Wavelet Methods for Time Series Analysis

Part IV: Wavelet Packets, Best Bases and Matching Pursuit

- discrete wavelet transforms (DWTs)
  - yields time/scale analysis of $\mathbf{X}$ of sample size N
  - need N to be a multiple of 2^{J_0} for partial DWT of level J_0
  - one partial DWT for each level $j = 1, \ldots, J_0$
  - scale τ_j related to frequencies in $(1/2^{j+1}, 1/2^j]$
  - scale λ_j related to frequencies in $(0, 1/2^{j+1}]$
  - splits $(0, 1/2]$ into octave bands
  - computed via pyramid algorithm
  - maximal overlap DWT also of interest

Wavelet Packet Transforms – Overview

- discrete wavelet packet transforms (DWPTs)
  - yields time/frequency analysis of $\mathbf{X}$
  - need N to be a multiple of 2^{J_0} for DWPT of level J_0
  - one DWPT for each level $j = 1, \ldots, J_0$
  - splits $(0, 1/2]$ into 2^j equal intervals
  - computed via modification of pyramid algorithm
  - can 'mix' parts of DWPTs of different levels j, leading to many more orthonormal transforms and to the notion of a 'best basis' for a particular $\mathbf{X}$
  - maximal overlap DWPT (MODWPT) also of interest

Wavelet Packets – Basic Concepts: I

- 1st stage of DWT pyramid algorithm:

$$\mathcal{P}_1 \mathbf{X} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{V}_1 \end{bmatrix} \equiv \begin{bmatrix} \mathbf{W}_{1,1} \\ \mathbf{W}_{1,0} \end{bmatrix}$$

  - $\mathbf{W}_{1,1} \equiv \mathbf{W}_1$ associated with $f \in (\frac{1}{4}, \frac{1}{2}]$
  - $\mathbf{W}_{1,0} \equiv \mathbf{V}_1$ associated with $f \in [0, \frac{1}{4}]$
- $\mathcal{P}_1$ is orthonormal:

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

- transform is $J_0 = 1$ partial DWT

Wavelet Packets – Basic Concepts: II

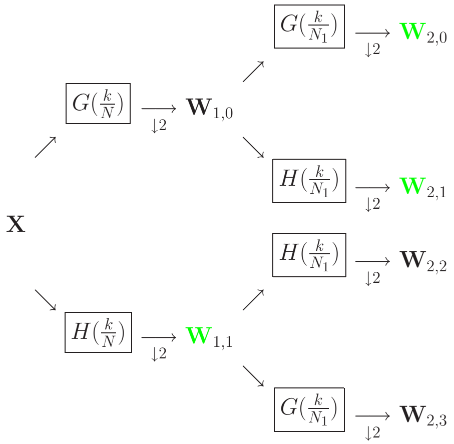
- likewise, 2nd stage defines $J_0 = 2$ partial DWT:

$$\begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \mathbf{V}_2 \end{bmatrix} \equiv \begin{bmatrix} \mathbf{W}_{1,1} \\ \mathbf{W}_{2,1} \\ \mathbf{W}_{2,0} \end{bmatrix}$$

  - $\mathbf{W}_{2,1} \equiv \mathbf{W}_2$ associated with $f \in (\frac{1}{8}, \frac{1}{4}]$
  - $\mathbf{W}_{2,0} \equiv \mathbf{V}_2$ associated with $f \in [0, \frac{1}{8}]$

Wavelet Packets – Basic Concepts: III

- flow diagram for transform from $\mathbf{X}$ to $\mathbf{W}_{2,0}$, $\mathbf{W}_{2,1}$, $\mathbf{W}_{2,2}$ and $\mathbf{W}_{2,3}$:

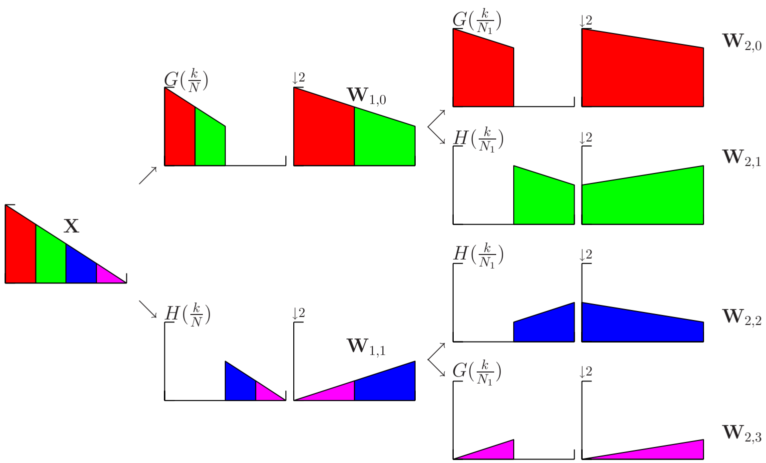


Wavelet Packets – Basic Concepts: IV

- can argue $\mathbf{W}_{2,0}$, $\mathbf{W}_{2,1}$, $\mathbf{W}_{2,2}$ and $\mathbf{W}_{2,3}$ are associated with $f \in [0, \frac{1}{8}]$, $(\frac{1}{8}, \frac{1}{4}]$, $(\frac{1}{4}, \frac{3}{8}]$ and $(\frac{3}{8}, \frac{1}{2}]$
- scheme sometimes called a 'regular' DWT because it splits $[0, \frac{1}{2}]$ split into 4 'regular' subintervals, each of width 1/8
- basis for argument is the following facts:
  - $\mathbf{V}_1$ related to $f \in [0, \frac{1}{4}]$ portion of $\mathbf{X}$
  - $\mathbf{W}_1$ related to $f \in (\frac{1}{4}, \frac{1}{2}]$ portion of $\mathbf{X}$ *but with reversal of order of frequencies*

Wavelet Packets – Basic Concepts: V

- flow diagram in frequency domain:



Wavelet Packets – Basic Concepts: VI

- transform from $\mathbf{X}$ to $\mathbf{W}_{2,0}$, $\mathbf{W}_{2,1}$, $\mathbf{W}_{2,2}$ and $\mathbf{W}_{2,3}$ is called a level $j = 2$ discrete wavelet packet transform
  - abbreviated as DWPT
  - splitting of $[0, \frac{1}{2}]$ similar to DFT
  - unlike DFT, DWPT coefficients localized
  - DWPT is 'time/frequency'; DWT is 'time/scale'
- because level $j = 2$ DWPT is an orthonormal transform, we obtain an energy decomposition:

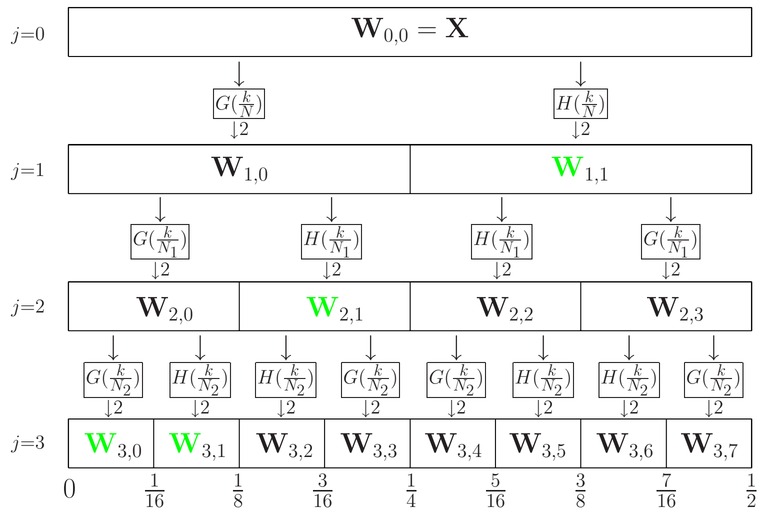
$$\|\mathbf{X}\|^2 = \sum_{n=0}^{3} \|\mathbf{W}_{2,n}\|^2$$

DWPTs of General Levels: I

- can generalize scheme to define DWPTs for levels $j = 0, 1, 2, 3, \ldots$ (with $\mathbf{W}_{0,0}$ defined to be $\mathbf{X}$)
- idea behind DWPT is to use $G(\cdot)$ and $H(\cdot)$ to split each of the 2^{j-1} vectors on level $j-1$ into 2 new vectors, ending up with a level j transform with 2^j vectors
- given $\mathbf{W}_{j-1,n}$'s, here is the rule for generating $\mathbf{W}_{j,n}$'s:
  - if n in $\mathbf{W}_{j-1,n}$ is even:
    - use $G(\cdot)$ to get $\mathbf{W}_{j,2n}$ by transforming $\mathbf{W}_{j-1,n}$
    - use $H(\cdot)$ to get $\mathbf{W}_{j,2n+1}$ by transforming $\mathbf{W}_{j-1,n}$
  - if n in $\mathbf{W}_{j-1,n}$ is odd:
    - use $H(\cdot)$ to get $\mathbf{W}_{j,2n}$ by transforming $\mathbf{W}_{j-1,n}$
    - use $G(\cdot)$ to get $\mathbf{W}_{j,2n+1}$ by transforming $\mathbf{W}_{j-1,n}$

DWPTs of General Levels: II

- example of rule, yielding level $j = 3$ DWPT in the bottom row



DWPTs of General Levels: III

- note: $\mathbf{W}_{j,0}$ and $\mathbf{W}_{j,1}$ correspond to vectors $\mathbf{V}_j$ and $\mathbf{W}_j$ in a jth level partial DWT
- $\mathbf{W}_{j,n}$, $n = 0, \ldots, 2^j - 1$, is associated with $f \in (\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}]$
- n is called the 'sequency' index
- in terms of circular filtering, we can write

$$W_{j,n,t} = \sum_{l=0}^{L-1} u_{n,l} W_{j-1,\lfloor \frac{n}{2} \rfloor, 2t+1-l \bmod N/2^j}, \quad t = 0, \ldots, \frac{N}{2^j} - 1,$$

where $W_{j,n,t}$ is the tth element of $\mathbf{W}_{j,n}$ and

$$u_{n,l} \equiv \begin{cases} g_l, & \text{if } n \bmod 4 = 0 \text{ or } 3; \\ h_l, & \text{if } n \bmod 4 = 1 \text{ or } 2. \end{cases}$$

DWPTs of General Levels: IV

- can also get $\mathbf{W}_{j,n}$ by filtering $\mathbf{X}$ and downsampling:

$$W_{j,n,t} = \sum_{l=0}^{L_j-1} u_{j,n,l} X_{2^j[t+1]-1-l \bmod N}, \quad t = 0, 1, \ldots, \frac{N}{2^j} - 1,$$

where $\{u_{j,n,l}\}$ is the equivalent filter associated with $\mathbf{W}_{j,n}$
- let $\{u_{j,n,l}\} \longleftrightarrow U_{j,n}(\cdot), n = 0, \ldots, 2^j - 1$
- to construct $U_{j,n}(\cdot)$, define $M_0(f) = G(f)$ & $M_1(f) = H(f)$
- let $\mathbf{c}_{1,0} \equiv [0]$ & $\mathbf{c}_{1,1} \equiv [1]$ &, for $j > 1$, create $\mathbf{c}_{j,n}$ recursively
  - by appending 0 to $\mathbf{c}_{j-1,\lfloor \frac{n}{2} \rfloor}$ if $n \bmod 4 = 0$ or 3 or
  - by appending 1 to $\mathbf{c}_{j-1,\lfloor \frac{n}{2} \rfloor}$ if $n \bmod 4 = 1$ or 2

DWPTs of General Levels: V

- letting $c_{j,n,m}$ be mth element of $\mathbf{c}_{j,n}$, then

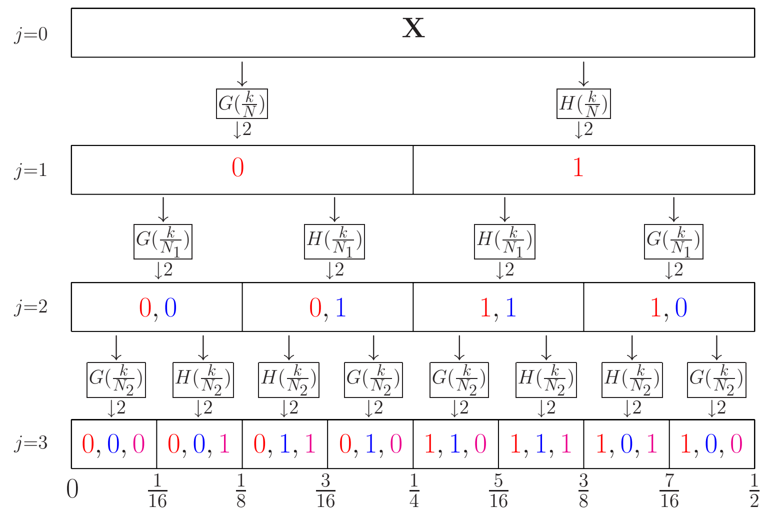
$$U_{j,n}(f) = \prod_{m=0}^{j-1} M_{c_{j,n,m}}(2^m f)$$

- example: $\mathbf{c}_{3,3} = [0, 1, 0]^T$ says

$$U_{3,3}(f) = M_0(f)M_1(2f)M_0(4f) = G(f)H(2f)G(4f)$$

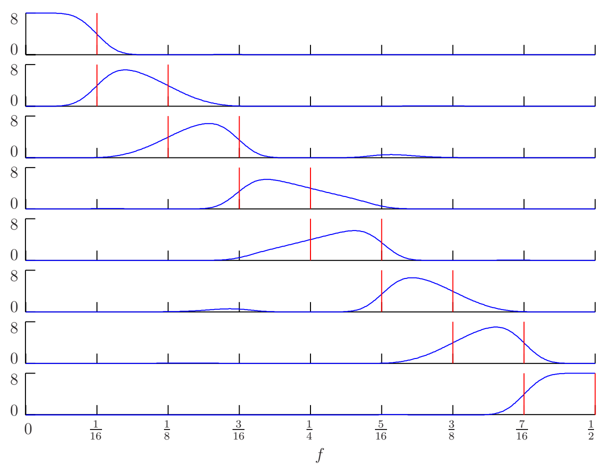
DWPTs of General Levels: VI

- contents of $\mathbf{c}_{j,n}$ for $j = 1, 2$ & 3 and $n = 0, \ldots, 2^j - 1$



DWPTs of General Levels: VII

- squared gain functions $|U_{3,n}(\cdot)|^2$ using LA(8) $\{g_l\}$ & $\{h_l\}$



- note overlap in $n = 3$ and 4 bands – not well separated

DWPTs of General Levels: VIII

- $\mathbf{W}_{j,n}$ nominally associated with bandwidth $1/2^{j+1}$ (corresponding frequency interval is $\mathcal{I}_{j,n} \equiv (\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}]$)
- $\mathbf{W}_{j,0}$ same as $\mathbf{V}_j$ in level j partial DWT
- since $\mathbf{V}_j$ has scale $\lambda_j = 2^j$, can say $\mathbf{W}_{j,0}$ has 'time width' λ_j
- each $\{u_{j,n,l}\}$ has width L_j, so each $\mathbf{W}_{j,n}$ has time width λ_j
- $j = 0$: time width is unity and bandwidth is $1/2$
- $j = J$: time width is $N = 2^J$ and bandwidth is $1/2N$
- note that time width $\times$ bandwidth is constant, which is an example of 'reciprocity'

Wavelet Packet Tables/Trees: I

- collection of DWPTs called a wavelet packet table (or tree), with the tree nodes being labeled by the doublets (j, n):

<table>
<tr><td>j=0</td><td colspan="8">$\mathbf{W}_{0,0} = \mathbf{X}$</td></tr>
<tr><td>j=1</td><td colspan="4">$\mathbf{W}_{1,0}$</td><td colspan="4">$\mathbf{W}_{1,1}$</td></tr>
<tr><td>j=2</td><td colspan="2">$\mathbf{W}_{2,0}$</td><td colspan="2">$\mathbf{W}_{2,1}$</td><td colspan="2">$\mathbf{W}_{2,2}$</td><td colspan="2">$\mathbf{W}_{2,3}$</td></tr>
<tr><td>j=3</td><td>$\mathbf{W}_{3,0}$</td><td>$\mathbf{W}_{3,1}$</td><td>$\mathbf{W}_{3,2}$</td><td>$\mathbf{W}_{3,3}$</td><td>$\mathbf{W}_{3,4}$</td><td>$\mathbf{W}_{3,5}$</td><td>$\mathbf{W}_{3,6}$</td><td>$\mathbf{W}_{3,7}$</td></tr>
</table>

$0 \quad \frac{1}{16} \quad \frac{1}{8} \quad \frac{3}{16} \quad \frac{1}{4} \quad \frac{5}{16} \quad \frac{3}{8} \quad \frac{7}{16} \quad \frac{1}{2}$

- nodes $\mathcal{C} \equiv \{(j, n) : n = 0, \ldots, 2^j - 1\}$ for row j form a DWPT
- nonoverlapping complete covering of $[0, \frac{1}{2}]$ yields coefficients for an orthonormal transform $\mathbf{O}$ ('disjoint dyadic decomposition')
- let's consider 2 sets of doublets yielding such a decomposition

Wavelet Packet Tables/Trees: II

- $\mathcal{C} = \{(3,0), (3,1), (2,1), (1,1)\}$ yields the DWT:

<table>
<tr><td>j=0</td><td colspan="8"></td></tr>
<tr><td>j=1</td><td colspan="4"></td><td colspan="4">$\mathbf{W}_{1,1}$</td></tr>
<tr><td>j=2</td><td colspan="2"></td><td colspan="2">$\mathbf{W}_{2,1}$</td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>j=3</td><td>$\mathbf{W}_{3,0}$</td><td>$\mathbf{W}_{3,1}$</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

- $\mathcal{C} = \{(2,0), (3,2), (3,3), (1,1)\}$ yields another $\mathbf{O}$:

<table>
<tr><td>j=0</td><td colspan="8"></td></tr>
<tr><td>j=1</td><td colspan="4"></td><td colspan="4">$\mathbf{W}_{1,1}$</td></tr>
<tr><td>j=2</td><td colspan="2">$\mathbf{W}_{2,0}$</td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>j=3</td><td></td><td></td><td>$\mathbf{W}_{3,2}$</td><td>$\mathbf{W}_{3,3}$</td><td></td><td></td><td></td><td></td></tr>
</table>

Optimal Orthonormal Transform: I

- WP table yields *many* $\mathbf{O}$'s: is one 'optimal'?
- Coifman & Wickerhauser (1992) proposed notion of 'best basis'
- form WP table out to level J, and assign 'cost' to $\mathbf{W}_{j,n}$ via

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

where $m(\cdot)$ is real-valued cost function (require $m(0) = 0$)

- let $\mathcal{C}$ be any collection of indices in the set $\mathcal{N}$ of all possible indices forming an orthonormal transform
- 'optimal' such transform satisfies

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

Optimal Orthonormal Transform: II

- consider following 2 unit norm vectors:

$$\mathbf{W}_{j,n}^{(1)} = \left[\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}\right]^T \quad \text{and} \quad \mathbf{W}_{j,n}^{(2)} = [1, 0, 0, 0]^T$$

- example: 'entropy-based' cost function

$$m(|W_{j,n,t}|) = -W_{j,n,t}^2 \log(W_{j,n,t}^2)$$

(since $|x| \log(|x|) \to 0$ as $x \to 0$, will interpret $0 \log(0)$ as 0)

- here $M(\mathbf{W}_{j,n}^{(1)}) = 4 \cdot (-\frac{1}{4} \log \frac{1}{4}) > 0$ and $M(\mathbf{W}_{j,n}^{(2)}) = 0$ (lower cost if energy is concentrated in a few $|W_{j,n,t}|$'s)

Optimal Orthonormal Transform: III

- continue looking at $\mathbf{W}_{j,n}^{(1)} = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right]^T$ & $\mathbf{W}_{j,n}^{(2)} = [1, 0, 0, 0]^T$
- 2nd example: threshold cost function

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

  if $\delta = 1/4$, $M(\mathbf{W}_{j,n}^{(1)}) = 4$ and $M(\mathbf{W}_{j,n}^{(2)}) = 1$
  (lower cost if there are only a few large $|W_{j,n,t}|$'s)
- 3rd example: ℓ_p cost function $m(|W_{j,n,t}|) = |W_{j,n,t}|^p$
  if $p = 1$, $M(\mathbf{W}_{j,n}^{(1)}) = 2$ and $M(\mathbf{W}_{j,n}^{(2)}) = 1$
  (same pattern as before)
- once costs assigned, need to find optimal transform

Optimal Orthonormal Transform: IV

- let $\mathbf{X}$ be following series of length $N = 8$:

$$\mathbf{X} = \sqrt{2}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} + \sqrt{8}\begin{bmatrix} \frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \\ -2 \\ 2 \end{bmatrix}$$

Optimal Orthonormal Transform: V

- Haar DWPT coefficients, levels $j = 1, 2$ and 3 (three underlined coefficents correspond to basis vectors used in forming $\mathbf{X}$):

j=0	$\mathbf{X} = [2, 0, -1, 1, 0, 0, -2, 2]^T$							
j=1	$[\underline{\sqrt{2}}, 0, 0, 0]$				$[-\sqrt{2}, \sqrt{2}, 0, \sqrt{8}]$			
j=2	$[1, 0]$		$[-1, 0]$		$[2, 2]$		$[0, \underline{2}]$	
j=3	$[\frac{1}{\sqrt{2}}]$	$[-\frac{1}{\sqrt{2}}]$	$[\frac{1}{\sqrt{2}}]$	$[-\frac{1}{\sqrt{2}}]$	$[\underline{\sqrt{8}}]$	$[0]$	$[\sqrt{2}]$	$[\sqrt{2}]$

Optimal Orthonormal Transform: VI

- cost table using $-W_{j,n,t}^2 \log(W_{j,n,t}^2)$ cost function:

j=0	1.45							
j=1	0.28				0.88			
j=2	0.19		0.19		0.72		0.36	
j=3	**<u>0.12</u>**	**<u>0.12</u>**	**<u>0.12</u>**	**<u>0.12</u>**	**<u>0.32</u>**	**<u>0.00</u>**	**<u>0.28</u>**	**<u>0.28</u>**

- algorithm to find 'best' basis
  - mark all costs of 'childern' nodes at bottom
  - compare cost of children with their 'parent'
    - if parent cheaper, mark parent node
    - if children cheaper, replace cost of parent
  - repeat for each level; when done, look for top-marked nodes

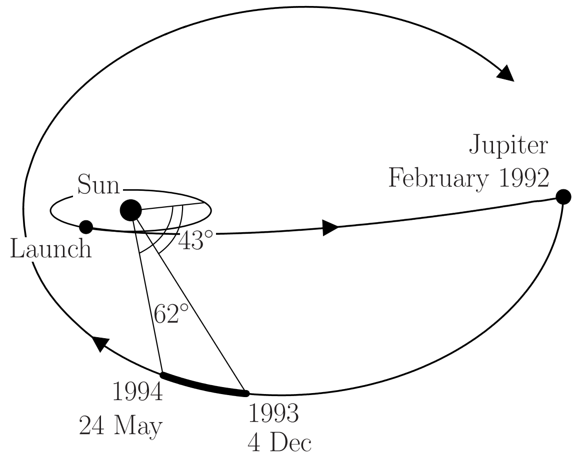
Optimal Orthonormal Transform: VII

- final step (best basis includes 3 vectors forming **X**):

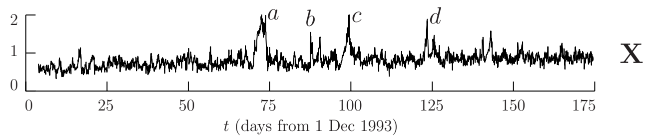
<table>
<tr><td>j=0</td><td colspan="8">0.96</td></tr>
<tr><td>j=1</td><td colspan="4"><u>0.28</u></td><td colspan="4">0.68</td></tr>
<tr><td>j=2</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.32</td><td colspan="2"><u>0.36</u></td></tr>
<tr><td>j=3</td><td><u>0.12</u></td><td><u>0.12</u></td><td><u>0.12</u></td><td><u>0.12</u></td><td><u>0.32</u></td><td><u>0.00</u></td><td><u>0.28</u></td><td><u>0.28</u></td></tr>
</table>

Example – Analysis of Solar Physics Data: I

- path of Ulysses spacecraft (records magnetic field of heliosphere)

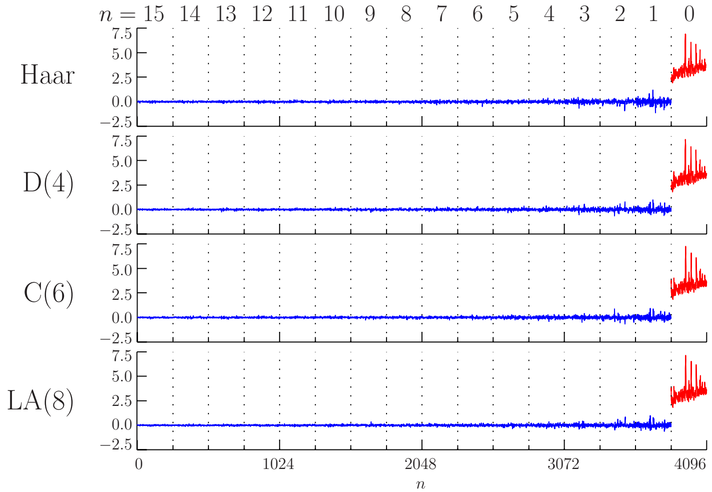


Example – Analysis of Solar Physics Data: II



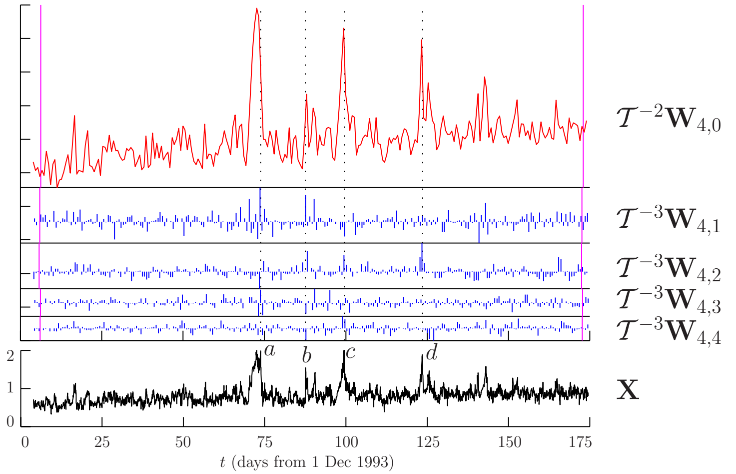
- magnetic field measurements of polar region of sun recorded hourly from 4 Dec 1993 to 24 May 1994 ($\Delta t = 1/24$ day)
- Ulysses moved from 4 AU to 3 AU (explains upward trend)
- a, b, c, d are fast solar wind streams from polar coronal holes
- two classifications for these 'shocks'
  - corotating interaction regions (CIRs) – recur every solar rotation (about 25 days)
  - fast coronal mass ejections (CMEs) – transient in nature

Example – Analysis of Solar Physics Data: III



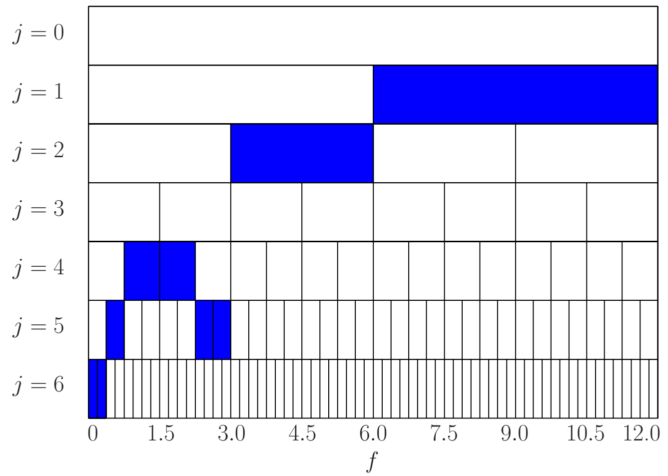
- 4 different level $j = 4$ DWPTs, each partitioning $(0, 1/2\,\Delta t]$ into 16 intervals

Example – Analysis of Solar Physics Data: IV



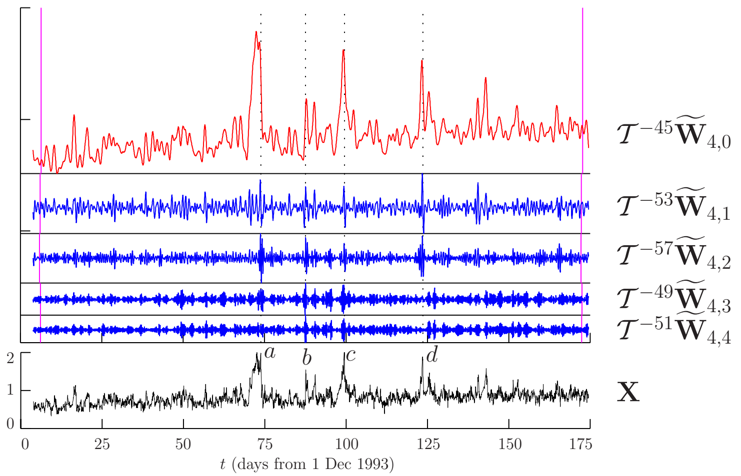
- level $j = 4$ LA(8) DWPT coefficients $\mathbf{W}_{4,n}$, $n = 0, \ldots, 4$, after time alignments (derived from study of phase functions)

Example – Analysis of Solar Physics Data: V



- best basis transform using LA(8) filter and $-W_{j,n,t}^2 \log(W_{j,n,t}^2)$ cost function

Example – Analysis of Solar Physics Data: VI

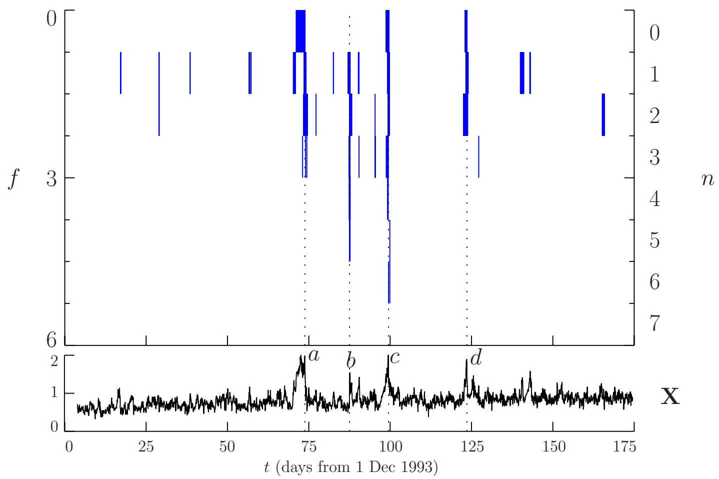


- level $j = 4$ LA(8) MODWPT coefficients $\mathbf{W}_{4,n}$, $n = 0, \ldots, 4$, after time alignments

Example – Analysis of Solar Physics Data: VII

- will summarize using a modified time/frequency plot, which indicates locations of
  - 100 largest values in $\mathcal{T}^{-|\nu_{4,0}|}\widetilde{\mathbf{W}}_{4,0}$
  - 100 largest values in $\mathcal{T}^{-|\nu_{4,1}|}\widetilde{\mathbf{W}}_{4,1}$
  - 100 largest values in $\mathcal{T}^{-|\nu_{4,n}|}\widetilde{\mathbf{W}}_{4,n}$, $n = 2, \ldots, 15$ (in fact these all occur in $n = 2, \ldots, 6$)

Example – Analysis of Solar Physics Data: VIII



- 4 events coherently broad-band; events a, c, d are recurrent; b is transient; a might be two events (recurrent & transient)

Matching Pursuit – Basics

- idea: approximate $\mathbf{X}$ using a few # of 'time/frequency' vectors from large set of such vectors (cf. best basis)
- form 'dictionary' of vectors $\mathcal{D} \equiv \{\mathbf{d}_\gamma : \gamma \in \Gamma\}$
  - $\mathbf{d}_\gamma = \left[d_{\gamma,0}, d_{\gamma,1}, \ldots, d_{\gamma,N-1}\right]^T$
  - each vector has unit norm: $\|\mathbf{d}_\gamma\|^2 = \sum_{l=0}^{N-1} d_{\gamma,l}^2 = 1$
  - γ is vector of parameters connecting $\mathbf{d}_\gamma$ to time/frequency; e.g., $\gamma = [j, n, t]^T$ for WP table dictionary
  - Γ = finite set of possible values for γ
  - $\mathcal{D}$ contains basis for $\mathcal{R}^N$, but can be highly redundant (helps identify time/frequency content in $\mathbf{X}$)
- matching pursuit successively approximates $\mathbf{X}$ with orthogonal projections onto elements of $\mathcal{D}$

Background Material

- recall that we can reconstruct a time series $\mathbf{X}$ from its DWT coefficients $\mathbf{W}$ via $\mathbf{X} = \mathcal{W}^T\mathbf{W}$, where $\mathbf{W} \equiv \mathcal{W}\mathbf{X}$
- jth coefficient in $\mathbf{W}$ is $\langle \mathbf{X}, \mathcal{W}_{j\bullet}\rangle$, i.e., the inner product of $\mathbf{X}$ & a column vector $\mathcal{W}_{j\bullet}$ whose elements are the jth row of $\mathcal{W}$
- hence we can write

$$\mathbf{X} = \mathcal{W}^T\mathbf{W} = [\mathcal{W}_{0\bullet}, \mathcal{W}_{1\bullet}, \ldots, \mathcal{W}_{N-1\bullet}] \begin{bmatrix} \langle \mathbf{X}, \mathcal{W}_{0\bullet}\rangle \\ \langle \mathbf{X}, \mathcal{W}_{1\bullet}\rangle \\ \vdots \\ \langle \mathbf{X}, \mathcal{W}_{N-1\bullet}\rangle \end{bmatrix} = \sum_{j=0}^{N-1} \langle \mathbf{X}, \mathcal{W}_{j\bullet}\rangle \mathcal{W}_{j\bullet}$$

- regard $\langle \mathbf{X}, \mathcal{W}_{j\bullet}\rangle \mathcal{W}_{j\bullet}$ as approximation to $\mathbf{X}$ based on just $\mathcal{W}_{j\bullet}$

Matching Pursuit Algorithm: I

- for $\mathbf{d}_{\gamma_0} \in \mathcal{D}$, form $\langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle \mathbf{d}_{\gamma_0}$, and define residual vector: $\mathbf{R}^{(1)} \equiv \mathbf{X} - \langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle \mathbf{d}_{\gamma_0}$ so that $\mathbf{X} = \langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle \mathbf{d}_{\gamma_0} + \mathbf{R}^{(1)}$
- Exer. [240] says that $\mathbf{d}_{\gamma_0}$ and $\mathbf{R}^{(1)}$ orthogonal:

$$\begin{aligned} \langle \mathbf{d}_{\gamma_0}, \mathbf{R}^{(1)}\rangle &= \langle \mathbf{d}_{\gamma_0}, \mathbf{X} - \langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle \mathbf{d}_{\gamma_0}\rangle \\ &= \langle \mathbf{d}_{\gamma_0}, \mathbf{X}\rangle - \langle \mathbf{d}_{\gamma_0}, \langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle \mathbf{d}_{\gamma_0}\rangle \\ &= \langle \mathbf{d}_{\gamma_0}, \mathbf{X}\rangle - \langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle = 0 \end{aligned}$$

- hence $\langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle \mathbf{d}_{\gamma_0}$ & $\mathbf{R}^{(1)}$ are also orthogonal, showing that

$$\|\mathbf{X}\|^2 = \|\langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle \mathbf{d}_{\gamma_0}\|^2 + \|\mathbf{R}^{(1)}\|^2 = \left|\langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle\right|^2 + \|\mathbf{R}^{(1)}\|^2$$

- minimize energy in residuals by choosing $\gamma_0 \in \Gamma$ such that

$$\left|\langle \mathbf{X}, \mathbf{d}_{\gamma_0}\rangle\right| = \max_{\gamma \in \Gamma} \left|\langle \mathbf{X}, \mathbf{d}_\gamma\rangle\right|$$

Matching Pursuit Algorithm: II

- after first step of algorithm, second step is to treat the residuals in the same manner as $\mathbf{X}$ was treated in first step, yielding
$$\mathbf{R}^{(1)} = \langle \mathbf{R}^{(1)}, \mathbf{d}_{\gamma_1} \rangle \mathbf{d}_{\gamma_1} + \mathbf{R}^{(2)},$$
with $\mathbf{d}_{\gamma_1}$ picked such that
$$\left| \langle \mathbf{R}^{(1)}, \mathbf{d}_{\gamma_1} \rangle \right| = \max_{\gamma \in \Gamma} \left| \langle \mathbf{R}^{(1)}, \mathbf{d}_{\gamma} \rangle \right|$$
- letting $\mathbf{R}^{(0)} \equiv \mathbf{X}$, after m such steps, have additive decomposition:
$$\mathbf{X} = \sum_{k=0}^{m-1} \langle \mathbf{R}^{(k)}, \mathbf{d}_{\gamma_k} \rangle \mathbf{d}_{\gamma_k} + \mathbf{R}^{(m)}$$

Matching Pursuit Algorithm: III

- also have an energy decomposition:
$$\|\mathbf{X}\|^2 = \sum_{k=0}^{m-1} \| \langle \mathbf{R}^{(k)}, \mathbf{d}_{\gamma_k} \rangle \mathbf{d}_{\gamma_k} \|^2 + \|\mathbf{R}^{(m)}\|^2$$
$$= \sum_{k=0}^{m-1} |\langle \mathbf{R}^{(k)}, \mathbf{d}_{\gamma_k} \rangle|^2 + \|\mathbf{R}^{(m)}\|^2$$
- note: as m increases, $\|\mathbf{R}^{(m)}\|^2$ must decrease (must reach zero under certain conditions)

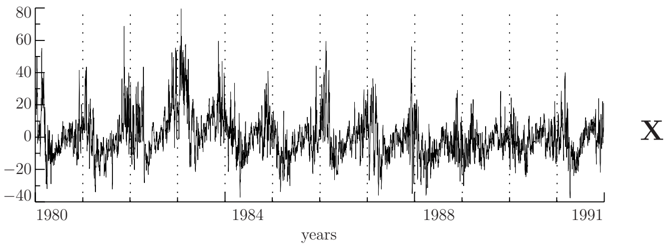
Matching Pursuit Dictionaries: I

- key to matching pursuit is dictionary
- simplest dictionary: DWT dictionary
  - $\mathcal{D}$ contains $\mathbf{d}_\gamma \equiv \mathcal{W}_{j\bullet}$, $j = 0, \ldots, N-1$
  - $\gamma = [j]$ associates $\mathcal{W}_{j\bullet}$ with time/scale
  - $\langle \mathbf{X}, \mathbf{d}_\gamma \rangle = W_j$ is jth DWT coefficient
  - 1st step picks W_j with largest magnitude:
$$\mathbf{X} = W_{(0)} \mathbf{W}_{(0)} + \mathbf{R}^{(1)} \text{ with } \mathbf{R}^{(1)} = \sum_{j \neq (0)} W_j \mathbf{W}_{j\bullet}$$
  - 2nd step picks out W_j with 2nd largest $|W_j|$
  - for any orthonormal $\mathcal{D}$, matching pursuit approximates $\mathbf{X}$ using coefficients with largest magnitudes

Matching Pursuit Dictionaries: II

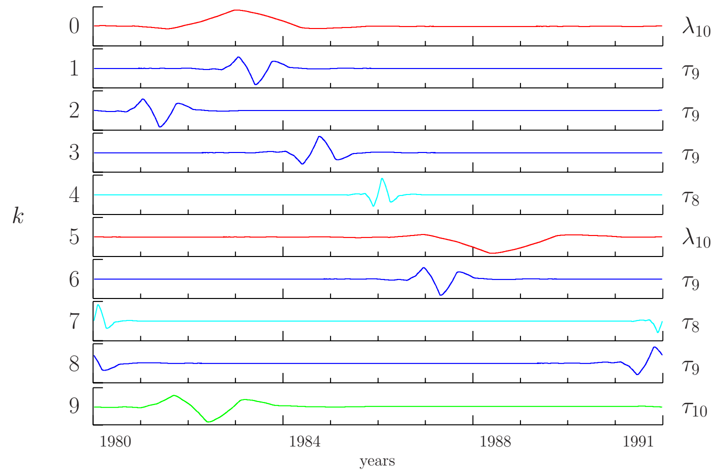
- larger dictionary: wavelet packet table dictionary (more flexible than best basis)
- even larger dictionary: above combined with basis vectors corresponding to a discrete Fourier transform (DFT)
- level J_0 MODWT dictionary
  - works for all N, shift invariant, redundant
  - $\mathcal{D}$ contains vectors whose elements are either
    * normalized rows of $\widetilde{\mathcal{W}}_j$, $j = 1, \ldots, J_0$, or
    * normalized rows of $\widetilde{\mathcal{V}}_{J_0}$

Example – Subtidal Sea Levels: I



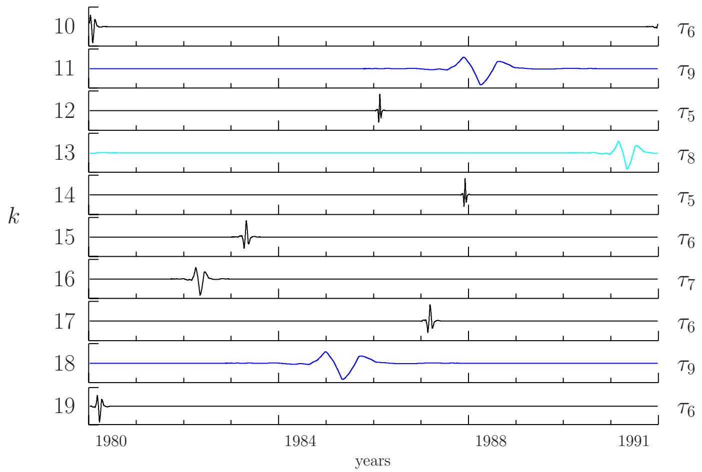
- recall subtidal sea level series **X** for Crescent City, CA

Example – Subtidal Sea Levels: II



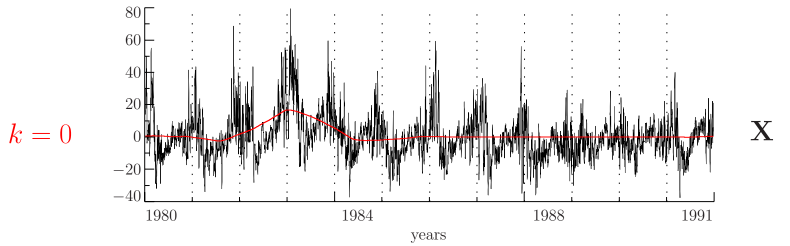
- use $J_0 = 10$ LA(8) MODWT dictionary (96,206 vectors in all)
- above shows first 10 vectors picked by matching pursuit ($\times \pm 1$)

Example – Subtidal Sea Levels: III



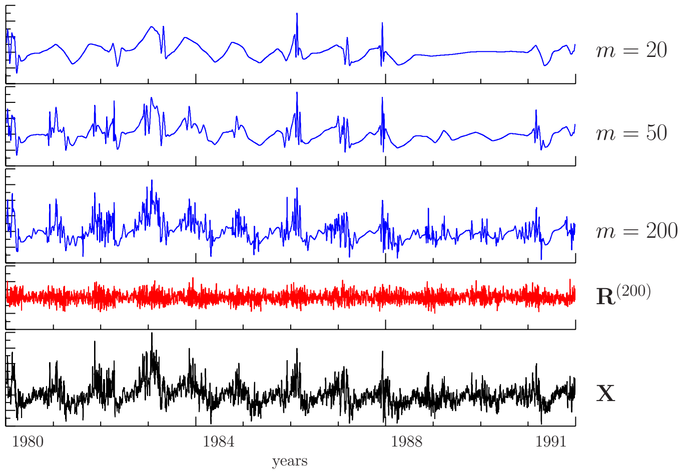
- next 10 vectors picked by matching pursuit ($\times \pm 1$)

Example – Subtidal Sea Levels: IV



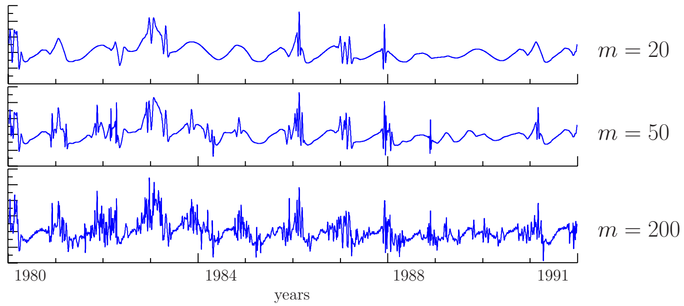
- very first ($k = 0$) associated with overall increase in 1982–3
- first 10 are for $\tau_8 \, \Delta t = 64$ to $\lambda_{10} \, \Delta t = 512$ days
- 7 of first 20 are associated with $\tau_9 \, \Delta t = 128$ days (needed to account for seasonal variabilty)
- $k = 3$ has inverted sign & picks out gradual dip in Spring, 1984 (cf. 1981, 3, 5, 7 & 8); $k = 8$ also inverted, but is a boundary effect

Example – Subtidal Sea Levels: V



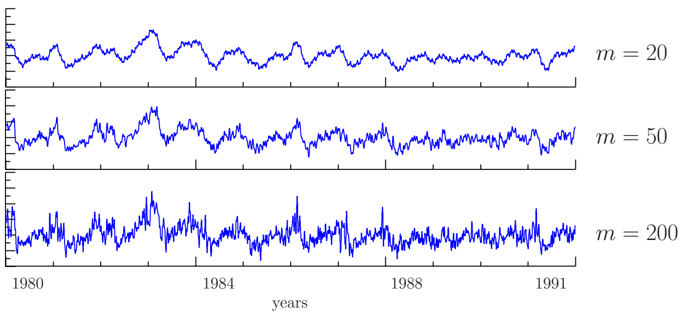
- matching pursuit approximations of orders $m = 20, 50$ and 200, along with residuals for $m = 200$

Example – Subtidal Sea Levels: VI



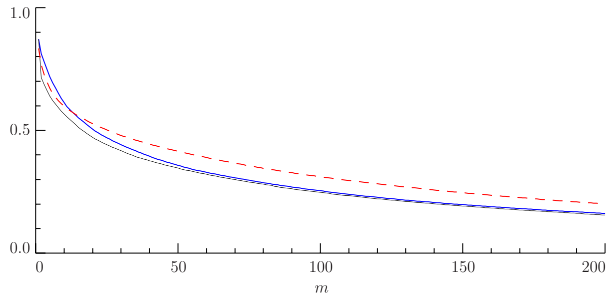
- matching pursuit approximations of orders $m = 20, 50$ and 200, but now using a dictionary augmented to include basis vectors corresponding to the DFT
- $k = 0$ choice same as before, but $k = 1$ choice is DFT vector with period close to one year
- for $2 \le k < 200$, only $k = 65, 84$ and 192 are DFT vectors

Example – Subtidal Sea Levels: VII



- matching pursuit approximations of orders $m = 20, 50$ and 200, but now using a dictionary consisting of just the basis vectors corresponding to the DFT

Example – Subtidal Sea Levels: VIII



- normalized residual sum of squares $\|\mathbf{R}^{(\mathbf{m})}\|^2/\|\mathbf{X}\|^2$ versus number of terms m in matching pursuit approximation using the MODWT dictionary (thick curve), the DFT-based dictionary (dashed) and both dictionaries combined (thin)
- combined dictionary does best for small m, but MODWT dictionary by itself becomes competitive as m increases