CDI15

6. *Haar wavelets (1D)*

1027, 1104, 1110 | 414, 416, 428 SXD

Notations

6.1. The Haar transforms

6.2. Haar wavelets

6.3. Multiresolution analysis

6.4. Compression/decompression

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

Notations

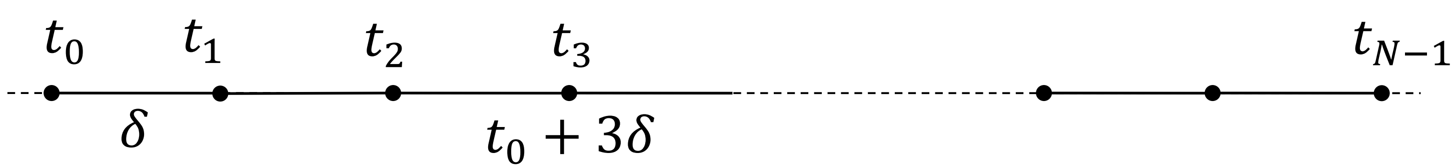
Discrete signal (of length N): $\boldsymbol{f} = (f_0, f_1, \dots, f_{N-1}) \in \mathbb{R}^N$. The choice of indexing, starting with 0, is not the usual convention in mathematics texts, but we adopt it as it facilities the relation with computations.

Usually $\boldsymbol{f}$ is a *sample* of a function (or **continuous signal**) $\varphi = \varphi(t)$:

$$f_j = \varphi(t_j),\, t_0 < t_1 < \cdots < t_{N-1}.$$

If not said otherwise, the t_j are assumed to be equally spaced:

$$t_j = t_0 + j\delta,\, \delta > 0 \text{ a constant (\textit{sampling spacing}).}$$



Energy. The norm squared,

$$|\boldsymbol{f}|^2 = f_0^2 + f_1^2 + \cdots + f_{N-1}^2,$$

is also called the *energy* of the signal. Alternative notation: $\mathcal{E}(\boldsymbol{f})$.

Trend and fluctuation signals

The (first) ***trend*** *signal* $\boldsymbol{a}^1(\boldsymbol{f})$ is defined by the formula

$$\boldsymbol{a}^1(\boldsymbol{f}) = (a_0, a_1, \dots, a_{N/2-1}),$$

$$a_j = (f_{2j} + f_{2j+1})/\sqrt{2}$$

The (first) ***fluctuation*** *or* ***difference*** *signal* $\boldsymbol{d}^1(\boldsymbol{f})$ is defined by

$$\boldsymbol{d}^1(\boldsymbol{f}) = (d_0, d_1, \dots, d_{N/2-1}),$$

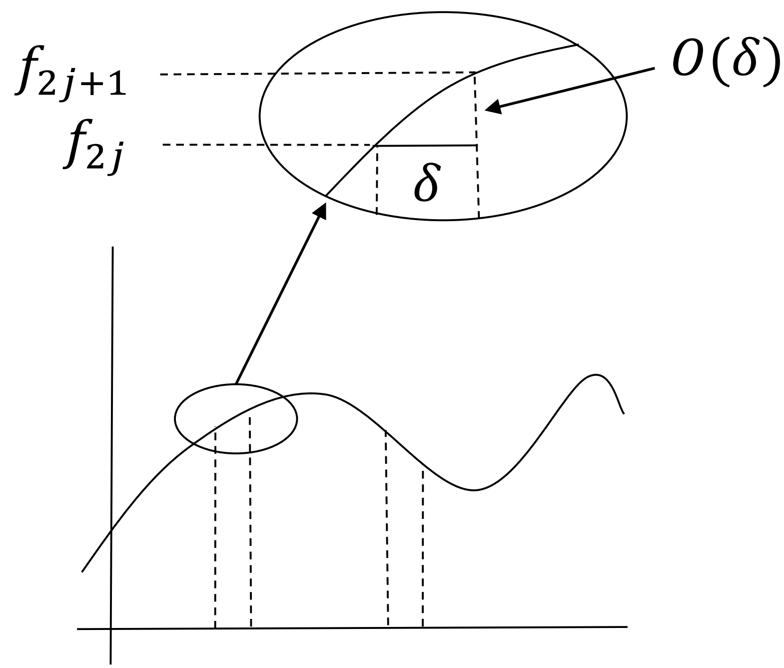
$$d_j = (f_{2j} - f_{2j+1})/\sqrt{2}$$

6. 1. The Haar transforms

The *first level* *Haar transform* of $\boldsymbol{f}$ is the signal defined by

$$H_1(\boldsymbol{f}) = \boldsymbol{a}^1(\boldsymbol{f}) \mid \boldsymbol{d}^1(\boldsymbol{f}).$$

Remark (small fluctuations). If $\boldsymbol{f}$ is a sampling of φ, with a *small* sampling increment δ, then



$$f_{2j+1} \simeq f_{2j} + \varphi'(t_{2j})\delta,$$

and hence

$$|d_j| = O(\delta).$$

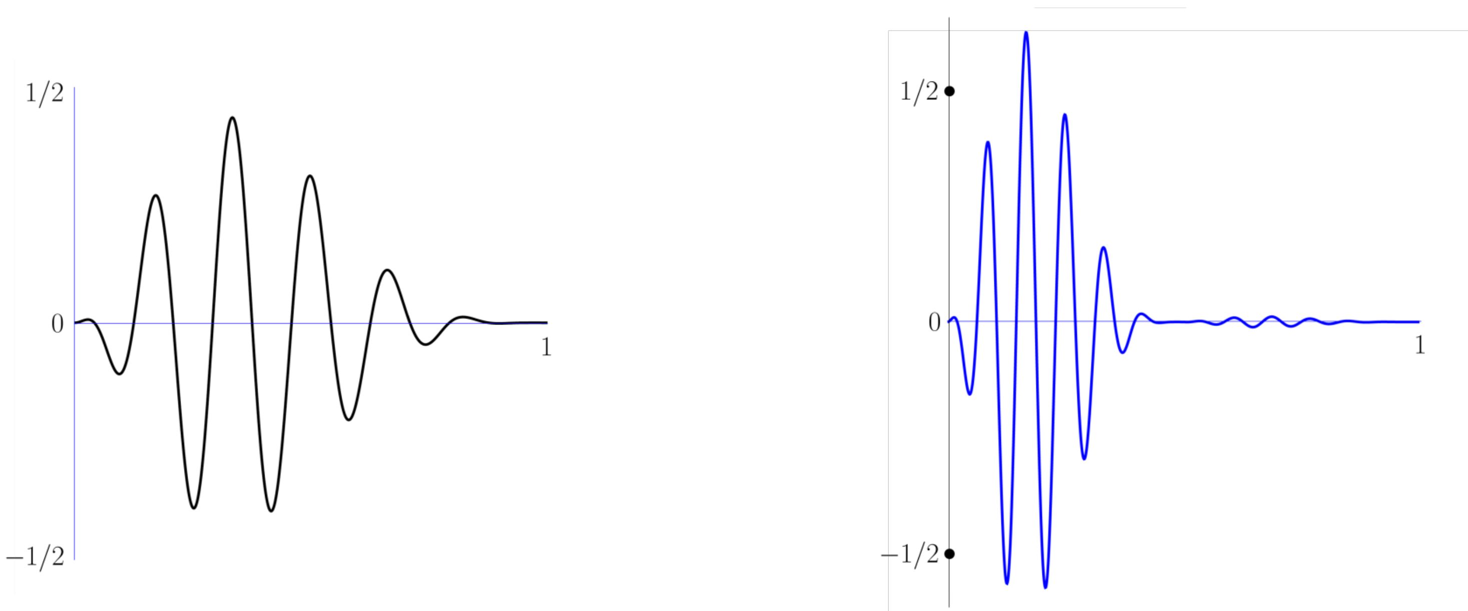
On the other hand, up to $O(\delta)$, $a_j \simeq \sqrt{2} f_{2j} \simeq \sqrt{2} f_{2j+1}$. In other words, the components of $\boldsymbol{d}^1(\boldsymbol{f})$ are small and the signal $\boldsymbol{a}^1(\boldsymbol{f})$ is like $\boldsymbol{f}$ at even (or odd) times, scaled by $\sqrt{2}$.

Proposition (inverse of the Haar transform). It is the map given by

$$\left(a_0, a_1, \ldots, a_{N/2-1}, d_0, d_1, \ldots, d_{N/2-1}\right) \mapsto (f_0, f_1, \ldots, f_{N-1}),$$

where $f_{2j} = \frac{a_j + d_j}{\sqrt{2}}, f_{2j+1} = \frac{a_j - d_j}{\sqrt{2}}, j = 0, \ldots, N/2 - 1.$

Example: Sampling $\varphi(x) = 20x^2(1-x)^4\cos(12\pi x), N = 2^{10}$, and its Haar transform.



Left: φ sampled at N points. Right: Haar transform of the left sample.

Proposition. $\mathcal{E}(\boldsymbol{f}) = \mathcal{E}(\boldsymbol{H}_1(\boldsymbol{f}))$.

Proof: A short computation.[N1]

In the example, $\mathcal{E}(\boldsymbol{f}) = \mathcal{E}\big(\boldsymbol{H}_1(\boldsymbol{f})\big) = 31.83$, while $\mathcal{E}\big(\boldsymbol{a}^1(\boldsymbol{f})\big) = 31.82$ and $\mathcal{E}\big(\boldsymbol{d}^1(\boldsymbol{f})\big) = 0.01$.

Remark (compaction of energy). $\mathcal{E}(\boldsymbol{f}) = \mathcal{E}\big(\boldsymbol{a}^1(\boldsymbol{f})\big) + \mathcal{E}(\boldsymbol{d}^1(\boldsymbol{f}))$ and $\mathcal{E}(\boldsymbol{d}^1(\boldsymbol{f}))$ tends to be small (see Remark on small fluctuations), in which case $\mathcal{E}\big(\boldsymbol{a}^1(\boldsymbol{f})\big)$ amounts to a large percentage of $\mathcal{E}(\boldsymbol{f})$.

The *Haar transform of level* r, $\boldsymbol{H}_r(\boldsymbol{f})$, requires that N is divisible by 2^r and is defined recursively by

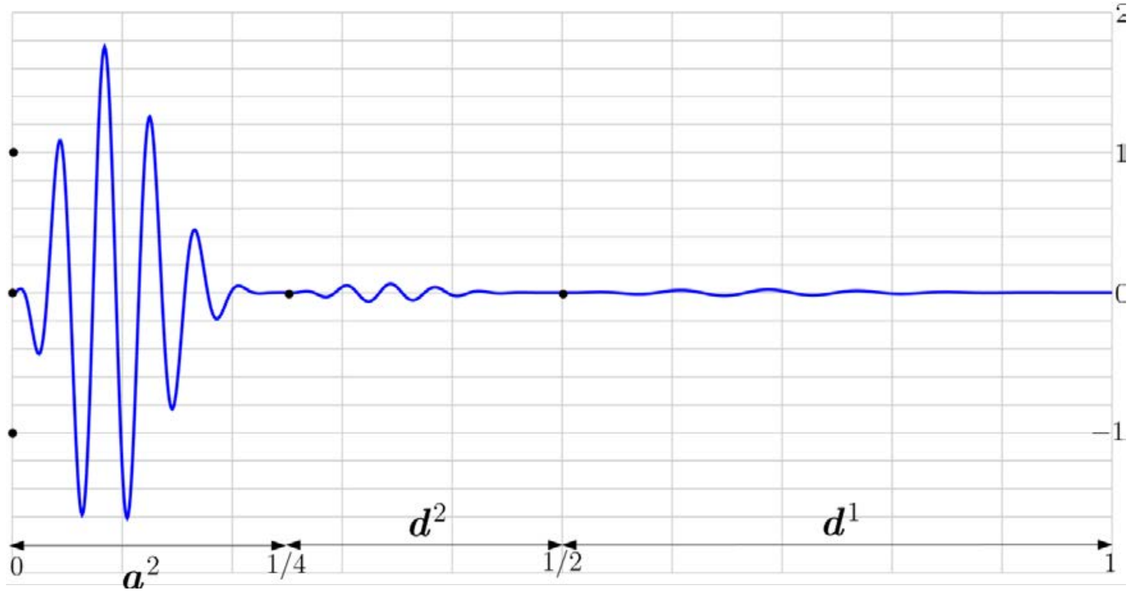
$$H_r(\boldsymbol{f}) = \boldsymbol{a}^r(\boldsymbol{f})|\boldsymbol{d}^r(\boldsymbol{f})|\boldsymbol{d}^{r-1}(\boldsymbol{f})|\cdots|\boldsymbol{d}^1(\boldsymbol{f}), \text{ where}$$

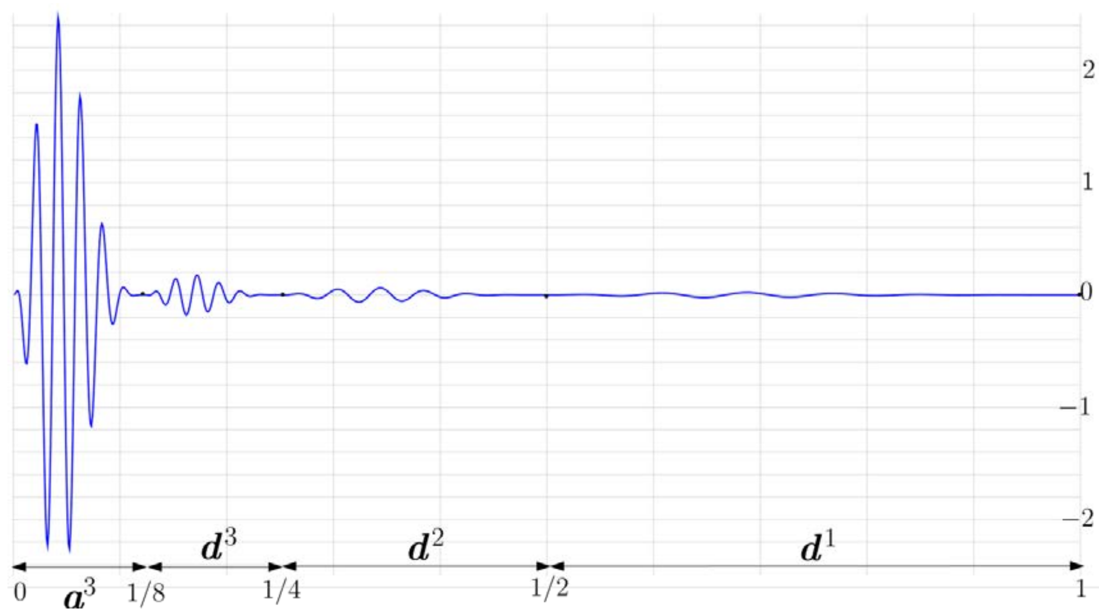
$$\boldsymbol{a}^r(\boldsymbol{f}) = \boldsymbol{a}^1\left(\boldsymbol{a}^{r-1}(\boldsymbol{f})\right) \equiv \left(a_0^r, a_1^r, \ldots, a_{N/2^r-1}^r\right),\ a_j^r = \frac{a_{2j}^{r-1} + a_{2j+1}^{r-1}}{\sqrt{2}}$$

$$\boldsymbol{d}^r(\boldsymbol{f}) = \boldsymbol{d}^1\left(\boldsymbol{a}^{r-1}(\boldsymbol{f})\right) \equiv \left(d_0^r, d_1^r, \ldots, d_{N/2^r-1}^r\right),\ d_j^r = \frac{a_{2j}^{r-1} - a_{2j+1}^{r-1}}{\sqrt{2}}.$$

So $H_r(\boldsymbol{f}) = H_{r-1}\big(\boldsymbol{a}^1(\boldsymbol{f})\big)|\ \boldsymbol{d}^1(\boldsymbol{f})$. [N2]

Examples: 2-level and 3-level Haar transforms of a sample of the function defined on page 5.



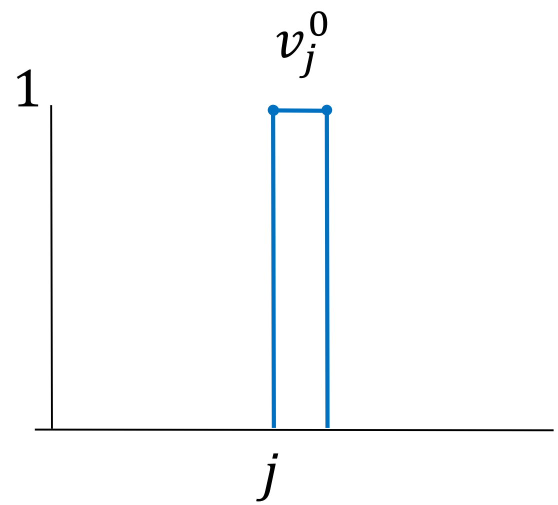


6. 2. Haar Wavelets

Let $\mathbf{v}_0^0 = (1,0,\ldots,0)$, $\mathbf{v}_1^0 = (0,1,\ldots,0)$, …, $\mathbf{v}_{N-1}^0 = (0,0,\ldots,0,1)$, which are called the 0-level *scaling signals*. Hence

$$\boldsymbol{f} = f_0\mathbf{v}_0^0 + f_1\mathbf{v}_1^0 + \cdots + f_{N-1}\mathbf{v}_{N-1}^0,$$

and $f_j = \boldsymbol{f} \cdot \mathbf{v}_j^0$.

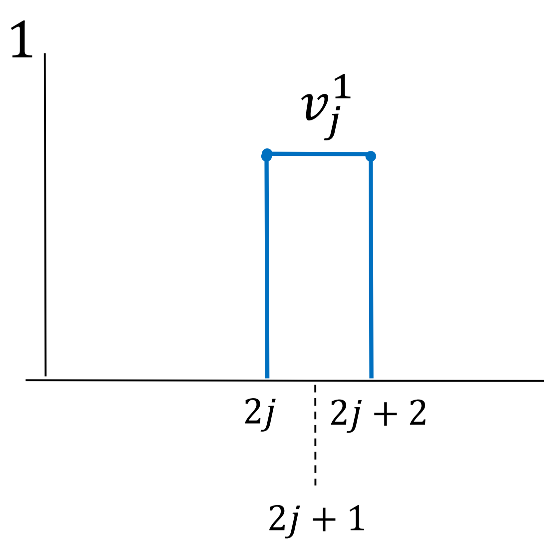


Introduce now what are called 1-level Haar *scaling signals*

$$\mathbf{v}_j^1 = \frac{1}{\sqrt{2}}(\mathbf{v}_{2j}^0 + \mathbf{v}_{2j+1}^0),\, j = 0,\ldots,N/2-1,$$

and the 1-level *Haar wavelets*

$$\mathbf{w}_j^1 = \frac{1}{\sqrt{2}}(\mathbf{v}_{2j}^0 - \mathbf{v}_{2j+1}^0),\, j = 0,\ldots,N/2-1.$$



Proposition. $a_j = \boldsymbol{f} \cdot \mathbf{v}_j^1$, $d_j = \boldsymbol{f} \cdot \mathbf{w}_j^1$.

Remark. $\{\mathbf{v}_1^1, \mathbf{v}_2^1, \dots, \mathbf{v}_{N/2}^1\}$ and $\{\mathbf{w}_1^1, \mathbf{w}_2^1, \dots, \mathbf{w}_{N/2}^1\}$ are orthonormal systems that are orthogonal to each other: $\mathbf{v}_j^1 \cdot \mathbf{w}_k^1 = 0$ for all j, k.

Assuming that N is divisible by 2^r, $r \geq 2$, then the r-level Haar *scaling functions* and wavelets are defined as

$$\mathbf{v}_j^r = \frac{1}{\sqrt{2}}(\mathbf{v}_{2j}^{r-1} + \mathbf{v}_{2j+1}^{r-1}),\ j = 0, \dots, N/2^r - 1,$$

$$\mathbf{w}_j^r = \frac{1}{\sqrt{2}}(\mathbf{v}_{2j}^{r-1} - \mathbf{v}_{2j+1}^{r-1}), j = 0, \dots, N/2^r - 1.$$

Proposition. We have the following formulas:

$$a_j^r = \boldsymbol{f} \cdot \mathbf{v}_j^r,\ \ d_j^r = \boldsymbol{f} \cdot \mathbf{w}_j^r,\ \ j = 0, 1, \dots, N/2^r - 1.$$

Proof: Indeed, for the case $r = 1$, see the Proposition on page 7.

For $r > 1$,

$$a_j^r = \frac{1}{\sqrt{2}}\left(a_{2j}^{r-1} + a_{2j+1}^{r-1}\right) = \frac{1}{\sqrt{2}}\left(\boldsymbol{f} \cdot \mathbf{v}_{2j}^{r-1} + \boldsymbol{f} \cdot \mathbf{v}_{2j+1}^{r-1}\right) = \boldsymbol{f} \cdot \mathbf{v}_j^r, \text{ and}$$

$$d_j^r = \frac{1}{\sqrt{2}}\left(a_{2j}^{r-1} - a_{2j+1}^{r-1}\right) = \frac{1}{\sqrt{2}}\left(\boldsymbol{f} \cdot \mathbf{v}_{2j}^{r-1} - \boldsymbol{f} \cdot \mathbf{v}_{2j+1}^{r-1}\right) = \boldsymbol{f} \cdot \mathbf{w}_j^r.$$

Proposition. The vectors $\mathbf{v}_j^r$, $j \in N_r = \{0, \dots, N/2^r - 1\}$, can be computed directly from the $\mathbf{v}_k^0$ as follows. If J is a subset of N_r, let

$$\mathbf{v}_J^0 = \sum_{j \in J} \mathbf{v}_j^0 \text{ (1's in the positions } J\text{, 0 otherwise).}$$

Then

$$\mathbf{v}_j^r = 2^{-r/2} \mathbf{v}_{[2^r j, 2^r (j+1))}^0$$

(2^r 1's starting at $2^r j$, scaled by $2^{-r/2}$).

In a similar way we have, for $j \in \{1, \dots, N/2^r - 1\}$,

$$\mathbf{w}_j^r = 2^{-r/2} \left(\mathbf{v}_{[2^r j,\, 2^r j + 2^{r-1} - 1)}^0 - \mathbf{v}_{[2^r j + 2^{r-1},\, 2^r (j+1))}^0 \right)$$

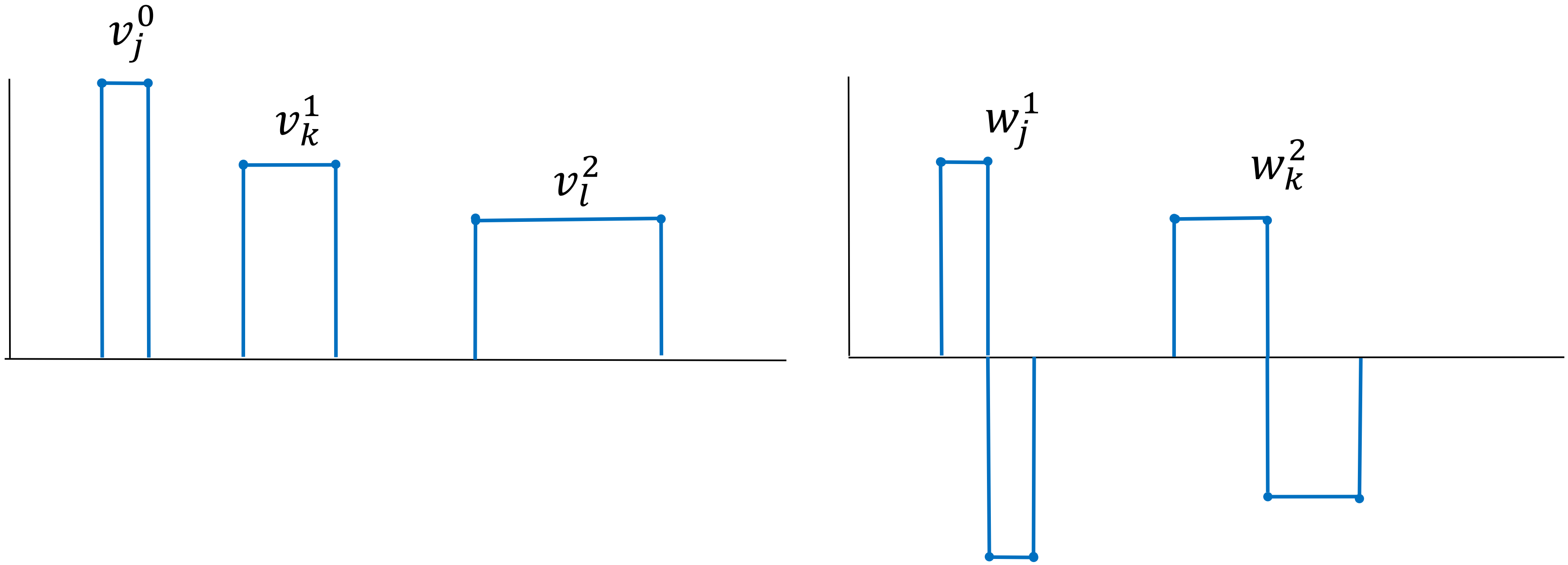
(2^{r-1} 1's starting at $2^r j$ followed by 2^{r-1} (-1)'s, all scaled by $2^{-r/2}$)

In particular it follows that

$$a_j^r = 2^{-r/2} (f_{2^r j} + f_{2^r j + 1} + \cdots + f_{2^r j + 2^r - 1})$$

$$d_j^r = 2^{-r/2} \left(f_{2^r j} + \cdots + f_{2^r j + 2^{r-1} - 1} \right) - 2^{-r/2} \left(f_{2^r j + 2^{r-1}} + \cdots + f_{2^r j + 2^r - 1}) \right)$$

Examples of scaling and wavelet vectors:



Notations. $\mathbb{R}^N = \langle \mathbf{v}_{\mathbf{0}}^0, \mathbf{v}_1^0, \dots, \mathbf{v}_{N-1}^0 \rangle$ will be denoted $\mathcal{V}^0$ and we set, for any r such that N is divisible by 2^r,

$$\mathcal{V}^r = \langle \mathbf{v}_0^r, \mathbf{v}_1^r, \dots, \mathbf{v}_{N/2^r-1}^r \rangle,\ \mathcal{W}^r = \langle \mathbf{w}_0^r, \mathbf{w}_1^r, \dots, \mathbf{w}_{N/2^r-1}^r \rangle$$

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

$$\mathcal{V}^0 \supset \mathcal{V}^1 \supset \cdots \supset \mathcal{V}^r \text{ and } \mathcal{V}^{r-1} = \mathcal{V}^r \perp \mathcal{W}^r.$$

The inclusion of $\mathcal{V}^r$ and $\mathcal{W}^r$ in $\mathcal{V}^{r-1}$, and the fact that $\mathcal{V}^r$ and $\mathcal{W}^r$ are orthogonal, are direct consequences of the defining relations, namely

$$\mathbf{v}_j^r = \tfrac{1}{\sqrt{2}}(\mathbf{v}_{2j}^{r-1} + \mathbf{v}_{2j+1}^{r-1}),\ \mathbf{w}_j^r = \tfrac{1}{\sqrt{2}}(\mathbf{v}_{2j}^{r-1} - \mathbf{v}_{2j+1}^{r-1}).$$

These expressions also show that $\mathbf{v}_0^r, \dots, \mathbf{v}_{N/2^r-1}^r, \mathbf{w}_0^r, \dots, \mathbf{w}_{N/2^r-1}^r$ form a basis of $\mathcal{V}^r$, for we also have

$$\mathbf{v}_{2j}^{r-1} = \tfrac{1}{\sqrt{2}}(\mathbf{v}_j^r + \mathbf{w}_j^r),\ \ \mathbf{v}_{2j+1}^{r-1} = \tfrac{1}{\sqrt{2}}(\mathbf{v}_j^r - \mathbf{w}_j^r).$$

6. 3. Multiresolution analysis

First average and detail signals:

$$\begin{aligned}
\boldsymbol{A}^1(\boldsymbol{f}) &= \frac{1}{\sqrt{2}}\left(a_0, a_0, a_1, a_1, \dots, a_{N/2-1}, a_{N/2-1}\right) \\
&= a_0\mathbf{v}_0^1 + a_1\mathbf{v}_1^1 + \cdots + a_{N/2-1}\mathbf{v}_{N/2-1}^1 \\
&= (\boldsymbol{f}\cdot\mathbf{v}_0^1)\mathbf{v}_0^1 + (\boldsymbol{f}\cdot\mathbf{v}_1^1)\mathbf{v}_1^1 + \cdots + \left(\boldsymbol{f}\cdot\mathbf{v}_{N/2-1}^1\right)\mathbf{v}_{N/2-1}^1 \\
&= \Pi_{\mathcal{V}^1}(\boldsymbol{f}).
\end{aligned}$$

[N3]

$$\begin{aligned}
\boldsymbol{D}^1(\boldsymbol{f}) &= \frac{1}{\sqrt{2}}\left(d_0, -d_0, d_1, -d_1, \dots, d_{N/2-1}, -d_{N/2-1}\right) \\
&= d_0\mathbf{w}_0^1 + d_1\mathbf{w}_1^1 + \cdots + d_{N/2-1}\mathbf{w}_{N/2-1}^1 \\
&= (\boldsymbol{f}\cdot\mathbf{w}_0^1)\mathbf{w}_0^1 + (\boldsymbol{f}\cdot\mathbf{w}_1^1)\mathbf{w}_1^1 + \cdots + \left(\boldsymbol{f}\cdot\mathbf{w}_{N/2-1}^1\right)\mathbf{w}_{N/2-1}^1 \\
&= \Pi_{\mathcal{W}^1}(\boldsymbol{f}).
\end{aligned}$$

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

Proof : Simple calculation. For example (see Remark on the Haar inverse),

$$\frac{1}{\sqrt{2}}(a_0 + d_0) = \frac{1}{2}(f_0 + f_1 + f_0 - f_1) = f_0 \text{ and}$$

$$\frac{1}{\sqrt{2}}(a_0 - d_0) = \frac{1}{2}(f_0 + f_1 - f_0 + f_1) = f_1 \,.$$

In general, we define (while N is divisible by 2^r) the r-th average and detail signals by

$$\begin{aligned}
\boldsymbol{A}^r(\boldsymbol{f}) &= \Pi_{\mathcal{V}^r}(\boldsymbol{f}) \\
&= (\boldsymbol{f} \cdot \mathbf{v}_0^r)\mathbf{v}_0^r + (\boldsymbol{f} \cdot \mathbf{v}_1^r)\mathbf{v}_1^r + \cdots + \left(\boldsymbol{f} \cdot \mathbf{v}_{N/2^r-1}^r\right)\mathbf{v}_{N/2^r-1}^r \\
&= a_0^r\mathbf{v}_0^r + a_1^r\mathbf{v}_1^r + \cdots + a_{N/2^r-1}^r\mathbf{v}_{N/2^r-1}^r, \\
\boldsymbol{D}^r(\boldsymbol{f}) &= \Pi_{\mathcal{W}^r}(\boldsymbol{f}) \\
&= (\boldsymbol{f} \cdot \mathbf{w}_0^r)\mathbf{w}_0^r + (\boldsymbol{f} \cdot \mathbf{w}_1^r)\mathbf{w}_1^r + \cdots + \left(\boldsymbol{f} \cdot \mathbf{w}_{N/2^r-1}^r\right)\mathbf{w}_{N/2^r-1}^r \\
&= d_0^r\mathbf{w}_0^r + d_1^r\mathbf{w}_1^r + \cdots + d_{N/2^r-1}^r\mathbf{w}_{N/2^r-1}^r.
\end{aligned}$$

Proposition (*Multiresolution* of $\boldsymbol{f}$)

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

Proof: It is enough to show that

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

for then the expression results by induction. But this follows from the definitions and the facts established so far:

$$\begin{aligned} \boldsymbol{A}^r(\boldsymbol{f}) + \boldsymbol{D}^r(\boldsymbol{f}) &= \Pi_{\mathcal{V}^r}(\boldsymbol{f}) + \Pi_{\mathcal{W}^r}(\boldsymbol{f}) \\ &= \Pi_{\mathcal{V}^{r-1}}(\boldsymbol{f}) \qquad \text{(because } \mathcal{V}^{r-1} = \mathcal{V}^r \perp \mathcal{W}^r\text{)}^{\textbf{N4}} \\ &= \boldsymbol{A}^{r-1}(\boldsymbol{f}). \end{aligned}$$

Computations

$\boldsymbol{A}^r(\boldsymbol{f})$ and $\boldsymbol{D}^r(\boldsymbol{f})$ are computed by the functions

```
high_filter(f,r), low_filter(f,r).
```

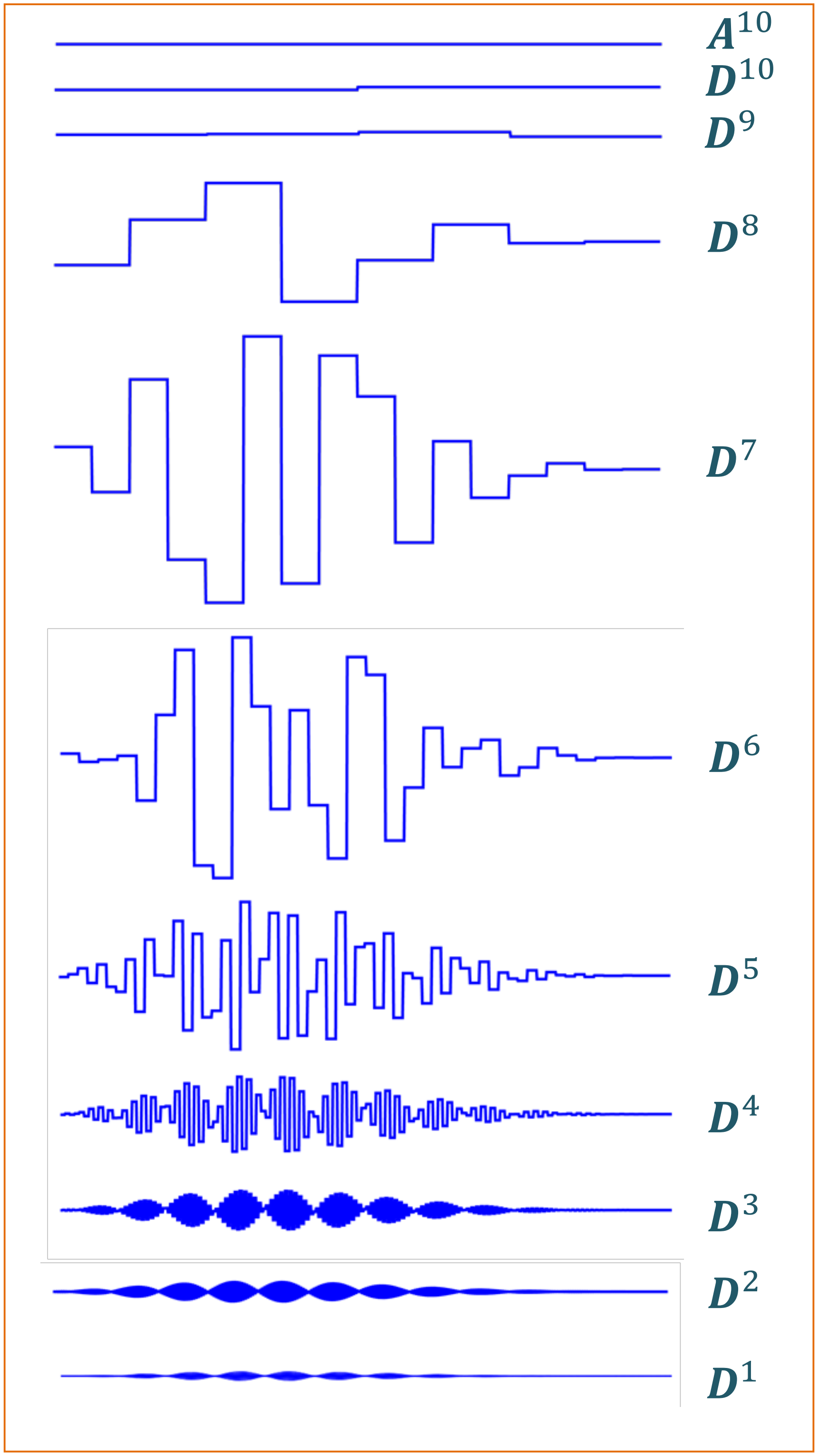
The image on the left of next page shows the graphs of $\boldsymbol{A}^{10}(\boldsymbol{f})$ and $\boldsymbol{D}^{j}(\boldsymbol{f})$**,** $j = 1, \ldots, 10$, where $\boldsymbol{f}$ is the signal of the Example on page 5. By what we have seen so far, the sum of these signals,

$$\boldsymbol{A}^{10} + \sum_{j=1}^{j=10} \boldsymbol{D}^{j},$$

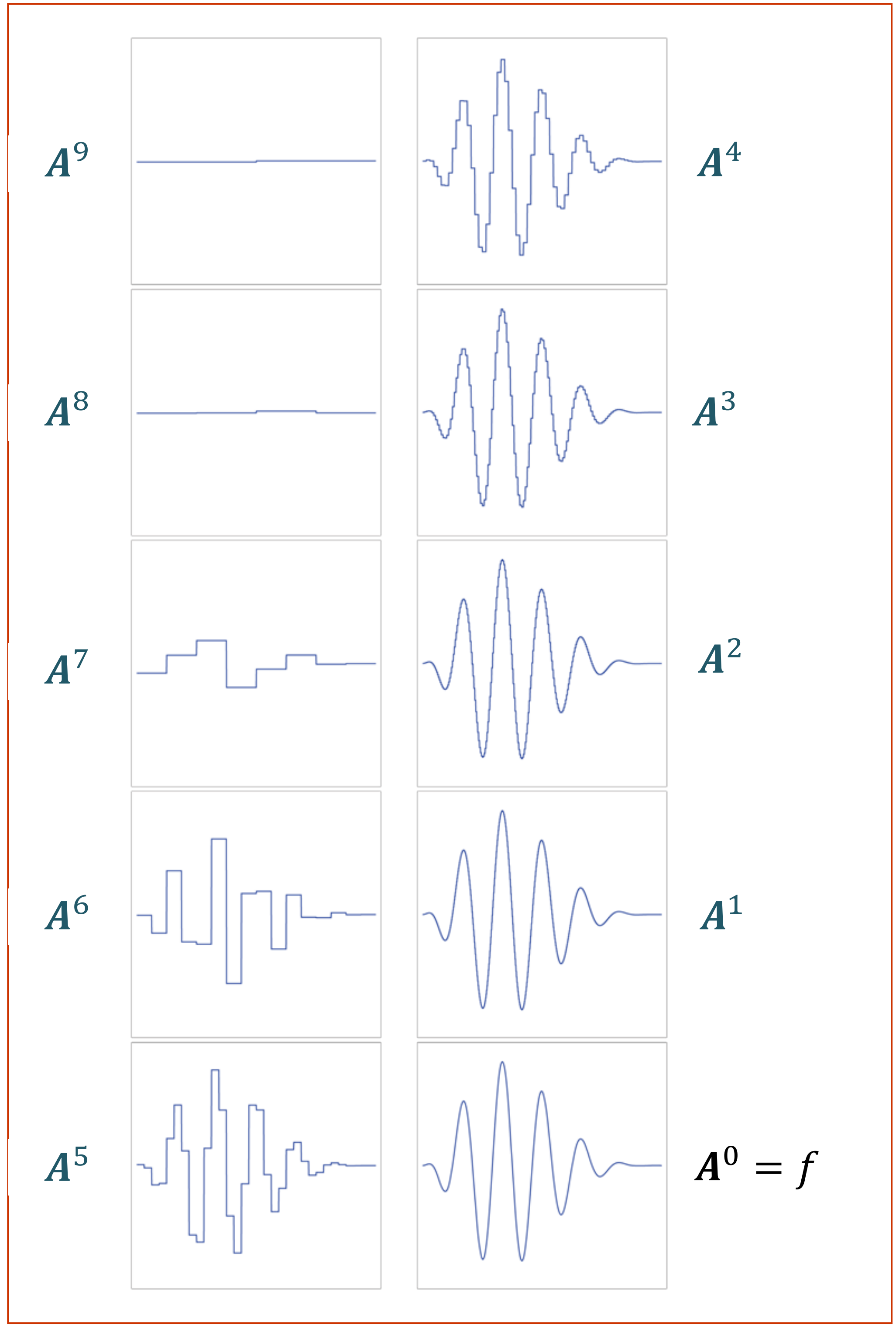
agrees with $\boldsymbol{f}$.

That `high_filter(f,r)` and `low_filter(f,r)` compute $\boldsymbol{A}^{r}(\boldsymbol{f})$ and $\boldsymbol{D}^{r}(\boldsymbol{f})$ is a straightforward observation based on the definitions and the actual coding of these functions.

On the right of next page we include images of the $\boldsymbol{A}^{r}(\boldsymbol{f})$, for $r = 1, \ldots, 9$.



A^{10}
D^{10}
D^9
D^8
D^7
D^6
D^5
D^4
D^3
D^2
D^1



A^9
A^4
A^8
A^3
A^7
A^2
A^6
A^1
A^5
$A^0 = f$

6.4 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. Provided by the formula on top of page 13, and its generalization to any level. We are going to see how this works in general (Haar and Daubechies wavelets) at the end of T7.

Notes

N1 (p. 6). We have:

$$a_j^2 = \left(\frac{f_{2j} + f_{2j+1}}{\sqrt{2}}\right)^2 = \frac{1}{2}\left(f_{2j}^2 + f_{2j+1}^2 + 2f_{2j}f_{2j+1}\right),$$

$$d_j^2 = \left(\frac{f_{2j} - f_{2j+1}}{\sqrt{2}}\right)^2 = \frac{1}{2}\left(f_{2j}^2 + f_{2j+1}^2 - 2f_{2j}f_{2j+1}\right)$$

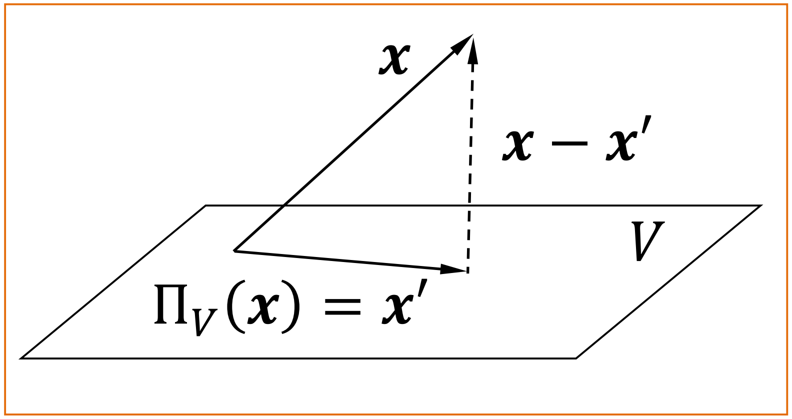
and hence $a_j^2 + d_j^2 = f_{2j}^2 + f_{2j+1}^2$. Summing for $j = 0, \dots, N/2 - 1$, the left hand side yields $\mathcal{E}(H_1(\boldsymbol{f}))$ and the right hand side $\mathcal{E}(\boldsymbol{f})$.

N2 (p. 6). H_r also preserves energy: $\mathcal{E}\big(H_r(\boldsymbol{f})\big) = \mathcal{E}(\boldsymbol{f})$. This follows immediately from the recursion definition: from

$$\mathcal{E}\big(H_r(f)\big) = \mathcal{E}(H_{r-1}(\boldsymbol{a}^1(\boldsymbol{f}))|\boldsymbol{d}^1(\boldsymbol{f})) = \mathcal{E}\big(H_{r-1}(\boldsymbol{a}^1(\boldsymbol{f}))\big) + \mathcal{E}(\boldsymbol{d}^1(\boldsymbol{f}))$$

and induction we get $\mathcal{E}\big(H_r(\boldsymbol{f})\big) = \mathcal{E}(\boldsymbol{a}^1(\boldsymbol{f})) + \mathcal{E}(\boldsymbol{d}^1(\boldsymbol{f})) = \mathcal{E}(\boldsymbol{f})$.

N3 (p. 13). If $V \subseteq \mathbb{R}^N$ is a linear subspace, and $\boldsymbol{x} \in \mathbb{R}^N$, there is a unique vector $\boldsymbol{x}' \in V$ such that $\boldsymbol{x} - \boldsymbol{x}'$ is orthogonal to V. This vector $\boldsymbol{x}'$ is called the *orthogonal projection* of $\boldsymbol{x}$ to V and is denoted $\Pi_V(\boldsymbol{x})$.



This vector can be calculated quite easily if we know an orthonormal basis $\mathbf{u}_1, \dots, \mathbf{u}_k$ of V. Indeed, in this case we have

$$\boldsymbol{x}' = (\boldsymbol{x} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\boldsymbol{x} \cdot \mathbf{u}_k)\mathbf{u}_k,$$

for the right hand side clearly belongs to V and $\boldsymbol{x} - \boldsymbol{x}'$ is orthogonal to all the $\mathbf{u}_j$ (for $\boldsymbol{x}' \cdot \mathbf{u}_j = \boldsymbol{x} \cdot \mathbf{u}_j$ because of the relations $\mathbf{u}_i \cdot \mathbf{u}_j = \delta_{ij}$ that hold for an orthonormal system).

The same argument can be adapted to show the uniqueness of $\boldsymbol{x}'$.

Remark. The computation of $\boldsymbol{x}'$ can also be carried out if we know any basis $\mathbf{v}_1, \dots, \mathbf{v}_k$ of V. In this case it is enough to impose that a vector

$$\boldsymbol{x}' = t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k \in V$$

satisfies the conditions $\boldsymbol{x}' \cdot \mathbf{v}_j = \boldsymbol{x} \cdot \mathbf{v}_j$ for $j = 1, \dots, k$. But these conditions are equivalent to the system of linear equations

$$(\mathbf{v}_1 \cdot \mathbf{v}_j)t_1 + \cdots + (\mathbf{v}_k \cdot \mathbf{v}_j)t_k = \boldsymbol{x} \cdot \mathbf{v}_j,\ \ j = 1, \dots, k$$

in the unknowns $t_1, \dots, t_k$. The solution of this system, which is unique, gives then the orthogonal projection of $\boldsymbol{x}$ to V.

N4 (p. 15). In general, if V and W are linear subspaces of $\mathbf{R}^n$, and $V \perp W$, then $\Pi_{V+W}(f) = \Pi_V(f) + \Pi_W(f)$. Indeed, the relations

$$\begin{aligned} f - \big(\Pi_V(f) + \Pi_W(f)\big) &= \big(f - \Pi_V(f)\big) - \Pi_W(f) \\ &= \big(f - \Pi_W(f)\big) - \Pi_V(f) \end{aligned}$$

show that the left-hand side is orthogonal to V (because $f - \Pi_V(f)$ and $\Pi_W(f)$ are orthogonal to V) and orthogonal to W (similar reason). Therefore it is orthogonal to $V + W$ and hence $\Pi_V(f) + \Pi_W(f)$ is the orthogonal projection of f on $V + W$.