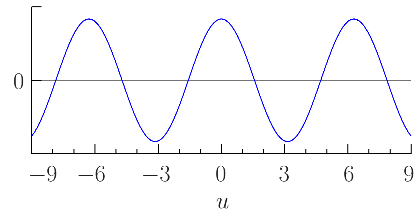
Wavelet Methods for Time Series Analysis

Part I: Introduction to Wavelets and Wavelet Transforms

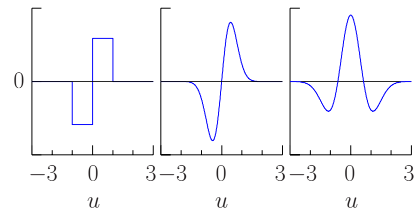
- wavelets are analysis tools for time series and images
- as a subject, wavelets are
  - relatively new (1983 to present)
  - a synthesis of old/new ideas
  - keyword in 29, 826+ articles and books since 1989 (4032 more since 2005: an inundation of material!!!)
- broadly speaking, there have been two waves of wavelets
  - continuous wavelet transform (1983 and on)
  - discrete wavelet transform (1988 and on)
- will introduce subject via CWT & then concentrate on DWT

What is a Wavelet?

- sines & cosines are 'big waves'

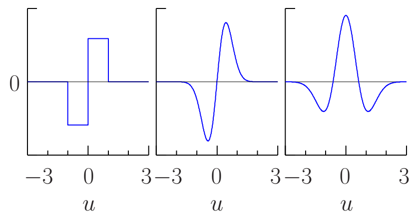


- wavelets are 'small waves' (left-hand is Haar wavelet $\psi^{(\mathrm{H})}(\cdot)$)



Technical Definition of a Wavelet: I

- real-valued function $\psi(\cdot)$ defined over real axis is a wavelet if
  1. integral of $\psi^2(\cdot)$ is unity: $\int_{-\infty}^{\infty} \psi^2(u)\,du = 1$ (called 'unit energy' property, with apologies to physicists)
  2. integral of $\psi(\cdot)$ is zero: $\int_{-\infty}^{\infty} \psi(u)\,du = 0$ (technically, need an 'admissibility condition,' but this is almost equivalent to integration to zero)

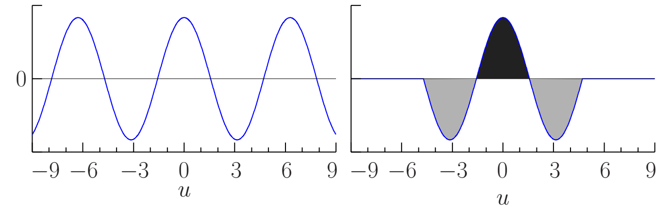


Technical Definition of a Wavelet: II

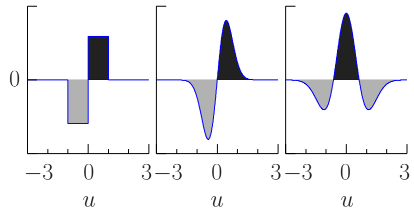
- $\int_{-\infty}^{\infty} \psi^2(u)\,du = 1$ & $\int_{-\infty}^{\infty} \psi(u)\,du = 0$ give a wavelet because:
  - by property 1, for every small $\epsilon > 0$, have
  $$\int_{-\infty}^{-T} \psi^2(u)\,du + \int_{T}^{\infty} \psi^2(u)\,du < \epsilon$$
  for some finite T
  - 'business' part of $\psi(\cdot)$ is over interval $[-T, T]$
  - width $2T$ of $[-T, T]$ might be huge, but will be insignificant compared to $(-\infty, \infty)$
  - by property 2, $\psi(\cdot)$ is balanced above/below horizontal axis
- matches intuitive notion of a 'small' wave

Two Non-Wavelets and Three Wavelets

- two failures: $f(u) = \cos(u)$ & same limited to $[-3\pi/2, 3\pi/2]$:



- Haar wavelet $\psi^{(H)}(\cdot)$ and two of its friends:



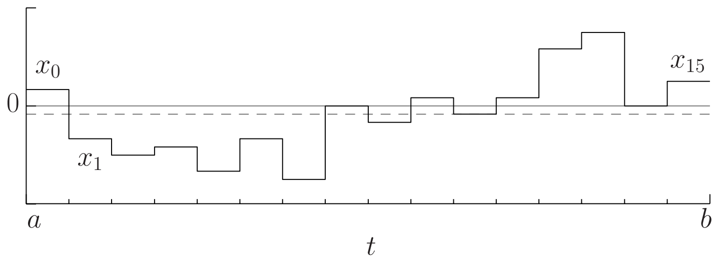
What is Wavelet Analysis?

- wavelets tell us about variations in local averages
- to quantify this description, let $x(\cdot)$ be a 'signal'
  - real-valued function of t defined over real axis
  - will refer to t as time (but it need not be such)
- consider 'average value' of $x(\cdot)$ over $[a, b]$:

$$\frac{1}{b-a}\int_a^b x(t)\,dt$$

Example of Average Value of a Signal

- let $x(\cdot)$ be step function taking on values $x_0, x_1, \ldots, x_{15}$ over 16 equal subintervals of $[a, b]$:



- here we have

$$\frac{1}{b-a}\int_a^b x(t)\,dt = \frac{1}{16}\sum_{j=0}^{15} x_j = \text{ height of dashed line}$$

Average Values at Different Scales and Times

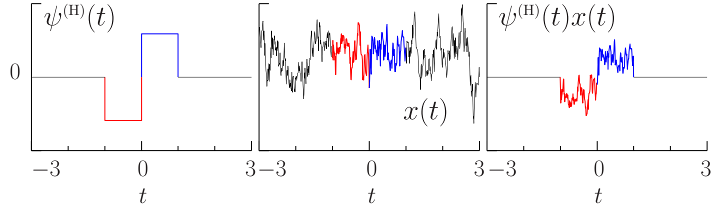
- define the following function of λ and t

$$A(\lambda, t) \equiv \frac{1}{\lambda}\int_{t-\frac{\lambda}{2}}^{t+\frac{\lambda}{2}} x(u)\,du$$

  - λ is width of interval – refered to as 'scale'
  - t is midpoint of interval
- $A(\lambda, t)$ is average value of $x(\cdot)$ over scale λ centered at t
- average values of signals have wide-spread interest
  - one second average temperatures over forest
  - ten minute rainfall rate during severe storm
  - yearly average temperatures over central England

Defining a Wavelet Coefficient W

- multiply Haar wavelet & time series $x(\cdot)$ together:



- integrate resulting function to get 'wavelet coefficient' $W(1,0)$:

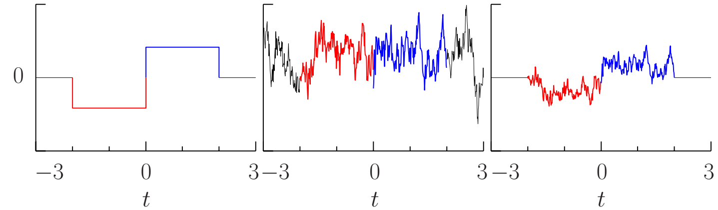
$$\int_{-\infty}^{\infty} \psi^{(H)}(t)x(t)\,dt = W(1,0)$$

- to see what $W(1,0)$ is telling us about $x(\cdot)$, note that

$$W(1,0) \propto \frac{1}{1}\int_0^1 x(t)\,dt - \frac{1}{1}\int_{-1}^0 x(t)\,dt = A(1,\tfrac{1}{2}) - A(1,-\tfrac{1}{2})$$

Defining Wavelet Coefficients for Other Scales

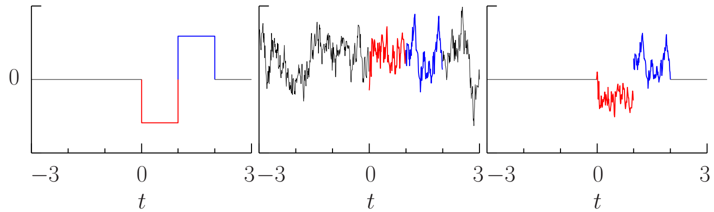
- $W(1,0)$ proportional to difference between averages of $x(\cdot)$ over $[-1,0]$ & $[0,1]$, i.e., two unit scale averages before/after $t=0$
  - '1' in $W(1,0)$ denotes scale 1 (width of each interval)
  - '0' in $W(1,0)$ denotes time 0 (center of combined intervals)
- stretch or shrink wavelet to define $W(\tau,0)$ for other scales τ:



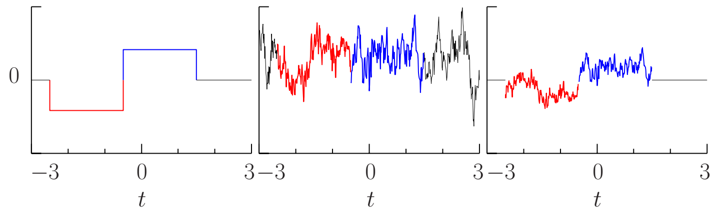
yields $W(2,0)$

Defining Wavelet Coefficients for Other Locations

- relocate to define $W(\tau,t)$ for other times t:



yields $W(1,1)$



yields $W(2,-\frac{1}{2})$

Haar Continuous Wavelet Transform (CWT)

- for all $\tau > 0$ and all $-\infty < t < \infty$, can write

$$W(\tau,t) = \frac{1}{\sqrt{\tau}}\int_{-\infty}^{\infty} x(u)\psi^{(H)}\left(\frac{u-t}{\tau}\right) du$$

  - $\frac{u-t}{\tau}$ does the stretching/shrinking and relocating
  - $\frac{1}{\sqrt{\tau}}$ needed so $\psi^{(H)}_{\tau,t}(u) \equiv \frac{1}{\sqrt{\tau}}\psi^{(H)}\left(\frac{u-t}{\tau}\right)$ has unit energy
  - since it also integrates to zero, $\psi^{(H)}_{\tau,t}(\cdot)$ is a wavelet
- $W(\tau,t)$ over all $\tau > 0$ and all t is Haar CWT for $x(\cdot)$
- analyzes/breaks up/decomposes $x(\cdot)$ into components
  - associated with a scale and a time
  - physically related to a difference of averages

Other Continuous Wavelet Transforms: I

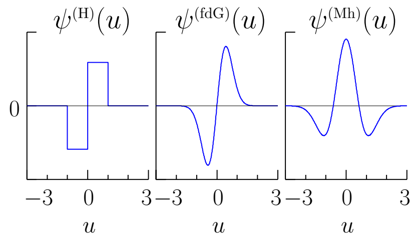
- can do the same for wavelets other than the Haar
- start with basic wavelet $\psi(\cdot)$
- use $\psi_{\tau,t}(u) = \frac{1}{\sqrt{\tau}}\psi\left(\frac{u-t}{\tau}\right)$ to stretch/shrink & relocate
- define CWT via

$$W(\tau,t) = \int_{-\infty}^{\infty} x(u)\psi_{\tau,t}(u)\,du = \frac{1}{\sqrt{\tau}}\int_{-\infty}^{\infty} x(u)\psi\left(\frac{u-t}{\tau}\right)\,du$$

- analyzes/breaks up/decomposes $x(\cdot)$ into components
  - associated with a scale and a time
  - physically related to a difference of *weighted* averages

Other Continuous Wavelet Transforms: II

- consider two friends of Haar wavelet



- $\psi^{(fdG)}(\cdot)$ proportional to 1st derivative of Gaussian PDF
- 'Mexican hat' wavelet $\psi^{(Mh)}(\cdot)$ proportional to 2nd derivative
- $\psi^{(fdG)}(\cdot)$ looks at difference of adjacent weighted averages
- $\psi^{(Mh)}(\cdot)$ looks at difference between weighted average and sum of weighted averages occurring before & after

First Scary-Looking Equation

- CWT equivalent to $x(\cdot)$ because we can write

$$x(t) = \int_0^{\infty}\left[\frac{1}{C\tau^2}\int_{-\infty}^{\infty} W(\tau,u)\frac{1}{\sqrt{\tau}}\psi\left(\frac{t-u}{\tau}\right)\,du\right]\,d\tau,$$

  where C is a constant depending on specific wavelet $\psi(\cdot)$
- can synthesize (put back together) $x(\cdot)$ given its CWT; i.e., nothing is lost in reexpressing signal $x(\cdot)$ via its CWT
- regard stuff in brackets as defining 'scale τ' signal at time t
- says we can reexpress $x(\cdot)$ as integral (sum) of new signals, each associated with a particular scale
- similar additive decompositions will be one central theme

Second Scary-Looking Equation

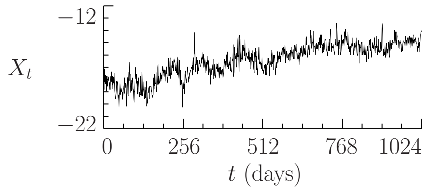
- energy in $x(\cdot)$ is reexpressed in CWT because

$$\text{energy} = \int_{-\infty}^{\infty} x^2(t)\,dt = \int_0^{\infty}\left[\frac{1}{C\tau^2}\int_{-\infty}^{\infty} W^2(\tau,t)\,dt\right]\,d\tau$$

- can regard $x^2(t)$ versus t as breaking up the energy across time (i.e., an 'energy density' function)
- regard stuff in brackets as breaking up the energy across scales
- says we can reexpress energy as integral (sum) of components, each associated with a particular scale
- function defined by $W^2(\tau,t)/C\tau^2$ is an energy density across both time and scale
- similar energy decompositions will be a second central theme

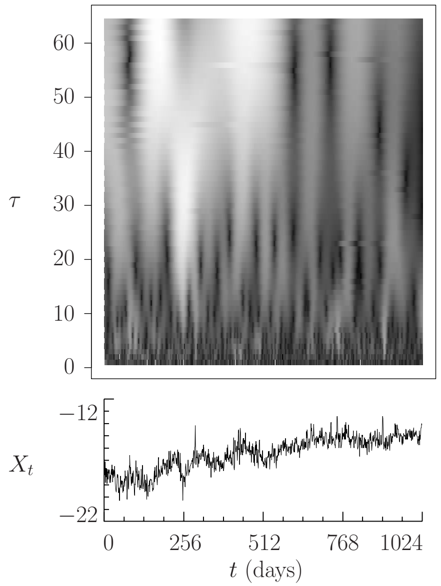
Example: Atomic Clock Data

- example: average daily frequency variations in clock 571

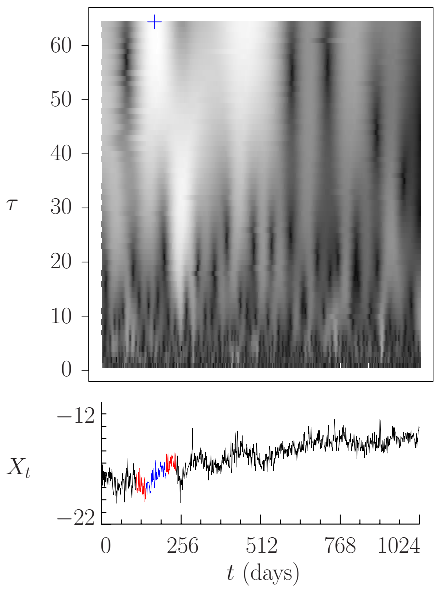


- t is measured in days (one measurment per day)
- plot shows X_t versus integer t
- $X_t = 0$ for all t would say that clock 571 keeps time perfectly
- $X_t < 0$ implies that clock is losing time systematically
- can easily adjust clock if X_t were constant
- inherent quality of clock related to changes in averages of X_t

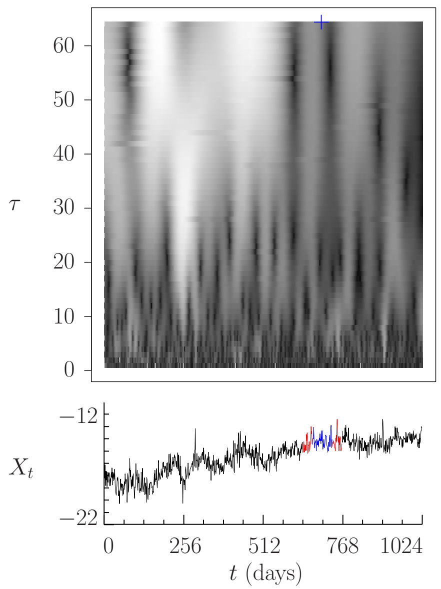
Mexican Hat CWT of Clock Data: I



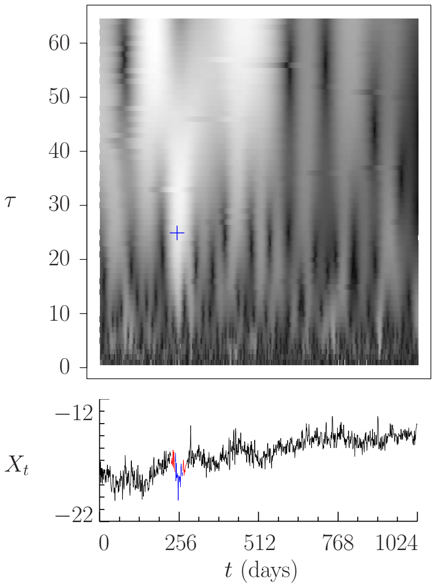
Mexican Hat CWT of Clock Data: II



Mexican Hat CWT of Clock Data: III

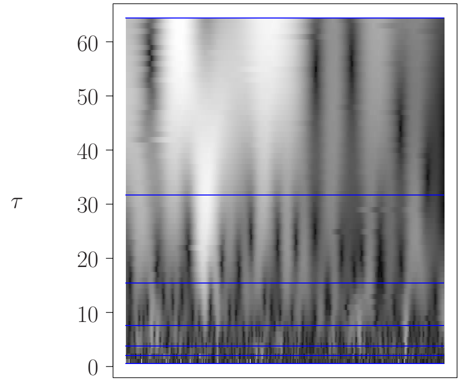


Mexican Hat CWT of Clock Data: IV



Beyond the CWT: the DWT

- can often get by with subsamples of $W(\tau, t)$
- leads to notion of discrete wavelet transform (DWT) (can regard as discretized 'slices' through CWT)



Rationale for the DWT

- DWT has appeal in its own right
  - most time series are sampled as discrete values (can be tricky to implement CWT)
  - can formulate as orthonormal transform (makes meaningful statistical analysis possible)
  - tends to decorrelate certain time series
  - standardization to dyadic scales often adequate
  - generalizes to notion of wavelet packets
  - can be faster than the fast Fourier transform
- will concentrate primarily on DWT for remainder of course

Qualitative Description of DWT

- will give precise definition of DWT in Part II
- let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ be a vector of N time series values (note: 'T' denotes transpose; i.e., $\mathbf{X}$ is a column vector)
- need to assume $N = 2^J$ for some positive integer J (restrictive!)
- DWT is a linear transform of $\mathbf{X}$ yielding N DWT coefficients
- notation: $\mathbf{W} = \mathcal{W}\mathbf{X}$
  - $\mathbf{W}$ is vector of DWT coefficients (jth component is W_j)
  - $\mathcal{W}$ is $N \times N$ orthonormal transform matrix; i.e., $\mathcal{W}^T\mathcal{W} = I_N$, where I_N is $N \times N$ identity matrix
- inverse of $\mathcal{W}$ is just its transpose, so $\mathcal{W}\mathcal{W}^T = I_N$ also

Implications of Orthonormality: I

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

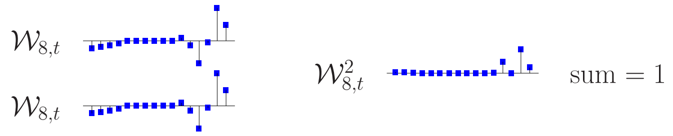
$$\langle \mathcal{W}_{j\bullet}, \mathcal{W}_{k\bullet} \rangle \equiv \sum_{l=0}^{N-1} \mathcal{W}_{j,l}\mathcal{W}_{k,l} = \begin{cases} 1, & \text{when } j = k, \\ 0, & \text{when } j \neq k \end{cases}$$

  – $\langle \mathcal{W}_{j\bullet}, \mathcal{W}_{k\bullet} \rangle$ is inner product of jth & kth rows

  – $\|\mathcal{W}_{j\bullet}\|^2 \equiv \langle \mathcal{W}_{j\bullet}, \mathcal{W}_{j\bullet} \rangle$ is squared norm (energy) for $\mathcal{W}_{j\bullet}$

Implications of Orthonormality: II

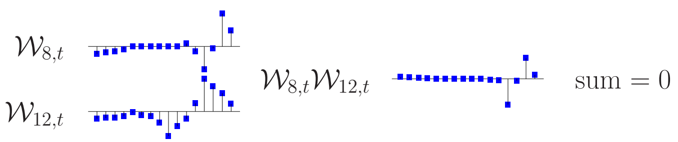
- example from $\mathcal{W}$ of dimension 16×16 we'll see later on
  – inner product of row 8 with itself (i.e., squared norm):



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

Implications of Orthonormality: III

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

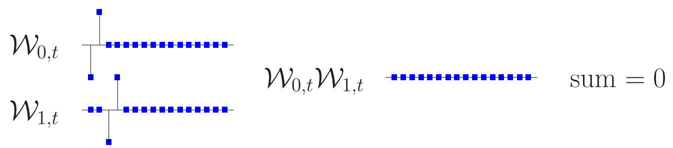


  – rows 8 & 12 said to be orthogonal since inner product is 0

The Haar DWT: I

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

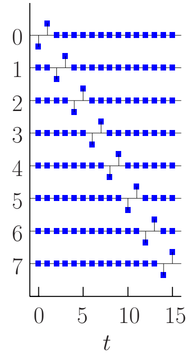
  note: $\|\mathcal{W}_{0\bullet}\|^2 = \frac{1}{2} + \frac{1}{2} = 1$ & hence has required unit energy
- row $j = 1$: $\left[0, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \underbrace{0, \ldots, 0}_{N-4 \text{ zeros}}\right] \equiv \mathcal{W}_{1\bullet}^T$
- $\mathcal{W}_{0\bullet}$ and $\mathcal{W}_{1\bullet}$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



The Haar DWT: II

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

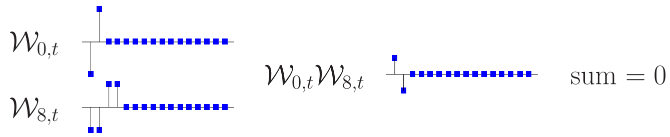
$N = 16$ example



The Haar DWT: III

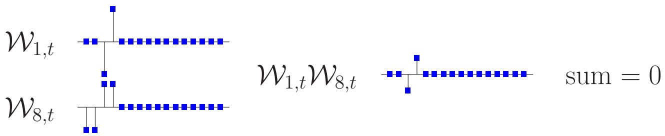
- to form next row, stretch $[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \ldots, 0]$ out by a factor of two and renormalize to preserve unit energy
- $j = \frac{N}{2}$: $[-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \underbrace{0, \ldots, 0}_{N-4 \text{ zeros}}] \equiv \mathcal{W}^T_{\frac{N}{2}\bullet}$

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

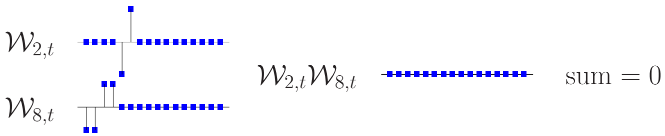


The Haar DWT: IV

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

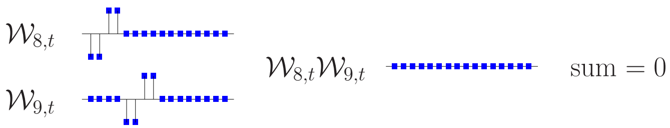


- $\mathcal{W}_{2\bullet}$ and $\mathcal{W}_{\frac{N}{2}\bullet}$ are orthogonal



The Haar DWT: V

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



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

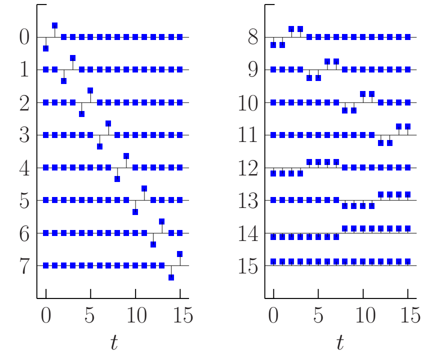
The Haar DWT: VI

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

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

The Haar DWT: VII

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



Haar DWT Coefficients: I

- obtain Haar DWT coefficients $\mathbf{W}$ by premultiplying $\mathbf{X}$ by $\mathcal{W}$:

$$\mathbf{W} = \mathcal{W}\mathbf{X}$$

- jth coefficient W_j is inner product of jth row $\mathcal{W}_{j\bullet}$ and $\mathbf{X}$:

$$W_j = \langle \mathcal{W}_{j\bullet}, \mathbf{X} \rangle$$

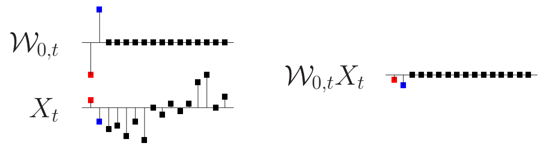
- can interpret coefficients as difference of averages
- to see this, let

$$\overline{X}_t(\lambda) \equiv \frac{1}{\lambda}\sum_{l=0}^{\lambda-1} X_{t-l} = \text{'scale } \lambda\text{' average}$$

  – note: $\overline{X}_t(1) = X_t$ = scale 1 ‘average’
  – note: $\overline{X}_{N-1}(N) = \overline{X}$ = sample average

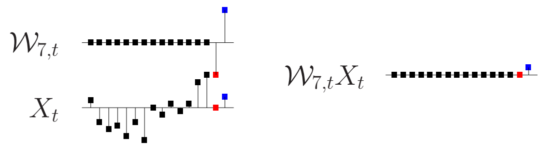
Haar DWT Coefficients: II

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



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

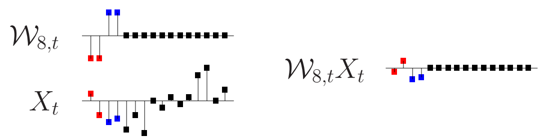
- similar interpretation for $W_1, \ldots, W_{\frac{N}{2}-1} = W_7 = \langle \mathcal{W}_{7\bullet}, \mathbf{X} \rangle$:



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

Haar DWT Coefficients: III

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

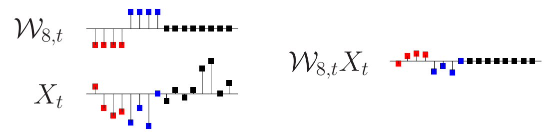


sum $\propto \overline{X}_3(2) - \overline{X}_1(2)$

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

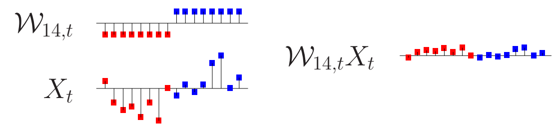
Haar DWT Coefficients: IV

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



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

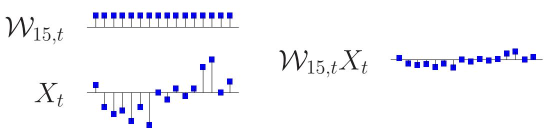
- continuing in this manner, come to $W_{N-1} = \langle \mathcal{W}_{14\bullet}, \mathbf{X} \rangle$:



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

Haar DWT Coefficients: V

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



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

- structure of rows in $\mathcal{W}$
  - first $\frac{N}{2}$ rows yield W_j's $\propto$ *changes* on scale 1
  - next $\frac{N}{4}$ rows yield W_j's $\propto$ *changes* on scale 2
  - next $\frac{N}{8}$ rows yield W_j's $\propto$ *changes* on scale 4
  - next to last row yields $W_j \propto$ *change* on scale $\frac{N}{2}$
  - last row yields $W_j \propto$ *average* on scale N

Structure of DWT Matrices

- $\frac{N}{2\tau_j}$ wavelet coefficients for scale $\tau_j \equiv 2^{j-1}$, $j = 1, \ldots, J$
  - $\tau_j \equiv 2^{j-1}$ is standardized scale
  - $\tau_j\,\Delta t$ is physical scale, where Δt is sampling interval
- each W_j localized in time: as scale $\uparrow$, localization $\downarrow$
- rows of $\mathcal{W}$ for given scale τ_j:
  - circularly shifted with respect to each other
  - shift between adjacent rows is $2\tau_j = 2^j$
- similar structure for DWTs other than the Haar
- differences of averages common theme for DWTs
  - simple differencing replaced by higher order differences
  - simple averages replaced by weighted averages

Two Basic Decompositions Derivable from DWT

- additive decomposition
  - reexpresses $\mathbf{X}$ as the sum of $J+1$ new time series, each of which is associated with a particular scale τ_j
  - called multiresolution analysis (MRA)
  - related to first 'scary-looking' CWT equation
- energy decomposition
  - yields analysis of variance across J scales
  - called wavelet spectrum or wavelet variance
  - related to second 'scary-looking' CWT equation

Partitioning of DWT Coefficient Vector W

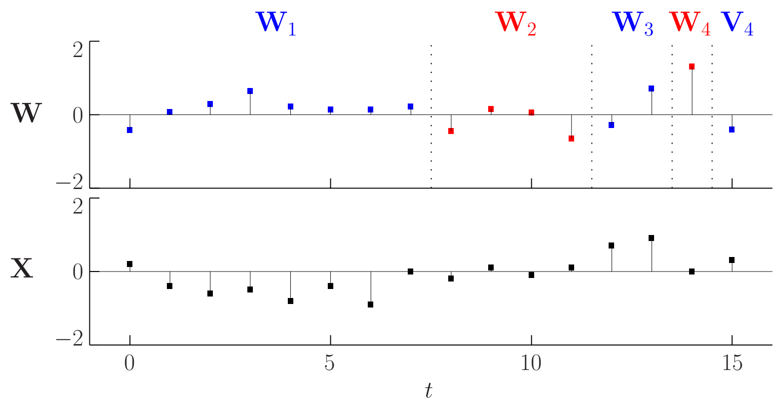
- decompositions are based on partitioning of $\mathbf{W}$ and $\mathcal{W}$
- partition $\mathbf{W}$ into subvectors associated with scale:

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

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

Example of Partitioning of W

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



Partitioning of DWT Matrix $\mathcal{W}$

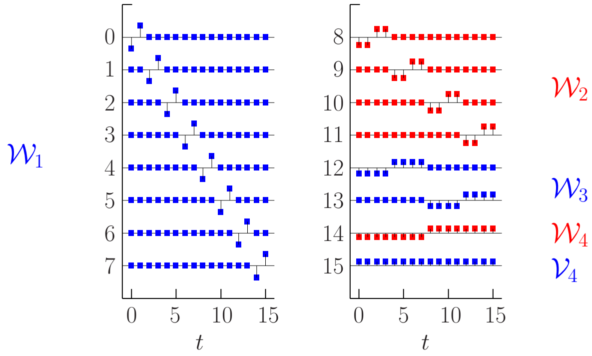
- partition $\mathcal{W}$ commensurate with partitioning of $\mathbf{W}$:

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

- $\mathcal{W}_j$ is $\frac{N}{2^j} \times N$ matrix (related to scale $\tau_j = 2^{j-1}$ changes)
- $\mathcal{V}_J$ is $1 \times N$ row vector (each element is $\frac{1}{\sqrt{N}}$)

Example of Partitioning of $\mathcal{W}$

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



- two properties: (a) $\mathbf{W}_j = \mathcal{W}_j \mathbf{X}$ and (b) $\mathcal{W}_j \mathcal{W}_j^T = I_{\frac{N}{2^j}}$

DWT Analysis and Synthesis Equations

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

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

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

Multiresolution Analysis: I

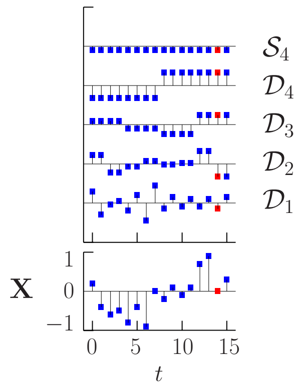
- synthesis equation leads to additive decomposition:

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

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

Multiresolution Analysis: II

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



- adding values for, e.g., $t = 14$ in $\mathcal{D}_1, \ldots, \mathcal{D}_4$ & $\mathcal{S}_4$ yields X_{14}

Energy Preservation Property of DWT Coefficients

- define 'energy' in $\mathbf{X}$ as its squared norm:

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

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

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

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

Wavelet Spectrum (Variance Decomposition): I

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

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

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

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

Wavelet Spectrum (Variance Decomposition): II

- define discrete wavelet power spectrum:

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

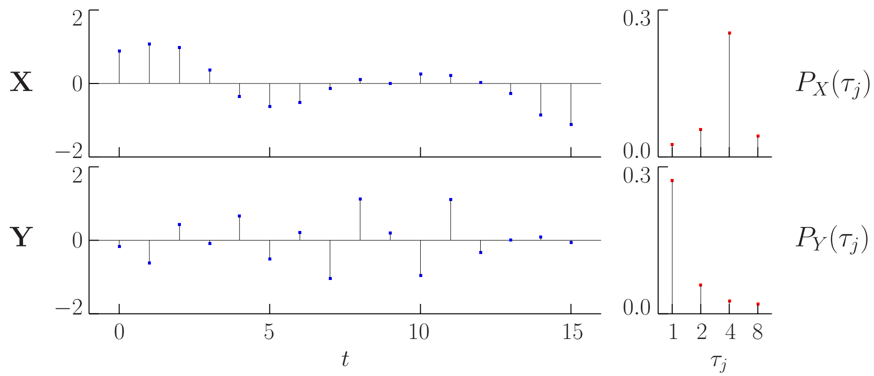
- gives us a scale-based decomposition of the sample variance:

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

- in addition, each $W_{j,t}$ in $\mathbf{W}_j$ associated with a portion of $\mathbf{X}$; i.e., $W_{j,t}^2$ offers scale- & time-based decomposition of $\hat{\sigma}_X^2$

Wavelet Spectrum (Variance Decomposition): III

- wavelet spectra for time series $\mathbf{X}$ and $\mathbf{Y}$ of length $N = 16$, each with zero sample mean and same sample variance



Summary of Qualitative Description of DWT

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