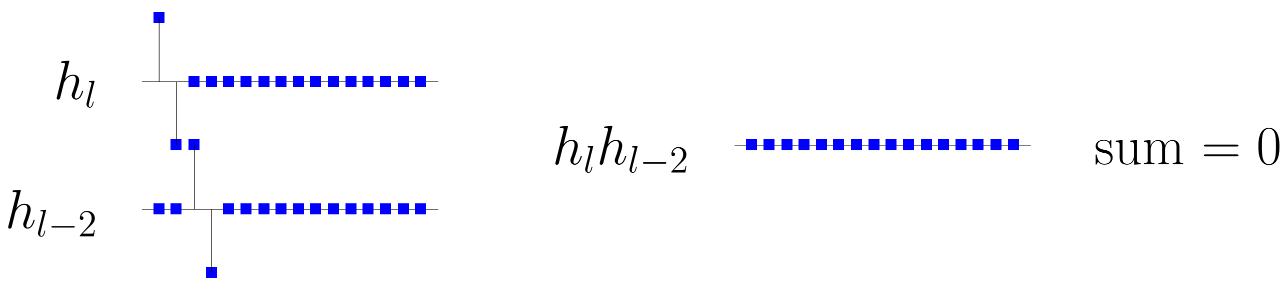
$$\sum_{l=-\infty}^{\infty} h_l h_{l+2n} = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

The Wavelet Filter: VII

- summation to zero and unit energy relatively easy to achieve (analogous to conditions imposed on wavelet functions $\psi(\cdot)$)
- orthogonality to even shifts is key property
- orthogonality hardest to satisfy, and is reason L must be even
  - consider filter $\{h_0, h_1, h_2\}$ of width $L = 3$
  - width 3 requires $h_0 \neq 0$ and $h_2 \neq 0$
  - orthogonality to a shift of 2 requires $h_0 h_2 = 0$ – impossible!

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

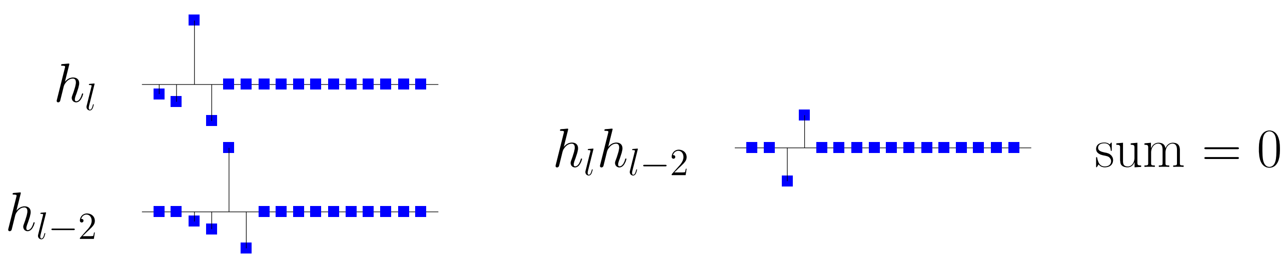


D(4) Wavelet Filter: I

- next simplest wavelet filter is D(4), for which $L = 4$:

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

  - ‘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 ± 2 case:



D(4) Wavelet Filter: II

- Q: what is rationale for D(4) filter?
- consider $X_t^{(1)} \equiv X_t - X_{t-1} = a_0 X_t + a_1 X_{t-1}$, where $\{a_0 = 1, a_1 = -1\}$ defines 1st difference filter:

$$\{X_t\} \longrightarrow \boxed{\{1, -1\}} \longrightarrow \{X_t^{(1)}\}$$

  - 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:

$$\{X_t\} \longrightarrow \boxed{\{1, -1\}} \longrightarrow \boxed{\{1, -1\}} \longrightarrow \{X_t^{(2)}\}$$

- equivalent filter defines 2nd difference filter:

$$\{X_t\} \longrightarrow \boxed{\{1, -2, 1\}} \longrightarrow \{X_t^{(2)}\}$$

D(4) Wavelet Filter: III

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

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

(mentioned as being highly discretized Mexican hat wavelet)

- consider '2 point weighted average' followed by 2nd difference:

$$\{X_t\} \longrightarrow \boxed{\{a, b\}} \longrightarrow \boxed{\{1, -2, 1\}} \longrightarrow \{Y_t\}$$

- D(4) wavelet filter based on equivalent filter for above:

$$\{X_t\} \longrightarrow \boxed{\{h_0, h_1, h_2, h_3\}} \longrightarrow \{Y_t\}$$

D(4) Wavelet Filter: IV

- using conditions

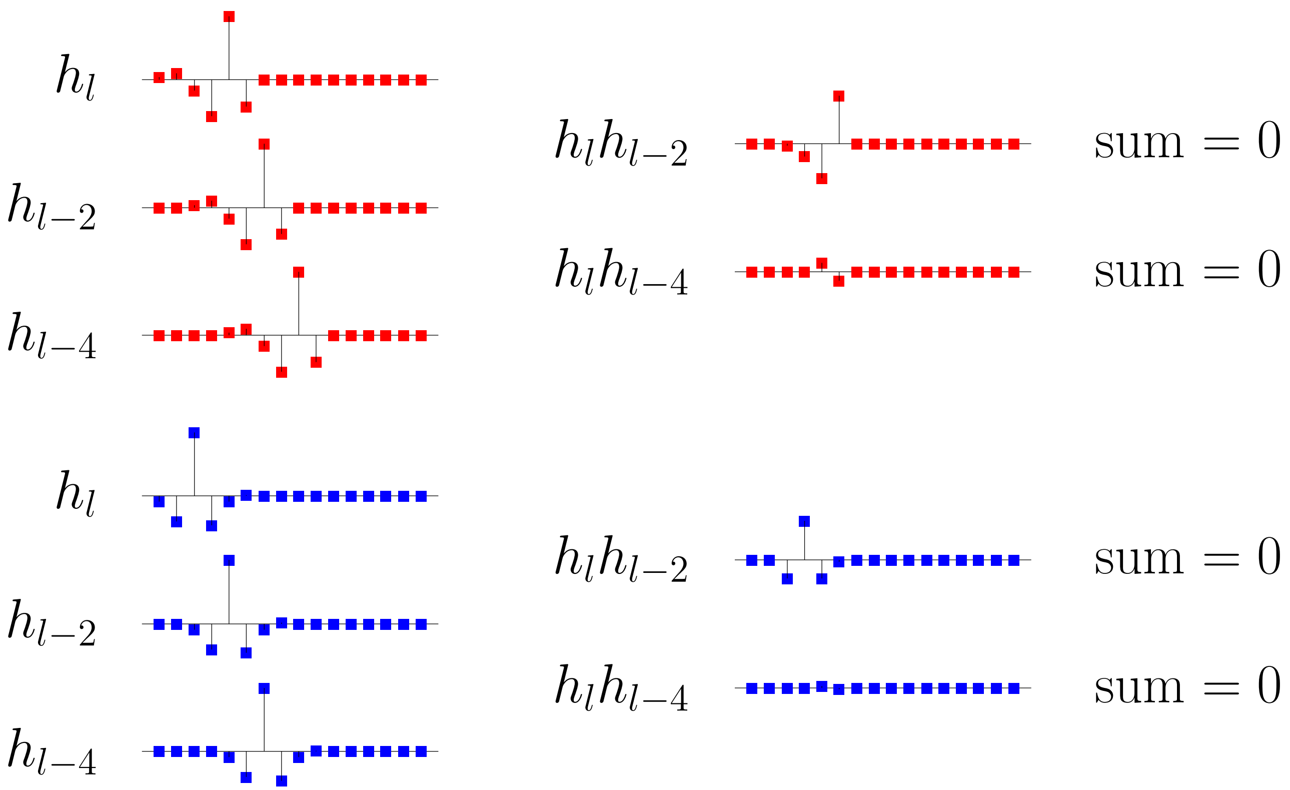
  1. summation to zero: $h_0 + h_1 + h_2 + h_3 = 0$
  2. unit energy: $h_0^2 + h_1^2 + h_2^2 + h_3^2 = 1$
  3. orthogonality to even shifts: $h_0 h_2 + h_1 h_3 = 0$

  can solve for feasible values of a and b (Exer. [4.1])

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

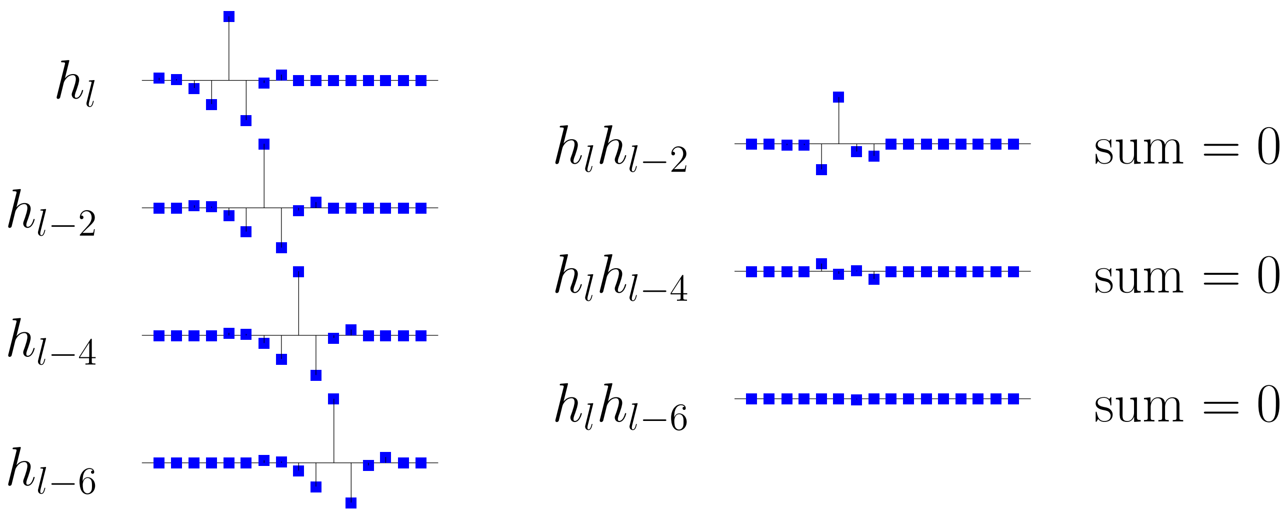
A Selection of Other Wavelet Filters: I

- lots of other wavelet filters exist – here are three we'll see later
- D(6) wavelet filter (top) and C(6) wavelet filter (bottom)



A Selection of Other Wavelet Filters: II

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



- all 3 wavelet filters resemble Mexican hat (somewhat)
- can interpret each filter as cascade consisting of
  - weighted average of effective width of 1
  - higher order differences
- filter outputs 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 $N = 2^J$, goal is to define matrix $\mathcal{W}$ for computing $\mathbf{W} = \mathcal{W}\mathbf{X}$
- periodize $\{h_l\}$ to length N to form $h_0^\circ, h_1^\circ, \ldots, h_{N-1}^\circ$
- circularly filter $\mathbf{X}$ using $\{h_l^\circ\}$ to yield output

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

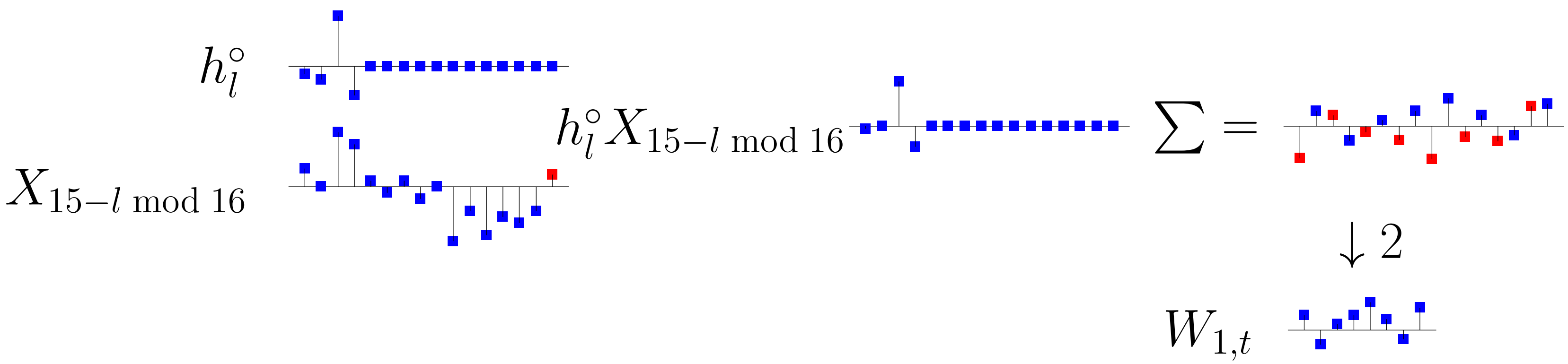
- starting with $t = 1$, take every other value of output to define

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

$\{W_{1,t}\}$ formed by *downsampling* filter output by a factor of 2

First Level Wavelet Coefficients: II

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



- note: '$\downarrow 2$' denotes 'downsample by two' (take every 2nd value)

First Level Wavelet Coefficients: III

- $\{W_{1,t}\}$ are unit scale wavelet coefficients
  - j in $W_{j,t}$ indicates a particular group of wavelet coefficients
  - $j = 1, 2, \ldots, J$ (upper limit tied to sample size $N = 2^J$)
  - will refer to index j as the level
  - thus $W_{1,t}$ is associated with level $j = 1$
  - $W_{1,t}$ also associated with scale 1
  - level j is associated with scale 2^{j-1} (more on this later)
- $\{W_{1,t}\}$ forms first $N/2$ elements of $\mathbf{W} = \mathcal{W}\mathbf{X}$
- first $N/2$ elements of $\mathbf{W}$ form subvector $\mathbf{W}_1$
- $W_{1,t}$ is tth element of $\mathbf{W}_1$
- also have $\mathbf{W}_1 = \mathcal{W}_1\mathbf{X}$, with $\mathcal{W}_1$ being first $N/2$ rows of $\mathcal{W}$

Upper Half of DWT Matrix: I

- setting $t = 0$ in definition for $W_{1,t}$ yields

$$\begin{aligned} W_{1,0} &= \sum_{l=0}^{N-1} h_l^\circ X_{1-l \bmod N} \\ &= h_0^\circ X_1 + h_1^\circ X_0 + h_2^\circ X_{N-1} + \cdots + h_{N-2}^\circ X_3 + h_{N-1}^\circ X_2 \\ &= h_1^\circ X_0 + h_0^\circ X_1 + h_{N-1}^\circ X_2 + h_{N-2}^\circ X_3 + \cdots + h_2^\circ X_{N-1} \end{aligned}$$

- recall $W_{1,0} = \langle \mathcal{W}_{0\bullet}, \mathbf{X} \rangle$, where $\mathcal{W}_{0\bullet}^T$ is first row of $\mathcal{W}$ & of $\mathcal{W}_1$
- comparison with above says that

$$\mathcal{W}_{0\bullet}^T = [h_1^\circ, h_0^\circ, h_{N-1}^\circ, h_{N-2}^\circ, \ldots, h_5^\circ, h_4^\circ, h_3^\circ, h_2^\circ]$$

Upper Half of DWT Matrix: II

- similar examination of $W_{1,1}, \ldots W_{1,\frac{N}{2}}$ shows following pattern
  - circularly shift $\mathcal{W}_{0\bullet}$ by 2 to get 2nd row of $\mathcal{W}$:

$$\mathcal{W}_{1\bullet}^T = [h_3^\circ, h_2^\circ, h_1^\circ, h_0^\circ, h_{N-1}^\circ, h_{N-2}^\circ, \ldots, h_5^\circ, h_4^\circ]$$

  - form $\mathcal{W}_{j\bullet}$ by circularly shifting $\mathcal{W}_{j-1\bullet}$ by 2, ending with

$$\mathcal{W}_{\frac{N}{2}-1\bullet}^T = [h_{N-1}^\circ, h_{N-2}^\circ, \ldots, h_5^\circ, h_4^\circ, h_3^\circ, h_2^\circ, h_1^\circ, h_0^\circ]$$

- if $L \leq N$ (usually the case), then

$$h_l^\circ \equiv \begin{cases} h_l, & 0 \leq l \leq L-1 \\ 0, & \text{otherwise} \end{cases}$$

Example: Upper Half of Haar DWT Matrix

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

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

- rows obviously orthogonal to each other

Example: Upper Half of D(4) DWT Matrix

- when $L = 4$ & $N = 16$, $\mathcal{W}_1$ (i.e., upper half of $\mathcal{W}$) looks like

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

- rows orthogonal because $h_0h_2 + h_1h_3 = 0$
- note: $\langle \mathcal{W}_{0\bullet}, \mathbf{X} \rangle$ yields $W_{1,0} = h_1X_0 + h_0X_1 + h_3X_{14} + h_2X_{15}$
- unlike other coefficients from above, this 'boundary' coefficient depends on circular treatment of $\mathbf{X}$ (a curse, not a feature!)

Orthonormality of Upper Half of DWT Matrix: I

- if $L \leq N$, orthonormality of rows of $\mathcal{W}_1$ follows readily from orthonormality of $\{h_l\}$
- as example of $L > N$ case (comes into play at higher levels), consider $N = 4$ and $L = 6$:

$$h_0^\circ = h_0 + h_4;\ h_1^\circ = h_1 + h_5;\ h_2^\circ = h_2;\ h_3^\circ = h_3$$

- $\mathcal{W}_1$ is:

$$\begin{bmatrix} h_1^\circ & h_0^\circ & h_3^\circ & h_2^\circ \\ h_3^\circ & h_2^\circ & h_1^\circ & h_0^\circ \end{bmatrix} = \begin{bmatrix} h_1 + h_5 & h_0 + h_4 & h_3 & h_2 \\ h_3 & h_2 & h_1 + h_5 & h_0 + h_4 \end{bmatrix}$$

- inner product of two rows is

$$h_1h_3 + h_3h_5 + h_0h_2 + h_2h_4 + h_1h_3 + h_3h_5 + h_0h_2 + h_2h_4$$
$$= 2(h_0h_2 + h_1h_3 + h_2h_4 + h_3h_5) = 0$$

because $\{h_l\}$ is orthogonal to $\{h_{l+2}\}$ (an even shift)

Orthonormality of Upper Half of DWT Matrix: II

- will now show that, for all L and even N,

$$W_{1,t} = \sum_{l=0}^{N-1} h_l^\circ X_{2t+1-l \bmod N}, \text{ or, equivalently, } \mathbf{W}_1 = \mathcal{W}_1 \mathbf{X}$$

  forms *half* an orthonormal transform; i.e.,

$$\mathcal{W}_1 \mathcal{W}_1^T = I_{\frac{N}{2}}$$

- need to show that rows of $\mathcal{W}_1$ have unit energy and are pairwise orthogonal

Orthonormality of Upper Half of DWT Matrix: III

- recall what first row of $\mathcal{W}_1$ looks like:

$$\mathcal{W}_{0\bullet}^T = [h_1^\circ, h_0^\circ, h_{N-1}^\circ, h_{N-2}^\circ, \ldots, h_2^\circ]$$

- last $\frac{N}{2} - 1$ rows formed by circularly shift above by 2, 4, ...
- orthonormality follows if we can show

$$\sum_{n=0}^{N-1} h_n^\circ h_{n+l \bmod N}^\circ \equiv h^\circ \star h_l^\circ = \begin{cases} 1, & \text{if } l = 0; \\ 0, & \text{if } l = 2, 4, \ldots, N-2. \end{cases}$$

- Exer. [33] says $\{h_l^\circ\} \longleftrightarrow \{H(\frac{k}{N})\}$
- implies $\{h^\circ \star h_l^\circ\} \longleftrightarrow \{|H(\frac{k}{N})|^2 = \mathcal{H}(\frac{k}{N})\}$

Orthonormality of Upper Half of DWT Matrix: IV

- inverse DFT relationship says that

$$
\begin{aligned}
h^\circ \star h^\circ_{2l} &= \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{H}(\tfrac{k}{N}) e^{i2\pi(2l)k/N} \\
&= \frac{1}{N} \left(\sum_{k=0}^{\frac{N}{2}-1} \mathcal{H}(\tfrac{k}{N}) e^{i4\pi lk/N} + \sum_{k=0}^{\frac{N}{2}-1} \mathcal{H}(\tfrac{k}{N} + \tfrac{1}{2}) e^{i4\pi l(\frac{k}{N}+\frac{1}{2})} \right) \\
&= \frac{1}{N} \sum_{k=0}^{\frac{N}{2}-1} \left[\mathcal{H}(\tfrac{k}{N}) + \mathcal{H}(\tfrac{k}{N} + \tfrac{1}{2}) \right] e^{i4\pi lk/N}
\end{aligned}
$$

- orthonormality property for $\{h_l\}$ says $\mathcal{H}(\frac{k}{N}) + \mathcal{H}(\frac{k}{N} + \frac{1}{2}) = 2$

Orthonormality of Upper Half of DWT Matrix: V

- thus have

$$h^\circ \star h^\circ_{2l} = \frac{2}{N} \sum_{k=0}^{\frac{N}{2}-1} e^{i4\pi lk/N} = \begin{cases} 1, & \text{if } l = 0; \\ 0, & \text{if } l = 1, 2, \ldots, \frac{N}{2} - 1, \end{cases}$$

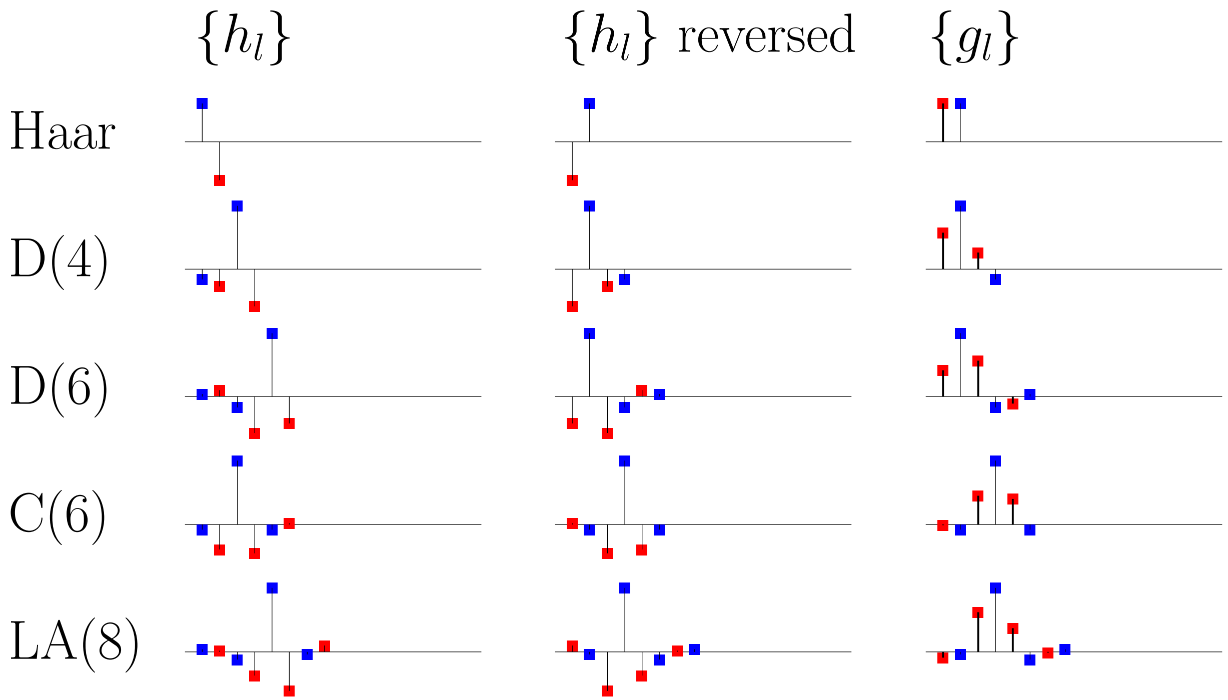
where the last part follows from an application of

$$\sum_{k=0}^{\frac{N}{2}-1} z^k = \frac{1 - z^{N/2}}{1 - z} \text{ with } z = e^{i4\pi l/N}, \text{ so } z^{N/2} = e^{i2\pi l} = 1$$

- $\mathcal{W}_1$ is thus half of the desired orthonormal DWT matrix
- Q: how can we construct the other half of $\mathcal{W}$?

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



- 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

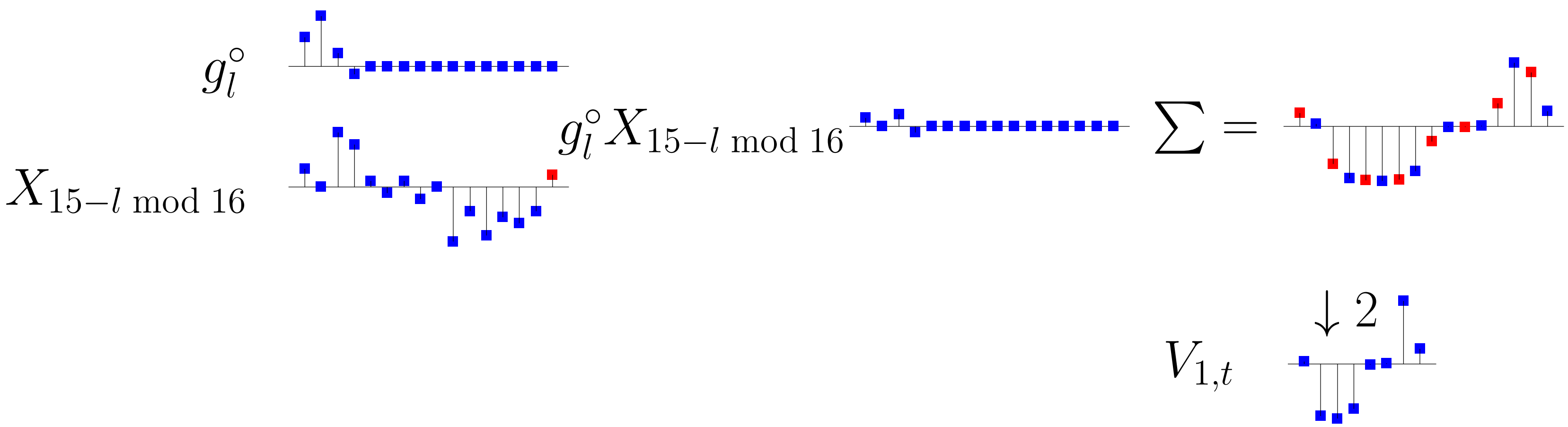
First Level Scaling Coefficients: I

- orthonormality property of $\{h_l\}$ was all we needed to prove that $\mathcal{W}_1$ 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
- periodize $\{g_l\}$ to length N to form $g_0^\circ, g_1^\circ, \ldots, g_{N-1}^\circ$
- circularly filter $\mathbf{X}$ using $\{g_l^\circ\}$ and downsample to define

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

First Level Scaling Coefficients: II

- example of formation of $\{V_{1,t}\}$



- $\{V_{1,t}\}$ are scaling coefficients for level $j = 1$
- place these $N/2$ coefficients in vector called $\mathbf{V}_1$

First Level Scaling Coefficients: III

- define $\mathcal{V}_1$ in a manner analogous to $\mathcal{W}_1$ so that $\mathbf{V}_1 = \mathcal{V}_1 \mathbf{X}$
- when $L = 4$ and $N = 16$, $\mathcal{V}_1$ looks like

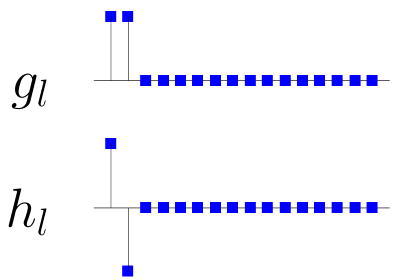
$$\begin{bmatrix}
g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 \\
g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 &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

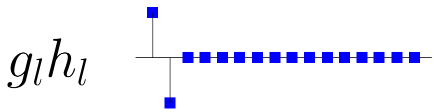
- $\mathcal{V}_1$ obeys same orthonormality property as $\mathcal{W}_1$:

$$\text{similar to } \mathcal{W}_1 \mathcal{W}_1^T = I_{\frac{N}{2}}, \text{ have } \mathcal{V}_1 \mathcal{V}_1^T = I_{\frac{N}{2}}$$

Orthonormality of $\mathcal{V}_1$ and $\mathcal{W}_1$: I

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

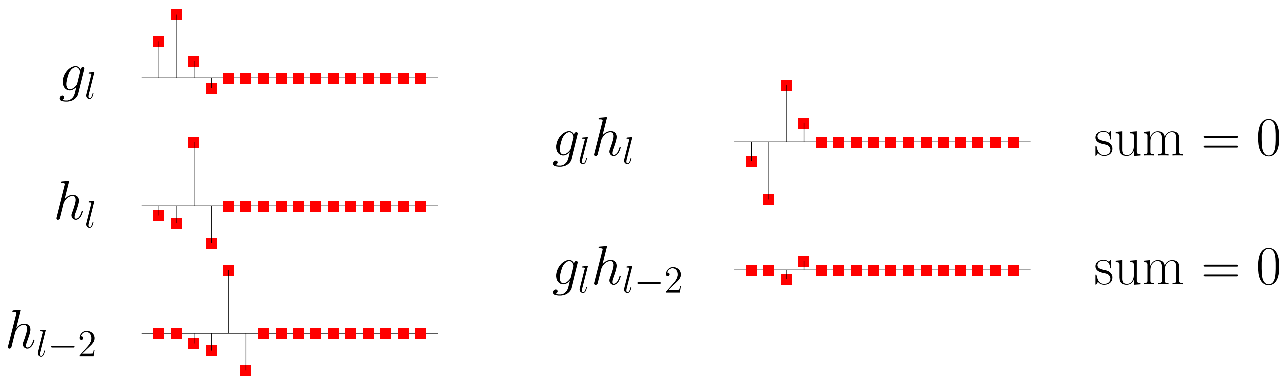




sum $= 0$

Orthonormality of $\mathcal{V}_1$ and $\mathcal{W}_1$: II

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



- before proving claim, need to introduce matrices for circularly shifting vectors

Matrices for Circularly Shifting Vectors

- define $\mathcal{T}$ and $\mathcal{T}^{-1}$ to be $N \times N$ matrices that circularly shift $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ either right or left one unit:

$$\mathcal{T}\mathbf{X} = [X_{N-1}, X_0, X_1, \ldots, X_{N-3}, X_{N-2}]^T$$
$$\mathcal{T}^{-1}\mathbf{X} = [X_1, X_2, X_3, \ldots, X_{N-2}, X_{N-1}, X_0]^T$$

- for $N = 4$, here are what these matrices look like:

$$\mathcal{T} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \ \& \ \mathcal{T}^{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

- note that $\mathcal{T}\mathcal{T}^{-1} = I_N$
- define $\mathcal{T}^2 = \mathcal{T}\mathcal{T}$, $\mathcal{T}^{-2} = \mathcal{T}^{-1}\mathcal{T}^{-1}$ etc.
- for all integers j & k, have $\mathcal{T}^j\mathcal{T}^k = \mathcal{T}^{j+k}$, with $\mathcal{T}^0 \equiv I_N$

Orthonormality of $\mathcal{V}_1$ and $\mathcal{W}_1$: III

- $[\mathcal{T}^{2t}\mathcal{V}_{0\bullet}]^T$ and $[\mathcal{T}^{2t}\mathcal{W}_{0\bullet}]^T$ are tth rows of $\mathcal{V}_1$ & $\mathcal{W}_1$
- for $0 \le t \le \frac{N}{2} - 1$ and $0 \le t' \le \frac{N}{2} - 1$, need to show that

$$\langle \mathcal{T}^{2t}\mathcal{V}_{0\bullet}, \mathcal{T}^{2t'}\mathcal{W}_{0\bullet} \rangle = 0$$

- letting $n = t' - t$, have, for $n = 0, \ldots, \frac{N}{2} - 1$,

$$\begin{aligned}\langle \mathcal{T}^{2t}\mathcal{V}_{0\bullet}, \mathcal{T}^{2t'}\mathcal{W}_{0\bullet} \rangle &= \mathcal{V}_{0\bullet}^T \mathcal{T}^{-2t}\mathcal{T}^{2t'}\mathcal{W}_{0\bullet} \\ &= \mathcal{V}_{0\bullet}^T \mathcal{T}^{2n}\mathcal{W}_{0\bullet} = \sum_{l=0}^{N-1} g_l^\circ h_{l+2n \bmod N}^\circ\end{aligned}$$

- example for $n = 1$, $L = 4$ and $N = 16$:

$$\begin{aligned}\mathcal{V}_{0\bullet}^T &= \begin{bmatrix} g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 \end{bmatrix} \\ \mathcal{T}^2\mathcal{W}_{0\bullet} &= \begin{bmatrix} h_3 & h_2 & h_1 & h_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\end{aligned}$$

Frequency Domain Properties of Scaling Filter

- needs some facts about frequency domain properties of $\{g_l\}$
- define transfer and squared gain functions for $\{g_l\}$

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

- Exer. [76a]: $G(f) = e^{-i2\pi f(L-1)} H(\frac{1}{2} - f)$, so

$$\mathcal{G}(f) = |e^{-i2\pi f(L-1)}|^2 |H(\tfrac{1}{2} - f)|^2 = \mathcal{H}(\tfrac{1}{2} - f)$$

- evenness of $\mathcal{H}(\cdot)$ yields $\mathcal{G}(f) = \mathcal{H}(f - \frac{1}{2})$
- unit periodicity of $\mathcal{H}(\cdot)$ yields $\mathcal{G}(f) = \mathcal{H}(f + \frac{1}{2})$
- $\mathcal{H}(f) + \mathcal{H}(f + \frac{1}{2}) = 2$ implies

$$\mathcal{H}(f) + \mathcal{G}(f) = 2 \;\text{ and also }\; \mathcal{G}(f) + \mathcal{G}(f + \tfrac{1}{2}) = 2$$

Orthonormality of $\mathcal{V}_1$ and $\mathcal{W}_1$: IV

- to establish orthogonality of $\mathcal{V}_1$ and $\mathcal{W}_1$, need to show

$$\sum_{l=0}^{N-1} g_l^\circ h_{l+2n \bmod N}^\circ = g^\circ \star h_{2n}^\circ = 0 \text{ for } n = 0, \ldots, \frac{N}{2} - 1,$$

  where $\{g^\circ \star h_l^\circ\}$ is cross-correlation of $\{g_l^\circ\}$ & $\{h_l^\circ\}$

- since $\{g_l^\circ\} \longleftrightarrow \{G(\frac{k}{N})\}$ and $\{h_l^\circ\} \longleftrightarrow \{H(\frac{k}{N})\}$, have

$$\{g^\circ \star h_l^\circ\} \longleftrightarrow \{G^*(\tfrac{k}{N})H(\tfrac{k}{N})\}$$

Orthonormality of $\mathcal{V}_1$ and $\mathcal{W}_1$: V

- Exer. [78]: use inverse DFT of $\{G^*(\frac{k}{N})H(\frac{k}{N})\}$ to argue that

$$g^\circ \star h^\circ_{2n} = \frac{1}{N} \sum_{k=0}^{\frac{N}{2}-1} \left[G^*(\tfrac{k}{N})H(\tfrac{k}{N}) + G^*(\tfrac{k}{N}+\tfrac{1}{2})H(\tfrac{k}{N}+\tfrac{1}{2}) \right] e^{i4\pi nk/N}$$

  and then argue that

$$G^*(\tfrac{k}{N})H(\tfrac{k}{N}) + G^*(\tfrac{k}{N}+\tfrac{1}{2})H(\tfrac{k}{N}+\tfrac{1}{2}) = 0,$$

  which establishes orthonormality

- thus $\mathcal{W}_1$ & $\mathcal{V}_1$ are jointly orthonormal:

  $\mathcal{W}_1\mathcal{V}_1^T = \mathcal{V}_1\mathcal{W}_1^T = 0_{\frac{N}{2}}$ in addition to $\mathcal{V}_1\mathcal{V}_1^T = \mathcal{W}_1\mathcal{W}_1^T = I_{\frac{N}{2}}$,

  where $0_{\frac{N}{2}}$ is an $\frac{N}{2} \times \frac{N}{2}$ matrix, all of whose elements are zeros

Orthonormality of $\mathcal{V}_1$ and $\mathcal{W}_1$: VI

- implies that

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

is an $N \times N$ orthonormal matrix since

$$\begin{aligned} \mathcal{P}_1 \mathcal{P}_1^T &= \begin{bmatrix} \mathcal{W}_1 \\ \mathcal{V}_1 \end{bmatrix} \begin{bmatrix} \mathcal{W}_1^T, \mathcal{V}_1^T \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{W}_1 \mathcal{W}_1^T & \mathcal{W}_1 \mathcal{V}_1^T \\ \mathcal{V}_1 \mathcal{W}_1^T & \mathcal{V}_1 \mathcal{V}_1^T \end{bmatrix} = \begin{bmatrix} I_{\frac{N}{2}} & 0_{\frac{N}{2}} \\ 0_{\frac{N}{2}} & I_{\frac{N}{2}} \end{bmatrix} = I_N \end{aligned}$$

- 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

Three Comments

- if N even (i.e., don't need $N = 2^J$), then $\mathcal{P}_1$ is well-defined and can be of interest by itself
- rather than defining $g_l = (-1)^{l+1} h_{L-1-l}$, could use alternative definition $g_l = (-1)^{l+1} h_{1-l}$ (definitions are same for Haar)
  - $g_{-(L-2)}, \ldots, g_1$ would be nonzero rather than $g_0, \ldots, g_{L-1}$
  - structure of $\mathcal{V}_1$ would then not parallel that of $\mathcal{W}_1$
  - useful for wavelet filters with infinite widths
- scaling and wavelet filters are often called 'father' and 'mother' wavelet filters, but Strichartz (1994) notes that this terminology

  '... shows a scandalous misunderstanding of human reproduction; in fact, the generation of wavelets more closely resembles the reproductive life style of amoebas.'

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

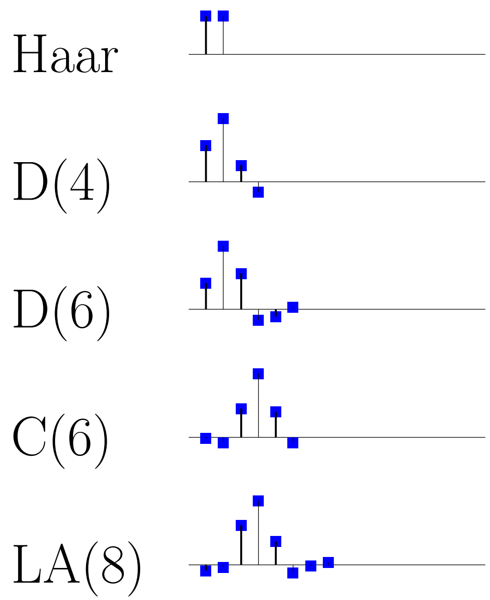
$$\begin{bmatrix}
g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 &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

- since $\mathbf{V}_1 = \mathcal{V}_1 \mathbf{X}$, each $V_{1,t}$ is proportional to a 2 point average:

$$V_{1,0} = g_1 X_0 + g_0 X_1 = \frac{1}{\sqrt{2}} X_0 + \frac{1}{\sqrt{2}} X_1 \propto \overline{X}_1(2) \text{ and so forth}$$

Interpretation of Scaling Coefficients: II

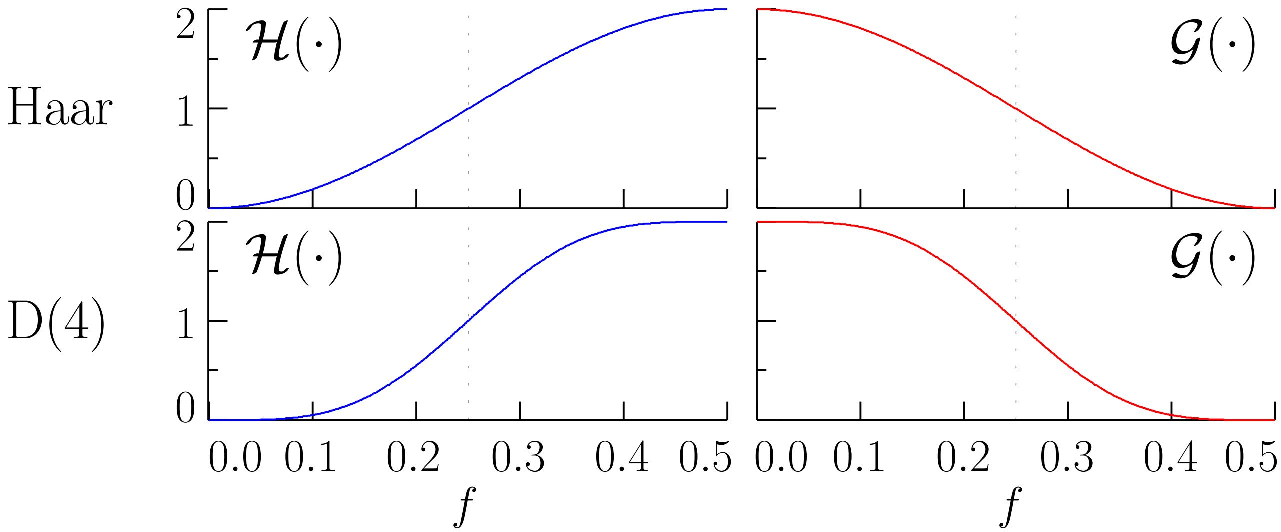
- reconsider shapes of $\{g_l\}$ seen so far:



- 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 $\{h_l\}$ and $\{g_l\}$

- since $\mathbf{W}_1$ and $\mathbf{V}_1$ contain (downsampled) output from filters, let's look at their squared gain functions
- example: $\mathcal{H}(\cdot)$ and $\mathcal{G}(\cdot)$ 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 $[0, 1/4]$

What Kind of Process is $\{V_{1,t}\}$?: I

- letting $\{X_t\} \longleftrightarrow \{\mathcal{X}_k\}$ & $f_k = k/N$, use inverse DFT to get

$$X_t = \frac{1}{N}\sum_{k=0}^{N-1} \mathcal{X}_k e^{i2\pi f_k t} = \frac{1}{N}\sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} \mathcal{X}_k e^{i2\pi f_k t},$$

  where the change in the limits of summation is OK because $\{\mathcal{X}_k\}$ and $\{e^{i2\pi f_k t}\}$ are both periodic with a period of N

- since $\{g_l\} \longleftrightarrow G(f) = |G(f)|e^{i\theta^{(G)}(f)}$, where $|G(f)| \approx \sqrt{2}$ for $|f| \in [-\frac{1}{4}, \frac{1}{4}]$ and $|G(f)| \approx 0$ for $|f| \in (\frac{1}{4}, \frac{1}{2}]$, can argue

$$\sum_{l=0}^{L-1} g_l X_{t-l \bmod N} \approx \frac{\sqrt{2}}{N}\sum_{k=-\frac{N}{4}+1}^{\frac{N}{4}} \mathcal{X}_k e^{i\theta^{(G)}(f_k)} e^{i2\pi f_k t}$$

What Kind of Process is $\{V_{1,t}\}$?: II

- with downsampling,

$$V_{1,t} \approx \frac{\sqrt{2}}{N} \sum_{k=-\frac{N}{4}+1}^{\frac{N}{4}} \mathcal{X}_k e^{i\theta^{(G)}(f_k)} e^{i2\pi f_k(2t+1)}, \ \ 0 \le t \le \frac{N}{2} - 1$$

$$= \frac{2}{N} \sum_{k=-\frac{N}{4}+1}^{\frac{N}{4}} \frac{1}{\sqrt{2}} \mathcal{X}_k e^{i\theta^{(G)}(f_k)} e^{i2\pi f_k} \times e^{i2\pi (2f_k)t}$$

$$\equiv \frac{1}{N'} \sum_{k=-\frac{N'}{2}+1}^{\frac{N'}{2}} \mathcal{X}'_k e^{i2\pi f'_k t}, \ \ 0 \le t \le N' - 1$$

if we let $N' \equiv \frac{N}{2}$, $\mathcal{X}'_k \equiv \frac{1}{\sqrt{2}} \mathcal{X}_k e^{i\theta^{(G)}(f_k)} e^{i2\pi f_k}$ and $f'_k \equiv 2f_k$

What Kind of Process is $\{V_{1,t}\}$?: III

- let's study the above result:

$$V_{1,t} \approx \frac{1}{N'} \sum_{k=-\frac{N'}{2}+1}^{\frac{N'}{2}} \mathcal{X}'_k e^{i2\pi f'_k t}, \;\; 0 \le t \le N'-1$$

- $\mathcal{X}'_k$ is associated with $f'_k = 2f_k = \frac{2k}{N} = \frac{k}{N/2} = \frac{k}{N'}$
- since $-\frac{N'}{2} + 1 \le k \le \frac{N'}{2}$, have $-\frac{1}{2} < f'_k \le \frac{1}{2}$
- whereas result of filtering $\{X_t\}$ with $\{g_l\}$ is a 'half-band' (low-pass) process involving approximately just $f_k \in [-\frac{1}{4}, \frac{1}{4}]$ down-sampled process $\{V_{1,t}\}$ is 'full-band' involving $f'_k \in [-\frac{1}{2}, \frac{1}{2}]$

What Kind of Process is $\{W_{1,t}\}$?: I

- in a similar manner, because $h_l \approx$ high pass, can argue that

$$\sum_{l=0}^{L-1} h_l X_{t-l \bmod N} \approx \frac{\sqrt{2}}{N} \left(\sum_{k=-\frac{N}{2}+1}^{-\frac{N}{4}} + \sum_{k=\frac{N}{4}+1}^{\frac{N}{2}} \right) \mathcal{X}_k e^{i\theta^{(H)}(f_k)} e^{i2\pi f_k t}$$

- with downsampling,

$$W_{1,t} \approx \frac{1}{N'} \sum_{k=-\frac{N'}{2}+1}^{\frac{N'}{2}} \mathcal{X}'_k e^{i2\pi f'_k t}, \ \ 0 \leq t \leq N' - 1,$$

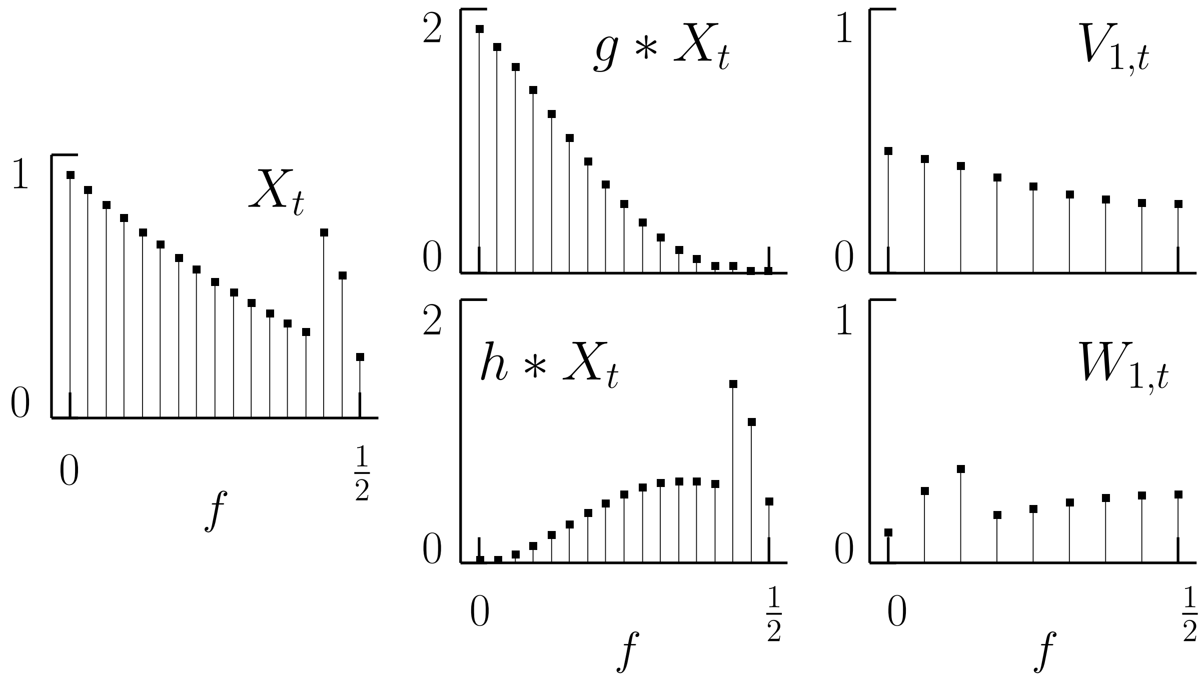
where now $\mathcal{X}'_k = -\frac{1}{\sqrt{2}} \mathcal{X}_{k+\frac{N}{2}} e^{i\theta^{(H)}(f_k + \frac{1}{2})} e^{i2\pi f_k}$

What Kind of Process is $\{W_{1,t}\}$?: II

- note that $|\mathcal{X}'_k| \propto |\mathcal{X}_{k+\frac{N}{2}}| = |\mathcal{X}_{k-\frac{N}{2}}|$ because $\{\mathcal{X}_k\}$ is periodic
- since X_t is real-valued, $|\mathcal{X}_{-k}| = |\mathcal{X}_k|$ and hence $|\mathcal{X}'_k| \propto |\mathcal{X}_{\frac{N}{2}-k}|$
- as before, $\mathcal{X}'_k$ is associated with $f'_k = 2f_k$
- $\mathcal{X}_{\frac{N}{2}-k}$ is associated with $f_{\frac{N}{2}-k} = \frac{1}{2} - f_k$
- conclusion: the coefficient for $W_{1,t}$ at f'_k is related to the coefficient for X_t at $\frac{1}{2} - f_k$
- in particular, coefficients for $f'_k \in [0, \frac{1}{2}]$ are related to those for $f_k \in [\frac{1}{4}, \frac{1}{2}]$, but in a *reversed* direction
- whereas filtering $\{X_t\}$ with $\{h_l\}$ yields a 'half-band' (high-pass) process, the downsampled process $\{W_{1,t}\}$ is 'full-band'

Example: $\{V_{1,t}\}$ and $\{W_{1,t}\}$ as Full-Band Processes

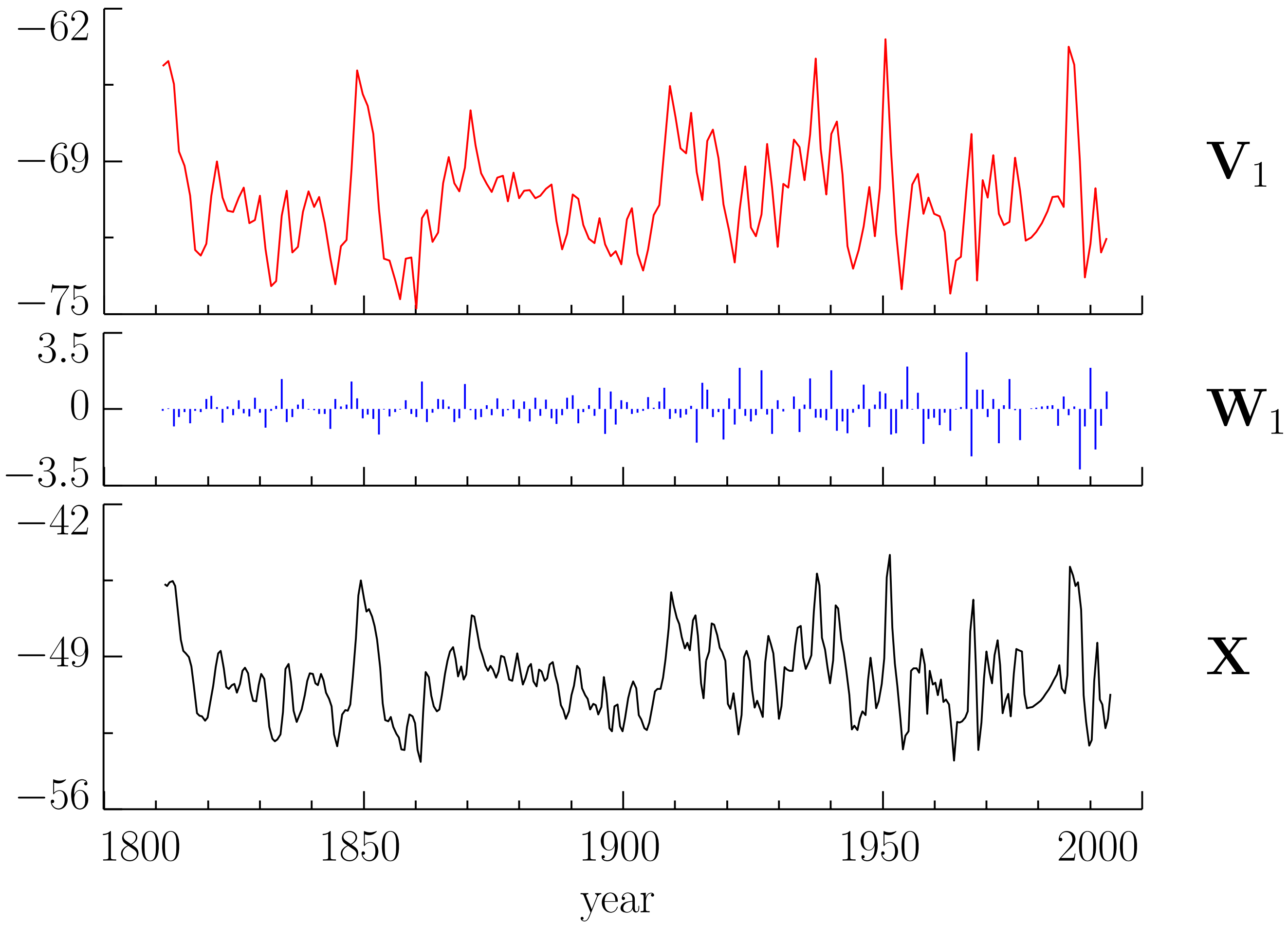
- $\{V_{1,t}\}$ and $\{W_{1,t}\}$ formed using Haar DWT



- plots are of magnitude squared DFTs for $\{X_t\}$ etc.

Example of Decomposing $\mathbf{X}$ into $\mathbf{W}_1$ and $\mathbf{V}_1$: I

- oxygen isotope records $\mathbf{X}$ from Antarctic ice core



Example of Decomposing $\mathbf{X}$ into $\mathbf{W}_1$ and $\mathbf{V}_1$: II

- oxygen isotope record series $\mathbf{X}$ has $N = 352$ observations
- spacing between observations is $\Delta t \doteq 0.5$ years
- used Haar DWT, obtaining 176 scaling and wavelet coefficients
- scaling coefficients $\mathbf{V}_1$ related to averages on scale of $2\Delta t$
- wavelet coefficients $\mathbf{W}_1$ related to changes on scale of Δt
- 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 $\mathbf{X}$ from $\mathbf{W}_1$ and $\mathbf{V}_1$

- in matrix notation, form wavelet & scaling coefficients via

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

- recall that $\mathcal{P}_1^T\mathcal{P}_1 = I_N$ because $\mathcal{P}_1$ is orthonormal
- since $\mathcal{P}_1^T\mathcal{P}_1\mathbf{X} = \mathbf{X}$, premultiplying both sides by $\mathcal{P}_1^T$ yields

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

- $\mathcal{D}_1 \equiv \mathcal{W}_1^T\mathbf{W}_1$ 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

Construction of First Level Detail: I

- consider $\mathcal{D}_1 = \mathcal{W}_1^T \mathbf{W}_1$ for $L = 4$ & $N > L$:

$$\mathcal{D}_1 = \begin{bmatrix} h_1 & h_3 & 0 & \cdots & 0 & 0 \\ h_0 & h_2 & 0 & \cdots & 0 & 0 \\ 0 & h_1 & h_3 & \cdots & 0 & 0 \\ 0 & h_0 & h_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & h_1 & h_3 \\ 0 & 0 & 0 & \cdots & h_0 & h_2 \\ h_3 & 0 & 0 & \cdots & 0 & h_1 \\ h_2 & 0 & 0 & \cdots & 0 & h_0 \end{bmatrix} \begin{bmatrix} W_{1,0} \\ W_{1,1} \\ W_{1,2} \\ \vdots \\ W_{1,N/2-2} \\ W_{1,N/2-1} \end{bmatrix}$$

note: $\mathcal{W}_1^T$ is $N \times \frac{N}{2}$ & $\mathbf{W}_1$ is $\frac{N}{2} \times 1$

- $\mathcal{D}_1$ is *not* result of filtering $W_{1,t}$'s with $\{h_0, h_1, h_2, h_3\}$

Construction of First Level Detail: II

- augment $\mathcal{W}_1$ to $N \times N$ and $\mathbf{W}_1$ to $N \times 1$:

$$\mathcal{D}_1 = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & h_0 & h_1 & h_2 & h_3 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & h_0 & h_1 & h_2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & h_1 & h_2 & h_3 \\ h_3 & 0 & 0 & 0 & 0 & 0 & \cdots & h_0 & h_1 & h_2 \\ h_2 & h_3 & 0 & 0 & 0 & 0 & \cdots & 0 & h_0 & h_1 \\ h_1 & h_2 & h_3 & 0 & 0 & 0 & \cdots & 0 & 0 & h_0 \end{bmatrix} \begin{bmatrix} 0 \\ W_{1,0} \\ 0 \\ W_{1,1} \\ 0 \\ W_{1,2} \\ \vdots \\ W_{1,N/2-2} \\ 0 \\ W_{1,N/2-1} \end{bmatrix}$$

- can now regard the above as equivalent to use of a filter

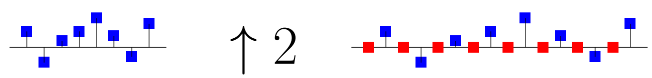
Construction of First Level Detail: III

- formally, define *upsampled* (by 2) version of $W_{1,t}$'s:

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

- example of upsampling:

$W_{1,t}$ $\uparrow 2$ $W^{\uparrow}_{1,t}$



- note: '$\uparrow 2$' denotes 'upsample by 2' (put 0's before values)

Construction of First Level Detail: IV

- can now write

$$\mathcal{D}_{1,t} = \sum_{l=0}^{N-1} h_l^\circ W^{\uparrow}_{1,t+l \bmod N}, \quad t = 0, 1, \ldots, N-1$$

- doesn't look exactly like filtering, which would look like

$$\sum_{l=0}^{N-1} h_l^\circ W^{\uparrow}_{1,t-l \bmod N};$$ i.e., direction of $W^{\uparrow}_{1,t}$ not reversed

- form that $\mathcal{D}_{1,t}$ takes is what engineers call 'cross-correlation'
- if $\{h_l\} \longleftrightarrow H(\cdot)$, cross-correlating $\{h_l\}$ & $\{W^{\uparrow}_{1,t}\}$ is equivalent to filtering $\{W^{\uparrow}_{1,t}\}$ using filter with transfer function $H^*(\cdot)$
- $\mathcal{D}_1$ formed by circularly filtering $\{W^{\uparrow}_{1,t}\}$ with filter $\{H^*(\frac{k}{N})\}$

Synthesis (Reconstruction) of $\mathbf{X}$

- can also write the tth element of first level smooth $\mathcal{S}_1$ as

$$\mathcal{S}_{1,t} = \sum_{l=0}^{L-1} g_l V^{\uparrow}_{1,t+l \bmod N}, \quad t = 0, 1, \ldots, N-1$$

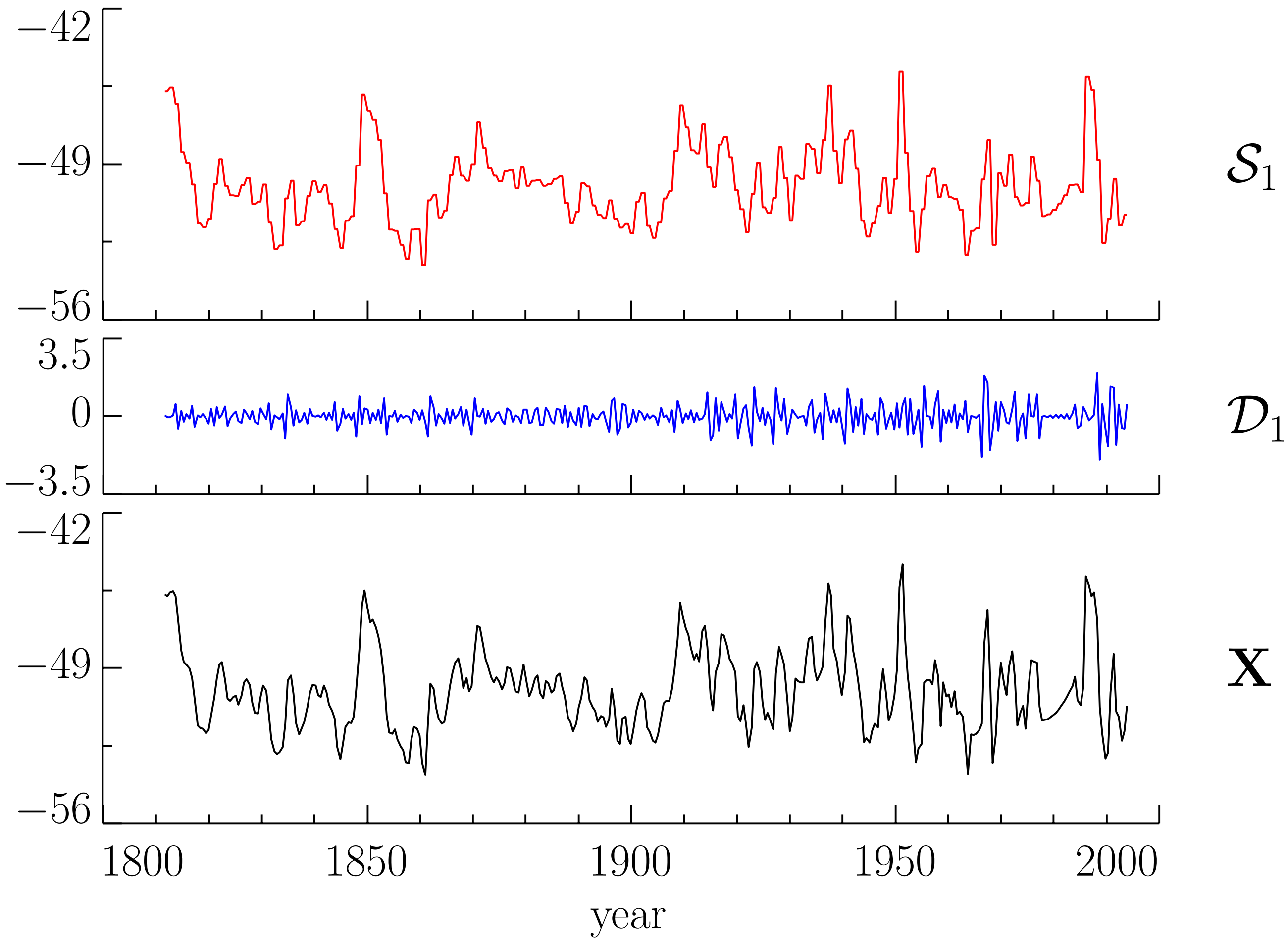
- since $\{g_l\} \longleftrightarrow G(\cdot)$, cross-correlating $\{g_l\}$ & $\{V^{\uparrow}_{1,t}\}$ is the same as circularly filtering $\{V^{\uparrow}_{1,t}\}$ using the filter $\{G^*(\frac{k}{N})\}$
- since $\mathbf{X} = \mathcal{S}_1 + \mathcal{D}_1$, can write

$$X_t = \sum_{l=0}^{N-1} h_l^\circ W^{\uparrow}_{1,t+l \bmod N} + \sum_{l=0}^{N-1} g_l^\circ V^{\uparrow}_{1,t+l \bmod N},$$

which is the filtering version of $\mathbf{X} = \mathcal{W}_1^T \mathbf{W}_1 + \mathcal{V}_1^T \mathbf{V}_1$

Example of Synthesizing X from $\mathcal{D}_1$ and $\mathcal{S}_1$

- 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

$$\mathcal{P}_1\mathbf{X} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{V}_1 \end{bmatrix} \text{ and hence } \|\mathcal{P}_1\mathbf{X}\|^2 = \|\mathbf{W}_1\|^2 + \|\mathbf{V}_1\|^2$$

- leads to a decomposition of the sample variance for $\mathbf{X}$:

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

$$= \frac{1}{N}\|\mathbf{W}_1\|^2 + \frac{1}{N}\|\mathbf{V}_1\|^2 - \overline{X}^2$$

First Level Variance Decomposition: II

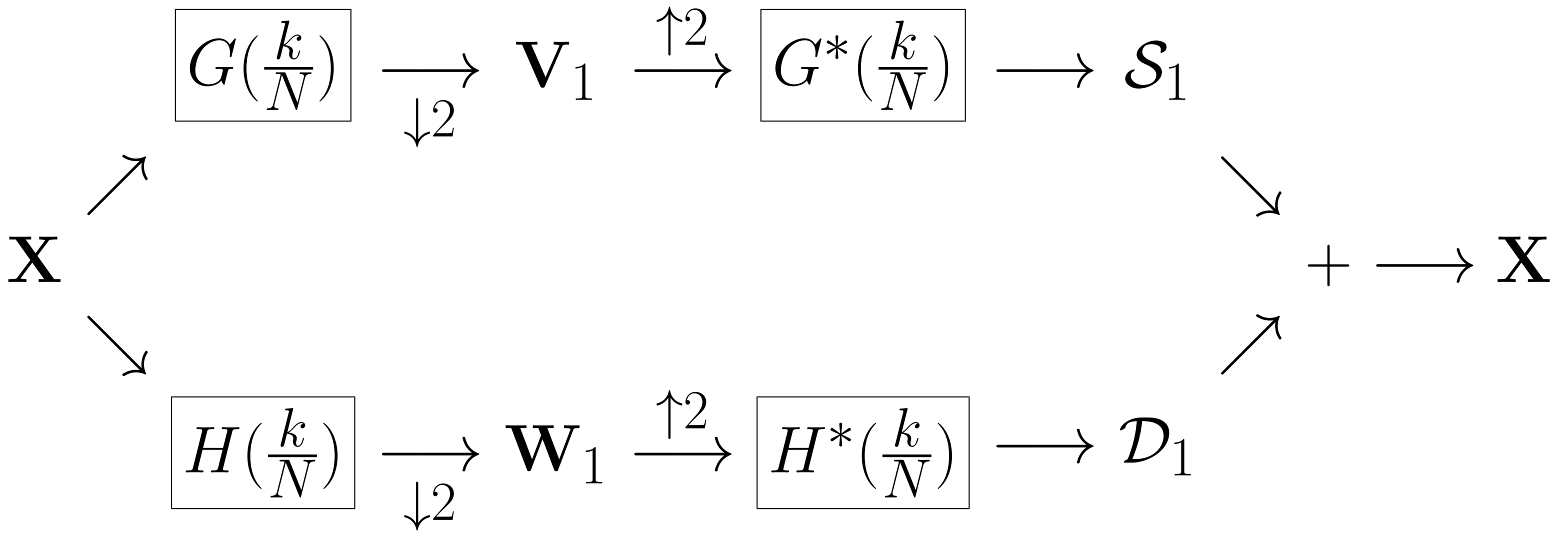
- breaks up $\hat{\sigma}^2_X$ into two pieces:
  1. $\frac{1}{N}\|\mathbf{W}_1\|^2$, attributable to changes in averages over scale 1
  2. $\frac{1}{N}\|\mathbf{V}_1\|^2 - \overline{X}^2$, attributable to averages over scale 2
- Haar-based example for oxygen isotope records
  - first piece: $\frac{1}{N}\|\mathbf{W}_1\|^2 \doteq 0.295$
  - second piece: $\frac{1}{N}\|\mathbf{V}_1\|^2 - \overline{X}^2 \doteq 2.909$
  - sample variance: $\hat{\sigma}^2_X \doteq 3.204$
  - changes on scale of $\Delta t \doteq 0.5$ years account for 9% of $\hat{\sigma}^2_X$ (standardized scale of 1 corresponds to physical scale of Δt)

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:
  - $\mathbf{W}_1$, 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 $\mathcal{D}_1$
  - $\mathcal{W}_1$, an $\frac{N}{2} \times N$ matrix consisting of first $\frac{N}{2}$ rows of $\mathcal{W}$
- $N/2$ scaling coefficients $\{V_{1,t}\}$ associated with:
  - $\mathbf{V}_1$, a vector of length $N/2$
  - averages on scale 2 and nominal frequencies $0 \leq f \leq \frac{1}{4}$
  - first level smooth $\mathcal{S}_1$
  - $\mathcal{V}_1$, an $\frac{N}{2} \times N$ matrix spanning same subspace as last $N/2$ rows of $\mathcal{W}$

Level One Analysis and Synthesis of X

- can express analysis/synthesis of $\mathbf{X}$ as a flow diagram



Constructing Remaining DWT Coefficients: I

- have regarded time series X_t as 'one point' averages $\overline{X}_t(1)$ over
  - physical scale of Δt (sampling interval between observations)
  - standardized scale of 1
- first level of basic algorithm transforms $\mathbf{X}$ of length N into
  - $N/2$ wavelet coefficients $\mathbf{W}_1 \propto$ changes on a scale of 1
  - $N/2$ scaling coefficients $\mathbf{V}_1 \propto$ averages of X_t on a scale of 2
- in essence basic algorithm takes length N series $\mathbf{X}$ related to scale 1 averages and produces
  - length $N/2$ series $\mathbf{W}_1$ associated with the same scale
  - length $N/2$ series $\mathbf{V}_1$ related to averages on double the scale

Constructing Remaining DWT Coefficients: II

- Q: what if we now treat $\mathbf{V}_1$ in the same manner as $\mathbf{X}$?
- basic algorithm will transform length $N/2$ series $\mathbf{V}_1$ into
  - length $N/4$ series $\mathbf{W}_2$ associated with the same scale (2)
  - length $N/4$ series $\mathbf{V}_2$ related to averages on twice the scale
- by definition, $\mathbf{W}_2$ contains the level 2 wavelet coefficients
- Q: what if we treat $\mathbf{V}_2$ in the same way?
- basic algorithm will transform length $N/4$ series $\mathbf{V}_2$ into
  - length $N/8$ series $\mathbf{W}_3$ associated with the same scale (4)
  - length $N/8$ series $\mathbf{V}_3$ related to averages on twice the scale
- by definition, $\mathbf{W}_3$ contains the level 3 wavelet coefficients

Constructing Remaining DWT Coefficients: III

- continuing in this manner defines remaining subvectors of $\mathbf{W}$ (recall that $\mathbf{W} = \mathcal{W}\mathbf{X}$ 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 *define* the subvectors $\mathbf{W}_1$, $\mathbf{W}_2$, ..., $\mathbf{W}_J$, $\mathbf{V}_J$ of DWT coefficient vector $\mathbf{W}$
- overall scheme is known as the 'pyramid' algorithm
- item [1] of Comments and Extensions to Sec. 4.6 has pseudo code for DWT pyramid algorithm

Scales Associated with DWT Coefficients

- jth level of algorithm transforms scale 2^{j-1} averages into
  - differences of averages on scale 2^{j-1}, i.e., $\mathbf{W}_j$, the wavelet coefficients
  - averages on scale $2 \times 2^{j-1} = 2^j$, i.e., $\mathbf{V}_j$, the scaling coefficients
- let $\tau_j \equiv 2^{j-1}$ be standardized scale associated with $\mathbf{W}_j$
  - for $j = 1, \ldots, J$, takes on values $1, 2, 4, \ldots, N/4, N/2$
  - physical (actual) scale given by $\tau_j \, \Delta t$
- let $\lambda_j \equiv 2^j$ be standardized scale associated with $\mathbf{V}_j$
  - takes on values $2, 4, 8, \ldots, N/2, N$
  - physical scale given by $\lambda_j \, \Delta t$

Matrix Description of Pyramid Algorithm: I

- define $\frac{N}{2^j} \times \frac{N}{2^{j-1}}$ matrix $\mathcal{B}_j$ in same way as $\frac{N}{2} \times N$ matrix $\mathcal{W}_1$; i.e., rows contain $\{h_l\}$ periodized to length $N/2^{j-1}$
- for $N/2^j = 8$ and $N/2^{j-1} = 16$ when $L = 4$, have

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

- matrix gets us jth level wavelet coefficients via $\mathbf{W}_j = \mathcal{B}_j \mathbf{V}_{j-1}$

Matrix Description of Pyramid Algorithm: II

- define $\frac{N}{2^j} \times \frac{N}{2^{j-1}}$ matrix $\mathcal{A}_j$ in same way as $\frac{N}{2} \times N$ matrix $\mathcal{V}_1$; i.e., rows contain $\{g_l\}$ periodized to length $N/2^{j-1}$
- for $N/2^j = 8$ and $N/2^{j-1} = 16$ when $L = 4$, have

$$\mathcal{A}_j = \begin{bmatrix} g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 \\ g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 \end{bmatrix}$$

- matrix gets us jth level scaling coefficients via $\mathbf{V}_j = \mathcal{A}_j \mathbf{V}_{j-1}$

Matrix Description of Pyramid Algorithm: III

- if we define $\mathbf{V}_0 = \mathbf{X}$ and let $j = 1$, then

$$\mathbf{W}_j = \mathcal{B}_j \mathbf{V}_{j-1} \text{ reduces to } \mathbf{W}_1 = \mathcal{B}_1 \mathbf{V}_0 = \mathcal{B}_1 \mathbf{X} = \mathcal{W}_1 \mathbf{X}$$

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

- likewise, when $j = 1$,

$$\mathbf{V}_j = \mathcal{A}_j \mathbf{V}_{j-1} \text{ reduces to } \mathbf{V}_1 = \mathcal{A}_1 \mathbf{V}_0 = \mathcal{A}_1 \mathbf{X} = \mathcal{V}_1 \mathbf{X}$$

  because $\mathcal{A}_1$ has the same definition as $\mathcal{V}_1$

Formation of Submatrices of $\mathcal{W}$: I

- using $\mathbf{V}_j = \mathcal{A}_j \mathbf{V}_{j-1}$ repeatedly and $\mathbf{V}_1 = \mathcal{A}_1 \mathbf{X}$, can write

$$\begin{aligned} \mathbf{W}_j &= \mathcal{B}_j \mathbf{V}_{j-1} \\ &= \mathcal{B}_j \mathcal{A}_{j-1} \mathbf{V}_{j-2} \\ &= \mathcal{B}_j \mathcal{A}_{j-1} \mathcal{A}_{j-2} \mathbf{V}_{j-3} \\ &= \mathcal{B}_j \mathcal{A}_{j-1} \mathcal{A}_{j-2} \cdots \mathcal{A}_1 \mathbf{X} \equiv \mathcal{W}_j \mathbf{X}, \end{aligned}$$

where $\mathcal{W}_j$ is $\frac{N}{2^j} \times N$ submatrix of $\mathcal{W}$ responsible for $\mathbf{W}_j$

- likewise, can get $1 \times N$ submatrix $\mathcal{V}_J$ responsible for $\mathbf{V}_J$

$$\begin{aligned} \mathbf{V}_J &= \mathcal{A}_J \mathbf{V}_{J-1} \\ &= \mathcal{A}_J \mathcal{A}_{J-1} \mathbf{V}_{J-2} \\ &= \mathcal{A}_J \mathcal{A}_{J-1} \mathcal{A}_{J-2} \mathbf{V}_{J-3} \\ &= \mathcal{A}_J \mathcal{A}_{J-1} \mathcal{A}_{J-2} \cdots \mathcal{A}_1 \mathbf{X} \equiv \mathcal{V}_J \mathbf{X} \end{aligned}$$

- $\mathcal{V}_J$ is the last row of $\mathcal{W}$, & all its elements are equal to $1/\sqrt{N}$

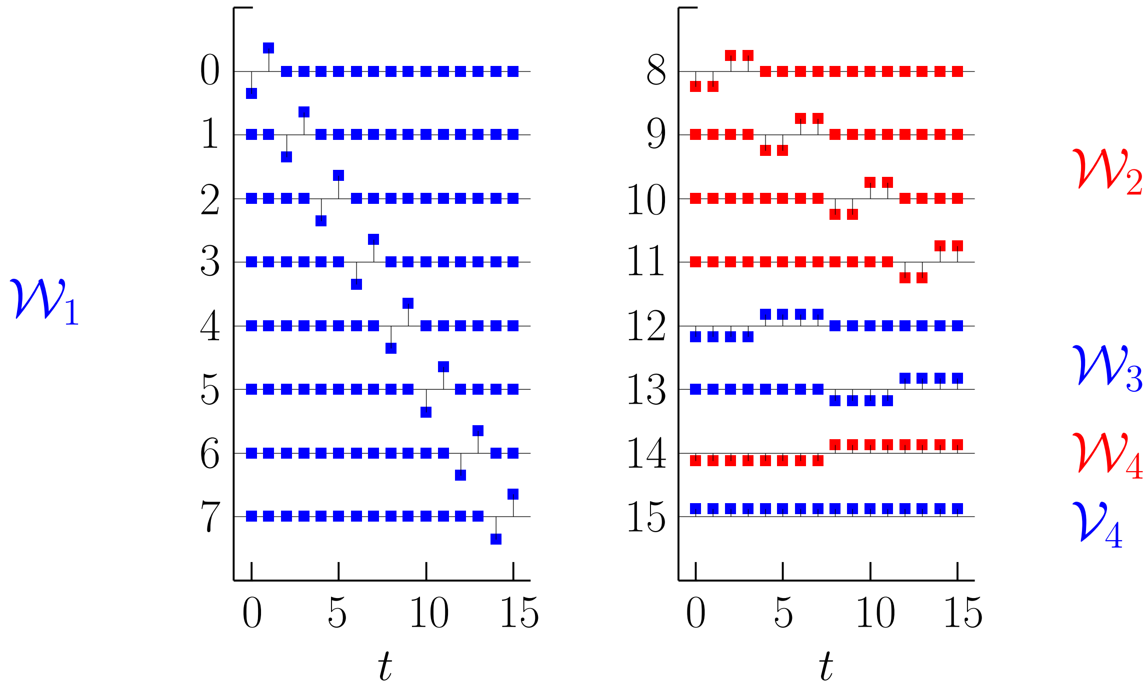
Formation of Submatrices of $\mathcal{W}$: II

- 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

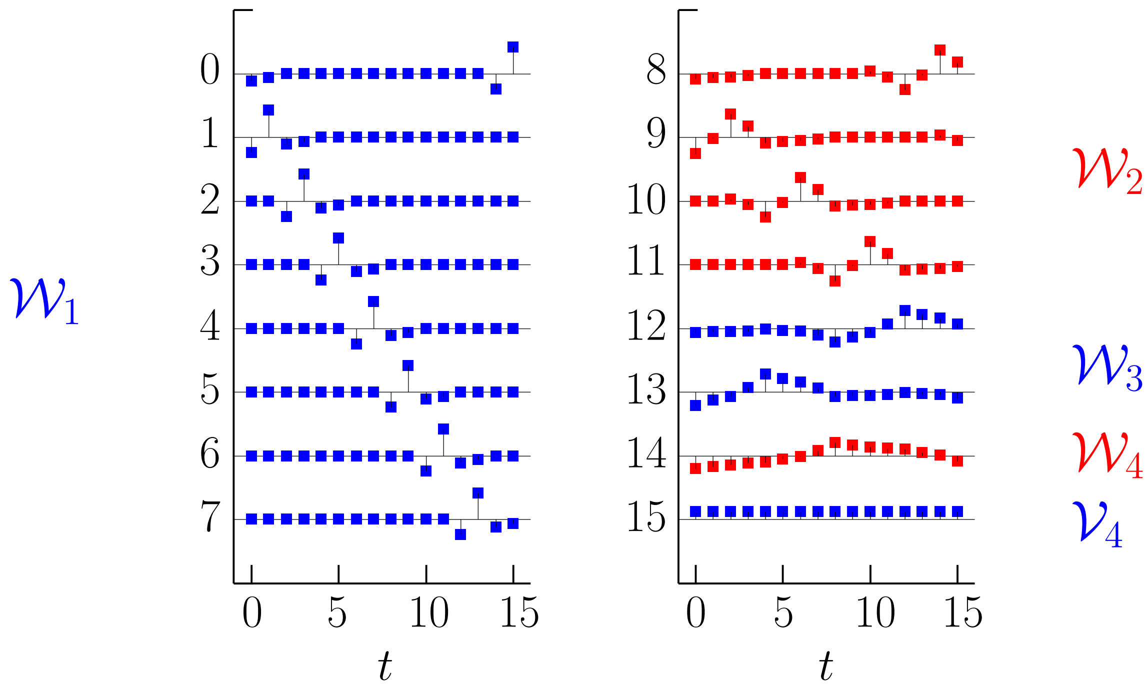
- $N = 16$ case for Haar DWT matrix $\mathcal{W}$



- 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

Matrix Description of Multiresolution Analysis: I

- just as we could reconstruct $\mathbf{X}$ from $\mathbf{W}_1$ and $\mathbf{V}_1$ using
$$\mathbf{X} = \mathcal{W}_1^T \mathbf{W}_1 + \mathcal{V}_1^T \mathbf{V}_1,$$
so can we reconstruct $\mathbf{V}_{j-1}$ from $\mathbf{W}_j$ and $\mathbf{V}_j$ using
$$\mathbf{V}_{j-1} = \mathcal{B}_j^T \mathbf{W}_j + \mathcal{A}_j^T \mathbf{V}_j$$
(recall the correspondences $\mathbf{V}_0 = \mathbf{X}$, $\mathcal{B}_1 = \mathcal{W}_1$ and $\mathcal{A}_1 = \mathcal{V}_1$)

- we can thus write
$$\begin{aligned}
\mathbf{X} &= \mathcal{B}_1^T \mathbf{W}_1 + \mathcal{A}_1^T \mathbf{V}_1 \\
&= \mathcal{B}_1^T \mathbf{W}_1 + \mathcal{A}_1^T (\mathcal{B}_2^T \mathbf{W}_2 + \mathcal{A}_2^T \mathbf{V}_2) \\
&= \mathcal{B}_1^T \mathbf{W}_1 + \mathcal{A}_1^T \mathcal{B}_2^T \mathbf{W}_2 + \mathcal{A}_1^T \mathcal{A}_2^T \mathbf{V}_2 \\
&= \mathcal{B}_1^T \mathbf{W}_1 + \mathcal{A}_1^T \mathcal{B}_2^T \mathbf{W}_2 + \mathcal{A}_1^T \mathcal{A}_2^T (\mathcal{B}_3^T \mathbf{W}_3 + \mathcal{A}_3^T \mathbf{V}_3) \\
&= \mathcal{B}_1^T \mathbf{W}_1 + \mathcal{A}_1^T \mathcal{B}_2^T \mathbf{W}_2 + \mathcal{A}_1^T \mathcal{A}_2^T \mathcal{B}_3^T \mathbf{W}_3 + \mathcal{A}_1^T \mathcal{A}_2^T \mathcal{A}_3^T \mathbf{V}_3
\end{aligned}$$

Matrix Description of Multiresolution Analysis: II

- studying the bottom line

$$\mathbf{X} = \mathcal{B}_1^T \mathbf{W}_1 + \mathcal{A}_1^T \mathcal{B}_2^T \mathbf{W}_2 + \mathcal{A}_1^T \mathcal{A}_2^T \mathcal{B}_3^T \mathbf{W}_3 + \mathcal{A}_1^T \mathcal{A}_2^T \mathcal{A}_3^T \mathbf{V}_3$$

  says jth level detail should be $\mathcal{D}_j \equiv \mathcal{A}_1^T \mathcal{A}_2^T \cdots \mathcal{A}_{j-1}^T \mathcal{B}_j^T \mathbf{W}_j$

- likewise, letting jth level smooth be $\mathcal{S}_j \equiv \mathcal{A}_1^T \mathcal{A}_2^T \cdots \mathcal{A}_j^T \mathbf{V}_j$ yields, for $1 \leq k \leq J$,

$$\mathbf{X} = \sum_{j=1}^{k} \mathcal{D}_j + \mathcal{S}_k \text{ and, in particular, } \mathbf{X} = \sum_{j=1}^{J} \mathcal{D}_j + \mathcal{S}_J$$

- above are multiresolution analyses (MRAs) for levels k and J; i.e., additive decomposition (first of two basic decompositions derivable from DWT)

Matrix Description of Energy Decomposition: I

- just as we can recover the energy in $\mathbf{X}$ from $\mathbf{W}_1$ & $\mathbf{V}_1$ using
$$\|\mathbf{X}\|^2 = \|\mathbf{W}_1\|^2 + \|\mathbf{V}_1\|^2,$$
so can we recover the energy in $\mathbf{V}_{j-1}$ from $\mathbf{W}_j$ & $\mathbf{V}_j$ using
$$\|\mathbf{V}_{j-1}\|^2 = \|\mathbf{W}_j\|^2 + \|\mathbf{V}_j\|^2$$
(recall the correspondence $\mathbf{V}_0 = \mathbf{X}$)
- we can thus write
$$\begin{aligned} \|\mathbf{X}\|^2 &= \|\mathbf{W}_1\|^2 + \|\mathbf{V}_1\|^2 \\ &= \|\mathbf{W}_1\|^2 + \|\mathbf{W}_2\|^2 + \|\mathbf{V}_2\|^2 \\ &= \|\mathbf{W}_1\|^2 + \|\mathbf{W}_2\|^2 + \|\mathbf{W}_3\|^2 + \|\mathbf{V}_3\|^2 \end{aligned}$$

Matrix Description of Energy Decomposition: II

- generalizing from the bottom line

$$\|\mathbf{X}\|^2 = \|\mathbf{W}_1\|^2 + \|\mathbf{W}_2\|^2 + \|\mathbf{W}_3\|^2 + \|\mathbf{V}_3\|^2$$

  indicates that, for $1 \leq k \leq J$, we can write

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{k} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_k\|^2$$

  and, in particular,

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_J\|^2$$

- above are energy decompositions for levels k and J (second of two basic decompositions derivable from DWT)

Partial DWT: I

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

$$\begin{bmatrix} \mathcal{W}_1 \\ \mathcal{W}_2 \\ \vdots \\ \mathcal{W}_j \\ \vdots \\ \mathcal{W}_{J_0} \\ \mathcal{V}_{J_0} \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{B}_2\mathcal{A}_1 \\ \vdots \\ \mathcal{B}_j\mathcal{A}_{j-1}\cdots\mathcal{A}_1 \\ \vdots \\ \mathcal{B}_{J_0}\mathcal{A}_{J_0-1}\cdots\mathcal{A}_1 \\ \mathcal{A}_{J_0}\mathcal{A}_{J_0-1}\cdots\mathcal{A}_1 \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_j \\ \vdots \\ \mathbf{W}_{J_0} \\ \mathbf{V}_{J_0} \end{bmatrix}$$

- $\mathcal{V}_{J_0}$ 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:

$$\mathbf{X} = \sum_{j=1}^{J_0} \mathcal{D}_j + \mathcal{S}_{J_0}$$

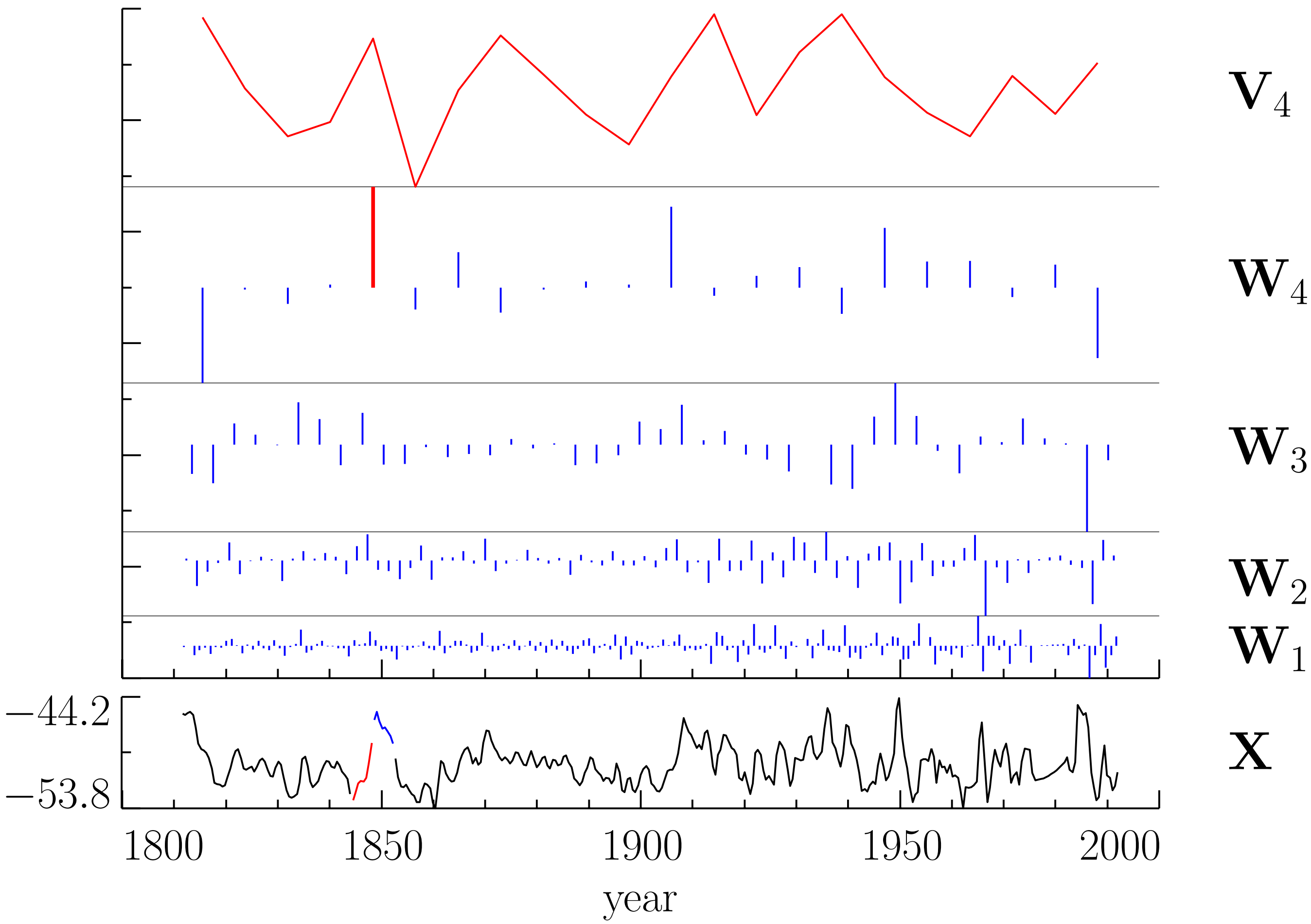
$\mathcal{S}_{J_0}$ represents averages on scale $\lambda_{J_0} = 2^{J_0}$ (includes $\overline{X}$)

- analysis of variance for partial DWT:

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \frac{1}{N}\|\mathbf{V}_{J_0}\|^2 - \overline{X}^2$$

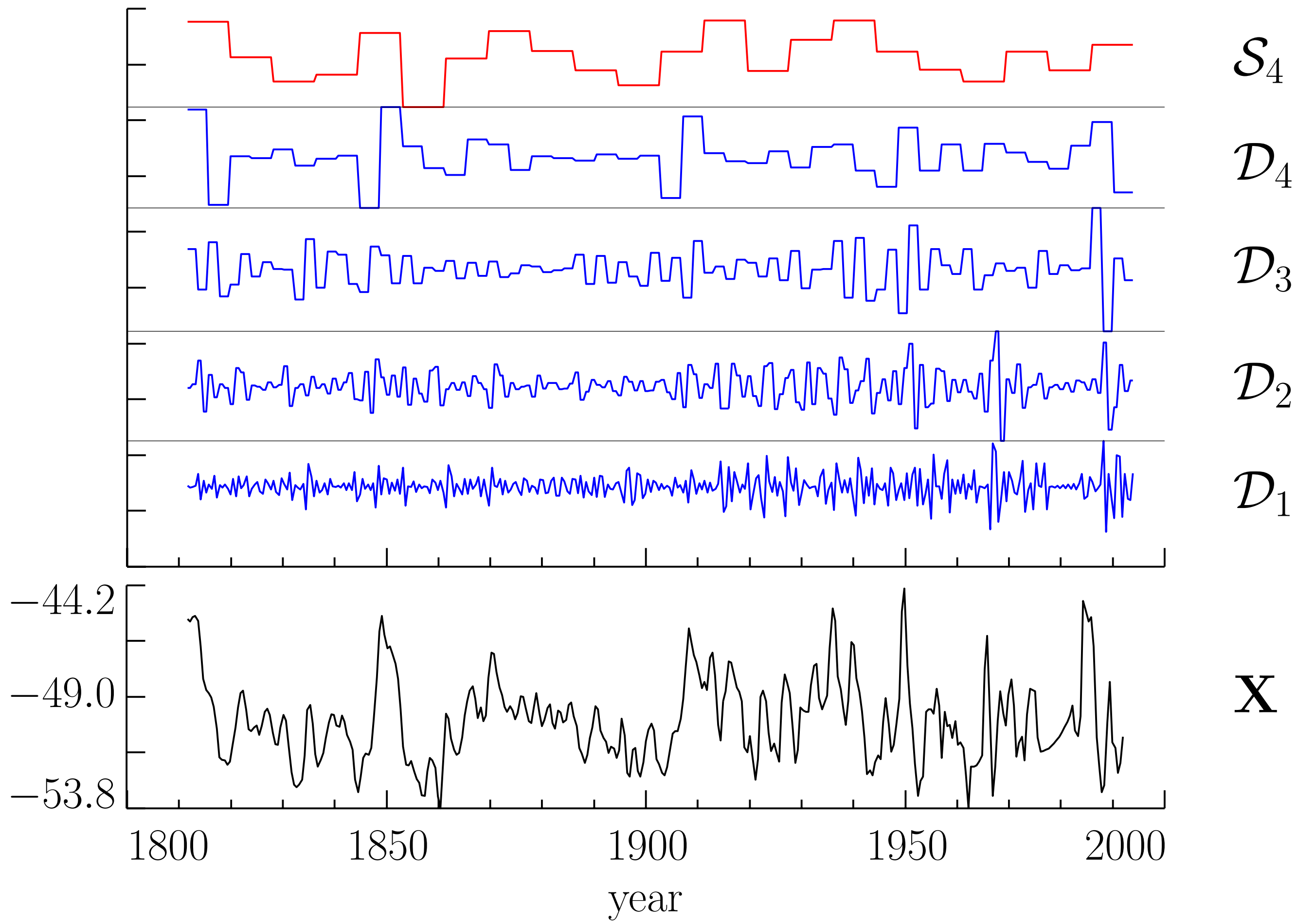
Example of $J_0 = 4$ Partial Haar DWT

- oxygen isotope records **X** from Antarctic ice core



Example of MRA from $J_0 = 4$ Partial Haar DWT

- oxygen isotope records **X** from Antarctic ice core



Example of Variance Decomposition

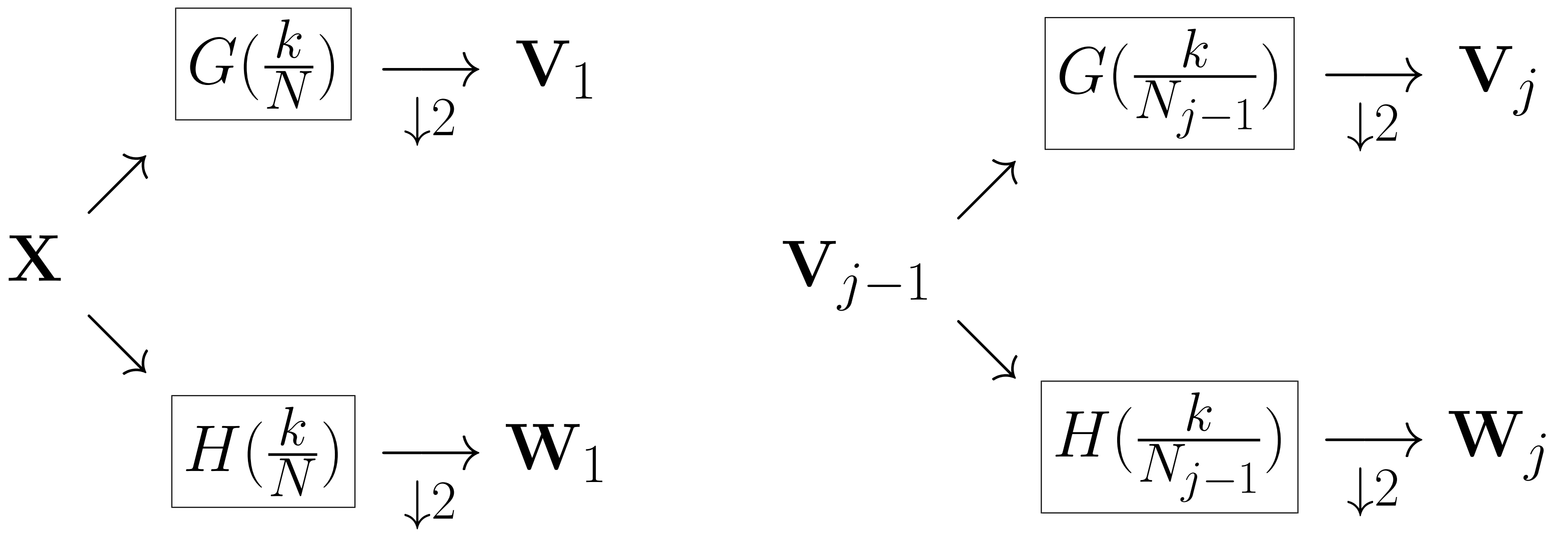
- decomposition of sample variance from $J_0 = 4$ partial DWT

$$\hat{\sigma}_X^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} \left(X_t - \overline{X}\right)^2 = \sum_{j=1}^{4} \frac{1}{N} \|\mathbf{W}_j\|^2 + \frac{1}{N} \|\mathbf{V}_4\|^2 - \overline{X}^2$$

- Haar-based example for oxygen isotope records
  - 0.5 year changes: $\frac{1}{N}\|\mathbf{W}_1\|^2 \doteq 0.295$ ($\doteq$ 9.2% of $\hat{\sigma}_X^2$)
  - 1.0 years changes: $\frac{1}{N}\|\mathbf{W}_2\|^2 \doteq 0.464$ ($\doteq$ 14.5%)
  - 2.0 years changes: $\frac{1}{N}\|\mathbf{W}_3\|^2 \doteq 0.652$ ($\doteq$ 20.4%)
  - 4.0 years changes: $\frac{1}{N}\|\mathbf{W}_4\|^2 \doteq 0.846$ ($\doteq$ 26.4%)
  - 8.0 years averages: $\frac{1}{N}\|\mathbf{V}_4\|^2 - \overline{X}^2 \doteq 0.947$ ($\doteq$ 29.5%)
  - sample variance: $\hat{\sigma}_X^2 \doteq 3.204$

Filtering Description of Pyramid Algorithm

- flow diagrams for analyses of $\mathbf{X}$ at level 1 and of $\mathbf{V}_{j-1}$ at level j are quite similar:



- in the above $N_j \equiv N/2^j$ (also recall $\mathbf{V}_0 = \mathbf{X}$ by definition)

Equivalent Wavelet Filter for Level $j = 3$

- consider flow diagram for extracting $\mathbf{W}_3$ from $\mathbf{X}$:

$$\mathbf{X} \longrightarrow \boxed{G(\tfrac{k}{N})} \underset{\downarrow 2}{\longrightarrow} \boxed{G(\tfrac{k}{N_1})} \underset{\downarrow 2}{\longrightarrow} \boxed{H(\tfrac{k}{N_2})} \underset{\downarrow 2}{\longrightarrow} \mathbf{W}_3$$

- can be regarded as filter cascade, but must adjust for '$\downarrow 2$'
- equivalent filter for cascade can be represented by
  - impulse response sequence $\{h_{3,l}\}$
  - transfer function $H_3(f) \equiv G(f)G(2f)H(4f)$, where, as usual, $\{h_{3,l}\} \longleftrightarrow H_3(\cdot)$
- in above, '$2f$' and '$4f$' adjust for downsampling (Exer. [91])
- with the equivalent filter, flow diagram becomes

$$\mathbf{X} \longrightarrow \boxed{H_3(\tfrac{k}{N})} \underset{\downarrow 8}{\longrightarrow} \mathbf{W}_3$$

Equivalent Scaling Filter for Level $j = 3$

- similar results hold for transforming $\mathbf{X}$ into $\mathbf{V}_3$:

$$\mathbf{X} \longrightarrow \boxed{G(\tfrac{k}{N})} \underset{\downarrow 2}{\longrightarrow} \boxed{G(\tfrac{k}{N_1})} \underset{\downarrow 2}{\longrightarrow} \boxed{G(\tfrac{k}{N_2})} \underset{\downarrow 2}{\longrightarrow} \mathbf{V}_3$$

- equivalent filter for cascade can be represented by
  - impulse response sequence $\{g_{3,l}\}$
  - transfer function $G_3(f) \equiv G(f)G(2f)G(4f)$, where, once again, $\{g_{3,l}\} \longleftrightarrow G_3(\cdot)$
- with the equivalent filter, flow diagram becomes

$$\mathbf{X} \longrightarrow \boxed{G_3(\tfrac{k}{N})} \underset{\downarrow 8}{\longrightarrow} \mathbf{V}_3$$

Equivalent Wavelet & Scaling Filters for Level j

- results generalize in an obvious way to other levels j
- jth level equivalent wavelet filter can be represented by
  - impulse response sequence $\{h_{j,l}\} \longleftrightarrow H_j(\cdot)$
  - transfer function $H_j(f) \equiv H(2^{j-1}f) \prod_{l=0}^{j-2} G(2^l f)$
- jth level equivalent scaling filter can be represented by
  - impulse response sequence $\{g_{j,l}\} \longleftrightarrow G_j(\cdot)$
  - transfer function $G_j(f) \equiv \prod_{l=0}^{j-1} G(2^l f)$
- convenient to define $H_1(f) = H(f)$ and $G_1(f) = G(f)$
- flow diagrams become

$$\mathbf{X} \longrightarrow \boxed{H_j(\tfrac{k}{N})} \underset{\downarrow 2^j}{\longrightarrow} \mathbf{W}_j \quad \text{and} \quad \mathbf{X} \longrightarrow \boxed{G_j(\tfrac{k}{N})} \underset{\downarrow 2^j}{\longrightarrow} \mathbf{V}_j$$

Relating Filtering and Matrix Descriptions

- because $\mathbf{W}_j = \mathcal{W}_j\mathbf{X}$ and because

$$\mathbf{X} \longrightarrow \boxed{H_j(\tfrac{k}{N})} \underset{\downarrow 2^j}{\longrightarrow} \mathbf{W}_j$$

  can argue that

  – rows of $\mathcal{W}_j$ must contain values dictated by $\{h_{j,l}\}$ after periodization to length N

  – adjacent rows are circularly shifted by 2^j units

- from $\mathbf{V}_j = \mathcal{V}_j\mathbf{X}$ & related flow diagram, can also argue that

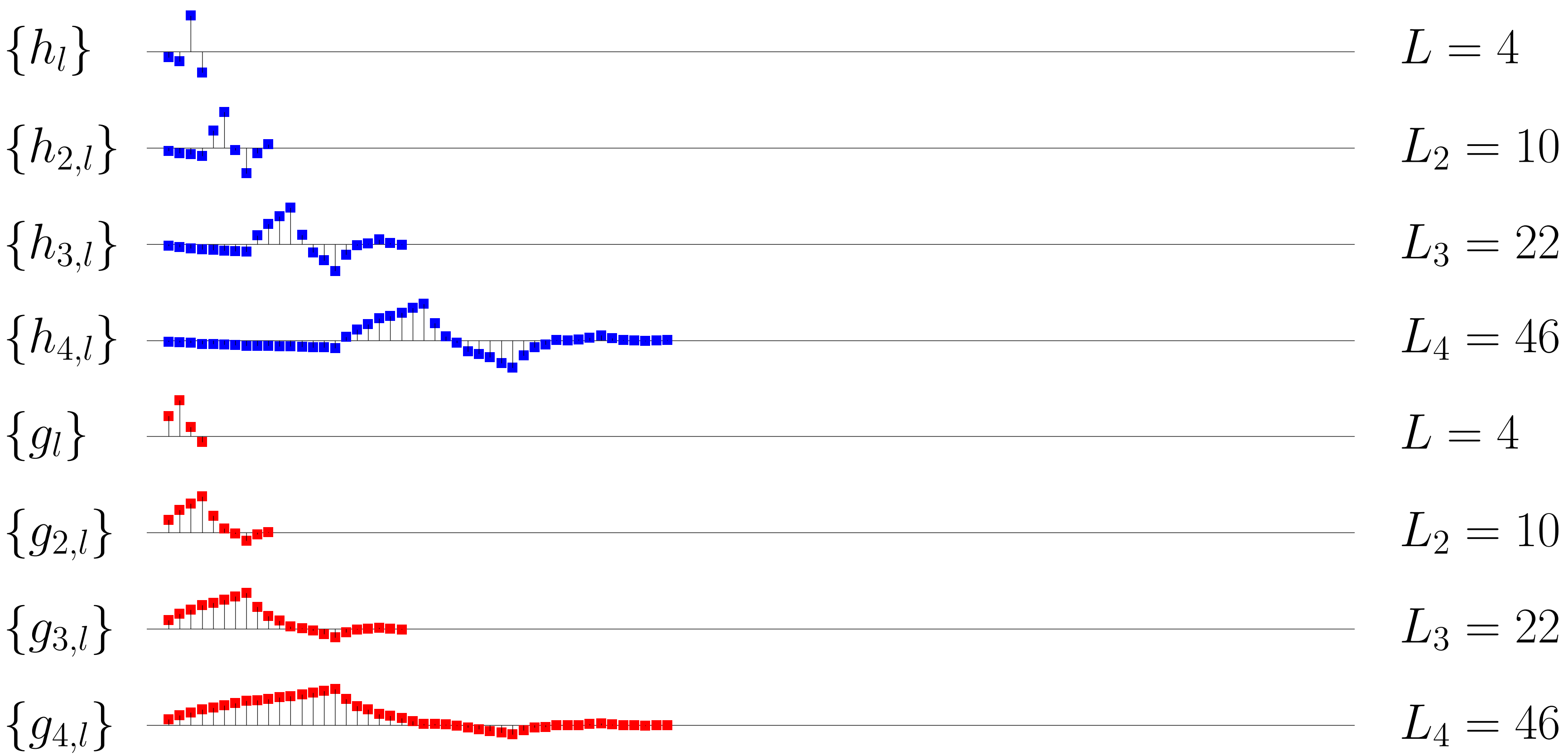
  – rows of $\mathcal{V}_j$ must contain values dictated by $\{g_{j,l}\}$ after periodization to length N

  – adjacent rows are circularly shifted by 2^j units

Haar Equivalent Wavelet & Scaling Filters

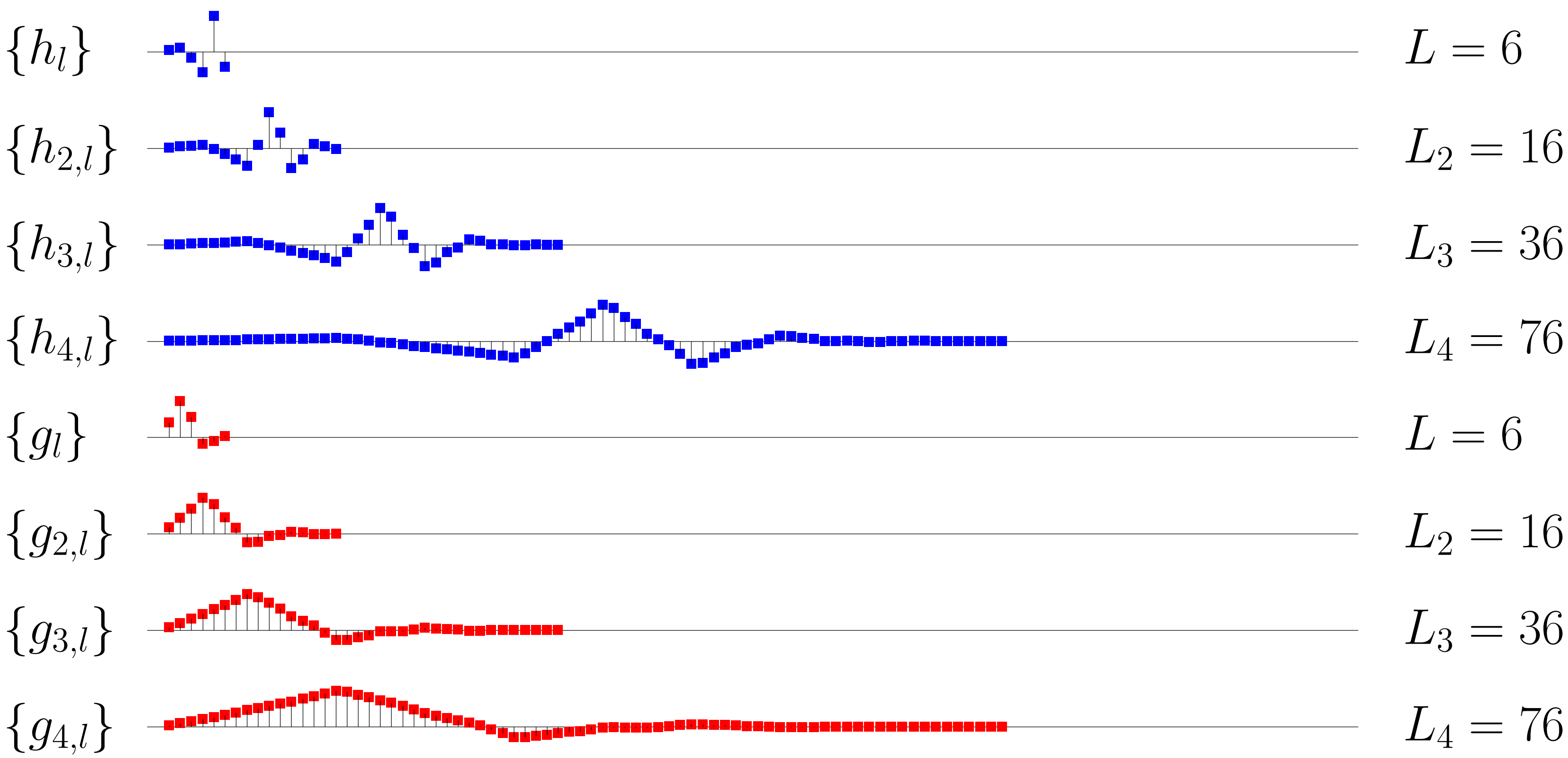
- $L_j = 2^j$ is width of $\{h_{j,l}\}$ and $\{g_{j,l}\}$

D(4) Equivalent Wavelet & Scaling Filters



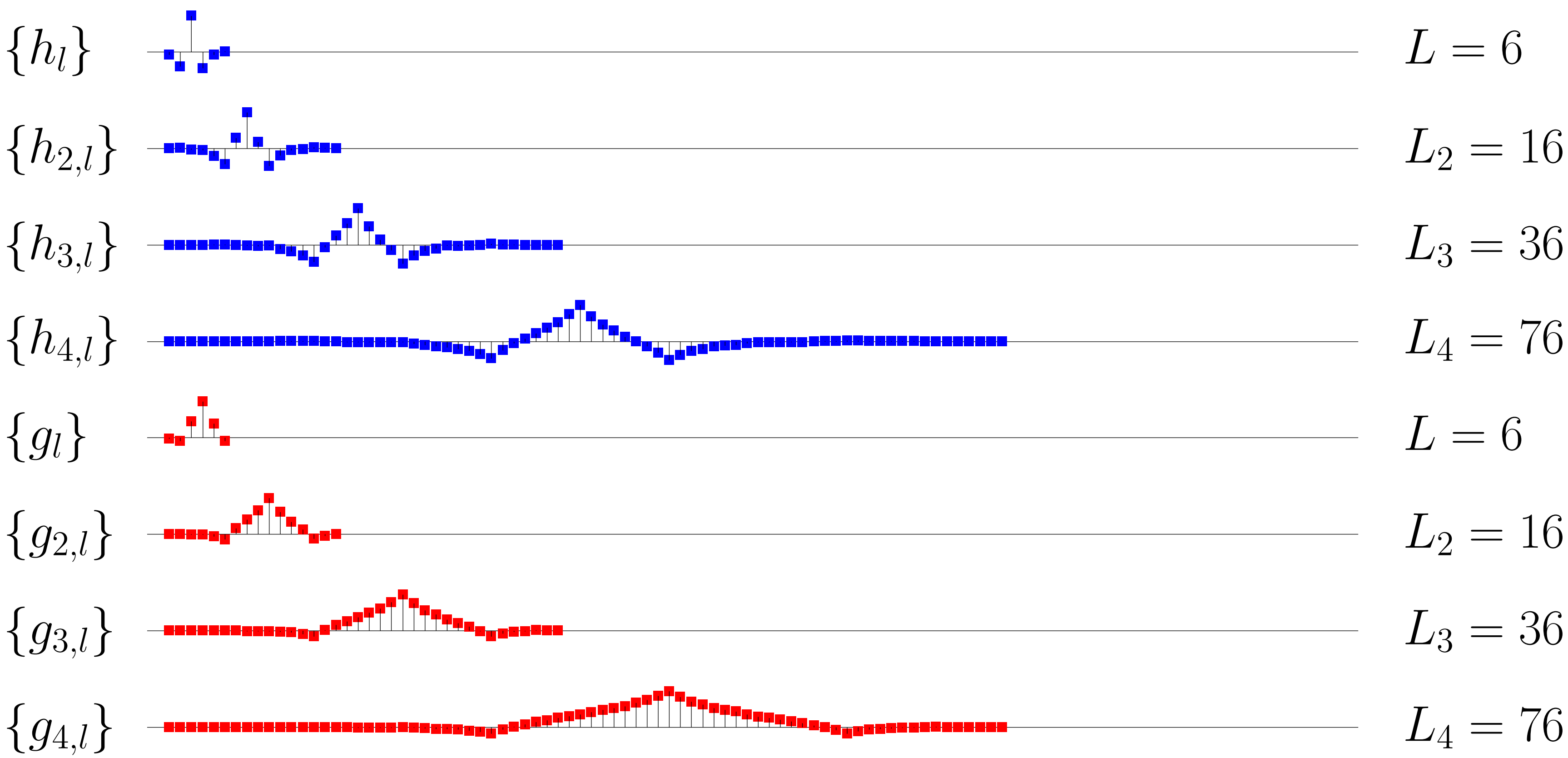
- L_j dictated by general formula $L_j = (2^j - 1)(L - 1) + 1$, but can argue that *effective* width is 2^j (same as Haar L_j)

D(6) Equivalent Wavelet & Scaling Filters



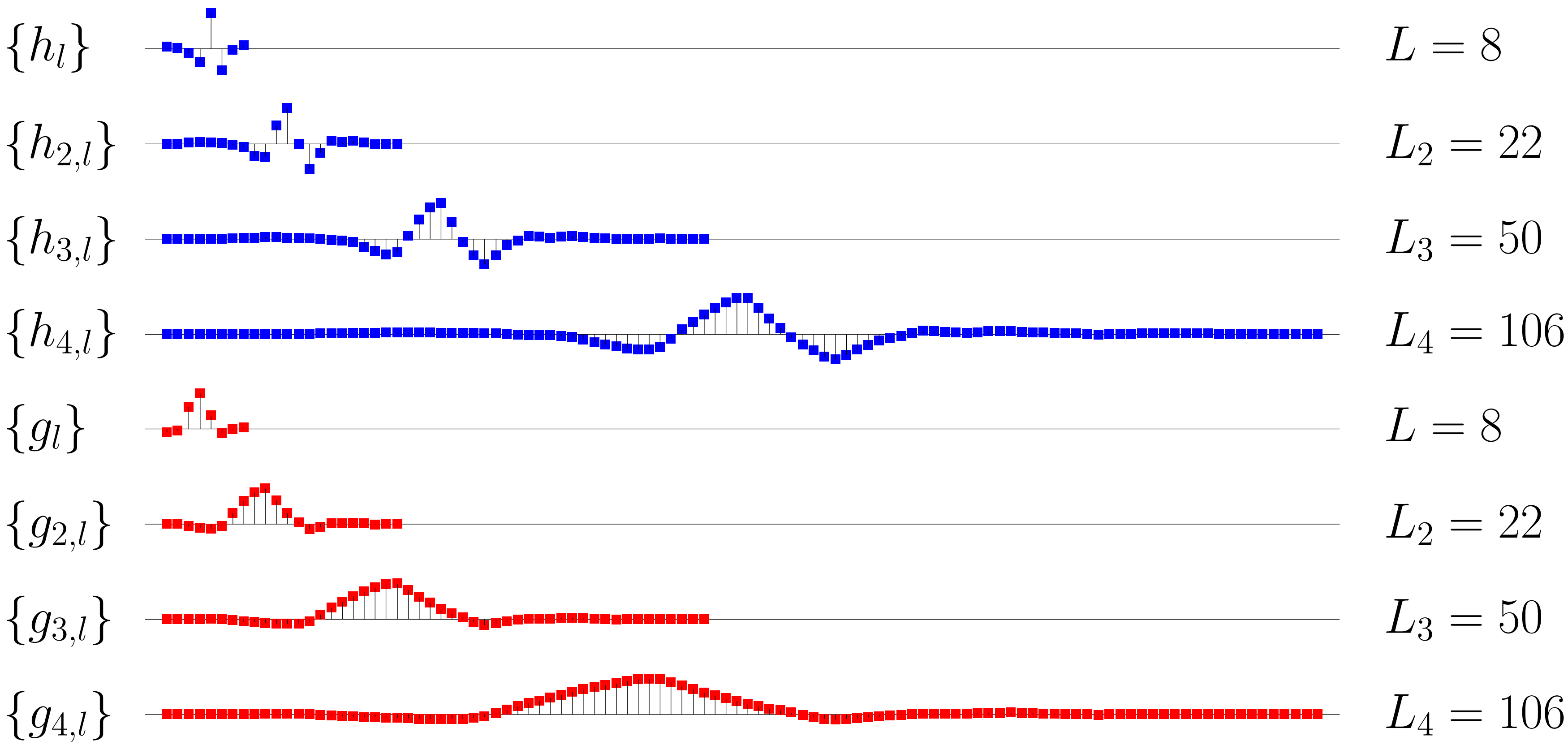
- $\{h_{4,l}\}$ resembles discretized version of Mexican hat wavelet

C(6) Equivalent Wavelet & Scaling Filters



- $\{g_{j,l}\}$ yields 'triangularly' weighted average (effective width 2^j)

LA(8) Equivalent Wavelet & Scaling Filters



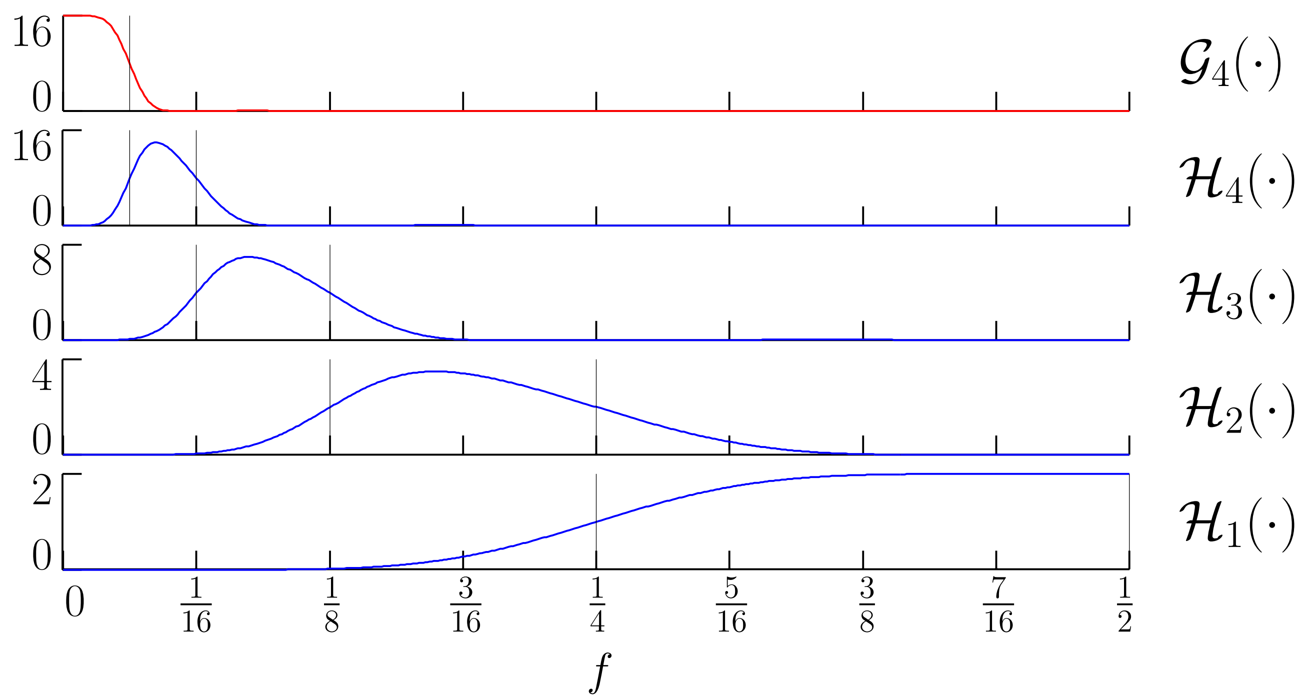
- $\{h_{j,l}\}$ resembles discretized version of Mexican hat wavelet, again with an effective width of 2^j

Squared Gain Functions for Filters

- squared gain functions give us frequency domain properties:

$$\mathcal{H}_j(f) \equiv |H_j(f)|^2 \text{ and } \mathcal{G}_j(f) \equiv |G_j(f)|^2$$

- example: squared gain functions for LA(8) $J_0 = 4$ partial DWT



Summary of Key Points about the DWT: I

- 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 $\mathbf{X}$ whose length N is an integer multiple of 2^{J_0} into DWT coefficients $\mathbf{W}_1$, $\mathbf{W}_2$, ..., $\mathbf{W}_{J_0}$ and $\mathbf{V}_{J_0}$ (sizes of vectors are $\frac{N}{2}$, $\frac{N}{4}$, ..., $\frac{N}{2^{J_0}}$ and $\frac{N}{2^{J_0}}$, 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 $\mathbf{X}$ is reexpressed as

$$\mathbf{X} = \sum_{j=1}^{J_0} \mathcal{D}_j + \mathcal{S}_{J_0},$$

where $\mathcal{D}_j$ is a time series reflecting variations in $\mathbf{X}$ on scale τ_j, while $\mathcal{S}_{J_0}$ is a series reflecting its λ_{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 $\mathbf{X}$:

$$\begin{aligned} \hat{\sigma}_X^2 &= \frac{1}{N} \sum_{t=0}^{N-1} \left(X_t - \overline{X}\right)^2 \\ &= \frac{1}{N} \sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \frac{1}{N} \|\mathbf{V}_{J_0}\|^2 - \overline{X}^2 \end{aligned}$$