CDI15

7. *Daubechies wavelets (1D)*

1113, 1127, 1130; 428 SXD

7.1. Daubechies scaling and wavelet vectors

7.2. Daubechies transform D_1 (level 1)

7.3. Daubechies transform D_r (level r)

7.4. Multiresolution analysis (MRA) and compression

7.5. Other wavelets

Summary

James S. Walker
A primer on wavelets and their scientific application
Chapman and Hall, 1999

7.1. Daubechies scaling and wavelet vectors

Let

$$\alpha_1 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \alpha_2 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \alpha_3 = \frac{3-\sqrt{3}}{4\sqrt{2}}, \alpha_4 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$

(we will see later where these numbers come from). Approximately,

$$\alpha_1 \simeq 0.48296, \alpha_2 \simeq 0.83652, , \alpha_3 \simeq 0.22414 \ \ , \alpha_4 \simeq -0.12941.$$

Note that

$$\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 = 1, \qquad [1]$$

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = \sqrt{2}, \qquad [2]$$

$$a_1\alpha_3 + \alpha_2\alpha_4 = 0, \qquad [3]$$

$$\alpha_4 - \alpha_3 + \alpha_2 - \alpha_1 = 0. \qquad [4]$$

$$0\alpha_4 - 1\alpha_3 + 2\alpha_2 - 3\alpha_1 = 0. \qquad [5]$$

Scaling vectors

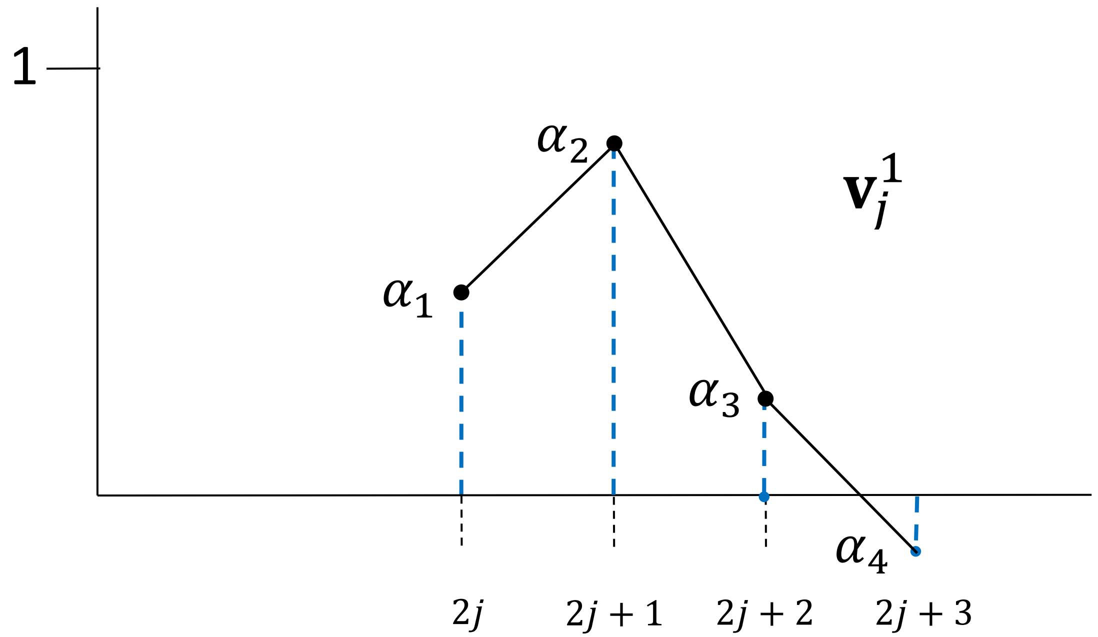
Define, for $j = 0, \dots, \frac{N}{2} - 1$, the 1[st] level scaling vectors $\mathbf{v}_j^1$ by

$$\mathbf{v}_j^1 = \alpha_1 \mathbf{v}_{2j}^0 + \alpha_2 \mathbf{v}_{2j+1}^0 + \alpha_3 \mathbf{v}_{2j+2}^0 + \alpha_4 \mathbf{v}_{2j+3}^0,$$

with the convention that $N + k \equiv k$ (wrap-around). In detail,

$$\mathbf{v}_0^1 = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, 0, \dots, 0),$$
$$\mathbf{v}_1^1 = (0,0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, 0, \dots, 0),$$
$$\dots$$
$$\mathbf{v}_{N/2-2}^1 = (0, \dots, 0, \alpha_1, \alpha_2, \alpha_3, \alpha_4),$$
$$\mathbf{v}_{N/2-1}^1 = (\alpha_3, \alpha_4, 0, \dots, 0, \alpha_1, \alpha_2).$$

Remark. If we define $\tau_s(\boldsymbol{f}) = (f_{N-s}, \dots, f_{N-1}, f_0, \dots, f_{N-s-1})$, which means shifting $\boldsymbol{f}$ by s units, then $\mathbf{v}_{j+1}^1 = \tau_2(\mathbf{v}_j^1)$, for $j = 0, \dots, N/2 - 1$.



For $r > 1$, the r th level scaling vectors $\mathbf{v}_j^r$, $j = 0, \ldots, N/2^r - 1$, are defined recursively (assuming N is divisible by 2^r) as

$$\mathbf{v}_j^r = \alpha_1 \mathbf{v}_{2j}^{r-1} + \alpha_2 \mathbf{v}_{2j+1}^{r-1} + \alpha_3 \mathbf{v}_{2j+2}^{r-1} + \alpha_4 \mathbf{v}_{2j+3}^{r-1}.$$

Proposition. The scaling vectors of any level form an orthonormal system.
Proof. That $|\mathbf{v}_j^1| = 1$ follows from [1] on page 2, and that $\mathbf{v}_j^1 \cdot \mathbf{v}_k^1 = 0$ follows from [3] on page 2 when $k = \pm 1$ and it is obvious otherwise. The statement for $r > 1$ can be shown by induction.

Wavelets

The 1st level wavelet vectors $\mathbf{w}_j^1, j = 1, \dots, N/2$, are defined by

$$\mathbf{w}_j^1 = \alpha_4 \mathbf{v}_{2j}^0 - \alpha_3 \mathbf{v}_{2j+1}^0 + \alpha_2 \mathbf{v}_{2j+2}^0 - \alpha_1 \mathbf{v}_{2j+3}^0,$$

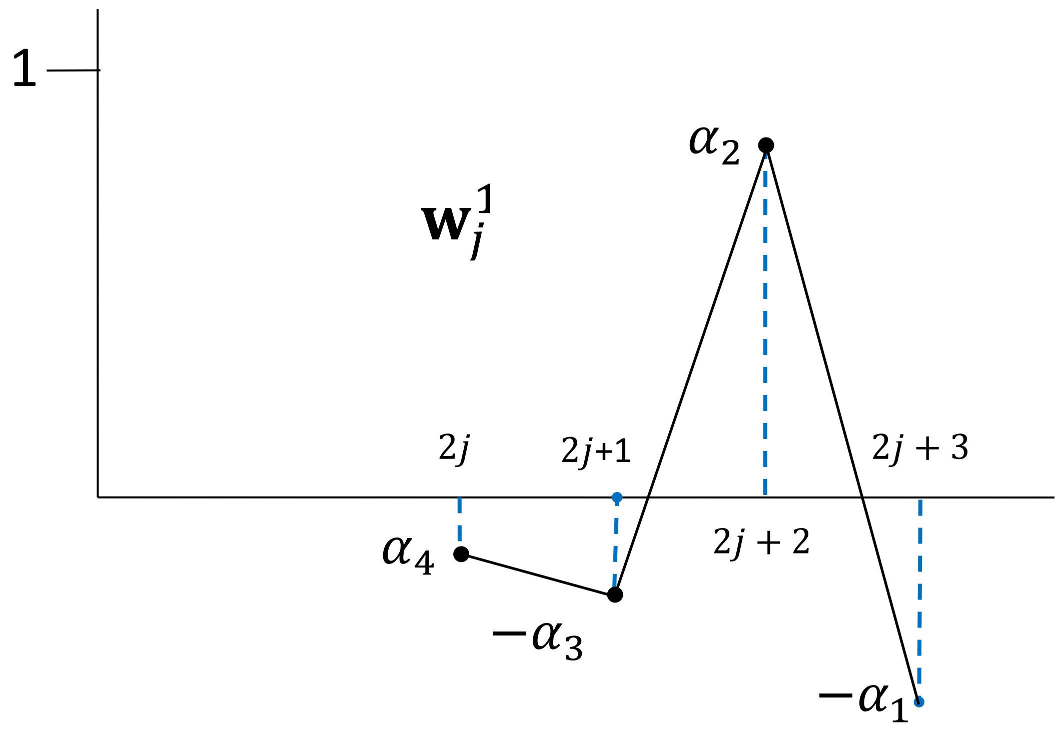
again with the wrap-around convention that $N + k \equiv k$.

In detail,

$$\mathbf{w}_0^1 = (\alpha_4, -\alpha_3, \alpha_2, -\alpha_1, 0, \dots, 0), \dots, \mathbf{w}_{N/2-1}^1 = (\alpha_2, -\alpha_1, 0, \dots, 0, \alpha_4, -\alpha_3).$$

For $r > 1$, the r th level wavelet vectors $\mathbf{w}_j^r$, $j = 0, \dots, N/2^r - 1$, are defined recursively (assuming N is divisible by 2^r) as

$$\mathbf{w}_j^r = \alpha_4 \mathbf{v}_{2j}^{r-1} - \alpha_3 \mathbf{v}_{2j+1}^{r-1} + \alpha_2 \mathbf{v}_{2j+2}^{r-1} - \alpha_1 \mathbf{v}_{2j+3}^{r-1}.$$



Computation of the scaling and wavelet vectors, and applications to compression:

The result are arrays V and W such that $V[r]$ is the matrix of the scaling vectors of level r ($r = 0,1, \dots, n$) and $W[r-1]$ the matrix of the wavelet vectors of level r ($r = 1, \dots, n$), where n is the highest integer such that $2^n|N$

Proposition. The wavelet vectors of level r form an orthonormal system which is orthogonal to the scaling vectors of level r.

Proof. It is similar to the preceding proof.

If we let $\mathcal{V}^r$ and $\mathcal{W}^r$ denote the spaces spanned by the $\mathbf{v}_j^r$ and $\mathbf{w}_j^r$, respectively, then we have

$$\mathcal{V}^r, \mathcal{W}^r \subset \mathcal{V}^{r-1},$$

$$\dim \mathcal{V}^r = N/2^r = \dim \mathcal{W}^r,$$

$$\mathcal{V}^{r-1} = \mathcal{V}^r \perp \mathcal{W}^r.$$

$$\begin{aligned} \mathcal{V}^0 &= \mathcal{V}^1 \perp \mathcal{W}^1 \\ &= \mathcal{V}^2 \perp \mathcal{W}^2 \perp \mathcal{W}^1 \\ &= \mathcal{V}^r \perp \mathcal{W}^r \perp \cdots \perp \mathcal{W}^2 \perp \mathcal{W}^1. \end{aligned}$$

This is the basis of the MRA developed in section 7.4 below.

7.2. Daubechies transform D_1

It is defined by

$$D_1 \colon \boldsymbol{f} \mapsto \boldsymbol{a}^1 | \boldsymbol{d}^1, \text{ with } a_j^1 = \boldsymbol{f} \cdot \mathbf{v}_j^1,\ d_j^1 = \boldsymbol{f} \cdot \mathbf{w}_j^1.$$

We call $\boldsymbol{a}^1$ and $\boldsymbol{d}^1$ the 1st *level* (Daubechies) *trend* and *fluctuation* (or *difference*) respectively.

Actually, a_j^1 is a (weighted) average of four successive values of $\boldsymbol{f}$ multiplied by $\sqrt{2}$ (by [2] on page 2).

By contrast, d_j^1 can be seen, by the relation [4] on page 2, as differencing operation on $\boldsymbol{f}$, involving four successive values of $\boldsymbol{f}$. The main property of $\boldsymbol{d}$ is the following:

> *If $\boldsymbol{f}$ is (approximately) linear over the support of the Daubechies wavelet $\mathbf{w}_j^1$, then d_j^1 is (approximately) zero.*

This follows readily from identities [4] and [5] on page 2.[N1]

The usefulness of D_1, far better than the Haar transform, stems from the property above and the following:

Proposition. D_1 preserves energy: $\mathcal{E}(\boldsymbol{f}) = \mathcal{E}(D_1(\boldsymbol{f}))$.

Proof. Let $\beta_1 = \alpha_4, \beta_2 = -\alpha_3, \beta_3 = \alpha_2, \beta_4 = -\alpha_1$. Consider the $N \times N$ matrix

$$D_N = \begin{pmatrix} \mathbf{v}^1 \\ \mathbf{w}^1 \\ \mathbf{v}^2 \\ \mathbf{w}^2 \\ \vdots \\ \mathbf{v}^{N/2} \\ \mathbf{w}^{N/2} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & 0 & 0 & 0 & 0 \cdots & 0 & 0 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & 0 & 0 & 0 & 0 \cdots & 0 & 0 \\ 0 & 0 & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & 0 & 0 \cdots & 0 & 0 \\ 0 & 0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & 0 & 0 \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_3 & \alpha_4 & 0 & 0 & 0 & 0 & 0 & 0 \cdots & \alpha_1 & \alpha_2 \\ \beta_3 & \beta_4 & 0 & 0 & 0 & 0 & 0 & 0 \cdots & \beta_1 & \beta_2 \end{pmatrix}$$

This matrix is orthonormal ($D_N D_N^T = I_N$) and

$$\left(a_0, d_0, a_1, d_1, \ldots, a_{N/2-1}, d_{N/2-1}\right)^T = D_N \boldsymbol{f}^T.$$

Therefore,

$$a_0^2 + d_0^2 + \cdots + a_{N/2-1}^2 + d_{N/2-1}^2 = (D_N \boldsymbol{f}^T)^T D_N \boldsymbol{f}^T = \boldsymbol{f} D_N^T D_N \boldsymbol{f}^T = \boldsymbol{\mathcal{E}}(\boldsymbol{f}),$$

and it is clear that $a_0^2 + d_0^2 + \cdots + a_{N/2-1}^2 + d_{N/2-1}^2 = \mathcal{E}\big(D_1(\boldsymbol{f})\big)$.

7.3 Daubechies transform D_r

The higher level Daubechies transform D_r is defined recursively by

$$D_r\colon \boldsymbol{f} \mapsto \boldsymbol{a}^r|\boldsymbol{d}^r|\boldsymbol{d}^{r-1}|\cdots|\boldsymbol{d}^1, \text{ with } \boldsymbol{a}^r|\boldsymbol{d}^r = D_1(\boldsymbol{a}^{r-1}).$$

We call $\boldsymbol{a}^r$ and $\boldsymbol{d}^r$ the r^{th} *level* (Daubechies) *trend* and *fluctuation* (or *difference*), respectively, of the signal $\boldsymbol{f}$.

Proposition. As for the Haar wavelets, we have that

$$a_j^r = \boldsymbol{f}\cdot\mathbf{v}_j^r,\ d_j^r = \boldsymbol{f}\cdot\mathbf{w}_j^r.$$ **N2**

Proposition. The transform D_r preserves energy.**N3**

Remark. It turns out that a_j^2 is a (weighted) average of ten successive values of $\boldsymbol{f}$ multiplied by 2. This observation continues for higher r: a_j^r is a weighted average over longer intervals as r increases, multiplied by $2^{r/2}$.

Remark. Let us indicate where the values $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ come from. Set, as before,

$$\beta_1 = \alpha_4, \beta_2 = -\alpha_3, \beta_3 = \alpha_2, \beta_4 = -\alpha_1,$$

so that

$$\mathbf{w}_j^1 = \beta_1 \mathbf{v}_{2j}^0 + \beta_2 \mathbf{v}_{2j+1}^0 + \beta_3 \mathbf{v}_{2j+2}^0 + \beta_4 \mathbf{v}_{2j+3}^0.$$

Since we want that these form an orthonormal set, we require that

$$\beta_1^2 + \beta_2^2 + \beta_3^2 + \beta_4^2 = 1 \text{ and } \beta_1\beta_3 + \beta_2\beta_4 = 0.$$

We also need that

$$\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0 \text{ and } 0\beta_1 + 1\beta_2 + 2\beta_3 + 3\beta_4 = 0.$$

Now we can solve these relations and we obtain $\pm$ the values quoted on page two, or $\pm$ those values reordered as $\alpha_3, \alpha_4, \alpha_1, \alpha_2$. If we want that the sum be positive (actually $\sqrt{2}$), we have two solutions: $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and $\alpha_3, \alpha_4, \alpha_1, \alpha_2$. We choose the first because $\alpha_1 < \alpha_2$.

7.4. MRA and compression

As in the case of the Haar wavelets, define

$$A^r(\boldsymbol{f}) = \Pi_{\mathcal{V}^r}(\boldsymbol{f}) = \sum_j (\boldsymbol{f} \cdot \mathbf{v}_j^r)\mathbf{v}_j^r = \sum_j a_j^r \, \mathbf{v}_j^r \ (r = 0,1, \dots, n),$$

$$D^r(\boldsymbol{f}) = \Pi_{\mathcal{W}^r}(\boldsymbol{f}) = \sum_j (\boldsymbol{f} \cdot \mathbf{w}_j^r)\mathbf{w}_j^r = \sum_j d_j^r \, \mathbf{w}_j^r \ (r = 1, \dots, n).$$

We say that $A^r(\boldsymbol{f})$ is the *level* (or *scale*) r approximation of $\boldsymbol{f}$. By the formula above, it can be computed from the level r trend vector $\boldsymbol{a}^r$ as $\sum_j a_j^r \, \mathbf{v}_j^r$, provided we also know the level r scaling vectors $\mathbf{v}_j^r$. Similarly, $D^r(\boldsymbol{f})$ is the *level* (or *scale*) r difference (or detail) vector and it can be computed from the level r fluctuation vector $\boldsymbol{d}^r$ as $\sum_j d_j^r \, \mathbf{w}_j^r$, provided we also know the level r wavelet vectors $\mathbf{w}_j^r$.

Taking into account the fact that $\mathcal{V}^{r-1} = \mathcal{V}^r \perp \mathcal{W}^r$ (page 7), we have

$$A^{r-1}(\boldsymbol{f}) = A^r(\boldsymbol{f}) + D^r(\boldsymbol{f}).$$

So the level $r-1$ approximation can be decomposed as the sum of the (coarser) level r approximation and the level r difference vector.

Iterating, we see that

$$\boldsymbol{f} = A^0(\boldsymbol{f}) = A^1(\boldsymbol{f}) + D^1(\boldsymbol{f}) = A^2(\boldsymbol{f}) + D^2(\boldsymbol{f}) + D^1(\boldsymbol{f})$$
$$= A^r(\boldsymbol{f}) + D^r(\boldsymbol{f}) + \cdots + D^2(\boldsymbol{f}) + D^1(\boldsymbol{f}).$$

for any allowable r.

Compression/decompression. We can regard the level r trend vector $\boldsymbol{a}^r$ as a (lossy) *compression* of $\boldsymbol{f}$ (the compression factor is $1/2^r$) and the (iterative) function `D4trend(f,r)` provides and efficient means to compute it. The vector $A^r(\boldsymbol{f})$ is then the *decompression* of $\boldsymbol{a}^r$ and one way to compute it is to use the formula $\sum_j a_j^r\,\mathbf{v}_j^r$, which presupposes knowing the matrix $V[r]$ of the level r scaling vectors $\mathbf{v}_j^r$.

Fast decompression. We are going to describe a procedure to compute $A^r(\boldsymbol{f})$ in terms $\boldsymbol{a}^r$ that does not depend on the matrix $V[r]$.

To keep the indices in a convenient range, define $h = [h_0, h_1, h_2, h_3] = [\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ and $g = [g_0, g_1, g_2, g_3] = [\beta_1, \beta_2, \beta_3, \beta_4]$.

Now we need to introduce the *high filter* H. This is the linear map

$$H: \mathbf{R}^N \to \mathbf{R}^N,\ H(\boldsymbol{x})_k = \sum h_j x_{k-j},$$

with the convention that the components with index out of range are taken to be 0. The *low filter* G is defined in a similar way, but using the g_j instead of the h_j.

We also need the *upsampling* operator

$$U: \mathbf{R}^m \to \mathbf{R}^{2m},\ \ (x_0, x_1, \dots, x_{m-1}) \mapsto (x_0, 0, x_1, 0, \dots, x_{m-1}, 0).$$

Theorem. The vector $A^r(\boldsymbol{f})$ is the result of applying r times HU to $\boldsymbol{a}^r$. The vector $D^r(\boldsymbol{f})$ is the result of applying $r-1$ times the operator HU to the vector $GU\boldsymbol{d}^r$.[N4]

7.5. Other wavelets

There are other Daubechies wavelets, mainly Daub*J*, with $J = 6,8, \dots ,20$ (we have worked out Daub4), and Coif*I* for $I = 6,12,18,24,30$.

Let us consider, for definiteness, the case $J = 6$. In this case there are six α:

$$\alpha_0 = 0.332670552950083, \qquad \alpha_1 = 0.806891509311092,$$

$$\alpha_2 = 0.459877502118491, \qquad \alpha_3 = -0.135011020010255,$$

$$\alpha_4 = -0.0854412738820267, \quad \alpha_5 = 0.0352262918857095.$$

They satisfy the identities

$$\sum_l \alpha_l^2 = 1, \sum_l \alpha_l = \sqrt{2}.$$

With them we can define scaling vectors $\mathbf{v}_j^1$ as for Daub4 (in this case there are two wraparounds at the end).

Now with $\beta_l = (-1)^l \alpha_{5-l}$, $l = 1, \ldots, 6$, we can define wavelet vectors $\mathbf{w}_j^1$ and the machinery can be copied to this case. What is the advantage? Basically the relations

$$\sum_l \beta_l = 0,\ \sum_l l\beta_l = 0,\ \sum_l l^2 \beta_l = 0.$$

These identities insure that

If $\boldsymbol{f}$ is (approximately) quadratic over the support of the Daubechies wavelet $\mathbf{w}_j^1$, then d_j^1 is (approximately) zero. And similarly for d_j^r.

In general,

If $\boldsymbol{f}$ is (approximately) a polynomial of degree $< J/2$ over the support of the Daubechies wavelet $\mathbf{w}_j^1$, then d_j^1 is (approximately) zero. And similarly for d_j^r.

The framework developed in the preceding pages for the system Daub4 can be adapted to the Daub6 system in a straightforward way.

Summary

Data. A vector $\boldsymbol{f} \in \mathbf{R}^N$. Usually obtained by sampling an analog signal φ at N points. We let n denote the highest integer such that 2^n divides N.

Lossy compression of level $r \leq n$. The trend vector $\boldsymbol{a}^r$ of level r. The compression factor is $1/2^r$.

Decompression. The level r approximation of $\boldsymbol{f}$, $A^r(\boldsymbol{f}) = \sum_j a_j^r \mathbf{v}_j^r$. This can be obtained by first calculating the array V of scaling vectors.

Fast decompression algorithm. Apply r times the operator HU to $\boldsymbol{a}^r$.

Differences. The difference vector $D^r(\boldsymbol{f}) = A^{r-1}(\boldsymbol{f}) - A^r(\boldsymbol{f})$ can also be computed efficiently from the level r fluctuation $\boldsymbol{d}^r$: apply $r-1$ times the operator HU to $GU\boldsymbol{d}^r$.

Multiresolution. $\boldsymbol{f} = A^r(\boldsymbol{f}) + D^r(\boldsymbol{f}) + \cdots + D^2(\boldsymbol{f}) + D^1(\boldsymbol{f})$.

Notes

N1 (p. 8). Since a sample of a function φ with small difference δ is locally approximately linear, their 1st level Daubechies fluctuation is approximately zero. We can quantify this a bit more as follows. For $l = 0,1,2,3$,

$$f_{2j+l} = \varphi(t_{2j+l}) = \varphi(t_{2j} + l\delta) = f_{2j} + \varphi'(t_{2j})l\delta + O(\delta^2)$$

Consequently

$$\alpha_4 f_{2j} - \alpha_3 f_{2j+1} + \alpha_2 f_{2j+2} - \alpha_1 f_{2j+3}$$

$$= f_{2j}(\alpha_4 - \alpha_3 + \alpha_2 - a_1) + \varphi'(t_{2j})\delta(0\alpha_4 - 1\alpha_3 + 2\alpha_2 - 3a_1) + O(\delta^2)$$

$$= O(\delta^2).$$

We have used the relations [4] and [5] on page 2.

N2 (p. 10)**.** The proof follows readily by induction:

$$\begin{aligned} a_j^r &= \alpha_1 a_{2j}^{r-1} + \alpha_2 a_{2j+1}^{r-1} + \alpha_3 a_{2j+2}^{r-1} + \alpha_1 a_{2j+3}^{r-1} \\ &= \alpha_1\left(\boldsymbol{f} \cdot \mathbf{v}_{2j}^{r-1}\right) + \alpha_2\left(\boldsymbol{f} \cdot \mathbf{v}_{2j+1}^{r-1}\right) + \alpha_3\left(\boldsymbol{f} \cdot \mathbf{v}_{2j+2}^{r-1}\right) + \alpha_4\left(\boldsymbol{f} \cdot \mathbf{v}_{2j+3}^{r-1}\right) \\ &= \boldsymbol{f} \cdot \left(\alpha_1 \mathbf{v}_{2j}^{r-1} + \alpha_2 \mathbf{v}_{2j+1}^{r-1} + \alpha_3 \mathbf{v}_{2j+2}^{r-1} + \alpha_4 \mathbf{v}_{2j+3}^{r-1}\right) \\ &= \boldsymbol{f} \cdot \mathbf{v}_j^r. \end{aligned}$$

The proof of the relation $d_j^r = \boldsymbol{f} \cdot \mathbf{w}_j^r$ is similar.

N3 (p. 10)**.** The proof is by induction again. For $r = 1$, see the proposition on page 9. For $r > 1$, we have:

$$\begin{aligned} &\mathcal{E}(\boldsymbol{a}^r|\boldsymbol{d}^r|\boldsymbol{d}^{r-1}|\cdots|\boldsymbol{d}^2|\boldsymbol{d}^1) \\ &\quad= \mathcal{E}(\boldsymbol{a}^r|\boldsymbol{d}^r) + \mathcal{E}(\boldsymbol{d}^{r-1}|\cdots|\boldsymbol{d}^2|\boldsymbol{d}^1) \\ &\quad= \mathcal{E}\big(D_1(\boldsymbol{a}^{r-1})\big) + \mathcal{E}(\boldsymbol{d}^{r-1}|\cdots|\boldsymbol{d}^2|\boldsymbol{d}^1) \\ &\quad= \mathcal{E}(\boldsymbol{a}^{r-1}) + \mathcal{E}(\boldsymbol{d}^{r-1}|\cdots|\boldsymbol{d}^2|\boldsymbol{d}^1) \\ &\quad= \mathcal{E}(\boldsymbol{a}^{r-1}|\boldsymbol{d}^{r-1}|\cdots|\boldsymbol{d}^2|\boldsymbol{d}^1) = \mathcal{E}(\boldsymbol{f}). \end{aligned}$$

N4 (p. 14). For the first part, the key point is that if $\boldsymbol{a} = \sum_{j=0}^{N/2^r-1} a_j \mathbf{v}_j^r \in \mathcal{V}^r$ then $HU(a_0, \dots, a_{N/2^r-1})$ are the components of $\boldsymbol{a}$ with respect to the basis $\mathbf{v}_k^{r-1}$ of $\mathcal{V}^{r-1}$ (remember that $\mathcal{V}^r \subset \mathcal{V}^{r-1}$). To see this, it is enough to expand the $\mathbf{v}_j^r$ in terms of the $\mathbf{v}_k^{r-1}$ ($\mathbf{v}_j^r = \sum_k h_k \mathbf{v}_{2j+k}^{r-1}$, with wraparound for the index $2j+k$) and rearrange the expression:

$$\boldsymbol{a} = \sum_{j=1}^{2^r-1} a_j \mathbf{v}_j^r = \sum_j \sum_k a_j \, h_k \mathbf{v}_{2j+k}^{r-1} = \sum_l \sum_j a_j \, h_{l-2j} \mathbf{v}_l^{r-1},$$

which shows that the l-th component of $\boldsymbol{a} \in \mathcal{V}^{r-1}$ is $b_l = \sum_j a_j h_{l-2j}$. If we now introduce the vector $\boldsymbol{u} = (a_0, 0, a_1, 0, \dots, a_{2^r-1}, 0)$, we have

$$b_l = \sum_j u_{2j} h_{l-2j} = \sum_j u_j h_{l-j} = H\left(u_0, \dots, u_{N/2^{r-1}-1}\right)_l,$$

and this proves the claim.

The second part is proved similarly. The key point is that if $\boldsymbol{d} = \sum_{j=0}^{N/2^r-1} d_j \mathbf{w}_j^r \in \mathcal{W}^r$, then $GU(d_0, \dots, d_{N/2^r-1})$ are the components of $\boldsymbol{d}$ with respect to the basis $\mathbf{v}_k^{r-1}$ of $\mathcal{V}^{r-1}$. And this is proved much in the same way, inasmuch as $\mathbf{w}_j^r = \sum_k g_k \mathbf{v}_{2j+k}^{r-1}$. And then we only need to apply $r-1$ times the operator HU.