Wavelet LP Filter Design

(Revised November 5, 2012)

Robert Murray, Ph.D.
Omicron Research Institute

8.0 Contents

8.1 DWT and MODWT Wavelet Filters

The computation of the wavelet transform is accomplished by means of a recursive procedure known as the Pyramid Algorithm [PW]. The data set X_n of length *N*, where *N* must be a power of 2, is circularly filtered with the wavelet filters $h_{j,l}$ and $g_{j,l}$ to yield the wavelet coefficients $\tilde{K}_{j;\nu}$ and scaling coefficients $\tilde{J}_{j;\nu}$ [PW, pp.96-97]:

$$\tilde{K}_{j;\nu} \equiv \sum_{t=0}^{N-1} \tilde{W}_{j;\nu,t} X_t \equiv 2^{-j/2} \sum_{l=0}^{L_j-1} h_{j;l} X_{[\nu-l]\bmod N} \qquad (\nu = 0,1,\ldots,N-1)$$

$$\tilde{J}_{J;\nu} \equiv \sum_{t=0}^{N-1} \tilde{V}_{J;\nu,t} X_t \equiv 2^{-J/2} \sum_{l=0}^{L_J-1} g_{J;l} X_{[\nu-l]\bmod N} \qquad (\nu = 0,1,\ldots,N-1)$$

$$N_j \equiv N/2^j \qquad L_j \equiv \left(2^j - 1\right)\left(L-1\right)+1 \qquad \left(j = 1,\ldots,J \quad N = 2^{J_0}\right)$$

The j values for the wavelet coefficients range up to some maximum value J, and the scaling coefficients are only for this same value of J, so there are the same number of scaling coefficients as there are wavelet coefficients on the highest level, J.

Wavelet and Scaling Filters

The basic wavelet filter h_l satisfies the following definitions [PW, p.69]:

$$\sum_{l=0}^{L-1} h_l = 0, \qquad \sum_{l=0}^{L-1} h_l^2 = 1, \qquad \text{and} \qquad \sum_{l=0}^{L-1} h_l h_{l+2n} = 0 \quad (n \neq 0)$$

The last equation means that the wavelet filters are orthogonal to even shifts. The basic scaling filter g_l is the "quadrature mirror filter" corresponding to h_l [PW, p.75]:

$$g_l \equiv (-1)^{l+1} h_{L-1-l} \qquad \Leftrightarrow \qquad h_l \equiv (-1)^l g_{L-1-l}$$

Then we have:

$$\sum_{l=0}^{L-1} g_l = \pm\sqrt{2}, \qquad \sum_{l=0}^{L-1} g_l^2 = 1, \qquad \text{and} \qquad \sum_{l=0}^{L-1} g_l g_{l+2n} = 0 \quad (n \neq 0)$$

The basic wavelet filters are of length L as shown above, and are nonzero in the range $\{0,\cdots,L-1\}$. However, we may extend the range of the index to the range $\{-\infty,\cdots,+\infty\}$ and defined the *periodized* wavelet and scaling filters as follows [PW, p.32]:

$$h_l^{\circ} \equiv \sum_{n=-\infty}^{+\infty} h_{l+nN}, \qquad g_l^{\circ} \equiv \sum_{n=-\infty}^{+\infty} g_{l+nN}, \qquad l = 0,\cdots,N-1$$

Then, for example, the filtering equations can be written symbolically:

$$\tilde{K}_\nu \equiv \sum_{l=0}^{L-1} h_l X_{[\nu-l]\bmod N} = \sum_{l=-\infty}^{+\infty} h_l X_{[\nu-l]\bmod N} = \sum_{n=-\infty}^{+\infty} \sum_{l=nN}^{(n+1)N-1} h_l X_{[\nu-l]\bmod N} =$$

$$= \sum_{n=-\infty}^{+\infty} \sum_{l=0}^{N-1} h_{l+nN} X_{[\nu-l]\bmod N} = \sum_{l=0}^{N-1} \left[\sum_{n=-\infty}^{+\infty} h_{l+nN} \right] X_{[\nu-l]\bmod N} \equiv \sum_{l=0}^{N-1} h_l^{\circ} X_{[\nu-l]\bmod N}$$

There is an analogous equation for g_l°. We then have for the periodized filters:

$$\sum_{l=0}^{N-1} h_l^\circ h_{l+2n \bmod N}^\circ = \sum_{l=0}^{N-1} g_l^\circ g_{l+2n \bmod N}^\circ = \begin{cases} 1 & \text{if} \quad n = 0 \\ 0 & \text{if} \quad n = 1, 2, \cdots, \dfrac{N}{2} - 1 \end{cases}$$

$$\sum_{l=0}^{N-1} h_l^\circ g_{l+2n \bmod N}^\circ = \sum_{l=0}^{N-1} g_l^\circ h_{l+2n \bmod N}^\circ = 0 \quad \text{for} \quad n = 0, 1, 2, \cdots, \frac{N}{2} - 1$$

This once again shows the orthogonality of the periodized wavelet and scaling filters, and their orthogonality to even shifts.

The wavelet filter $h_{j;l}$ for level *j* is formed by convolving the basic wavelet and scaling filters h_l and g_l in a certain way [PW, p.102]. A recursive definition for these is given by the following two different methods:

$$h_{j;l} \equiv \sum_{k=0}^{L-1} h_k g_{j-1;l-2^{j-1}k} \qquad g_{j;l} \equiv \sum_{k=0}^{L-1} g_k g_{j-1;l-2^{j-1}k} \qquad \left(l = 0, \ldots, L_j - 1\right)$$

$$h_{j;l} \equiv \sum_{k=0}^{L_{j-1}-1} g_{l-2k} h_{j-1;k} \qquad g_{j;l} \equiv \sum_{k=0}^{L_{j-1}-1} g_{l-2k} g_{j-1;k} \qquad L_j \equiv \left(2^j - 1\right)\left(L - 1\right) + 1$$

(Note: $h_{1;l} \equiv h_l$ and $g_{1;l} \equiv g_l$ are the wavelet and scaling filters.)

Here, the index *j* ranges from 1 to *J*, while the index *l* ranges from 0 to $L_j - 1$. Notice that the wavelet coefficients are obtained from the time series vector by a procedure of filtering by the basic wavelet and scaling filters followed by *down sampling*, as can be seen from the factors of 2^j in front of the summation index. Conversely, the inverse (transposed) transformation from wavelet coefficients to time series vector involves an *up sampling* procedure [PW]. We merely denote this latter procedure by the transpose of the former, in view of the fact that these are actually orthogonal matrices.

DWT Wavelet Filter

For the DWT filter the wavelet coefficients $K_{j;\nu}$ and scaling coefficients $J_{j;\nu}$ are obtained by down-sampling $\tilde{K}_{j;\nu}$ and $\tilde{J}_{j;\nu}$ *j* times:

$$K_{j;\nu} \equiv 2^{j/2} \tilde{K}_{j;2^j(\nu+1)-1} \qquad J_{j;\nu} \equiv 2^{j/2} \tilde{J}_{j;2^j(\nu+1)-1}$$

We must therefore also define:

$$W_{j;\nu,t} \equiv \tilde{W}_{j;2^j(\nu+1)-1,t} \qquad V_{j;\nu,t} \equiv \tilde{V}_{j;2^j(\nu+1)-1,t}$$

The wavelet coefficients for a given $N : t = 0, 1, \ldots, N-1$ can then be written as:

$$K_{j;\nu} \equiv \sum_{t=0}^{N-1} W_{j;\nu,t} X_t \equiv \sum_{l=0}^{L_j-1} h_{j;l} X_{\left[2^j(\nu+1)-1-l\right] \bmod N} \qquad \left(\nu = 0,1,\ldots,N_j - 1\right)$$
$$J_{J;\nu} \equiv \sum_{t=0}^{N-1} V_{J;\nu,t} X_t \equiv \sum_{l=0}^{L_j-1} g_{J;l} X_{\left[2^J(\nu+1)-1-l\right] \bmod N} \qquad \left(\nu = 0,1,\ldots,N_J - 1\right)$$
$$N_j \equiv N/2^j \qquad L_j \equiv \left(2^j - 1\right)\left(L-1\right)+1 \qquad \left(j = 1,\ldots,J \quad N = 2^{J_0}\right)$$

This is equivalent to filtering with the periodized filters [PW, p.96]:

$$K_{j;\nu} \equiv \sum_{l=0}^{N-1} h^{\circ}_{j;l} X_{\left[2^j(\nu+1)-1-l\right] \bmod N} \quad \left(h^{\circ}_{j;l} \equiv \text{periodized DWT wavelet filter}\right)$$
$$J_{J;\nu} \equiv \sum_{l=0}^{N-1} g^{\circ}_{J;l} X_{\left[2^j(\nu+1)-1-l\right] \bmod N} \quad \left(g^{\circ}_{J;l} \equiv \text{periodized DWT scaling filter}\right)$$

Let us now define a new summation variable:

$$t \equiv \left[2^j(\nu+1)-1-l\right] \bmod N \quad \Rightarrow \quad l \equiv \left[2^j(\nu+1)-1-t\right] \bmod N$$

Thus, $\bmod N$, in terms of this new variable we have:

$$K_{j;\nu} \equiv \sum_{t=0}^{N-1} W_{j;\nu,t} X_t \equiv \sum_{t=0}^{N-1} h^{\circ}_{j;\left[2^j(\nu+1)-1-t\right] \bmod N} X_t \qquad \left(\nu = 0,1,\ldots,N_j - 1\right)$$
$$J_{J;\nu} \equiv \sum_{t=0}^{N-1} V_{J;\nu,t} X_t \equiv \sum_{t=0}^{N-1} g^{\circ}_{J;\left[2^J(\nu+1)-1-t\right] \bmod N} X_t \qquad \left(\nu = 0,1,\ldots,N_J - 1\right)$$

Thus we find the following expressions for the DWT matrices:

$$W_{j;\nu,t} \equiv h^{\circ}_{j;\left[2^j(\nu+1)-1-t\right] \bmod N} \qquad \left(\nu = 0,1,\ldots,N_j - 1, t = 0,1,\ldots,N\right)$$
$$V_{J;\nu,t} \equiv g^{\circ}_{J;\left[2^J(\nu+1)-1-t\right] \bmod N} \qquad \left(\nu = 0,1,\ldots,N_J - 1, t = 0,1,\ldots,N\right)$$

The set of all wavelet and scaling coefficients for the index set $\left(j;\nu \text{ and } t\right)$ form an orthogonal matrix. Each set of indices has *N* values in all. The highest value of *j* is denoted *J*, with $J \le J_0$. If this highest value is J_0, then it can be seen that $N_{J_0} = 1$ and there is only one scaling coefficient. Otherwise there are $N_J = 2^{J_0 - J}$ scaling coefficients and an equal number of wavelet coefficients on the highest level. In this case the total number of wavelet and scaling coefficients can be seen to be:

$$\sum_{j=1}^{J} \frac{N}{2^j} + 2^{J_0 - J} = 2^{J_0}\left[\sum_{j=1}^{J}\left(\frac{1}{2}\right)^j + 2^{-J}\right] = 2^{J_0}\left[\left(1 - \frac{1}{2^J}\right)\sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^j + 2^{-J}\right] = 2^{J_0} = N$$

Thus for any value of *J* the total number of wavelet plus scaling coefficients is *N*.

MODWT Wavelet Filter

Let us consider the MODWT wavelet transform. These are the same definitions as given above. At level *j*, the wavelet and scaling coefficients are given by:

$$\tilde{K}_{j;\nu} \equiv \sum_{t=0}^{N-1} \tilde{W}_{j;\nu,t} X_t \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j;l} X_{[\nu-l] \bmod N} \qquad (\nu = 0,1,\ldots,N-1)$$

$$\tilde{J}_{J;\nu} \equiv \sum_{t=0}^{N-1} \tilde{V}_{J;\nu,t} X_t \equiv \sum_{l=0}^{L_j-1} \tilde{g}_{J;l} X_{[\nu-l] \bmod N} \qquad (\nu = 0,1,\ldots,N-1)$$

$$L_j \equiv \left(2^j - 1\right)\left(L-1\right)+1 \qquad \left(j = 1,\ldots,J \quad N = 2^{J_0}\right)$$

The MODWT filters are defined in terms of the DWT filters by:

$$\tilde{h}_{j,l} = 2^{-j/2} h_{j,l} \qquad \tilde{g}_{j,l} = 2^{-j/2} g_{j,l}$$

This is equivalent to filtering with the periodized filters [PW, p.170]:

$$\tilde{K}_{j;\nu} \equiv \sum_{l=0}^{N-1} \tilde{h}^{\circ}_{j;l} X_{[\nu-l] \bmod N} \qquad \left(\tilde{h}^{\circ}_{j;l} \equiv \text{periodized MODWT wavelet filter}\right)$$

$$\tilde{J}_{J;\nu} \equiv \sum_{l=0}^{N-1} \tilde{g}^{\circ}_{J;l} X_{[\nu-l] \bmod N} \qquad \left(\tilde{g}^{\circ}_{J;l} \equiv \text{periodized MODWT scaling filter}\right)$$

Note that the MODWT wavelet coefficient vector $\tilde{K}_{j;\nu}$ and the scaling coefficient vector $\tilde{J}_{J;\nu}$ are both *N*-dimensional vectors. Let us now define a new summation variable:

$$t \equiv \left[\nu - l\right] \bmod N \quad \Rightarrow \quad l \equiv \left[\nu - t\right] \bmod N$$

Thus, $\bmod N$, in terms of this new variable we have:

$$\tilde{K}_{j;\nu} \equiv \sum_{t=0}^{N-1} \tilde{W}_{j;\nu,t} X_t \equiv \sum_{t=0}^{N-1} \tilde{h}^{\circ}_{j;[\nu-t] \bmod N} X_t \qquad (\nu = 0,1,\ldots,N-1)$$

$$\tilde{J}_{J;\nu} \equiv \sum_{t=0}^{N-1} \tilde{V}_{J;\nu,t} X_t \equiv \sum_{t=0}^{N-1} \tilde{g}^{\circ}_{J;[\nu-t] \bmod N} X_t \qquad (\nu = 0,1,\ldots,N-1)$$

Thus we find the following expressions for the MODWT matrices:

$$\tilde{W}_{j;\nu,t} \equiv \tilde{h}^{\circ}_{j;[\nu-t] \bmod N} \qquad (\nu = 0,1,\ldots,N-1, t = 0,1,\ldots,N)$$

$$\tilde{V}_{J;\nu,t} \equiv \tilde{g}^{\circ}_{J;[\nu-t] \bmod N} \qquad (\nu = 0,1,\ldots,N-1, t = 0,1,\ldots,N)$$

In this case the set of all wavelet and scaling coefficients for the index set $\left(j;\nu \text{ and } t\right)$ do not form an orthogonal matrix – they are an over-complete set.

Phase Lag of LA(8) Filter

The DWT and MODWT wavelet transforms introduce a phase lag, which are given as follows [PW, pp.114-115]. The wavelet coefficients at a given time index are associated with the original time series at a shifted index, where the shift is given by $\nu_j^{(H)}$ for the wavelet filter and by $\nu_J^{(G)}$ for the scaling filter. To achieve approximately zero phase for the DWT filter, we must associate:

$$K_{j;\nu} \equiv \sum_{l=0}^{L_j-1} h_{j;l} X_{\left[2^j(\nu+1)-1-l\right]\bmod N} \quad \Rightarrow \quad K_{j;\nu} \sim X_{\left[2^j(\nu+1)-1-\left|\nu_j^{(H)}\right|\right]\bmod N}$$

$$J_{J;\nu} \equiv \sum_{l=0}^{L_j-1} g_{J;l} X_{\left[2^J(\nu+1)-1-l\right]\bmod N} \quad \Rightarrow \quad J_{J;\nu} \sim X_{\left[2^J(\nu+1)-1-\left|\nu_J^{(G)}\right|\right]\bmod N}$$

To achieve approximately zero phase for the MODWT filter, we must associate:

$$\tilde{K}_{j;\nu} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j;l} X_{[\nu-l]\bmod N} \quad \Rightarrow \quad \tilde{K}_{j;\nu} \sim X_{\left[\nu-\left|\nu_j^{(H)}\right|\right]\bmod N}$$

$$\tilde{J}_{J;\nu} \equiv \sum_{l=0}^{L_j-1} \tilde{g}_{J;l} X_{[\nu-l]\bmod N} \quad \Rightarrow \quad \tilde{J}_{J;\nu} \sim X_{\left[\nu-\left|\nu_J^{(G)}\right|\right]\bmod N}$$

For the LA(8) wavelet filter, the phase shifts for the wavelet levels are given by:

$$\text{LA(8):} \quad \nu_j^{(H)} \equiv -\left|\nu_j^{(H)}\right| = -\frac{L_j}{2} = -\frac{\left(2^j-1\right)\left(L-1\right)+1}{2} = -\frac{7}{2}\cdot 2^j + 3$$

For the LA(8) scaling filter, the phase shift for the scaling level is:

$$\text{LA(8):} \quad \nu_J^{(G)} \equiv -\left|\nu_J^{(G)}\right| = -\frac{\left(L_J-1\right)\left(L-2\right)}{2\left(L-1\right)} = -\frac{\left(2^J-1\right)\left(L-2\right)}{2} = -\frac{6}{2}\cdot 2^J + 3$$

(Note that for the top two levels, the phase shift is greater than the data width.) This indicates the amount by which the wavelet coefficients must be shifted (to the left) to bring them into correspondence with the data elements. For the LA(8) filter and $J = 8$ we find the absolute value of these phase shifts given by:

$$\left\{\left|\nu_j^{(H)}\right|, \left|\nu_J^{(G)}\right|\right\} = \{4, 11, 25, 53, 109, 221, 445, 893, 765\} \qquad \left(j = 1, \ldots, J; \quad J = 8\right)$$

These phase shifts also affect the degree by which the "wrap-around" affects the circular filtering, related to the width of the wavelet and scaling filters.

8.2 *DWT and MODWT Filter Kernels*

Definition of Filter Kernels

We have found the following expressions for the DWT matrices:

$$W_{j;\nu,n} \equiv h^{\circ}_{j;\left[2^j(\nu+1)-1-n\right]\bmod N} \qquad \left(\nu = 0,1,\ldots,N_j-1,\ n = 0,1,\ldots,N\right)$$

$$V_{J;\nu,n} \equiv g^{\circ}_{J;\left[2^J(\nu+1)-1-n\right]\bmod N} \qquad \left(\nu = 0,1,\ldots,N_J-1,\ n = 0,1,\ldots,N\right)$$

Let us define quantities that we call the DWT filter kernels, as follows:

$$F_{j;n,n'} \equiv \sum_{\nu=0}^{N_j} W^T_{j;n,\nu} W_{j;\nu,n'} = \sum_{\nu=0}^{N_j} h^{\circ}_{j;\left[2^j(\nu+1)-1-n\right]\bmod N}\, h^{\circ}_{j;\left[2^j(\nu+1)-1-n'\right]\bmod N}$$

$$G_{J;n,n'} \equiv \sum_{\nu=0}^{N_J} V^T_{J;n,\nu} V_{J;\nu,n'} = \sum_{\nu=0}^{N_J} g^{\circ}_{J;\left[2^J(\nu+1)-1-n\right]\bmod N}\, g^{\circ}_{J;\left[2^J(\nu+1)-1-n'\right]\bmod N}$$

Note that the sum over the index in the DWT filter kernel is only nonzero when the index is within the range 0 to L_j–1, corresponding to the width of the wavelet filter.

We have found the following expressions for the MODWT matrices:

$$\tilde{W}_{j;\nu,n} \equiv \tilde{h}^{\circ}_{j;[\nu-n]\bmod N} \qquad \left(\nu = 0,1,\ldots,N-1,\ n = 0,1,\ldots,N\right)$$

$$\tilde{V}_{J;\nu,n} \equiv \tilde{g}^{\circ}_{J;[\nu-n]\bmod N} \qquad \left(\nu = 0,1,\ldots,N-1,\ n = 0,1,\ldots,N\right)$$

Let us then define the MODWT filter kernels, as follows:

$$\tilde{F}_{j;n,n'} \equiv \sum_{\nu=0}^{N} \tilde{W}^T_{j;n,\nu} \tilde{W}_{j;\nu,n'} = \sum_{\nu=0}^{N} \tilde{h}^{\circ}_{j;[\nu-n]\bmod N}\, \tilde{h}^{\circ}_{j;[\nu-n']\bmod N}$$

$$\tilde{G}_{J;n,n'} \equiv \sum_{\nu=0}^{N} \tilde{V}^T_{J;n,\nu} \tilde{V}_{J;\nu,n'} = \sum_{\nu=0}^{N} \tilde{g}^{\circ}_{J;[\nu-n]\bmod N}\, \tilde{g}^{\circ}_{J;[\nu-n']\bmod N}$$

These correspond to an averaging of the DWT filter kernels over all circular shifts [PW, p.204]. As a consequence, in the MODWT basis the filter kernels are invariant under circular shifts, unlike the DWT basis, and so the filter kernels in the MODWT basis are zero-phase filters. They are also translation invariant. Since the DWT MRA is *not* invariant under circular shifts, this shows that the DWT filter kernels are *not* translation invariant.

Note that the DWT matrix in the filter kernel can be calculated using the DWT Pyramid Algorithm. Starting with an N-dimensional vector and calculating the DWT, the Pyramid Algorithm computes each row of the DWT matrix by setting successive elements the vector to unity. This must then be done N times in order to find the whole matrix, once for each value of the index. Likewise, the MODWT matrix may be found using the MODWT Pyramid Algorithm. An equivalent procedure would be to find all $J+1$ of the wavelet and smooth filters directly by the recursive procedure, and save these in $J+1$ vectors. Then the DWT and MODWT matrix and filter kernels can be found by direct multiplication as given above. Or, the DWT and MODWT filter kernels can be found directly by running the DWT and MODWT Multi-Resolution Analysis a total of N times, with each run giving a row of the matrix.

DWT Filter

In matrix notation the linear DWT transformation may be written:

$$\mathbf{K}_j = [W_j]\mathbf{X} \qquad \mathbf{J}_J = [V_J]\mathbf{X}$$

The matrix $[W_j]$ may be written in terms of elementary DWT matrices:

$$[W_j] \equiv [B_j][A_{j-1}]\cdots[A_1]$$

It is an $N_j \times N$ matrix satisfying:

$$[W_j][W_j]^T = I_{N_j}$$

The vector $\mathbf{K}_j$ is of length N_j, and we define a vector $\mathbf{D}_j$ of length N called the DWT wavelet detail, by:

$$\mathbf{D}_j \equiv [W_j]^T[W_j]\mathbf{X} \equiv [F_j]\mathbf{X} \equiv [W_j]^T\mathbf{K}_j$$

Similarly, the matrix $[V_J]$ may be written in terms of elementary DWT matrices:

$$[V_J] \equiv [A_J][A_{J-1}]\cdots[A_1]$$

It is an $N_J \times N$ matrix satisfying:

$$[V_J][V_J]^T = I_{N_J}$$

The vector $\mathbf{J}_J$ is of length N_J, and we define a vector $\mathbf{S}_J$ of length N called the DWT wavelet smooth, by:

$$\mathbf{S}_J \equiv [V_J]^T[V_J]\mathbf{X} \equiv [G_J]\mathbf{X} \equiv [V_J]^T\mathbf{J}_J$$

We also have the following orthogonality relation:

$$[W_J][V_J]^T = [V_J][W_J]^T = 0_{N_J}$$

Using these orthogonality relations, we may prove the following Multi-Resolution Analysis (MRA) [PW, p.104]:

$$\mathbf{X} = \sum_{j=1}^{J} \mathbf{D}_j + \mathbf{S}_J = \sum_{j=1}^{J} [F_j]\mathbf{X} + [G_J]\mathbf{X}$$

Thus we have the following identity:

$$\sum_{j=1}^{J} [W_j]^T [W_j] + [V_J]^T [V_J] \equiv \sum_{j=1}^{J} [F_j] + [G_J] = I_N$$

From the above we find the following "energy decomposition" of the MRA:

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J} \|\mathbf{K}_j\|^2 + \|\mathbf{J}_J\|^2 = \sum_{j=1}^{J} \|\mathbf{D}_j\|^2 + \|\mathbf{S}_J\|^2$$

The first expression is a sum over all levels of the wavelet variances, while the second is the sum over all levels of the "energy" of the MRA levels, and for the DWT these two sums are equivalent. The DWT MRA thus satisfies the Analysis of Variance (ANOVA), indicated by the equality of the first expression above with the second.

We also have the following identities:

$$[W_{j'}]\mathbf{X} \equiv \mathbf{K}_{j'} = \sum_{j=1}^{J} [W_{j'}][W_j]^T \mathbf{K}_j + [W_{j'}][V_J]^T \mathbf{J}_J$$

$$[V_J]\mathbf{X} \equiv \mathbf{J}_J = \sum_{j=1}^{J} [V_J][W_j]^T \mathbf{K}_j + [V_J][V_J]^T \mathbf{J}_J$$

Thus we find the following orthogonality relations, since we know that all the wavelet and smooth coefficient vectors form a linearly independent set:

$$[W_{j'}][W_j]^T = \delta_{jj'} I_{N_j}^{(j)} \quad [V_J][V_J]^T = I_{N_J}^{(J+1)}$$

$$[W_{j'}][V_J]^T = [V_J][W_j]^T = 0$$

Here, the matrix $I_{N_j}^{(j)}$ means the identity matrix in the N_j-dimensional space of the wavelet coefficient vectors $\mathbf{K}_j$, while the matrix $I_{N_J}^{(J+1)}$ means the identity matrix in the N_J-dimensional space of the smooth coefficient vector $\mathbf{J}_J$. Note that the complete space of the DWT wavelet coefficients is *N*-dimensional, and is the direct sum of all the N_j-dimensional and N_J-dimensional subspaces of the wavelet and smooth coefficient vectors. Hence we have the following "decomposition of unity":

$$\sum_{j=1}^{J}[W_j][W_j]^T + [V_J][V_J]^T = \sum_{j=1}^{J} I_{N_j}^{(j)} + I_{N_J}^{(J)} = I_N^{(\mathrm{all})}$$

This gives the orthogonality relation for the DWT matrices in their *N*-dimensional space.

MODWT Filter

In matrix notation the linear MODWT transformation may be written [PW, p.171,201]:

$$\tilde{\mathbf{K}}_j = [\tilde{W}_j]\mathbf{X} \qquad \tilde{\mathbf{J}}_J = [\tilde{V}_J]\mathbf{X}$$

The vector $\tilde{\mathbf{K}}_j$ is of length *N*, and we define a vector $\tilde{\mathbf{D}}_j$ of length *N* called the MODWT wavelet detail, by:

$$\tilde{\mathbf{D}}_j \equiv [\tilde{W}_j]^T[\tilde{W}_j]\mathbf{X} \equiv [\tilde{F}_j]\mathbf{X} \equiv [\tilde{W}_j]^T\tilde{\mathbf{K}}_j$$

The vector $\tilde{\mathbf{J}}_J$ is of length *N*, and we define a vector $\tilde{\mathbf{S}}_J$ of length *N* called the MODWT wavelet smooth, by:

$$\tilde{\mathbf{S}}_J \equiv [\tilde{V}_J]^T[\tilde{V}_J]\mathbf{X} \equiv [\tilde{G}_J]\mathbf{X} \equiv [\tilde{V}_J]^T\tilde{\mathbf{J}}_J$$

Using these relations, we may prove the following Multi-Resolution Analysis (MRA) [PW, p.169] for the MODWT, which follows from the properties of the periodized MODWT filters:

$$\mathbf{X} = \sum_{j=1}^{J}\tilde{\mathbf{D}}_j + \tilde{\mathbf{S}}_J = \sum_{j=1}^{J}[\tilde{F}_j]\mathbf{X} + [\tilde{G}_J]\mathbf{X}$$

Thus we have the following identity:

$$\sum_{j=1}^{J}[\tilde{W}_j]^T[\tilde{W}_j] + [\tilde{V}_J]^T[\tilde{V}_J] \equiv \sum_{j=1}^{J}[\tilde{F}_j] + [\tilde{G}_J] = I_N$$

Likewise, the MODWT decomposition of variance is given by:

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J}\left\|\tilde{\mathbf{K}}_j\right\|^2 + \left\|\tilde{\mathbf{J}}_J\right\|^2 \neq \sum_{j=1}^{J}\left\|\tilde{\mathbf{D}}_j\right\|^2 + \left\|\tilde{\mathbf{S}}_J\right\|^2$$

However, unlike the DWT, the MODWT details and smooths do not form an ANOVA.

We also have the following identities:

$$[\tilde{W}_{j'}]\mathbf{X} \equiv \tilde{\mathbf{K}}_{j'} = \sum_{j=1}^{J} [\tilde{W}_{j'}][\tilde{W}_j]^T \tilde{\mathbf{K}}_j + [\tilde{W}_{j'}][\tilde{V}_J]^T \tilde{\mathbf{J}}_J$$
$$[\tilde{V}_J]\mathbf{X} \equiv \tilde{\mathbf{J}}_J = \sum_{j=1}^{J} [\tilde{V}_J][\tilde{W}_j]^T \tilde{\mathbf{K}}_j + [\tilde{V}_J][\tilde{V}_J]^T \tilde{\mathbf{J}}_J$$

Thus we find the following orthogonality relations, assuming all the wavelet and smooth coefficient vectors on different levels are orthogonal:

$$[\tilde{W}_{j'}][\tilde{W}_j]^T \approx \delta_{jj'} I_N^{(j)} \quad [\tilde{V}_J][\tilde{V}_J]^T \approx I_N^{(J+1)}$$
$$[\tilde{W}_{j'}][\tilde{V}_J]^T = [\tilde{V}_J][\tilde{W}_j]^T = 0$$

Here, the matrix $I_N^{(j)}$ means the identity matrix in the N-dimensional space of the wavelet coefficient vectors $\tilde{\mathbf{K}}_j$, while the matrix $I_N^{(J+1)}$ means the identity matrix in the N-dimensional space of the smooth coefficient vector $\tilde{\mathbf{J}}_J$. Note that the complete space of the MODWT wavelet coefficients is $(J+1)N$ - dimensional, and is the direct sum of all the N-dimensional subspaces of the wavelet and smooth coefficient vectors. Hence we have the following "decomposition of unity":

$$\sum_{j=1}^{J} [\tilde{W}_j][\tilde{W}_j]^T + [\tilde{V}_J][\tilde{V}_J]^T \approx \sum_{j=1}^{J} I_N^{(j)} + I_N^{(J+1)} \approx I_{(J+1)N}^{(\text{all})}$$

This gives a (pseudo-)orthogonality relation for the MODWT matrices in their $(J+1)N$ - dimensional space, as compared to orthogonality for the DWT matrices in their N-dimensional space.

However, the orthogonality cannot be completely correct, because the MODWT wavelet and scaling coefficient vectors form an over-complete set, hence they cannot be orthogonal, at least not within the same level. All we can say is that on each level, the above matrices must be linearly equivalent (in the original N-dimensional vector space) to the unit matrix acting on the MODWT wavelet and scaling coefficient vectors, since these vectors do not form a linearly independent set. Hence the right-hand side above cannot in general be the unit matrix, and the transpose of the MODWT matrix must be described as the "Moore-Penrose generalized inverse".

Orthogonality of Filter Kernels

Let us make the following definitions for convenience:

$$G_{J;n,n'} \equiv F_{J+1;n,n'} \qquad V_{J;\nu,n'} \equiv W_{J+1;\nu,n'}$$

Then, due to the orthogonality of the DWT, we should have:

$$\begin{aligned}
\sum_{n''=0}^{N-1} F_{j;n,n''} F_{j';n'',n'} &\equiv \sum_{n''=0}^{N-1} \left[\sum_{\nu=0}^{N_j} W^T_{j;n,\nu} W_{j;\nu,n''} \right] \left[\sum_{\nu'=0}^{N_{j'}} W^T_{j';n'',\nu'} W_{j';\nu',n'} \right] \\
&= \sum_{\nu=0}^{N_j} \sum_{\nu'=0}^{N_{j'}} W^T_{j;n,\nu} \left[\sum_{n''=0}^{N-1} W_{j;\nu,n''} W^T_{j';n'',\nu'} \right] W_{j';\nu',n'} \\
&= \sum_{\nu=0}^{N_j} \sum_{\nu'=0}^{N_{j'}} W^T_{j;n,\nu} \left[\delta_{jj'} \delta_{\nu\nu'} \right] W_{j';\nu',n'} \\
&= \delta_{jj'} \left[\sum_{\nu=0}^{N_j} W^T_{j;n,\nu} W_{j;\nu,n'} \right] = \delta_{jj'} F_{j;n,n'}
\end{aligned}$$

Therefore the product of two DWT filter kernels is equal to another DWT filter kernel of the same level. This product rule indicates that the filter kernels are actually projection operators onto each DWT level. This would generally be the case if they were ideal bandpass filters. However, also due to orthogonality, the sum over j of the filter kernels (taking into account the above definition) gives a delta function:

$$\sum_{j=1}^{J+1} \sum_{n''=0}^{N-1} F_{j;n,n''} F_{j;n'',n'} = \sum_{j=1}^{J+1} F_{j;n,n'} \equiv \sum_{j=1}^{J} F_{j;n,n'} + G_{J;n,n'} = \delta_{nn'}$$

This gives us a means of easily inverting a covariance matrix expressed in terms of the filter kernel. This is because when the covariance matrix is expressed in terms of the DWT filter kernel, it is automatically in SVD form.

For example, we may define a stationary covariance matrix by multiplying the DWT filter kernels by the level average wavelet variance, and we may define an inverse covariance matrix by multiplying the DWT filter kernel by the inverse level average wavelet variance. Then we have:

$$\sum_{n''=0}^{N-1} \left[\sum_{j=1}^{J+1} \left\langle K_j^2 \right\rangle^{-1} F_{j;n,n''} \right] \left[\sum_{k=1}^{J+1} \left\langle K_k^2 \right\rangle^{+1} F_{k;n'',n'} \right] = \sum_{j=1}^{J+1} \sum_{k=1}^{J+1} \frac{\left\langle K_k^2 \right\rangle}{\left\langle K_j^2 \right\rangle} \delta_{jk} F_{j;n,n'} = \sum_{j=1}^{J+1} F_{j;n,n'} = \delta_{nn'}$$

Thus the covariance matrix can easily be inverted now, at least if we are allowed to sum over all n, including future as well as past values.

However, it should be noted that the above construction only works for the DWT filter kernels, because only in the DWT case is the wavelet transform matrix orthogonal. This then leads directly to the Singular Value Decomposition form given above for the covariance matrix, in terms of the orthogonal DWT wavelet matrices. The crucial step in the product rule is the use of the orthogonality relation for the DWT matrices in the above calculation. This step would not apply to the MODWT matrices, since the MODWT matrices are not orthogonal, and the MODWT filter kernels evidently do not obey the above product rule. So this product rule cannot be applied to inverting a covariance matrix based on the MODWT filter kernels. This must also be related to the fact that the MRA does not form an ANOVA for the MODWT. Thus the covariance matrix must be based on the DWT filter kernels in order to get the SVD form and the simple expression for the inverse covariance matrix.

8.3 Covariance Matrix – Wavelet Form

The Discrete Wavelet Transform (DWT) of the *N*-dimensional vector X_t has been defined previously. The wavelet coefficients for a given $N : t = 0,1,\ldots,N-1$ can then be written as:

$$
\begin{aligned}
K_{j;\nu} &\equiv \sum_{t=0}^{N-1} W_{j;\nu,t} X_t \equiv \sum_{l=0}^{L_j-1} h_{j;l} X_{\left[2^j(\nu+1)-1-l\right] \bmod N} && \left(\nu = 0,1,\ldots,N_j-1\right) \\
J_{J;\nu} &\equiv \sum_{t=0}^{N-1} V_{J;\nu,t} X_t \equiv \sum_{l=0}^{L_j-1} g_{J;l} X_{\left[2^J(\nu+1)-1-l\right] \bmod N} && \left(\nu = 0,1,\ldots,N_J-1\right) \\
N_j &\equiv N/2^j \qquad L_j \equiv \left(2^j-1\right)\left(L-1\right)+1 && \left(j = 1,\ldots,J \quad N = 2^{J_0}\right)
\end{aligned}
$$

We then have that the inverse DWT transform is given by:

$$
X_n \equiv \sum_{j=1}^{J} \sum_{\nu=0}^{N_j-1} W_{j;n,\nu}^T K_{j;\nu} + \sum_{\nu=0}^{N_J-1} V_{J;n,\nu}^T J_{J;\nu} \equiv \sum_{j=1}^{J} \sum_{\nu=0}^{N_j-1} K_{j;\nu} W_{j;\nu,n} + \sum_{\nu=0}^{N_J-1} J_{J;\nu} V_{J;\nu,n}
$$

The important characteristic of the DWT transform is that it approximately *decorrelates* or *diagonalizes* the covariance matrix under a wide variety of situations. This property is useful if the covariance matrix is difficult to estimate, and especially if it is expected to be non-stationary. The DWT wavelet decomposition is especially well adapted to *non-stationary* time series, as opposed to the Fourier series decomposition, which implicitly assumes stationarity. Using this DWT transform, it is possible to make a simple assumption that the DWT wavelet coefficients are independent random variables, and thereby arrive at a simple approximation that describes a time-dependent covariance matrix.

Definition of DWT Covariance Matrix

We now make the fundamental assumption that in the DWT wavelet basis, the wavelet and scaling coefficients are *statistically independent*, and hence their covariance matrix is diagonal in this basis:

$$\left\langle K_{j;\nu}, K_{j';\nu'} \right\rangle \equiv \left\langle K_{j;\nu}^2 \right\rangle \delta_{jj'} \delta_{\nu\nu'}, \qquad \left\langle J_{J;\nu}, J_{J;\nu'} \right\rangle \equiv \left\langle J_{J;\nu}^2 \right\rangle \delta_{\nu\nu'}, \qquad \left\langle K_{j;\nu}, J_{J;\nu'} \right\rangle = 0$$

Using this, we may now write down an expression for the covariance matrix in the time basis in terms of the wavelet basis:

$$\Gamma_{m,n} \equiv \left\langle X_m, X_n \right\rangle = \sum_{j=1}^{J} \sum_{\nu=0}^{N_j-1} W_{j;m,\nu}^T \left\langle K_{j;\nu}^2 \right\rangle W_{j;\nu,n} + \sum_{\nu=0}^{N_J-1} V_{J;m,\nu}^T \left\langle J_{J;\nu}^2 \right\rangle V_{J;\nu,n}$$

Note that this covariance matrix is given in Singular Value Decomposition (SVD) form, given that the DWT matrices are orthogonal. The covariance matrix can likewise be expressed in a similar form in terms of the MODWT matrices, but since these matrices are not orthogonal, the result is not in SVD form.

The quantities $\left\langle K_{j;\nu}^2 \right\rangle$ and $\left\langle J_{J;\nu}^2 \right\rangle$ are called the *wavelet variance*. They are functions of the level index *j* and the time index ν. In general, the wavelet variance is not constant on each level and depends on the time index – it is the variance of each wavelet coefficient regarded as a random variable. However, the true wavelet variance components are not simply given by the square of the DWT wavelet and scaling coefficients, because they must be estimated by a smoothing procedure to reduce the stochastic noise, just as in the case of the periodogram. Nevertheless, for the DWT wavelet variance we will use the square of the wavelet coefficients as the best estimate o the past wavelet variance, and the level averages for the estimate of the future wavelet variance. This is on the assumption that the computation of the wavelet decomposition itself represents a smoothing procedure. If we were using the MODWT wavelet variance, then the expressions for the wavelet variance would necessarily imply that some sort of smoothing procedure has been implemented to estimate the true value of the variance. (Since the MODWT coefficients change sign, the corresponding MODWT wavelet variance must go through zero at some points. The conclusion is that the MODWT wavelet variance is not an accurate representation of the true wavelet variance, unless it is smoothed.) Notice that if the wavelet variance were constant on each level, then the above expressions for the covariance matrix would reduce to this constant level wavelet variance times the filter kernels defined previously.

Estimation of DWT Wavelet Variance

For the DWT wavelet variance in the above expressions for the covariance matrix in the DWT basis, we must determine which DWT wavelet coefficient corresponds to the present point in time, and then replace all the wavelet variances to the future of this time with the average level variances. The wavelet variance is estimated from the past data using reflection boundary conditions, the wavelet variances in the past left alone, and the future DWT wavelet variances replaced with the average level variances.

To do this we note the following index correspondences between the DWT wavelet coefficients $K_{j;\nu}$ and scaling coefficients $J_{J;\nu}$, and the original time series, where we define the length of the original time series *X* to be 2*N*:

$$K_{j;\nu} \sim X_{\left[2^j(\nu+1)-1-\left|\nu_j^{(H)}\right|\right]\bmod 2N} \quad \text{with} \quad -\left|\nu_j^{(H)}\right| = -\frac{7}{2}\cdot 2^j + 3$$

$$J_{J;\nu} \sim X_{\left[2^J(\nu+1)-1-\left|\nu_J^{(G)}\right|\right]\bmod 2N} \quad \text{with} \quad -\left|\nu_J^{(G)}\right| = -\frac{6}{2}\cdot 2^J + 3$$

This means that the index *v* of the DWT wavelet or scaling coefficient corresponds to the indicated index of the original time series *X*. To find which DWT index corresponds to the present time $N-1$, we set the index of the time series to this value and solve for ν_{present}. Thus for the level *j* wavelet coefficients we have:

$$N-1 \equiv 2^j(\nu+1)-1+\left(-\frac{7}{2}\cdot 2^j+3\right) = 2^j\left(\nu-\frac{5}{2}\right)+2$$

$$\nu_{\text{present}}[K_j] = \frac{(N-3)}{2^j}+\frac{5}{2} \equiv N_j+\frac{5}{2}-\frac{3}{2^j} = \left(N_j-1\right)+3.5-\frac{3}{2^j}$$

For the level *J* scaling coefficients we have:

$$N-1 \equiv 2^J(\nu+1)-1+\left(-\frac{6}{2}\cdot 2^J+3\right) = 2^J\left(\nu-\frac{4}{2}\right)+2$$

$$\nu_{\text{present}}[J_J] = \frac{(N-3)}{2^J}+\frac{4}{2} \equiv N_j+\frac{4}{2}-\frac{3}{2^J} = \left(N_J-1\right)+3.0-\frac{3}{2^J}$$

With the wavelet coefficients on level *j* (*J* for the scaling coefficients) ranging from 0 to $2N_j-1$, each coefficient in this range corresponds to a time interval of 2^j in the original data series. To align the wavelet coefficients with the original series, the present time at index $N-1$ should correspond to index N_j-1 of the wavelet coefficient. If bars of width 2^j represent the wavelet coefficients, then the position of the wavelet coefficient will be in the middle of the bar.

It can be seen from the above that, for the wavelet coefficients, the present point in time is slightly to the left of the position 3.5 index units to the future of the point N_j-1. For the scaling coefficients the present point in time is slightly to the left of the position 3.0 index units to the future of the point N_j-1. Thus if the wavelet coefficients are shifted to the left by 3.5 bars, minus 3 units of the original time series, the present points of both time series will line up, and the boundary between two bars will lie 3 time units to the right of the present. If the scaling coefficients are shifted to the left by 3.0 bars, minus 3 units of the original time series, the present points of both time series will line up, and the middle of the bar, corresponding to the position of the scaling coefficient, will lie 3 time units to the right of the present.

We thus find it convenient to define the future values of wavelet variance to be shifted by 4 units from the point in both cases:

$$\nu_{\text{future}}[K_j] \geq \left(N_j-1\right)+4.0 \quad \text{and} \quad \nu_{\text{future}}[J_J] \geq \left(N_J-1\right)+4.0$$

So these values of wavelet variance to the future of the present time, corresponding to index $\left(N_j-1\right)$ or $\left(N_J-1\right)$, respectively, are replaced by the level average values, corresponding to the best estimate of these future, unknown values. (This provides a little greater emphasis to the present value of the scaling coefficient than to the present values of the wavelet coefficients.)

Inverse of DWT Covariance Matrix

In this SVD form it is an easy matter to define the inverse covariance matrix, at least over the whole index set of past and future time indices. The inverse covariance matrix is given by:

$$\Gamma^{-1}_{m,n} = \sum_{j=1}^{J}\sum_{\nu=0}^{N_j-1} W^T_{j;m,\nu}\left\langle K^2_{j;\nu}\right\rangle^{-1} W_{j;\nu,n} + \sum_{\nu=0}^{N_J-1} V^T_{J;m,\nu}\left\langle J^2_{J;\nu}\right\rangle^{-1} V_{J;\nu,n}$$

This inverse is well defined as long as the wavelet variance is greater than zero for all index values. Thus we have, using the orthogonality of the DWT matrices:

$$\begin{aligned}
\sum_{n''=0}^{N-1}\Gamma^{-1}_{n,n''}\Gamma_{n'',n'} &= \sum_{n''=0}^{N-1}\left[\sum_{j=1}^{J}\sum_{\nu=0}^{N_j-1}W^T_{j;n,\nu}\left\langle K^2_{j;\nu}\right\rangle^{-1}W_{j;\nu,n''}\right]\left[\sum_{j'=1}^{J}\sum_{\nu'=0}^{N_{j'}-1}W^T_{j';n'',\nu'}\left\langle K^2_{j';\nu'}\right\rangle^{+1}W_{j';\nu',n'}\right]\\
&+\sum_{n''=0}^{N-1}\left[\sum_{\nu=0}^{N_J-1}V^T_{J;n,\nu}\left\langle J^2_{J;\nu}\right\rangle^{-1}V_{J;\nu,n''}\right]\left[\sum_{\nu'=0}^{N_J-1}V^T_{J;n'',\nu'}\left\langle J^2_{J;\nu'}\right\rangle^{+1}V_{J;\nu',n'}\right]\\
&=\sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{\nu=0}^{N_j-1}\sum_{\nu'=0}^{N_{j'}-1}W^T_{j;n,\nu}\left\langle K^2_{j;\nu}\right\rangle^{-1}\left[\sum_{n''=0}^{N-1}W_{j;\nu,n''}W^T_{j';n'',\nu'}\right]\left\langle K^2_{j';\nu'}\right\rangle^{+1}W_{j';\nu',n'}\\
&+\sum_{\nu=0}^{N_J-1}\sum_{\nu'=0}^{N_J-1}V^T_{J;n,\nu}\left\langle J^2_{J;\nu}\right\rangle^{-1}\left[\sum_{n''=0}^{N-1}V_{J;\nu,n''}V^T_{J;n'',\nu'}\right]\left\langle J^2_{J;\nu'}\right\rangle^{+1}V_{J;\nu',n'}\\
&=\sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{\nu=0}^{N_j-1}\sum_{\nu'=0}^{N_{j'}-1}W^T_{j;n,\nu}\left\langle K^2_{j;\nu}\right\rangle^{-1}\delta_{jj'}\delta_{\nu\nu'}\left\langle K^2_{j';\nu'}\right\rangle^{+1}W_{j';\nu',n'}\\
&+\sum_{\nu=0}^{N_J-1}\sum_{\nu'=0}^{N_J-1}V^T_{J;n,\nu}\left\langle J^2_{J;\nu}\right\rangle^{-1}\delta_{\nu\nu'}\left\langle J^2_{J;\nu'}\right\rangle^{+1}V_{J;\nu',n'}\\
&=\sum_{j=1}^{J}\sum_{\nu=0}^{N_j-1}W^T_{j;n,\nu}W_{j;\nu,n'}+\sum_{\nu=0}^{N_J-1}V^T_{J;n,\nu}V_{J;\nu,n'}=\delta_{nn'}
\end{aligned}$$

The crucial step in the above calculation was the use of the orthogonality relation for the DWT matrices. This step would not hold for the MODWT matrices, and so the definition of the inverse covariance matrix as given above would not hold in terms of MODWT matrices. (However, it is possible that a valid *approximate* expression for the inverse covariance matrix could be given in terms of MODWT matrices.)

Now let us consider the problem of inverting the covariance matrix of the past data. The inverse of the total covariance matrix of the past and future data is known from the SVD form given above. Let us now consider the total data space to be of width $2N$ instead of N, with the first N data consisting of the past data, and the second N data consisting of future projected data and zero padding. Let us also redefine the indices to run from $N-1\geq n\geq 0$ for the past data and $0>-h\geq -N$ for the future data, with the index $n=0$ corresponding to the present time. We denote the total covariance matrix $\tilde{\Gamma}_{m,n}$ and divide the covariance matrix into partitions between past and future as follows:

$$\tilde{\Gamma}\equiv\begin{bmatrix}\Gamma_{m,n} & \hat{\Gamma}_{m,-h}\\ \hat{\Gamma}^T_{-k,n} & \hat{\Gamma}_{-k,-h}\end{bmatrix}\qquad\left(N-1\geq m,n\geq 0>-k,-h\geq -N\right)$$

Now, we define the total inverse matrix $\tilde{\Delta} \equiv \tilde{\Gamma}^{-1}$, which is known from the SVD decomposition, to have a similar partition:

$$\tilde{\Delta} \equiv \begin{bmatrix} \Delta_{m,n} & \hat{\Delta}_{m,-h} \\ \hat{\Delta}^{T}_{-k,n} & \overset{\frown}{\Delta}_{-k,-h} \end{bmatrix} \qquad \left(N-1 \geq m,n \geq 0 > -k,-h \geq -N\right)$$

Now, we may verify the following formula [NR, p.77]:

$$\Gamma^{-1}_{m,n} = \Delta_{m,n} - \sum_{h=1}^{N}\sum_{k=1}^{N} \hat{\Delta}_{m,-h} \overset{\frown}{\Delta}^{-1}_{-h,-k} \hat{\Delta}^{T}_{-k,n}$$

The sum is taken only over the future indices, and the corresponding off-diagonal blocks of the covariance matrix are expected to be small. Then the second term is of second order in small quantities. To verify this formula, we have:

$$\begin{aligned} I_{2N} &\equiv \tilde{\Gamma}\tilde{\Delta} \\ 0 &= \Gamma\hat{\Delta} + \hat{\Gamma}\overset{\frown}{\Delta} && \Rightarrow && \hat{\Gamma} = -\Gamma\hat{\Delta}\overset{\frown}{\Delta}^{-1} \\ I_N &= \Gamma\Delta + \hat{\Gamma}\hat{\Delta}^{T} && \Rightarrow && I_N = \Gamma\Delta - \Gamma\hat{\Delta}\overset{\frown}{\Delta}^{-1}\hat{\Delta}^{T} \\ 0 &= \hat{\Gamma}^{T}\Delta + \overset{\frown}{\Gamma}\hat{\Delta}^{T} \\ I_N &= \hat{\Gamma}^{T}\hat{\Delta} + \overset{\frown}{\Gamma}\overset{\frown}{\Delta} && \therefore && \Gamma^{-1} = \Delta - \hat{\Delta}\overset{\frown}{\Delta}^{-1}\hat{\Delta}^{T} \end{aligned}$$

Thus knowing the SVD form of the inverse of the total covariance matrix, which is easy to find, we arrive at an expression for the inverse of the past covariance matrix. In order to find this, we must invert the future part of the (inverse) total covariance matrix, but this may easily be done by approximating it in the SVD form with constant level wavelet variances for the future estimated covariance matrix (between future values and future values). So we may approximate the inverse of the past covariance matrix as the past part of the inverse of the total covariance matrix, plus a correction term of second order in smallness, which depends only on the SVD form of the inverse of the total covariance matrix.

Writing the above formula out explicitly, we have:

$$\tilde{\Delta}_{m,n} \equiv \sum_{j=1}^{J} \sum_{\nu=N_j-1}^{-N_j} W^T_{j;m,\nu} \left\langle K^2_{j;\nu} \right\rangle^{-1} W_{j;\nu,n} + \sum_{\nu=N_J-1}^{-N_J} V^T_{J;m,\nu} \left\langle J^2_{J;\nu} \right\rangle^{-1} V_{J;\nu,n}$$

$$\Delta_{m,n} \equiv \sum_{j=1}^{J} \sum_{\nu=N_j-1}^{-N_j} W^T_{j;m,\nu} \left\langle K^2_{j;\nu} \right\rangle^{-1} W_{j;\nu,n} + \sum_{\nu=N_J-1}^{-N_J} V^T_{J;m,\nu} \left\langle J^2_{J;\nu} \right\rangle^{-1} V_{J;\nu,n}$$

$$\hat{\Delta}_{m,-h} \equiv \sum_{j=1}^{J} \sum_{\nu=N_j-1}^{-N_j} W^T_{j;m,\nu} \left\langle K^2_{j;\nu} \right\rangle^{-1} W_{j;\nu,-h} + \sum_{\nu=N_J-1}^{-N_J} V^T_{J;m,\nu} \left\langle J^2_{J;\nu} \right\rangle^{-1} V_{J;\nu,-h}$$

$$\hat{\Delta}^T_{-k,n} \equiv \sum_{j=1}^{J} \sum_{\nu=N_j-1}^{-N_j} W^T_{j;-k,\nu} \left\langle K^2_{j;\nu} \right\rangle^{-1} W_{j;\nu,n} + \sum_{\nu=N_J-1}^{-N_J} V^T_{J;-k,\nu} \left\langle J^2_{J;\nu} \right\rangle^{-1} V_{J;\nu,n}$$

$$\widehat{\Delta}_{-k,-h} \approx \sum_{j=1}^{J} \sum_{\nu=N_j-1}^{-N_j} W^T_{j;-k,\nu} \left\langle K^2_{j} \right\rangle^{-1} W_{j;\nu,-h} + \sum_{\nu=N_J-1}^{-N_J} V^T_{J;-k,\nu} \left\langle J^2_{J} \right\rangle^{-1} V_{J;\nu,-h}$$

$$\widehat{\Delta}^{-1}_{-h,-k} \approx \sum_{j=1}^{J} \sum_{\nu=N_j-1}^{-N_j} W^T_{j;-h,\nu} \left\langle K^2_{j} \right\rangle^{+1} W_{j;\nu,-k} + \sum_{\nu=N_J-1}^{-N_J} V^T_{J;-h,\nu} \left\langle J^2_{J} \right\rangle^{+1} V_{J;\nu,-k}$$

It will be most convenient to compute these expressions directly using the DWT Pyramid Algorithm and its inverse, acting directly on the appropriate 2N-dimensional DWT vectors. Projection operators are used to restrict the summation over time indices to future values only. Even though the summation over the indices is not over the complete range, we still assume as an approximation that the product of wavelet matrices between different levels is zero, so the products are done level-by-level, with each level independent of the others.

8.4 Linear Prediction Equation – Wavelet Form

From now on we will do all calculations relative to a time $n = 0$, which corresponds to the present time. The new index corresponding to time steps will be positive going backward into the past, and negative going forward into the future. We use a data set of length 2N, consisting of past data $N-1 \ge n \ge 0$ of length N and future projected data and zero padding $0 > n \ge -N$ of length N according to the new index. Transposing this to the old index set ranging from 0 to $2N-1$, with $N \equiv 2048$ and the index values in time order with 0 the distant past and $2N-1$ the distant future, the present time $n = 0$ corresponds to index $N-1 = 2047$.

Solution of the Linear Prediction Equation

Let us then start again with the *h*-step linear prediction equation, in which the data *h* steps ahead are estimated using the past *N* known data values. This is a linear regression equation, in which the future data *h* steps ahead are regressed on past *N* known data values:

$$X_{-h} \equiv \hat{X}_{-h} + \varepsilon_{-h} \qquad \hat{X}_{-h} \equiv \sum_{n=0}^{N-1} X_n \phi_{n(h)}$$

The quantity $\hat{X}_{-h}$ represents the estimated (future) *signal*, and ε_{-h} represents the *residual*, which is random white noise. In terms of the time-dependent covariance matrix, which depends on both indices separately (rather than just their difference if it is a Toeplitz matrix, which is the case for a stationary process), this becomes:

$$\left\langle X_m, X_n \right\rangle \equiv \Gamma_{m,n} \equiv \hat{\Gamma}_{m,n} + \left\langle \varepsilon^2 \right\rangle \delta_{m,n}$$
$$\sum_{n=0}^{N-1} \Gamma_{m,n} \phi_{n(h)} \equiv \sum_{n=0}^{N-1} \left[\hat{\Gamma}_{m,n} + \left\langle \varepsilon^2 \right\rangle \delta_{m,n} \right] \phi_{n(h)} = \hat{\Gamma}_{m,-h}$$

Thus the LP coefficients may be found by inverting the time-dependent covariance matrix over the past *N* data values:

$$\phi_{n(h)} = \sum_{m=0}^{N-1} \Gamma^{-1}_{n,m} \hat{\Gamma}_{m,-h} = \sum_{m=0}^{N-1} \hat{\Gamma}^T_{-h,m} \Gamma^{-1}_{m,n}$$

Here we have taken the transpose since the covariance matrix is symmetric. Now we may use this to write the future estimated time series as a matrix acting on the past series:

$$\boxed{\hat{X}_{-h} \equiv \sum_{n=0}^{N-1} \phi_{n(h)} X_n = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \hat{\Gamma}^T_{-h,m} \Gamma^{-1}_{m,n} X_n}$$

Taking the inverse of the covariance matrix only over non-negative values of the index, and multiplying by covariance matrix elements with a negative value –*h* corresponding to the future prediction, results in non-zero values of the LP coefficients. We call this the Linear Prediction (LP) equation.

However, the matrix $\tilde{\Gamma}_{m,n}$, defined for all values of indices both *past* and *future*, is what we estimate in the diagonal DWT wavelet basis by estimating the past and future time-dependent DWT level wavelet variance. It is very convenient therefore to be able to invert the covariance matrix in this diagonal basis. We may then define a new inverse matrix over the whole data set, both past and future, as follows:

$$\sum_{n=N-1}^{-N} \tilde{\Gamma}_{m,n}\tilde{\Delta}_{n,k} = \sum_{n=N-1}^{-N} \tilde{\Delta}_{m,n}\tilde{\Gamma}_{n,k} = \delta_{m,k}$$

We found previously that the total covariance matrix $\tilde{\Gamma}_{m,n}$ could be written in block-diagonal form, separated between past and future index values, which is easy to compute from its SVD decomposition:

$$\tilde{\Gamma} \equiv \begin{bmatrix} \Gamma_{m,n} & \hat{\Gamma}_{m,-h} \\ \hat{\Gamma}^T_{-k,n} & \widehat{\Gamma}_{-k,-h} \end{bmatrix} \qquad \left(N-1 \geq m,n \geq 0 > -k,-h \geq -N\right)$$

We also found the total inverse matrix $\tilde{\Delta} \equiv \tilde{\Gamma}^{-1}$, which is known from the SVD decomposition, to have a similar partition:

$$\tilde{\Delta} \equiv \begin{bmatrix} \Delta_{m,n} & \hat{\Delta}_{m,-h} \\ \hat{\Delta}^T_{-k,n} & \widehat{\Delta}_{-k,-h} \end{bmatrix} \qquad \left(N-1 \geq m,n \geq 0 > -k,-h \geq -N\right)$$

Then we found the following formula [NR, p.77] for the inverse covariance matrix over the past index values in terms of this SVD form of the total inverse covariance matrix:

$$\boxed{\Gamma^{-1}_{m,n} = \Delta_{m,n} - \sum_{h=1}^{N}\sum_{k=1}^{N} \hat{\Delta}_{m,-h}\widehat{\Delta}^{-1}_{-h,-k}\hat{\Delta}^T_{-k,n}}$$

Substituting this expression into the formula for the LP coefficients yields the solution, where the matrices involved are all computed from their SVD form using the estimated DWT wavelet variance and successive DWT operations.

Calculation of Covariance Matrices

Let us return to the original definition of the total covariance matrix in terms of the DWT coefficients:

$$\tilde{\Gamma}_{m,n} \equiv \sum_{j=1}^{J} \sum_{\nu=N_j-1}^{-N_j} W^T_{j;m,\nu} \left\langle K^2_{j;\nu} \right\rangle W_{j;\nu,n} + \sum_{\nu=N_J-1}^{-N_J} V^T_{J;m,\nu} \left\langle J^2_{J;\nu} \right\rangle V_{J;\nu,n}$$

Let us once again adopt the following shorthand notation for the wavelet smooth matrices and wavelet smooth variance:

$$V_{J;\nu,n} \equiv W_{J+1;\nu,n}, \qquad J^2_{J;\nu} \equiv K^2_{J+1;\nu}$$

Then the covariance matrix can be written:

$$\tilde{\Gamma}_{m,n} \equiv \sum_{j=1}^{J+1} \sum_{\nu=N_j-1}^{-N_j} W_{j;m,\nu}^T \left\langle K_{j;\nu}^2 \right\rangle W_{j;\nu,n}$$

The component matrices in the LP equation are all expressed in terms of the block decomposition of this total covariance matrix and its inverse in SVD form. Hence they can be calculated directly by means of the DWT and its inverse.

Thus the matrix $\hat{\Gamma}_{m,-h}$ in the LP equation is in the off-diagonal block of the total covariance matrix, which is thus given by:

$$\hat{\Gamma}_{m,-h} \equiv \sum_{j=1}^{J+1} \sum_{\nu=N_j-1}^{-N_j} W_{j;m,\nu}^T \left\langle K_{j;\nu}^2 \right\rangle^{+1} W_{j;\nu,-h}$$

The inverse matrix $\Delta_{n,m}$ can also be written in this form:

$$\Delta_{n,m} \equiv \sum_{j=1}^{J+1} \sum_{\nu=N_j-1}^{-N_j} W_{j;n,\nu}^T \left\langle K_{j;\nu}^2 \right\rangle^{-1} W_{j;\nu,m}$$

The inverse matrix $\hat{\Delta}_{n,-k}$ can be written likewise:

$$\hat{\Delta}_{n,-k} \equiv \sum_{j=1}^{J+1} \sum_{\nu=N_j-1}^{-N_j} W_{j;n,\nu}^T \left\langle K_{j;\nu}^2 \right\rangle^{-1} W_{j;\nu,-k}$$

Finally, the matrix $\widehat{\Delta}_{-k,-l}^{-1}$ can be written likewise:

$$\widehat{\Delta}_{-k,-l}^{-1} \approx \sum_{j=1}^{J+1} \sum_{\nu=N_j-1}^{-N_j} W_{j;-k,\nu}^T \left\langle K_{j;\nu}^2 \right\rangle^{+1} W_{j;\nu,-l}$$

This latter matrix is approximately, but not exactly, the stationary covariance matrix computed from the level average future DWT wavelet variances. (There is some overlap with the past values of DWT wavelet variance.) It is easily seen that the transpose matrices assume the same form.

To compute these matrix products in terms of the DWT, let us start with the following product:

$$\begin{aligned}
\sum_{m=0}^{N-1} \hat{\Gamma}_{-h,m}^T \Delta_{m,n} &= \sum_{m=0}^{N-1} \left[\sum_{j=1}^{J+1} \sum_{\nu=N_j-1}^{-N_j} W_{j;-h,\nu}^T \left\langle K_{j;\nu}^2 \right\rangle^{+1} W_{j;\nu,m} \right] \left[\sum_{j=1}^{J+1} \sum_{\nu=N_j-1}^{-N_j} W_{j;m,\nu}^T \left\langle K_{j;\nu}^2 \right\rangle^{-1} W_{j;\nu,n} \right] \\
&= \sum_{j=1}^{J+1} \sum_{j'=1}^{J+1} \sum_{\nu=N_j-1}^{-N_j} \sum_{\nu'=N_{j'}-1}^{-N_{j'}} W_{j;-h,\nu}^T \left\langle K_{j;\nu}^2 \right\rangle^{+1} \left[\sum_{m=0}^{N-1} W_{j;\nu,m} W_{j';m,\nu'}^T \right] \left\langle K_{j';\nu'}^2 \right\rangle^{-1} W_{j';\nu',n}
\end{aligned}$$

The product of matrices in brackets is summed only over past index values. Thus let us define the following *projection vectors* to project onto the subspaces of past (+) and future (–) indices:

$$P_m^{(+)} \equiv \begin{cases} 1 & (m \geq 0) \\ 0 & (m < 0) \end{cases} \qquad P_m^{(-)} \equiv \begin{cases} 0 & (m \geq 0) \\ 1 & (m < 0) \end{cases} \qquad \begin{matrix} \Rightarrow & \text{past} \\ \Rightarrow & \text{future} \end{matrix}$$

Then we may express the matrix product as a sum over the whole index set, utilizing the projection vectors:

$$\sum_{m=0}^{N-1} W_{j;\nu,m} W_{j';m,\nu'}^T = \sum_{m=N-1}^{-N} W_{j;\nu,m} P_m^{(+)} W_{j';m,\nu'}^T \equiv \left[W P^{(+)} W^T \right]_{j,j';\nu,\nu'}$$

Starting with a vector on the right in the DWT wavelet coefficient space, this may be computed by successively taking the inverse DWT, multiplying by the projection vector in time series space, then taking the DWT back into the wavelet coefficient space. Thus, written in matrix notation, the expression for the matrix product above becomes:

$$\sum_{m=0}^{N-1} \hat{\Gamma}_{-h,m}^T \Delta_{m,n} = \left[P^{(-)} \left(W^T \left\langle K^2 \right\rangle^{+1} W \right) P^{(+)} \left(W^T \left\langle K^2 \right\rangle^{-1} W \right) P^{(+)} \right]_{-h,n}$$

Similarly, we may write the other term of the LP coefficient equation as:

$$\sum_{h=1}^{N} \sum_{k=1}^{N} \hat{\Delta}_{m,-h} \widehat{\Delta}_{-h,-k}^{-1} \hat{\Delta}_{-k,n}^T \equiv \left[P^{(+)} \hat{\Delta} \; P^{(-)} \widehat{\Delta}^{-1} P^{(-)} \hat{\Delta}^T P^{(+)} \right]_{m,n}$$

$$\hat{\Delta} \equiv \hat{\Delta}^T \equiv \left(W^T \left\langle K^2 \right\rangle^{-1} W \right), \qquad \widehat{\Delta}^{-1} \equiv \left(W^T \left\langle K^2 \right\rangle^{+1} W \right)$$

This term is then multiplied on the left by the expression for $\hat{\Gamma}_{-h,m}^T$ given above, and subtracted from the previous term, to yield the complete expression for the LP coefficients (in matrix form).

It should be noted that in the *stationary* approximation, the wavelet variance on each level is constant in time and hence does not depend on the index *ν*. Hence the covariance matrix over the whole data set reduces to the following form:

$$\tilde{\Gamma}_{m,n}^{(\text{stat.})} \equiv \sum_{j=1}^{J+1} \left\langle K_j^2 \right\rangle \sum_{\nu=N_j-1}^{-N_j} W_{j;m,\nu}^T W_{j;\nu,n} \equiv \sum_{j=1}^{J+1} \left\langle K_j^2 \right\rangle F_{j;m,n}$$

The quantity $F_{j;m,n}$ is the *filter kernel* and is the same function that operates on the data set to give each level *j* of the DWT MRA. Thus the stationary covariance matrix acting on a vector X_n over the whole data set is equivalent to taking the MRA of the data, and multiplying each level *j* by the constant level wavelet variance $\left\langle K_j^2 \right\rangle$.

Optimal Filtering and Smoothing

We may define a generalized Linear Prediction and Smoothing that includes an estimation of the *true signal* values of the past data along with the projected future data. Thus we start again with the original LP equation for the projected future data $\hat{X}_{-h}$, except that now we write it for a general index *k* that can be either a future or a past index:

$$\hat{X}_k \equiv \sum_{n=0}^{N-1} X_n \tilde{\phi}_{n(k)} = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \hat{\Gamma}_{k,m}^T \Gamma_{m,n}^{-1} X_n$$

If *k* were a past index value and the matrix $\hat{\Gamma}_{k,m}^T$ were the full covariance matrix, then obviously the product of the past covariance matrix with its inverse would yield the identity matrix. In other words, the best estimate of the past signal would be simply the value it assumed. But the smoothing operation enters because $\hat{\Gamma}_{k,m}^T$ is an estimate of the *true* covariance matrix, in which the noise covariance has been removed. If *k* were a future index, then there would be no noise component anyway, but if *k* is a past index, then there will be a diagonal noise component in addition to the covariance matrix of the signal, which has diagonal and off-diagonal components. Thus to arrive at this estimate of the *true* covariance matrix, we must estimate the noise component and subtract it from the diagonal components of the covariance matrix. Then the product of this de-noised covariance matrix with the inverse of the original covariance matrix is not the identity, and it results in an estimate of the *true* signal, after de-noising. This filtering operation on the past data is also called *optimal filtering* and *smoothing*.

8.5 References

[PW] Donald B. Percival & Andrew T. Walden,
Wavelet Methods for Time Series Analysis,
Cambridge University Press, Cambridge, UK (2000)