Wavelet Methods for Time Series Analysis

Part III: Basic Theory for Discrete Wavelet Transform (DWT)

- precise definition of DWT requires a few basic concepts from Fourier analysis and theory of linear filters
- will start with discussion/review of:
  - convolution/filtering of infinite sequences
  - filter cascades
  - Fourier theory for finite sequences
  - circular convolution/filtering of finite sequences
  - periodization of a filter

Basic Concepts of Filtering: I

- convolution & linear time invariant filtering are same concepts:
  - $\{b_t : t = \ldots, -1, 0, 1, \ldots\} = \{b_t\}$ is the input to the filter
  - $\{a_t\}$ represents the filter
  - $\{c_t\}$ is the output from the filter
- flow diagram for filtering:

$$\{b_t\} \longrightarrow \boxed{\{a_t\}} \longrightarrow \{c_t\} \text{ or } \{b_t\} \longrightarrow \boxed{a_t} \longrightarrow \{c_t\}$$

- $\{a_t\}$ is called the impulse response sequence for the filter
- the discrete Fourier transform (DFT) of $\{a_t\}$ is given by

$$A(f) \equiv \sum_{t=-\infty}^{\infty} a_t e^{-i2\pi ft}$$

  and is known as the transfer function for the filter

Basic Concepts of Filtering: II

- the inverse DFT allows us to recover $\{a_t\}$ from $A(\cdot)$:

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

- since $A(-f) = A^*(f)$ (the complex conjugate of $A(f)$) when $\{a_t\}$ is real-valued, we really just need to consider nonnegative Fourier frequencies f for studying $A(\cdot)$
- $\{a_t\}$ and $A(\cdot)$ form a Fourier transform pair, a fact that is denoted by $\{a_t\} \longleftrightarrow A(\cdot)$ or, less formally, by $a_t \longleftrightarrow A(f)$
- in general $A(\cdot)$ is complex-valued, so write $A(f) = |A(f)|e^{i\theta(f)}$
  - $|A(f)|$ defines gain function
  - $\mathcal{A}(f) \equiv |A(f)|^2$ defines squared gain function
  - $\theta(f)$ called phase function (well-defined at f if $|A(f)| > 0$)

Basic Concepts of Filtering: III

- given $\{a_t\} \longleftrightarrow A(\cdot)$ and $\{b_t\} \longleftrightarrow B(\cdot)$, their convolution

$$c_t \equiv \sum_{u=-\infty}^{\infty} a_u b_{t-u}, \quad t = \ldots, -1, 0, 1, \ldots,$$

  has a DFT given by

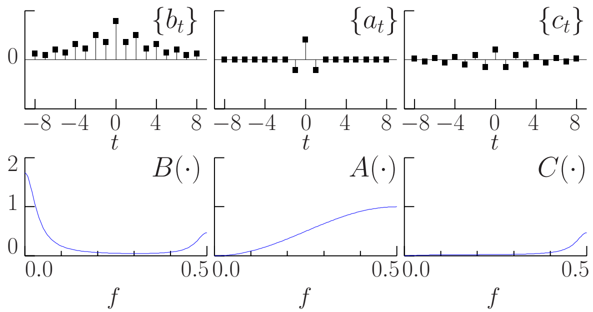
$$C(f) \equiv \sum_{t=-\infty}^{\infty} c_t e^{-i2\pi ft} = A(f)B(f)$$

- $\{c_t\}$ is output from filter with impulse response sequence $\{a_t\}$ and transfer function $A(\cdot)$ related by $\{a_t\} \longleftrightarrow A(\cdot)$
- since $A(\cdot)$ is equivalent to $\{a_t\}$, can express flow diagram as

  either $\{b_t\} \longrightarrow \boxed{\{a_t\}} \longrightarrow \{c_t\}$ or $\{b_t\} \longrightarrow \boxed{A(\cdot)} \longrightarrow \{c_t\}$

Example of a High-Pass Filter

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



Cascade of Filters: I

- idea: output from one filter becomes input to another
- flow diagram for cascade with 2 filters (can have more!):

$$\{b_t\} \longrightarrow \boxed{A_1(\cdot)} \xrightarrow{1.} \boxed{A_2(\cdot)} \xrightarrow{2.} \{c_t\}$$

if $\{b_t\} \longleftrightarrow B(\cdot)$ and $\{c_t\} \longleftrightarrow C(\cdot)$, then

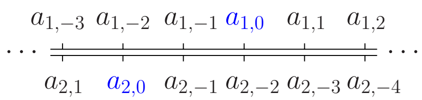
1. output from $A_1(\cdot)$ has DFT $A_1(f)B(f)$
2. output from $A_2(\cdot)$ has DFT $A_2(f)A_1(f)B(f)$
   so $C(f) = A_2(f)A_1(f)B(f)$

- let $A(f) \equiv A_2(f)A_1(f)$
- can reexpress overall effect of filter cascade as

$$\{b_t\} \longrightarrow \boxed{A(\cdot)} \longrightarrow \{c_t\}$$

Cascade of Filters: II

- $A(\cdot)$ is transfer function for equivalent filter for cascade
- let $\{a_t\} \longleftrightarrow A(\cdot)$, $\{a_{1,t}\} \longleftrightarrow A_1(\cdot)$ and $\{a_{2,t}\} \longleftrightarrow A_2(\cdot)$
- to form $\{a_t\}$, just need to convolve $\{a_{1,t}\}$ and $\{a_{2,t}\}$ (reverse one filter, multiply by other; shift and repeat)
- example: $a_{1,t} = \begin{cases} -\frac{1}{2}, & t = -1 \\ \frac{1}{2}, & t = 0 \\ 0, & \text{otherwise} \end{cases}$ & $a_{2,t} = \begin{cases} \frac{1}{2}, & t = 0 \\ -\frac{1}{2}, & t = 1 \\ 0, & \text{otherwise} \end{cases}$



$a_{-2} = \sum_{u=-\infty}^{\infty} a_{1,u} a_{2,-2-u}$

Fourier Theory for Finite Sequences

- $\{a_t : t = 0, 1, \ldots, N-1\} = \{a_t\}$ has DFT

$$A_k \equiv \sum_{t=0}^{N-1} a_t e^{-i2\pi f_k t}, \quad \text{with } f_k \equiv \frac{k}{N} \;\&\; k = 0, 1, \ldots, N-1$$

- inverse DFT says that

$$a_t = \frac{1}{N} \sum_{k=0}^{N-1} A_k e^{i2\pi f_k t}, \quad t = 0, 1, \ldots, N-1$$

- relationship between $\{a_t\}$ and $A(\cdot)$ denoted by

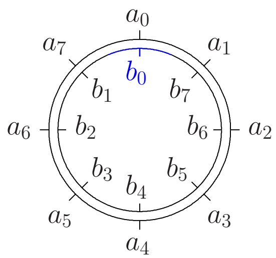
$$\{a_t\} \longleftrightarrow \{A_k\} \text{ or, less formally, by } a_t \longleftrightarrow A_k$$

Convolution/Filtering of Finite Sequences: I

- given $\{a_t\}$ & $\{b_t\}$ of length N with DFTs $\{A_k\}$ & $\{B_k\}$, their convolution is defined to be
$$c_t \equiv \sum_{u=0}^{N-1} a_u b_{t-u \bmod N}, \quad t = 0, 1, \ldots, N-1,$$
where $k \bmod N \equiv k$ if $0 \le k \le N-1$;
if not, $k \bmod N \equiv k + nN$, where n is unique integer yielding $0 \le k + nN \le N-1$; thus
$b_{0 \bmod N} = b_0$, $b_{-1 \bmod N} = b_{N-1}$, $b_{-2 \bmod N} = b_{N-2}$ etc
- $\{c_t\}$ is output from circular filtering operation expressible as
$$\{b_t\} \longrightarrow \boxed{a_t} \longrightarrow \{c_t\} \text{ or } \{b_t\} \longrightarrow \boxed{A_k} \longrightarrow \{c_t\}$$

Convolution/Filtering of Finite Sequences: II

- sequence $\{c_t\}$ is called a circular (cyclic) convolution:



$c_0 = \sum_{u=0}^{7} a_u b_{-u \bmod 8}$

- DFT $\{C_k\}$ of $\{c_t\}$ again has a simple form, namely,
$$C_k = \sum_{t=0}^{N-1} c_t e^{-i2\pi f_k t} = A_k B_k;$$
i.e., $\{c_t\} \longleftrightarrow \{A_k B_k\}$

Convolution/Filtering of Finite Sequences: III

- suppose $\{a_t\}$ has width M with $a_t = 0$ for $t < 0$ and $t \ge M$
- given $\{b_t\}$ of length N, can reexpress
$$c_t = \sum_{u=0}^{M-1} a_u b_{t-u \bmod N}, \quad t = 0, \ldots, N-1,$$
as a circular convolution using a 'periodized' filter of length N:
$$c_t = \sum_{u=0}^{N-1} a_u^{\circ} b_{t-u \bmod N}, \text{ where } a_u^{\circ} \equiv \sum_{n=-\infty}^{\infty} a_{u+nN}, \quad u = 0, \ldots, N-1$$
- DFT of $\{a_t^{\circ}\}$ given by $A(\frac{k}{N})$, $k = 0, \ldots, N-1$, where
$$A(f) \equiv \sum_{t=-\infty}^{\infty} a_t e^{-i2\pi f t} = \sum_{t=0}^{M-1} a_t e^{-i2\pi f t}$$

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

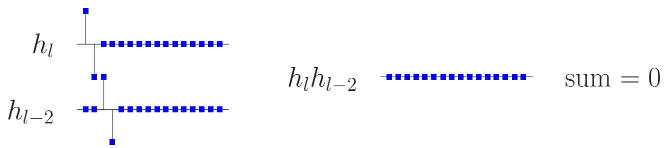
- 2 and 3 together are called the orthonormality property

The Wavelet Filter: III

- 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
- orthogonal to even shifts
  orthogonality to even shifts also readily apparent

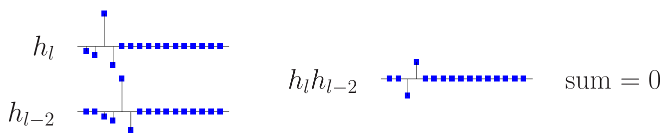


D(4) Wavelet Filter: I

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

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

  - 'D' stands for Daubechies
  - $L = 4$ width member of her 'extremal phase' wavelets
- computations show $\sum_l h_l = 0$ & $\sum_l h_l^2 = 1$, as required
- orthogonal to even shifts 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:

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

- 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:

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

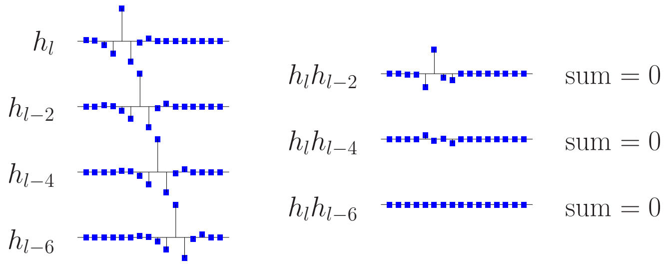
D(4) Wavelet Filter: IV

- 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
- one solution is $a = \frac{1+\sqrt{3}}{4\sqrt{2}} \doteq 0.48$ and $b = \frac{-1+\sqrt{3}}{4\sqrt{2}} \doteq 0.13$ (3 other solutions, but these yield essentially the same filter)
- interpret D(4) filtered output as changes in weighted averages
  - 'change' now measured by 2nd difference (1st for Haar)
  - average is now 2 point weighted average (1 point for Haar)
  - can argue that effective scale of weighted average is one

Another Wavelet Filter

- LA(8) wavelet filter ('LA' stands for 'least asymmetric')



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 $h^\circ_0, h^\circ_1, \ldots, h^\circ_{N-1}$
- circularly filter $\mathbf{X}$ using $\{h^\circ_l\}$ to yield output

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

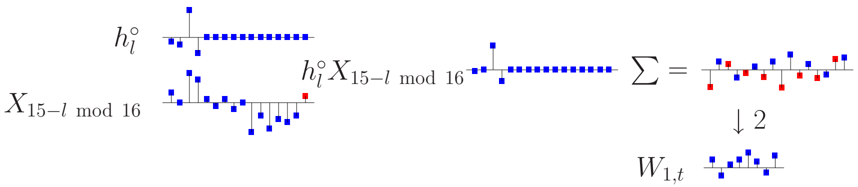
- starting with $t = 1$, take every other value of output to define

$$W_{1,t} \equiv \sum_{l=0}^{N-1} h^\circ_l X_{2t+1-l \bmod N}, \quad t = 0, \ldots, \tfrac{N}{2} - 1;$$

  $\{W_{1,t}\}$ 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 $h_0 h_2 + h_1 h_3 = 0$
- note: $\langle \mathcal{W}_{0\bullet}, \mathbf{X} \rangle$ yields $W_0 = h_1 X_0 + h_0 X_1 + h_3 X_{14} + h_2 X_{15}$
- unlike other coefficients from above, this 'boundary' coefficient depends on circular treatment of $\mathbf{X}$ (a curse, not a feature!)

Upper Half of DWT Matrix: III

- if $L \le N$, 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)

Upper Half of DWT Matrix: IV

- can argue that, for all L and even N,

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

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

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

- Q: how can we construct the other half of $\mathcal{W}$?

The Scaling Filter

- scaling filter defined by $g_l \equiv (-1)^{l+1} h_{L-1-l}$
- $\{g_l\}$ 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

- only orthonormality property of $\{h_l\}$ 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

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

First Level Scaling Coefficients: II

- 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 $\mathcal{W}_1$:

similar to $\mathcal{W}_1 \mathcal{W}_1^T = I_{\frac{N}{2}}$, 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?
- can show scaling filter obeys important fourth property

4. orthogonality to $\{h_l\}$ and its even shifts: for all n have

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

- implies any row in $\mathcal{V}_1$ orthogonal to any row in $\mathcal{W}_1$
- implies $\mathcal{W}_1$ & $\mathcal{V}_1$ are jointly orthonormal:

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

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

- 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 much interest!), in fact $\mathcal{P}_1 = \mathcal{W}$
- if $N > 2$, $\mathcal{P}_1$ is intermediate step on way to $\mathcal{W}$
  - $\mathcal{V}_1$ spans same subspace as lower half of $\mathcal{W}$
  - $\mathcal{P}_1$ can be of interest by itself (just needs N even)

Three Comments

- if N even, then $\mathcal{P}_1$ is well-defined (don't need $N = 2^J$)
- rather than defining $g_l = (-1)^{l+1} h_{L-1-l}$, could use alternative definition $g_l = (-1)^{l-1} h_{1-l}$
  - 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

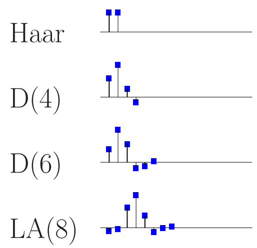
$$\begin{bmatrix} g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 &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

- 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

- here are plots of $\{g_l\}$ corresponding to four different wavelet filters:



- 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 $\{g_l\}$: I

- since $\mathbf{W}_1$ and $\mathbf{V}_1$ contain (downsampled) output from filters, let's consider frequency domain properties of $\{h_l\}$ & $\{g_l\}$
- define transfer and squared gain functions for wavelet filter:

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

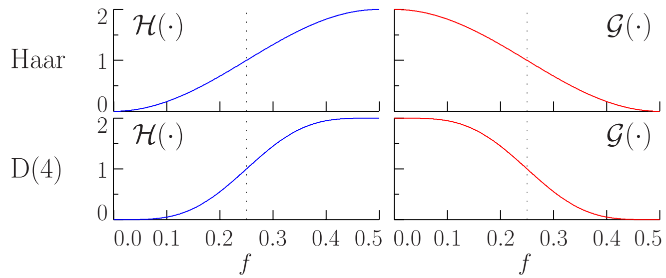
- define similar functions for scaling filter:

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

- effect of $\{h_l\}$ & $\{g_l\}$ on $\mathbf{X}$ can be deduced from $\mathcal{H}(\cdot)$ & $\mathcal{G}(\cdot)$

Frequency Domain Properties of $\{h_l\}$ and $\{g_l\}$: II

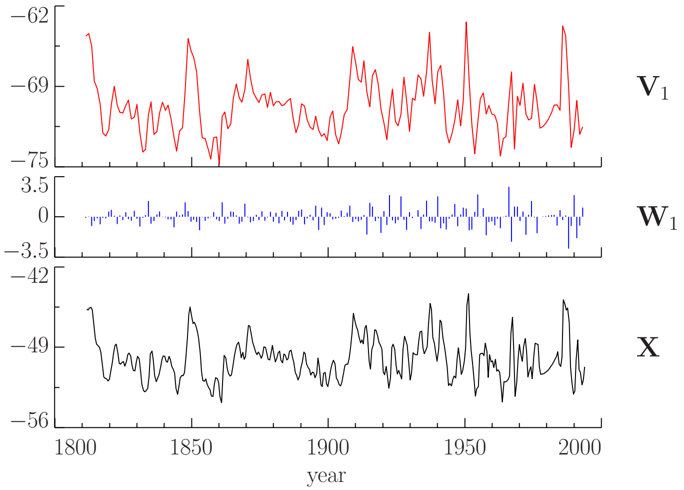
- example: $\mathcal{H}(\cdot)$ and $\mathcal{G}(\cdot)$ for Haar & D(4) filters



- $\{h_l\}$ is high-pass filter with nominal pass-band $[1/4, 1/2]$
- $\{g_l\}$ is low-pass filter with nominal pass-band $[0, 1/4]$
- same true for all Daubechies wavelet and scaling filters
- orthonormality condition equivalent to $\mathcal{H}(f) + \mathcal{G}(f) = 2$

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

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



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

- 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 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 $\frac{N}{2} \times 1$

- $\mathcal{D}_1$ *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 $N \times 1$:

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

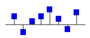
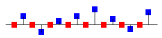
- can now regard the above as equivalent to use of a filter

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

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

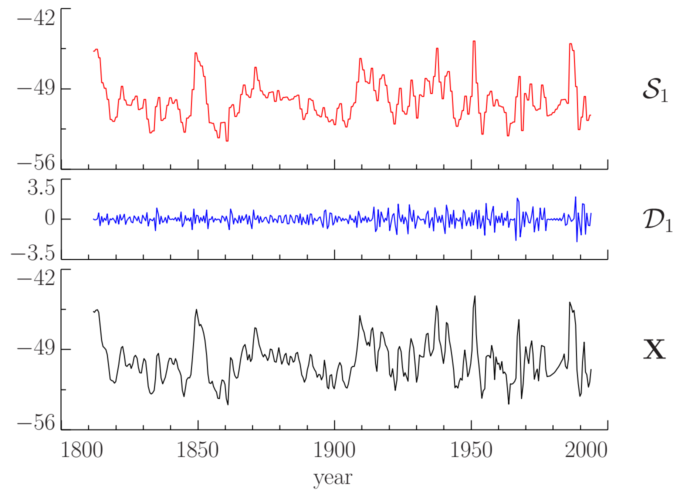
- doesn't look like exactly like filtering, which would look like

$$\sum_{l=0}^{N-1} h^{\circ}_l W^{\uparrow}_{1,t-l \bmod N};$$ 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 $\{H^*(\frac{k}{N})\}$

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

- Haar-based decomposition for oxygen isotope records $\mathbf{X}$



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

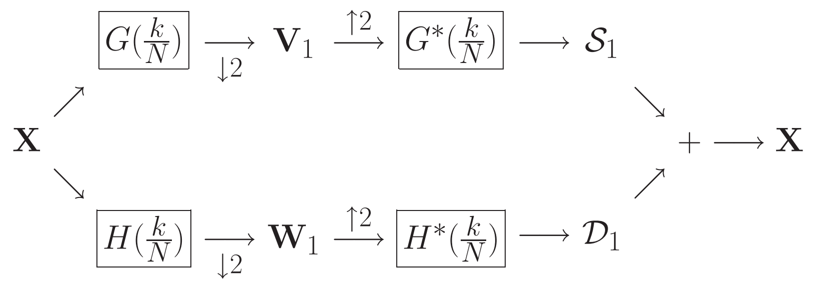
- breaks up $\hat{\sigma}_X^2$ into two pieces:
  1. $\frac{1}{N}\|\mathbf{W}_1\|^2$, attributable to changes in averages over scale 1
  2. $\frac{1}{N}\|\mathbf{V}_1\|^2 - \overline{X}^2$, attributable to averages over scale 2
- Haar-based example for oxygen isotope records
  - first piece: $\frac{1}{N}\|\mathbf{W}_1\|^2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

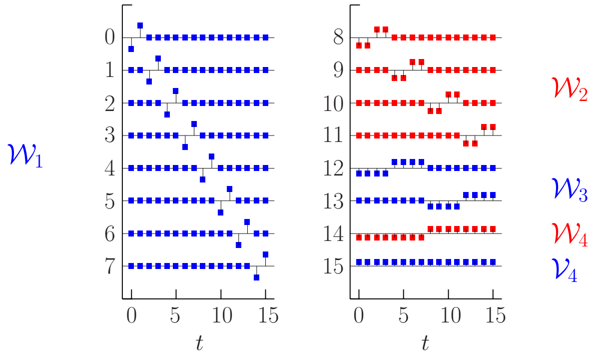
- continuing in this manner defines remaining subvectors of $\mathbf{W}$ (recall that $\mathbf{W} = \mathcal{W}\mathbf{X}$ is the vector of DWT coefficients)
- at each level j, outputs $\mathbf{W}_j$ and $\mathbf{V}_j$ from the basic algorithm are each half the length of the input $\mathbf{V}_{j-1}$
- 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 *defines* the remaining subvectors $\mathbf{W}_2, \ldots, \mathbf{W}_J, \mathbf{V}_J$ of DWT coefficient vector $\mathbf{W}$
- overall scheme is known as the 'pyramid' algorithm

Scales Associated with DWT Coefficients

- 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 $\lambda_j \, \Delta t$

Examples of $\mathcal{W}$ and its Partitioning: I

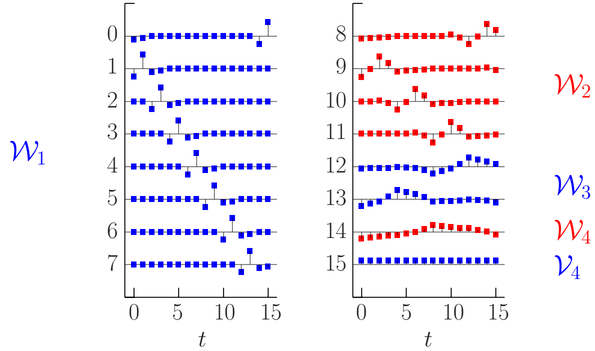
- $N = 16$ case for Haar DWT matrix $\mathcal{W}$



- 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

Multiresolution Analysis

- $\mathcal{D}_j \equiv \mathcal{W}_j^T \mathbf{W}_j$ is the jth level detail
- $\mathcal{S}_j \equiv \mathcal{V}_j^T \mathbf{V}_j$ is the jth level 'smooth'
- we get multiresolution analyses (MRAs) for levels k and J: for $1 \leq k \leq J$,

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

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

- stop at $J_0 < J$ repetitions — a level J_0 'partial' DWT
- only requires N to be integer multiple of 2^{J_0}
- choice of J_0 is application dependent
- multiresolution analysis for partial DWT:

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

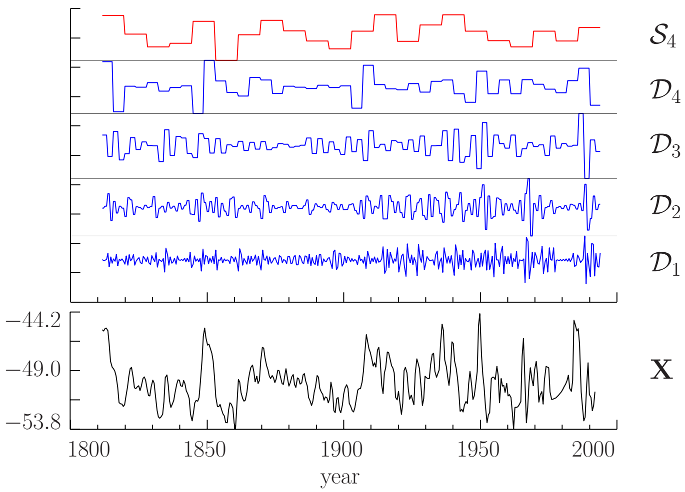
$\mathcal{S}_{J_0}$ represents averages on scale $\lambda_{J_0} = 2^{J_0}$ (includes $\overline{X}$)

- 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 MRA from $J_0 = 4$ Partial Haar DWT

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



Assigning Times to Wavelet Coefficients

- LA class of wavelet and scaling filters designed to exhibit 'near symmetry' about some point in the filter
- makes it easier to align $W_{j,t}$ and $V_{J_0,t}$ with values in $\mathbf{X}$
- some gory details: if X_t is associated with actual time $t_0 + t\,\Delta t$, LA wavelet coefficient $W_{j,t}$ should be plotted at time

$$t_0 + (2^j(t+1) - 1 - |\nu_j^{(H)}| \bmod N)\,\Delta t$$

e.g., $|\nu_j^{(H)}| = [7(2^j - 1) + 1]/2$ for LA(8) wavelet. For $N = 16$

coefficient	$W_{1,0}$	$W_{1,1}$	$W_{1,2}$	$W_{1,3}$	$W_{1,4}$	$W_{1,5}$	$W_{1,6}$	$W_{1,7}$
associated time	13	15	1	3	5	7	9	11

- order in which elements of $\mathbf{W}_1$ should be displayed is thus

$$W_{1,2}, W_{1,3}, W_{1,4}, W_{1,5}, W_{1,6}, W_{1,7}, W_{1,0}, W_{1,1}$$

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:

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

- define $\mathcal{T}^{-2} = \mathcal{T}^{-1}\mathcal{T}^{-1}$, $\mathcal{T}^{-3} = \mathcal{T}^{-1}\mathcal{T}^{-1}\mathcal{T}^{-1}$ and so forth

Circularly Shifting a Vector and Time Alignment

- can express reordering elements of

$$\mathbf{W}_1 = [W_{1,0}, W_{1,1}, W_{1,2}, W_{1,3}, W_{1,4}, W_{1,5}, W_{1,6}, W_{1,7}]^T$$

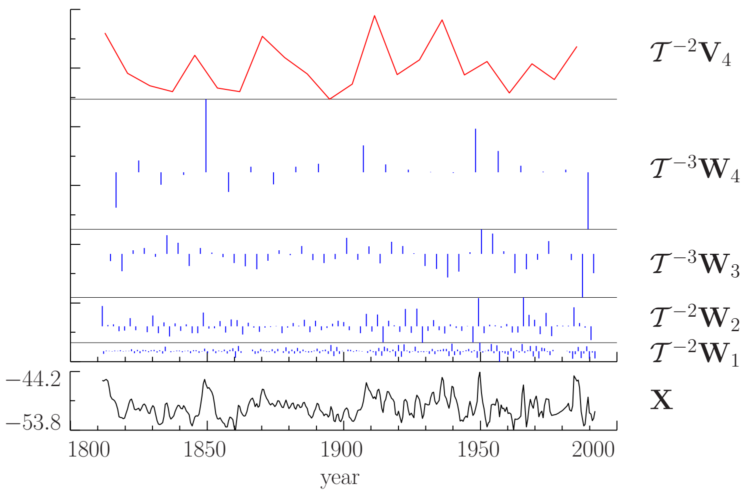
as they occur in time using

$$\mathcal{T}^{-2}\mathbf{W}_1 = [W_{1,2}, W_{1,3}, W_{1,4}, W_{1,5}, W_{1,6}, W_{1,7}, W_{1,0}, W_{1,1}]^T$$

- can use to time-align wavelet coefficients
- note that the details and smooths do not need to be time-aligned as the associated filters do not cause a time shift

Example of $J_0 = 4$ Partial LA(8) DWT

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



Summary of Key Points about the DWT: I

- the DWT $\mathcal{W}$ is orthonormal, i.e., satisfies $\mathcal{W}^T\mathcal{W} = I_N$
- construction of $\mathcal{W}$ starts with a wavelet filter $\{h_l\}$ of even length L that by definition
  1. sums to zero; i.e., $\sum_l h_l = 0$;
  2. has unit energy; i.e., $\sum_l h_l^2 = 1$; and
  3. is orthogonal to its even shifts; i.e., $\sum_l h_l h_{l+2n} = 0$
- 2 and 3 together called orthonormality property
- wavelet filter defines a scaling filter via $g_l = (-1)^{l+1} h_{L-1-l}$
- scaling filter satisfies the orthonormality property, but sums to $\sqrt{2}$ and is also orthogonal to $\{h_l\}$; i.e., $\sum_l g_l h_{l+2n} = 0$
- while $\{h_l\}$ is a high-pass filter, $\{g_l\}$ is a low-pass filter

Summary of Key Points about the DWT: II

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

Summary of Key Points about the DWT: III

- after J_0 repetitions, this 'pyramid' algorithm transforms time series $\mathbf{X}$ whose length N is an integer multiple of 2^{J_0} into DWT coefficients $\mathbf{W}_1$, $\mathbf{W}_2$, $\ldots$, $\mathbf{W}_{J_0}$ and $\mathbf{V}_{J_0}$ (sizes of vectors are $\frac{N}{2}$, $\frac{N}{4}$, $\ldots$, $\frac{N}{2^{J_0}}$ 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 $\mathbf{X}$:

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