Applications of Mathematics

Jaroslav Kautský

An algebraic construction of discrete wavelet transforms

Applications of Mathematics, Vol. 38 (1993), No. 3, 169–193

Persistent URL: `http://dml.cz/dmlcz/104545`

Terms of use:

© Institute of Mathematics AS CR, 1993

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* `http://dml.cz`

AN ALGEBRAIC CONSTRUCTION OF DISCRETE WAVELET TRANFORMS

JAROSLAV KAUTSKY, Adelaide

(Received April 14, 1992)

Summary. Discrete wavelets are viewed as linear algebraic transforms given by banded orthogonal matrices which can be built up from small matrix blocks satisfying certain conditions. A generalization of the finite support Daubechies wavelets is discussed and some special cases promising more rapid signal reduction are derived.

Keywords: orthogonal transform, wavelet, pyramidal algorithm

AMS classification: primary 15A04, secondary 65F25, 65T99

1. INTRODUCTION

Wavelets are discussed explicitly, or in various disguises, in a wide range of literature on signal and image processing. Their mathematical foundation has been firmed in the language of functional analysis through the works of Meyer [3], Daubechies [1] and Mallat [2], to name just a few. Strang gives an excellent overview in [4].

Our aim in this paper is to give an algebraic exposition of discrete wavelet transforms. After all, in any practical application (including those of signal and image processing) we finish up working with finite sequences of values, often restricted to a fairly limited range. We attempt to motivate the properties required of the transformation of data and show that these algebraic conditions fully determine the linear transformations equivalent to those emerging as the discrete implementation of the analytically derived wavelets. Furthermore, we propose a generalization which, although mentioned in [1], does not appear to be treated in an analytical context, and show the existence of such generalized wavelet transforms derived by simple algebraic means.

The paper is organized as follows. In the two parts of §2 we motivate the algebraic approach both by recalling the discretization of Mallat's multiresolution analysis as well as by presenting a self-contained argument for the need for linear transforms with certain properties. In §3 we give a formal definition of the discrete transform by introducing a banded matrix X based on a small $k \times n$ kernel matrix W and formulate the basic requirements of orthogonality and regularity. In expressing the requirements on X in terms of W we allow for the ratio between the lengths of the original and reduced signals to be $k > 1$—that means that faster reduction rates may be possible with wavelets with number of rows k larger than 2. In §4 we discuss the case $k = 2$ and we reconstruct the Daubechies finite support wavelets. In the following §5 we derive some basic properties important particularly for the case $k > 2$ and in §6 we use them to find explicitly the 3×5 and 3×6 wavelets. In §7 we discuss, for general k, the $k \times 2k$ wavelets for which we indicate the maximal achievable regularity and derive, after numerical exploration, explicit formulae for wavelets of fixed regularity. In §9 we show that wavelets with larger k are potentially superior to those with $k = 2$, by both theoretical considerations as well as by presenting numerical results demonstrating their performance. The last §10 contains conclusions.

Although we refer to the analytical exposition of the wavelet transforms the paper is essentially self-contained; it is hoped it demonstrates that wavelets can be treated by simple algebraic means.

2. Motivation of the algebraic approach

2.1. Matrix representation of the multiresolution analysis.

In multiresolution analysis (see, e.g. [2]), a continuous signal is represented by a doubly infinite sequence of coefficients f_{jk} which arise from projecting the signal into a sequence of imbedded subspaces $\mathcal{V}_j \subset \mathcal{V}_{j+1}$ with suitable bases. The discretization is immediate—for fixed j, the coefficients can be viewed as the signal values sampled at a certain rate. For every resolution level j, the sampling rate is different, usually double of that of level $j-1$. The continuous signal can be approximated by forming the series using the coefficients and the corresponding basis functions; this, however, is not relevant to our considerations here. Some information about the signal is lost when moving from $f_{j+1,k}$ to f_{jk}, that is from the finer discretization level $j+1$ to the coarser level j. This lost information is represented by the complementary sequence of *detail* coefficients d_{jk} which are in fact coefficients of the signal expansion in the orthogonal complement of $\mathcal{V}_j$ in $\mathcal{V}_{j+1}$. The basis functions in this orthogonal complement last subspaces are actually called wavelets.

Multiresolution analysis, that is the subspaces, their bases, etc., can be characterized in several ways. The one which is directly relevant to the discretization is through two discrete filter sequences $H = \{h_j\}$ and $G = \{g_j\}$ about which it is shown ([2]) that

$$f_{jk} = \sum_i h_{2k-i} f_{j+1,i} \tag{2.1}$$

$$d_{jk} = \sum_i g_{2k-i} f_{j+1,i}. \tag{2.2}$$

The filters H and G are also called *quadrature mirror filters* or *low-pass* and *high-pass* filters in different contexts. They are closely related (in fact one determines the other) and satisfy certain properties of which the following are of interest here.

- Complete recovery, i.e. invertibility of (2.1) and (2.2). In other words, the sequence $f_{j+1,k}$ can be obtained from the two sequences f_{jk} and d_{jk}.
- Orthogonality, which means that the recovery is achieved essentially by the same filters H and G.
- Regularity (of order p) which means that d_{jk} vanishes whenever $f_{j+1,k}$ is a polynomial (of degree less than p) in k.

This last property applies locally (in k) if the filters have finite support (i.e. only finite number of nonzero coefficients), or at least approximately locally, if the filters decay sufficiently fast and are truncated to finite support.

For practical implementation, only finite discrete signal sequences are available, which may be considered as a discretization at some resolution level J: $\boldsymbol{f}^J = (f_{J1}\ f_{J2}\ \dots\ f_{JN})^T$ and the processing involves the calculation of f_{jk} for several lower resolution levels $j = J-1, J-2, \dots$. To preserve, for the finite case, the above properties which hold for infinite sequences of coefficients, it suffices to extend the signal periodically (with or without symmetry depending on the symmetry of the filters). Nevertheless, from N (even) values $f_{j+1,k}$ we obtain only $\frac{1}{2}N$ values of both f_{jk} and d_{jk} due to the double shift in centering the filter convolution (called also *undersampling*).

With this periodicity and undersampling taken into account, the equations (2.1) and (2.2) describe a finite dimensional linear algebraic transform $\boldsymbol{F}^j = X\boldsymbol{f}^{j+1}$ where $\boldsymbol{F}^j$ contains both the signal and detail at the j-th level and X is a sparse square matrix the detailed description of which is given in §3 in terms of a kernel matrix formed by the coefficients h_i and g_i of the filters H and G.

2.2. Algebraic motivation of the wavelet transform.

A discretized signal, that is a vector $\boldsymbol{f}$ of a potentially large size N, may contain redundancies which could be represented in a more efficient way than through a large number of data values. The simplest redundancies are blocks of repeated values or, more generally, blocks of values representable as polynomials of low degrees. To identify such redundancies automatically, and thus, if required, pre-process the signal for compression, we seek a linear transform $\boldsymbol{g} = X\boldsymbol{f}$ with the following properties.

The nonsingular matrix X is blocked into two parts producing

$$\boldsymbol{g} = \begin{pmatrix} \boldsymbol{f}_r \\ \boldsymbol{f}_d \end{pmatrix} = \begin{pmatrix} X_r \\ X_d \end{pmatrix} \boldsymbol{f} = X\boldsymbol{f}$$

and is chosen in such a way that the *detail* $\boldsymbol{f}_d$ is small or *vanishes* on the redundacies, e.g. polynomials. The significant information contained in the original signal $\boldsymbol{f}$ is then concentrated in the *reduced* signal $\boldsymbol{f}_r$.

The original signal can be reconstructed by the inverse transformation $\boldsymbol{f} = X^{-1}\boldsymbol{g}$. This reconstruction may be only approximate for several reasons:

- round-off errors,
- using an approximate inverse and
- replacing $\boldsymbol{g}$, particularly the detail part $\boldsymbol{f}_d$, by its approximation (e.g. as a result of data compression or transmission errors).

The transformation may be repeated, with smaller matrices X and in a tree-like (pyramidal) fashion, on both parts of $\boldsymbol{g}$, i.e. on both the reduced signal $\boldsymbol{f}_r$ and the detail $\boldsymbol{f}_d$. The reconstruction then follows the appropriate reversed path.

The requirements on matrix X are as follows:

- The multiplication $X\boldsymbol{f}$ should not be expensive - X should be sparse, banded or block banded (possibly after permutation).
- The inverse should be simple to obtain and numerically stable - X could have orthogonal rows, $XX^T = D_X^2$, a diagonal matrix. We can then either use the orthogonal matrix $D_X^{-1}X$ or, simply, $X^{-1} = X^T D_X^{-2}$.
- The detail part of X should satisfy $X_d C = 0$ for some given constant matrix C of a set of simple signals, e.g. polynomials.

As an example, consider, for $N = 2m$ even, the transformation

$$X = P\,\mathrm{diag}(\underbrace{W_2, W_2, \ldots, W_2}_{m\times}), \quad W_2 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

where P is a permutation matrix comprising the odd numbered rows of an identity followed by the even numbered rows. Clearly, X is a multiple of an orthogonal matrix (the scaling chosen preserves the magnitude of the signal values) so that the inverse operation (recovery) is performed by $2X^T$. The multiplication $X\boldsymbol{f}$ requires $O(N)$ operations due to the very sparse structure of X and the detail vanishes whenever two adjacent values in $\boldsymbol{f}$ are equal. In what comes later, we call this W_2 a 2×2

orthogonal order 1 *kernel wavelet matrix* (or just *orthogonal wavelet*, for brevity). In this case, W_2 being square, there is no need for periodic extension of the signal.

3. Matrix formulation of the wavelet transform

In this section we show how to construct the transformation matrix X, introduced intuitively in the previous section, starting from a small building block, a kernel wavelet $k \times n$ matrix W which, for brevity, we may also call a wavelet. We formulate the requirements of regularity (polynomial order) and orthogonality of X in terms of this kernel wavelet matrix. We are going to consider general $k \geqslant 2$ which corresponds to undersampling (see section 2.1) by a factor k, or, in other words, while the first row of W is the low-pass filter H, *all* other $k-1$ rows of W play the role of the high-pass filters G.

In what follows, I and O will denote the identity and zero matrices, respectively, of appropriate sizes. We also use $\boldsymbol{o}$ for zero vector, $\boldsymbol{e}_j$ for the j-th column of the identity matrix and $\mathbf{1}$ to denote a vector of 1's of a suitable length. Euclidean norm of vectors will be denoted by $\|\cdot\|$.

We partition the $k \times n$ matrix W into

$$(3.1) \qquad \begin{pmatrix} W & O \end{pmatrix} = \begin{pmatrix} A_1 & A_2 & \ldots & A_q \end{pmatrix}$$

where $0 \leqslant kq - n < k$ and all A_j are square $k \times k$ matrices (A_q may have some zero columns when k does not divide n). When we express the convolution of W with the finite discrete signal $\boldsymbol{f}$ (that is the repeated inner products) in matrix notation, the wavelet matrix W must be shifted k positions to the right after each application (the undersampling). Furthermore, to take in account the periodical extension of $\boldsymbol{f}$, observe that we have for, say, $n = 2k$ and $N = 3k$ (length of $\boldsymbol{f}$)

$$\begin{pmatrix} A_1 & A_2 & O & O & O & O \\ O & A_1 & A_2 & O & O & O \\ O & O & A_1 & A_2 & O & O \end{pmatrix} \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{f} \end{pmatrix} = \begin{pmatrix} A_1 & A_2 & O \\ O & A_1 & A_2 \\ A_2 & O & A_1 \end{pmatrix} \boldsymbol{f}.$$

Not surprisingly, a nonsquare matrix of shifted wavelets is converted, due to the assumed periodicity of the signal, to a circulant matrix. This leads us to the following definition.

Definition 3.1. Given N divisible by k and W partitioned as in (3.1) we denote by $X_{\text{cir}}(W, N)$ the circulant $N \times N$ matrix

$$(3.2) \qquad X_{\text{cir}}(W,N) = \begin{pmatrix} A_1 & A_2 & \dots & A_q & O & \dots & O & O \\ O & A_1 & \dots & A_{q-1} & A_q & \dots & O & O \\ & & \dots & & & \dots & & \\ A_3 & A_4 & \dots & O & O & \dots & A_1 & A_2 \\ A_2 & A_3 & \dots & O & O & \dots & O & A_1 \end{pmatrix}.$$

All that remains now is to separate the reduced signal from the detail. As we shall see later, it is useful to separate the detail generated by different rows of W as well. This is expressed as follows.

Definition 3.2. Given again $N = mk$, let $P_{m,k}$ be the $N \times N$ permutation matrix selecting first the 1 mod k rows, then the 2 mod k rows, etc. Given, also, a $k \times n$ wavelet matrix W we define the $N \times N$ wavelet transform matrix by

$$X(W, N) = P_{m,k} X_{\text{cir}}(W, N)$$

where the circulant matrix $X_{\text{cir}}(W, N)$ is defined as in (3.2).

The purpose of the permutation matrix $P_{m,k}$ is, as indicated above, to *split* the reduced signal and the detail, as produced by the rows of the kernel wavelet W. The inverse operation

$$\tilde{g} = P_{m,k}^T \begin{pmatrix} \boldsymbol{f}_r \\ \boldsymbol{f}_d \end{pmatrix}$$

merges the separated reduced signal and details into an N-vector ordered for multiplication by the inverse of the circulant matrix above. We repeat that this splitting and merging corresponds to what is also called undersampling and oversampling.

We now wish to express the conditions imposed on the transform matrix X in §2.2 in terms of the generating matrix W.

Firstly, the sparseness is assured by definition and the multiplication $X(W, N)\boldsymbol{f}$ is of $O(Nn)$ complexity; even lower complexity may be achievable by exploiting special properties of the elements of W.

Secondly, the orthogonality depends on W having rows and also partial rows orthogonal to each other. This can be formulated as follows.

Theorem 3.1. *Let W be a $k \times n$ matrix partitioned as in* (3.1). *The transform matrix $X(W, N)$ has orthogonal rows if and only if the following conditions hold:*

$$(3.3) \qquad WW^T = D_W^2 = \operatorname{diag}(w_1^2, w_2^2, \dots, w_k^2), \quad w_j = \|W^T e_j\|, \quad j = 1, \dots, k$$

$$(3.4) \qquad \sum_{i=1}^{j} A_i A_{q+i-j}^T = O, \quad j = 1, 2, \dots, q-1.$$

Furthermore, $X(W, N)$ will be orthogonal if also $w_1 = \ldots = w_k = 1$, i.e. if W has normalized rows.

Finally, the concept of polynomial order, or regularity, is of critical importance to the concept of wavelets. We introduce the following definition to make it specific.

Definition 3.3. The $N \times N$ transform $X = (\, X_r^T X_d^T \,)^T$ is said to have polynomial order p if $X_d C_{p,N} = 0$ where the $N \times p$ matrix $C_{p,N} = (\, c_{jk} \,)$ has elements

$$c_{jk} = j^{k-1}, \quad j = 1, \ldots, N, \quad k = 1, \ldots, p.$$

We note that, as X is banded, the elements of $\boldsymbol{f}_d$ vanish even if only a part of the signal $\boldsymbol{f}$ (of length equal to the bandwidth) has polynomial character.

It is now obvious that the polynomial order requirement translates directly into a constraint on the wavelet matrix.

Theorem 3.2. *The transform matrix $X(W, N)$ has polynomial order p if the $k \times n$ kernel wavelet matrix W satisfies, for some vector $\boldsymbol{x}$,*

$$WC_{p,n} = \begin{pmatrix} \boldsymbol{x}^T \\ O \end{pmatrix} \tag{3.5}$$

where $C_{p,n}$ is defined as in Definition 3.3.

We will refer to (3.3) as the *main orthogonality* conditions, to (3.4) as the *shifted orthogonality* conditions and to (3.5) as the *regularity* or *order* conditions. A matrix W satisfying the three sets of conditions above will be referred to as an *orthogonal wavelet of order p* or a *p-order orthogonal wavelet* (*normalized*, if it also has normalized rows).

Our aim is to investigate the existence of orthogonal wavelets of various sizes and of a given order. We have an immediate simple result.

Theorem 3.3. *If $n = k(q-1) + 1$ then an orthogonal wavelet W of size $k \times n$ has, in fact, size $k \times n - 1$.*

Proof. For $n = 1 \bmod k$ the matrix A_q in (3.1) vanishes but for the first column, i.e. $A_q = \boldsymbol{a}\boldsymbol{e}_1^T$ where $\boldsymbol{a} = A_q \boldsymbol{e}_1 = W\boldsymbol{e}_n$ is the last column of the wavelet. The shifted orthogonality condition (3.4) for $j = 1$ is now $A_1 A_q^T = A_1 \boldsymbol{e}_1 \boldsymbol{a}^T = 0$ so that either the first or last column of W must vanish. That makes it a $k \times n - 1$ wavelet. □

More generally, we may compare the number of free parameters with the number of algebraic conditions required to hold for an orthogonal wavelet of order p. There are

- $\frac{1}{2}k(k-1)$ main orthogonality conditions,
- $(q-1)k^2$ shifted orthogonality conditions and
- $p(k-1)$ order conditions

making a total of $\frac{1}{2}k(2(kq+p)-k-1)-p$ conditions.

Considering that the rows have yet to be normalized, there are $(n-1)k$ free parameters to choose in a $k \times n$ wavelet W. In §5 we will show that, for $k > 2$, there are in fact $\frac{1}{2}(k-2)(k-1)$ fewer free parameters. Subtracting the number of conditions we find that there are $(k-1)(1-p)-k(kq-n)$ more parameters to choose than conditions specifying a p-order orthogonal wavelet. This appears to imply that, even for $n = kq$, the maximal possible order of orthogonal wavelets is $p = 1$. Wavelets of higher order exist, indicating that the orthogonality conditions are not independent. In the following sections we use an algebraic technique to unravel this dependency and to find some higher order wavelets.

4. $k = 2$: Daubechies wavelets

In this section we discuss $2 \times n$ orthogonal wavelets. By Theorem 3.3 there is no need to consider odd n so we assume that $n = 2q$ is even. The $W_{2,2q}$ wavelets are related to Daubechies orthogonal wavelets D_{2q} with compact support.

With arbitrary normalization we have $4q-2$ free parameters and $4q+p-3$ conditions for a p-order orthogonal $2 \times 2q$ wavelet

$$W_{2,2q} = \{w_{ij}\}_{i=1}^{2}\,{}_{j=1}^{2q}.$$

This appears to imply that only order $p = 1$ is achievable for any q (trivial for $q = 1$ with $W_{2,2} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$). However, Daubechies [1] shows that $p = q$ is possible.

We denote $W = (\, A_1 \quad A_2 \quad \dots \quad A_q \,)$ and $Z = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The shifted orthogonality condition (3.4) states that

$$\sum_{i=1}^{j} A_i A_{q+i-j}^T = O, \quad j = 1, \dots, q-1. \tag{4.1}$$

The D_{2q} wavelets are determined [1] by the signal (low-pass filter) coefficients w_{1j}. The detail (high-pass filter) coefficients satisfy, possibly up to a factor, $w_{2j} =$

$(-1)^j w_{1,n-j+1}$ which we may write as

$$e_2^T A_j = e_1^T A_{q-j+1} Z, \quad j = 1, \ldots, q. \tag{4.2}$$

This special relation cuts the number of free parameters in $W_{2,2q}$ by half, i.e. to $2q-1$ but has the most dramatic effect on the $4q-3$ nonlinear orthogonality relations—it satisfies $2q-1$ of them explicitly while making the remaining $2q-2$ pair-wise identical. Thus only $q-1$ nonlinear equations

$$\sum_{i=1}^{j} e_s^T A_i A_{q+i-j}^T e_s = 0, \quad j = 1, \ldots, q-1, \quad s = 1 \text{ or } 2 \tag{4.3}$$

remain. It is therefore obvious that we can subject the $2q-1$ free parameters to $p = q$ linear equations (3.5) (with $k = 2$) and to obtain the order q wavelet $W_{2,2q}$ by solving the $q-1$ nonlinear equations (4.3). For $q = 2$ this can be done explicitly to obtain the Daubechies D_4 wavelet. For larger q we used a simple *MATLAB* program to find, without great difficulty, orthogonal wavelets up to size $q = 8$, not necessarily the same as in [1], depending on the initial value supplied to the solver.

The choice (4.2) is ingenious and its motivation can be found in the analytical exposition of wavelets (e.g. [1], [2]). We observe that it is not clear that the orthogonality conditions imply this relation between the first and second rows of $W_{2,2q}$. We may formulate this explicitly.

Proposition 4.1. *If a wavelet $W_{2,2q} = (w_{ij})_{i=1}^{2}{}_{j=1}^{2q}$ satisfies the main and shifted othogonality conditions then it also satisfies $w_{2j} = \gamma(-1)^j w_{1,2q-j+1}$ for some constant γ.*

It is possible to prove this proposition by considering certain polynomials associated with the coefficients w_{ij}. Here we will show it true for $q \leqslant 3$ by a more algebraic argument. It remains, however, an open question whether such a proof can be easily extended to wavelets of order $q > 3$. The relation (4.2) between the free parameters gives little insight about what to do in case $k > 2$. The alternative approach to proving Proposition 4.1 gives us such an insight.

We start with (4.1) for $j = 1$, that is $A_1 A_q^T = 0$. If either of the matrices is nonsingular, the other must vanish; for nontrivial solution both must therefore have rank 1. With the normalization $A_1 e_1 = \mathbf{1} = (1 \quad 1)^T$ and noting that $\boldsymbol{x}^T Z \boldsymbol{x} = 0$ for any 2- vector $\boldsymbol{x}$ we conclude that

$$A_1 = \mathbf{1}\boldsymbol{w}^T, \quad A_q = \boldsymbol{v}\boldsymbol{w}^T Z \tag{4.4}$$

for some vectors $w = (1 \quad \alpha_1)^T$, $v = (\alpha_2 \quad \alpha_3)^T$. So we have lost 3 out of 6 parameters in matrices A_1, A_q but satisfied 4 equations. Moreover, we have structural information about these matrices.

For $q = 2$ the main orthogonality condition (3.3) implies $1 + \alpha_2\alpha_3 = 0$ from which, and from (4.4), the Proposition 4.1 follows. A simple calculation shows that we get the same result as D_2 by solving the two regularity equations for the wavelet of order 2.

For $q = 3$, a similar rank argument using (4.1) for $j = 2$ leads to $A_2 = \alpha_4(A_1 + A_3)Z$ and again, $1 + \alpha_2\alpha_3 = 0$ from (3.3), and the Proposition 4.1 follows. The three order conditions can again be solved.

For $q > 3$ the matrix argument using the shifted orthogonality condition to obtain structural information on matrices A_j, $j = 2, \ldots, q - 1$ becomes complicated. We observe that, contrary to the choice (4.2) which resolves the main orthogonality conditions and *some* shifted orthogonality conditions, we are starting here with *all* the shifted orthogonality conditions. The advantage of the rank argument is that it extends easily to the $k > 2$ wavelets.

There is one more property of Daubechies wavelets left to mention. If the order of a $2 \times n$ orthogonal wavelet W is at least 1 then the sum of the signal coefficients $\sigma = e_1^T W\mathbf{1}$ is related to its normalization $w_1 = \|e_1^T W\| = \sigma/\sqrt{2}$. This is important in signal processing as the reduced signal of values, say, in interval $[0, 1]$ will have values in $[0, \sigma]$. We prove this relation, by algebraic means, for the general ($k \geqslant 2$) orthogonal wavelets in the next section.

5. Some properties of generalized wavelets

In this section we discuss properties of $k \times n$ wavelets with $k \geqslant 2$. We first derive the relation between the sum $\sigma = e_1^T W\mathbf{1}$ and the normalization of the wavelet mentioned in the previous section.

Theorem 5.1. *If W is a $k \times n$ orthogonal wavelet of order at least 1 then*

$$\sigma = e_1^T W\mathbf{1} = \sqrt{k}\|e_1^T W\|. \tag{5.1}$$

Proof. We apply Theorem 3.1 with $N = kq$, $0 \leqslant kq - n < k$ and note that the circulant matrix

$$X_{\text{cir}} = X_{\text{cir}}(W, kq) = \begin{pmatrix} A_1 & A_2 & A_3 & \ldots & A_{q-1} & A_q \\ A_2 & A_3 & A_4 & \ldots & A_q & A_1 \\ & & & \ldots & & \\ A_q & A_1 & A_2 & \ldots & A_{q-1} & A_{q-1} \end{pmatrix}$$

has orthogonal rows. Denoting, as before, $WW^T = D_W^2$, D_W the diagonal matrix of norms of rows of W, we have $X_{\text{cir}}X_{\text{cir}}^T = D^2$, D a diagonal matrix consisting of D_W repeated q times. As $W\mathbf{1} = \sigma e_1$ by the order assumption, we have

$$kq = \mathbf{1}^T\mathbf{1} = \mathbf{1}^T X_{\text{cir}}^T D^{-2} X_{\text{cir}}\mathbf{1} = q\mathbf{1}^T W^T D_W^{-2} W\mathbf{1} = q\sigma^2 e_1^T D_W^{-2} e_1$$

from which the result follows. □

We now present two more properties typical for $k > 2$. It is useful to partition W into the reduced signal w_r and detail W_d parts (i.e. $W = (\, w_r \quad W_d^T\,)^T$). First we observe the following:

Lemma 5.2. *If W is a $k \times n$ normalized orthogonal wavelet of order p and if Q is an orthogonal $k \times k$ matrix such that $Qe_1 = e_1$ then QW is also a $k \times n$ normalized orthogonal wavelet of order p.*

The proof is obvious but it is worth noting that the normalization is essential here. We have therefore a range of equivalent orthogonal wavelets of the same order, nontrivial for $k > 2$ (for $k = 2$ the only possible Q is the identity). The equivalence of these wavelets means that they all produce the same reduced signal and that the homogeneous properties of W_d (like order p) are preserved.

Theorem 5.3. *For any normalized orthogonal wavelet W of order p there exists an orthogonal $k \times k$ matrix Q such that $Qe_1 = e_1$, and the equivalent wavelet $W_H = QW$ is upper Hessenberg, i.e. $e_i^T W_H e_j = 0$ whenever $i > j + 1$.*

P r o o f. Use the QR decomposition of W_d (i.e. W without the first row) and Lemma 5.2. □

This theorem implies that to investigate the existence of orthogonal wavelets of certain size $k \times n$ and order p it is sufficient to consider those with $\frac{1}{2}(k-2)(k-1)$ zeros in the bottom left triangular part of W. This, in turn, considering that the rows are yet to be normalized, means that we have $nk - 1 - \frac{1}{2}k(k-1)$ free parameters to choose in a $k \times n$ wavelet W as envisaged in Section 4.

We introduce a more detailed concept of polynomial regularity or order—for $k > 2$ some rows of W may be more regular than others:

Definition 5.1. Let $\boldsymbol{p} = (p_2, p_3, \ldots, p_k)$ be a vector of integers. We shall say that W is of order $\boldsymbol{p}$ (or a $\boldsymbol{p}$-order wavelet) if

$$e_j^T W C_{p_j,n} = 0^T \quad j = 2, 3, \ldots, k. \tag{5.2}$$

As the order of all rows in W except the first is not significant, we may assume that the elements of $\boldsymbol{p}$ are ordered. We now have a result concerning the maximal $\boldsymbol{p}$ for a $\boldsymbol{p}$-order generalized wavelets, i.e. for those with $k > 2$.

Theorem 5.4. *For any normalized orthogonal wavelet W of order p there exists an orthogonal $k \times k$ matrix Q such that $Qe_1 = e_1$ and the equivalent wavelet $W_{\max} = QW$ is a $\tilde{\boldsymbol{p}}$-order wavelet where $\tilde{\boldsymbol{p}} = (p, p+1, \ldots, p+k-2)$.*

Proof. We choose the nontrivial part of Q from the QR decomposition of the last $k-2$ columns of $W_d C_{p+k-2,n}$ (the first p columns vanish by the order assumption). ☐

This theorem asserts that we can modify any p-order orthogonal wavelet to increase the polynomial order of some parts of the detail.

As the first step in the discussion of the existence of orthogonal wavelets of various sizes we conclude this section with the special case $n = k$. Here $X_{\mathrm{cir}}(W_{k,k}, N)$ is block diagonal (no overlap in (3.2)).

Theorem 5.5. *For any $k > 1$ there exists an orthogonal $k \times k$ $\tilde{\boldsymbol{p}}$-order wavelet $W_{\max}$, where $\tilde{\boldsymbol{p}} = (1, 2, \ldots, k-1)$.*

There are no p-order orthogonal $k \times k$ wavelets for $p > 1$.

Proof. As $q = 1$ in (3.1) there are no shift orthogonality conditions on $W_{k,k}$. Using the upper Hessenberg form of Theorem 5.3 we obtain the $k \times k$ orthogonal wavelet,

$$W_H = \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 & 1 \\ 1-k & 1 & \ldots & 1 & 1 & 1 \\ 0 & 2-k & \ldots & 1 & 1 & 1 \\ & & & \ldots & & \\ 0 & 0 & \ldots & 0 & -1 & 1 \end{pmatrix}$$

which obviously has order 1, and this order can not be improved. We obtain $W_{\max}$ from the normalized W_H by Theorem 5.4. ☐

Matrix $W_{\max}$ can be also constructed directly. It appears that in non-normalized form it can have integer elements. For example, the last row contains binomial coefficients with alternating signs.

6. Special cases of $k > 2$ wavelets

In this section we discuss two particular cases of generalized wavelets. For $k = 3$ the smallest interesting case has $n = 5$.

6.1. The W_{35} wavelet.

We will partition the wavelet as

$$W_{35} = (\, A \quad \boldsymbol{c} \quad B \,)$$

where A and B are 3×2 matrices and we may choose, for normalization (taking into account Theorem 5.3 with some modification), $\boldsymbol{c} = (\, 1 \quad 1 \quad 0 \,)^T$.

The shifted orthogonality condition $AB^T = O$ implies that any nontrivial A and B must both have rank one and the form

$$A = \boldsymbol{a}\boldsymbol{x}^T, \quad B = \boldsymbol{b}\boldsymbol{x}^T Z, \quad Z = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

where, to avoid duplicating free parameters, we denote $\boldsymbol{a} = (\, 1 \quad \alpha \quad \beta \,)^T$, $\boldsymbol{b} = (\, \gamma \quad \delta \quad 1 \,)^T$ and $\boldsymbol{x} = (\, \lambda \quad \mu \,)^T$. The wavelet thus has the form

$$W_{35} = \begin{pmatrix} \lambda & \mu & 1 & \gamma\mu & -\gamma\lambda \\ \alpha\lambda & \alpha\mu & 1 & \delta\mu & -\delta\lambda \\ \beta\lambda & \beta\mu & 0 & \mu & -\lambda \end{pmatrix}$$

which satisfies the shifted orthogonality conditions explicitly. We have six free parameters and three main othogonality conditions so that order 2 regularity, which requires 4 additional conditions, is not likely to be possible.

We will further describe only the results of the calculations which were performed using the symbolic calculation package MAPLE.

We use the two order 1 linear relations to eliminate λ and μ. The two main orthogonality conditions involving the third row of W_{35} have simpler form than the third one and give immediately $\beta = -\gamma$ and $\delta = \alpha\gamma$. Substituting all this into the third main orthogonality condition we obtain $\alpha = -\frac{1}{2}$. All these conclusions are necessary for nontrivial solutions. We thus obtain a one parameter family of 3×5 order 1 wavelets in the form

$$W_{35} = \begin{pmatrix} 1-\gamma & 1+\gamma & 1+\gamma^2 & (1+\gamma)\gamma & (1-\gamma)\gamma \\ \gamma-1 & -\gamma-1 & 2(1+\gamma^2) & -(1+\gamma)\gamma & (1-\gamma)\gamma \\ (1-\gamma)\gamma & (1+\gamma)\gamma & 0 & -1-\gamma & 1-\gamma \end{pmatrix}. \tag{6.1}$$

Applying Theorem 5.4 we can now construct a family of 3×5 wavelets of order $(\, 1 \quad 2 \,)$. The expressions for the second and third rows of these wavelets involve polynomials in γ of degree up to 5 with integer coefficients. As the third row is already of order 2 we may now try to choose γ to satisfy the order 2 condition for the second row - that would lead to an order 2 wavelet W_{35}. However, further MAPLE calculation shows that this condition simplifies to

$$13\gamma^4 - 48\gamma^3 + 134\gamma^2 - 48\gamma + 13 = 0$$

which has only complex roots. We have therefore established the following result.

Theorem 6.1. *There exist a family of* 3×5 *wavelets of order* 1 *the Hessenberg form of which is given in* (6.1) *but no real* 3×5 *wavelet of order* 2.

We note that for $\gamma = 1$ the W_{35} reduces, but for a multiplicative factor, to the 3×3 wavelet of Theorem 5.5 with $k = 3$.

6.2. The W_{36} wavelet.

Our first attempt to find a 3×6 wavelet of order 2, without the benefits of Theorems 5.3 and 5.4 and without using the rank argument, was to find the 18 unknowns from the 4 linear order equations and 15 nonlinear orthonormalizing equations by least square minimization. This approach failed, as did the first attempts using symbolic calculation on the non-Hessenberg form of the wavelet. However, the following approach leads to the solution.

We will partition the wavelet as

$$W_{36} = \begin{pmatrix} A & B \end{pmatrix}$$

where A and B are 3×3 matrices.

The shifted orthogonality condition $AB^T = O$ implies that nontrivial A and B must have rank at most two and one, respectively. We choose A to be in the Hessenberg form (Theorem 5.3) and therefore B to have rank one. Choosing the normalization to simplify the wavelet we obtain the form

$$W_{36} = \begin{pmatrix} 1 & \alpha - \beta & \gamma - \delta & \sigma\lambda & -\sigma\delta & \sigma\beta \\ 1 & \alpha & \gamma & \tau\lambda & -\tau\delta & \tau\beta \\ 0 & \beta & \delta & \varrho\lambda & -\varrho\delta & \varrho\beta \end{pmatrix}$$

which satisfies the shifted orthogonality conditions explicitly for $\lambda = \alpha\delta - \beta\gamma$. We have seven free parameters and three main othogonality conditions so that order 2 regularity, which requires four additional conditions, is likely to be possible.

Again, we only describe the results of the MAPLE calculations leading to the solution. We use the two order 2 linear relations to eliminate τ and ϱ obtaining

$$\tau = \frac{3+2\alpha+\gamma}{2\beta-\delta}, \quad \varrho = \frac{2\beta+\delta}{2\beta-\delta}.$$

A nontrivial combination of the two order 1 linear relations, which eliminates λ, yields

$$\beta = \frac{\alpha+2}{\gamma-1}\delta.$$

The two main orthogonality conditions involving the first row of W_{36} give, after a not so short calculation,

$$\sigma = \frac{5+2\alpha-\gamma}{(3+2\alpha+\gamma)(1-2\alpha+\gamma)}, \quad \delta = \frac{(1-2\alpha+\gamma)(1-\gamma)}{5+2\alpha-\gamma}.$$

We still have one main orthogonality condition and either of the identical order 1 conditions, both rational in α and γ, with numerators of degree 4 and 2, respectively. To our surprise and delight, MAPLE produced an explicit solution of the form

$$\alpha = \frac{293-135\gamma-130\gamma^2}{200\gamma-214}, \quad 950\gamma^2-475\gamma-571=0$$

which yields two real solutions $\gamma_{1,2} = 1/4 \pm 41\sqrt{57}/380$.

The Hessenberg form of the 3×6 wavelet can thus be obtained by back substitution in the above formulae. We have used MAPLE to evaluate it to 17 digits and MATLAB to produce the normalized, maximal order (2 3) form, presented in the tables below (the above formulae can, of course, be implemented in MATLAB independently of MAPLE). We observe that the first rows of the two wavelets differ only in orientation while the others appear significantly different. The orthogonality and order conditions have been numerically verified.

W_{36}^T wavelet of order (2 3), version with γ_1		
first row	second row (order 2)	third row (order 3)
0.33838609728386	−0.18891776559956	0.37551320045549
0.53083618701374	0.02941808858961	−0.83095884420054
0.72328627674361	−0.29996861472639	0.35003638840610
0.23896417190576	0.78683311028654	0.18051189543004
0.04651408217589	0.15315609724545	0.03513641844631
−0.14593600755399	−0.48052091579565	−0.11023905853741

W_{36}^T wavelet of order (2 3), version with γ_1		
first row	second row (order 2)	third row (order 3)
−0.14593600755399	0.57180800871317	−0.28767215433970
0.04651408217589	−0.33573028308050	0.76068600730791
0.23896417190576	−0.69554601736901	−0.57842310830715
0.72328627674361	0.20868152180886	0.04787482447101
0.53083618701374	0.15315609724545	0.03513641844632
0.33838609728386	0.09763067268203	0.02239801242162

7. $W_{k,2k}$ WAVELETS: DISCOVERY BY COMPUTATION

In this section we discuss the $k \times 2k$ wavelets. We first investigate what polynomial order we may expect to achieve by comparing the number of conditions with the number of free parameters. Next we choose one form of the $W_{k,2k}$ wavelet and show how to obtain a $k-1$-parametric range of such wavelets of order 1. We then seek to choose these parameters to achieve polynomial order 2 and derive explicit formulae for such wavelets. The way towards these formulae was, however, through extensive computations, both numerical and symbolic, and we consider it of interest to indicate how the results have been achieved.

7.1. Maximal order considerations.

There are $2k^2$ elements in a $k \times 2k$ matrix but we know that such a parameter count is misleading. We have to consider the implications of the shifted orthogonality conditions and the multitude of orthogonal wavelets of given order expressed in Lemma 5.2.

We partition the wavelet

$$W = (A \quad B)$$

where A and B are $k \times k$ matrices. The shifted orthogonality condition $AB^T = O$ implies that $rank(A) + rank(B) \leqslant k$ and we may choose equality for the maximal number of free parameters. Let j be the number of linearly independent rows of A and assume that they are rows $2, 3, \ldots, j+1$. We further assume that it is the other rows (i.e. first row and last $k-j-1$ rows) of B which are linearly independent. Taking into account Theorem 5.3 we can assume A to be in Hessenberg form; this implies that the last $k-j-1$ rows of A vanish. This means that the last $k-j+1$ rows of B may be assumed to be in upper triangular form by application of Lemma 5.2. Finally, we normalize all rows chosen as linearly independent in both A and B by choosing the first (generically nonzero) element to be one. As an illustrative example, for $k=6$, $j=2$ the wavelet matrix has the form

$$(A \quad B) = \begin{pmatrix} d & d & d & d & d & d & 1 & x & x & x & x & x \\ 1 & x & x & x & x & x & d & d & d & d & d & d \\ 0 & 1 & x & x & x & x & d & d & d & d & d & d \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & x & x & x \end{pmatrix}.$$

where d denotes elements of the dependent rows and x of the independent ones.

As each dependent row of B is determined by $k - j$ coefficients, and the only dependent row of A by j coefficients, it is easy to count the number of free parameters to be

$$N_{FP} = \frac{1}{2}(k-1)(k+2) + 2j(k-j).$$

The number of main orthogonality and polynomial order conditions is as discussed in §3; for the shifted orthogonality conditions to be satisfied we only need the linearly independent rows of A and B to be pair-wise orthogonal. This gives

$$N_{cond} = (k-1)(q + \frac{1}{2}k) + j(k-j)$$

for polynomial order q. Subtracting we obtain that maximal achievable order should satisfy

$$q(k-1) \leqslant j(k-j) + k - 1. \tag{7.1}$$

This bound is obviously maximal if j is $\frac{1}{2}k$ (possibly rounded to an integer value). The following table indicates what maximal order, and number of free parameters, may be possible for the $k \times 2k$ wavelet matrix.

k	3	4	5	6	7	8	9	10	11	12	13
q	2	2	2	2	3	3	3	3	4	4	4
free param.	0	1	2	4	0	2	4	7	0	3	6

We leave as an open question the existence of higher order wavelets and concentrate, in the rest of this section, on the choice $j = k - 1$ where A is rank $k - 1$ upper unit Hessenberg and B is of rank one. We note that in this case (7.1) simplifies to $q \leqslant 2$ so that the best expected order, for all values of k is 2, with no further free parameters.

7.2. $W_{k,2k}$ wavelet of polynomial order 1.

We now discuss the wavelet

$$W = (A \quad B)$$

where A is a unit Hessenberg matrix of rank $k-1$ (the first row is a combination of the others) and $B = \boldsymbol{b}\boldsymbol{a}^T$ is of rank 1. For normalization we choose the first element of $\boldsymbol{b}$ and the last element of $\boldsymbol{a}$ to equal 1—this is not significantly different from the previous section. For later convenience we index the elements of A, $\boldsymbol{a}$ and $\boldsymbol{b}$ in reverse order, i.e. $\boldsymbol{e}_i^T A \boldsymbol{e}_j = r_{k-i+1,k-j+1}$, $\boldsymbol{a} = (a_k \ \dots \ a_2 \ a_1)^T$ and $\boldsymbol{b} = (b_k \ \dots \ b_2 \ b_1)^T$.

As each extra order requires $k-1$ constraints, we expect, from the maximal order considerations, to be able to choose $k-1$ parameters and yet obtain an order 1 orthogonal wavelet. Either vector $\boldsymbol{a}$ or $\boldsymbol{b}$ is thus a good candidate. We observe that the row $k-j+1$ of W is subjected to one polynomial order condition, one shifted orthogonality condition (which here reduces to $A\boldsymbol{a} = 0$) and, assuming the last rows are already known, a further $j-1$ main orthogonality conditions. Moreover, these conditions are linear equations in the $j+1$ unknown elements b_j and r_{ji}, $i = 1, \dots, j$, assuming the vector $\boldsymbol{a}$ is given. This applies to all rows but the first one ($j = k$), where there is no polynomial order condition, but we have $b_k = 1$. The following result is therefore indicated.

Proposition 7.1. *Given the normalized vector $\boldsymbol{a}$ there exists a unique $k \times 2k$ orthogonal wavelet W of order 1, in unit Hessenberg form. This wavelet can be constructed explicitly.*

We have implemented the above construction and used it extensively in computations described in §7.3. We did not encounter a situation where any of the k linear systems were singular but we are curious to know the precise conditions for their nonsingulsrity. Denoting the sum of elements in $\boldsymbol{a}$ by $\sigma = \mathbf{1}^T\boldsymbol{a}$, the solution of the first 2×2 system is (for general a_1)

$$r_{11} = \frac{-a_2}{a_1}, \quad b_1 = \frac{a_2 - a_1}{a_1\sigma} \tag{7.2}$$

as long as σ and a_1 do not vanish (this justifies our normalization $a_1 = 1$).

At first sight, as the elements of a larger linear system depend on the solutions of the previous smaller systems, we could not see how to make a general statement about the nonsingularity of all these systems. However, a few hours of using a symbolic computer program (such as MAPLE) reveals a simple pattern for the solution and $\sigma \neq 0$ is indeed sufficient for its existence.

Theorem 7.2. *For every vector $\boldsymbol{a}$ such that $\boldsymbol{e}_k^T\boldsymbol{a} = 1$ and $\mathbf{1}^T\boldsymbol{a} \neq 0$ there exists a unique $k \times 2k$ orthogonal wavelet $W = (\,A \quad \boldsymbol{b}\boldsymbol{a}^T\,)$ of order 1, with vector $\boldsymbol{b}$ and a* *unit Hessenberg A given by*

$$
\begin{aligned}
&b_j = \sigma_k(a_{j+1}\sigma_j - \varrho_j)/(\varrho_k d_j), \quad j = 1,2,\ldots,k-1, \quad b_k = 1,\\
&r_{ji} = (\sigma_j(a_i + a_{j+1}) - \varrho_j - (j + \sigma_k^2/\varrho_k)a_i a_{j+1})/d_j, \quad i = 1,2,\ldots,j, \quad j = 1,2,\ldots,k-1,\\
&r_{ki} = \varrho_k/\sigma_k - a_i, \quad i = 1,2,\ldots,k
\end{aligned}
$$

where

$$
\begin{aligned}
d_j &= (j + \sigma_k^2/\varrho_k)\varrho_j - \sigma_j^2, \quad j = 1,2,\ldots,k-1,\\
\sigma_j &= a_1 + a_2 + \ldots a_j,\\
\varrho_j &= a_1^2 + a_2^2 + \ldots a_j^2, \quad j = 1,2,\ldots,k.
\end{aligned}
$$

Proof. The proof of this Theorem (which also proves the previous Proposition) is by direct substitution. The d_j's, which are proportional to the determinants of the linear systems mentioned above, are positive because

$$
j\varrho_j - \sigma_j^2 = \sum_{1\leqslant s<t\leqslant j} (a_s - a_t)^2 > 0.
$$

□

7.3. $W_{k,2k}$ wavelet of polynomial order 2.

We have shown, for every $k > 1$, the existence of a $(k-1)$-parametric range of order 1 orthogonal $k \times 2k$ wavelets. The free parameters can be chosen to achieve further properties of the wavelet.

In particular, we may require the last $k-1$ rows of the wavelet to satisfy the order 2 conditions. We thus have $k-1$ nonlinear equations for the $k-1$ unknowns in the vector $\boldsymbol{a}$, the last element of which is 1. We used the standard MATLAB solver *fsolve* to find such order 2 wavelets for values $k < 9$. In fact, we performed an extensive search using random starting values to obtain some empirical evidence on the number of distinct solutions to this set of equations for $k > 3$; we know, from §3 and §6, that for $k = 2, 3$ we have exactly two distinct solutions. However, for all $k > 3$ we also obtained exactly two distinct solutions but the larger was k, the fewer starting values lead to convergence. Morever, the final values of $\boldsymbol{a}$ form, for different k, a triangle of interleaving numbers (for each branch of solutions) as listed in the following table.

Vectors $\boldsymbol{a}$ determining $k \times 2k$ wavelets of order 2					
First solutions, sizes $k = 2, 3, 4, 5, 6, 7$					
1.7320508	2.1374586	2.3899749	2.5615528	2.6855212	2.7792110
1	1.5687293	1.9266499	2.1711646	2.3484170	2.4826758
	1	1.4633250	1.7807764	2.0113127	2.1861407
		1	1.3903882	1.6742085	1.8896055
			1	1.3371042	1.5930703
				1	1.2965352
					1
Second solutions, sizes $k = 2, 3, 4, 5, 6, 7$					
−1.7320508	−1.6374586	−1.5899749	−1.5615528	−1.5426641	−1.5292110
1	−0.3187293	−0.7266499	−0.9211646	−1.0341313	−1.1076758
	1	0.1366750	−0.2807764	−0.52559845	−0.6861407
		1	0.3596118	−0.0170656	−0.2646055
			1	0.4914672	0.1569297
				1	0.5784648
					1

One can extrapolate from these tables to find the starting values for the nonlinear solver when seeking the solution for the next value of k. For example, fitting a two-dimensional, second degree polynomial to the first 5 columns of the "First solutions" above gave the last column within about 0.05 which was a sufficiently good starting value. By weighting the known 1 in the next column we brought the prediction of the solution to such efficiency that only about 20 function evaluations were needed to calculate the next larger size wavelet to full accuracy of the order 2 requirement.

This looked like the end of the story until we accidentally calculated the case $k = 11$. The result was $\boldsymbol{a} = (3\ 2.8\ 2.6\ \ldots\ 1.2\ 1)^T$! Has the reader noticed that the entries in each column in the above tables form an arithmetic progression? With this hindsight, and accepting the hypothesis, one can easily derive the following result by requiring the polynomial order 2 for just the last row of W and employing the formulae (7.2).

Theorem 7.3. *The elements of vector $\boldsymbol{a}$ determining the orthogonal wavelet $W = (\,A \quad \boldsymbol{b}\boldsymbol{a}^T\,)$ of order 2 are given by*

$$a_j = 1 + (j-1)x, \quad j = 1, 2, \ldots, k$$

where x is a root of the quadratic equation

$$(k^2 - 1)x^2 + 6x - 6 = 0$$

This gives, for any k, an explicit construction of two orthogonal wavelets of size $k \times 2k$, the $k-1$ detail rows of which have polynomial orders $2, 3, \ldots, k$. Whether one can do better, as indicated in §7.1 to be possible for $k > 6$, remains an open question.

8. An example of an application

When using discrete wavelet transforms, e.g. in pre-processing for data compression, we apply them repeatedly in a pyramidal fashion. Thus, say for $k = 3$, after first application to a signal $\boldsymbol{f}$ of length N, we obtain three vectors $\boldsymbol{f}_r$, $\boldsymbol{f}_d$, $\boldsymbol{f}_d$, each of length $\frac{1}{3}N$. In the second stage, assumming N is divisible by 9, we apply the transform to each of these vectors and obtain nine vectors $\boldsymbol{f}_{rr}$, $\boldsymbol{f}_{rd}$, $\boldsymbol{f}_{rd}$, $\boldsymbol{f}_{dr}$, $\boldsymbol{f}_{dd}$, $\boldsymbol{f}_{dd}$, $\boldsymbol{f}_{dr}$, $\boldsymbol{f}_{dd}$, $\boldsymbol{f}_{dd}$ where the subscripts indicate the reduced and detail parts of the signals.

In Figure 1 we demonstrate the two stages of applying the maximal order $W_{4,8}$ to the discrete signal (top graph marked 8) which happens to be a cubic spline. Each graph, which only appears to be a continuous line, represents 224 (as indicated) discrete values, equidistantly spaced. The first quarter (indexed 0–55) of the next graph (marked 6) is the reduced signal after the first application of $W_{4,8}$ (same shape as the original signal but undersampled by factor 4), the rest of that graph are the three details which, however, are too small to show in the same scale as the reduced signal. The next graph (marked 4) shows therefore the same information as the second graph, with the reduced signal zeroed and the three details (indexed 56–111, 112–167 and 168–223) magnified. We note both the different degree and magnitude reductions by the three high-pass filters of the maximal order 4×8 wavelet. The details are piece-wise linear, piece-wise constant and piece-wise zero, respectively. The last two graphs display the result of the second application of the wavelet (we use the full pyramid, i.e. we apply the transform to the details as well as to the reduced signal). After two applications we have one reduced signal and 15 details, all of length 14.

9. Comparisons of wavelets of different sizes

In comparing the effectiveness of different wavelets we take into account three things:

(a) The cost of applying the transform to a given signal.
(b) The size of the reduced signal.
(c) The polynomial order of the wavelet.

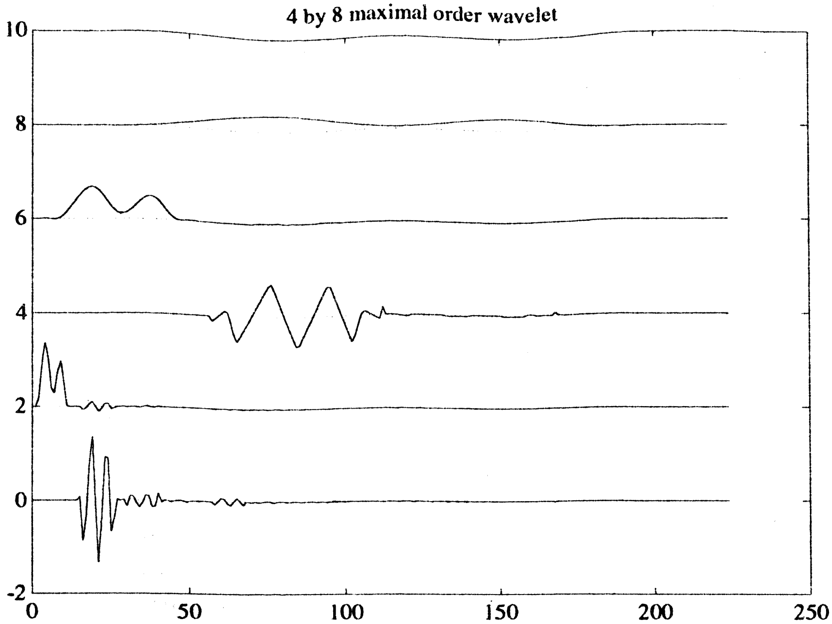

Figure 1. Stages 1 and 2 of decomposition.

For a $k \times n$ wavelet the cost of j stages applied to a signal of length N is proportional to Nnj and the reduced signal has length Nk^{-j}. Thus j stages with a $k \times n$ wavelet will cost the same and produce a reduced signal of the same length as j' stages with an $sk \times tn$ wavelet if

$$j = j't, \quad s = k^{t-1}. \tag{9.1}$$

To assess the overall effectiveness of applying wavelets of certain order is slightly more complicated. In the $k = 3$ example above, had the wavelet been of order p, the detail f_{dd} would have had polynomials of order less than $2p$ annihilated. If, however, the p-order wavelet was the maximal order wavelet constructed according to Theorem 5.4 then different "d"'s would represent a reduction of polynomial order by either p or $p + 1$.

We use, as a measure of effectiveness, the *average* cumulative order applied to the detail of the signal after several stages. For example, two stages with a $k = 2$ wavelet of order p have an average effect of $4p/3$ (two quarters with p and one with $2p$). Generally, j stages with an order p wavelet (maximized for $k > 2$ as indicated above) have an average order effect

$$\nu(k,j,p) = j(p + \frac{1}{2}k - 1)(k-1)\frac{k^{j-1}}{k^j - 1}.$$

We now want to compare the performance of $k \times 2k$ wavelets and Daubechies wavelets $W_{2,2p}$. For a fair comparison, we must consider such sizes k and p and such numbers of stages of decomposition (say j and j', respectively), so that the same total work leads to the same size of the reduced signals by both wavelets.

Using (9.1) we find that we should compare j stages of decomposition by a $k \times 2k$ wavelet with $j' = j \log_2 k$ applications of $W_{2,2p}$ with $p = k/\log_2 k$. This is possible practically (to get integral j' and p) only for very few $k = 2^{2^e} = 4, 16, 256, \ldots$ when $j' = 2^e j = 2j, 4j, 8j, \ldots$ and $p = 2^{2^e - e} = 2, 4, 32, \ldots$. However, comparing the polynomial average orders gives

$$\frac{\nu(k,j,2)}{\nu(2,j',p)} = \frac{(k+2)(k-1)}{k^2} = 1 + \frac{1}{k} - \frac{2}{k^2}$$

which reaches a maximum value of $9/8 = 1.125$ for $k = 4$. We conclude, that from the point of view of applying the highest polynomial order high-pass filter to the data for the same cost and signal reduction, j applications of $W_{4,8}$ are 12.5% better than $2j$ applications of the standard $W_{2,4}$.

To demonstrate this we show in Figure 2 the results of applying $W_{2,4}$ four times to the same data as in Figure 1. For easy comparison only the results of the second and fourth stages are shown. As explained above, two stages of the pyramid using $W_{2,4}$ require the same work and produce the same reduction as one stage with $W_{4,8}$. The reduced signals (graphs marked 6 and 2 in Figures 1 and 2) look practically identical. The most noticeable difference is in the third quarter of the graph marked 4. In Figure 2, two applications of $W_{2,4}$ produced only a piece-wise linear detail (type $\boldsymbol{f}_{dr}$ above) while, in Figure 1, the order 3 second high-pass filter of $W_{4,8}$ produced a piece-wise constant detail.

From the observations above, it does not appear profitable to use the $W_{k,2k}$ wavelets of order 2 for $k > 4$ as their advantage against the corresponding $W_{2,2p}$ would diminish with increasing k. The picture changes, however, should we be able to find higher order $W_{k,2k}$ as indicated in §7.1. Thus, for example, a $W_{16,32}$ of order 5 would be

$$\frac{\nu(16,j,5)}{\nu(2,4j,4)} = \frac{195}{128} \doteq 1.52$$

times better than $W_{2,8}$!

A further advantage of the $k > 2$ wavelets may be that they allow a greater flexibility in accomodating signals of length N divisible by factors other than powers of two.

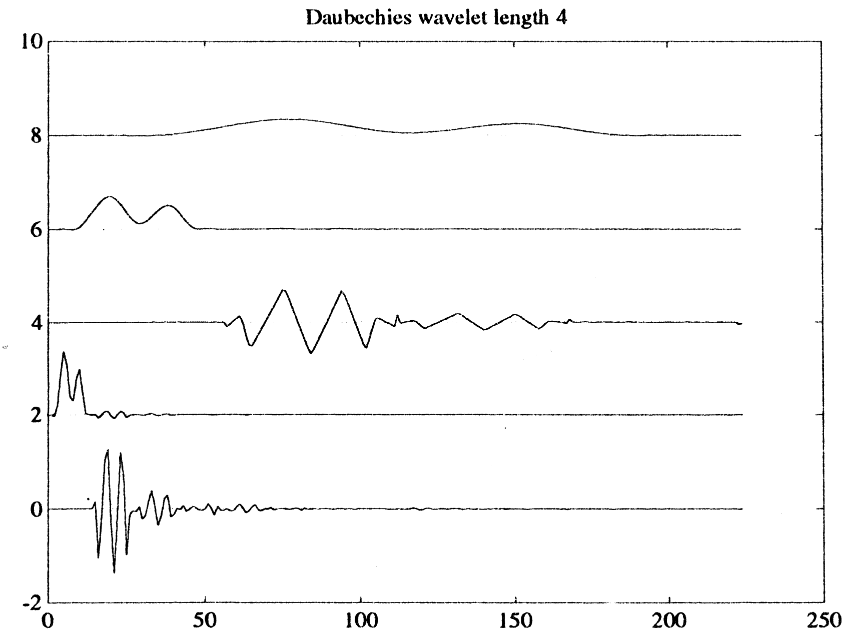

Figure 2. Stages 2 and 4 of decomposition.

10. Concluding remarks

We have presented an algebraic exposition of wavelets or, to be more precise, of linear transformations by banded matrices built up from small matrices with special properties, here also called wavelets. These transformations have applications in signal and image processing, for example for compression and reconstruction of digitized signals.

We have shown that by this algebraic approach we can derive the same wavelets as those known as "Daubechies finite support orthogonal wavelets" directly from their algebraic properties (orthogonality and polynomial order) and derived examples of similar, but more general, wavelets allowing undersampling (i.e. number of rows in our terminology) by factors larger than 2. We have suggested a way to compare the efficiency of transforms using wavelets of different sizes and shown that some of the generalized wavelets may perform better than those with $k = 2$. How much better will also depend on the application. What we find appealing in the algebraic approach is the explicit nature of the representation of wavelets and their properties—this also allows us to consider ranges of wavelets customized to particular purposes.

We have concentrated the attention here on the Daubechies-like asymmetric wavelets with smallest possible support. However, we believe that Mallat's [2] symmetric wavelets, though infinite, can be treated similarly through truncation.

I wish to comment on my experience with symbolic calculations which played a significant role in deriving explicit formulae for the $k \times 2k$ wavelets. I found it necessary to work much more interactively than in numerical calculations. I could not rely on proven programming modules to work with even slightly changed data—so frequent checks on the progress of the calculations, the form and size of current results and the adjustments of procedures were important. Very often a small manual improvement meant the difference between success and failure, i.e. between obtaining a simple, informative solution and being flooded by pages of cumbersome expressions—or the program running out of memory. Successful use of MAPLE was a novel experience for a numerical analyst.

References

[1] *I. Daubechies*: Orthonormal bases of compactly supported wavelets, Comm. Pure Appl. Math. *41* (1988), 909–996.

[2] *S. Mallat*: A theory for multiresolution signal decomposition: The wavelet representation, IEEE Trans. Pattern Anal. and Machine Intell. *11* (1989), 674–693.

[3] *Y. Meyer*: Ondelettes et Opèrateurs, Hermann, Paris, 1990.

[4] *G. Strang*: Wavelets and dilation equations: A brief introduction, SIAM Review *31(4)* (1989), 614–627.

Author's address: *Jaroslav Kautsky*, School of Information Science and Technology, Flinders University, GPO Box 2100, Adelaide, SA 5042, Australia.