Dyadic Wavelets Analysis

by D.S.G. Pollock
University of Leicester
Email: stephen_pollock@sigmapi.u-net.com

In the previous chapter, we have described the means of effecting a dyadic decomposition of the Nyquist frequency range that is associated with sampled data systems. In the present chapter, we shall consider the continuous-time processes that underlie the data.

No assumptions will be made regarding the statistical homogeneity or stationarity of the processes that generate the data. Therefore, a global parametric description of the statistical properties of the data cannot be provided. Instead, the requirement is for a set of basis functions that can give rise to representations of the underlying processes that fully reflect the local variations of the data.

The set of basis functions is provided by sequences of wavelets and scaling functions that are ordered both in time and in frequency. The coefficients attributed to these functions in representing the underlying continuous processes are none other than the sequences obtained by the procedures of digital filtering and dowsampling that have been described in the previous chapter.

The continuous-time wavelets and scaling functions are intimately associated with the filters; and the precise specifications of the filters and of the functions are matters that will be dealt with in the following chapter. For the present, we shall rely on the fourth-order Daubechies functions to provide the necessary examples.

We shall also take the opportunity to develop a matrix representation of the dyadic decomposition that can serve in place of the z-transform representation that has already been developed in the context of two-channel filter banks. The two representations closely linked. In dealing with finite data sequences, we shall resort to circulant matrices.

The Dyadic Decomposition of a Space of Functions

A discrete wavelet analysis is based on the supposition that the elements of the data sequence $\{y_k; k = 0, 1, 2, \ldots, T-1\}$ have been sampled from a continuous or piecewise continuous function $f(t)$ with $t \in [0, T)$. It is presumed that the function can be reconstituted, to some degree of approximation, by associating a scaling function kernel or father wavelet $\phi_{0,k}(t) = \phi_0(t-k)$ to each of these ordinates and by summing the result:

$$f(t) \simeq y(t) = \sum_{k=0}^{T-1} y_k \phi_0(t-k). \tag{1}$$

The scaling function $\phi(t-k)$ is defined on the real line, with $t \in \mathcal{R}$, and it is supported on an interval that may be finite or infinite. The nominal centre of the function is at $t = k$, and successive values of $k \in \mathcal{I} = \{0, \pm 1, \pm 2, \ldots\}$ represent successive displacements to the left.

The scaling functions are designed to constitute an orthonormal basis of the space $\mathcal{V}_0$ in which the function $f(t)$, or its approximation $y(t)$, resides. Therefore,

$$\int_t \phi_0(t-j)\phi_0(t-k)dt = \big\langle \phi_{0,j}(t), \phi_{0,k}(t) \big\rangle = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{if } j \neq k. \end{cases} \tag{2}$$

This orthogonality condition can be expressed both in the time domain and in the frequency domain as follows:

$$\big\langle \phi_{0,j}(t), \phi_{0k}(t) \big\rangle = \delta(j-k) \longleftrightarrow \sum_k |\phi_0(\omega + 2k\pi|^2 = 1. \tag{3}$$

On the LHS, there is a discrete-time sequence with a unit impulse at time $\tau = j - k = 0$ and with zeros elsewhwere. The discrete-time Fourier transform gives rise to the 2π-periodic function that is seen on the RHS, where the sum is over all positive an negative integers. Also evident on the RHS is the fact that the transform of the unit-impulse sequence is a unit-valued constant function that is defined for all positive and negative frequency values $\omega \in \mathcal{R}$.

The elements of the data sequence, which are the amplitude coefficients of the associated scaling functions, are given by

$$y_k = \int_t y(t)\phi_0(t-k)dt = \big\langle y(t), \phi_{0.k}(t) \big\rangle. \tag{4}$$

The basis $\phi_{0,k}(t) = \phi_0(t-k); k = 0, 1, \ldots, T-1$, which is ordered in time, may be described as the initial basis of scaling functions. Scaling functions have nominal frequency contents that extend from a limiting frequency of π radians per sampling interval down to the zero frequency.

In a dyadic wavelets analysis, the T amplitude coefficients of equation (1), which are associated with the initial basis, and which are the sampled values, are transformed into a hierarchy of T coefficients that are associated with an alternative basis, which is ordered both according to the temporal locations of the wavelets and according to their frequency contents. This constitutes the final basis.

The hierarchy of wavelets within the final basis can be described with reference to a so-called mosaic diagram that defines a partitioning of the time-frequency plane, which corresponds to the space $\mathcal{V}_0$. This is illustrated for a sample of size $T = 128 = 2^7$ by Figure 1. In the figure, the height of a cell corresponds to a bandwidth in the frequency domain, whereas its width denotes a temporal duration.

The highest frequency in the mosaic diagram is the Nyquist frequency of π radians per sample interval, which represents the maximum frequency that is

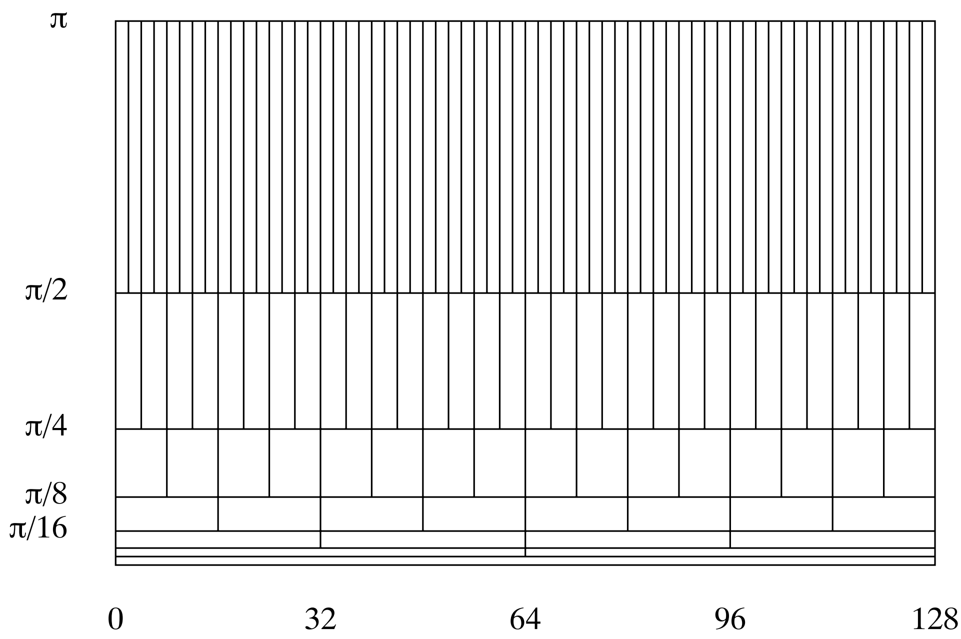

Figure 1. The partitioning of the time–frequency plane according to a multiresolution analysis of a data sequence of $128 = 2^7$ points.

detectable via the process of discrete sampling. Centred on each cell, but liable to extend beyond its temporal boundaries, there is a wavelet. The frequency contents of the wavelet is also liable to extend beyond the nominal bandwidth that is indicated in the figure.

Given that the wavelets and scaling functions extend beyond the temporal boundaries of the cells to which they are assigned, there are bound to be cases where they extend beyond the bounds of the sample.

A data sequence of infinite length, which would support all wavelets, could be generated by the infinite periodic extension of the sample. However, to achieve the same effect, it is also appropriate to envisage wrapping the data around a circle of a circumference T, equal to the number of data points, such that the end of the sample is adjacent to its beginning. This creates a circulant process, which is equivalent to a periodic process.

The wavelets can then be wrapped likewise around the circle. If the lengths of the individual wavelets exceed that of the circumference, then they can be wrapped around the circle the appropriate number of times, and their overlying ordinates can be added.

The effect of wapping the scaling function $\phi(t)$ around the circle it to create a circulant function

$$\phi_0^\circ(t) = \phi_0(t) + \sum_k \bigl\{\phi_0(t + kT) + \phi_0(t - kT)\bigr\}, \tag{5}$$

where it is to be expected that, if they do no vanish as k increases, the terms within the summation will decrease rapidly.

In the process of wrapping the scaling functions, the tails of those associated with the lower end of the sample will make incursions into the upper end. Conversely, heads of the functions from the upper end will impinge on the lower end. However, the sequential orthogonality of the functions will ensure that this will have no untoward effect.

Apart from the final cell in the vertical hierarchy of the mosaic diagram, which stretches across the width of the diagram and which is bounded by the zero frequency, the cells and bounded above and below by positive frequencies. These cells are occupied by mother wavelets, which have a different form from that of the scaling functions.

The Divisions of the Frequency Scale

The horizontal bands of the mosaic diagram are obtained by successive divisions of the frequency range. First, the range of frequencies $[0, \pi]$ of the space $\mathcal{V}_0$ is divided into the equal subintervals $[0, \pi/2]$ and $(\pi/2, \pi]$. The upper frequency interval will have $T/2$ wavelet functions, denoted by $\psi_{1,k}(t); k = 0, 1, \ldots, [T/2] - 1$, separated one from the next by two time intervals. These wavelets will constitute a basis for a space denoted by $\mathcal{W}_1$.

The lower frequency interval will have the same number $T/2$ of scaling functions, denoted by $\phi_{1,k}(t); k = 0, 1, \ldots, [T/2] - 1$, also separated by two intervals. These scaling functions will constitute a basis for a space denoted by $\mathcal{V}_1$. The division of $\mathcal{V}_0$ is such that its two subspaces $\mathcal{W}_1$ and $\mathcal{V}_1$ are mutually orthogonal. Their orthogonality, which can be denoted by writing $\mathcal{V}_1 \perp \mathcal{W}_1$, entails the fact that $\mathcal{V}_1 \cap \mathcal{W}_1 = 0$.

The direct sum of the two subspaces is $\mathcal{W}_1 \oplus \mathcal{V}_1 = \mathcal{V}_0$. This means that any element in $y(t) \in \mathcal{V}_0$ can be expressed as $y(t) = w_1(t) + v_1(t)$ with $w_1(t) \in \mathcal{W}_1$ and $v_1(t) \in \mathcal{V}_1$, which is the sum of two orthogonal functions.

In the next stage of the decomposition of $\mathcal{V}_0$, the lower interval is subdivided into the intervals $[0, \pi/4]$ and $(\pi/4, \pi/2]$, which are filled, respectively, with $T/4$ scaling functions, denoted by $\phi_{2,k}(t); k = 0, 1, \ldots, [T/4] - 1$, and $T/4$ wavelets, denoted by $\psi_{2,k}(t); k = 0, 1, \ldots, [T/4] - 1$, separated by four time intervals. These will constitute the basis functions, respectively, of the spaces $\mathcal{V}_2$ and $\mathcal{W}_2$, which are mutually orthogonal subspaces of $\mathcal{V}_1 = \mathcal{W}_2 \oplus \mathcal{V}_2$.

The process of subdivision continues, by dividing successively the lower subintervals, until it can go no further. If there are $T = 2^n$ points in the sample, then T can be divided n times, and there will be a total of $n + 1$ horizontal bands, with the cells of all but the final band filled with wavelets. The final band will contain a single scaling function.

The process of subdivision generates a nested sequence of vector spaces, each of which is spanned by a set of scaling functions:

$$\mathcal{V}_0 \supset \mathcal{V}_1 \supset \cdots \supset \mathcal{V}_n. \tag{6}$$

The jth stage of the process, which generates $\mathcal{V}_j$, also generates the accompanying space $\mathcal{W}_j$ of wavelet functions, which is its orthogonal complement within $\mathcal{V}_{j-1}$. The complete process can be summarised by displaying the successive decompositions of $\mathcal{V}_0$:

$$\begin{aligned}\mathcal{V}_0 &= \mathcal{W}_1 \oplus \mathcal{V}_1\\ &= \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \mathcal{V}_2\\ &\vdots\\ &= \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \cdots \oplus \mathcal{W}_n \oplus \mathcal{V}_n.\end{aligned} \tag{7}$$

The elements the final expression correspond to the successive horizontal bands of the mosaic diagram.

Given the decomposition of $\mathcal{V}_0$ as a sum of mutually orthogonal subspaces, represented by equation (7), and given that $y(t) \in \mathcal{V}_0$, it is possible to represent the function $y(t)$ as a sum of orthogonal components residing in the subspaces. Thus

$$y(t) = w_1(t) + w_2(t) + \cdots + w_n(t) + v_n(t), \tag{8}$$

with $w_j(t) \in \mathcal{W}_j$ for $j = 1, \ldots, n$ and with $v_n(t) \in \mathcal{V}_n$.

The generic component of this decomposition may be represented, relative to the basis functions $\psi_{j,k}(t); k = 0, 1, \ldots, [T/2^j] - 1$ of $\mathcal{W}_j$, by

$$w_j(t) = \sum_{k=0}^{[T/2^j]-1} \beta_{jk}\psi_{j,k}(t). \tag{9}$$

Here, β_{jk} is an amplitude coefficient of the kth wavelet function. The two final elements of the decomposition of (8) are the wavelet function $w_n(t) = \beta_{n0}\psi_{n,0}(t)$ and the scaling function $v_n(t) = \gamma_{n0}\phi_{n,0}(t)$, which have been scaled by the amplitude coefficients β_{n0} and γ_{n0}, respectively.

It follows that $y(t) \in \mathcal{V}_0$ can be expressed as

$$y(t) = \sum_{j=0}^{n-1}\left\{\sum_{k(j)} \beta_{jk}\psi_{j,k}(t)\right\} + \beta_{n0}\psi_{n,0}(t) + \gamma_{n0}\phi_{n,0}(t). \tag{10}$$

Here, j is an index of the frequency level, or of the level of resolution, and k is a temporal index. Since the range of k depends on the level of resolution, it has been written as $k = k(j)$ for the subscript of the summation sign.

The means by which the data vector $y = [y_0, y_1, \ldots, y_{T-2}, y_{T-1}]'$ of T elements is converted to the vector $\beta = [\beta_{0,1}, \beta_{0,2}, \ldots \beta_{n,0}, \gamma_{n,0}]'$ of T amplitude coefficients is described as a Discrete Wavelet Transform (DWT).

The transformation entails an orthonormal matrix Q such that $QQ' = Q'Q = I_T$. Then, $\beta = Q'y$, and the corresponding inverse transformation that maps from the amplitude coefficients to the data is $y = Q\beta$. This is the matrix version of equation (10).

Given that the function $y(t)$ is represented, in practice, by its T sampled ordinates $y_k; k = 0, 1, 2, \ldots, T-1$, it would be appropriate to express the components of the decomposition of (8) in terms of their ordinates sampled at the integer points. This would generate the expression

$$y_k = w_{1k} + w_{2k} + \cdots + w_{nk} + v_{nk}; \quad k = 0, 1, \ldots, T-1, \tag{11}$$

which is the discrete-time counterpart of equation (8). However, unless they can be given analytic expressions, it it is difficult to obtain the ordinates of the wavelets at the points in question. One recourse is to use to the so-called Maximum-Overlap Discrete Wavelet Transform (MODWT) that generates full sequences of T coefficients at each level of resolution.

A common purpose is to remove from $f(t)$ the traces of an additive noise contamination. If the noise resides within a limited set of wavelets bands, which are liable to be the high-frequency bands, then the signal can be enhanced by removing the corresponding components from the sum.

If the only part of the signal that is of interest resides within a limited set of adjacent bands, then it can be isolated in a straightforward way by forming the partial sum of the corresponding components.

Another common purpose in a wavelets analysis is to achieve a measure of data compression. If the absolute value of the amplitude coefficient β_{jk} associated with the wavelet basis function $\psi_{j,k}(t) \in \mathcal{W}_j$ is below a predetermined level of significance, then it can be set to zero. In this way, it may become possible to convey the essential information of the signal in far fewer than the T coefficients that are present in equation (1), which are the data points.

The Two-Scale Dilation Equations

The scaling functions and the wavelets in successive bands represent dilated or stretched versions of the functions in the bands above. Let $\mathcal{V}_0$ be the space spanned by the scaling functions $\phi_{0,k}(t) = \phi_0(t-k)$, which constitute an orthonormal basis, and let $\mathcal{V}_1 \subset \mathcal{V}_0$ be the subspace containing functions at half the resolution. Then, $\mathcal{V}_1$ will be spanned by the basis functions $\phi_{1,k}(t)$, which represent versions of the functions $\phi_{0,k}(t) = \phi_0(t-k)$ that have been stretched by a factor of 2. Thus, if $\phi_{0,k}(t)$ is supported on a finite interval, then $\phi_{1,k}(t)$ will be supported on an interval of double the length.

The relationship between the two sets of functions is such that

$$\phi_{1,k}(t) = \phi_1(t-k) = 2^{-1/2}\phi_0(2^{-1}t - k) \tag{12}$$

which implies that

$$\phi_{0,k}(t) = \phi_0(t-k) = 2^{1/2}\phi_1(2t - k) \tag{13}$$

Replacing t by $2^{-1}t$ means that a basis function $\phi_{1,k}(t)$ of $\mathcal{V}_1$ of (12) will evolve at half the rate of a basis functions $\phi_{0,k}(t)$ of $\mathcal{V}_0$. Multiplying the functions by the factor $2^{-1/2}$ ensures that their squares will continue to integrate to unity, which is a necessary normalisation. This follows from observing that

$$\int \phi_1^2(t)dt = \frac{1}{2}\int \phi_0^2(2^{-1}t)dt = \frac{1}{2}\int \phi_0^2(\tau)\frac{dt}{d\tau}d\tau = 1, \tag{14}$$

where $t = 2\tau$ and $dt/\tau = 2$, and where the integral of $\phi_0^2(\tau)$ is unity in consequence of (2). Evidently, the area of a scaling function that has been stretched by a factor of two has been doubled; and, by applying the factor $2^{-1/2}$ to the function, the area is reduced to unity.

Also observe that the basis functions of $\mathcal{V}_1$ are separated one from the next by intervals of 2 points. Thus, whereas $\phi_1(t-k) = 2^{-1/2}\phi_0(2^{-1}t-k)$ will have its centre at the point $t = 2k$, which is the solution of $2^{-1}t - k = 0$, the succeeding function $\phi_{1,k+1}(t)$ will have its centre at the point $t = 2k+2$.

Equation (12) may be generalised to give

$$\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k), \tag{15}$$

which is a basis function of $\mathcal{V}_j$. It should be noted that, whereas the present notation has $\mathcal{V}_1 \subset \mathcal{V}_0$, some authors reverse the order of the indices so that the space of higher dimension acquires the higher index.

It is possible to express the scaling function $\phi_{1,0}(t) \in \mathcal{V}_1$ as a linear combination of the elements $\phi_{0,k}(t)$ of the basis of a space $\mathcal{V}_0$ of twice the resolution. An appropriate expression of this relationship is via the equation

$$\phi_{1,0}(t) = \sum_k g_k \phi_{0,k}(t), \tag{16}$$

where

$$g_k = \langle \phi_{1,0}(t), \phi_{0,k}(t) \rangle = \int_{-\infty}^{\infty} \phi_1(t)\phi_0(t-k)dt \tag{17}$$

An equivalent expression of (16) is

$$\phi_1(t) = \sum_k g_k \phi_0(t-k). \tag{18}$$

Also, there is

$$\phi_0(t) = 2^{1/2}\sum_k g_k \phi_0(2t-k), \tag{19}$$

which makes use of a lagged version of equation (13) whereby $\phi_{-1,k}(t) = 2^{1/2}\phi_0(2t-k)$.

Equation (16) is the so-called two-scale dilation equation of the scaling function. The coefficients g_k of the dilation equation are also the coefficients of a lowpass filter. More generally, the relationship between the basis elements of $\mathcal{V}_j$ and those of $\mathcal{V}_{j-1}$ is indicated by

$$\phi_{j,0}(t) = \sum_k g_k \phi_{j-1,k}(t). \tag{20}$$

The orthogonal complement within $\mathcal{V}_0$ of the space $\mathcal{V}_1$ of scaling functions is the space $\mathcal{W}_1$ of wavelets functions. The $\mathcal{W}_1$ space is spanned by the wavelets functions $\psi_{1,k}(t); k = 0, 1, \ldots, [T/2]-1$, which constitute an orthonormal basis.

Since $\mathcal{W}_1 \subset \mathcal{V}_0$, it is possible to express the wavelet function $\psi_1(t)$ as a linear combination of the elements of the basis of $\mathcal{V}_0$. The appropriate expression is

$$\psi_{1,0}(t) = \sum_k h_k \phi_{0,k}(t), \tag{21}$$

where

$$h_k = \langle \psi_{1,0}(t), \phi_{0,k}(t) \rangle = \int_{-\infty}^{\infty} \psi_1(t)\phi_0(t-k)dt \tag{22}$$

Equation (21) is the two-scale dilation equation of the wavelet function. The coefficients h_k of the equation are also the coefficients of a highpass filter that is complementary to the lowpass filter that entails the coefficients g_k.

More generally, there is

$$\psi_{j,0}(t) = \sum_k h_k \phi_{j-1,k}(t), \tag{23}$$

and, in parallel with equation (15), there is

$$\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k). \tag{24}$$

Restrictions on the Scaling Coefficients

Various conditions must be imposed on the coefficients of the two-scale equations. The first condition concerns the sum of the coefficients, which must be

$$\sum_k g_k = 2^{1/2}. \tag{25}$$

The necessity of this condition is established by integrating both sides of equation (16). On the one side, there is

$$\begin{aligned} \int \phi_{1,0}(t)dt &= 2^{-1/2}\int \phi_0(2^{-1}t)dt \\ &= 2^{-1/2}\int \phi_0(\tau)\frac{dt}{d\tau}d\tau = 2^{1/2}\int \phi_0(\tau)d\tau. \end{aligned} \tag{26}$$

Here, the variable of integration has been changed from t to $\tau = t/2$, which accounts for the factor $dt/d\tau = 2$. On the other side, there is

$$\int \sum_k g_k \phi_{0,k}(t)dt = \sum_k g_k \int \phi_0(t-k)dt, \tag{27}$$

where it has been assumed that integration and the summation can be commuted. The integrals of $\phi_0(\tau)$ and $\phi_0(t-k)$ are of equal value, regardless of the displacement k. Therefore, the equality of (26) and (27) implies the condition of (25).

The next condition is that the sum of squares of the coefficients of the dilation is unity:

$$\sum_k g_k^2 = 1. \tag{28}$$

This follows from the fact that the scaling functions at all levels constitute orthonormal bases. Thus, at level 1, there is

$$\begin{aligned} 1 &= \int_t \phi_{1,0}^2(t)dt = \int_t \Big\{ \sum_k g_k \phi_0(t-k) \Big\}^2 dt \\ &= \sum_j \sum_k g_j g_k \int_t \phi_0(t-j)\phi_0(t-k)dt = \sum_k g_k^2, \end{aligned} \tag{29}$$

where the final equality follows from the conditions of orthonormality of (2).

A further important condition affecting the coefficients is that

$$\sum_k g_k g_{k+2m} = 0. \tag{30}$$

This also follows from the orthogonality of the scaling functions and from the dilation equation. The orthogonality of any two separate scaling functions at level 0 implies that, if $m \neq 0$, then, in view or (15), there is

$$\begin{aligned} 0 &= \int_t \phi(2^{-1}t)\phi(2^{-1}t - m)dt \\ &= 2\int_t \sum_j \sum_k g_j g_k \phi(t-j)\phi([t-2m]-k)dt \\ &= 2\sum_j \sum_k g_j g_k \int_t \phi(t-j)\phi(t-[2m+k])dt. \end{aligned} \tag{31}$$

The integral within the final expression will be zero-valued unless $j = k + 2m$. In that case, the expression will deliver the term $2g_k g_{k+2m}$, which must be equal to zero. It follows that equation (30) is a necessary condition for the sequential orthogonality of the scaling functions.

The coefficients of the dilation equation of the wavelets must fulfil conditions that are equivalent to those that affect the scaling function dilation. Thus

$$p_0 = \sum_k h_k^2 = 1 \quad \text{and} \quad p_{2m} = \sum_k h_k h_{k+2m} = 0. \tag{32}$$

These are necessary conditions for the *sequential* orthogonality of separate wavelets within the band in question; and they represent restrictions on an autocovariance function. In addition, it is required that

$$\sum_k h_k = 0. \tag{33}$$

This is sufficient to ensure that the integrals of the wavelets are zero.

Conditions must also be imposed to ensure that the wavelets are orthogonal to the scaling functions. This is described as *lateral* orthogonality. To ensure that the scaling function $\phi_j(t)$ and the wavelet $\psi_j(t-m)$ that are at different displacements will be mutually orthogonal, it is sufficient to impose the condition that

$$\sum_k g_k h_{k+2m} = 0. \tag{34}$$

It is also necessary to ensure the orthogonality of wavelets and scaling functions that are at the same displacement. It is assumed that the two dilation equations contain the same number M of coefficients, and that this is an even number. Then, a sufficient condition for the orthogonality is that

$$\sum_k g_k h_k = 0. \tag{35}$$

If the coefficients of the scaling function dilation equation are $g_0, g_1, \ldots, g_{M-1}$, then the conditions of (33) and (34) can be realised by setting

$$h_k = (-1)^k g_{M-1-k}, \quad \text{which implies that} \quad g_k = (-1)^{k+1} h_{M-1-k}. \tag{36}$$

An example is provided by the case where $M = 4$. Then, there are

$$\begin{aligned} &g_0, && h_0 = g_3, \\ &g_1, && h_1 = -g_2, \\ &g_2, && h_2 = g_1, \\ &g_3, && h_3 = -g_0; \end{aligned} \tag{37}$$

and the conditions of (33) and (34) are clearly satisfied.

The z-Transforms of the Filters

It can be helpful to express the relationships between the coefficients of the two-scale dilations in terms of the z-transforms of their sequences. These are

$$\begin{aligned} G(z) &= g_0 + g_1 z + g_2 z^2 + g_3 z^3 = z^3 H(-z^{-1}), \\ H(z) &= g_3 - g_2 z + g_1 z^2 - g_0 z^3 = -z^3 G(-z^{-1}). \end{aligned} \tag{38}$$

The autocovariance generating functions formed from the coefficients of the dilation equation of the scaling function and the wavelet are

$$P(z) = G(z)G(z^{-1}) \quad \text{and} \quad Q(z) = H(z)H(z^{-1}) = G(-z)G(-z^{-1}) \tag{39}$$

The conditions of sequential orthogonality affecting the scaling function imply that the coefficients of $P(z)$ associated with the even powers of z must be zeros. The coefficients in question are comprised by the function $P(z) + P(-z)$, from which the odd powers of z are absent. On taking account of the condition of

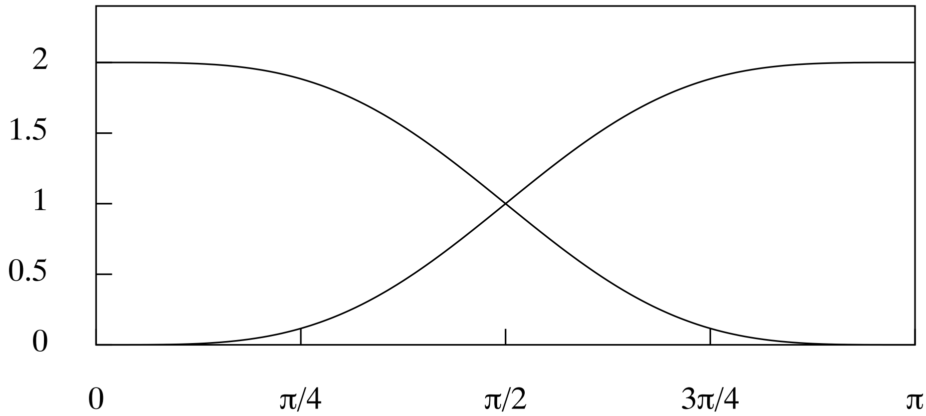

Figure 2. The squared gains of the complementary lowpass and highpass D4 Daubechies filters.

(28) that $p_0 = 1$, it can be seen that the condition for sequential orthogonality is that

$$\begin{aligned} 2 = P(z) + P(-z) &= G(z)G(z^{-1}) + G(-z)G(-z^{-1}) \\ &= G(z)G(z^{-1}) + H(z)H(z^{-1}) \\ &= H(-z)H(-z^{-1}) + H(z)H(z^{-1}) = Q(-z) + Q(z). \end{aligned} \tag{40}$$

The second line of this equations indicates the complementary nature of the highpass and lowpass filters that are derived from the coefficients of the dilation equations. Setting $z = \exp\{\mathrm{i}\omega\}$ with $\omega \in [-\pi, \pi]$ within $H(z)H(z^{-1})$ and $G(-z)G(-z^{-1})$ gives the squared gains of the filters. These are plotted in Figure 2 for the case of the Daubechies D4 filters that are to be specified in the section that follows. The third line of equation indicates the sequential orthogonality of the wavelets.

The cross-covariance generating function formed from the coefficients of the highpass and lowpass filters is $R(z) = G(z)H(z^{-1})$. The condition of (34), which imposes the mutual orthogonality of the wavelets and the scaling functions at displacements that are multiples of two points, is equivalent to the condition that

$$R(z) + R(-z) = G(z)H(z^{-1}) + G(-z)H(-z^{-1}) = 0. \tag{41}$$

Given that

$$\begin{aligned} G(-z) = g_0 - g_1 z + g_2 z^2 - g_3 z^3 &= -z^3 H(z^{-1}) \quad \text{and} \\ H(-z^{-1}) = g_3 + g_2 z^{-1} + g_1 z^{-2} + g_0 z^{-3} &= z^{-3} G(z), \end{aligned} \tag{42}$$

it follows that (41) is satisfied if the coefficients of $G(z)$ and $H(z)$ satisfy the conditions of (37), or more, generally, if they satisfy the conditions of (36).

Combining (36) and (37) and allowing for the substitution of $-z$ for z gives

$$\begin{bmatrix} G(z) & G(-z) \\ H(z) & H(-z) \end{bmatrix} \begin{bmatrix} G(z^{-1}) & H(z^{-1}) \\ G(-z^{-1}) & H(-z^{-1}) \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}. \tag{43}$$

This equation can be rendered in a matrix form. To achieve this, the powers of z can be replaced by powers of a circulant matrix $K_T = [e_1, e_2, \ldots, e_{T-1}, e_0]$. This matrix is formed from the identity matrix $I_T = [e_0, e_1, e_2, \ldots, e_{T-1}]$ by moving the leading vector e_0 to the end of the array. Then, $K_T^0 = I_T$ corresponds to z^0 and K_T' corresponds to z^{-1}; and it will be seen that (43) corresponds to the product of an orthogonal matrix and its transpose.

The conditions that ensure the mutual orthogonality of the elements of the bases of $\mathcal{V}_1$ and $\mathcal{W}_1$, which are associated with the first round of the dyadic decomposition, will guarantee the mutually orthogonality of all of the elements of the final basis that reside in different bands. The orthogonality of such elements is described as *lateral* orthogonality.

Generating Wavelets and Scaling Functions

In the majority of cases, there are no analytic functions to represent the wavelets and the scaling functions in the time domain. Therefore, iterative procedures must be used for generating graphical representations of these functions. Such iterative procedures are based on the appropriate dilation equations.

Consider the equation (19), which can be used to express the scaling function $\phi_1(t)$ in terms of M scaling functions $\phi_1(2t-k); k = 0, 1, \ldots, M-1$ of twice the resolution to give the first expansion:

$$\phi_1(t) = 2^{1/2} \sum_{k=0}^{M-1} g_k \phi_1(2t-k). \tag{44}$$

Each of the overlapping scaling functions on the RHS of (44) are supported on intervals of half the width of the interval supporting $\phi(t)$ and their centres are separated one from the next by distances of 1/2 a unit. The amplitude coefficients $g_0, g_1, \ldots, g_{M-1}$ form a discrete sequence of which the elements can be attributed to the central points of the corresponding wavelets. The profile of this sequence is roughly indicative of that of $\phi(t)$.

The scaling functions on the RHS of (44) are themselves amenable to a second expansion via the dilation equation. Thus

$$\phi(2t-k) = 2^{1/2} \sum_{j=0}^{M-1} g_j \phi(2[2t-k]-j). \tag{45}$$

When equation (45) is substituted into the RHS of equation (44), for all values of k, the result is an expression for $\phi(t)$ that contains M^2 contracted scaling functions. Each of these is supported on an interval that has 1/4th of the length of the support of $\phi(t)$.

In the resulting expression for $\phi(t)$, the centres of the scaling functions on the RHS of (45) are separated one from the next by distances of 1/4 of a unit, as are the corresponding amplitude coefficients. Some of these contracted functions have overlapping supports; and, together, they are supported on the same interval as $\phi(t)$.

The amplitude coefficients come in M batches of M elements at a time. Successive batches, indexed by k, are separated by distances of 1/2 a unit. There are M^2 products $g_k g_j$. Where they are coincident, they are added together, and the profile of the resulting sequence of products is a more refined representation of the profile of $\phi(t)$ than is the original sequence.

The manner in which the coefficients of the second expansion are generated may be illustrated by the case where $M = 4$. The coefficients are the products of the following multiplications:

$$\begin{bmatrix} g_0 & 0 & 0 & 0 \\ g_1 & 0 & 0 & 0 \\ g_2 & g_0 & 0 & 0 \\ g_3 & g_1 & 0 & 0 \\ 0 & g_2 & g_0 & 0 \\ 0 & g_3 & g_1 & 0 \\ 0 & 0 & g_2 & g_0 \\ 0 & 0 & g_3 & g_1 \\ 0 & 0 & 0 & g_2 \\ 0 & 0 & 0 & g_3 \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} g_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ g_1 & g_0 & 0 & 0 & 0 & 0 & 0 \\ g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 \\ g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 \\ 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 \\ 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 \\ 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 \\ 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 \\ 0 & 0 & 0 & 0 & 0 & g_3 & g_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_3 \end{bmatrix} \begin{bmatrix} g_0 \\ 0 \\ g_1 \\ 0 \\ g_2 \\ 0 \\ g_3 \end{bmatrix}. \tag{46}$$

The expression on the LHS corresponds to the manner of forming the coefficients that has already been described. That is to say, four batches of the four coefficients from the second expansion, separated by a fixed interval, are shifted successively by a double shift before being multiplied in turn by the coefficients g_0, g_1, g_2 and g_3 of the first expansion.

The expression on the RHS embodies the lower-triangular Toeplitz matrix of a linear filter. The sequence that is subject to the filter is obtained by interpolating zeros between the elements of the pre-existing vector of derived amplitude coefficients. This is described as an upsampling operation.

Successive expansions of the sum of wavelets can proceed in the manner indicated by the expression on the RHS of (46), by upsampling the sequence of amplitude coefficients derived in the previous expansion and then by subjecting the result to a process of filtering.

The coefficients of the successive expansions gives rise to a sequence of z-transform polynomials of increasing degrees. The filter polynomial $G(z)$, which has M coefficients, constitutes a polynomial of degree $p = M - 1$.

The interpolation of zeros, occcasioned by the upsampling, doubles the degree of the z-transform of the coefficient sequence. The filtering adds p elements to the sequence and it adds p to the degree of the z-transform. Therefore, the degrees of the successive z-transforms are given by the formula

$$q(r) = 2q(r-1) + p; \quad r = \{1, 2, \ldots\}, \quad \text{with} \quad q(0) = p. \tag{47}$$

The solution to this differnce equation is $q(r) = (2^r - 1)p$, which can be confirmed by substituting the solution into the equation.

The solution is also the degree of the product $G(z)G(z^2)\cdots G(z^{2^{r-1}})$. This can be explained by invoking the result of (32), which indicates that the sequence of operations denoted by

$$\delta(z) \longrightarrow (\uparrow 2) \longrightarrow G(z) \longrightarrow \cdots \longrightarrow (\uparrow 2) \longrightarrow G(z^{2^{r-1}}), \tag{48}$$

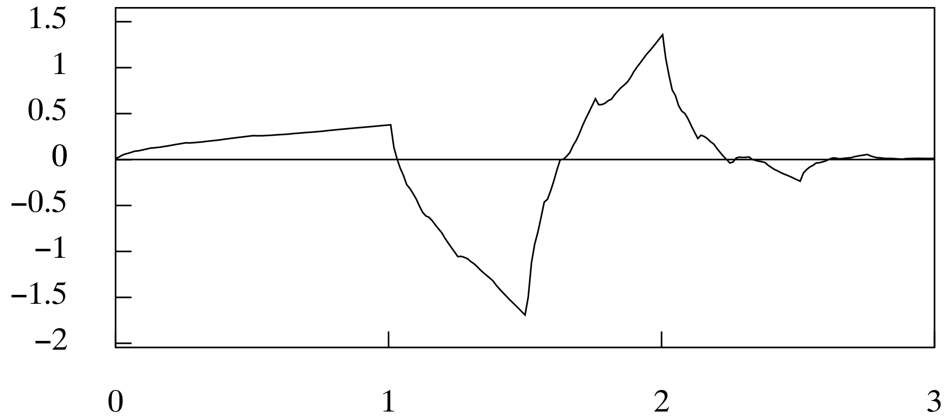

Figure 3. The Daubechies D4 wavelet function calculated via a recursive method.

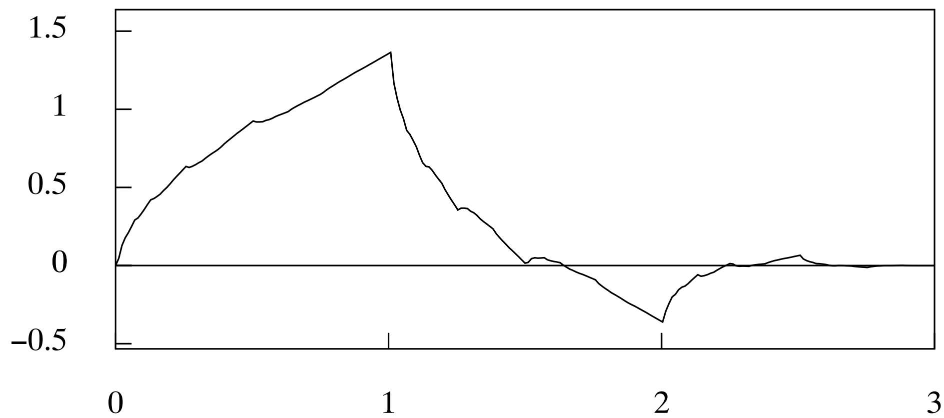

Figure 4. The Daubechies D4 scaling function calculated via a recursive method.

which comprises r expansions, is equivalent to the following sequence of operations

$$\delta(z) \longrightarrow (\uparrow 2^r) \longrightarrow G(z)G(z^2)\cdots G(z^{2^{r-1}}). \tag{49}$$

Here, $\delta(z)$ denotes the impulse function consisting of a single unit preceded and followed by an indefinite number of zeros. Upsampling this function r times has no noticable effect.

The number of coefficients in each expansion is one more than the degree of the corresponding z-transform, For the rth expansion, the number is

$$M(r) = (2^r - 1)(M - 1) + 1 \leq 2^r(M - 1). \tag{50}$$

In the case of $M = 4$, the number of coefficients in succesive expansions increases as the sequence $\{1, 4, 19, 22, 46.\ldots\}$. Here, the unit that begins the sequence refers to the impulse function $\delta(z)$, which comprises a single coefficient associated with z^0.

As the number of expansions increases, an increasing number of amplitude coefficients are mapped into the interval that supports $\phi(t)$. In the process, the supports of the wavelets associated with the coefficients are successively diminished. Eventually, the wavelets will be adequately represented by spikes of unit area based on a point, which are Dirac delta functions. By that stage, the profile of $\phi(t)$ will be well represented by the closely spaced sequence of the derived amplitude coefficients.

For the expansions to give rise to a sequence of functions that converges on $\phi(t)$, rectangles of the appropriate height and width must be associated to each of the coefficients. The height must increase and width must diminish with the order of the expansion.

To begin, a unit rectangle is associated with the impulse function $\delta(z)$. Next, to each of the coefficients of $G(z)$ is associated a rectangle of height $\sqrt{2}$ and of width 1/2. Within the rth expansion, the coefficients are associated with rectangles of height $2^{r/2}$ supported on a bases of width $1/2^r$. Thus, if $g_k^{(r)}$ represents the kth coefficient within the rth expansion, then the piecewise constant representation of $\phi(t)$ will be

$$\phi^{(r)}(t) = 2^{r/2} g_k^{(r)} \quad \text{when} \quad \frac{k}{2^r} \leq t < \frac{k+1}{2^r}, \tag{51}$$

and then $\phi^{(r)}(t) \to \phi(t)$ as $r \to \infty$. Moreover. the functions $\phi^{(r)}(t)$ will be confined to the interval $[0, M-1]$, with supports that tend to the upper bound as $r \to \infty$.

The dilation equation for the wavelets functions at level 1, which is given in (24), can be represented as follows:

$$\psi_1(t) = 2^{1/2} \sum_{k=0}^{M-1} h_k \phi_1(2t - k). \tag{52}$$

This can be expanded in the same way as the scaling function to generate a sequence of closely spaced coefficients that will represent the profile of the wavelet.

An example of a pair of wavelet and scaling functions that have dilation equations of four coefficients is provided by Daubechies' D4 functions. In this case, there are

$$\begin{aligned} g_0 &= (1+\sqrt{3})/(4\sqrt{2}), \qquad & g_1 &= (3+\sqrt{3})/(4\sqrt{2}), \\ g_2 &= (3-\sqrt{3})/(4\sqrt{2}), \qquad & g_3 &= (1-\sqrt{3})/(4\sqrt{2}), \end{aligned} \tag{53}$$

and there are $h_0 = g_3$, $h_1 = -g_2$, $h_2 = g_1$ and $h_3 = -g_0$, in accordance with (37). The profiles of the functions are represented in Figures 3 and 4.

The Decomposition of a Function in $\mathcal{V}_0$

The equations of the two-scale dilation provide the means for pursuing an hierarchical decomposition of a function $y(t)$ into a sum of orthogonal components residing in the subspaces spanned by the wavelets basis functions. Such a decomposition is represented by equation (8).

If $y(t) \in \mathcal{V}_0$, which is to say that the function truly resides in the space $\mathcal{V}_0$, then equation (1) is no longer an approximation, and it becomes

$$y(t) = \sum_k \langle y(t), \phi_{0,k}(t) \rangle \phi_{0,k}(t) = \sum_k y_k \phi_0(t-k). \tag{54}$$

This equation represents the projection of $y(t)$ on the basis vectors of $\mathcal{V}_0$. Since $\mathcal{V}_0 = \mathcal{W}_1 \oplus \mathcal{V}_1$, an alternative representation of $y(t)$ is obtained by projecting it on the conjunction of the level-1 basis vectors of $\mathcal{W}_1$ and $\mathcal{V}_1$, to generate the orthogonal components $v_1(t)$ and $w_1(t)$ of $y(t) = w_1(t) + v_1(t)$.

The projection on the level-1 scaling functions gives

$$v_1(t) = \sum_m \langle y(t), \phi_{1,m}(t) \rangle \phi_{1,m}(t) = \sum_m \gamma_{1,m} \phi_{1,m}(t), \tag{55}$$

and we require to evaluate the coefficient $\gamma_{1,m}$. In view of the dilation equations of (12) and(16), there is

$$\begin{aligned} \phi_{1,m}(t) = 2^{-1/2}\phi_0(2^{-1}t - m) &= \sum_k g_k \phi_0(2[2^{-1}t - m] - k) \\ &= \sum_k g_k \phi_0(t - [2m + k]) \end{aligned} \tag{56}$$

Therefore, the coefficient associated with the basis function $\phi_{1,m}(t)$ is

$$\begin{aligned} \gamma_{1,m} = \langle y(t), \phi_{1,m}(t) \rangle &= \sum_k g_k \langle y(t), \phi_{0,2m+k}(t) \rangle \\ &= \sum_k g_k y_{2m+k}. \end{aligned} \tag{57}$$

The projection of $y(t)$ on the level-1 wavelets gives

$$w_1(t) = \sum_m \langle y(t), \phi_{1,m}(t) \rangle \psi_{1,m}(t) = \sum_m \beta_{1,m} \psi_{1,m}(t). \tag{58}$$

In this case, it can be shown, as in the case of the scaling functions, that the coefficient associated with $\psi_{1,m}(t)$ is

$$\beta_{1m} = \langle y(t), \psi_{1,m}(t) \rangle = \sum_k h_k y_{2m+k}. \tag{59}$$

The equations

$$\gamma_{1m} = \sum_{k=0}^{T-1} g_k y_{2m+k}; \quad m = 0, 1, \ldots, [T/2] - 1, \tag{60}$$

$$\beta_{1m} = \sum_{k=0}^{T-1} h_k y_{2m+k}; \quad m = 0, 1, \ldots, [T/2] - 1, \tag{61}$$

which deliver the amplitude coefficients of the level-1 scaling functions and wavelets respectively, can be construed as the equations of a pair of complementary linear filters that are applied to a common data sequence $y_0, y_1, \ldots, y_{T-1}$ of T elements. Equation (60) describes a lowpass filter and equation (61) describes a highpass filter.

These filters move through the sample in step with the index $2m$, which is to say that they take steps of two points at a time. When this index is replaced by one with unit increments, it becomes necessary to select alternate values of the filtered outputs via a process that is commonly described as down sampling.

Since the data sequence is finite, there will be problems in applying the filters at the ends of the sample where data are required that lie beyond the ends. To overcome the problem, the filter can be applied to the data via a process of circular convolution, which is equivalent to applying the filter to the periodic extension of the data.

To accommodate this adaptation within equation (60) and (61), it is sufficient to replace y_k by $y_{k \ mod \ T}$. When $t \in \{0, T-1\}$ there will be $y_{k \ mod \ T} = y_k$. Otherwise, when it appears to lie outside the sample, y_k will be replaced by a value from within the sample.

The second stage of the decomposition, as well as all subsequent stages, can be modelled on the first stage. Thus, the coefficients of the second stage are given by

$$\gamma_{2n} = \sum_{k=0}^{[T/2]-1} g_k \gamma_{1,2n+k}; \quad n = 0, 1, \ldots, [T/4] - 1, \tag{62}$$

$$\beta_{2n} = \sum_{k=0}^{[T/2]-1} h_k \gamma_{1,2n+k}; \quad n = 0, 1, \ldots, [T/4] - 1. \tag{63}$$

The complete process of decomposition is best represented using a matrix notation.

A Matrix Formulation for Wavelets Computations

Let $y = [y_0, \ldots, y_{T-1}]'$, where $T = 2^n$, represent the vector of observations, which are associated with the scaling functions of the initial basis, and let β represent the vector of order T of the coefficients associated with the wavelets of the final basis. The mapping from y to β, denoted by $\beta = Q'y$, is effected by an orthonormal matrix Q such that $QQ' = Q'Q = I_T$.

Since $(Q')^{-1} = Q$, it follows that there is an inverse transformation from the wavelet coefficients to the data of the form $Q\beta = y$. This mapping from β to y effects a wavelet synthesis. If β contains a single nonzero element representing the amplitude coefficient of a solitary wavelet, then the mapping of β via Q will generate the vector, corresponding to a single column of Q, containing elements that approximate the ordinates of that wavelet, sampled at unit intervals.

The T elements of the vector β can be ordered in a manner that corresponds to a dyadic decomposition, such as is illustrated in Figure 1. Then, within β, there will be a succession of subvectors, which contain the coefficients associated with the hierarchy of the wavelet functions of the final basis. The subvectors of

$$\beta = [\beta'_{(1)}, \beta'_{(2)}, \ldots, \beta'_{(n)}, \gamma'_{(n)}]' \tag{64}$$

are

$$\begin{aligned} \beta_{(1)} &= [\beta_{10}, \beta_{11}, \ldots, \beta_{1,[T/2]-1}]', \\ \beta_{(2)} &= [\beta_{20}, \beta_{21}, \ldots, \beta_{2,[T/4]-1}]', \\ &\vdots \\ \beta_{(n-1)} &= [\beta_{n-1,0}, \beta_{n-1,1}]', \\ \beta_{(n)} &= [\beta_{n0},], \\ \gamma(n) &= [\gamma_{n0}]. \end{aligned} \tag{65}$$

The procedure that generates the succession of wavelet amplitude coefficients has been described as the pyramid algorithm, albeit that the sequence of coefficients form an inverted pyramid.

A linear filter can be applied to a finite data sequence via a matrix transformation of the vector y of the data. Let the z-transform of the filter be represented by the polynomial in z, such that the polynomial corresponding to the lowpass filter is $G(z) = g_0 + g_1 z + \cdots + g_{M-1} z^{M-1}$, and assume that the filter is applied to a data vector of T elements via a process of circular convolution.

Then, the matrix transformation that implements the filter can obtained by replacing the powers of z by powers of a circulant matrix $K_T = [e_{T-1}, e_0, \ldots, e_{T-2}]$. This matrix is formed from the identity matrix $I_T = [e_0, e_1, e_2, \ldots, e_{T-1}]$ by moving the last vector e_{T-1} to the beginning of the array. The resulting matrix is

$$G(K_T) = g_0 I_T + g_1 K_T + \cdots + g_{M-1} K_T^{M-1}, \tag{66}$$

where $I_T = K_T^0$, and the filtered vector is given by $Gy = G(K_T)y$.

A process of down sampling can also be affected by a matrix transformation. The down sampling matrix is $\mathrm{V} = \Lambda' = [e_0, e_2, e_4, \ldots, e_{T-2}]'$, which is obtained by deleting alternate rows from the identity matrix I_T.

To see in detail how the wavelet amplitude coefficients can be generated in this manner, it is best to take a specific example. In the example, there are $T = 8 = 2^3$ data points and there are $M = 4$ coefficients in the dilation equations of (16) and (21), which constitute the coefficients of a lowpass and a high pass filter. Each stage of the process that converts the data into the wavelet coefficients involves the application of a circulant linear filter followed by a process of down sampling.

The highpass filter that is to be applied to the data in the first round of the wavelets decomposition has the following circulant matrix representation:

$$H_{(1)} = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 \\ h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 \end{bmatrix}. \tag{67}$$

This matrix $H_{(1)} = H(K_8)$ is obtained by replacing the argument z within the polynomial $H(z) = h_0 + h_1 z + h_2 z^2 + h_3 z^3$ by the circulant matrix $K_8 = [e_7, e_0, \ldots, e_5, e_6]$. Premultiplying $H_{(1)}$ by the down sampling matrix is a matter of deleting alternate rows:

$$M_{(1)} = \mathrm{V}H_{(1)} = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 \end{bmatrix}. \tag{68}$$

When this matrix is combined with the matrix $L_{(1)} = \mathrm{V}G_{(1)}$, which is the down sampled version of the lowpass filter matrix $G_{(1)} = G(K_8)$, and when the data vector y is mapped through the combined matrix, the result is

$$\begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \hline \gamma_{10} \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \end{bmatrix} = \left[\begin{array}{cccccccc} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ \hline g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_0 & g_1 & g_2 & g_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 \\ g_2 & g_3 & 0 & 0 & 0 & 0 & g_0 & g_1 \end{array}\right] \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix}. \tag{69}$$

An alternative arrangement of the equations of (69) is described as a Lapped Orthogonal Transform (LOT):

$$\begin{bmatrix} \beta_{10} \\ \gamma_{10} \\ \hline \beta_{11} \\ \gamma_{11} \\ \hline \beta_{12} \\ \gamma_{12} \\ \hline \beta_{13} \\ \gamma_{13} \end{bmatrix} = \left[\begin{array}{cccccccc} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & g_0 & g_1 & g_2 & g_3 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 \\ \hline h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ g_2 & g_3 & 0 & 0 & 0 & 0 & g_0 & g_1 \end{array}\right] \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix}. \tag{70}$$

We shall have occasion to use this arrangement at a later stage when we deal with the transform coding that is used commonly in connection with digital images.

The equations of (69) can be represented, in summary notation, by

$$\begin{bmatrix} \beta_{(1)} \\ \gamma_{(1)} \end{bmatrix} = \begin{bmatrix} \mathrm{V}H_{(1)} \\ \mathrm{V}G_{(1)} \end{bmatrix} y. \tag{71}$$

This equation will serve to represent the general case, where $T = 2^n$ and $M = 2q$ are unspecified values.

The conditions of sequential and lateral othogonality imply that the matrix of (71) is an orthonomal matrix of which the transpose is the inverse. Therefore,

$$\begin{bmatrix} H'_{(1)}\Lambda & G'_{(1)}\Lambda \end{bmatrix} \begin{bmatrix} \mathrm{V}H_{(1)} \\ \mathrm{V}G_{(1)} \end{bmatrix} = I. \tag{72}$$

It follows that the data vector y can be recovered from $\gamma_{(1)}$ and $\beta_{(1)}$ via

$$y = H'_{(1)}\Lambda\beta_{(1)} + G'_{(1)}\Lambda\gamma_{(1)}. \tag{73}$$

In terms of the example with $T = 8$ and $M = 4$, this can be rendered as

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix} = \begin{bmatrix} h_0 & 0 & 0 & h_2 \\ h_1 & 0 & 0 & h_3 \\ h_2 & h_0 & 0 & 0 \\ h_3 & h_1 & 0 & 0 \\ 0 & h_2 & h_0 & 0 \\ 0 & h_3 & h_1 & 0 \\ 0 & 0 & h_2 & h_0 \\ 0 & 0 & h_3 & h_1 \end{bmatrix} \begin{bmatrix} \gamma_{10} \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \end{bmatrix} + \begin{bmatrix} g_0 & 0 & 0 & g_2 \\ g_1 & 0 & 0 & g_3 \\ g_2 & g_0 & 0 & 0 \\ g_3 & g_1 & 0 & 0 \\ 0 & g_2 & g_0 & 0 \\ 0 & g_3 & g_1 & 0 \\ 0 & 0 & g_2 & g_0 \\ 0 & 0 & g_3 & g_1 \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \end{bmatrix}. \tag{74}$$

It will be recognised that the component $G'_{(1)}\Lambda\gamma_{(1)}$ within this equation is closely related to the expression on the LHS of equation (46), which represents the first stage in the iterative dilation algorithm that leads, eventually, to a representation of the scaling function. The expressions differ in consequence of the circularity that has been imposed on equation (74) and in the replacement of the vector $[g_0, g_1, g_2, g_3]'$ in (46) by $[\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}]'$, in (74).

In the second round of the wavelets decomposition, the coefficients associated with the level-1 wavelets are preserved and the coefficients associated with the level-1 scaling functions are subject to a further decomposition:

$$\begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \hline \beta_{20} \\ \beta_{21} \\ \hline \gamma_{20} \\ \gamma_{21} \end{bmatrix} = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ 0 & 0 & 0 & 0 & h_2 & h_3 & h_0 & h_1 \\ \hline 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 \\ 0 & 0 & 0 & 0 & g_2 & g_3 & g_0 & g_1 \end{array}\right] \begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \hline \gamma_{10} \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \end{bmatrix}. \tag{75}$$

The summary notation for this is

$$\begin{bmatrix} \beta_{(1)} \\ \beta_{(2)} \\ \gamma_{(2)} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \mathrm{V}H_{(2)} \\ 0 & \mathrm{V}G_{(2)} \end{bmatrix} \begin{bmatrix} \beta_{(1)} \\ \gamma_{(1)} \end{bmatrix}. \tag{76}$$

The first row within $\mathrm{V}G_{(2)}$ is derived by subtracting a sufficient number of trailing zeros from the first row of $G_{(1)}$ to halve its length. Subsequent rows are obtained by a double shift to the right of the preceding row, while taking account of is circulant nature. The rows of $\mathrm{V}H_{(2)}$ are derived likewise.

The effect of the downsampling upon the circular filter can be see in equation (75). The two filters are defined on four points and, at this level, only four data points are available. There are no zeros remaining within the matrices $\mathrm{V}H_{(2)}$ and $\mathrm{V}G_{(2)}$.

In the next and final round of filtering, there are only two data points to be mapped through the filters. The consequence is that γ_{20} and γ_{21} must be used twice in the third and final transformation. This can represented equally by

$$\begin{bmatrix} \gamma_{30} \\ \beta_{30} \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 \\ g_0 & g_1 & g_2 & g_3 \end{bmatrix} \begin{bmatrix} \gamma_{20} \\ \gamma_{21} \\ \gamma_{20} \\ \gamma_{21} \end{bmatrix} \tag{77}$$

or by

$$\begin{bmatrix} \gamma_{30} \\ \beta_{30} \end{bmatrix} = \begin{bmatrix} h_0 + h_2 & h_1 + h_3 \\ g_0 + g_2 & g_1 + g_3 \end{bmatrix} \begin{bmatrix} \gamma_{20} \\ \gamma_{21} \end{bmatrix}. \tag{78}$$

On the LHS is a vector containing the amplitude coefficients, respectively, of a wavelet and a scaling function stretching the length of the data sequence.

A general expression can now be given for the set of amplitude coefficients associated with the projection of the function $y(t)$ onto the basis of the subspace $\mathcal{W}_j$. These coefficients are contained in the jth vector of the sequence of (65), which is given by

$$\beta_{(j)} = \mathrm{V}H_{(j)}\mathrm{V}G_{(j-1)}\cdots\mathrm{V}G_{(1)}y = Q'_{(j)}y. \tag{79}$$

In order to relieve the burden of notation, the subscripts have been omitted from the succession of down sampling matrices that would indicate their orders. Reading from right to left, the first down sampling matrix is $\mathrm{V}_{(1)}$ of order $T/2 \times T$. The penultimate matrix is $\mathrm{V}_{(j-1)}$ of order $T/2^{j-1} \times T/2^{j-2}$ and the final matrix is $\mathrm{V}_{(j)}$ of order $T/2^j \times T/2^{j-1}$.

On the RHS of (79) is the matrix $Q'_{(j)}$, which represents a submatrix formed from a set of adjacent rows of the matrix Q', which is entailed in the mapping $\beta = Q'y$ from the sampled ordinates of $y(t)$ to the amplitude coefficients of the final basis. Given that $QQ' = I_T$, it follows that $Q\beta = y$ represents the synthesis of the vector y from the amplitude coefficients.

The vector $\beta_{(j)}$ of (79) is entailed in the synthesis of the component vector $w_j = [w_{0j}, w_{1j}, \ldots, w_{T-1,j}]'$ of the decomposition of $y = w_1 + \cdots + w_n + v_n$. The synthesis can be represented by

$$w_j = Q_{(j)}\beta_{(j)} = G'_{(1)}\Lambda\cdots G'_{(j-1)}\Lambda H'_{(j)}\Lambda\beta_{(j)}, \tag{80}$$

where $\Lambda = \mathrm{V}'$ represents the upsampling matrix, which interpolates zeros between the elements of any vector than it premultiplies.

The Two-Phase Representation

There is an alternative way of formulating the equations of the analysis and the synthesis stages of a wavelet transformation that indicates a significant gain in efficiency, when the processes are conducted in real time, but which has no significant advantage when the calculations are performed off-line. This is a two-phase formulation that depends on a separations the data sequence into two subsequences of which one contains the elements with even-valued indices and the other contains the remaining elements with odd-valued indices.

The two-phase formulation can be illustrated by taking, once more, the case where $T = 8$ and $M = 4$. An alternative way of formatting the equations of (69), which represent the first stage of the analysis transformation, is as follows:

$$\begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \hline \gamma_{10} \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \end{bmatrix} = \begin{bmatrix} h_0 & h_2 & 0 & 0 \\ 0 & h_0 & h_2 & 0 \\ 0 & 0 & h_0 & h_2 \\ h_2 & 0 & 0 & h_0 \\ \hline g_0 & g_2 & 0 & 0 \\ 0 & g_0 & g_2 & 0 \\ 0 & 0 & g_0 & g_2 \\ g_2 & 0 & 0 & g_0 \end{bmatrix} \begin{bmatrix} y_0 \\ y_2 \\ y_4 \\ y_6 \end{bmatrix} + \begin{bmatrix} h_1 & h_3 & 0 & 0 \\ 0 & h_1 & h_3 & 0 \\ 0 & 0 & h_1 & h_3 \\ h_3 & 0 & 0 & h_1 \\ \hline g_1 & g_3 & 0 & 0 \\ 0 & g_1 & g_3 & 0 \\ 0 & 0 & g_1 & g_3 \\ g_3 & 0 & 0 & g_1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_3 \\ y_5 \\ y_7 \end{bmatrix}. \tag{81}$$

Omitting some of the subscripts that indicate the level of the decomposition, this can be represented, in summary notation, by

$$\begin{bmatrix} \beta_{(1)} \\ \gamma_{(1)} \end{bmatrix} = \begin{bmatrix} H^e \\ G^e \end{bmatrix} y^e + \begin{bmatrix} H^o \\ G^o \end{bmatrix} y^o, \tag{82}$$

where $y^e = [y_0, y_2, y_4, y_6]' = \mathrm{V}y$ contains the elements of y with even indices and $y^o = [y_1, y_3, y_5, y_7]' = \mathrm{V}Ky$ contains those with odd indices. Here, $K = [e_{T-1}, e_0, \ldots, e_{T-2}]$ is the circulant lag operator, which is obtained from the identity matrix $I = [e_0, e_1, e_2, \ldots, e_{T-1}]$ by moving the last column to the beginning of the array. Premultiplying a matrix by K displaces it rows upwards, throwing the first row to the bottom. (Postmultiplying by a matrix by K move its first column to the back of the array.)

Defining $\Lambda = \mathrm{V}'$, which is the upsampling matrix, gives

$$\begin{aligned} G^e &= \mathrm{V}G_{(1)}\Lambda = g_0 I_{T/2} + g_2 K_{T/2} \quad \text{and} \\ G^o &= \mathrm{V}G_{(1)}K\Lambda = g_1 I_{T/2} + g_3 K_{T/2}, \end{aligned} \tag{83}$$

which are the circulant filter matrices of a reduced order that are based on the coefficients of even and odd indices respectively of the filter polynomial $G(z) = g_0 + g_1 z + g_2 z^2 + g_3 z^2$.

Now consider again the synthesis equation (73), which represents the inverse of the first stage of the analysis:

$$y = H'_{(1)}\Lambda\beta_{(1)} + G'_{(1)}\Lambda\gamma_{(1)}. \tag{84}$$

This has been exemplified in the case of $T = 8$ and $m = 4$ by equation (74). An alternative way of formatting equation (74) is

$$\begin{bmatrix} y_0 \\ y_2 \\ y_4 \\ y_6 \\ \hline y_1 \\ y_3 \\ y_5 \\ y_7 \end{bmatrix} = \begin{bmatrix} h_0 & 0 & 0 & h_2 \\ h_2 & h_0 & 0 & 0 \\ 0 & h_2 & h_0 & 0 \\ 0 & 0 & h_2 & h_0 \\ \hline g_0 & 0 & 0 & g_2 \\ g_2 & g_0 & 0 & 0 \\ 0 & g_2 & g_0 & 0 \\ 0 & 0 & g_2 & g_0 \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \end{bmatrix} + \begin{bmatrix} h_1 & 0 & 0 & h_3 \\ h_3 & h_1 & 0 & 0 \\ 0 & h_3 & h_1 & 0 \\ 0 & 0 & h_3 & h_1 \\ \hline g_1 & 0 & 0 & g_3 \\ g_3 & g_1 & 0 & 0 \\ 0 & g_3 & g_1 & 0 \\ 0 & 0 & g_3 & g_1 \end{bmatrix} \begin{bmatrix} \gamma_{10} \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \end{bmatrix}. \tag{85}$$

In the summary notation, this becomes

$$\begin{bmatrix} y^e \\ y^o \end{bmatrix} = \begin{bmatrix} H^{e\prime} \\ H^{o\prime} \end{bmatrix} \beta_{(1)} + \begin{bmatrix} G^{e\prime} \\ G^{o\prime} \end{bmatrix} \gamma_{(1)}. \tag{86}$$

The advantage of the two-phase formulation, when the processing in real time, is due to the splitting of the data into odd and even subsequences prior to the transformation of the subsequences. In the direct formulation, the data sequence in its entirety is liable to be transformed by the high pass filter and the low pass filters before downsampling. This is twice the labour that is entailed in transforming the subsequences via the two-phase procedure.

The Maximal Overlap Discrete Wavelet Transform (MODWT)

A Maximal Overlap Discrete Wavelet Transform (MODWT) arises from a dyadic Discrete Wavelet Transform (DWT) when the operation of down sampling, which accompanies each stage of the pyramid algorithm, is omitted. In place of the $T/2^j$ wavelet amplitude coefficients arising from the jth stage of the DWT, the MODWT generates a full vector β_j of T coefficients. Morever, there is no longer any pupose in the restriction that $T = 2^n$.

A further transformation, analogous to the inverse DWT, can be applied to the vector β_j to generate a vector w_j, which is a component of the data that corresponds to the jth band in a dyadic decomposition of the frequency spectrum. In contrast to those generated by a DWT, the MODWT components are not mutually orthogonal.

In describing the MODWT and in comparing it with the DWT, attention can be concentrated on the first stage of a multistage decomposition, from which, initially, the subscripts, indicating that stage can be omitted. To find the products of the MODWT transformation, we use the undecimated matrices $G_1 = G_{(1)}$ and $H_1 = H_{(1)}$, which represent the half-band lowpass filter and the half-band highpass filter, respectively. Then, instead of the equations of (71), there are

$$\beta_1 = \frac{1}{\sqrt{2}} H_1 y \qquad \text{and} \qquad \gamma_1 = \frac{1}{\sqrt{2}} G_1 y. \tag{87}$$

which lead to

$$\begin{aligned} v_1 &= \frac{1}{\sqrt{2}} G_1' \gamma = \frac{1}{2} G_1' G_1 y, \\ w_1 &= \frac{1}{\sqrt{2}} H_1' \beta = \frac{1}{2} H_1' H_1 y. \end{aligned} \tag{88}$$

Here, the scale factor of $1/2$ is to ensure that the addition of the two components gives the original data vector, which is to say that

$$w_1 + v_1 = y_1. \tag{89}$$

To confirm this identity, the matrices G_1 and H_1 may be depicted in terms of the downsampled matrices $L_1 = \mathrm{V} G_1$ and $M_1 = \mathrm{V} H_1$ and of the matrices which are their complements within G_1 and H_1, which may be denoted by $L_{1\triangledown}$ and $H_{1\triangledown}$.

To derive the $L_{1\triangledown}$ and $H_{1\triangledown}$ from their parent matrices, a second downsampling matrix is considered, which is defined by $\nabla = [e_1, e_3, e_5, \ldots, e_{T-1}]'$ and which selects the alternate rows of a matrix that are excluded by the matrix $\mathrm{V} = [e_0, e_2, e_4, \ldots, e_{T-2}]'$ defined previously. The transpose of ∇ is $\Delta = \nabla'$. Observe that, whereas $\Lambda \mathrm{V} = [e_0, 0, e_2, \ldots, e_{T-2}, 0]$, there is $\Delta\nabla = [0, e_1, 0, \ldots, 0, e_{T-1}]$. Therefore, $\Delta\nabla + \Lambda\mathrm{V} = I_T$.

It is easy to see that $L_{1\triangledown} = \nabla G_1$ and $M_{1\triangledown} = \nabla H_1$. It follows that

$$M_{1\triangledown}' M_{1\triangledown} + M_1' M_1 = H_1'(\Delta\nabla + \Lambda\mathrm{V}) H_1 = H_1' H_1. \tag{90}$$

Likewise, there is

$$L_{1\triangledown}' L_{1\triangledown} + L_1' L_1 = G_1'(\Delta\nabla + \Lambda\mathrm{V}) G_1 = G_1' G_1. \tag{91}$$

Observe that, in terms of $L_1 = \mathrm{V} G_1$ and $M_1 = \mathrm{V} H_1$, the identity of (72) can be written as

$$M_1' M_1 + L_1' L_1 = I_T. \tag{92}$$

Now it will be recognised that, likewise, there is the identity

$$M_{1\triangledown}' M_{1\triangledown} + L_{1\triangledown}' L_{1\triangledown} = I_T. \tag{93}$$

It follows from (92) to (93) that

$$H_1' H_1 + G_1' G_1 = (M_{1\triangledown}' M_{1\triangledown} + L_{1\triangledown}' L_{1\triangledown}) + (M_1' M_1 + L_1' L_1) = 2 I_T. \tag{94}$$

This justifies the assertion that

$$v_1 + w_1 = \frac{1}{2}(H_1' H_1 + G_1' G_1) y = y. \tag{95}$$

It should be observed that, in general $v_j'w_j \neq 0$, which is to say that, in contrast to those of the DWT, the vectors of the MODWT are not subject to a condition of orthogonality. However, from (87) and (94), it also follows that

$$y'y = \gamma_1'\gamma_1 + \beta_1'\beta_1. \tag{96}$$

It may also be noted that the outcome of the MODWT is invariant with respect to a circular shift in the data relative to the filter matrix or vice versa. First, if the matrices K^qG_1 and K^qH_1 are used in place of G_1 and H_1, then there is no effect on v_1 and w_1, since $G_1'(K')^qK^qG_1 = G_1'G_1$ and $H_1'(K')^qK^qH_1 = H_1'H_1$. Therefore, the equations of (88) are not affected.

Next, in view of the commutativity in multiplication of circulant matrices, there is $K^qG_1 = G_1K^q$ and $K^qH_1 = H_1K^q$. It follows that $v_1 = (K')^q\frac{1}{2}G_1'G_1K^qy$ whence, on premultiplying by K^q, we get $K^qv = \frac{1}{2}G_1'G_1(K^qy)$, which shows that, if y maps into v_1, then K^qy maps into K^qv_1. An analogous result holds in respect of the mapping of y into w_1.

We may now proceed to the second round of the MODWT. For this, we require to define the lowpass filter matrix

$$G_2 = \begin{bmatrix} g_0 & 0 & g_1 & 0 & g_2 & 0 & g_3 & 0 \\ 0 & g_0 & 0 & g_1 & 0 & g_2 & 0 & g_3 \\ g_3 & 0 & g_0 & 0 & g_1 & 0 & g_2 & 0 \\ 0 & g_3 & 0 & g_0 & 0 & g_1 & 0 & g_2 \\ g_2 & 0 & g_3 & 0 & g_0 & 0 & g_1 & 0 \\ 0 & g_2 & 0 & g_3 & 0 & g_0 & 0 & g_1 \\ g_1 & 0 & g_2 & 0 & g_3 & 0 & g_0 & 0 \\ 0 & g_1 & 0 & g_2 & 0 & g_3 & 0 & g_0 \end{bmatrix}, \tag{97}$$

together with the corresponding highpass filter matrix H_2.

The first row within G_2 is derived by that of G_1 by interpolating zeros between its elements and, thereafter, by subtracting a sufficient number of trailing zeros to create a vector of T elements. Subsequent rows are obtained by a single shift to the right of the preceding row, while taking account of its circulant nature. The matrix H_2 is derived from H_1 in the same manner.

A comparison with the matrix of the second stage of the DWT, depicted in (75) and (76), shows that $\mathrm{V}G_{(2)} = \mathrm{V}G_2\Lambda$. In effect, the zeros that have been interpolated within the rows of G_2 are to accommodate the fact that there has been no downsampling of the vector generated by the first round of the MODWT. The extra rows of G_2 generate the extra coefficients that are interpolated between the elements of the vector of amplitude coefficients of the DWT.

At the end of the previous chapter it has been shown that the product G_2G_1 will result in quarter-band lowpass filter, albeit that the demonstration had been in terms of z-transform polynomials instead of circulant matrices.

For the next stags of the decomposition there are.

$$\beta_2 = \frac{1}{\sqrt{2}}H_2\gamma_1 = \frac{1}{2}H_2G_1y \qquad \text{and} \qquad \gamma_2 = \frac{1}{\sqrt{2}}G_2\gamma_1 = \frac{1}{2}G_2G_1y. \tag{99}$$

which lead to

$$\begin{aligned} v_2 &= \frac{1}{4} G_1' G_2' G_2 G_1 y, \\ w_2 &= \frac{1}{4} G_1' H_2' H_2 G_1 y. \end{aligned} \tag{99}$$

since $v_1 = w_2 + v_2$, equation (25) gives

$$w_1 + w_2 + v_2 = y. \tag{100}$$

There is also

$$y'y = \gamma_2'\gamma_2 + \beta_2'\beta_2 + \beta_1'\beta_1. \tag{101}$$

It is easy to see how the decomposition can be pursued through further rounds.

An advantage that can be claimed for the undecimated MODWT, in comparison of the ordinary DWT, is the ease with which the products of the various stages of the decomposition can be aligned with each other to reveal the contents of the data within the frequency bands.

In the case of the DWT, the intervals between successive wavelet amplitude coefficients is doubled in the descent from one band to the next. This can hinder a visual analysis.

It should be recognised that the MODWT is nothing other than a filter bank that employs filters that are peculiar to a wavelets analysis and that labours under the imposition of a dyadic filter-band structure. Except when the MODWT is to be employed as an accompaniment of a DWT analysis, there is no inherent advantage in these restrictions.