E9 285 Biomedical imaging-Inverse problems

Chapter 2.b (3)
Representative reconstruction methods for fluorescence microscopy

The Wavelet Thresholded-Lendweber Method

M. A. T. Figueiredo and R. D. Nowak, “An EM algorithm for wavelet-based image restoration,” *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Aug. 2004.

Lendeweber method for unregularized case

$\mathbf{x}$: scanned vector of the required image

$\mathbf{H}$: convolution matrix of the microscope

$\mathbf{W}$: matrix corresponding to wavelet decomposition

The method minimizes the following cost:

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2.$$

The iteration is given by:

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \tau \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)}).$$

Wavelet regularized denoising

The method minimizes the following cost:

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1.$$

$$\begin{aligned} J(\mathbf{x}) &= \|\mathbf{W}\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1 \\ &= \sum_{n=1}^{N} |w_{\mathbf{y}n} - w_{\mathbf{x}n}|^2 + \lambda |w_{\mathbf{x}n}| \end{aligned}$$

The solution is obtained by the following function:

$$\mathbf{x} = \mathbf{W}^T \mathcal{T}_{\lambda/2}\{\mathbf{W}\mathbf{y}\}$$

$$\mathcal{T}_{\lambda/2}(w) = \operatorname{sgn}(w)(|w| - \lambda/2)_+$$

where $(\,\cdot\,)_+$ is the positive-part function

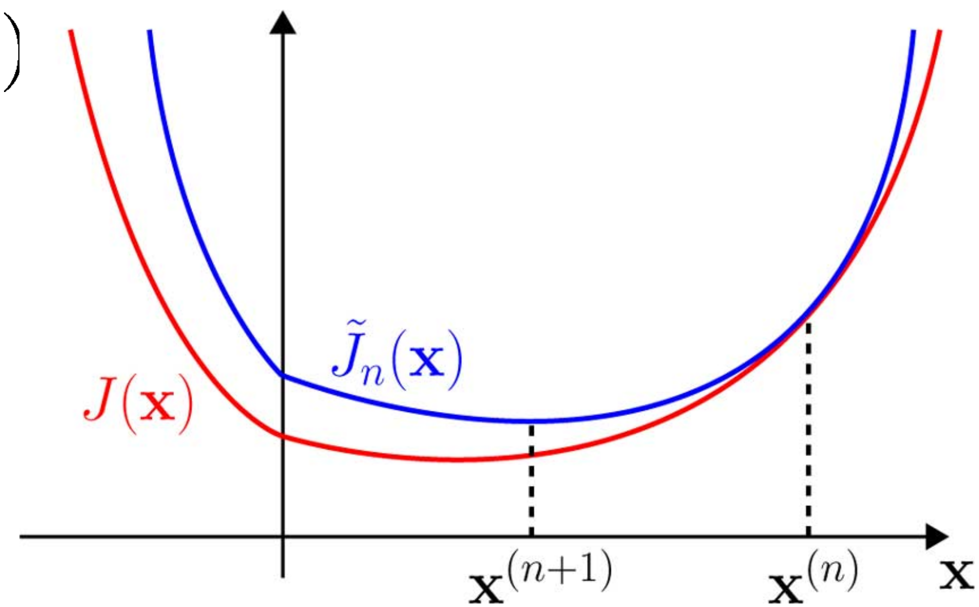
$$(t)_+ = \begin{cases} t, & \text{if } t > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The bounded optimization principle

Using the current estimate $\mathbf{x}^{(n)}$, the key idea is to construct an auxiliary functional $\tilde{J}_n(\mathbf{x})$ with the following properties:

- when $\mathbf{x} = \mathbf{x}^{(n)}$, $\tilde{J}_n(\mathbf{x})$ coincides with $J(\mathbf{x})$;
- when $\mathbf{x} \neq \mathbf{x}^{(n)}$, $\tilde{J}_n(\mathbf{x})$ upper-bounds $J(\mathbf{x})$.

$$\mathbf{x}^{(n+1)} = \arg\min_{\mathbf{x}} \tilde{J}_n(\mathbf{x})$$



$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1.$$

The method of Daubechies et al.

Daubechies *et al.* [11] proposed to use functionals of the form

$$\tilde{J}_n(\mathbf{x}) = \alpha \left\| \mathbf{x}^{(n)} - \mathbf{x} \right\|_2^2 + J(\mathbf{x}) - \left\| \mathbf{H}\mathbf{x}^{(n)} - \mathbf{H}\mathbf{x} \right\|_2^2. \quad (8)$$

Here, the (real and positive) scalar α must be chosen strictly larger than the spectral radius of $\mathbf{H}^T\mathbf{H}$

$$\alpha > \rho(\mathbf{H}^T\mathbf{H}) \quad \text{where} \quad \rho(\mathbf{H}^T\mathbf{H}) = \max_{\|\mathbf{v}\|_2 = 1} \|\mathbf{H}\mathbf{v}\|_2^2.$$

Equivalently, since $\mathbf{H}$ is a convolution matrix, $\rho(\mathbf{H}^T\mathbf{H})$ is the largest squared modulus of the DFT coefficients of $\mathbf{H}$.

[11] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Aug. 2004.

Property 1: Assume that $\alpha > \rho(\mathbf{H}^T\mathbf{H})$ holds in (8). Then $\tilde{J}_n(\mathbf{x}) > J(\mathbf{x})$, except at $\mathbf{x} = \mathbf{x}^{(n)}$, where $\tilde{J}_n(\mathbf{x}) = J(\mathbf{x})$.

Proof: The inequality $\alpha > \rho(\mathbf{H}^T\mathbf{H})$ ensures that $\alpha\mathbf{I} - \mathbf{H}^T\mathbf{H}$ is positive-definite. This means that, when $\mathbf{x} \neq \mathbf{x}^{(n)}$, we have

$$\alpha \left\|\mathbf{x}^{(n)} - \mathbf{x}\right\|_2^2 - \left\|\mathbf{H}\mathbf{x}^{(n)} - \mathbf{H}\mathbf{x}\right\|_2^2 \\ = \left(\mathbf{x}^{(n)} - \mathbf{x}\right)^T (\alpha\mathbf{I} - \mathbf{H}^T\mathbf{H}) \left(\mathbf{x}^{(n)} - \mathbf{x}\right) > 0$$

whence it follows that $\tilde{J}_n(\mathbf{x})$ is a strict upper bound of $J(\mathbf{x})$, except at $\mathbf{x} = \mathbf{x}^{(n)}$ where the equality is met. ■

Property 2: Definition (8) is equivalent to

$$\tilde{J}_n(\mathbf{x}) = \alpha \left\| \mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T \left(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)} \right) - \mathbf{x} \right\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1 + c$$

where c is a constant that does not depend on $\mathbf{x}$.

$$\begin{aligned}\tilde{J}_n(\mathbf{x}) &= \alpha \left\|\mathbf{x}^{(n)} - \mathbf{x}\right\|_2^2 + \|\mathbf{y} - \mathbf{Hx}\|_2^2 \\ &\quad - \left\|\mathbf{Hx}^{(n)} - \mathbf{Hx}\right\|_2^2 + \lambda\|\mathbf{Wx}\|_1 \\ &= \alpha \left\|\mathbf{x}^{(n)}\right\|_2^2 - 2\alpha\mathfrak{Re}\left\{\mathbf{x}^T\mathbf{x}^{(n)}\right\} + \alpha\|\mathbf{x}\|_2^2 \\ &\quad + \|\mathbf{y}\|_2^2 - 2\mathfrak{Re}\{\mathbf{x}^T\mathbf{H}^T\mathbf{y}\} + \|\mathbf{Hx}\|_2^2 \\ &\quad - \|\mathbf{Hx}^{(n)}\|_2^2 + 2\mathfrak{Re}\left\{\mathbf{x}^T\mathbf{H}^T\mathbf{Hx}^{(n)}\right\} \\ &\quad - \|\mathbf{Hx}\|_2^2 + \lambda\|\mathbf{Wx}\|_1\end{aligned}$$

$$
\begin{aligned}
&= \alpha||\mathbf{x}||_2^2 - 2\alpha \\
&\quad \times \mathfrak{Re}\left\{\mathbf{x}^T\left(\mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T\left(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)}\right)\right)\right\} \\
&\quad + \lambda||\mathbf{W}\mathbf{x}||_1 + c'
\end{aligned}
$$

where $c' = \alpha||\mathbf{x}^{(n)}||_2^2 + ||\mathbf{y}||_2^2 - ||\mathbf{H}\mathbf{x}^{(n)}||_2^2$ is a constant that does not depend on $\mathbf{x}$. We complete the proof by adding the constant $\alpha||\mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})||_2^2$ (which does not depend on $\mathbf{x}$) so as to complete the quadratic term. ■

The algorithm

- compute the Landweber iteration $\mathbf{z}^{(n)} = \mathbf{x}^{(n)} + \tau \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})$, with step-size τ;
- perform the wavelet-domain denoising operation $\mathbf{x}^{(n+1)} = \mathbf{W}^T \mathcal{T}_{\lambda\tau/2}\{\mathbf{W}\mathbf{z}^{(n)}\}$, with threshold level $\lambda\tau/2$.

The Fast Wavelet Thresholded-Lendweber Method
(using Shannon wavelets)

C. Vonesch and M. Unser, “A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution,” *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 539–549, Apr. 2008.

Recap: Upsampler and downsampler

$$x(n) \rightarrow \downarrow M \rightarrow y(n)$$

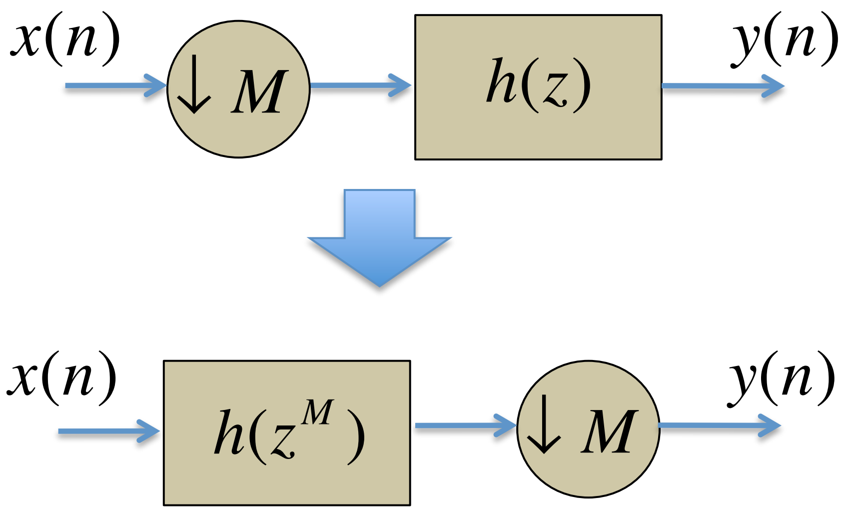
$$y(n) = x(nM)$$

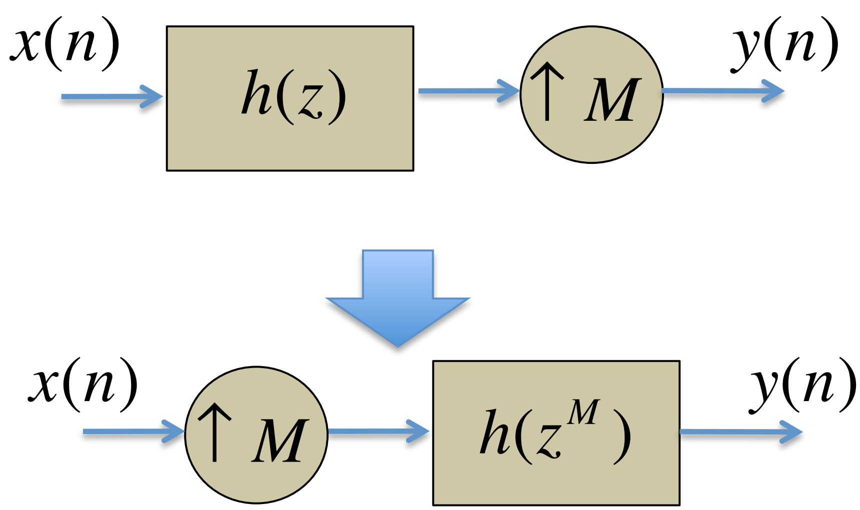
$$x(n) \rightarrow \uparrow M \rightarrow y(n)$$

$$y(n) = \begin{cases} x(n/M), & \text{for } n/M \text{ being} \\ \text{an integer} & \\ 0, & \text{Otherwise} \end{cases}$$

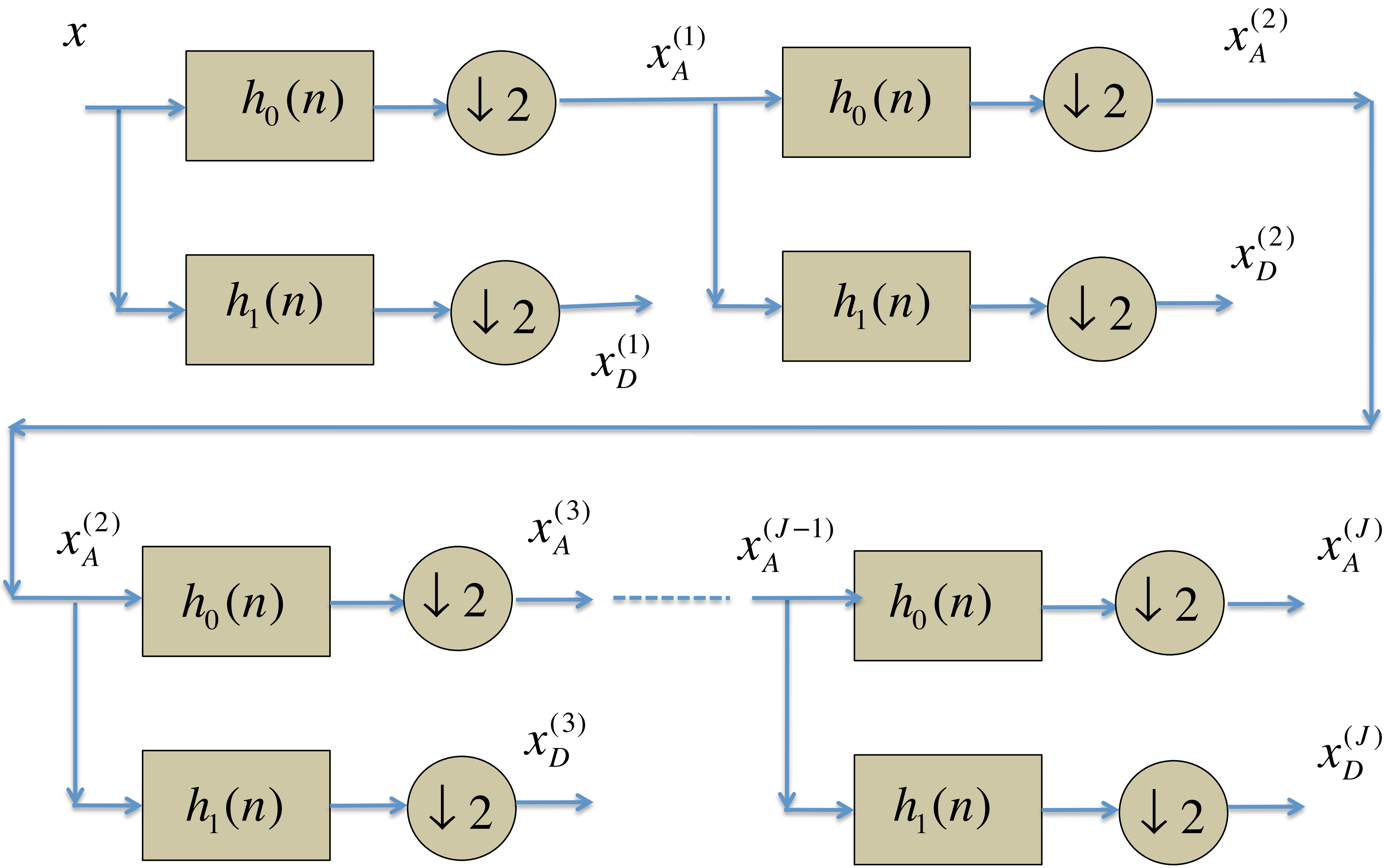
$$\Rightarrow y(z) = x(z^M)$$

Recap: Noble (??) identities



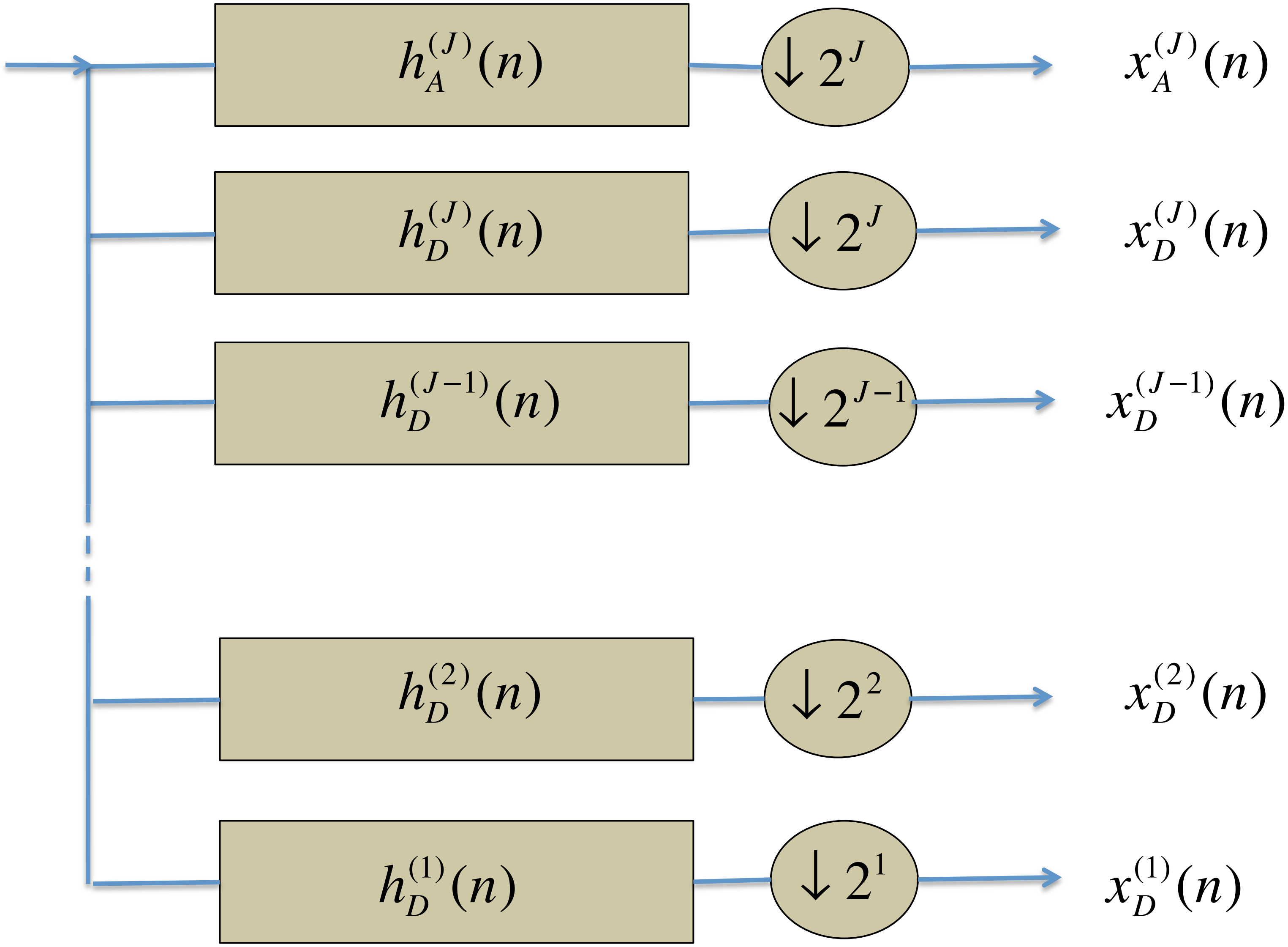


Recap – The wavelet transform:



$$\left\langle h_0(n-2l), h_0(n-2l)\right\rangle = \delta(l-m) \Rightarrow H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1}) = 1$$

$$h_1(n) = (-1)^n h_0(n)$$



$$x_A^{(J)}(n) = \sum_m h_A^{(J)}(m - 2^J n)x(m) \qquad x_D^{(j)}(n) = \sum_m h_D^{(j)}(m - 2^j n)x(m)$$

$$j = 1, \ldots, J$$

$$h_D^{(1)}(z) = h_1(z)$$

$$h_D^{(2)}(z) = h_0(z)h_1(z^2)$$

$$h_D^{(3)}(z) = h_0(z)h_0(z^2)h_1(z^4)$$

$$h_D^{(j)}(z) = h_0(z)h_0(z^2)\cdots h_0(z^{2^{j-1}})h_1(z^{2^j})$$

$$h_D^{(J)}(z) = h_0(z)h_0(z^2)\cdots h_0(z^{2^{J-1}})h_1(z^{2^J})$$

$$h_A^{(J)}(z) = h_0(z)h_0(z^2)\cdots h_0(z^{2^J})$$

W_j : vector subspace containing $x_D^{(j)}$, for $j = 1,...,J$

W_{J+1} : vector subspace containing $x_A^{(J)}$

$\{\mathbf{W}_j\}_{j=1}^{J+1}$ are matrices such that $\mathbf{x}_D^{(j)} = \mathbf{W}_j \mathbf{x}$, for $j = 1,..,J$, $\mathbf{x}_A^{(J)} = \mathbf{W}_{J+1} \mathbf{x}$,

(1) For $j = 1,..,J$, $\mathbf{W}_j$ is the derived from the convolution matrix of $h_D^{(j)}(n)$ after downsampling the rows by 2^j

(2) $\mathbf{W}_{J+1}$ is the derived from the convolution matrix of $h_A^{(J)}(n)$ after downsampling the rows by 2^J

$$\left\langle h_0(n-2l), h_0(n-2m)\right\rangle = \delta(l-m) \quad h_1(n) = (-1)^n h_0(n)$$

$$\Downarrow$$

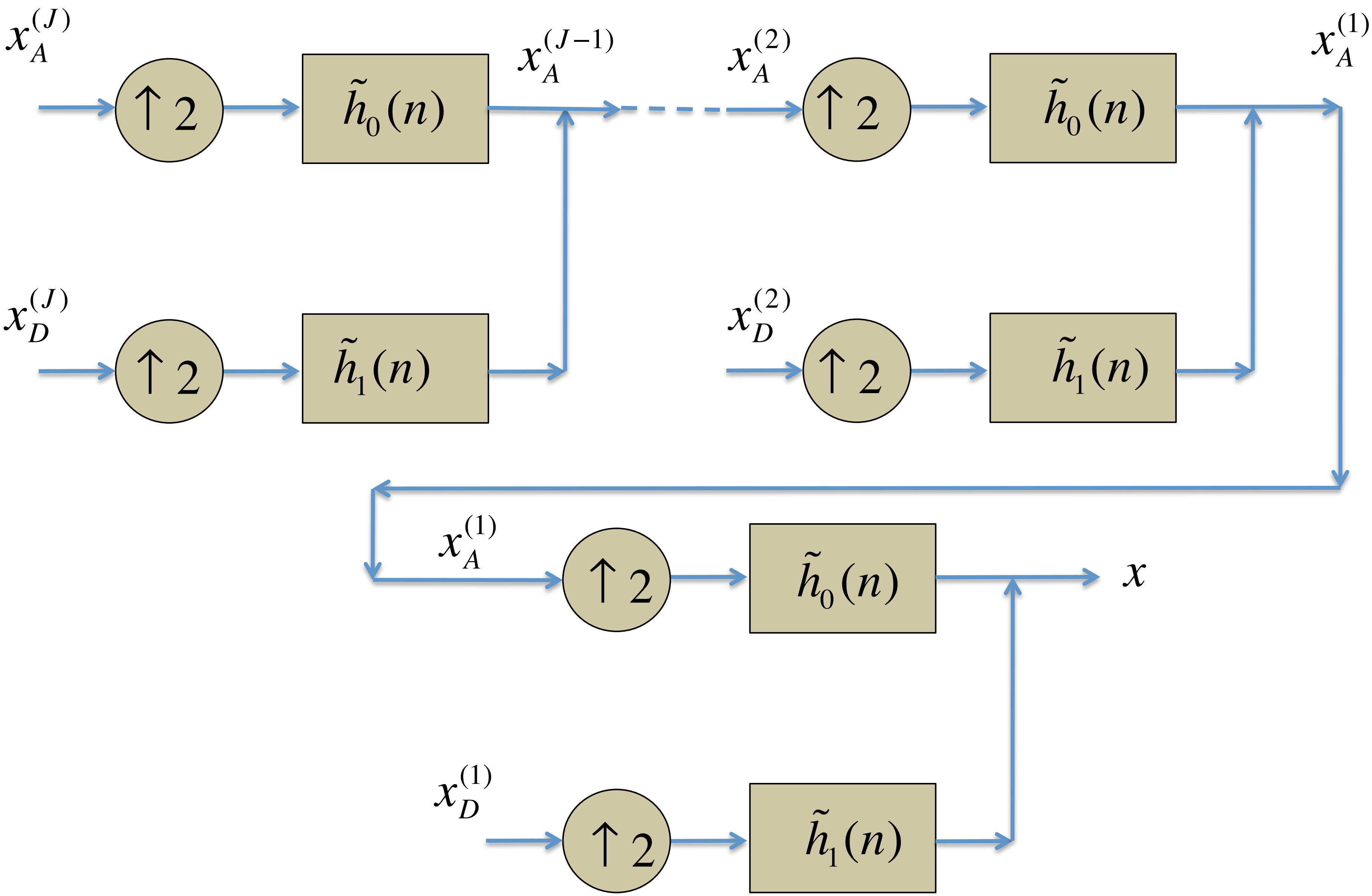
$$\left\langle h_A^{(J)}(n-2^J l), h_A^{(J)}(n-2^J m)\right\rangle = \delta(l-m)$$

$$\left\langle h_D^{(j)}(n-2^j l), h_D^{(k)}(n-2^k M)\right\rangle = \delta(j-k)\delta(l-m)$$

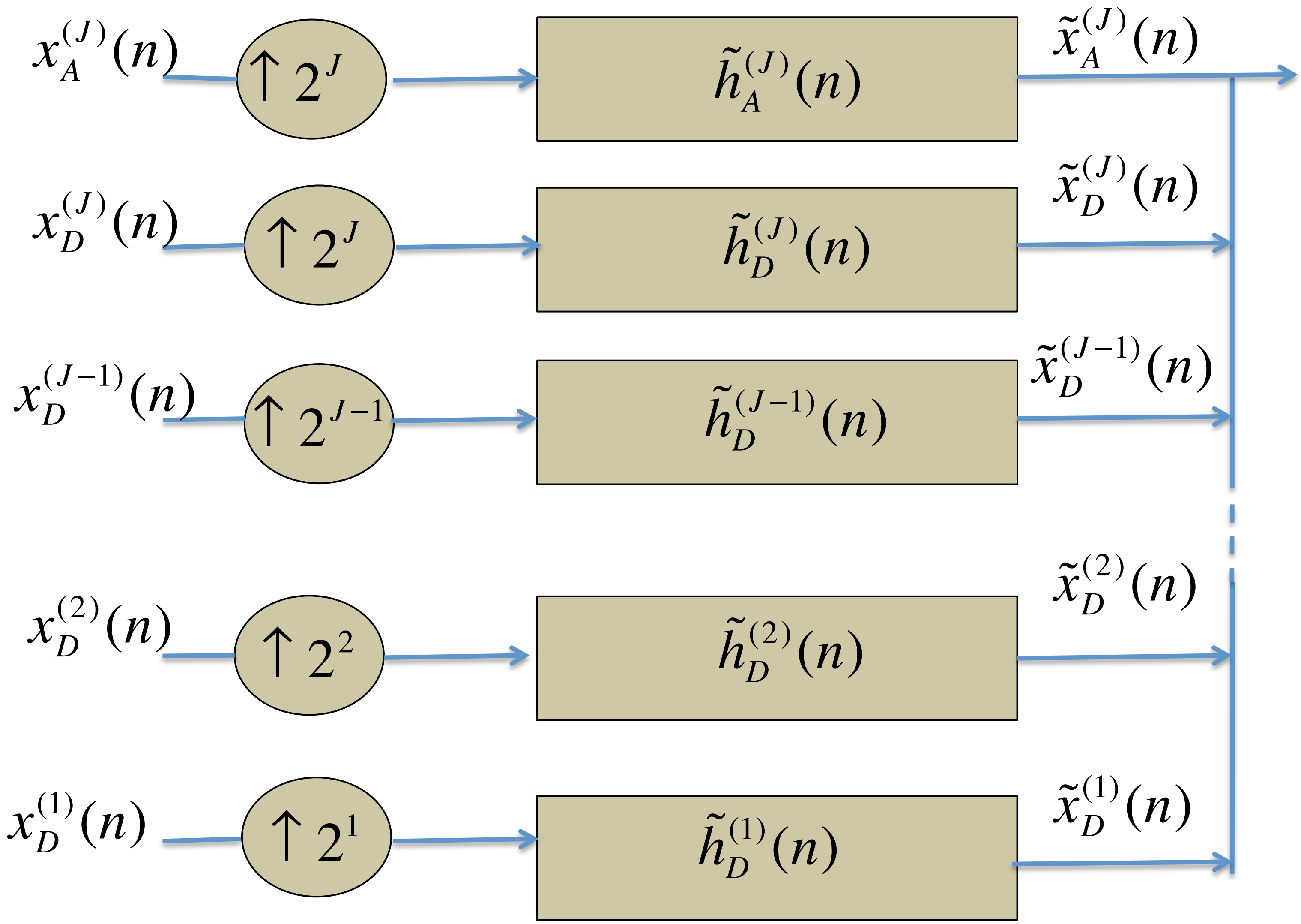
$$\left\langle h_D^{(j)}(n-2^j l), h_A^{(J)}(n-2^J m)\right\rangle = 0$$

$$\Downarrow$$

The matrix defined in the equation $\begin{bmatrix} \mathbf{x}_A^{(J)} \\ \mathbf{x}_{AD}^{(J)} \\ \mathbf{x}_D^{(J-1)} \\ \vdots \\ \mathbf{x}_D^{(2)} \\ \mathbf{x}_D^{(1)} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{W}_{J+1} \\ \mathbf{W}_J \\ \mathbf{W}_{J-1} \\ \vdots \\ \mathbf{W}_2 \\ \mathbf{W}_1 \end{bmatrix}}_{\mathbf{W}} \mathbf{x}$ is orthogonal.



$x_A^{(J)}$
$x_D^{(J)}$
$\uparrow 2$
$\tilde{h}_0(n)$
$\tilde{h}_1(n)$
$x_A^{(J-1)}$
$x_A^{(2)}$
$x_D^{(2)}$
$x_A^{(1)}$
$x_D^{(1)}$
x



$$\tilde{x}_A^{(J)}(n) = \sum_m \tilde{h}_A^{(J)}(2^J m - n) x_A^{(J)}(m)$$

$$\tilde{x}_D^{(j)}(n) = \sum_m \tilde{h}_D^{(j)}(2^j m - n) x_D^{(j)}(m)$$

$$j = 1, \ldots, J$$

$$\tilde{h}_D^{(j)}(n) = h_D^{(j)}(-n), j = 1,\ldots,J$$

$$\tilde{h}_A^{(J)}(z) = h_A^{(J)}(-n)$$

$\{\tilde{\mathbf{W}}_j\}_{j=1}^{J+1}$ are matrices such that

$$\tilde{\mathbf{x}}_D^{(j)} = \tilde{\mathbf{W}}_j \mathbf{x}_D^{(j)}, \text{ for } j = 1,..,J, \quad \tilde{\mathbf{x}}_A^{(J)} = \tilde{\mathbf{W}}_{J+1} \mathbf{x}_A^{(J)},$$

(1) For $j = 1,..,J$, $\tilde{\mathbf{W}}_j$ is the derived from the convolution matrix of $\tilde{h}_D^{(j)}(n)$ after downsampling the columns by 2^j

(2) $\tilde{\mathbf{W}}_{J+1}$ is the derived from the convolution matrix of $\tilde{h}_A^{(J)}(n)$ after downsampling the columns by 2^J

What is the relashionship between $\mathbf{W}_j$ and $\tilde{\mathbf{W}}_j$?

$$\tilde{\mathbf{W}}_j = \mathbf{W}_j^T$$

$$\left\langle \tilde{h}_0(n-2l), \tilde{h}_0(n-2m) \right\rangle = \delta(l-m) \quad \tilde{h}_1(n) = (-1)^n \tilde{h}_0(n)$$

$$\Downarrow$$

$$\left\langle \tilde{h}_A^{(J)}(n-2^J l), \tilde{h}_A^{(J)}(n-2^J m) \right\rangle = \delta(l-m)$$

$$\left\langle \tilde{h}_D^{(j)}(n-2^j l), \tilde{h}_D^{(k)}(n-2^k M) \right\rangle = \delta(j-k)\delta(l-m)$$

$$\left\langle \tilde{h}_D^{(j)}(n-2^j l), \tilde{h}_A^{(J)}(n-2^J m) \right\rangle = 0$$

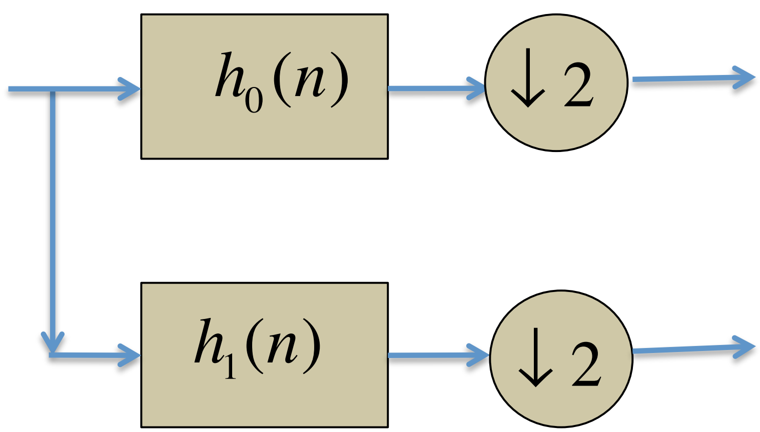
$$\Downarrow$$

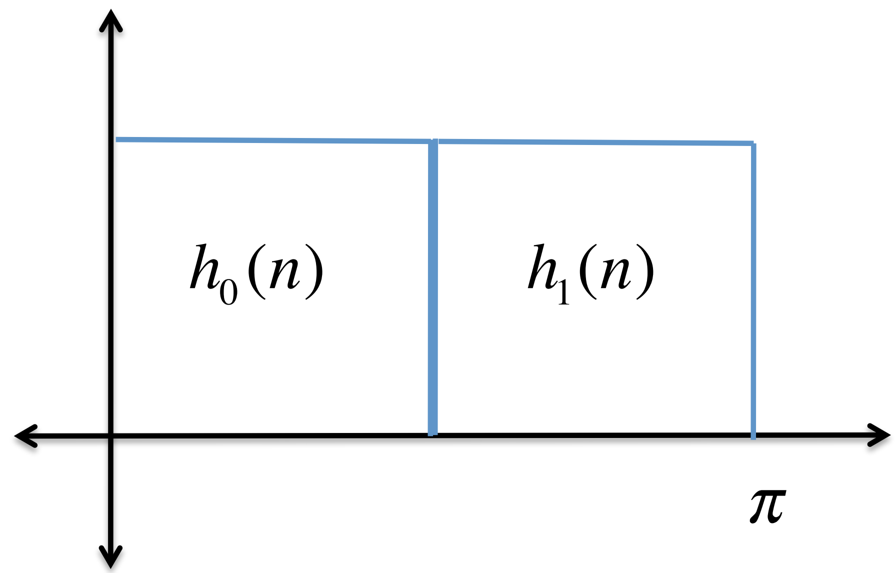
The matrix defined in the equation

$$\mathbf{x} = \underbrace{\begin{bmatrix} \tilde{\mathbf{W}}_{J+1} & \tilde{\mathbf{W}}_J & \tilde{\mathbf{W}}_{J-1} & \cdots & \tilde{\mathbf{W}}_2 & \tilde{\mathbf{W}}_1 \end{bmatrix}}_{\tilde{\mathbf{W}}} \begin{bmatrix} \mathbf{x}_A^{(J)} \\ \mathbf{x}_{AD}^{(J)} \\ \mathbf{x}_D^{(J-1)} \\ \vdots \\ \mathbf{x}_D^{(2)} \\ \mathbf{x}_D^{(1)} \end{bmatrix}$$

is orthogonal $\Rightarrow \tilde{\mathbf{W}} = \mathbf{W}^T \Rightarrow \tilde{\mathbf{W}}_j = \mathbf{W}_j^T$

Shannon Wavelet filters





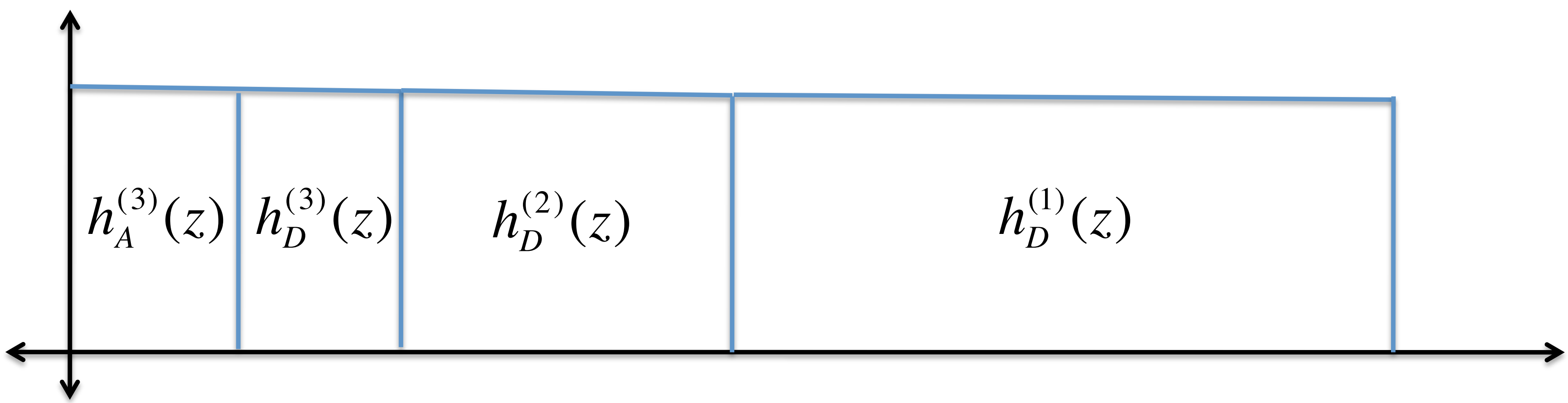
Ex: Three level decomposition

$$h_D^{(1)}(z) = h_1(z)$$

$$h_D^{(2)}(z) = h_0(z)h_1(z^2)$$

$$h_D^{(3)}(z) = h_0(z)h_0(z^2)h_1(z^4)$$

$$h_A^{(3)}(z) = h_0(z)h_0(z^2)h_0(z^4)$$



Derivation of the FTL algorithm

Projection operators

$$\mathbf{x} = \sum_{j \in S} \underbrace{\mathbf{W}_j^T \mathbf{W}_j}_{\mathbf{P}_j} \mathbf{x}.$$

$S = \{1, 2, ..., J + 1\}$

Property 3: Let $\mathbf{C}$ be a block-circulant matrix. Then, for the Shannon wavelet basis

$$\mathbf{C}\mathbf{P}_j = \mathbf{P}_j \mathbf{C}.$$

Splitting the cost function

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}\mathbf{x}\|_1.$$

Property 4: When using the Shannon wavelet basis

$$J(\mathbf{x}) = \sum_{j \in S} \|\mathbf{P}_j\mathbf{y} - \mathbf{H}\mathbf{P}_j\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}_j\mathbf{x}\|_1$$

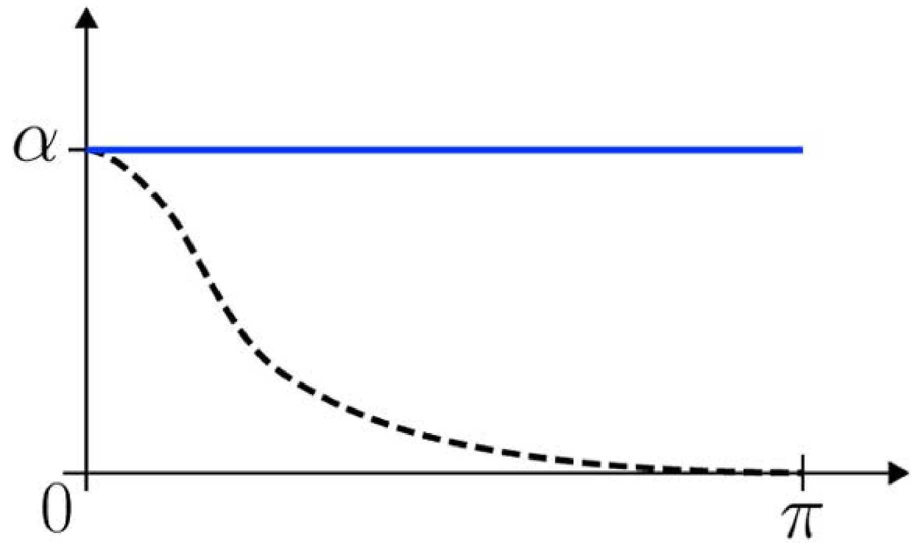
where $\mathbf{P}_j$ is the projection operator on the jth subband.

The split auxiliary function

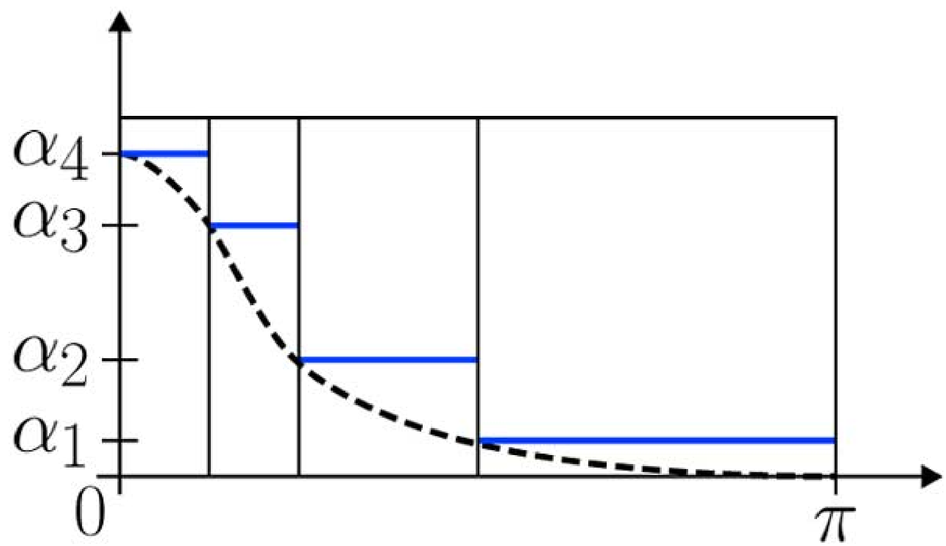
$$\tilde{J}_n(\mathbf{x}) = J(\mathbf{x}) + \sum_{j \in S} \alpha_j \left\| \mathbf{P}_j \mathbf{x}^{(n)} - \mathbf{P}_j \mathbf{x} \right\|_2^2 - \left\| \mathbf{H}\mathbf{P}_j \mathbf{x}^{(n)} - \mathbf{H}\mathbf{P}_j \mathbf{x} \right\|_2^2 \tag{10}$$

$$\alpha_j > \rho_j(\mathbf{H}^T\mathbf{H})$$

$$\text{where} \quad \rho_j(\mathbf{H}^T\mathbf{H}) = \max_{\|\mathbf{v}\|_2 = 1, \mathbf{v} \in W_j} \|\mathbf{H}\mathbf{v}\|_2^2.$$



(a)



The FTL algorithm

Property 5: Assume that $\alpha_j > \rho_j(\mathbf{H}^T\mathbf{H})$ holds for every $j \in S$ in definition (10). Then $\tilde{J}_n(\mathbf{x}) > J(\mathbf{x})$, except at $\mathbf{x} = \mathbf{x}^{(n)}$, where $\tilde{J}_n(\mathbf{x}) = J(\mathbf{x})$. Moreover, (10) is equivalent to

$$\tilde{J}_n(\mathbf{x}) = \sum_{j \in S} \alpha_j \Big\| \mathbf{P}_j \Big[\mathbf{x}^{(n)} + \alpha_j^{-1}\mathbf{H}^T \times \Big(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)}\Big)\Big] - \mathbf{P}_j\mathbf{x}\Big\|_2^2 + \lambda\|\mathbf{W}_j\mathbf{x}\|_1 + c_j$$

where the constants $(c_j)_{j\in S}$ do not depend on $\mathbf{x}$.

- Compute the Landweber iteration

$$\mathbf{z}^{(n)} = \mathbf{x}^{(n)} + \sum_{j \in S} \tau_j \mathbf{P}_j \mathbf{H}^T \left(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)} \right)$$

with subband-dependent step sizes τ_j.
- Perform the wavelet-domain denoising operation

$$\mathbf{x}^{(n+1)} = \sum_{j \in S} \mathbf{W}_j^T \mathcal{T}_{\lambda \tau_j / 2} \left\{ \mathbf{W}_j \mathbf{z}^{(n)} \right\}$$

with subband-dependent thresholds $\lambda \tau_j / 2$.