Chapter 2

Wavelets

In this chapter we introduce wavelets and the discrete wavelet transform from the classical viewpoint, based on the concept multi-resolution analysis. The Fast Wavelet Transform (FWT) allows to calculate a wavelet transform in an efficient way. We will not go into all deep details, but we will limit ourselves to the topics that are important for the rest of this thesis.

2.1 What are Wavelets?

The main idea behind wavelet analysis is to decompose a signal f into a basis of functions Ψ_i:

$$f = \sum_i a_i \Psi_i.$$

To have an efficient representation of the signal f using only a few coefficients a_i, it is very important to use a suitable family of functions Ψ_i. The functions Ψ_i should match the features of the data we want to represent.

Real-world signals usually have the following features: they are both limited in time (time-limited)[1] and limited in frequency (band-limited). Time-limited signals can be represented efficiently using a basis of block functions (Dirac delta functions for infinitesimal small blocks). But block signals are not limited in frequency. Band-limited signals can be represented efficiently using a Fourier basis, but sines and cosines are not limited in time.

What we need is a compromise between the pure time-limited and band-limited basis functions, a compromise that combines the best of both worlds: wavelets ("small waves").

[1] Or space-limited in the case of images.

2.2 Why wavelets?

One of the main features of wavelets that are important for the applications that we shall consider is their good decorrelation:

- Wavelets are localized in both the space/time and scale/frequency domains. Hence they can easily detect local features in a signal.
- Wavelets are based on a multi-resolution analysis. A wavelet decomposition allows to analyze a signal at different resolution levels (scales).
- Wavelets are smooth, which can be characterized by their number of vanishing moments. A function defined on the interval $[a,b]$ has n vanishing moments if
$$\int_a^b f(x)x^i\,dx = 0,$$
for $i = 0,1,\ldots n-1$. The higher the number of vanishing moments, the better smooth signals can be approximated in a wavelet basis.

Furthermore there exist fast ($O(n)$) and stable (wavelets can be orthogonal or biorthogonal) algorithms to calculate the discrete wavelet transform and its inverse.

2.3 Wavelets and Multi-Resolution Analysis

Classically, discrete wavelets are introduced in the concept of a multi-resolution analysis. We introduce this concept first.

2.3.1 Multi-Resolution Analysis

Consider the space $\mathbf{L}^2$, the vector space of square integrable functions in $\mathbf{R}$:

$$\mathbf{L}^2 = \left\{ f : \int_{-\infty}^{+\infty} f^2(x)\,dx < \infty \right\}.$$

In a multi-resolution analysis [34] we decompose $\mathbf{L}^2$ in nested subspaces V_j

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots$$

such that the closure of their union is $\mathbf{L}^2$

$$\overline{\bigcup_{j=-\infty}^{+\infty} V_j} = \mathbf{L}^2,$$

and their intersection contains only the zero-function

$$\bigcap_{j=-\infty}^{+\infty} V_j = \{0\}.$$

In the dyadic case, i.e. when each subspace V_j is twice as large as V_{j-1}, a function $f(x)$ that belongs to one of these subspaces V_j has the following properties:

$$f(x) \in V_j \quad \Leftrightarrow \quad \text{dilation } f(2x) \in V_{j+1}, \tag{2.1}$$

$$f(x) \in V_0 \quad \Leftrightarrow \quad \text{translation } f(x+1) \in V_0. \tag{2.2}$$

If we can find a function $\varphi(x) \in V_0$ such that the set of functions consisting of $\varphi(x)$ and its integer translates

$$\{\varphi(x-k)\}_{k\in\mathbf{Z}}$$

form a basis for the space V_0, we call it a *scaling function* or *father function*. For the other subspaces V_j (with $j \neq 0$) we define:

$$\varphi_{j,k}(x) \equiv 2^{j/2}\varphi\left(2^j x - k\right).$$

2.3.2 Wavelet Functions

Because the subspaces V_j are nested:

$$V_j \subset V_{j+1},$$

we can decompose V_{j+1} in V_j and W_j, the orthogonal complement of V_j in V_{j+1}:

$$\begin{aligned} V_j \oplus W_j &= V_{j+1}, \\ W_j &\perp V_j. \end{aligned}$$

The direct sum of the subspaces W_j is equal to $\mathbf{L}^2$:

$$\overline{\bigcup_{j=-\infty}^{+\infty} V_j} = \overline{\bigoplus_{j=-\infty}^{+\infty} W_j} = \mathbf{L}^2.$$

This means that V_j is a "coarse-resolution" representation of V_{j+1}, while W_j carries the "high-resolution" difference information between V_{j+1} and V_j.

If we can find a function $\psi(x) \in W_0$ that obeys the translation property (2.2), i.e.

$$\psi(x) \in W_0 \Leftrightarrow \text{translation } \psi(x+1) \in W_0,$$

and such that the set of functions consisting of $\psi(x)$ and its integer translates

$$\{\psi(x-k)\}_{k\in\mathbf{Z}}$$

form a basis for the space W_0, we call it a *wavelet function* or *mother function*. For the other subspaces W_j (with $j \neq 0$) we define:

$$\psi_{j,k}(x) \equiv 2^{j/2}\,\psi\big(2^j x - k\big)\,.$$

2.3.3 The Fast Wavelet Transform (FWT)

Because both V_0 and W_0 are subspaces of V_1:

$$V_0 \subset V_1 \text{ and } W_0 \subset V_1,$$

we can express $\varphi(x)$ and $\psi(x)$ in terms of the basis functions of V_1:

$$\begin{aligned}
\varphi(x) &= 2\sum_k h_k\,\varphi(2x-k),\\
\psi(x) &= 2\sum_k g_k\,\varphi(2x-k).
\end{aligned}$$

Due to the multi-resolution analysis, these relations are also valid between V_{j+1}, V_j and W_j for arbitrary j.

We call the h_k and g_k the filter coefficients that uniquely define the scaling function $\varphi(x)$ and the wavelet $\psi(x)$.

Since $V_{j+1} = V_j \oplus W_j$, we can express a function $f(x)$ that is written in terms of the basis functions of V_{j+1} in terms of the basis functions of V_j and W_j also:

$$\begin{aligned}
f(x) &= \sum_k \lambda_{j+1,k}\,\varphi_{j+1,k}(x),\\
&= \sum_l \lambda_{j,l}\,\varphi_{j,l}(x) + \sum_l \gamma_{j,l}\,\psi_{j,l}(x).
\end{aligned}$$

with the transform coefficients $\lambda_{j,l}$ and $\gamma_{j,l}$ defined by:

$$\begin{aligned}
\lambda_{j,l} &= \sqrt{2}\sum_k h_{k-2l}\,\lambda_{j+1,k},\\
\gamma_{j,l} &= \sqrt{2}\sum_k g_{k-2l}\,\lambda_{j+1,k}.
\end{aligned}$$

This operation is known as the *Fast Wavelet Transform* (FWT). It is called "fast" because its complexity is $O(n)$, i.e. the amount of work is proportional to the signal length. The filter h_k is a low-pass filter, while g_k is a high-pass filter.

The inverse wavelet transform can be obtained in a similar way. This algorithm is known as the filter bank algorithm (Figure 2.1).

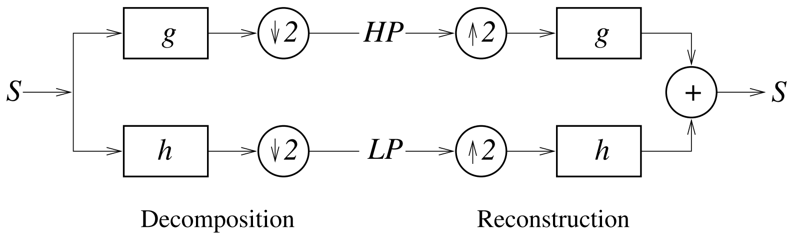

Figure 2.1: The filter bank algorithm for orthogonal wavelets: the signal S is filtered and downsampled to get a low-pass signal LP and a high-pass signal HP. It can be reconstructed by upsampling and filtering with the correct filters.

2.3.4 Orthogonal Wavelets

If the $\varphi_{j,k}(x)$ and $\psi_{j,k}(x)$ are orthonormal:

$$\begin{aligned} V_j &\perp W_j, \\ \langle \varphi_{j,l}, \varphi_{j,l'} \rangle &= \delta_{l-l'}, \\ \langle \psi_{j,l}, \psi_{j',l'} \rangle &= \delta_{j-j'}\, \delta_{l-l'}, \end{aligned}$$

then we can calculate the coefficients of the decomposition

$$f(x) = \sum_l \lambda_{j,l}\, \varphi_{j,l}(x) + \sum_l \gamma_{j,l}\, \psi_{j,l}(x)$$

by taking inner products with the scaling and wavelet functions:

$$\begin{aligned} \lambda_{j,l} &= \langle f, \varphi_{j,l} \rangle, \\ \gamma_{j,l} &= \langle f, \psi_{j,l} \rangle. \end{aligned}$$

The decomposition in the wavelet basis is guaranteed to be stable: if the function $f(x)$ is slightly changed, this will cause only a slight change in the coefficients $\lambda_{j,l}$ and $\gamma_{j,l}$. Parseval's identity holds:

$$\begin{aligned} \|f\|^2 &= \sum_k \lambda^2_{j+1,k}, && (2.3) \\ &= \sum_l \lambda^2_{j,l} + \sum_l \gamma^2_{j,l}. && (2.4) \end{aligned}$$

Examples of orthogonal wavelets are the family of orthogonal wavelets constructed by Daubechies [12]. The scaling function $\varphi(x)$ and wavelet function $\psi(x)$ for the member with 2 vanishing moments (also known as "D4" because the corresponding wavelet filter has 4 taps or filter coefficients) is shown in Figure 2.2.

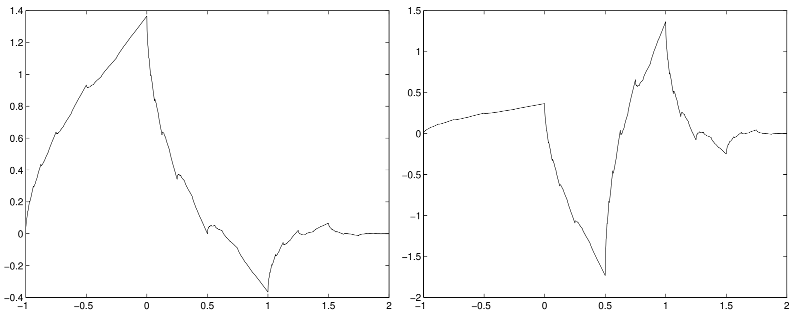

Figure 2.2: Orthogonal scaling function $\varphi(x)$ and wavelet $\psi(x)$ with 2 vanishing moments, as constructed by Daubechies.

2.3.5 Biorthogonal Wavelets

If we relax the orthogonality conditions to biorthogonality conditions we can obtain wavelets with some special properties that are not possible with orthogonal wavelets. In the biorthogonal case we have two multi-resolution analyses, a primal and a dual:

- primal: V_j, W_j, $\varphi_{j,k}$, $\psi_{j,k}$,
- dual: $\widetilde{V}_j$, $\widetilde{W}_j$, $\widetilde{\varphi}_{j,k}$, $\widetilde{\psi}_{j,k}$.

The biorthogonality conditions imply:

$$\begin{aligned} \widetilde{V}_j &\perp W_j, \\ V_j &\perp \widetilde{W}_j, \\ \langle \widetilde{\varphi}_{j,l}, \varphi_{j,l'} \rangle &= \delta_{l-l'}, \\ \langle \widetilde{\psi}_{j,l}, \psi_{j',l'} \rangle &= \delta_{j-j'}\delta_{l-l'}. \end{aligned}$$

Because the biorthogonal wavelets form a Riesz basis

$$A\|f\|^2 \leq \sum_l \lambda_{j,l}^2 + \sum_l \gamma_{j,l}^2 \leq B\|f\|^2, \tag{2.5}$$

with A and B positive constants[2], the decomposition in the biorthogonal wavelet basis is still stable and the coefficients $\lambda_{j,l}$ and $\gamma_{j,l}$ of the decomposition are now obtained by calculating inner products with the dual basis functions:

$$\begin{aligned} \lambda_{j,l} &= \langle f, \widetilde{\varphi}_{j,l} \rangle, \\ \gamma_{j,l} &= \langle f, \widetilde{\psi}_{j,l} \rangle. \end{aligned}$$

[2] In the orthogonal case we have $A = B = 1$, i.e. we have Parseval's identity.

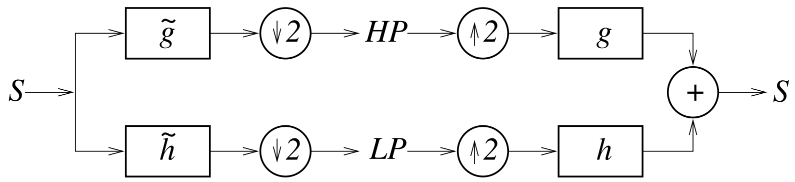

Figure 2.3: The filter bank algorithm for biorthogonal wavelets.

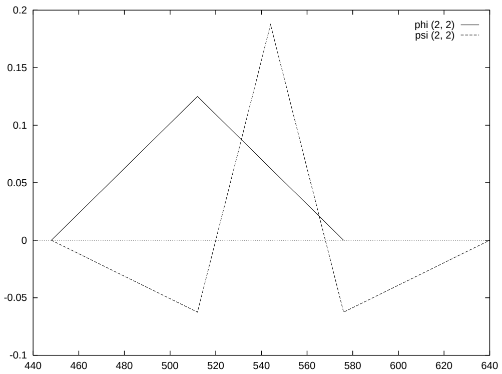

Figure 2.4: The primal scaling function $\varphi(x)$ and wavelet $\psi(x)$ for the Cohen-Daubechies-Feauveau $(2,2)$ biorthogonal wavelets. The dual scaling function and wavelet are not shown because of their irregular nature.

We can still use the filter bank algorithm if we use the dual filter pair $(\widetilde{h},\widetilde{g})$ (related to the dual multi-resolution analysis) for the decomposition and the primal filter pair (h,g) (related to the primal multi-resolution analysis) for the reconstruction (Figure 2.3).

Examples of biorthogonal wavelets are the family of biorthogonal wavelets constructed by Cohen, Daubechies and Feauveau [9]. The primal scaling and wavelet functions for the member with 2 vanishing moments for both the primal and the dual wavelet are shown in Figure 2.4.

2.3.6 The Wavelet Decomposition Tree

The subspaces V_j are nested, and each of them can be split in two subspaces V_{j-1} and W_{j-1}. If we start with the highest resolution subspace V_0, we perform the following decompositions:

$$V_0 \quad = \quad V_{-1} \oplus W_{-1},$$

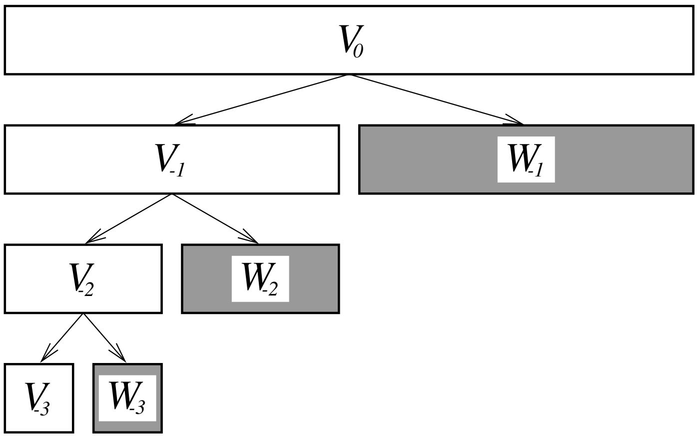

Figure 2.5: The wavelet decomposition tree.

$$
\begin{aligned}
&= V_{-2} \oplus W_{-2} \oplus W_{-1}, \\
&= V_{-3} \oplus W_{-3} \oplus W_{-2} \oplus W_{-1}, \\
&\vdots
\end{aligned}
$$

and we obtain a decomposition tree as in Figure 2.5.

Since the amount of data to transform is halved at each resolution level, the total complexity of the full wavelet decomposition is still $O(n)$. As a comparison, the Fast Fourier Transform (FFT) has a complexity of $O(n \log n)$.

2.3.7 Higher Dimensions

The above wavelets were defined on the real line **R**, i.e. on a one-dimensional domain. To create wavelets on higher dimensional domains, we can follow two approaches:

- We can perform the wavelet transform independently for each dimension. Because the wavelet transform can be written as a multiplication with a wavelet transform matrix and due to the associativity of the matrix product, the exact order is not important.

  In the two-dimensional case we get a decomposition as shown in Figure 2.6. The square variant is most used and in that case the basis functions are tensor products of the one-dimensional basis functions, i.e. after one transform step we have:

$$
\begin{array}{c|c}
\varphi(x) \otimes \varphi(x) & \varphi(x) \otimes \psi(x) \\
\hline
\psi(x) \otimes \varphi(x) & \psi(x) \otimes \psi(x)
\end{array}
$$

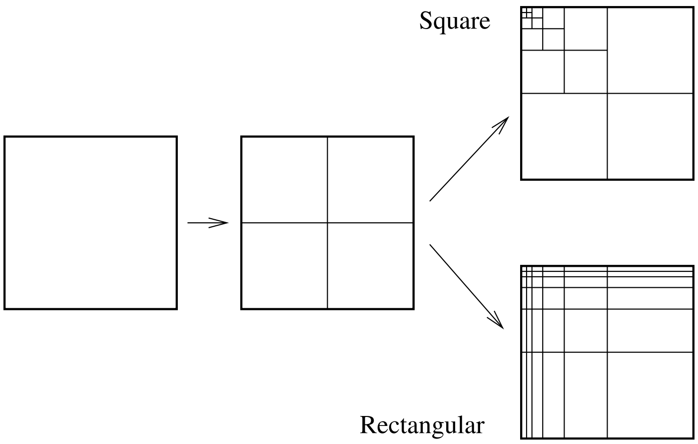

Figure 2.6: The two-dimensional wavelet transform: the square and the rectangular variant.

- An alternative is to use real multi-dimensional wavelets, which are less anisotropic with respect to directional information in the original data, e.g. Red-Black wavelets (see Section 7.2).

2.4 The Wavelet Transform and Translation Invariance

Due to the subsampling in the filter bank algorithm a wavelet transform is not translation invariant: if a signal is delayed or advanced its wavelet transform is not simply a delayed or advanced version of the wavelet transform of the original signal. Only if the delay or advance is a multiple of 2^n, with n the number of transform levels, the wavelet transform will be a delayed or advanced version of the original transformed signal. In two dimensions, it is even worse because this condition needs to be true in both the horizontal and vertical directions.

If each transform level is considered independently, one step of a wavelet transform is translation invariant if the translation is limited to an even number of pixels. For the one-dimensional case and one transform level, this is shown in Figure 2.7.

2.5 The Redundant Wavelet Transform

The wavelet transform can be converted into a translation invariant transform by removing the subsampling step. This is called the *redundant wavelet transform* (or

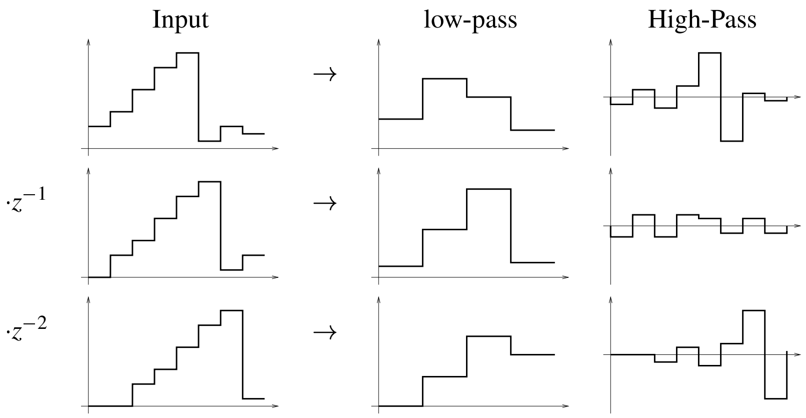

Figure 2.7: Translation invariance of the wavelet transform. When delaying the input signal by one time step, the resulting wavelet transform is completely different, while a delay of the input signal by two time steps results in a delayed version (by one time step at the coarser level) of the transformed original signal.

stationary, or non-decimated wavelet transform) [37]. It is important for applications that can suffer from translation variance, like e.g. feature detection and noise reduction.

In the redundant wavelet transform one gets rid of the decimation step, causing all subbands to have the same size as the size of the input data set. At each resolution level, the filters have to be upsampled[3] to keep a consistent multi-resolution analysis.

Because all subbands have the same size, the computational complexity is no longer $O(n)$, but $O(pn)$, with $p \leq \log n$ the number of transform levels. The redundant transform also requires more memory ($O(pn)$) to store the subbands. For the inverse transform, we have to calculate some kind of mean value due to the redundancy in the forward transform.

2.6 Wavelet Packets

In every step of a wavelet transform the low-frequency data of the signal is split in a low- and a high-frequency part, as can be seen in Figure 2.5. After decomposition the low-frequency part with the lowest index and the high-frequency parts will be stored.

In a wavelet packet transform, one also splits the high-frequency part in a "low"

[3] A filter is upsampled by putting zeros between the successive filter coefficients.

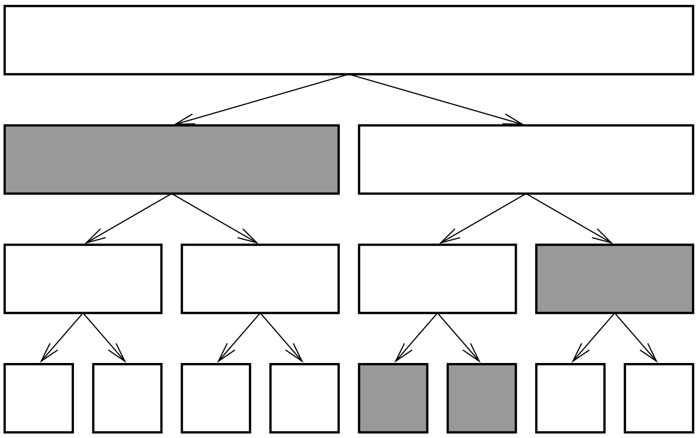

Figure 2.8: A wavelet packet decomposition tree.

and a "high" frequency part [10]. This produces a decomposition tree as shown in the Figures 2.8 (1D) and 2.9 (2D), and leads to a redundant tree of possible basis functions. A wavelet packet transform can be calculated with a complexity of $O(n\log n)$, compared to $O(n)$ for an ordinary wavelet transform.

The aim of wavelet packet analysis is to choose a "best basis", i.e. to find the set of functions that best decorrelates the input data and that is a subset of the basis functions in the decomposition tree [61].

The selection of the best basis is equivalent to the answer to the question "Do we best split this part into a low- and a high-frequency part or not?". We can make these decisions by starting at the bottom level of the tree in Figure 2.8 or 2.9 and comparing the "cost" of two (1D) respectively four (2D) neighboring "child" squares in this level with the "cost" of the "parent" square in the higher level. The "cost" is a measure for the number of bits we would need to store the data in the corresponding basis, i.e. a measure for the achieved decorrelation.

We retain the cheapest squares and proceed in the same way with the higher level, until we arrive at the top level. The basis functions that correspond to the retained squares form the best basis.

To calculate the "cost" of a matrix $v = [v_{ij}]$, several possibilities are available. We mention a few of them:

- **1-Norm**-based cost function: the cost of storing a matrix is proportional the the 1-norm of the matrix:

$$\text{cost} = \sum_{i,j} |v_{ij}|.$$

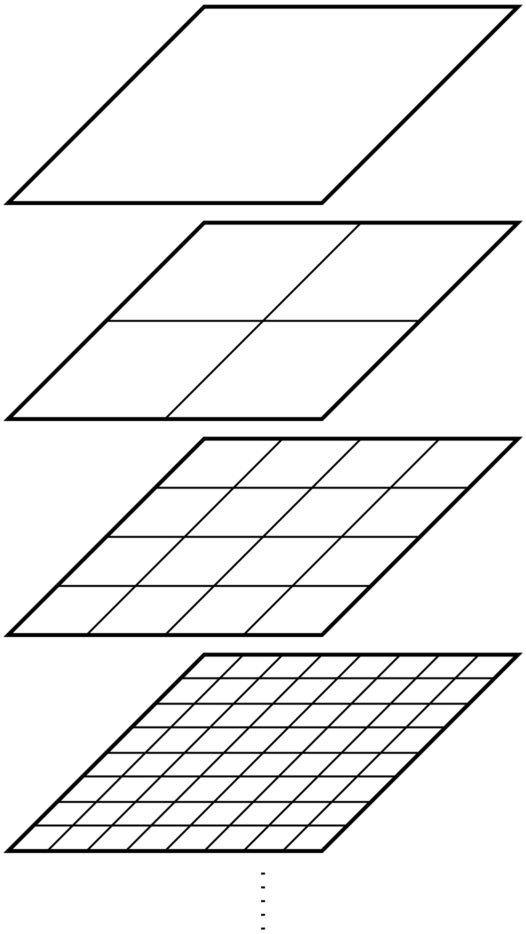

Figure 2.9: The two-dimensional wavelet packet decomposition tree.

- **First order Shannon Entropy**-based cost function: the first order entropy is the theoretical number of bits per symbol needed to encode a string of symbols by considering individual symbols [41]:

$$\text{cost} = -\sum_{i,j} v_{ij} \log |v_{ij}|.$$

- **Threshold**-based cost function:

$$\text{cost} = \#V_T \qquad \text{with } V_T = \left\{ v_{ij} \mid |v_{ij}| > T \right\},$$

with T a given threshold.

- **Logarithm of Energy**-based cost function:

$$\text{cost} = \sum_{i,j} \log(v_{ij})^2,$$

which is equivalent with

$$\text{cost} = \sum_{i,j} \log |v_{ij}|.$$

This is a variant of the first order Shannon entropy.