Chapter 2

Discrete Wavelet Transform, Lifting, and Image Coding: An Overview

This second chapter is an overview of the relevant issues required in the development of the Ph.D. thesis dissertation and a description of the state-of-the-art surrounding the lifting scheme. Section 2.1 starts with an introduction to wavelet theory and discrete wavelet transforms. Their connection to filter banks and subband coding is established. The employment of the wavelet transform in image compression is then justified. Section 2.2 introduces lifting scheme and describes its use and advantages in image coding. Adaptive lifting is the foundation stone of the nonlinear transforms proposed in this dissertation. It is briefly described in section 2.2.6 waiting for the more detailed analysis in chapter 4. Section 2.3 is a state-of-the-art review on lifting filters design and optimization techniques. Section 2.4 refers to some related nonlinear decompositions. Finally, section 2.5 is a review of another fundamental aspect for this work, the wavelet-based image coders.

2.1 Discrete Wavelet Transform

A main goal of wavelet research [Dau88, Coh92, Mal98, Bur98] is to create a set of expansion functions and transforms that give informative, efficient, and useful description of a function or signal. In applications working on discrete signals, one never has to directly deal with expansion functions. Discrete wavelet transform (DWT) is obtained simply by passing a discrete signal through a filter bank (figure 2.1). Wavelet theory can be understood and developed only by using such digital filters. This is the meeting point between wavelets and subband coding and the origin of two different nomenclatures for the same concepts. In fact, wavelet transform and subband coding are so closely connected that both terms are often used interchangeably.

Filter banks [Vai92, Vet95] are structures that allow a signal to be decomposed into subsignals through digital filters, typically at a lower sampling rate. Figure 2.1 shows a two-band filter bank.

It is formed by the analysis filters ($H_i(z)$, $i = 0, 1$) and the synthesis filters $G_i(z)$, for $i = 0, 1$. Filters $H_0(z)$ and $G_0(z)$ are low-pass filters. In an M-band filter bank, $H_i(z)$ and $G_i(z)$ for $0 < i < M - 1$ are band-pass filters, and $H_{M-1}(z)$ and $G_{M-1}(z)$ are high-pass filters. For a two-band filter bank, $M = 2$ and $H_1(z)$ and $G_1(z)$ are high-pass filters.

If the input signal can be recovered without errors from the subsignals, the filter bank is said to be a perfect reconstruction (PR) or a reversible filter bank. To enable PR, the analysis and synthesis filters have to satisfy a set of bilinear constraints. A polyphase characterization of perfect reconstruction is derived in §2.2.2.

Every finite impulse response (FIR) filter bank with an additional linear constraint on the low-pass filter is associated with a wavelet basis. The low-pass synthesis filter $G_0(z)$ is associated with the scaling function, and the remaining band-pass synthesis filters ($G_1(z)$ in the 2-band case) are each associated with the wavelet functions. Analysis low-pass filter $H_0(z)$ is associated with the so-called dual scaling function and analysis band-pass filters with the dual wavelet functions.

The notion of *channel* refers to each of the filter bank branches. A channel is the branch of the 1-D scaling coefficients (or approximation signal) and also each branch of the wavelet coefficients (or detail signals). The concept of *band* involves the concept of frequency representation, but it is commonly used in image processing to refer to each set of samples which are the output of the same 2-D filter. In 1-D linear processing both concepts are interchangeable.

The coefficients of the discrete wavelet expansion of a signal may be computed using a tree-structure where the filter bank is applied recursively along the low-pass filter channel. Every recurrence output is a different resolution level, which is a coarser scale representation of the original signal. In summary, a DWT is a dyadic tree-structured transform with a multi-resolution structure.

An alternative approach to the filter bank structure for computing DWT is the lifting scheme (LS). Lifting is more flexible and may be applied to more general problems. It is studied in detail in section 2.2.

2.1.1 Multi-resolution Analysis

Wavelet theory has a firm mathematical foundation on the multi-resolution analysis (MRA) axiomatic approach [Mal89]. This section starts with the definition of the multi-resolution hierarchy of nested sub-spaces which is then connected to real-valued wavelet basis functions and finally, wavelets are related to the filter bank structure.

A multi-resolution analysis on $L^2(\mathbb{R})$ is defined as a set of nested sub-spaces

$$\ldots \subseteq \mathcal{V}^{(2)} \subseteq \mathcal{V}^{(1)} \subseteq \mathcal{V}^{(0)} \subseteq \mathcal{V}^{(-1)} \subseteq \mathcal{V}^{(-2)} \subseteq \ldots$$

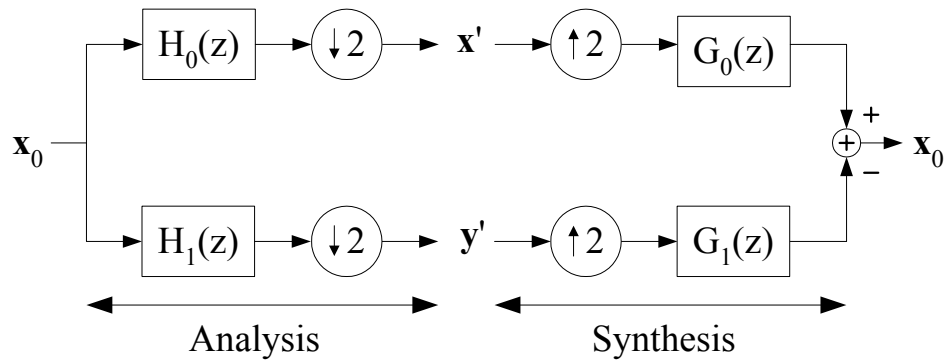


Figure 2.1: One-level two-band perfect reconstruction filter bank.

satisfying a set of five multi-resolution properties:

1. Upward completeness: $\overline{\bigcup_{j\in\mathbb{Z}} \mathcal{V}^{(j)}} = L^2(\mathbb{R})$.
2. Downward completeness: $\bigcap_{j\in\mathbb{Z}} \mathcal{V}^{(j)} = \{0\}$.
3. Shift invariance: if $f(t) \in \mathcal{V}^{(0)} \Leftrightarrow f(t-n) \in \mathcal{V}^{(0)}, \ \forall n \in \mathbb{Z}$.
4. Scale invariance: if $f(t) \in \mathcal{V}^{(0)} \Leftrightarrow f(2^{-j}t) \in \mathcal{V}^{(j)}, \ \forall j \in \mathbb{Z}$.
5. Basis existence: There exists $\varphi(t)$ such that the set of functions $\{\varphi(t-n)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $\mathcal{V}^{(0)}$.

The function $\varphi(t)$ is called the scaling function. The set of its integer translates $\{\varphi(t-n)\}_{n\in\mathbb{Z}}$ form a Riesz basis of $\mathcal{V}^{(0)}$. The dilated and normalized scaling function is denoted by

$$\varphi_{j,n}(t) = \sqrt{2^{-j}}\varphi(2^{-j}t - n).$$

The dilated set $\{\varphi_{j,n}(t)\}_{n\in\mathbb{Z}}$ is a Riesz basis of $\mathcal{V}^{(j)}$ for every j. The sub-space $\mathcal{W}^{(j)}$ is the orthogonal complement of $\mathcal{V}^{(j)}$ in $\mathcal{V}^{(j-1)}$, that is

$$\mathcal{V}^{(j-1)} = \mathcal{V}^{(j)} \oplus \mathcal{W}^{(j)}, \quad \forall j \in \mathbb{Z}.$$

A consequence of the MRA is the existence of the wavelet function $\psi(t)$. The set of the integer translates of the wavelet $\{\psi_{j,n}(t)\}_{n\in\mathbb{Z}}$ forms a Riesz basis for $\mathcal{W}^{(j)}$, being

$$\psi_{j,n}(t) = \sqrt{2^{-j}}\psi(2^{-j}t - n).$$

Also, the set $\{\psi_{j,n}(t)\}_{n,j\in\mathbb{Z}}$ forms an orthonormal wavelet basis for $L^2(\mathbb{R})$.

Since $\varphi(t), \psi(t) \in \mathcal{V}^{(-1)}$, they can be expressed as linear combination of the basis vectors $\{\varphi_{-1,n}(t)\}_{n\in\mathbb{Z}}$ of $\mathcal{V}^{(-1)}$, i.e., the scaling and wavelet function satisfy each one the so-called two-scale or refinement relation

$$\varphi(t) = \sqrt{2}\sum_n g_0[n]\varphi(2t-n), \tag{2.1}$$

$$\psi(t) = \sqrt{2}\sum_n g_1[n]\varphi(2t-n), \tag{2.2}$$

being g_0 and g_1 the synthesis low- and high-pass filter, respectively. Scaling and wavelet functions are related to the coefficients of a discrete filter through equations (2.1) and (2.2). This permits the interpretation of the MRA from a strict digital processing point of view despite of the underlying real-function setting. Let us illustrate this junction with an example. Assume $f(t) \in \mathcal{V}^{(j-1)}$. Then, the function $f(t)$ is completely described by the coefficients of the inner products of $f(t)$ with the scaling basis functions

$$x_{j-1}[n] =< f(t), \varphi_{j-1,n}(t) > \tag{2.3}$$

through the linear decomposition

$$f(t) = \sum_n x_{j-1}[n]\varphi_{j-1,n}(t). \tag{2.4}$$

Wavelets are necessary to describe the decomposition of $f(t)$ at the lower resolution scale j, because the *detail* is not available at scale j. The decomposition in the sub-spaces $\mathcal{V}^{(j)}$ and $\mathcal{W}^{(j)}$ may be also obtained with the corresponding continuous inner products as in (2.3),

$$f(t) = \sum_n x_j[n]\varphi_{j,n}(t) + \sum_n y_j[n]\psi_{j,n}(t).$$

However, the two-scale relation (2.1) applied on the scaling function in the inner product creates the decomposition working directly on the discrete domain samples $x[n]$ and $y[n]$:

$$\begin{aligned} x_j[n] &= < f(t), \varphi_{j,n}(t) > \\ &= \int f(t)\sqrt{2^{-j}}\varphi(2^{-j}t-n)dt \\ &= \sum_m g_0[m-2n]\int f(t)\sqrt{2^{-(j-1)}}\varphi(2^{-(j-1)}t-m)dt \\ &= \sum_m g_0[m-2n]x_{j-1}[m]. \end{aligned}$$

Therefore, the relation between the scaling coefficients of $f(t)$ at two consecutive resolution levels is

$$x_j[n] = \sum_m g_0[m-2n]x_{j-1}[m] = \bar{g}_0[2n] * x_{j-1}[n].$$

Equivalently, the corresponding relationship for the wavelet coefficients is

$$y_j[n] = \sum_m g_1[m-2n]x_{j-1}[m] = \bar{g}_1[2n] * x_{j-1}[n],$$

where $\bar{g}_0[n] = g_0[-n]$ and $\bar{g}_1[n] = g_1[-n]$.

These relations are precisely the low- and high-pass filtering of the decomposition coefficients at $\mathcal{V}^{(j-1)}$. By these means, a continuous decomposition of a function in $L^2(\mathbb{R})$ is related to a subband filtering of a $l^2(\mathbb{Z})$ sequence.

Every MRA gives rise to an orthonormal basis and an underlying filter bank that provides a vehicle for implementing the wavelet transform.

For image compression applications, there are two important features that transforms may have: linear phase and finite support. However, there is only one two-band orthonormal subband transform with these two properties, the so-called Haar wavelet. The corresponding filters to the Haar transform have two-taps and thus, this wavelet is hardly useful for many applications due to its short length. Linear phase and finite support are obtained at the same time by means of the biorthogonal systems.

2.1.2 Biorthogonal Wavelets

Biorthogonality is obtained by slightly relaxing the fifth MRA property, i.e., the existence of an integer-translate orthonormal basis of $\mathcal{V}^{(0)}$. The distinctive characteristic of biorthogonal systems is that the decomposition bases at the analysis and synthesis side are different. Besides the scaling and wavelet functions, biorthogonal systems have a dual scaling function and a dual wavelet function that also generate a multi-resolution analysis. The dual scaling function is denoted by $\tilde{\varphi}(t)$ and the dual wavelet by $\tilde{\psi}(t)$. Functions $\tilde{\varphi}(t)$ and $\tilde{\psi}(t)$ satisfy the two-scale relations (2.1) and (2.2) with the coefficients $h_0[n]$ and $h_1[n]$, respectively. The dual functions are biorthogonal to the primal functions $\varphi(t)$ and $\psi(t)$ in the sense that

$$< \tilde{\varphi}(t), \psi(t-n) > = < \tilde{\psi}(t), \varphi(t-n) > = 0,$$

$$< \tilde{\varphi}(t), \varphi(t-n) > = < \tilde{\psi}(t), \psi(t-n) > = \delta[n].$$

A biorthogonal wavelet system extends the orthogonal ones, it is more flexible, and generally easier to design. The advantages of a biorthogonal system w.r.t. an orthogonal one are described in the list below.

- Orthogonal system filters must be of the same length, and the length must be even. This restriction is relaxed for biorthogonal systems.

- Biorthogonal wavelets may be symmetric and thus, filter frequency response may be linear phase. This is the main reason of the inclusion of biorthogonal systems in the JPEG2000 standard and their widespread use. On the other hand, as mentioned above, there are no two-band orthogonal transforms having FIR linear phase with more than 2 non-zero coefficients in any filter.
- In a biorthogonal system, analysis and synthesis filters may be switched and the resulting system is still sound. Therefore, the appropriate arrangement may be chosen for the application at hand. For image compression, it has been observed that the use of the smoother filter in the reconstruction of the coded image leads to better visual appearance.

Biorthogonal systems also have some disadvantages:

- Parseval theorem no longer holds for biorthogonal wavelets. This means that the norm of the coefficients is not the same as the norm of the functions being spanned. Many design efforts have been devoted to making near orthogonal systems.
- White gaussian noise remains white after an orthogonal transform, but becomes correlated after a non-orthogonal transform. This may be considered when biorthogonal systems are employed in estimation or detection applications.

2.1.2.1 Vanishing Moments

Vanishing moments is a core concept of wavelet theory. In fact, the number of vanishing moments was a more important factor than spectral considerations for the choice of the wavelet transforms for the JPEG2000 standard.

To vanish the n^{th} moment means that given a polynomial input up to degree n the filter output is zero. A wavelet has N vanishing moments if

$$\int_{-\infty}^{\infty} t^n \psi(t) dt = 0, \quad \text{for } 0 \leq n < N.$$

The same definition applies for the dual wavelet to have $\tilde{N}$ vanishing moments. N and $\tilde{N}$ are also the multiplicity of the origin as a root of the Fourier transform of the synthesis and analysis high-pass filter, respectively. Also, it is the multiplicity of the regularity factor $(1+z^{-1})$ in $H_1(z)$ and $G_1(z)$ (which are the z-transform of the filters $h_1[n]$ and $g_1[n]$).

2.1.3 Discrete Wavelet Transform in Image Compression

For images, filter banks and lifting filters are usually developed for the 1-D case and then they are extended to the separable 2-D case by a succession of a vertical and an horizontal

1-D filtering. This structure leads to a 4 band per resolution level decomposition (figure 2.2). The decomposition may be iterated on the LL band (the vertically and horizontally low-pass filtered band). The bands with high frequency components (the HL, LH, and HH bands) are not recursively filtered.

One of the wavelet transform advantages concerning data compression is that it tends to compact the input signal energy into a relatively small number of wavelet coefficients. Lowest resolution band coefficients have high energy. High frequency bands represent the majority of the transformed samples. Most of high frequency band coefficients are zero or have low energy. The exceptions are samples lying near strong edges relative to the band orientation. For instance, a vertical edge produces significant wavelet coefficients in the HL band, which is obtained by applying an horizontal high-pass filter and thus, the edge high frequency components are not eliminated. At the same time, the LH band is not affected by such an edge. Equivalent statements are valid for the other bands: in general, horizontal and diagonal edges produce significant coefficients in the HL and the HH bands, respectively. This phenomenon is illustrated in figure 2.3. The 256x256 image *crosses*[1] (figure 2.3a) is decomposed with the Haar wavelet (figure 2.3b shows the transformed image). The Haar transform analysis filters are $H_0(z) = 2^{-1/2}(1+z)$ and $H_1(z) = 2^{-1/2}(1-z^{-1})$. The distribution of the high energy coefficients in the different bands (the darker and lighter pixels) may be observed in the transformed image.

Therefore, the few high frequency samples with significant energy in a band are generally clustered together. Besides, there is relevant inter-band statistical dependance. Specifically, high energy coefficients often cluster around the same location across several scales. These properties are studied in detail in [Liu01]. They are exploited by the wavelet-based image coders to achieve excellent compression results. Section §2.5 outlines the strategies followed by these kinds of image coders.

2.2 Lifting Scheme

This section introduces the original lifting scheme due to Sweldens, its properties, and applications §2.2.1. The polyphase domain analysis in §2.2.2 and §2.2.3 provides the mathematical reason of the LS structural perfect reconstruction property and the connection between filter banks and LS. The section takes a more state-of-the-art review flavor in §2.2.4, which describes the use of wavelets and lifting in the image coding standard JPEG2000, a fact that has prompted the interest in LS. Space-varying lifting is described in §2.2.5 and §2.2.6 is an introduction to adaptive LS.

[1] Images employed in the experiments throughout this Ph.D. thesis are described in appendix A.

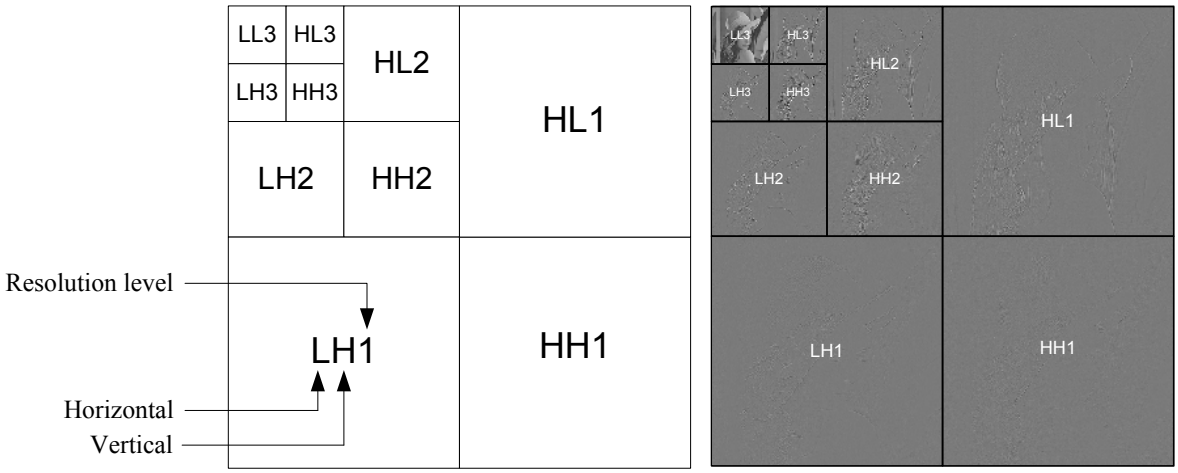
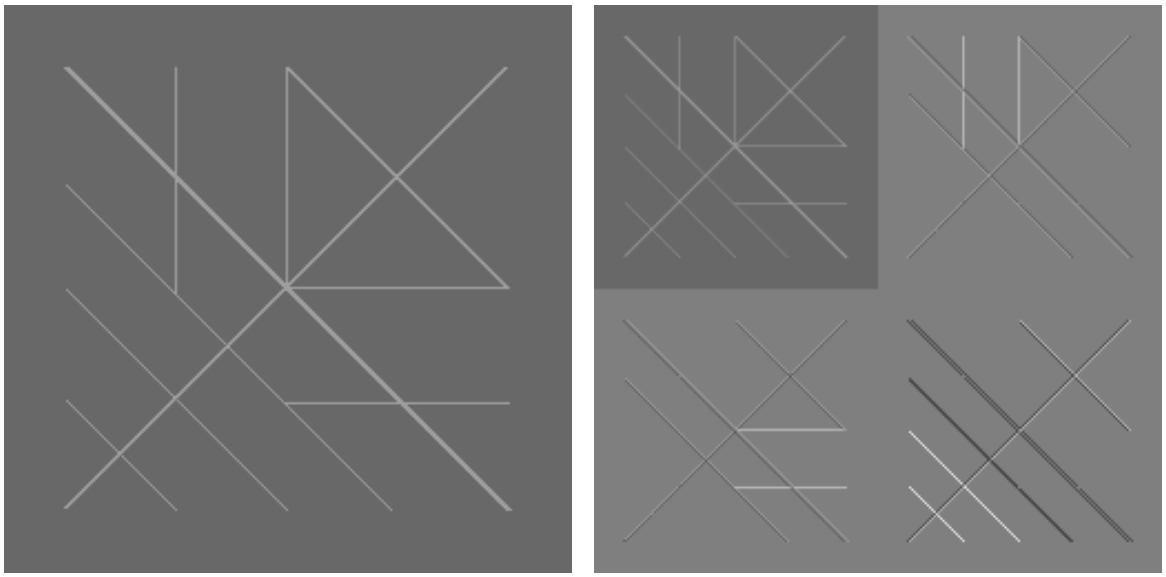


Figure 2.2: On the left, notation for a 3-level 2-D separable wavelet decomposition of an image. Every resolution level have 4 bands: LL, HL, LH, and HH, where L stands for low-pass filtered and H for high-pass filtered. On the right, a decomposition example with the image Lenna.



(a) (b)

Figure 2.3: (a) Image "crosses" and **(b)** its 2-level Haar wavelet transform.

2.2.1 Classical Lifting

The lifting scheme (figure 2.4) formally introduced in [Swe96, Swe97] by W. Sweldens is a well-known method to create biorthogonal wavelet filters from other ones. The scheme comprises the following parts:

(a) Input data $\mathbf{x}_0$.

(b) Polyphase decomposition (or lazy wavelet transform, LWT) of $\mathbf{x}_0$ into two subsignals:

- An approximation signal $\mathbf{x}$ formed by the even samples of $\mathbf{x}_0$.
- A detail signal $\mathbf{y}$ formed by the odd samples of $\mathbf{x}_0$.

(c) Lifting steps:

- Prediction P (or dual) lifting step that predicts the detail signal samples using the approximation samples $\mathbf{x}$,

$$y'[n] = y[n] - P(\mathbf{x}[n]). \tag{2.5}$$

- Update U (or primal) lifting step that updates the approximation signal with the detail samples $\mathbf{y}'$,

$$x'[n] = x[n] + U(\mathbf{y}'[n]). \tag{2.6}$$

(d) Output data: the transform coefficients $\mathbf{x}'$ and $\mathbf{y}'$.

Possibly, there are scaling factors K_1 and K_2 at the end of each channel in order to normalize the transform coefficients $\mathbf{x}'$ and $\mathbf{y}'$, respectively.

The inversion of the scheme is straightforward. The same prediction lifting step (PLS) and update lifting step (ULS) are employed and only the sign of the addition is changed. Finally, subsignals are merged into the higher rate signal to recover the original data $\mathbf{x}_0$.

Lifting steps improve the initial lazy wavelet properties. Alternatively, input data may be any other wavelet transform $(\mathbf{x}, \mathbf{y})$ with some properties to improve. The transform may be a multi-channel or M-band decomposition, leading to several detail subsignals $\mathbf{y}'_i$, for $i = 1, \ldots, M-1$. Also, several steps (a prediction followed by an update step or vice versa) may be concatenated in order to reach the desired properties for the wavelet basis.

The prediction and update operators may be a linear combination of $\mathbf{x}$ and $\mathbf{y}$, respectively, or any nonlinear operation, since by construction the LS is always reversible.

Given a wavelet decomposition that maps the input signal $\mathbf{x}_0$ to the output approximate subsignal $\mathbf{x}'$ and the output detail subsignal $\mathbf{y}'$, a multi-resolution decomposition of $\mathbf{x}_0$ is built

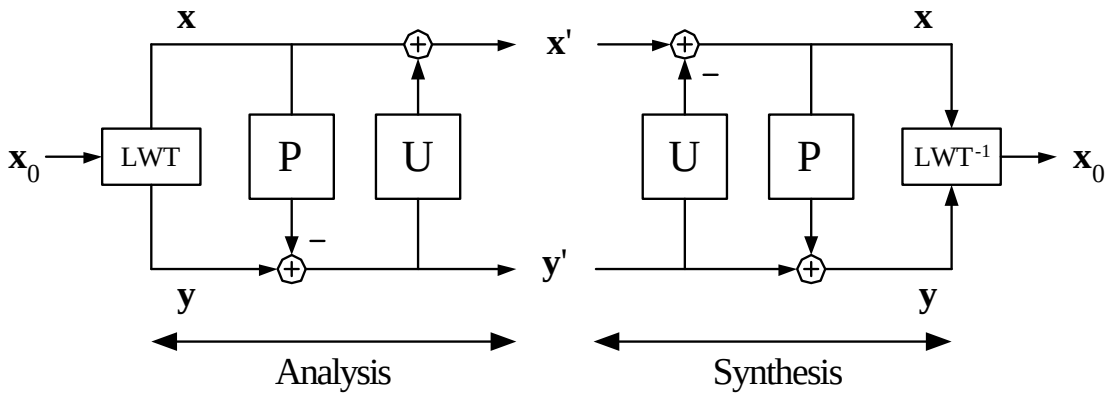


Figure 2.4: Classical lifting scheme.

by the concatenation of the lifting decomposition blocks on the approximate subsignal, exactly like the recursive filter bank tree-structure does. Subsignals $\mathbf{x}''$ and $\mathbf{y}''$ are obtained plugging $\mathbf{x}'$ into another decomposition block formed by the lifting steps. The process may be repeated on $\mathbf{x}''$, and so on. Thus, the concatenation of K such blocks yields a K-level wavelet decomposition. A 1-D signal decomposed with a 2-band K-level DWT results in the subsignals

$$\mathbf{x}_0 \rightarrow (\mathbf{x}', \mathbf{y}') \rightarrow (\mathbf{x}'', \mathbf{y}'', \mathbf{y}') \rightarrow \ldots \rightarrow (\mathbf{x}^{(K)}, \mathbf{y}^{(K)}, \ldots, \mathbf{y}'', \mathbf{y}').$$

LS is known as the second generation wavelet because it has many advantages with respect to the classical method for the construction of wavelets based on the Fourier transform. These advantages are itemized in the following list:

1. *Inverse existence.* Every lifting step is reversible by structure so there always exists the inverse wavelet transform constructed with lifting steps.

2. *Critical down-sampling assured.* Initial wavelet is modified with existing samples so no additional information (redundancy) is added.

3. *Transform direct spatial interpretation.* When constructing a new transform, the lifting structure permits us to consider how the output coefficients of a lifting filter affect the channel being filtered in a quite "visual" manner and without any spectral consideration. The reason is that the lifting structure itself performs a biorthogonal wavelet decomposition despite of the prediction and update filters. The alternative is to first construct the wavelet through Fourier methods and only then observe how it exactly acts on the signal in the spatial (or time) domain.

4. *Computational cost reduction.* Asymptotically, lifting reduces to one-half the computational cost of the standard filter implementation.

5. *Memory savings.* In-place lifting computation avoids auxiliary memory requirements since lifting outputs from one channel may be saved directly in the other channel. Such implementation considerations are explained in [Tau02a].

6. *FIR decomposition.* Daubechies et al. demonstrated in [Dau98] that every wavelet transform with FIR filters can be decomposed into a finite number of lifting steps.

7. *Boundary extensions.* Lifting significantly reduces the casuistry in the boundary treatment w.r.t. filter banks. Also, lifting implementation does not require explicit signal extension at boundaries.

Furthermore, LS has many applications in the wavelets field:

- New wavelets construction.
- Existing wavelets improvement.
- Wavelet construction on irregular grids. For instance, non-separable lifting on quincunx sampled images have been developed [Gou00].
- Easy replacement of linear filters by nonlinear/morphological filters (e.g. [Hei00]).
- Design of space-varying and adaptive decompositions (e.g. [Ger00]).

LS flexibility is profited for video coding applications as outlined in §2.3.3. Lately it has found applications like the coding of multi-view images [Ana05], the coding of 3-D mesh data [Hon05], or even the construction a multiscale manifold representation from noisy point clouds [Cho05]. Other signal processing applications as reversible image rotation and lossless data embedding in images [Kam05] have also appeared.

This Ph.D. dissertation is centered on the more "classical" applications of LS in the context of image coding: the analysis and improvement of existing wavelets as in §3.4.3 and §3.3.2, respectively, the construction of new linear transforms §3.3.3, wavelet construction on irregular grids §3.4.4, the design of adaptive decompositions §4.3, and the creation of new nonlinear decompositions in chapter 5.

2.2.2 Polyphase Characterization of Perfect Reconstruction

The polyphase domain analysis of filter banks permits us to obtain PR conditions on the filters and naturally leads to the lifting scheme as a built-in PR decomposition.

The lazy wavelet or polyphase transform separates odd and even samples. The two polyphase components $x_e[n] = x[2n]$ and $x_o[n] = x[2n+1]$ are obtained through a delay and two downsamplings (figure 2.5). The z-transform of $x[n]$ can be expressed as the z-transform of the

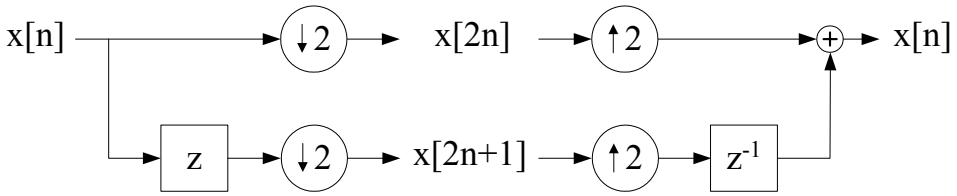


Figure 2.5: Polyphase decomposition or lazy wavelet transform, followed by the inverse process.

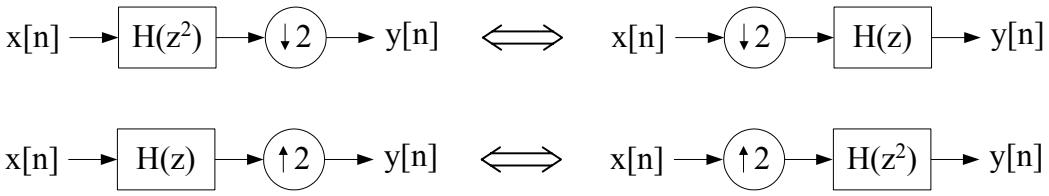


Figure 2.6: Noble multi-rate identities, **(top)** First Noble identity for the down-sampling and **(bottom)** Second Noble identity for the up-sampling

polyphase components: $X(z) = X_e(z^2) + z^{-1}X_o(z^2)$. The filters are also split in the same way (with the convenient delay for the odd samples):

$$\begin{aligned} H_0(z) &= H_{0e}(z^2) + zH_{0o}(z^2), \\ H_1(z) &= H_{1e}(z^2) + zH_{1o}(z^2), \\ G_0(z) &= G_{0e}(z^2) + z^{-1}G_{0o}(z^2), \\ G_1(z) &= G_{1e}(z^2) + z^{-1}G_{1o}(z^2). \end{aligned}$$

Filter $H_e(z^2)$ (resp. $H_o(z^2)$) contains the even (resp. odd) samples of the impulse response of $H(z)$ interpolated with zeros, one zero after each coefficient. Such zero-padded filters allow the interchanging of down-sampling and filtering, thus reaching an equivalent structure. This is known as the first Noble multi-rate identity. Figure 2.6 depicts the two Noble identities. The first identity is applied to the even and odd channels. The process is shown in the analysis part of figures 2.7 and 2.8. By these means, down-sampling is performed first and it is followed by the filter $H_e(z)$ (resp. $H_o(z)$). The structure in figure 2.8 is computationally more efficient but mathematically equivalent to the structures in figures 2.1 and 2.7.

The transformed coefficients $\mathbf{x}'$ and $\mathbf{y}'$ may be expressed as function of the polyphase components of the input signal and filters:

$$\begin{aligned} X'(z) &= H_{0e}(z)X_e(z) + H_{0o}(z)X_o(z), \\ Y'(z) &= H_{1e}(z)X_e(z) + H_{1o}(z)X_o(z). \end{aligned}$$

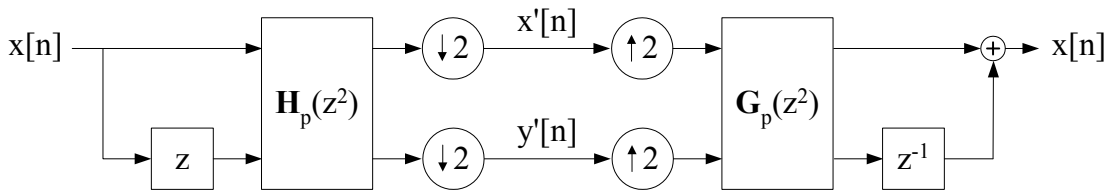


Figure 2.7: Polyphase structure of the filter bank before applying the first Noble identity.

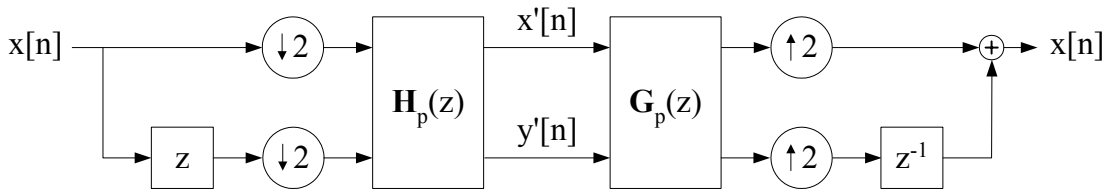


Figure 2.8: Polyphase characterization of a filter bank.

Using the matrix notation the two previous equations are expressed as

$$\begin{pmatrix} X'(z) \\ Y'(z) \end{pmatrix} = \underbrace{\begin{pmatrix} H_{0e}(z) & H_{0o}(z) \\ H_{1e}(z) & H_{1o}(z) \end{pmatrix}}_{\mathbf{H}_p(z)} \begin{pmatrix} X_e(z) \\ X_o(z) \end{pmatrix},$$

where $\mathbf{H}_p(z)$ is the analysis polyphase matrix of the filter bank. Filters $H_0(z)$ and $H_1(z)$ are given by

$$\begin{pmatrix} H_0(z) \\ H_1(z) \end{pmatrix} = \mathbf{H}_p(z^2) \begin{pmatrix} 1 \\ z \end{pmatrix}.$$

In general, an M-channel filter bank structure is compactly represented by an MxM component polyphase matrix, in which the element $[\mathbf{H}_p(z)]_{i,j}$ is the j^{th} polyphase component of the i^{th} filter.

The synthesis part of the filter bank may also be described with a polyphase matrix. A reconstructed signal $\hat{X}(z)$ is obtained,

$$\hat{X}(z) = G_0(z)X'(z^2) + G_1(z)Y'(z^2).$$

The expression of the synthesis filters with their polyphase components is introduced,

$$\hat{X}(z) = (G_{0e}(z^2) + z^{-1}G_{0o}(z^2))X'(z^2) + (G_{0e}(z^2) + z^{-1}G_{0o}(z^2))Y'(z^2). \tag{2.7}$$

The second multi-rate identity (figure 2.6) is applied to (2.7), thus obtaining the equivalent

synthesis structure in figure 2.8. Equation (2.7) written in a matrix form is

$$\hat{X}(z) = \begin{pmatrix} 1 & z^{-1} \end{pmatrix} \underbrace{\begin{pmatrix} G_{0e}(z^2) & G_{1e}(z^2) \\ G_{0o}(z^2) & G_{1o}(z^2) \end{pmatrix}}_{\mathbf{G}_p(z^2)} \begin{pmatrix} X'(z^2) \\ Y'(z^2) \end{pmatrix},$$

where $\mathbf{G}_p(z)$ is the synthesis polyphase matrix of the filter bank. Then, the output signal is related to input polyphase components through the filter bank polyphase matrices:

$$\hat{X}(z) = \begin{pmatrix} 1 & z^{-1} \end{pmatrix} \mathbf{G}_p(z^2)\mathbf{H}_p(z^2) \begin{pmatrix} X_e(z^2) \\ X_o(z^2) \end{pmatrix}. \tag{2.8}$$

PR is attained when the output signal is a delayed and scaled version of the input signal. By inspection of (2.8), it is observed that if the condition

$$\mathbf{G}_p(z)\mathbf{H}_p(z) = \mathbf{I} \tag{2.9}$$

holds, then the reconstructed signal is $X(z)$, since the structure in figure 2.8 reduces to the LWT of figure 2.5 and thus, PR is attained, i.e.,

$$\hat{X}(z) = X_e(z^2) + z^{-1}X_o(z^2) = X(z).$$

The determinants of the polyphase matrices of a PR FB are related. The polyphase analysis and synthesis matrices of any two channel FIR PR filter bank must satisfy

$$\begin{aligned} \det(\mathbf{H}_p(z)) &= \alpha z^{-k}, \\ \det(\mathbf{G}_p(z)) &= \alpha^{-1} z^{k}, \end{aligned} \tag{2.10}$$

for some arbitrary delay $k \in \mathbb{Z}$ and non-zero $\alpha \in \mathbb{R}$. The proof is straightforward. Since (2.9) must hold, the product of $\det(\mathbf{H}_p(z))$ and $\det(\mathbf{G}_p(z))$ must be equal to 1. Also, both determinants are finite polynomials in z because filters are FIR. Consequently, the determinants must be monomials of the form given by (2.10). The determinants are 1 with a proper filter scaling.

2.2.3 Polyphase Characterization of Lifting Scheme

Consider a filter bank from figure 2.1 satisfying the PR property. A "new" high-pass filter $H_1^{new}(z)$ is obtained by adding a prediction or dual lifting step after the down-sampling. The new high-pass filter is related to the "old" filter $H_1(z)$ by

$$H_1^{new}(z) = H_1(z) - H_0(z)P(z^2).$$

The high-pass channel is being lifted (improved) with the help of the low-pass channel. The high-pass filter is improved with an appropriate choice of P. The polyphase components of the

new filter are

$$\begin{aligned} H_{1e}^{new}(z) &= H_{1e}(z) - H_{0e}(z)P(z), \\ H_{1o}^{new}(z) &= H_{1o}(z) - H_{0o}(z)P(z). \end{aligned}$$

Therefore, the new and old polyphase matrices are related by

$$\mathbf{H}_p^{new}(z) = \begin{pmatrix} 1 & 0 \\ -P(z) & 1 \end{pmatrix} \mathbf{H}_p(z).$$

An inverse PLS should performed in the synthesis part in order to preserve the PR property of the new filter bank. The inverse is trivial, since it is the same matrix with the sign of $P(z)$ changed. The new synthesis polyphase matrix is

$$\mathbf{G}_p^{new}(z) = \mathbf{G}_p(z) \begin{pmatrix} 1 & 0 \\ P(z) & 1 \end{pmatrix}.$$

PR is preserved simply because

$$\mathbf{G}_p^{new}(z)\mathbf{H}_p^{new}(z) = \mathbf{G}_p(z) \begin{pmatrix} 1 & 0 \\ P(z) & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -P(z) & 1 \end{pmatrix} \mathbf{H}_p(z) = \mathbf{G}_p(z)\mathbf{H}_p(z) = \mathbf{I}.$$

A similar procedure is done in order to lift the properties of the low-pass channel: the update or primal lifting step. In the polyphase domain, the ULS is an upper triangular matrix with a positive $U(z)$ at analysis and with a negative sign at synthesis:

$$\begin{aligned} \mathbf{H}_p^{new}(z) &= \begin{pmatrix} 1 & U(z) \\ 0 & 1 \end{pmatrix} \mathbf{H}_p(z), \\ \mathbf{G}_p^{new}(z) &= \mathbf{G}_p(z) \begin{pmatrix} 1 & -U(z) \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

As said above, several lifting steps may be concatenated. The most frequent choice for the filters to begin the LS is the LWT, i.e., the polyphase decomposition with $\mathbf{H}_p(z) = \mathbf{I}$ and $\mathbf{G}_p(z) = \mathbf{I}$. In fact, it was shown in [Dau98] that any FIR wavelet may be decomposed into a finite number of lifting steps, an initial LWT, and a magnitude scaling:

$$\begin{aligned} \mathbf{H}_p(z) &= \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix} \left\{ \prod_{i=1}^{m} \begin{pmatrix} 1 & U_i(z) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -P_i(z) & 1 \end{pmatrix} \right\}, \\ \mathbf{G}_p(z) &= \left\{ \prod_{i=m}^{1} \begin{pmatrix} 1 & 0 \\ P_i(z) & 1 \end{pmatrix} \begin{pmatrix} 1 & -U_i(z) \\ 0 & 1 \end{pmatrix} \right\} \begin{pmatrix} \frac{1}{K_1} & 0 \\ 0 & \frac{1}{K_2} \end{pmatrix}. \end{aligned} \tag{2.11}$$

This decomposition corresponds to the polyphase matrix factorization into elementary matrices. A lifting step becomes an elementary matrix, that is, a triangular matrix (lower or upper) with all diagonal entries equal to one. A well known result in matrix algebra states that any matrix with polynomial entries and determinant one can be factored into such elementary matrices.

The crucial point is that the determinant of a triangular matrix with all diagonal elements equal to one is one and so, its inverse always exists independently of the value or form of the non-diagonal matrix entry. The inverse is simply obtained by changing the sign of the non-diagonal element. Therefore, despite of the lifting step added to the filter bank, the resulting analysis polyphase matrix determinant is not changed and an inverse step may be added to the synthesis polyphase matrix.

The proof of the factorization existence in [Dau98] relies on the Euclidean algorithm that can be used because the z-transform of a filter is a Laurent polynomial. Concretely, the filter z-transform $H(z) = \sum_{i=k_0}^{k_1} h_i z^{-i}$ is a Laurent polynomial in z of order $|H| = k_1 - k_0$, being z^{-k} a polynomial of order 0. The set of all Laurent polynomials with real coefficients has a commutative ring structure. In general, exact division within a ring is not possible. However, division with remainder is possible for Laurent polynomials. Polyphase matrix entries are Laurent polynomials, so they also form a ring structure. If the determinant of such a matrix is a monomial, then the matrix is invertible (2.10). Thus, the Euclidean algorithm can be applied for the decomposition of the polyphase matrix.

However, the long division of Laurent polynomials is not necessarily unique, so various decompositions are possible for the same filter bank, i.e., a wavelet transform has several lifting versions. The selection of the polyphase matrix factorization has practical relevance because the finite precision representation of the lifting filter and coefficients has effects on performance. Quantization deviates the lifting implementation from the theoretical transform properties.

Various factorization criteria has been envisaged: the minimum number of lifting steps, the ratio between the lifting coefficients of maximum and minimum magnitude [Ada98], the closeness of the scaling factor K to one [Cal98], or the minimum nonlinear iterated graphic function [Gra02], which is a figure of the difference between the wavelet function and the quantized LS impulse response. The latter criterion seems to perform the best.

2.2.4 Lifting in JPEG2000-LS

The JPEG2000 [ISO00] is an ISO/ITU-T image compression standard. JPEG2000 is a wavelet-based image coder largely derived from the EBCOT coder (§2.5.3) that achieves excellent lossy and lossless results. It has several interesting features, like the support of different types of scalability (§2.5.1). The wavelet transform in JPEG2000 is computed via lifting scheme.

The JPEG2000 choice for the lossy-to-lossless compression algorithm (JPEG2000-LS) is the DWT known as LeGall 5/3, spline 5/3, or (2,2). The low- and high-pass analysis filters have 5 and 3 taps, respectively. It was introduced by D. Le Gall [Gal88] in the subband coding domain, seeking short symmetric kernels for PR image coding purposes. Cohen, Daubechies, and Feauveau [Coh92] developed families of biorthogonal transforms involving linear phase filters using

Fourier arguments. The shortest biorthogonal scaling and wavelet function with 2 regularity factors (or vanishing moments) at analysis and synthesis, denoted (2,2), is attained with the filter bank proposed by Le Gall. Indeed, the LeGall 5/3 synthesis scaling function is a linear B-spline, which is the reason for the name spline 5/3. Figure 2.9 shows the scaling and wavelet functions. Sweldens [Swe96] proposed the construction of an entire family of Deslauriers-Dubuc biorthogonal interpolating wavelets via lifting, using 2 steps. The LeGall 5/3 wavelet also belongs to this family.

The LeGall 5/3 wavelet analysis low-pass filter $H_0(z)$ and the high-pass filter $H_1(z)$ are

$$\begin{aligned} H_0(z) &= \frac{-z^{-2}+2z^{-1}+6+2z^{1}-z^{2}}{8}, \\ H_1(z) &= z^{-1}\frac{-z^{-1}+2-z^{1}}{2}. \end{aligned} \tag{2.12}$$

For lossless coding, an integer-to-integer transform [Cal98] is preferred. Lifting with a rounding after each step attains this kind of transform straightforwardly. In this way, any FIR filter bank can be implemented as an integer-to-integer transform by placing the rounding operation after each filter and before the addition or substraction because of the stated FIR filter bank factorization property into lifting steps [Dau98]. For instance, the lifting steps

$$\begin{aligned} P(x[n], x[n+1]) &= \left\lceil \frac{x[n]+x[n+1]}{2} \right\rfloor, \\ U(y'[n-1], y'[n]) &= \left\lceil \frac{y'[n-1]+y'[n]}{4} \right\rfloor, \end{aligned} \tag{2.13}$$

realize the integer-to-integer transform of the filter bank (2.12). At low bit rates, reversible integer-to-integer transforms and their conventional counterparts often yield results of comparable quality [Ada00].

If the initial wavelet is the LWT, then the low- and high-pass filters are related to the linear prediction and update through

$$\begin{aligned} H_0(z) &= 1 + H_1(z)U(z^2), \\ H_1(z) &= z^{-1} - P(z^2). \end{aligned} \tag{2.14}$$

The analysis polyphase matrix of the LeGall 5/3 wavelet is

$$\mathbf{H}_p(z) = \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{4}z+\frac{1}{4} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{2}-\frac{1}{2}z^{-1} & 1 \end{pmatrix}.$$

Adopting the convention that the low- and high-pass analysis filters are normalized to have unit gain respectively at $\omega = 0$ and $\omega = \pi$, then the final scaling factors are $K_1 = 1$ and

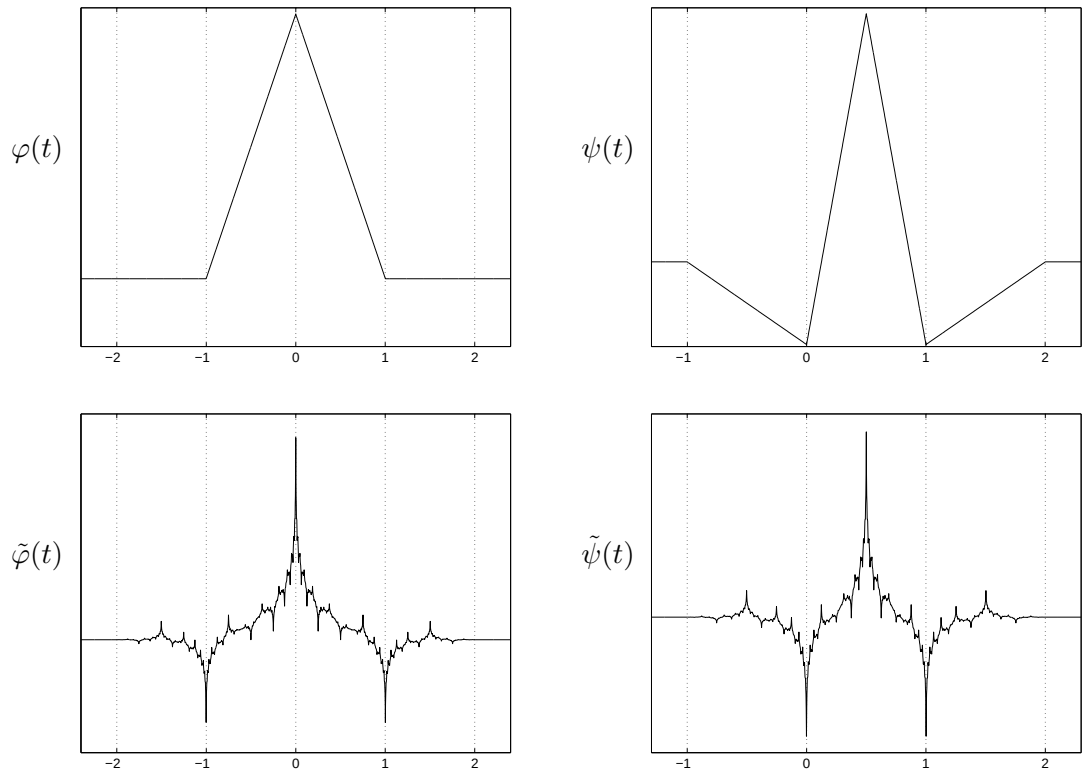


Figure 2.9: Synthesis and analysis scaling and wavelet functions for the LeGall 5/3 transform.

$K_2 = -\frac{1}{2}$. The synthesis filters are

$$\begin{aligned} G_0(z) &= \frac{z^{-1} + 2 + z^1}{2}, \\ G_1(z) &= z\frac{-z^{-2} - 2z^{-1} + 6 - 2z^1 - z^2}{8}. \end{aligned}$$

Interestingly, the lossless performance is almost independent of the normalization being performed or omitted. However, if the scaling factors are omitted, a performance degradation appear in lossy compression because the transform deviates from unitary and thus, the information content of the coefficients is not directly related to its magnitude.

JPEG2000 Part 1 standard supports the LeGall 5/3 wavelet for reversible transformations and the Daubechies 9/7 wavelet [Ant92] as irreversible transformation for lossy compression purposes. The choice of the LeGall 5/3 wavelet is not casual; it has several interesting mathematical properties, which are analyzed in [Uns03] with those of the Daubechies 9/7. LeGall 5/3 wavelet is the shortest symmetrical (to avoid boundary artifacts) biorthogonal wavelet with two vanishing moments. In addition, it has been shown that the LeGall 5/3 wavelet has the maximum vanishing moments for its support. It may be obtained by factorizing a maximally flat Daubechies or Dubuc-Deslaurier half-band filter [Dau88].

Notice that the order in which the filters are applied (analysis vs. synthesis) is important (figure 2.9): the shortest and most regular basis functions are placed on the synthesis side. This is consistent with the principle of maximizing the approximation power of the representation. Intuitively, the smoothest basis functions should be on the synthesis side in order to minimize perceptual artifacts. The reason is that the output is a weighted sum of the synthesis functions.

The wavelet transform evaluation work [Ada00] shows that the LeGall 5/3 wavelet fares well considering its very low computational complexity. For the case of images containing significant amount of high-frequency content, the 5/3 tends to obtain better lossless results than all the other longer transforms considered in the work and often by a remarkable margin. However, there is no single transform that performs consistently better for all images. That is the reason of the research effort towards signal and locally adaptive transforms as the works reviewed in section 2.3.

Since the main part of this Ph.D. thesis work is devoted to lossless compression, LeGall 5/3 wavelet constitutes the appropriate benchmark for comparisons.

2.2.5 Space-Varying Lifting

Spatial adaptivity is introduced into the lifting structure with the so-called space-varying lifting. This kind of LS chooses a lifting filter at each sample n according to the signal local characteristics (LC),

$$\begin{aligned} P(\mathbf{x}) &= P(\mathbf{x}, LC(\mathbf{x})), \\ U(\mathbf{y}') &= U(\mathbf{y}', LC(\mathbf{y}')). \end{aligned}$$

In general, there is no need to code side-information to indicate the chosen filter at sample n since lifting steps depend on the same samples as the classical (non-varying) case and so, coder and decoder have the same information available for the filter selection. This is the most significant difference w.r.t. the adaptive lifting §2.2.6 and the generalized lifting proposed in this Ph.D. thesis in chapter 4.

Typically, the operator $LC(\cdot)$ indicates flat zones, edges, or textures. The corresponding filters may vary in many ways. Some simple examples are given in the following list. Further modifications are explained in §2.3.1 and §2.3.2.

- *Coefficient values* may be modified in order to vary the number of vanishing moments [Ser00] or to perform an LMS update of the filter coefficients depending on previous estimation errors [Ger00].
- *Filter length* [Cla97, Cla03] may be altered to take into account an edge or other structures.

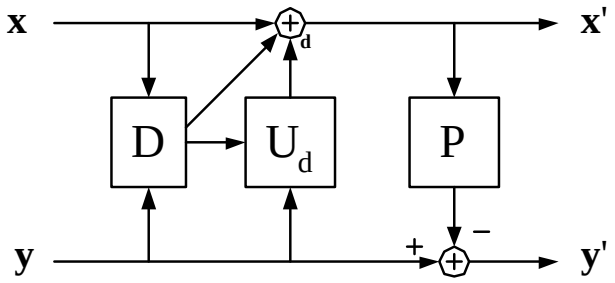


Figure 2.10: Adaptive update lifting step followed by a classical prediction.

The goal is to avoid to predict through such structures, which produces worse results than the prediction in a homogeneous region.

- *Filter type.* In [Egg95], the filter is chosen as linear or morphological in order to obtain good texture representation.

2.2.6 Adaptive Lifting

The adaptive LS is a modification of the classical lifting. A simple version was stated in [Pie01b, Pie01a, PP02] and then, a wider framework and 2-D extensions were presented in [Hei01, Hei05a, Pie05], and a lossy coding version in [PP03, Pie04, Hei05b].

Figure 2.10 shows an example of an adaptive ULS followed by a fixed prediction. At each sample n, an update operator is chosen according to a decision function $D(x[n], \mathbf{y})$. The crucial point is that $D(x[n], \mathbf{y})$ depends on $\mathbf{y}$, as in the classical and space-varying lifting, but it also depends on the sample being updated. For this reason a problem arises because the decoder does not dispose of the sample $x[n]$ used by the coder to take the decision. The decoder only knows $x'[n]$, which is an updated version of $x[n]$ through an unknown update filter. The challenge is to find a decision function and a set of filters that permit the reproduction of the decision $D(x[n], \mathbf{y})$ at the decoder,

$$D(x[n], \mathbf{y}) = D'(x'[n], \mathbf{y}), \tag{2.15}$$

thus obtaining a reversible decomposition scheme. This property is known as the decision conservation condition.

The range of D may indicate whether there exists an edge on $x[n]$ if D is the l^1-norm of the gradient

$$\begin{aligned} D: \ \mathbb{R} \times \mathbb{R}^k &\rightarrow \mathbb{R}_+ \\ (x[n], \ \mathbf{y}[n]) &\rightarrow d = \sum |y_i - x|, \end{aligned}$$

or whether $x[n]$ resides in a textured region if a texture detector D is used, like in [Egg95], or

further still, it may indicate other geometrical constraints. Also, the extension to two dimensions allows to take into account 2-D structures. Therefore, a suitable filter for the signal local characteristics at n (made evident by function D) is applied at sample n. Typically, long low-pass filters are chosen for smooth regions and short-support filters are selected around edges.

A relevant feature of the adaptive scheme is that it does not require any bookkeeping to enable PR although the filter may vary at each location using non-causal information (i.e., information not available at the decoder).

Adaptive lifting is extensively analyzed in chapter 4, which is useful in order to introduce the generalized lifting scheme. Generalized lifting contains all the possible reversible decomposition respecting equation (2.15).

2.3 Review of Lifting Algorithms

This section reviews different approaches for the construction of lifting steps. First part §2.3.1 describes works that aim to design lifting filters. Meanwhile, §2.3.2 reviews lifting optimization criteria and techniques. Finally, §2.3.3 outlines the use of lifting in video compression.

Sections 2.3.1 and 2.3.2 describe proposals in the lifting scheme domain. However, LS as part of the wavelet theory may profit and incorporate many interesting ideas from the framework of wavelet bases, discrete transforms, and filter banks design and optimization. The extension of these fields makes an attempt at an exhaustive state-of-the-art review impossible. Only some hints of such ideas applicable to lifting are detailed below.

- *Adaptivity.* Lifting offers an easy way to perform space-varying adaptive decompositions. However, adaptivity may be introduced in many different ways. For instance, [Don95] explains an idea for adaptive coding found in several works. A set of block transforms is constructed with an image training set and a LMS-like learning algorithm. First, input image is transformed with all bases. Afterwards, the bases for which the principal components give minimum MSE are selected for coding. Block transforms are updated for the image being coded. Similar systems may be envisaged using LS.

- *Topology* refers to the number of bands and the tree depth of the filter bank for each band. Many works, as [Ram96], optimize the filter bank topology instead of the filters themselves. Also, it is suggested that merging both optimizations (filters and topology at the same time) should lead to better results, but this seems to be a harder task.

- *Filter optimization.* A myriad of works are devoted to optimizing wavelet filters to attain a certain objective. [Del92] represents a standard approach minimizing detail signal variance, which is a common criterion. The search for optimal filters is usually limited to a filter

subset and/or to a signal model. In this case, the subset is the orthonormal FIR filters with a given length and the input signal second-order statistics are considered. The particularity of the proposal is that the algorithm is iterative since it progressively refines the filter by factoring the orthonormal matrices in rotation matrices.

- *Optimization criterion.* Besides filter optimization techniques and algorithms, it is interesting to focus on the optimization criteria. The usual criteria are the variance, entropy, and energy minimization. However, there are works that propose processes to optimize FB with more direct criteria, as bit-rate, a rate-distortion function, or the number of bits required for representing a signal window [Sin93]. All these criteria may also be employed in LS. Other criteria suggested for wavelet systems design are the coding gain, filters frequency selectivity, the number of vanishing moments, and the smoothness of the synthesis functions.

Many other works, as well as the above examples, are applicable to LS since it is composed of filters with a more or less clear objective. The following two sections §2.3.1 and §2.3.2 review works directly involved in the construction of lifting steps.

2.3.1 Methods for Lifting Design

Several works presented decomposition structures resembling LS before it was formally exposed by Sweldens in [Swe96]. For instance, [Bru92] introduces a ladder network scheme related to lifting. Also, [Egg95] proposes a lifting-type scheme that switches the high-pass analysis and the low-pass synthesis filter between a linear and a morphological filter according to a texture detector. The result is a biorthogonal filter bank capable of filtering each kind of region with the suited filter: textures with a linear filter and homogeneous regions and edges with a morphological one. The half-band type decomposition structure guarantees PR irrespective of the decomposition function and indeed, it is maximally decimated.

Equivalent structures are used in other nonlinear subband decompositions as [Que95, Que98]. Also [Flo94] describes a decomposition that can be seen as a down-sampling followed by a PLS of the finer scale signal. The prediction is performed by a nonlinear weighted median. The goal is to limit the aliasing artifacts. The output is a decomposition made up of an approximation and several error signals. [Que98] explains a similar idea that works directly on the 2-D image polyphase components. The interpolative prediction is a hybrid median-mean filter [Pit90]. The resulting quantized transform is useful for lossy compression.

In the nonlinear decomposition methods proposed before [Flo96], one of the subsignals is always a simple down-sampling of the original signal. This work presents a more general framework composed of nonlinear elementary stages that has linear filter banks and a sort of lifting

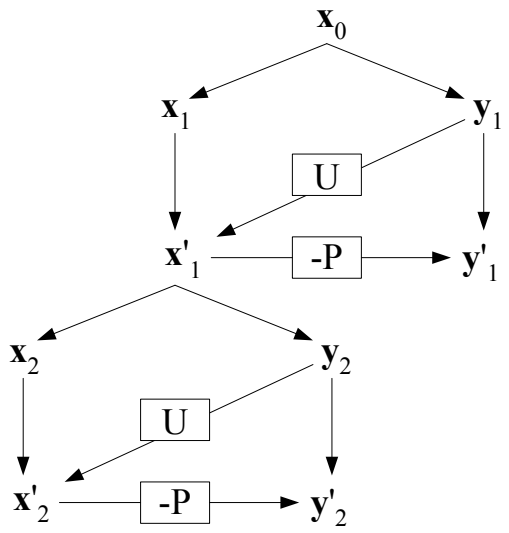


Figure 2.11: Two-level lifted wavelet transform with an update-first structure.

as particular cases. The goal is to produce a reversible and scalable approximate signal while profiting from nonlinear filtering advantages.

In [Ham96] one of the first nonlinear subband decompositions for image coding using lifting appears. A median prediction is proposed in order to reduce the ringing artifacts, which are typical of linear filters in lossy coding. [Ham96] is extended in the subsequent work [Ham98], which includes a theoretical analysis of a class of M-band nonlinear decompositions with maximal decimation and PR. An interesting particular case consists of a filter bank formed of injective operators and followed by lifting steps.

[Cla97] is already fully developed in the lifting framework. A set of linear predictors of different lengths are chosen according to a nonlinear selection function based on an edge detector. This avoids making a prediction based on data separated from the point of interest by a discontinuity. Prediction is preceded by a linear update (figure 2.11) in order to reach a stable transform throughout all resolution levels (since coarse scale coefficients linearly depend on the original signal) and to maintain a coherent space-frequency interpretation of the updated coefficients. [Cla03] extends the previous work and analyzes the transform reversibility, stability, and frequency characteristics. In addition, [Cla03] employs a 2-D non-separable window to make better prediction filter choices. The scheme is inherently devoted to lossy coding.

[Cla98] explains the possibility of applying a median predictor and keeping the track of the underlying basis for computing a median update, which attains a low-pass approximate signal if the appropriate constraints are considered. The scheme achieves nonlinear processing and multi-scale properties at the same time. The drawback is that side-information of the tracking is required except for the simplest case. Therefore, the scheme interest for compression applications

is reduced.

In the reversible integer-to-integer setting, [Ada99] establishes a set of criteria to be hold by the transform. The criteria are a minimum of two analysis and synthesis vanishing moments, to exceed a certain coding gain threshold, and low-pass and high-pass spectral restrictions. A systematic search for the filters respecting these criteria is proposed among the short filters having powers of two or dyadic rational coefficients. The algorithm outputs are several already known decompositions and other new low-complexity filters.

[Tau99] describes a sequence of nonlinear lifting steps that realize an orientation adaptive transform which reduces artifacts near edges. The resulting bit-stream is quite scalable because the edge detector is a function with low susceptibility to quantization errors. A similar idea is proposed in [Ger05, Ger06]. The approximation signal gradient is considered for the choice of the filtering direction of the LeGall 5/3 predictor. A detail sample is predicted through the direction with smaller gradient among the three possible directions, which are the horizontal, the top-left to down-right, and the down-left to top-right directions in the 1-D row-wise filtering case. Therefore, this is a multi-line filtering because data from neighboring rows may be employed for a row-wise filtering (the same occurs for the column-wise filtering). Multi-line lifting can be made computationally efficient as showed in [Tau99]. However, there exists the possibility of update leakage, i.e., information from other rows flows to the approximation signal and thus the subsequent ULS deviates from an anti-aliasing filter. To avoid this problem the strategy is the same as [Cla97], that is, the ULS is performed before the PLS.

An axiomatic framework of wavelet-type multi-resolution signal decompositions and nonlinear wavelets based on morphological operators and lifting is presented in [Hei00]. The work is not devoted to compression but the framework is used by [Li02] to create a statistical PLS that implicitly profits from local gradient information.

[Abh03a] also uses gradient information extracted from detail coefficients to modify the ULS in order to have a low-pass channel with adaptive smoothing that preserves edges and singularities.

In [Gou00], symmetrical lifting steps are constructed for quincunx sampled images. The method applies the McClellan transform to 1-D symmetrical filters to obtain 2-D filters that lie on a quincunx grid. [Kov00] provides a method for building wavelets via lifting in any dimension, for any type of lattice, and any number of vanishing moments. The construction involves an interpolative prediction based on the Neville filters and a running average update. A practical implementation of Neville filters for a quincunx grid appear in [Zee02]. The package includes nonlinear maxmin filters.

In [Sun04] a separable linear 2-D lifting is directly applied to the image, instead of applying the usual two 1-D filters on rows and then on columns. Its integer version is slightly different

from the usual filters and shows marginal lossless compression improvements with respect to it. The interest resides in the computational load gains.

[Jan04] is an original approach that introduces the adaptation in the LS at the signal splitting stage. Coarse scale samples are used to insert new samples close to edges. As a consequence, the number of coarse and detail samples is signal-dependant.

[Wan05] employs a lifting structure to perform a curved wavelet transform on the image domain. It improves the results of the separable transform in JPEG2000 despite of the required side-information to encode the curves along which the transform is computed.

[Zha04] introduces the orthogonality restriction into the lifting structure itself, instead of the weaker and usual biorthogonality. A class of IIR orthogonal filter banks is constructed by means of IIR all pass lifting filters. Lei et al. [Lei05] work on the design of 2-channel linear phase FB. Linear phase becomes a structural property of the LS in addition to the usual lifting PR property through a slight structure modification and a specific decomposition of the analysis polyphase matrix of the FIR FB. This permits an unconstrained optimization of the remaining FB free parameters.

This review shows different ways to profit from the flexibility given by the LS. The surveyed works employ the LS degrees of freedom to design space-varying, nonlinear, or 2-D non-separable decompositions, among others. Criteria are quite intuitive and logical in order to obtain good coding results, but none of the reviewed works states any objective criterion to optimize. The proposal in §5.1.1 remains within the philosophy of these approaches.

2.3.2 Methods for Lifting Optimization

[Sai96b] uses an initial S-transform refined by a second PLS. Three possibilities are given to compute an optimal second step: minimum variance (with Yule-Walker equations), minimum entropy (with Nelder-Mead simplex algorithm), and a frequency domain design (which reports the best results). [Yoo02] proposes 4-tap prediction and update filters that are reduced to be function of one parameter each after applying typical constraints: filter symmetry, zero DC gain of the high-pass filter, zero Nyquist gain of the low-pass, prediction coefficients summing up to one, and running average conservation. Therefore, the optimization is reduced to tune one parameter per lifting step. Several known wavelet transforms may be attained according to the parameter values. The optimization criterion is the weighted first-order entropy [Cal98]. Optimal values are found by an exhaustive search within the parameters interval.

[Dee03] starts from a given LS that is improved by means of a wavelet coefficients prediction from those already known. Prediction MSE is minimized considering the projection of the wavelet underlying vector onto the vectorial subspace spanned by the causal wavelet vectors and taking

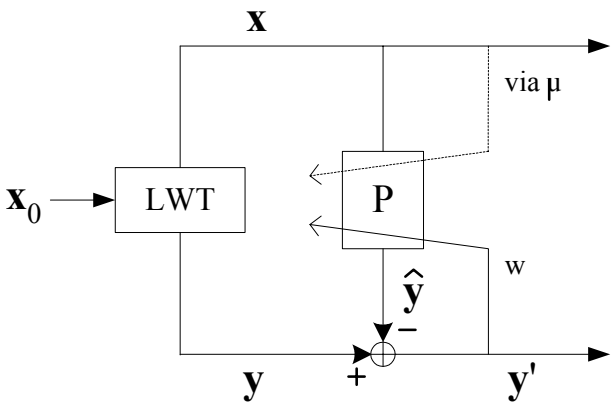


Figure 2.12: Analysis stage of the space-varying LMS filter bank.

into account the signal auto-correlation (a first-order auto-regressive model is used).

In [Bou01], input signal is assumed to be a stationary process with a known auto-correlation and spectral energy density. Prediction coefficients minimizing MSE are computed, obtaining the filter Fourier transform in function of the input spectral energy density. Optimal linear predictors are enhanced by directional nonlinear post-processing in the quincunx case and by adaptive-length nonlinear post-processing in the row-column sampling case. The ULS is fixed and linear.

In [Ger00] the space-varying decomposition scheme of figure 2.12 is proposed. The prediction is a linear filter,

$$\widehat{y}[n] = \sum_{k=-N}^{N} w_{n,k} x[n-k],$$

and its initial coefficients $w_{0,k}$ are progressively refined with an LMS-type algorithm according to prediction error,

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu \frac{\mathbf{x}_N}{\|\mathbf{x}_N\|^2} e[n].$$

The step size μ of the algorithm depends on the input values range. This scheme adaptively optimizes prediction coefficients and it is able to deal with the non-stationarity of images. A similar work is that of [Tra99], which aims to design a data-dependant prediction with the goal of minimizing the detail signal. This work distinguishes two approaches. One uses a local optimization criterion with space-varying prediction filter coefficients. The other approach is global in the sense that the l^1-norm of the entire detail signal is minimized through Wiener filter theory.

In [BB02], lifting is applied to multi-component images to eliminate intra- and inter-component correlation. First, the samples involved in filtering are determined. Then, a band and resolution level weighted entropy is minimized. Numeric optimization methods are used

(Nelder-Mead simplex) because entropy is an implicit function of the decomposition parameters. Yule-Walker equations solution is taken as the initial prediction filter. This method requires sending side-information.

[Ho99] defines several half-band filters. Each resolution level uses a linear combination of them. Combination weights minimize the sum of the prediction error for all resolution levels. [Abh03b] selects as lifting filters the Lagrange polynomial interpolators of degree 0, 1, 2, and 3 (which vanish 0, 1, 2, and 4 signal moments, respectively). Wiener-Hopf equations relating the auto-correlation and the filter coefficients are formulated. The solution gives the MSE minimizing coefficients. At each sample, the chosen interpolation filter is the one that applied to the Wiener-Hopf equation using the optimal filter coefficients gives the minimum MSE.

In [Kuz98] initial biorthogonal filters are modified by directly optimizing the ULS coefficients. The trick is that a very precise signal model is considered since the goal is to compress electrocardiography signals. Lagrange multipliers are used to minimize a cost function that leads to compact support wavelets. Two restrictions are imposed: the update step should be a high-pass filter and the detail signal equal to zero for the specified model. In [Fah02], for a concrete signal class, update coefficients are gradually improved until reaching the minimum MSE.

In [BB03] the image is partitioned into disjoint fixed blocks that are classified in several regions. Then, a predetermined couple of lifting steps is chosen for each block according to local statistics. The global entropy of the resulting pyramidal representation is minimized. Side-information indicating the filter choice is required. [Hat04] extends previous work by using a quadtree segmentation rule that gives flexibility to the input image block-partitioning stage. Detail signal statistics are modeled to minimize the entropy. Side-information containing the quadtree structure and the prediction coefficients is needed by the decoder. A further extension of the variable-size block-based procedure for the compression of multi-component images appear in [Hat05]. The quadtree partitioning rule takes into account simultaneously spatial and spectral redundancies.

[Gou01] optimizes prediction filter coefficients on a quincunx grid to minimize the detail signal variance. Then, it optimizes update coefficients to minimize the quadratic error between original image and the image reconstructed using a zero detail signal. This principle also aims at offering a better resolution and quality scalability. Filter coefficients are transmitted to the decoder in order to proceed to the inverse transform. [Gou04] generalizes [Gou01] to any kind of sampling grid. Filters are linear phase 2-D FIR.

In [Li05], the detail signal energy is minimized. It is expressed as a function of the auto-correlation of the image, the auto-correlation of image difference, or the auto-correlation of the image second-order difference. The kind of correlation shown by an image imposes the choice of the prediction among the three design criteria. Once the prediction coefficients are determined,

the ULS that maximizes the smoothness according to the Sobolev regularity is selected. Some nonlinear enhancements are proposed in order to improve the linear filter performance.

In summary, all the works reviewed in this section employ an optimization criterion to design lifting steps and an optimization technique to reach or to find such an optimal step. Common criteria are the variance, the (weighted) entropy, and the prediction MSE of the detail signal coefficients. Bit-rate and rate-distortion functions are used more frequently in the filter bank design than in LS. Optimization techniques such as Yule-Walker or Wiener-Hopf equations and the LMS algorithm are widespread. Often, the objective is a differentiable function and a closed-form solution may be derived. Another resource is the heuristic Nelder-Mead method. Decomposition topology is a variable that may also be modified to optimize any criterion. Assumptions on input data tend to simplify the design and to report better results than generic designs.

Several proposals throughout the Ph.D. dissertation follow this line to design new lifting steps (e.g., proposals in §3.3, §5.1.2, and §5.2.2).

2.3.3 Lifting in Video Compression

Subband motion-compensated temporal filtering video codecs have recently attracted attention due to their high compression performance comparable to state-of-the-art hybrid codecs based on the predictive feedback paradigm and due to their scalability features, which provide superior support for embedded, rate-scalable signal representations. Scalable video encoders work without knowledge of the actual decoded quality, resolution, or bit-rate. Scalability (cf. §2.5.1) is a desirable property for interactive multimedia applications, and it is a requirement to accommodate varying network bandwidths and different receiver capabilities. Scalable video coding also provides solutions for network congestion, video server design, and the protection for the transmission in error prone environments.

Initial subband-based video codecs computed the 3-D decomposition directly through the three spatio-temporal dimensions, leading to poor results w.r.t. the standard predictive coding. Currently, subband coding exploits the temporal inter-frame redundancy by applying a temporal wavelet transform in the motion direction over the frames of the video sequence raising performance to very competitive levels. LS is the considered realization for the adaptive motion-compensated temporal DWT.

An early attempt to introduce motion compensation into 3-D subband coding is due to Ohm et al [Ohm94]. Like most of the founding works in this domain, the temporal filter is limited to the two-tap Haar filter, which is unsatisfactory. In later works, the temporal transform support is extended by rendering the LeGall 5/3 wavelet applicable. LeGall 5/3 is spread out in motion-compensated video coding for the temporal filtering part.

Many works lean on block-based motion compensation, with fixed or variable block sizes, because of its simplicity and the long experience that has been developed around this tool. However, this technique leads to disconnected and multiple connected pixels, which may produce annoying coding artifacts. Disconnected and multiple connected pixels are those in the reference frame not used for temporal prediction and those used to predict more than one pixel in the current frame, respectively. This casuistry supposes extra difficulty to the subsequent ULS design. In [Ohm94], disconnected pixels are treated differently from connected pixels in order to maintain reversibility in integer-pixel accurate motion compensation. The use of mesh models for motion compensation may be a solution to the problem of processing these pixels. For example, [Tau94] spatially aligns video frames by arbitrary frame warping before applying the 3-D decomposition. In general, this warping is not invertible and therefore, it cannot achieve perfect reconstruction.

A research stream is the block-based motion estimation and the plead in favor of finding a satisfactory way to process the disconnected and multiple connected pixels in order to provide high-efficiency video codecs. Many works [Sec01, Meh03, Sec03, Gir05, Til05] are devoted to finding motion-compensated subband decompositions.

Efforts have also been led to the design and optimization of lifting steps for the video coding application. In lossy subband motion-compensated video coding, Girod et al. [Gir05] propose a prediction minimizing high-band energy and assert that an ULS design towards the same goal does not significantly contribute to reduce the global bit-rate. On the other hand, the inspection of the inverse transform reveals that the update greatly impacts in the distortion of the reconstructed frame sequences. Therefore, the ULS is optimized in order to minimize the reconstruction distortion assuming that the PLS and ULS are linear and the quantization errors introduced in the low- and high-pass frames are uncorrelated random vectors.

The same research line is followed in [Til05] but the assumptions are stronger: all coefficients in the detail frames are quantized with the same quantization step, they are independent and identically distributed, and the error is uncorrelated in the different bands. Then, the reconstruction error variance of an approximation pixel is related to the quantization errors variances. Thus, the optimal ULS coefficients are derived by the reconstruction error minimization. When the update filter coefficients are not restricted to sum up to one, the given solution is the same as the given in [Gir05]. Simple nonlinear ULS are also presented. The different alternatives lead to similar results.

ULS design is a difficult part in the video coding application, since the incorporation of this step contributes to the bit-stream scalable properties and to the decrease of the reconstruction error, but it may jeopardize compression rate and introduce disturbing artifacts in the low-pass temporal frames if the motion estimation fails. The trade-off is partly surmounted by adaptively varying the update weights according to the high-pass temporal frame energy [Meh03] or by the

related idea of changing the high-pass frames that the ULS employs.

2.4 Other Adaptive, Nonlinear, and Sparse Decompositions

There are many references, not directly related to (nonlinear) lifting, but more generally to new adaptive, directional, and sparse decompositions that is worth being mentioned since their goal is related to the proposed schemes. In the last years, a multitude of such representations has appeared: wedglets [Don97], ridgelets [Don98], bandelets [Pen00], curvelets [Sta02], contourlets, armlets, dual trees, etc.

They arise from the observation that in many applications the most interesting parts of a signal are its singularities. The consequent approach is to look for wavelets capable of tracking the discontinuities in shape. This idea leads to the construction of functions whose support has a shape adapted to the signal regularity. For example, Donoho [Don97] studied the optimal approximation of particular classes of indicator functions with an overcomplete collection of atoms, called wedgelets. This construction is based on a multi-scale organization of the edge data.

Another approach due to the same author are the ridgelets [Don98], which are elongated wavelets especially suited for object discontinuities across straight lines. Similarly, in [Pen00] a family of orthogonal wavelets capable of efficiently representing singularities along regular contours is combined with the Daubechies 9/7 wavelet transform. Such type of decompositions are called bandelets.

2.5 Wavelet-based Image Coders

This section is mainly devoted to explain wavelet-based image coders and their particular modular structure, but it also describes some other relevant image coders. The concepts of embedded bit-stream and scalability are introduced in §2.5.1. JPEG2000 standard is largely derived from the EBCOT coder. The widespread SPIHT coder and the EBCOT coder are described in more detail in §2.5.2 and §2.5.3.

This Ph.D. thesis aims to obtain "good" transforms for compression. Two image coders are mainly used to compare transforms: SPIHT without arithmetic encoder and EBCOT. Statistics drawn from transformed coefficients, as detail coefficients mean energy, variance, entropy, or even the number of zeros may also be indicators of a transform "goodness". Since a principal focus is lossless compression, a fair final comparison benchmark would seem to be JPEG-LS. For the sake of locating the proposed algorithms performance in a more global perspective, JPEG-LS compression results are sometimes given, but it is important to notice the functionalities loss of

this standard and that the goal is to compare different transforms, not image or entropy coders.

Wavelet-based image coders era began with the notable breakthrough of embedded zero-tree wavelet (EZW) coding by J. Shapiro [Sha93]. The EZW algorithm was able to exploit the multi-resolution properties of the DWT to produce a computationally simple algorithm with outstanding performance and quality embedded bit-stream. In this context, *embedded* means that every prefix of the compressed image bit-stream is itself a compressed bit-stream, but at a lower rate.

A number of wavelet coding methods have been proposed since the introduction of EZW. A common characteristic of these methods is that they use fundamental ideas found in the EZW algorithm. Most of image coders based on MRA have a similar modular structure:

1. *Multi-component decomposition* (MCT). Color or multi-spectral images have high correlation among their components that is reduced by an initial inter-component decomposition.

2. *Multi-resolution analysis.*

3. *Quantization* of the transformed coefficients. Common examples of quantization employed in practice are scalar uniform, trellis, and vectorial quantization. Scalar quantization with deadzone provides an embedded quantization. This means that the intervals of lower rate quantizers are partitioned to yield the intervals of higher rate quanitzers. Embedded quantization is very useful to obtain embedded bit-streams of the compressed image.

4. *Bit-plane encoding.* The array of quantized scalar indices coming from the quantization step are seen as a set binary arrays or bit-planes, the slices of the indices. The first bit-plane consists of the most significant bit (MSB) of each magnitude. The next bit-plane consists of the second MSB plane, and so on until the bit-plane for the least significant bit (LSB). There is also a bit-plane that consists of the sign bit of each index. Wavelet-based coders have a part dedicated to successively coding each of the bit-planes. EZW and SPIHT's main contribution is the specific way of bit-plane encoding. Bit-plane encoding appears in JPEG2000 in the so-called tier-1 (part 1 of the EBCOT).

5. *Entropy coder.* Context-based arithmetic coding is the common form of entropy coding applied to bit-planes. The sequence of symbols resulting from the bit-plane coding in EZW are losslessly compressed with the classical context-dependant arithmetic coding in [Wit87]. JPEG2000 uses the more sophisticated JBIG2 MQ-coder.

6. *Rate-distortion control.* SPIHT cuts the compressed bit-stream at a given point in order to achieve the desired rate or distortion. JPEG2000 tier-2 is devoted to the bit-stream organization with the goal of rate-distortion control, among others.

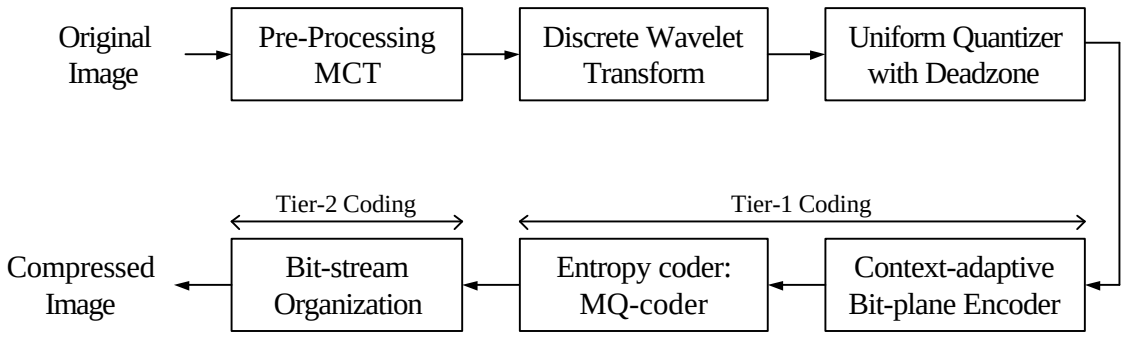


Figure 2.13: JPEG2000 block diagram.

In general, these points are a good representation of an MRA-based image coder. Blocks 2 to 5 are present in most of coders. The presence of block 1 depends on the target application. Block 6 is quite specific and one of the relevant characteristics of the EBCOT coder (§2.5.3). Quantization is inherently a lossy function, so it is not performed in lossless coding.

Figure 2.13 is an example of a wavelet-based image coder block diagram. The figure shows the JPEG2000 standard fundamental building blocks.

Scalar-quantization with deadzone is a function Q with the quantization step Δ of an input coefficient x that produces the quantization indices q:

$$q = Q(x) = \text{sign}(x) \begin{cases} \left\lfloor \frac{|x|}{\Delta} + \tau \right\rfloor, & \text{if } \frac{|x|}{\Delta} + \tau > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Parameter τ controls the width of the central deadzone. The most common values are $\tau = 1/2$, which amounts to a uniform quantizer, and $\tau = 0$, which corresponds to a deadzone width of 2Δ. JPEG2000 uses $\tau = 0$. Assuming that the magnitude of q is represented with N bits, then q may be written in sign-magnitude form as

$$q = Q(x) = s\ q_N\ q_{N-1} \ldots q_1,$$

where q_i is a magnitude bit and s is the sign bit. A bit-plane i is formed of the q_i bit of all the transformed coefficients.

Embedded bit-stream by means of bit-plane coding were used before the appearance of EZW. It was even included as part of the original JPEG standard [Pen92]. Before the appearance of EZW, coding systems followed a fixed scan pattern, p.e., a zig-zag scan order or a raster scan. Bit-plane coding in EZW permits data-dependant scanning in the form of zero-trees, which allows the coding of large numbers of zeros using very few compressed bits.

This idea is retaken and generalized in one of the more widespread methods, the Set Partitioning in Hierarchical Trees (SPIHT) algorithm introduced by Said and Pearlman in [Sai96a].

SPIHT became very popular since it was able to achieve equal or better performance than EZW without arithmetic encoder. SPIHT is extensively known in entropy coding and it is a good choice to compare transforms compression properties. It exploits the wavelet transform self-similarity across scales by using set partitioning. Wavelet coefficients are ordered into sets using a parent-child relationship and their significance at successively finer quantization levels. The produced bit-stream is SNR scalable. A more detailed explanation on SPIHT is given in §2.5.2. Appendix 5.B describes an extended 3-D version elaborated for the purposes of this Ph.D. thesis.

Another coding technique based on wavelet decomposition is the Embedded Block Coding with Optimized Truncation [Tau00] (EBCOT for short), which has been chosen as the basis of JPEG2000 standard [ISO00, Sko01, Use01]. JPEG2000 is the latest ISO/ITU-T standard for still image coding. A profound reading on the standard and surrounding technology is found in [Tau02a]. The standard is based on the discrete wavelet transform, scalar quantization with deadzone, arithmetic coding with context modeling, and post-compression rate allocation (figure 2.13). Section 2.5.3 explains the foundations of EBCOT (and JPEG2000).

MRA coders are the best for lossy compression. However, the concurrence is harder for lossless coding. There are algorithms that perform very well. Possibly, the state-of-the-art lossless coder is still CALIC [Wu97, Say00]. There exist specific coders that outperform CALIC in their target application, as JBIG2 [ISO99b] for binary images or PNG [W3C96], and PWC [Aus00] for palette and graphic images. However, for a wide range of images CALIC reports best results but at the cost of several functionalities found in JPEG2000 and other standards, such as support for scalability and error resilience. JPEG-LS [ISO99a] is the new ISO/ITU-T standard for lossless coding of still images. It is based on LOCO-I [Wei00] algorithm, a simplified CALIC. JPEG-LS uses adaptive prediction, context modeling, and Golomb coding. It reaches a compression efficiency very close to the best reported results.

2.5.1 Embedded Bit-Stream and Scalability

Any truncation of an embedded code achieves a decompressed image with a lower quality or resolution than the original. The application of embedded codes applied to image compression allows functionalities as remote browsing and interactive retrieval, and to accommodate different receiver capabilities.

Embedding arises from scalable coding. Scalability involves the generation of various layers from a single source or image. The lower layer provides a basic image representation and the subsequent layers successively enhance and refine the representation.

A compressed bit-stream is SNR scalable if the representation progressively improves the decoded image quality, i.e., reduces the distortion. The bit-streams produced by the EZW and

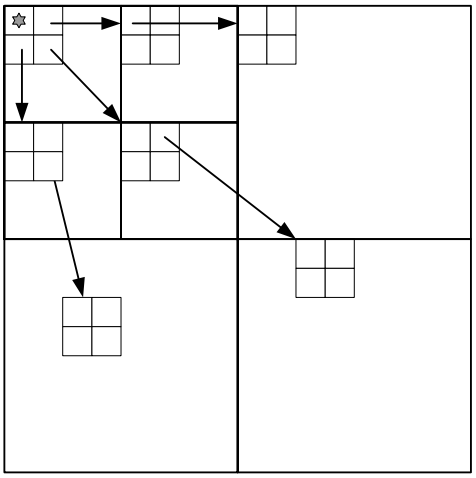


Figure 2.14: Parent-child relationships in SPIHT.

SPIHT are SNR scalable. A bit-stream is resolution scalable if it contributes to decode the bands of each resolution level successively from the lowest to the highest level.

An image coder may produce a bit-stream that is both, SNR and resolution scalable (e.g. JPEG2000). In this case, the bit-stream is composed by elements that successively improve the quality of the bands in each resolution level. These elements are part of the bit-stream that belong to both an SNR scalable and a resolution scalable bit-stream. There also exists the spatial scalability, in which image is decoded in a spatially ordered way. JPEG2000 admits component scalability for multi-component images, too.

A linear bit-stream only possesses one kind of scalability. Therefore, when two or more types of scalability are supported at the same time by a bit-stream, additional structural information is required in order to identify the location of the elements that permit the decoder to perform the appropriate kind of decompression according to the target application.

2.5.2 SPIHT coder

SPIHT exploits the dependencies existing among the transform coefficients in different bands. Concretely, coding gain arises from the representation with few bits of large regions of zero or near zero coefficients with related locations through scales and bands. SPIHT employs a parent-child relation among the coefficients of the same orientation bands. For example, a coefficient in the LH3 band is the parent of 4 children in the LH2 band. Coefficients in the HL1, LH1, and HH1 bands have no children. Also, one of every four coefficients in the lowest resolution band has no children. Figure 2.14 shows these parent-child relations. The coefficient marked with a star has no children.

SPIHT is described in terms of bit-plane coding of signed indices arising from a dead-zone scalar quantization of the transformed coefficients. A coefficient $c_{i,j}$ is quantized to $q = Q(c) = s\ q_N\ q_{N-1} \dots q_1$. Each q_i contributes to a bit-plane. The maximum magnitude for all the coefficients $N = \lfloor \log_2 \max_{(i,j)}(|c_{i,j}|) \rfloor$ determines the number of bit-planes. SPIHT encodes successively the bit-planes from $n = N$ to $n = 1$.

The assumption that the descendants of small coefficients tend to also be small is put in practice by the concept of significance. The significance of a coefficient determines if it is "large" or "small", which is made precise by comparing coefficients to a series of thresholds T_n, for $n = 1, \dots, N$. The initial threshold is $T_1 = 2^{N-1}$. All coefficients with $c_{i,j} \geq 2^{N-1}$ are significant w.r.t T_1, which means that they have a 1 at the MSB, i.e., in the bit-plane N. The following thresholds are defined by $T_n = 2^{N-n}$.

Zero-tree is the structure defined to signal that the descendants of a *root* coefficient have all 0 within a bit-plane. SPIHT employs two types of zero-trees. The first consists of a single root coefficient having all descendants 0 within the given bit-plane. The second type is similar, but excludes the four children of the root coefficient. If all descendants of a given root are insignificant, they comprise an insignificant set of type A. Similarly, the insignificant sets are of type B if they do not contain the root children. SPIHT employs three ordered lists which store the significance information of the different sets and coefficients:

- List of significant coefficients (LSC): contains the coordinates of all coefficients that are significant w.r.t. the given threshold.
- List of insignificant sets (LIS): contains the coordinates of the roots of insignificant sets of coefficients of type A and B.
- List of insignificant coefficients (LIC): contains the coordinates of all the root coefficients that are insignificant, but that do not reside within one of the two types of insignificant sets.

Each bit-plane is coded by a significance pass (also called sorting pass), followed by a refinement pass. The refinement pass codes a refinement bit for each coefficient that was significant at the end of the previous bit-plane following the order given by the LSC. Coefficients that become significant in the scan of the current bit-plane are not refined until the next bit-plane. Every coefficient in every band is initialized to the insignificant state. Then, the bit in the current bit-plane, starting with q_N, for every coefficient in the LIC list is checked. If it is one, the coefficient is significant, it is coded with the sign of the coefficient, and moved to the LSC. After that, each set in the LIS is examined, in the order of appearance in the list. If all coefficients in a set in the LIS are insignificant, a zero is coded for all the set. If there is a significant coefficient, the set is partitioned and the new sets and roots are sent to the corresponding list. This process

continues until all the LSB are coded. However, it may be stopped at any point if a desired rate or distortion is reached, since it is assured that the resulting bit-stream until the halting point is a lower rate-distortion code of the image.

Only significance bits are arithmetically coded in SPIHT. No gain is obtained to code the refinement or sign bits. Spatial significance bit dependencies are exploited by grouping them in 2x2 blocks. Different coding contexts are used depending on which bits are to be coded. The typical gain of SPIHT with arithmetic coding w.r.t. SPIHT without arithmetic coding is about 0.5 dB.

2.5.3 EBCOT coder

JPEG2000 supports several functionalities at the same time. Important examples are spatial, resolution, and SNR scalability, enhanced error resilience, random access, the possibility to encode images with arbitrarily shaped regions of interest, and open architecture. Many of these functionalities are inherent to wavelets and EBCOT coder. The EBCOT coder version in the JPEG2000 standard [Tau02b] has some differences w.r.t. the original [Tau00]. The former EBCOT [Tau02b] is used in this dissertation.

Wavelet coefficients become quantized indices through a deadzone quantization. Then, each bit-plane is split into blocks. A typical block size is 64x64. The coding proceeds from the most significant bit-plane to the least significant bit-plane. Blocks are coded independently using three coding passes: the significance propagation, the refinement, and the clean-up passes. The bit-plane coding procedure provides binary symbols and context labels to the MQ coder, which is a specific context-based arithmetic coder that converts the input symbols to compressed output bits. This is the JPEG2000 Tier 1 that generates a collection of bit-streams. One independent bit-stream is generated for each code-block, being each block bit-stream embedded.

Tier 1 is followed by the Tier 2 that multiplexes the bit-streams for their inclusion in the image code-stream and efficiently signals the ordering of the coded bit-plane passes. Tier 2 coded data may be easily parsed. Tier 2 also enables the SNR, resolution, spatial, ROI, and arbitrary progression and scalability.