Centrum voor Wiskunde en Informatica

REPORT RAPPORT

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

PNA

Building nonredundant adaptive wavelets by update lifting

H.J.A.M. Heijmans, B. Pesquet-Popescu, G. Piella

REPORT PNA-R0212 MAY 31, 2002

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

Building Nonredundant Adaptive Wavelets by Update Lifting

Henk J.A.M. Heijmans
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Béatrice Pesquet-Popescu
Signal and Image Processing Department
Ecole Nationale Supérieure des Télécommunications
46 rue Barrault, 75013 Paris, France

Gemma Piella
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

ABSTRACT

Adaptive wavelet decompositions appear useful in various applications in image and video processing, such as image analysis, compression, feature extraction, denoising and deconvolution, or optic flow estimation. For such tasks it may be important that the multiresolution representations take into account the characteristics of the underlying signal and do leave intact important signal characteristics such as sharp transitions, edges, singularities or other regions of interest. In this paper, we propose a technique for building adaptive wavelets by means of an extension of the lifting scheme. The classical lifting scheme provides a simple yet flexible method for building new, possibly nonlinear, wavelets from existing ones. It comprises a given wavelet transform, followed by a prediction and an update step. The update step in such a scheme computes a modification of the approximation signal, using information in the detail band. It is obvious that such an operation can be inverted, and therefore the perfect reconstruction property is guaranteed. In this paper we propose a lifting scheme including an adaptive update lifting and a fixed prediction lifting step. The adaptivity consists hereof that the system can choose between two different update filters, and that this choice is triggered by the local gradient of the original signal. If the gradient is large (in some seminorm sense) it chooses one filter, if it is small the other. In this paper we derive necessary and sufficient conditions for the invertibility of such an adaptive system for various scenarios.

2000 Mathematics Subject Classification: 42C40, 94A12
Keywords and Phrases: Perfect reconstruction filter bank, adaptive wavelet, lifting scheme, update lifting, seminorm, threshold criterion, quincunx scheme, denoising, derivative filter.
Note: This work was carried out under project PNA4.2 “Image Representation and Analysis”. The work of Heijmans and Pesquet-Popescu is partially supported by the IST Programme of the EU under contract number IST-2000-26467 (MASCOT). The work of Piella is sponsored by the Dutch Technology Foundation STW, project no. CWI 4616.

1. Introduction

Multiresolution representations, such as pyramids and wavelets, provide a powerful tool for the analysis of signals, images, and video sequences [2, 10, 25, 31]. The classical wavelet transforms, both continuous and discrete, are linear, and their constructions are often based on the ‘good old’ Fourier transform. The introduction by Sweldens [27–29] of the lifting scheme, however, has changed the ‘wavelet scene’ dramatically. This scheme, illustrated in Fig. 1, provides a general and flexible tool for the construction of new wavelets from existing ones. The general ingredients of the lifting scheme are an existing wavelet transform WT,

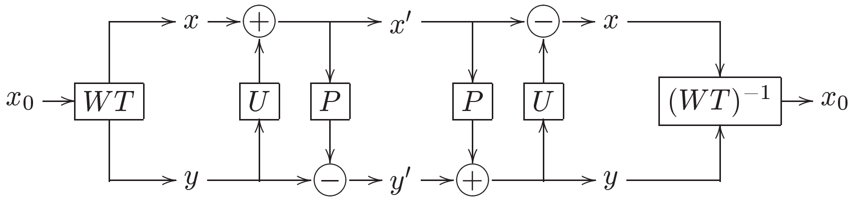

Figure 1: *General lifting scheme.*

an update map U, and a prediction map P. A decomposition of an input signal x_0 into bands x', y' is obtained in the following way. The original signal x_0 is first split into an approximation signal x and a detail signal y by a given wavelet transform WT (which may be a polyphase decomposition, also called 'lazy wavelet transform'). The update map U acting on y is used to modify x, resulting in a new approximation signal $x' = x + U(y)$. Subsequently, the prediction map P acting on x' is used to modify y, yielding a new detail signal $y' = y - P(x')$. At synthesis, the original signal x_0 is reconstructed by reversing the lifting steps and applying the inverse of WT.

It is important to observe that the invertibility of the scheme is guaranteed and does not require any condition on the lifting maps P and U. This flexibility has challenged researchers to develop various nonlinear wavelet transforms [5–9, 19–21], including morphological ones [4, 11, 14–17, 22].

The multiscale analyses deriving from classical multiresolution transforms, including many of the nonlinear transforms whose construction is based on the lifting scheme, lead to a uniform smoothing of the information contents in the image when going to lower resolutions. However, in a large number of applications in image and video processing it would be useful to have multiresolution representations that take into account the characteristics of the underlying signal and do leave intact or even enhance certain important signal characteristics such as sharp transitions, edges, singularities or other regions of interest. The importance of such "intelligent", "adaptive" or "data-driven" representations in image analysis, compression, denoising, or feature extraction, has been recognised by various researchers and has led to a wealth of new approaches in wavelet theory, such as bandelets [24], ridgelets [13], curvelets [3], wedgelets [12], etc.

In this paper we propose a general framework of adaptive wavelets constructed by means of an adaptive update lifting step. In the literature, one can find several other approaches for building adaptive wavelets [7,8,18,30]. In [26], some of these approaches, and their drawbacks, have been discussed in more detail. The adaptive update lifting scheme introduced in this paper is general in the sense that it is neither *causal*[1], nor does require any bookkeeping to enable perfect reconstruction.

This paper is a sequel to an earlier paper [26] by two of the authors. In § 2.2, we shall briefly discuss the adaptive scheme introduced in that paper and recall some of the main results derived there. In Section 2 we will present a framework which is much more general than the one in [26], in the sense that it allows more than two subbands, longer filters, and general seminorms in the decision map. In Section 3 we will derive necessary and sufficient conditions for perfect reconstruction. Then we will analyse several different choices for the seminorm, namely the l^1-norm and the l^∞-norm in Section 4, the quadratic seminorm in Section 5 and seminorms using derivative filters in Section 6. In Section 7, we perform

[1]Causality means that the computation of the detail signal at a given location depends 'only' on previously computed detail samples.

various experiments, both in one and two dimensions. Finally, in Section 8 we present our conclusions.

2. General framework for update lifting

In § 2.1 we introduce our adaptive update lifting scheme. This scheme generalises the one introduced by two of us in [26]; we briefly recall this scheme in § 2.2. Then in § 2.3 we discuss two major ingredients of our new scheme, the decision map and the update filters. In § 2.4 we formulate two conditions guaranteeing that the underlying scheme is truly adaptive.

2.1 The adaptive update lifting scheme

Assume we have an $(K+1)$-band filter bank decomposition with inputs $x, y^{(1)}, \dots, y^{(K)}$ (with $K \geq 1$) where $x, y^{(1)}, \dots, y^{(K)}$ generally represents polyphase components of the analyzed signal. The first signal x will be updated in order to obtain an approximation signal whereas $y^{(1)}, \dots, y^{(K)}$ will be further predicted so as to generate detail coefficients. Consider an adaptive update lifting scheme as depicted in Fig. 2. In this scheme D is a decision map

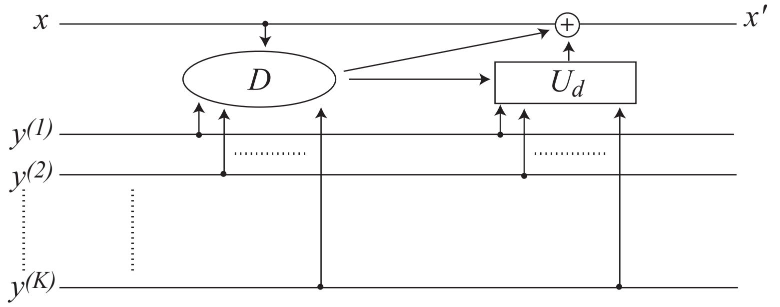

Figure 2: *K-band adaptive update lifting scheme.*

which uses inputs from all bands, i.e., $D = D(x, y^{(1)}, \dots, y^{(K)})$, and whose output is a decision parameter d. In this paper we are exclusively concerned with *binary* decision maps where d can only take the value 0 or 1. Note that in our previous paper [26] we have also treated the more general case where d can take values in a continuous interval.

The parameter d governs the choice of the update step. More precisely, if $d_{\boldsymbol{n}}$ is the output of D at location $\boldsymbol{n} \in \mathbb{Z}^d$, then the updated value $x'(\boldsymbol{n})$ is given by

$$x'(\boldsymbol{n}) = x(\boldsymbol{n}) \oplus_{d_{\boldsymbol{n}}} U_{d_{\boldsymbol{n}}}(y^{(1)}, \dots, y^{(K)})(\boldsymbol{n}) . \tag{2.1}$$

In the classical non-adaptive case, when there is no decomposition map and the update filter does not depend on x, the update step is given by

$$x'(\boldsymbol{n}) = x(\boldsymbol{n}) \oplus U(y^{(1)}, \dots, y^{(K)})(\boldsymbol{n}) ,$$

and can be inverted by means of

$$x(\boldsymbol{n}) = x'(\boldsymbol{n}) \ominus U(y^{(1)}, \dots, y^{(K)})(\boldsymbol{n}) ,$$

where $\ominus$ is the 'subtraction' which inverts the addition $\oplus$. In the adaptive case considered here, however, we need to know $d_{\boldsymbol{n}}$ at every location $\boldsymbol{n}$ to get perfect reconstruction. Since $d_{\boldsymbol{n}} = D(x, y^{(1)}, \dots, y^{(K)})(\boldsymbol{n})$ requires the original input signal x, which is not available at synthesis, recovery of $d_{\boldsymbol{n}}$ is an impossible task in most cases. However, under some special circumstances it is possible to recover $d_{\boldsymbol{n}}$ from x' and $y^{(1)}, \ldots, y^{(K)}$ by means of a so-called *posterior decision map* D'. Obviously, this map needs to satisfy

$$D'(x', y^{(1)}, \ldots, y^{(K)}) = D(x, y^{(1)}, \ldots, y^{(K)}),$$

for all inputs $x, y^{(1)}, \ldots, y^{(K)}$, with x' given by (2.1). Henceforth we assume that the value $d_{\boldsymbol{n}} = D(x, y^{(1)}, \ldots, y^{(K)})(\boldsymbol{n})$ depends on local information. In this paper, it will be assumed that it depends on the gradient vector determined by the values $x(\boldsymbol{n}) - y^{(p)}(\boldsymbol{n} + \boldsymbol{l})$, where $p = 1, \ldots, K$ and $\boldsymbol{l} \in L$; here $L \subseteq \mathbb{R}^d$ is a finite window around the origin. We present two examples to illustrate these concepts.

2.1 Example. First we consider filter banks for one-dimensional signals with one approximation band x and one detail band y. Thus $K = 1$ and we omit the superindex of y in this case. Assume that samples $x_0(2n), x_0(2n+1)$ of the original signal correspond with samples $x(n), y(n)$, respectively. In our paper [26] (see also the previous section) we assumed that d_n depends on the gradient vector with components $x(n) - y(n-1)$ and $x(n) - y(n)$. Thus, $L = \{-1, 0\}$ in this case.

2.2 Example. Next, we consider two-dimensional signals as depicted in Fig. 3, and we are interested in decompositions with $K = 3$ corresponding with a square (i.e., 2×2) sampling structure. The geometrical interpretation of the three last band signals is as follows (see

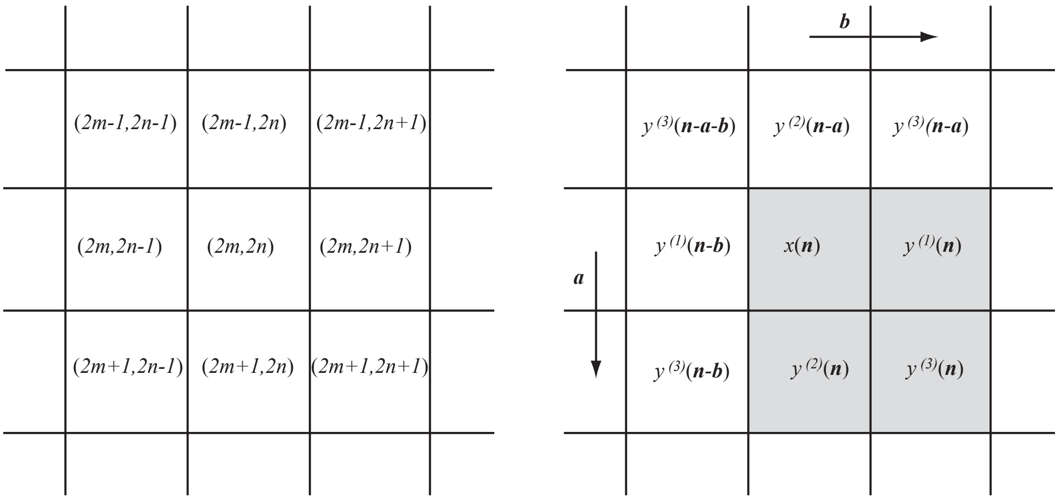

Figure 3: *Left: coordinatisation for two-dimensional signals. Right: location of the input signals x and $y^{(1)}, y^{(2)}, y^{(3)}$ after square sampling.*

also the right diagram in Fig. 3): after prediction, $y^{(1)}, y^{(2)}$ will represent the detail bands capturing *vertical* and *horizontal* details, respectively. The interpretation of $y^{(3)}$ is somewhat less intuitive. After prediction, it leads to what is sometimes called the *diagonal* detail band.

Let us assume that the decision $d_{\boldsymbol{n}}$ depends on the gradient vector associated with its 8 horizontal, vertical and diagonal neighbours. This involves, besides $x(\boldsymbol{n})$ itself, the samples (starting at the east and rotating counter-clockwise): $y^{(1)}(\boldsymbol{n})$, $y^{(3)}(\boldsymbol{n} - \boldsymbol{a})$, $y^{(2)}(\boldsymbol{n} - \boldsymbol{a})$, $y^{(3)}(\boldsymbol{n} - \boldsymbol{a} - \boldsymbol{b})$, $y^{(1)}(\boldsymbol{n} - \boldsymbol{b})$, $y^{(3)}(\boldsymbol{n} - \boldsymbol{b})$, $y^{(2)}(\boldsymbol{n})$, $y^{(3)}(\boldsymbol{n})$. Therefore $L = \{\boldsymbol{0}, -\boldsymbol{a}, -\boldsymbol{b}, -\boldsymbol{a} - \boldsymbol{b}\}$. Here $\boldsymbol{a}, \boldsymbol{b}$ are the unit row and column vectors $(1, 0)$ and $(0, 1)$.

We introduce some additional notation. Assume that the decision map at sample $\boldsymbol{n}$ depends only on the values $x(\boldsymbol{n}) - y^{(p_j)}(\boldsymbol{n} + \boldsymbol{l}_j)$ for $j = 1, \ldots, N$, with $p_j \in \{1, \ldots, K\}$ and $\boldsymbol{l}_j \in L$.

Obviously, $N \le K \cdot |L|$, where $|L|$ is the number of elements contained within window L. We define

$$y_j(\boldsymbol{n}) = y^{(p_j)}(\boldsymbol{n} + \boldsymbol{l}_j), \quad j = 1, \dots, N .$$

Note that we have some freedom in labeling the values $y^{(p)}(\boldsymbol{n} + \boldsymbol{l})$ by j. Fortunately, the specific choice of the labeling is of no importance. In Example 2.1(b), choosing a counter-clockwise labeling direction, we get $y_1(\boldsymbol{n}) = y^{(1)}(\boldsymbol{n})$, $y_2(\boldsymbol{n}) = y^{(3)}(\boldsymbol{n} - \boldsymbol{a})$, $y_3(\boldsymbol{n}) = y^{(2)}(\boldsymbol{n} - \boldsymbol{a})$, etc.

2.2 Previous results

In our previous work [26] we have been dealing exclusively with the one-dimensional case as discussed in Example 2.1(a). We have considered the same adaptive update lifting scheme as the one depicted in Fig. 2, but with one detail band only, i.e., $K = 1$. In this case, (2.1) reduces to

$$x'(n) = x(n) \oplus_{d_n} U_{d_n}(y)(n) ,$$

where $d_n = D(x, y)(n)$. In [26] we assume that the decision map D is of the form

$$D(x, y)(n) = q\big(|x(n) - y(n-1)| + |x(n) - y(n)|\big) ,$$

for some given function $q : \mathbb{R}_+ \to \mathbb{R}$. Furthermore, we only considered update filters U_d that have two taps:

$$U_d(y)(n) = \lambda_d y(n-1) + \mu_d y(n) .$$

For the function q we have considered two cases: the general case, meaning that the update filter coefficients depend on the l^1-gradient norm in a continuous fashion, and the binary threshold case with

$$q(s) = \begin{cases} 0 & \text{if } s \le T \\ 1 & \text{if } s > T . \end{cases}$$

We have derived necessary and sufficient conditions on the filter coefficients for perfect reconstruction in both cases.

2.3 Choice of decision map and update filter

We define the gradient vector $\boldsymbol{v}(\boldsymbol{n}) = (v_1(\boldsymbol{n}), \dots, v_N(\boldsymbol{n}))^T \in \mathbb{R}^N$ (where 'T' means transposition) by

$$v_j(\boldsymbol{n}) = x(\boldsymbol{n}) - y_j(\boldsymbol{n}), \quad j = 1, \dots, N . \tag{2.2}$$

As we said before, the decision map depends exclusively on the gradient vector $\boldsymbol{v}(\boldsymbol{n})$. Before we can give an explicit expression for the decision map, we need to introduce the concept of seminorm.

2.3 Definition. A function $p : \mathbb{R}^N \to \mathbb{R}_+$ is called a *seminorm* if the following two properties hold:

(*i*) $p(\lambda \boldsymbol{v}) = |\lambda| \cdot p(\boldsymbol{v}), \quad \boldsymbol{v} \in \mathbb{R}^N,\ \lambda \in \mathbb{R}$

(*ii*) $p(\boldsymbol{v}_1 + \boldsymbol{v}_2) \le p(\boldsymbol{v}_1) + p(\boldsymbol{v}_2), \quad \boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^N.$

This last inequality is called the *triangle inequality*.

A large class of seminorms on $\mathbb{R}^N$ is given by the expression

$$p(\boldsymbol{v}) = \Big(\sum_{i=1}^{I} |\boldsymbol{a}_i^T \boldsymbol{v}|^q\Big)^{1/q}, \tag{2.3}$$

where $\boldsymbol{a}_i \in \mathbb{R}^N,\ i = 1, \dots, I$ and $q \geq 1$. By $\boldsymbol{a}^T\boldsymbol{v}$ we mean the inner product of the vectors $\boldsymbol{a}$ and $\boldsymbol{v}$.

The seminorms given by

$$p(\boldsymbol{v}) = \left(\boldsymbol{v}^T M \boldsymbol{v}\right)^{1/2}, \tag{2.4}$$

where M is a symmetric positive semi-definite matrix, are called *quadratic seminorms*. It is not difficult to show that they belong to the family given by (2.3) with $q = 2$. Indeed, if M is a symmetric positive semi-definite matrix, we can write [23]:

$$M = \sum_{i=1}^{N} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T,\ \lambda_i \geq 0,$$

where $\{\lambda_i \mid 1 \leq i \leq N\}$ are the eigenvalues of M and $\{\boldsymbol{u}_i \mid 1 \leq i \leq N\}$ are the (orthonormal) eigenvectors of M. The expression (2.4) becomes:

$$p(\boldsymbol{v}) = \left|\boldsymbol{v}^T \Big(\sum_{i=1}^{N} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T\Big) \boldsymbol{v}\right|^{1/2} = \left(\sum_{i=1}^{N} \lambda_i |\boldsymbol{u}_i^T \boldsymbol{v}|^2\right)^{1/2}.$$

Now, if we take $\boldsymbol{a}_i = \sqrt{\lambda_i}\boldsymbol{u}_i$, we get (2.3), with $q = 2$ and $I = N$.

Recall that p is a norm if, in addition to (i)-(ii) in Definition 2.3, it satisfies $p(\boldsymbol{v}) = 0$ iff $\boldsymbol{v} = \boldsymbol{0}$. Obviously, every norm is a seminorm but not vice-versa. In particular, the seminorm given in (2.4) is a norm when M is a symmetric positive definite matrix.

In this paper we deal exclusively with binary decision maps of the form

$$D(x, y^{(1)}, \dots, y^{(K)})(\boldsymbol{n}) = \begin{cases} 1, & \text{if } p(\boldsymbol{v}(\boldsymbol{n})) > T \\ 0, & \text{if } p(\boldsymbol{v}(\boldsymbol{n})) \leq T, \end{cases} \tag{2.5}$$

where $\boldsymbol{v}(\boldsymbol{n})$ is given by (2.2), p is a seminorm, and $T > 0$ is a given threshold. Instead of (2.5) we may also use the shorthand notation

$$D(x, y^{(1)}, \dots, y^{(K)})(\boldsymbol{n}) = [\,p(\boldsymbol{v}(\boldsymbol{n})) > T\,], \tag{2.6}$$

where $[P]$ returns 1 if the predicate P is true, and 0 if it is false.

In the update step given by (2.1) we need to specify the 'addition' $\oplus_d$ as well as the update filter $U_d(y^{(1)}, \dots, y^{(K)})(\boldsymbol{n})$ for the values $d = 0, 1$. Henceforth we assume that the addition $\oplus_d$ is of the form

$$x \oplus_d u = \alpha_d(x + u), \tag{2.7}$$

with $\alpha_d \neq 0$. Such a choice means in particular that the operation $\oplus_d$ is invertible.

The update filter is taken to be of the form

$$U_{d_n}(y^{(1)}, \dots, y^{(K)})(\boldsymbol{n}) = \sum_{j=1}^{N} \lambda_{d_n,j} y_j(\boldsymbol{n}), \tag{2.8}$$

i.e., it is a linear combination of the values of the last N polyphase components inside the window L located at $\boldsymbol{n}$. The filter coefficients $\lambda_{d_{\boldsymbol{n}},j}$ depend on the decision $d_{\boldsymbol{n}}$ given by (2.6). Combination of (2.1), (2.7), and (2.8) yields that

$$x'(\boldsymbol{n}) = \alpha_{d_{\boldsymbol{n}}} x(\boldsymbol{n}) + \sum_{j=1}^{N} \beta_{d_{\boldsymbol{n}},j} y_j(\boldsymbol{n}) \,, \tag{2.9}$$

where

$$\beta_{d,j} = \alpha_d \lambda_{d,j} \,.$$

2.4 Adaptivity conditions

In the next section we will examine the question under which assumptions the lifting framework discussed previously is invertible. In other words, we show how to recover x from x' given by (2.9) and the original signals $y^{(1)}, \ldots, y^{(K)}$. If such an inversion is possible, then we say that the *perfect reconstruction condition* holds. Obviously, we can easily invert (2.9):

$$x(\boldsymbol{n}) = \frac{1}{\alpha_{d_{\boldsymbol{n}}}} \big(x'(\boldsymbol{n}) - \sum_{j=1}^{N} \beta_{d_{\boldsymbol{n}},j} y_j(\boldsymbol{n}) \big) \,, \tag{2.10}$$

presumed that the decision $d_{\boldsymbol{n}}$ is known at every location $\boldsymbol{n}$.

Not every seminorm can be used to model an adaptive scheme. For example, if p depends only on differences $v_i - v_j$, then the threshold criterion in (2.5) is independent of the value of $x(\boldsymbol{n})$, as can easily be seen by using (2.2). A simple condition on p which is necessary and sufficient for the adaptivity of the corresponding scheme is

$$p(\boldsymbol{u}) > 0 \,,$$

where $\boldsymbol{u} = (1, \ldots, 1)^T$ is a vector of length N. Indeed, it is easy to check that the condition $p(\boldsymbol{u}) = 0$ is equivalent to the condition

$$p(\boldsymbol{v} + \lambda \boldsymbol{u}) = p(\boldsymbol{v}), \;\; \boldsymbol{v} \in \mathbb{R}^N, \;\; \lambda \in \mathbb{R} \,.$$

Observe that the addition of λ to $x(\boldsymbol{n})$ while keeping all $y_j(\boldsymbol{n})$ constant, amounts to the addition of $\lambda \boldsymbol{u}$ to the gradient vector $\boldsymbol{v}(\boldsymbol{n})$. If such an addition does not affect the seminorm, then the corresponding decision criterion does not depend on $x(\boldsymbol{n})$, and hence the scheme is non-adaptive.

Adaptivity Condition for the Seminorm. The seminorm p on $\mathbb{R}^N$ satisfies

$$p(\boldsymbol{u}) > 0 \,, \tag{2.11}$$

where $\boldsymbol{u} = (1, \ldots, 1)^T$ is a vector of length N.

Obviously, we need a second condition to guarantee true adaptivity, namely that the update filters for $d = 0$ and $d = 1$ are different.

Adaptivity Condition for the Update Filters. The update filters for $d = 0$ and $d = 1$ do not coincide, i.e, $\beta_{0,j} \neq \beta_{1,j}$ for at least one j.

Henceforth we restrict ourselves to the case where both adaptivity conditions are satisfied. Define the value

$$\kappa_d = \alpha_d + \sum_{j=1}^{N} \beta_{d,j}, \qquad d = 0, 1 \,. \tag{2.12}$$

We have the following result.

2.4 Proposition. *Assume that p satisfies the adaptivity condition in* (2.11)*. A necessary condition for perfect reconstruction is $\kappa_0 = \kappa_1$.*

Proof. Assume that $\kappa_0 \neq \kappa_1$. Let $\xi \in \mathbb{R}$ be such that

$$|(\kappa_0 - \kappa_1)\xi| > \frac{\alpha_1 T}{p(\boldsymbol{u})} . \tag{2.13}$$

Let $\boldsymbol{n}$ be a given location and assume that $x(\boldsymbol{n}) = \xi$ and $y_k(\boldsymbol{n}) = \xi$ for $k = 1, \dots, N$. Obviously, $\boldsymbol{v}(\boldsymbol{n}) = \boldsymbol{0}$ hence $d_{\boldsymbol{n}} = 0$. It follows immediately that (2.9) gives $x'(\boldsymbol{n}) = \kappa_0\xi$. However, if we take $x(\boldsymbol{n}) = \xi + \eta$ and the same $y_k(\boldsymbol{n})$ as before, then $\boldsymbol{v}(\boldsymbol{n}) = \eta\boldsymbol{u}$. Therefore, if $|\eta| > T/p(\boldsymbol{u})$, then $d_{\boldsymbol{n}} = 1$ and we deduce that $x'(\boldsymbol{n}) = \kappa_1\xi + \alpha_1\eta$. If we choose $\eta = (\kappa_0 - \kappa_1)\xi/\alpha_1$, then, because of (2.13) the condition $|\eta| > T/p(\boldsymbol{u})$ is satisfied. For this particular choice, however, $\kappa_1\xi + \alpha_1\eta = \kappa_0\xi$. Thus we have shown that for the same values of $y_k(\boldsymbol{n})$, two different inputs for $x(\boldsymbol{n})$ may yield the same output. Obviously, perfect reconstruction is out of reach in such a case. $\square$

Throughout the remainder of this paper we normalise the filter coefficients so that

$$\kappa_0 = \kappa_1 = 1 . \tag{2.14}$$

Obviously, such a normalisation is possible only in the case where $\kappa_d \neq 0$. A system with $\kappa_d = 0$ would, in general, correspond to a prediction operator (i.e., "high-pass" filter), while the condition $\kappa_d \neq 0$ is more appropriate for update operators (corresponding to "low-pass" filters).

Unfortunately, the condition in (2.14) is far from being a sufficient condition for perfect reconstruction. In the following subsection we will be concerned with the derivation of sufficient conditions for perfect reconstruction.

To simplify notation, we will henceforth omit the argument $\boldsymbol{n}$ in our notation. Thus we write x, y_j instead of $x(\boldsymbol{n}), y_j(\boldsymbol{n})$, respectively, and $\boldsymbol{v} = (v_1, \dots, v_N)^T$ instead of $\boldsymbol{v}(\boldsymbol{n}) = (v_1(\boldsymbol{n}), \dots, v_N(\boldsymbol{n}))^T$. Now, the update lifting step in (2.9) can be written as

$$x' = \alpha_d x + \sum_{j=1}^{N} \beta_{d,j} y_j , \tag{2.15}$$

and the inversion in (2.10) reduces to

$$x = \frac{1}{\alpha_d}\big(x' - \sum_{j=1}^{N} \beta_{d,j} y_j\big) . \tag{2.16}$$

Subtraction of y_i at both sides of (2.15) yields

$$v_i' = (1 - \beta_{d,i})v_i - \sum_{j \neq i} \beta_{d,j} v_j , \tag{2.17}$$

where

$$v_i' = x' - y_i , \qquad i = 1, \dots, N . \tag{2.18}$$

Define the $N \times N$-matrix A_d by the right hand-side expression in (2.17), i.e.,

$$A_d = \begin{pmatrix} 1 - \beta_{d,1} & -\beta_{d,2} & -\beta_{d,3} & \dots & -\beta_{d,N} \\ -\beta_{d,1} & 1 - \beta_{d,2} & -\beta_{d,3} & \dots & \vdots \\ -\beta_{d,1} & -\beta_{d,2} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ -\beta_{d,1} & -\beta_{d,2} & -\beta_{d,3} & \dots & 1 - \beta_{d,N} \end{pmatrix} . \tag{2.19}$$

The adaptive update lifting step is described therefore by:

$$\begin{cases} \boldsymbol{v}' = A_d \boldsymbol{v} \\ d = [p(\boldsymbol{v}) > T] \,, \end{cases} \tag{2.20}$$

where $p : \mathbb{R}^N \to \mathbb{R}_+$ is a given seminorm satisfying the adaptivity condition (2.11).

3. When do we have perfect reconstruction?

In this section we will formulate conditions on the seminorm and the update filters which guarantee perfect reconstruction; see § 3.2. As a preparatory step, we will 'translate' the perfect reconstruction condition into an equivalent condition called the *Threshold Criterion*, stated in terms of the seminorm.

3.1 Seminorms and the threshold criterion

Before we formulate a necessary and sufficient condition for perfect reconstruction, we introduce some notions that we need in the sequel.

Let V be a vector space with seminorm p. For a linear operator $A : V \to V$ we define the *operator seminorm* $p(A)$ and the *inverse operator seminorm* $p^{-1}(A)$ as

$$\begin{aligned} p(A) &= \sup\{p(A\boldsymbol{v}) \mid \boldsymbol{v} \in V \text{ and } p(\boldsymbol{v}) = 1\} \\ p^{-1}(A) &= \sup\{p(\boldsymbol{v}) \mid \boldsymbol{v} \in V \text{ and } p(A\boldsymbol{v}) = 1\} \,. \end{aligned}$$

In the last expression we use the convention that $p^{-1}(A) = \infty$ if $p(A\boldsymbol{v}) = 0$ for all $\boldsymbol{v} \in V$, unless p is identically zero, in which case both $p(A)$ and $p^{-1}(A)$ are zero. Throughout this paper, we will discard the case where p is identically zero and, consequently, we will always have $p^{-1}(A) > 0$.

We list some properties of these two notions in the following proposition.

3.1 Proposition. *Let V be a Hilbert space, let $p : V \to \mathbb{R}_+$ be a seminorm and $A : V \to V$ be a bounded linear operator.*

(*a*) $p^{-1}(A) = p(A^{-1})$ *if A is invertible.*

(*b*) *The following two conditions are equivalent*

 (*i*) $p(A) < \infty$

 (*ii*) $p(\boldsymbol{v}) = 0$ *implies* $p(A\boldsymbol{v}) = 0$ *for* $\boldsymbol{v} \in V$

(*c*) *The following two conditions are also equivalent*

 (*i*) $p^{-1}(A) < \infty$

 (*ii*) $p(A\boldsymbol{v}) = 0$ *implies* $p(\boldsymbol{v}) = 0$ *for* $\boldsymbol{v} \in V$

(*d*) $p(A\boldsymbol{v}) \leq p(A)p(\boldsymbol{v})$ *if* $p(\boldsymbol{v}) \neq 0$.

(*e*) $p(\boldsymbol{v}) \leq p^{-1}(A)p(A\boldsymbol{v})$ *if* $p(A\boldsymbol{v}) \neq 0$.

Proof. The proofs of (a), (d) and (e) are straightforward. We prove (b) and (c).

(b): Assume (i), that is $p(A) < \infty$. Now suppose that there exists a $\boldsymbol{v} \in V$ such that $p(\boldsymbol{v}) = 0$ and $p(A\boldsymbol{v}) \neq 0$. We show that this gives rise to a contradiction. Fix a vector $\boldsymbol{w}$ with $p(\boldsymbol{w}) = 1$. Obviously,

$$p(\lambda\boldsymbol{v} + \boldsymbol{w}) \leq |\lambda| p(\boldsymbol{v}) + p(\boldsymbol{w}) = 1 \,,$$

and also

$$1 = p(\boldsymbol{w}) \leq p(\lambda\boldsymbol{v} + \boldsymbol{w}) + p(-\lambda\boldsymbol{v}) = p(\lambda\boldsymbol{v} + \boldsymbol{w}) \,,$$

which means that

$$p(\lambda \boldsymbol{v} + \boldsymbol{w}) = 1 \text{ for every } \lambda \in \mathbb{R}\,.$$

By definition,

$$p(A) \geq p(A(\lambda \boldsymbol{v} + \boldsymbol{w})) \geq p(\lambda A\boldsymbol{v}) - p(A\boldsymbol{w}) = |\lambda| p(A\boldsymbol{v}) - p(A\boldsymbol{w})\,.$$

Letting $|\lambda| \to \infty$, we arrive at the conclusion that $p(A) = \infty$, a contradiction.

Assume, on the other hand, that (ii) holds. Define $V_0 \subseteq V$ as $V_0 = \{\boldsymbol{v} \in V \mid p(\boldsymbol{v}) = 0\}$ and $V_1 = V_0^\perp$. It is easy to see that for any $\boldsymbol{v} \in V$ we have $p(\boldsymbol{v}) = p(\boldsymbol{v}_1)$ where $\boldsymbol{v}_1$ is the projection of $\boldsymbol{v}$ on V_1. Obviously, p defines a norm on the closed subspace V_1. The decomposition of V into V_0 and V_1 gives rise to a decomposition of the operator A into A_{ij} where A_{ij} maps V_i into V_j, for $i, j = 0, 1$. Thus we can write

$$A\boldsymbol{v} = (A_{00}\boldsymbol{v}_0 + A_{01}\boldsymbol{v}_1) + (A_{10}\boldsymbol{v}_0 + A_{11}\boldsymbol{v}_1)\,,$$

where the first and second expression between brackets lies in V_0 and V_1, respectively. The condition in (ii) obviously means that $A_{10} = 0$. It is then evident that

$$p(A) = \sup\{p(A_{11}\boldsymbol{v}_1) \mid p(\boldsymbol{v}_1) = 1\}\,,$$

and this coincides with the norm of A_{11} on V_1 which, by definition, is finite. This proves (b).

(c): This proof is very similar to that of (b). In the second part of the proof where it has to be shown that $p^{-1}(A) < \infty$, it is found that $A_{10} = 0$, A_{11} is invertible, and $p^{-1}(A) = p(A_{11}^{-1})$, which is finite.

$\square$

We return to the update lifting step described in the previous section. If $p(\boldsymbol{v}) \leq T$ at the analysis step, then the decision equals $d = 0$ and $\boldsymbol{v}' = A_0\boldsymbol{v}$. If, on the other hand, $p(\boldsymbol{v}) > T$, then $d = 1$ and $\boldsymbol{v}' = A_1\boldsymbol{v}$. To have perfect reconstruction we must be able to recover the decision d from the transformed gradient vector $\boldsymbol{v}'$. For simplicity we shall restrict ourselves to the case where d can be recovered by thresholding the seminorm $p(\boldsymbol{v}')$, i.e., the case that

$$d = [p(\boldsymbol{v}) > T] = [p(\boldsymbol{v}') > T']\,,$$

for some $T' > 0$. We formalize this condition in the following criterion.

Threshold Criterion. Given a threshold $T > 0$, there exists a (possibly different) threshold $T' > 0$ such that

- (i) if $p(\boldsymbol{v}) \leq T$ then $p(A_0\boldsymbol{v}) \leq T'$;
- (ii) if $p(\boldsymbol{v}) > T$ then $p(A_1\boldsymbol{v}) > T'$.

The following result is obvious.

3.2 Proposition. *If the threshold criterion holds then we have perfect reconstruction.*

The corresponding reconstruction algorithm is straightforward:

1. compute $\boldsymbol{v}'$ from (2.18);
2. if $p(\boldsymbol{v}') \leq T'$ then $d = 0$, otherwise $d = 1$;
3. compute x from (2.16).

Thus it remains to verify the validity of the threshold criterion. The following result, which we consider to be the main result of this section, provides necessary and sufficient conditions.

3.3 Proposition. *The threshold criterion holds if and only if the following three conditions are satisfied:*

$$p(A_0) < \infty \text{ and } p^{-1}(A_1) < \infty \tag{3.1}$$

$$p(A_0)p^{-1}(A_1) \le 1\,. \tag{3.2}$$

Proof. 'if': put $T' = p(A_0)T$; we show that the threshold criterion holds. To prove (i), assume that $p(\boldsymbol{v}) \le T$. If $p(\boldsymbol{v}) = 0$, then $p(A_0\boldsymbol{v}) = 0$ by (3.1) and Proposition 3.1(b). If $p(\boldsymbol{v}) > 0$, then we get from Proposition 3.1(d) that

$$p(A_0\boldsymbol{v}) \le p(A_0)p(\boldsymbol{v}) \le p(A_0)T = T'\,.$$

To prove (ii) assume that $p(\boldsymbol{v}) > T$. From the fact that $p^{-1}(A_1) < \infty$ and Proposition 3.1(c) we conclude that $p(A_1\boldsymbol{v}) \ne 0$ and we get from Proposition 3.1(e) that $p(\boldsymbol{v}) \le p^{-1}(A_1)p(A_1\boldsymbol{v})$. In combination with (3.2), this gives us

$$p(A_1\boldsymbol{v}) \ge \frac{p(\boldsymbol{v})}{p^{-1}(A_1)} \ge p(A_0)p(\boldsymbol{v}) > p(A_0)T = T'\,.$$

This concludes the proof of the 'if'-part.

'only if': to prove that $p(A_0) < \infty$, assume that $p(\boldsymbol{v}) = 0$ and $p(A_0\boldsymbol{v}) \ne 0$. We show that this will give rise to a contradiction. Choosing $\lambda > T'/p(A_0\boldsymbol{v})$ we have $p(A_0(\lambda\boldsymbol{v})) = |\lambda|\cdot p(A_0\boldsymbol{v}) > T'$. However $p(\lambda\boldsymbol{v}) = |\lambda|\cdot p(\boldsymbol{v}) = 0$, and we have a contradiction with (i) of the threshold criterion. The fact that $p^{-1}(A_1) < \infty$ is proved analogously. Thus it remains to prove (3.2). Choose $T = 1$ and let T' be the corresponding threshold given by the threshold criterion. We derive from (i) that $p(A_0) \le T'$. Now (ii) reads as follows: if $p(\boldsymbol{v}) > 1$ then $p(A_1\boldsymbol{v}) > T'$. Now suppose that (3.2) does not hold, i.e., $p(A_0)p^{-1}(A_1) > 1$, or equivalently, $p^{-1}(A_1) > (p(A_0))^{-1}$ ($p(A_0) \ne 0$, otherwise $p^{-1}(A_1)$ should be infinite). From the definition of $p^{-1}(A_1)$ it follows that there must be a vector $\boldsymbol{v} \in \mathbb{R}^N$ with $p(A_1\boldsymbol{v}) = 1$ and $p(\boldsymbol{v}) > (p(A_0))^{-1}$. Putting $\boldsymbol{v}' = p(A_0)\boldsymbol{v}$ we get $p(\boldsymbol{v}') > 1$ and $p(A_1\boldsymbol{v}') = p(A_0) \le T'$ which contradicts (ii) of the threshold criterion. Therefore, (3.2) must hold. □

Remarks.

- Without loss of generality, the threshold $T > 0$ could be normalized to 1 by redefining the semi-norm as p/T.
- The proof of the above proposition shows that it is sufficient to choose $T' = p(A_0)T$.

3.2 Necessary conditions for perfect reconstruction

Recall from § 2.3 that $\alpha_d \ne 0$ for $d = 0, 1$. Before specialising to certain classes of seminorms, we prove some general results related to the specific form of the linear operators under consideration. The matrix A_d in (2.19) can also be written as

$$A_d = I - \boldsymbol{u}\boldsymbol{\beta}_d^T, \tag{3.3}$$

where I is the $N \times N$ identity matrix, $\boldsymbol{u} = (1, \dots, 1)^T$ and $\boldsymbol{\beta}_d = (\beta_{d,1}, \dots, \beta_{d,N})^T$ are column vectors, both of length N. For its determinant we find, after simple algebra manipulations,

$$\det(A_d) = 1 - \boldsymbol{u}^T\boldsymbol{\beta}_d = 1 - \sum_{j=1}^{N}\beta_{d,j} = \alpha_d\,,$$

where we have used (2.12). Since we have assumed that $\alpha_d \neq 0$ for $d = 0, 1$, we may conclude that A_d is invertible. Moreover, one can easily show that

$$A_d^{-1} = I + \frac{1}{\alpha_d}\boldsymbol{u}\boldsymbol{\beta}_d^T = \begin{pmatrix} 1 + \frac{\beta_{d,1}}{\alpha_d} & \frac{\beta_{d,2}}{\alpha_d} & \frac{\beta_{d,3}}{\alpha_d} & \cdots & \frac{\beta_{d,N}}{\alpha_d} \\ \frac{\beta_{d,1}}{\alpha_d} & 1 + \frac{\beta_{d,2}}{\alpha_d} & \frac{\beta_{d,3}}{\alpha_d} & \cdots & \vdots \\ \vdots & \frac{\beta_{d,2}}{\alpha_d} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \frac{\beta_{d,1}}{\alpha_d} & \cdots & \cdots & \cdots & 1 + \frac{\beta_{d,N}}{\alpha_d} \end{pmatrix}.$$

Putting

$$\boldsymbol{\beta}'_d = -\boldsymbol{\beta}_d / \alpha_d$$

we find that A_d^{-1} takes a form similar to that of A_d :

$$A_d^{-1} = I - \boldsymbol{u}\boldsymbol{\beta}_d'^T .$$

We start with the following auxiliary result.

3.4 Proposition. *Let p be a seminorm and let V_0 be the "kernel" of p, i.e., the linear subspace of $\mathbb{R}^N$ given by*

$$V_0 = \{\boldsymbol{v} \in \mathbb{R}^N \mid p(\boldsymbol{v}) = 0\} .$$

Then $p(A_d) < \infty$ if and only if $\boldsymbol{\beta}_d \in V_0^\perp$.

Proof. 'if': assume that $\boldsymbol{\beta}_d \in V_0^\perp$. Following Proposition 3.1 we must show that $p(\boldsymbol{v}) = 0$ implies that $p(A_d\boldsymbol{v}) = 0$. If $p(\boldsymbol{v}) = 0$ then $\boldsymbol{v} \in V_0$ hence $\boldsymbol{\beta}_d^T\boldsymbol{v} = 0$. This implies that $A_d\boldsymbol{v} = \boldsymbol{v} - \boldsymbol{u}\boldsymbol{\beta}_d^T\boldsymbol{v} = \boldsymbol{v}$ and hence that $p(A_d\boldsymbol{v}) = 0$.

'only if': assume that $p(A_d) < \infty$ and $\boldsymbol{\beta}_d \notin V_0^\perp$. Thus there is a $\boldsymbol{v} \in V_0$ with $\boldsymbol{\beta}_d^T\boldsymbol{v} = 1$. Then $A_d\boldsymbol{v} = \boldsymbol{v} - \boldsymbol{u}\boldsymbol{\beta}_d^T\boldsymbol{v} = \boldsymbol{v} - \boldsymbol{u}$. Since p is a seminorm $p(\boldsymbol{u}) \neq 0$ (see (2.11)), we have $0 \neq p(\boldsymbol{u}) \leq p(\boldsymbol{u} - \boldsymbol{v}) + p(\boldsymbol{v}) = p(\boldsymbol{u} - \boldsymbol{v})$, and therefore $p(\boldsymbol{u} - \boldsymbol{v}) = p(\boldsymbol{v} - \boldsymbol{u}) = p(A_d\boldsymbol{v}) \neq 0$. Since $p(\boldsymbol{v}) = 0$ we conclude from Proposition 3.1 that $p(A_d) = \infty$, a contradiction. This concludes the proof. □

We now investigate the eigenvalue problem $A_d\boldsymbol{v} = \lambda\boldsymbol{v}$. This can be written as $\boldsymbol{v} - \boldsymbol{u}\boldsymbol{\beta}_d^T\boldsymbol{v} = \lambda\boldsymbol{v}$. We have to distinguish the cases $\lambda = 1$ and $\lambda \neq 1$. If $\lambda = 1$ we find $\boldsymbol{\beta}_d^T\boldsymbol{v} = 0$ and from $\lambda \neq 1$ we get that $\boldsymbol{v}$ is a multiple of $\boldsymbol{u}$. Thus we arrive at the following result.

3.5 Lemma. *(a) If $\alpha_d = 1$ then A_d has only one eigenvalue $\lambda = 1$; the eigenspace is the hyperplane $\boldsymbol{\beta}_d^T\boldsymbol{v} = 0$.*
(b) If $\alpha_d \neq 1$ then A_d has eigenvalues $1, \alpha_d$. The eigenspace associated with eigenvalue $\lambda = 1$ is the hyperplane $\boldsymbol{\beta}_d^T\boldsymbol{v} = 0$, and the eigenvector associated with $\lambda = \alpha_d$ is $\boldsymbol{u}$.

Note that in both the cases (*a*) and (*b*) we have $A_d\boldsymbol{u} = \alpha_d\boldsymbol{u}$.

Using that $p(\boldsymbol{u}) > 0$ (see (2.11)), we get that

$$\begin{aligned} p(A_d) &\geq p(A_d\boldsymbol{u})/p(\boldsymbol{u}) = |\alpha_d| \\ p^{-1}(A_d) &\geq p(\boldsymbol{u})/p(A_d\boldsymbol{u}) = |\alpha_d|^{-1} . \end{aligned}$$

On the other hand, if there exists a $\boldsymbol{v}$ with $\boldsymbol{\beta}_d^T\boldsymbol{v} = 0$ and $p(\boldsymbol{v}) \neq 0$ then

$$p(A_d) \geq 1 \text{ and } p^{-1}(A_d) \geq 1 . \tag{3.4}$$

Thus we arrive at the following necessary conditions for perfect reconstruction.

3.6 Proposition. *Assume that* $p(\boldsymbol{u}) \neq 0$.
(*a*) *The threshold criterion can only be satisfied if* $|\alpha_0| \leq |\alpha_1|$.
(*b*) *Assume in addition that* $p(\boldsymbol{v}_0) \neq 0$, $p(\boldsymbol{v}_1) \neq 0$ *for some vectors* $\boldsymbol{v}_d$ *with* $\boldsymbol{\beta}_d^T \boldsymbol{v}_d = 0$ *for* $d = 0, 1$, *then the threshold criterion can only be satisfied if* $|\alpha_0| \leq 1 \leq |\alpha_1|$.

Proof. The threshold criterion can only hold if (3.2) is satisfied, that is $p(A_0)p^{-1}(A_1) \leq 1$. If $p(\boldsymbol{u}) \neq 0$, then we have $p(A_0) \geq |\alpha_0|$ and $p^{-1}(A_1) \geq |\alpha_1|^{-1}$. Thus a necessary condition for (3.2) to be satisfied is $|\alpha_0| \cdot |\alpha_1|^{-1} \leq 1$. This proves (*a*).

To prove (*b*), assume that for $d = 0, 1$ we have $p(\boldsymbol{v}_d) \neq 0$ for some $\boldsymbol{v}_d$ with $\boldsymbol{\beta}_d^T \boldsymbol{v}_d = 0$. Since both $p(A_0)$ and $p^{-1}(A_1)$ are ≥ 1 by (3.4), we conclude that $|\alpha_0| \leq 1$ and $|\alpha_1|^{-1} \leq 1$. This concludes the proof. □

Before considering a number of special cases we observe that the problem becomes trivial if $N = 1$. In this case there is, apart from a multiplicative constant, only one seminorm, namely $p(v) = |v|$. Now the threshold criterion holds if and only if $|\alpha_0| \leq |\alpha_1|$. Henceforth we restrict ourselves to the case $N > 1$.

4. The l^1-norm and the l^∞-norm

We return to the specific situation described at the end of Section 2. Recall from (2.20) that the update lifting step is given by

$$\begin{cases} \boldsymbol{v}' = A_d \boldsymbol{v} \\ d = [p(\boldsymbol{v}) > T] \, , \end{cases}$$

where A_d is the matrix in (2.19).

In Section 3 we showed that a sufficient condition for perfect reconstruction is the threshold criterion, i.e. (3.1)-(3.2). In the particular case that p is a norm, (3.1) is trivially satisfied and the threshold condition reduces to (3.2). In the remainder of this subsection we concentrate on the case where p is the l^1-norm

$$p_1(\boldsymbol{v}) = \sum_{j=1}^{N} |v_j| \, ,$$

or the l^∞-norm

$$p_\infty(\boldsymbol{v}) = \max_{j=1,\ldots,N} |v_j| \, .$$

Observe that the condition $p(\boldsymbol{u}) \neq 0$ is satisfied in both cases. Furthermore, we assume that $N > 1$.

4.1 Proposition. *If* $p = p_1$, *then the threshold criterion holds if and only if* $N = 2$, $\beta_{0,1}, \beta_{0,2} \in [0, 1]$ *and either* $\beta_{1,1}, \beta_{1,2} \leq 0$ *or* $\beta_{1,1}, \beta_{1,2} \geq 1$.

Proof. We are interested in the cases where the threshold criterion holds, or equivalently, where $p(A_0)p^{-1}(A_1) \leq 1$ is satisfied. The l^1-norm of the matrix A_d is given by [1]

$$p_1(A_d) = \max_j \left(|1 - \beta_{d,j}| + (N-1)|\beta_{d,j}|\right) ,$$

and the norm of its inverse is

$$p_1^{-1}(A_d) = p_1(A_d^{-1}) = \max_j \left(|1 + \frac{\beta_{d,j}}{\alpha_d}| + (N-1)\frac{|\beta_{d,j}|}{|\alpha_d|} \right) .$$

Therefore, the condition $p_1(A_0)p_1^{-1}(A_1) \leq 1$ can be written as

$$\max_j \left(|1 - \beta_{0,j}| + (N-1)|\beta_{0,j}|\right) \cdot \max_j \left(|1 + \frac{\beta_{1,j}}{\alpha_1}| + (N-1)\frac{|\beta_{1,j}|}{|\alpha_1|} \right) \leq 1 \, .$$

Recall that $N \geq 2$. Let us first observe that for any $j = 1, \dots, N$

$$|1-\beta_{0,j}| + (N-1)|\beta_{0,j}| \quad = \quad \begin{cases} 1 + N|\beta_{0,j}| > 1 & \text{if } \beta_{0,j} < 0 \\ 1 + (N-2)|\beta_{0,j}| \geq 1 & \text{if } 0 \leq \beta_{0,j} \leq 1 \\ N|\beta_{0,j}| - 1 > 1 & \text{if } \beta_{0,j} > 1 \end{cases}$$

$$|1+\tfrac{\beta_{1,j}}{\alpha_1}| + (N-1)\tfrac{|\beta_{1,j}|}{|\alpha_1|} \quad = \quad \begin{cases} 1 + N\frac{|\beta_{1,j}|}{|\alpha_1|} > 1 & \text{if } \operatorname{sign}\beta_{1,j} = \operatorname{sign}\alpha_1 \\ 1 + (N-2)\frac{|\beta_{1,j}|}{|\alpha_1|} \geq 1 & \text{if } \operatorname{sign}\beta_{1,j} \neq \operatorname{sign}\alpha_1 \text{ and } |\alpha_1| \geq |\beta_{1,j}| \\ N\frac{|\beta_{1,j}|}{|\alpha_1|} - 1 > 1 & \text{if } \operatorname{sign}\beta_{1,j} \neq \operatorname{sign}\alpha_1 \text{ and } |\alpha_1| < |\beta_{1,j}|. \end{cases}$$

Thus, $p_1(A_0) \geq 1$ and $p_1^{-1}(A_1) \geq 1$. Consequently, condition $p_1(A_0)p_1^{-1}(A_1) \leq 1$ can only be satisfied when $p_1(A_0) = p_1^{-1}(A_1) = 1$. The equality $p_1(A_0) = 1$ implies that for any $j = 1, \dots, N$, either $\beta_{0,j} = 0$ or $N = 2$ and $0 \leq \beta_{0,j} \leq 1$. The equality $p_1^{-1}(A_1) = 1$ means that for any $j = 1, \dots, N$, either $\beta_{1,j} = 0$ or $N = 2$, $\operatorname{sign}\beta_{1,j} \neq \operatorname{sign}\alpha_1$ and $|\alpha_1| \geq |\beta_{1,j}|$. From these implications, Proposition 4.1 follows immediately. □

Next, we consider the l^∞-case. We will see that in this case the conditions on the filter coefficients are slightly more restrictive than in the previous case.

4.2 Proposition. *Assume $p = p_\infty$, then the threshold criterion holds if and only if $N = 2$, $\beta_{0,1} = \beta_{0,2} \in [0,1]$ and either $\beta_{1,1} = \beta_{1,2} \leq 0$ or $\beta_{1,1} = \beta_{1,2} \geq 1$.*

Proof. The l^∞-norm of the matrix A_d in (2.19) is given by

$$p_\infty(A_d) = \max_i \left(|1-\beta_{d,i}| + \sum_{j\neq i} |\beta_{d,j}| \right),$$

and the norm of its inverse is

$$p_\infty^{-1}(A_d) = \max_i \left(|1 + \frac{\beta_{d,i}}{\alpha_d}| + \sum_{j\neq i} \frac{|\beta_{d,j}|}{|\alpha_d|} \right).$$

Recall that $N \geq 2$, and the l^∞-norm of A_0 can be expressed as

$$p_\infty(A_0) = \begin{cases} 1 + \sum_j |\beta_{0,j}| > 1 & \text{if } \beta_{0,j} < 0 \text{ for some } j = 1, \dots, N \\ |1-\beta_{0,m}| + \sum_{j\neq m} |\beta_{0,j}| & \begin{cases} = 1 & \text{if } N = 2 \text{ and } 0 \leq \beta_{0,1} = \beta_{0,2} \leq 1 \\ = 1 & \text{if } \beta_{0,j} = 0 \text{ for all } j = 1, \dots, N \\ > 1 & \text{otherwise}, \end{cases} \end{cases}$$

where $m = \operatorname{argmin}_j \beta_{0,j}$. Likewise, the l^∞-norm of A_1^{-1} is

$$p_\infty(A_1^{-1}) = \begin{cases} 1 + \sum_j \frac{|\beta_{1,j}|}{|\alpha_1|} > 1 & \text{if } \operatorname{sign}\alpha_1 = \operatorname{sign}\beta_{1,j} \text{ for some } j = 1, \dots, N \\ |1 + \frac{\beta_{1,m}}{\alpha_1}| + \sum_{j\neq m} \frac{|\beta_{1,j}|}{|\alpha_1|} & \begin{cases} = 1 & \text{if } N = 2,\ \beta_{1,1} = \beta_{1,2},\ \operatorname{sign}\alpha_1 \neq \operatorname{sign}\beta_{1,j} \text{ and } |\beta_{1,j}| \leq |\alpha_1| \\ = 1 & \text{if } \beta_{1,j} = 0 \text{ for all } j = 1, \dots, N \\ > 1 & \text{otherwise}, \end{cases} \end{cases}$$

where $m = \operatorname{argmin}_j \frac{\beta_{1,j}}{\alpha_1}$. Thus, both $p_\infty(A_0)$ and $p_\infty(A_1^{-1})$ values are at least 1, which means that condition $p_\infty(A_0)p_\infty^{-1}(A_1) \leq 1$ holds only if $p_\infty(A_0) = p_\infty^{-1}(A_1) = 1$, which in turn is satisfied only under the conditions stated in the proposition. □

5. Quadratic seminorms

In this section we treat the case where p is the quadratic seminorm defined in (2.4) , i.e.,

$$p(\boldsymbol{v}) = \left(\boldsymbol{v}^T M \boldsymbol{v}\right)^{1/2}, \quad \boldsymbol{v} \in \mathbb{R}^N, \tag{5.1}$$

where M is a symmetric positive semi-definite matrix M. Before we treat this general case, we deal with the classical l^2-norm, also called the Euclidean norm. Thus $M = I$, where I is the $N \times N$ identity matrix. We start with the following auxiliary result. Recall that $\boldsymbol{u} = (1, \dots, 1)^T$.

5.1 Lemma. *Let p_2 be the quadratic norm given by $p_2(\boldsymbol{v}) = (v_1^2 + \cdots + v_N^2)^{\frac{1}{2}}$ and let A be the matrix $A = I - \boldsymbol{u}\boldsymbol{\beta}^T$, where $\boldsymbol{u}, \boldsymbol{\beta} \in \mathbb{R}^N$.*

(a) *If $\boldsymbol{u}, \boldsymbol{\beta}$ are collinear then*

$$p_2(A) = \|A\| = \max\{1, |\alpha|\},$$

where $\alpha = 1 - \boldsymbol{u}^T\boldsymbol{\beta}$.

(b) *If $\boldsymbol{u}, \boldsymbol{\beta}$ are not collinear, then $p_2(A) > 1$.*

Proof. (a) If $\boldsymbol{u}, \boldsymbol{\beta}$ are collinear, i.e., $\boldsymbol{\beta} = \mu\boldsymbol{u}$ for some constant $\mu \in \mathbb{R}$, then the matrix $A = I - \mu\boldsymbol{u}\boldsymbol{u}^T$ is symmetric and we get that $p_2(A) = \|A\|$ is the maximum absolute value of its eigenvalues. According to Lemma 3.5, these eigenvalues are 1 (with multiplicity $N-1$) and α. Thus $p_2(A) = \max\{1, |\alpha|\}$.

(b) If $\boldsymbol{u}, \boldsymbol{\beta}$ are not collinear, then we can decompose $\boldsymbol{\beta}$ as $\boldsymbol{\beta} = \mu\boldsymbol{u} + \boldsymbol{c}$ where $\boldsymbol{c} \neq \boldsymbol{0}$ is orthogonal to $\boldsymbol{u}$. Now

$$A\boldsymbol{c} = (I - \boldsymbol{u}\boldsymbol{\beta}^T)\boldsymbol{c} = \boldsymbol{c} - \boldsymbol{u}(\mu\boldsymbol{u} + \boldsymbol{c})^T\boldsymbol{c} = \boldsymbol{c} - (\boldsymbol{c}^T\boldsymbol{c})\boldsymbol{u} = \boldsymbol{c} - \|\boldsymbol{c}\|^2\boldsymbol{u},$$

whence we get that $\|A\boldsymbol{c}\|^2 = \|\boldsymbol{c}\|^2 + N\|\boldsymbol{c}\|^4$, where we have used that $\|\boldsymbol{u}\|^2 = N$. Therefore

$$p_2(A) \geq \|A\boldsymbol{c}\|/\|\boldsymbol{c}\| = (1 + N\|\boldsymbol{c}\|^2)^{\frac{1}{2}} > 1,$$

which concludes the proof. □

5.2 Proposition. *Let $p = p_2$ be the Euclidean norm. Then the threshold criterion holds if and only if $\boldsymbol{u}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1$ are collinear and $|\alpha_0| \leq 1 \leq |\alpha_1|$.*

Proof. We have $A_0 = I - \boldsymbol{u}\boldsymbol{\beta}_0^T$ and $A_1^{-1} = I - \boldsymbol{u}{\boldsymbol{\beta}'_1}^T$ where $\boldsymbol{\beta}'_1 = -\alpha_1^{-1}\boldsymbol{\beta}_1$. Thus the previous lemma yields that both $p(A_0) \geq 1$ and $p^{-1}(A_1) = p(A_1^{-1}) \geq 1$. Now Proposition 3.3 yields that the threshold criterion holds if and only if $p(A_0) = p(A_1^{-1}) = 1$. First, this requires that $\boldsymbol{u}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1$ are collinear. Then $p(A_0) = \max\{1, |\alpha_0|\}$ and $p(A_1^{-1}) = \max\{1, |\alpha_1|^{-1}\}$; here we have used that $1 - \boldsymbol{u}^T{\boldsymbol{\beta}'_1}^T = 1 - \alpha_1^{-1}(1 - \alpha_1) = \alpha_1^{-1}$. We obtain that the threshold criterion holds if and only if $|\alpha_0| \leq 1 \leq |\alpha_1|$. Reminding that $\alpha_0 \neq 0$, this proves the result. □

Now we are ready to consider the more general case in (2.4) with M an arbitrary symmetric positive semi-definite matrix. Thus, M can be decomposed as

$$M = Q\Lambda Q^T, \tag{5.2}$$

where Q is an orthogonal matrix (i.e., $Q^TQ = QQ^T = I$) and Λ is a diagonal matrix with nonnegative entries, the eigenvalues of M. The columns of Q are the (orthogonal) eigenvectors of M. Define n as

$$n = \operatorname{rank}(M) = \operatorname{rank}(\Lambda) \leq N.$$

Without loss of generality we can assume that

$$\Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

where Λ_{11} is an $n \times n$ diagonal matrix with strictly positive entries. Note that $\Lambda_{11} = \Lambda$ iff $n = N$. The corresponding decomposition of Q is given by

$$Q = (Q_1 \; Q_2)\,, \tag{5.3}$$

where Q_1, Q_2 are $N \times n$ and $N \times (N-n)$ matrices, respectively. (When $n = N$, we shall adopt the conventions: $Q = Q_1$ and $Q_2 = 0$.) Here the columns of Q_1 are the eigenvectors of M corresponding to the positive eigenvalues contained in Λ_{11}. Observe that, instead of (5.2), we can also write

$$M = Q_1 \Lambda_{11} Q_1^T\,.$$

The $(N \times n)$-matrix Q_1 is *semi-orthogonal* in the sense that $Q_1^T Q_1 = I$.

After these preparations we are able to formulate our results concerning the seminorm of an $N \times N$ matrix A.

5.3 Lemma. *Let the seminorm p be given by* (5.1) *and let A be an $N \times N$ matrix, then*

$$p(A) = \begin{cases} \|\Lambda_{11}^{\frac{1}{2}} Q_1^T A Q_1 \Lambda_{11}^{-\frac{1}{2}}\| & \textit{if } \; Q_1^T A Q_2 = 0 \\ +\infty & \textit{otherwise}\,. \end{cases} \tag{5.4}$$

Here $\|\cdot\|$ is the standard Euclidean norm.

In particular, if $\operatorname{rank}(M) = \operatorname{rank}(\Lambda) = N$, *then*

$$p(A) = \|\Lambda^{\frac{1}{2}} Q^T A Q \Lambda^{-\frac{1}{2}}\|\,. \tag{5.5}$$

Proof. Obviously, to compute $p(A)$, we have to maximize $(\boldsymbol{v}^T A^T M A \boldsymbol{v})^{\frac{1}{2}}$ under the constraint $\boldsymbol{v}^T M \boldsymbol{v} = 1$. Substituting $\boldsymbol{w} = Q^T \boldsymbol{v}$ this amounts to maximizing $(\boldsymbol{w}^T Q^T A^T Q \Lambda Q^T A Q \boldsymbol{w})^{\frac{1}{2}}$ under the constraint $\boldsymbol{w}^T \Lambda \boldsymbol{w} = 1$. Define the matrix $B = Q^T A Q$, then

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix} A \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} = \begin{pmatrix} Q_1^T A Q_1 & Q_1^T A Q_2 \\ Q_2^T A Q_1 & Q_2^T A Q_2 \end{pmatrix},$$

where B_{11} is an $n \times n$-matrix. The expression we have to maximize is $(\boldsymbol{w}^T B^T \Lambda B \boldsymbol{w})^{\frac{1}{2}}$. A simple computation shows that

$$B^T \Lambda B = \begin{pmatrix} B_{11}^T \Lambda_{11} B_{11} & B_{11}^T \Lambda_{11} B_{12} \\ B_{12}^T \Lambda_{11} B_{11} & B_{12}^T \Lambda_{11} B_{12} \end{pmatrix}.$$

Decomposing $\boldsymbol{w} = (\boldsymbol{w}_1 \; \boldsymbol{w}_2)^T$, with $\boldsymbol{w}_1 \in \mathbb{R}^n$ and $\boldsymbol{w}_2 \in \mathbb{R}^{N-n}$, we get

$$\boldsymbol{w}^T B^T \Lambda B \boldsymbol{w} = \boldsymbol{w}_1^T B_{11}^T \Lambda_{11} B_{11} \boldsymbol{w}_1 + 2 \boldsymbol{w}_1^T B_{11}^T \Lambda_{11} B_{12} \boldsymbol{w}_2 + \boldsymbol{w}_2^T B_{12}^T \Lambda_{11} B_{12} \boldsymbol{w}_2\,. \tag{5.6}$$

Furthermore, the constraint $\boldsymbol{w}^T \Lambda \boldsymbol{w} = 1$ amounts to

$$\boldsymbol{w}_1^T \Lambda_{11} \boldsymbol{w}_1 = 1\,.$$

This constraint only involves $\boldsymbol{w}_1$ and not $\boldsymbol{w}_2$. This means that maximisation of (5.6) yields $+\infty$ unless $B_{12} = Q_1^T A Q_2 = 0$. This proves the second equality in (5.4).

Let us henceforth assume that $B_{12} = 0$. Thus

$$\begin{aligned} (p(A))^2 &= \max\{\boldsymbol{w}_1^T B_{11}^T \Lambda_{11} B_{11} \boldsymbol{w}_1 | \boldsymbol{w}_1^T \Lambda_{11} \boldsymbol{w}_1 = 1\} \\ &= \max\{\boldsymbol{s}^T \Lambda_{11}^{-\frac{1}{2}} B_{11}^T \Lambda_{11}^{\frac{1}{2}} \Lambda_{11}^{\frac{1}{2}} B_{11} \Lambda_{11}^{-\frac{1}{2}} \boldsymbol{s} \mid \boldsymbol{s}^T \boldsymbol{s} = 1\} \\ &= \max\{\|\Lambda_{11}^{\frac{1}{2}} B_{11} \Lambda_{11}^{-\frac{1}{2}} \boldsymbol{s}\|^2 \mid \|\boldsymbol{s}\|^2 = 1\}, \end{aligned}$$

where we have substituted $\boldsymbol{s} = \Lambda_{11}^{\frac{1}{2}}\boldsymbol{w}_1$. This yields that

$$p(A) = \|\Lambda_{11}^{\frac{1}{2}} B_{11} \Lambda_{11}^{-\frac{1}{2}}\| = \|\Lambda_{11}^{\frac{1}{2}} Q_1^T A Q_1 \Lambda_{11}^{-\frac{1}{2}}\|$$

which had to be proved.

Finally, if $\operatorname{rank}(M) = N$ then $\Lambda_{11} = \Lambda$, $Q_1 = Q$ and $Q_2 = 0$, and thus (5.4) reduces to (5.5). □

We apply this result to the matrix A_d given by (2.19) or, alternatively, by (3.3). That is, A_d is of the form $A_d = I - \boldsymbol{u}\boldsymbol{\beta}_d^T$. Then

$$Q_1^T A_d Q_2 = Q_1^T Q_2 - Q_1^T \boldsymbol{u}(Q_2^T \boldsymbol{\beta}_d)^T = -Q_1^T \boldsymbol{u}(Q_2^T \boldsymbol{\beta}_d)^T$$

since $Q_1^T Q_2 = 0$ by the orthogonality of Q. Therefore $Q_1^T A_d Q_2 = 0$ if either (i) $Q_1^T \boldsymbol{u} = \boldsymbol{0}$ or (ii) $Q_2^T \boldsymbol{\beta}_d = \boldsymbol{0}$ for $d = 0, 1$. In case (i) we have $p(A_0) = p(A_1^{-1}) = \|\Lambda_{11}^{\frac{1}{2}} Q_1^T Q_1 \Lambda_{11}^{-\frac{1}{2}}\| = 1$ and the threshold criterion holds. Note however that we have $p(\boldsymbol{u}) = 0$ and consequently the adaptivity condition does not hold in this case. We now consider case (ii). Obviously, $Q_2^T \boldsymbol{\beta}_d = \boldsymbol{0}$ is equivalent to $\boldsymbol{\beta}_d \in \operatorname{Ran}(Q_2)^\perp = \operatorname{Ran}(Q_1)$. We compute $p(A_0)$ and $p(A_1^{-1})$ in this case:

$$p(A_0) = \|\Lambda_{11}^{\frac{1}{2}} Q_1^T (I - \boldsymbol{u}\boldsymbol{\beta}_0^T) Q_1 \Lambda_{11}^{-\frac{1}{2}}\| = \|I - \tilde{\boldsymbol{u}}\tilde{\boldsymbol{\beta}}_0^T\|\,,$$

where $\tilde{\boldsymbol{u}} = \Lambda_{11}^{\frac{1}{2}} Q_1^T \boldsymbol{u}$ and $\tilde{\boldsymbol{\beta}}_0 = \Lambda_{11}^{-\frac{1}{2}} Q_1^T \boldsymbol{\beta}_0$ are n-dimensional vectors. We conclude from Lemma 5.1 that $p(A_0) > 1$ if $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\beta}}_0$ are not collinear and that $p(A_0) = \max\{1, |\tilde{\alpha}_0|\}$, with $\tilde{\alpha}_0 = 1 - \tilde{\boldsymbol{u}}^T \tilde{\boldsymbol{\beta}}_0$, if $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\beta}}_0$ are collinear. Here we have assumed that $n > 1$. Substitution of $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\beta}}_0$ yields that

$$\tilde{\alpha}_0 = 1 - \boldsymbol{u}^T Q_1 Q_1^T \boldsymbol{\beta}_0\,.$$

A similar computation shows that $p(A_1^{-1}) > 1$ if $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\beta}}_1$ are not collinear, where $\tilde{\boldsymbol{\beta}}_1 = \Lambda_{11}^{-\frac{1}{2}} Q_1^T \boldsymbol{\beta}_1$, and that $p(A_1^{-1}) = \max\{1, |\tilde{\alpha}_1|^{-1}\}$ if $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\beta}}_1$ are collinear. Here

$$\tilde{\alpha}_1 = (1 + \frac{1}{\alpha_1} \boldsymbol{u}^T Q_1 Q_1^T \boldsymbol{\beta}_1)^{-1}\,.$$

5.4 Lemma. *When $Q_1^T \boldsymbol{u} \neq 0$, the following two assertions are equivalent:*

(i) $\boldsymbol{\beta}_d$ and $M\boldsymbol{u}$ are collinear.

(ii) $Q_2^T \boldsymbol{\beta}_d = \boldsymbol{0}$ and $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\beta}}_d$ are collinear.

Proof. Assume (i). We have $M\boldsymbol{u} \neq 0$ (otherwise $Q_1^T M \boldsymbol{u} = \Lambda_{11} Q_2^T \boldsymbol{u} = 0$) and then $\boldsymbol{\beta}_d = c \cdot M\boldsymbol{u} = c \cdot Q_1 \Lambda_{11} Q_1^T \boldsymbol{u}$, where $c \in \mathbb{R}$. Since $Q_2^T Q_1 = 0$ we find that $Q_2^T \boldsymbol{\beta}_d = \boldsymbol{0}$. Furthermore, $\tilde{\boldsymbol{\beta}}_d = \Lambda_{11}^{-\frac{1}{2}} Q_1^T \boldsymbol{\beta}_d = c \cdot \Lambda_{11}^{\frac{1}{2}} Q_1^T \boldsymbol{u} = c \cdot \tilde{\boldsymbol{u}}$. Here we have used that $Q_1^T Q_1 = I$.

Assume (ii): $Q_2^T \boldsymbol{\beta}_d = \boldsymbol{0}$ is equivalent to $\boldsymbol{\beta}_d \in \operatorname{Ran}(Q_1)$, i.e., $\boldsymbol{\beta}_d = Q_1 \boldsymbol{\xi}_d$. Since $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{\beta}}_d$ are collinear, we have $\tilde{\boldsymbol{\beta}}_d = c \cdot \tilde{\boldsymbol{u}}$, that is $\Lambda_{11}^{-\frac{1}{2}} Q_1^T \boldsymbol{\beta}_d = c \cdot \Lambda_{11}^{\frac{1}{2}} Q_1^T \boldsymbol{u}$, which yields $\boldsymbol{\xi}_d = c \cdot \Lambda_{11} Q_1^T \boldsymbol{u}$, and hence $\boldsymbol{\beta}_d = c \cdot Q_1 \Lambda_{11} Q_1^T \boldsymbol{u} = c \cdot M\boldsymbol{u}$. This concludes the proof. □

Now if $\boldsymbol{\beta}_d = cM\boldsymbol{u}$ with $c \in \mathbb{R}$, we get

$$\tilde{\alpha}_0 = 1 - \boldsymbol{u}^T Q_1 Q_1^T \boldsymbol{\beta}_0 = 1 - c\boldsymbol{u}^T Q_1 Q_1^T \Lambda_{11} Q_1^T \boldsymbol{u} = 1 - \boldsymbol{u}^T \boldsymbol{\beta}_0 = \alpha_0\,,$$

and

$$\tilde{\alpha}_1 = (1 + \frac{1}{\alpha_1} \boldsymbol{u}^T Q_1 Q_1^T \boldsymbol{\beta}_1)^{-1} = (1 + \frac{1}{\alpha_1} \boldsymbol{u}^T \boldsymbol{\beta}_1)^{-1} = (1 + \frac{1}{\alpha_1}(1 - \alpha_1))^{-1} = \alpha_1\,,$$

hence $p(A_0) = \max\{1, |\alpha_0|\}$ and $p^{-1}(A_1) = \max\{1, |\alpha_1|^{-1}\}$. Thus we arrive at the following result.

5.5 Proposition. *Let p be the quadratic norm given by* (5.1)*, let M be decomposed as in* (5.2)*, and assume that $n = \operatorname{rank}(M) \geq 2$. Then the threshold criterion holds in the following two cases:*

(*i*) $Q_1^T \boldsymbol{u} = \mathbf{0}$ *(in which case the adaptivity condition is not satisfied);*

(*ii*) $\boldsymbol{\beta}_d$ *and $M\boldsymbol{u}$ are collinear for $d = 0, 1$ and $|\alpha_0| \leq 1 \leq |\alpha_1|$.*

The case $n = 1$ corresponds to $M = \boldsymbol{a}\boldsymbol{a}^T$, where $\boldsymbol{a} \in \mathbb{R}^N$, $\boldsymbol{a} \neq \mathbf{0}$, i.e., $p(\boldsymbol{v}) = |\boldsymbol{a}^T \boldsymbol{v}|$ for all $\boldsymbol{v} \in \mathbb{R}^N$. The corresponding decision rule is similar to the one which will be studied in the next section.

Observe that Proposition 5.2 is only a special case of this last result. We consider two other cases in more detail.

5.6 Example. (*a*) Consider first the case where $M = \Lambda$ with Λ a diagonal matrix with strictly positive entries $M_{jj} = \lambda_j$ for $j = 1, \ldots, N$. The threshold criterion holds if and only if there are constants μ_0, μ_1 such that $\beta_{d,j} = \mu_d \lambda_j$ for $d = 0, 1$ and $j = 1, \ldots, N$, and

$$|1 - \mu_0(\lambda_1 + \cdots + \lambda_N)| \leq 1 \leq |1 - \mu_1(\lambda_1 + \cdots + \lambda_N)| . \tag{5.7}$$

If we assume the input signal to be contaminated by additive uncorrelated Gaussian noise, it is easy to show that we must take

$$\mu_0 = \frac{\sum_j \lambda_j}{\sum_j \lambda_j^2 + \left(\sum_j \lambda_j\right)^2} , \tag{5.8}$$

for minimizing the variance of the noise in the approximation signal. It is then obvious that the first inequality in condition (5.7) is satisfied; choosing, e.g., $\mu_1 = 0$ we do have perfect reconstruction.

(*b*) Consider the same case as in (*a*) but with $\lambda_1, \ldots, \lambda_n$ strictly positive and $\lambda_{n+1} = \cdots = \lambda_N = 0$. The threshold criterion requires that $\boldsymbol{\beta}_d$ is collinear with $M\boldsymbol{u}$. This means that $\beta_{d,n+1} = \cdots = \beta_{d,N} = 0$. In other words, the order of the update filter, initially assumed to be equal to N, is only n, and we are now back in the situation described in (*a*).

6. Seminorms based on derivative filters

Let us now consider the situation where p is given by the particular case of the seminorms defined in (2.3) with $q = 1$ and $I = 1$, i.e.,

$$p(\boldsymbol{v}) = \left|\boldsymbol{a}^T \boldsymbol{v}\right| , \quad \text{with } \boldsymbol{a} \neq \mathbf{0}. \tag{6.1}$$

The adaptivity condition holds if and only if $\boldsymbol{a}^T \boldsymbol{u} \neq 0$. We establish necessary and sufficient conditions for the threshold criterion.

6.1 Proposition. *If $p(\boldsymbol{v}) = |\boldsymbol{a}^T \boldsymbol{v}|$, with $\boldsymbol{a}^T \boldsymbol{u} \neq 0$, then the threshold criterion holds if and only if $\boldsymbol{\beta}_d$ and $\boldsymbol{a}$ are collinear and $|\alpha_0| \leq |\alpha_1|$.*

Proof. From the definition of a matrix seminorm we have that

$$p(A_d) = \sup\{\left|\boldsymbol{a}^T A_d \boldsymbol{v}\right| \mid \boldsymbol{v} \in \mathbb{R}^N \text{ and } \left|\boldsymbol{a}^T \boldsymbol{v}\right| = 1\}.$$

Therefore, in order to calculate this seminorm we have to find the supremum of $|\boldsymbol{a}^T A_d \boldsymbol{v}|$ under the constraint $\left|\boldsymbol{a}^T \boldsymbol{v}\right| = 1$. We distinguish two cases, namely $\boldsymbol{\beta}_d$ and $\boldsymbol{a}$ *are* or *are not* collinear.

(*i*) $\boldsymbol{\beta}_d$ collinear with $\boldsymbol{a}$. In this case we can write $\boldsymbol{\beta}_d = \eta_d\, \boldsymbol{a}$ for some constant $\eta_d \in \mathbb{R}$ and we get

$$|\boldsymbol{a}^T A_d \boldsymbol{v}| = |\boldsymbol{a}^T (I - \boldsymbol{u}\boldsymbol{\beta}_d^T)\boldsymbol{v}| = |\boldsymbol{a}^T \boldsymbol{v} - \eta_d \boldsymbol{a}^T \boldsymbol{u}\boldsymbol{a}^T \boldsymbol{v}| = |1 - \eta_d \boldsymbol{a}^T \boldsymbol{u}||\boldsymbol{a}^T \boldsymbol{v}| = |1 - \boldsymbol{\beta}_d^T \boldsymbol{u}| = |\alpha_d| .$$

This yields that $p(A_d) = |\alpha_d|$. In a similar way, we can show that $p(A_d^{-1}) = |\alpha_d|^{-1}$. Thus, from Proposition 3.3 we conclude that the threshold criterion holds if and only if $|\alpha_0| \le |\alpha_1|$.

(ii) $\boldsymbol{\beta}_d$ **not collinear with** $\boldsymbol{a}$. In this case we can express $\boldsymbol{\beta}_d = \eta_d \boldsymbol{a} + \boldsymbol{c}$ with $\boldsymbol{a}^T\boldsymbol{c} = 0$ and $\boldsymbol{c} \ne 0$. Let us choose $\boldsymbol{v}$ such that $\boldsymbol{a}^T\boldsymbol{v} = 0$ and $\boldsymbol{c}^T\boldsymbol{v} \ne 0$. Then, $p(\boldsymbol{v}) = |\boldsymbol{a}^T\boldsymbol{v}| = 0$ and

$$p(A_d\boldsymbol{v}) = |\boldsymbol{a}^T A_d \boldsymbol{v}| = |\boldsymbol{a}^T \boldsymbol{u}\boldsymbol{c}^T\boldsymbol{v}| \ne 0 .$$

From Proposition 3.1(b) we conclude that $p(A_d) = \infty$ and, consequently, the threshold condition cannot hold. □

Throughout the remainder of this section we deal with one-dimensional signals and we assume that we have one detail band y, that is, $K = 1$ in Fig. 2. As in Example 2.1(a) we take $x(n) = x_0(2n)$ and $y(n) = x_0(2n+1)$. In § 2.3 we have assumed that the gradient vector is indexed by $j = 1, 2, \ldots, N$, that is $\boldsymbol{v} = (v_1, \ldots, v_N)^T$. In this section we assume, for reasons that will become clear below, that $\boldsymbol{v} = (v_{-K}, v_{-K+1}, \ldots, v_{-1}, v_0, v_1, \ldots, v_{L-1}, v_L)^T$, and

$$v_j(n) = x(n) - y(n+j) \qquad \text{for } j = -K, \ldots, 0, \ldots, L . \tag{6.2}$$

We give an illustration in Fig. 4. With every coefficient vector $\boldsymbol{a} \in \mathbb{R}^{K+L+1}$ in (6.1) we can

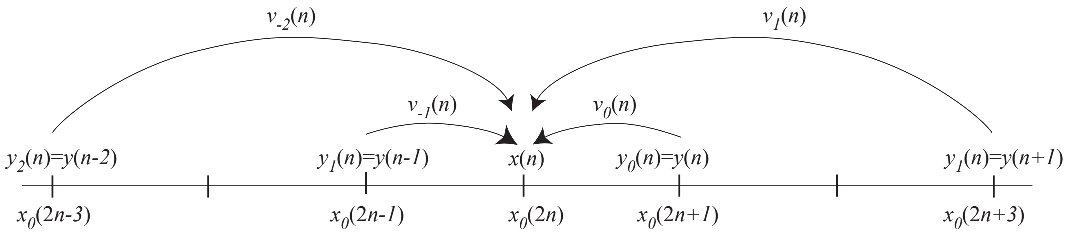

Figure 4: *Indexing of the gradient vector.*

associate a filter $\Delta_{\boldsymbol{a}}$ which maps an input vector $(y(n-K), \ldots, y(n-1), x(n), y(n), y(n+1), \ldots, y(n+L))$, or equivalently, $(x_0(2n-2K+1), \ldots, x_0(2n-1), x_0(2n), x_0(2n+1), \ldots, x_0(2n+2L+1))$ onto an output vector

$$\Delta_{\boldsymbol{a}}(x_0)(2n) = \sum_{j=-K}^{L} a_j v_j(n) = \sum_{j=-K}^{L} a_j \big(x_0(2n) - x_0(2n+2j+1)\big) . \tag{6.3}$$

It is possible to choose the coefficients in such a way that it corresponds with an N'th order discrete derivative filter for every N with

$$N \le L + K + 1 .$$

For $N = 1$ and $K = L = 0$, the value $v_0(n) = x_0(2n) - x_0(2n+1)$ is the first order derivative. For $N = 2$ (with $K = 1$ and $L = 0$) and $a_{-1} = a_0 = -1$, we arrive at the expression

$$\Delta(x_0)(2n) = -2x_0(2n) + x_0(2n-1) + x_0(2n+1) ,$$

which is a second-order derivative; see also Example 6.3 below.

We denote by $\mathcal{A}_N$, with $N \ge 1$, the coefficient vectors $\boldsymbol{a} \in \mathbb{R}^{K+L+1}$ for which the corresponding filter $\Delta_{\boldsymbol{a}}$ in (6.3) corresponds with an N'th order derivative filter, or equivalently, rejects signals that are polynomial of order $\le N-1$. The latter means that for all $n \in \mathbb{Z}$,

$$\sum_{j=-K}^{L} a_j \Big[(2n)^k - (2n+2j+1)^k\Big] = 0 \quad \text{for} \quad k = 0, \ldots, N-1 ,$$

which is satisfied if and only if either $N = 1$ or $N > 1$ and

$$\sum_{j=-K}^{L} a_j (2j+1)^k = 0 \text{ for } k = 1, 2, \ldots, N-1\,.$$

Consider the function $Q_{\boldsymbol{a}}$ given by

$$Q_{\boldsymbol{a}}(z) = \sum_{j=-K}^{L} a_j (1 - z^{2j+1})\,.$$

The proof of the following results is straightforward.

6.2 Lemma. $\boldsymbol{a} \in \mathcal{A}_N$ *if and only if* $Q_{\boldsymbol{a}}$ *has a zero at* $z = 1$ *with multiplicity* N.

We next consider the case $N > 1$. Obviously, $Q_{\boldsymbol{a}}$ has a zero of multiplicity N iff $Q'_{\boldsymbol{a}}$ has a zero of multiplicity $N-1$. Now

$$Q'_{\boldsymbol{a}}(z) = -\sum_{j=-K}^{L} a_j (2j+1) z^{2j}\,, \tag{6.4}$$

and if $Q'_{\boldsymbol{a}}$ has a zero at $z = 1$ with multiplicity $N-1$, then we can write

$$Q'_{\boldsymbol{a}}(z) = (z-1)^{N-1} z^{-2K} R(z)\,, \tag{6.5}$$

with

$$R(z) = \sum_{j=0}^{2(L+K)-N+1} r_j z^j\,.$$

From the fact that $Q'_{\boldsymbol{a}}$ is even (see (6.4)), we conclude that

$$(z-1)^{N-1} R(z) = (-1)^{N-1} (z+1)^{N-1} R(-z)\,.$$

This yields that R can be written as

$$R(z) = (z+1)^{N-1} \sum_{j=0}^{L+K+1-N} q_j z^{2j}\,. \tag{6.6}$$

Assume henceforth that

$$L + K + 1 \geq N \text{ (see before) and } L + K + 2 \leq 2N\,.$$

Substitution of (6.6) into (6.5) yields

$$\begin{aligned} Q'_{\boldsymbol{a}}(z) &= z^{-2K} (z^2 - 1)^{N-1} \sum_{i=0}^{L+K+1-N} q_i z^{2i} \\ &= z^{-2K} \sum_{j=0}^{N-1} \binom{N-1}{j} (-1)^{N-1-j} \sum_{i=0}^{L+K+1-N} q_i z^{2(i+j)}\,. \end{aligned}$$

Replacing the summation variables i, j by $l = i + j, j$ we get

$$\begin{aligned} Q'_{\boldsymbol{a}}(z) &= z^{-2K} \sum_{l=0}^{L+K} \sum_{j=\max\{0, l-L-K-1+N\}}^{\min\{N-1, l\}} \binom{N-1}{j} (-1)^{N-1-j} q_{l-j} z^{2l} \\ &= \sum_{l=-K}^{L} \sum_{j=\max\{0, l-L-1+N\}}^{\min\{N-1, l+K\}} \binom{N-1}{j} (-1)^{N-1-j} q_{l+K-j} z^{2l}\,. \end{aligned}$$

In combination with (6.4) this yields the following expression for the coefficients a_l:

$$-(2l+1)a_l = \sum_{j=\max\{0,l-L-1+N\}}^{\min\{N-1,l+K\}} \binom{N-1}{j} (-1)^{N-1-j} q_{l+K-j} \,. \tag{6.7}$$

If $N = L + K + 1$ this expression reduces to

$$-(2l+1)a_l = \binom{L+K}{l+K} (-1)^{L-l} q_0 \,.$$

In particular, if $K = L$ and $N = 2L + 1$ (odd length filter), we get (putting $q_0 = 1$):

$$a_l^{(N)} = \frac{(-1)^{L+l+1}}{2l+1} \binom{2L}{L+l} \,, \tag{6.8}$$

and if $K = L + 1$ and $N = 2L + 2$ (even length filter), we get

$$a_l^{(N)} = \frac{(-1)^{L+l+1}}{2l+1} \binom{2L+1}{L+l+1} \,. \tag{6.9}$$

Here the superindex 'N' indicates the order of the underlying derivative operator.

In the two previous cases, it can be shown that the adaptivity condition $\sum_{l=-K}^{L} a_l^{(N)} \neq 0$ is satisfied. Indeed, in both cases this expression represents the sum of an alternating series whose terms have decreasing absolute values. As the first term is positive, the sum is nonzero.

6.3 Example. (2nd order derivative) Consider the case where $K = 1,\ L = 0$ and $N = 2$. Then $\boldsymbol{a} = (a_{-1}, a_0)^T = (-1,-1)^T$ (choosing $q_0 = 1$). From Proposition 6.1 we conclude that the threshold criterion holds if $\boldsymbol{\beta}_0 = \gamma_0(1,1)^T$ and $\boldsymbol{\beta}_1 = \gamma_1(1,1)^T$ with

$$|1 - 2\gamma_0| \leq |1 - 2\gamma_1| \,.$$

Choosing $\boldsymbol{a}^{(N)}$ as in (6.8) or (6.9), as shown above, we have $A^{(N)} := \sum_j a_j^{(N)} \neq 0$. Following Proposition 6.1 we choose constants γ_0, γ_1 such that $|1 - \gamma_0 A^{(N)}| \leq |1 - \gamma_1 A^{(N)}|$ and $\beta_{d,j} = \gamma_d a_j^{(N)}$. We might, for example, choose $\gamma_1 = 0$. This corresponds with a trivial update step (i.e., no modification of the approximation signal) for decision $d = 1$. In homogeneous areas where $d = 0$, we choose γ_0 in such a way that $|1 - \gamma_0 A^{(N)}| \leq 1$ and a given noise rejection criterion is maximised. If we assume, as before, that a homogeneous region is a polynomial signal contaminated by uncorrelated Gaussian noise of variance σ^2, then we could minimise the variance of the noise in the approximation signal. As we have seen in Example 5.6(a) we must choose (putting $a_j = a_j^{(N)}$)

$$\gamma_0 = \frac{\sum_j a_j}{\sum_j a_j^2 + \left(\sum_j a_j\right)^2} \,. \tag{6.10}$$

This leads to $|1 - \gamma_0 A^{(N)}| = (\sum_j a_j^2)/\left(\sum_j a_j^2 + \left(\sum_j a_j\right)^2\right) \leq 1$, hence the conditions of Proposition 6.1 are satisfied and we do have perfect reconstruction.

7. Examples and Simulations

In this section, we present some simulation results to illustrate the adaptive decompositions described in the previous sections. We start with the 1D case in § 7.1. The remainder of the section, that is, § 7.2-7.4, is devoted to 2D images. In § 7.2 we treat the quincunx sampling scheme and in § 7.3-7.4 the 2×2 sampling scheme. The example treated in § 7.4 does not fit within the theoretical setting of the previous section since the decision map is based on *two* different seminorms.

The thresholds that we use in § 7.2-7.3 are normalised in such a way that in all experiments the decision map at a delta impulse $x(\boldsymbol{n}) = \delta(\boldsymbol{n} - \boldsymbol{n}_0)$ (i.e., $x(\boldsymbol{n}) = 1$ if $\boldsymbol{n} = \boldsymbol{n}_0$ and $x(\boldsymbol{n}) = 0$ otherwise) gives the same output. Obviously, the actual choice of the threshold value will depend on the application at hand as well as on the particular image.

7.1 1D case

First we consider the one-dimensional case. We assume that the gradient vector is indexed as in Section 6, i.e., $\boldsymbol{v} = (v_{-K}, \ldots, v_0, \ldots, v_L)^T$ with $v_j(n) = x(n) - y(n+j)$; see (6.2). In all cases, a fixed prediction step of the form $y'(n) = y(n) - \frac{1}{2}(x'(n) + x'(n+1))$ is applied after the update stage. The overall scheme can be iterated over the approximation signal yielding a multiresolution decomposition.

We carry out three different simulations. In the first one, we consider p to be a quadratic seminorm as described in Section 5. For the second and third simulations, we use a seminorm of the form (6.1). For all cases, we compute a three-level decomposition.

7.1.1. Experiment (Quadratic seminorm in 1D - Fig. 5)

We assume that the seminorm is given by

$$p(\boldsymbol{v}) = \frac{1}{3}v_{-2}^2 + v_{-1}^2 + v_0^2 + \frac{1}{3}v_1^2 .$$

Following Example 5.6(a) we see that the threshold criterion holds if $\boldsymbol{\beta}_d = \mu_d(\frac{1}{3}, 1, 1, \frac{1}{3})^T$, with

$$|1 - \frac{8}{3}\mu_0| \leq 1 \leq |1 - \frac{8}{3}\mu_1| .$$

We take $\mu_1 = 0$ and compute μ_0 from (5.8), which yields $\mu_0 = 2/7$. The input signal (a fragment of the 'leleccum' signal from the wavelet toolbox in Matlab) is shown in Fig. 5(a). The approximation and the detail signals are depicted in Fig. 5(b) and (c), respectively, for each level of the decomposition. The decision maps are also shown in Fig. 5(b), and they have been represented by vertical lines at those locations where $d = 1$. The decompositions obtained for both non-adaptive cases corresponding with fixed $d = 0$ and $d = 1$ are shown in Fig. 5(d)-(e) and Fig. 5(f)-(g), respectively.

Observe that the adaptive scheme tunes itself to the local structure of the signal: it yields a smoothed approximation signal except at locations where the gradient is large (i.e., $d = 1$). Such samples are treated as 'singularities' and as such they are not affected by the scheme. Consequently, the detail signal shows only a single peak near such discontinuities and avoids the oscillatory behaviour exhibited by the non-adaptive case with fixed $d = 0$. This oscillatory behaviour can be noticed by carefully inspecting the details at the finest resolution level.

7.1.2. Experiment (Seminorm $p(\boldsymbol{v}) = |\boldsymbol{a}^T\boldsymbol{v}|$ in 1D - Fig. 6)

Let p be given by (6.1) with $K = 2$, $L = 1$ and $\boldsymbol{a} = (1, 1, 1, 1)^T$. Proposition 6.1 yields that we must choose $\boldsymbol{\beta}_d = \gamma_d(1, 1, 1, 1)^T$ for some constants γ_0, γ_1. Now (2.12) and (2.14) yield that $\alpha_0 = 1 - 4\gamma_0$ and we get from Proposition 6.1 that

$$|1 - 4\gamma_0| \leq |1 - 4\gamma_1|$$

must hold. Choosing $\gamma_1 = 0$, i.e., we do not modify the approximation signal if $d = 1$ (high gradient regions), the condition on γ_0 reduces to $|1 - 4\gamma_0| \leq 1$. In fact, we will choose

$$\gamma_0 = 1/5 \text{ and hence } \beta_{0,j} = 1/5 \text{ for all } j .$$

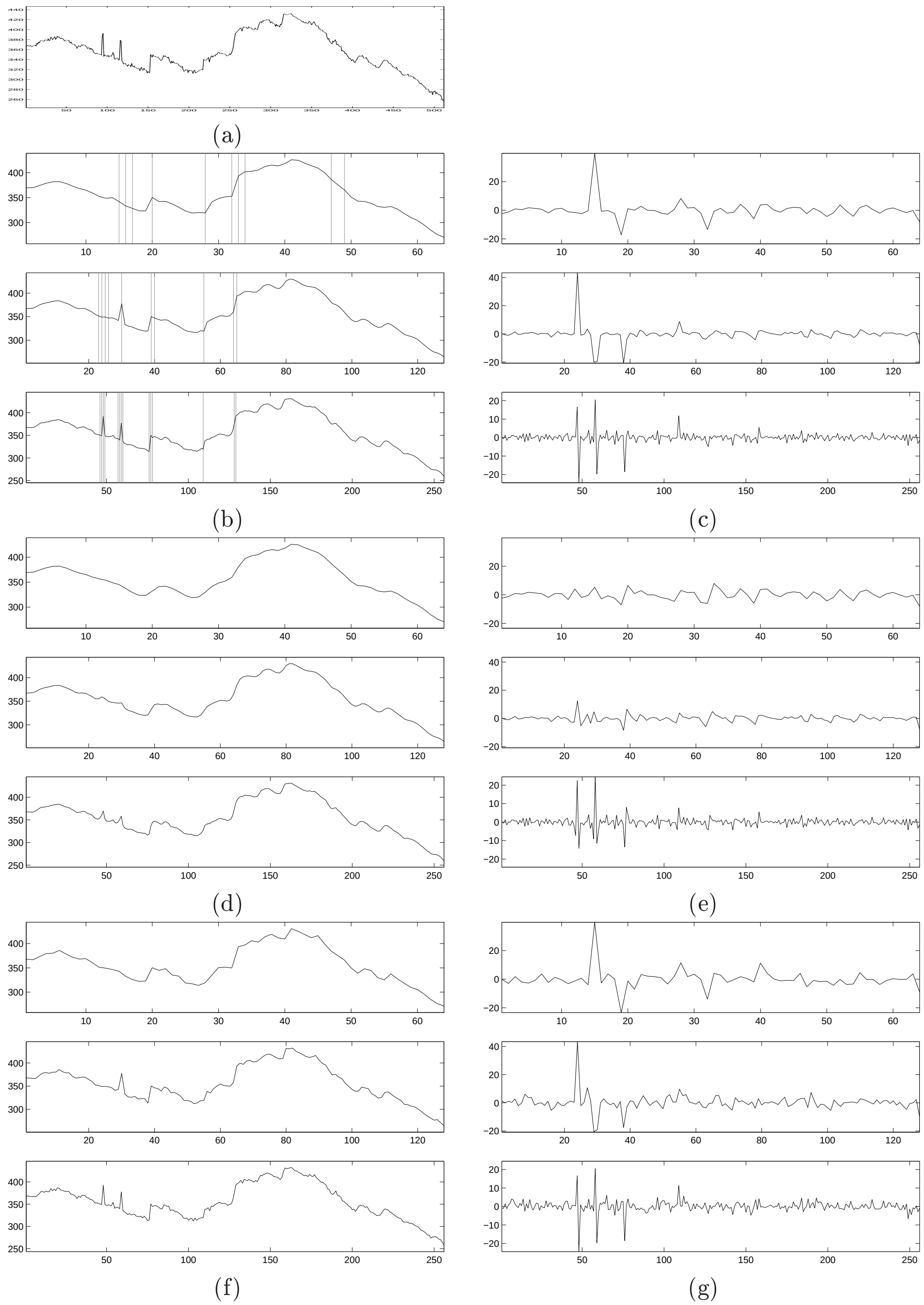

Figure 5: *Decompositions corresponding with Experiment 7.1.1. (a) Original signal; (b)-(c) approximation and detail signals in the adaptive case using a threshold $T = 18$; (d)-(e) approximation and detail signals in the non-adaptive case with $d = 0$; (f)-(g) approximation and detail signals in the non-adaptive case with $d = 1$.*

Again, we can observe from Fig. 6 that the adaptive scheme tunes itself to the local structure of the signal: it 'recognizes' and preserves the discontinuities, while smoothing the more homogeneous regions. This results in a detail signal which is small in homogeneous regions. Near singularities, however, the detail signal comprises a single peak, thus avoiding the oscillatory behaviour exhibited by the non-adaptive case using the update filter U_0 (i.e., fixed $d = 0$).

7.1.3. Experiment (Derivative seminorm in 1D - Fig. 7)
Next, we apply the results of Section 6 and choose $\boldsymbol{a}$ such that the decision map does not respond to polynomial regions of order 3. This gives $\boldsymbol{a}^{(4)} = (1/3, -3, -3, 1/3)^T$ and hence

$$\gamma_0 = 4/35 \text{ and } \boldsymbol{\beta_0} = (4/35)\boldsymbol{a}^{(4)} .$$

Consider an input signal (see Fig. 7) that contains constant, linear and quadratic parts, plus uncorrelated Gaussian noise with $\sigma_n = 0.1$. Figure 7 shows the original signal and the approximation signal at three subsequent levels of decomposition. As before, the vertical lines show the locations where the decision map equals $d = 1$. Like in the previous two experiments, we can observe that the adaptive scheme smooths the homogeneous regions but does not introduce intermediate points during sharp transitions. This allows removal of the noise while keeping the edges unaffected even at coarser scales.

7.2 2D QUINCUNX CASE
In this section we consider two-dimensional signals that are decomposed by a 2-band filter bank comprising one approximation band x and one detail band y, the sampling scheme being given by the quincunx grid depicted in Fig. 8. First we discuss two experiments, both using a quadratic seminorm, with a smaller and larger filter support, respectively. In a third experiment, again with the smaller filter support, we use the seminorm modeling the Laplacian derivative.

In all the three experiments, the adaptive update step will be followed by a fixed prediction step of the form

$$y' = y - \frac{1}{4}\sum_{j=1}^{4} x_j ,$$

where x_j, $j = 1, \ldots, 4$, are the updated coefficients at the four horizontal and vertical neighbouring positions of the point where y is evaluated. Repeated application of this scheme with respect to the approximation image yields an adaptive multiresolution decomposition. In our experiments we only show the decompositions at level 2; at this level the output images have been reduced by a factor 2 both in horizontal and vertical direction. However, for display purposes we will rescale them to their original size.

7.2.1. Experiment (Quadratic seminorm for 2D quincunx, 4 taps - Fig. 9)
In the first experiment of this subsection, we consider update filters with a support comprising the four pixels labeled by $y_1, \ldots, y_4$ in Fig. 8. The update step is of the form specified in (2.15), where y_j are the four horizontal and vertical neighbours $y_1, \ldots, y_4$. Thus:

$$x' = \alpha_d x + \beta_{d,1}y_1 + \beta_{d,2}y_2 + \beta_{d,3}y_3 + \beta_{d,4}y_4 ,$$

with $|\beta_{0,j}| + |\beta_{1,j}| \neq 0$, $j = 1, 2, 3, 4$. It is not difficult to see that the threshold criterion will not hold if p is one of the norms p_1 or p_∞. However, if $p = p_2$, the Euclidean norm, conditions for perfect reconstruction can easily be satisfied. Namely, Proposition 5.2 implies that $\beta_{d,j} = \beta_d$ for $j = 1, \ldots, 4$ and the condition $|\alpha_0| \leq 1 \leq |\alpha_1|$ reduces to

$$|1 - 4\beta_0| \leq 1 \leq |1 - 4\beta_1| .$$

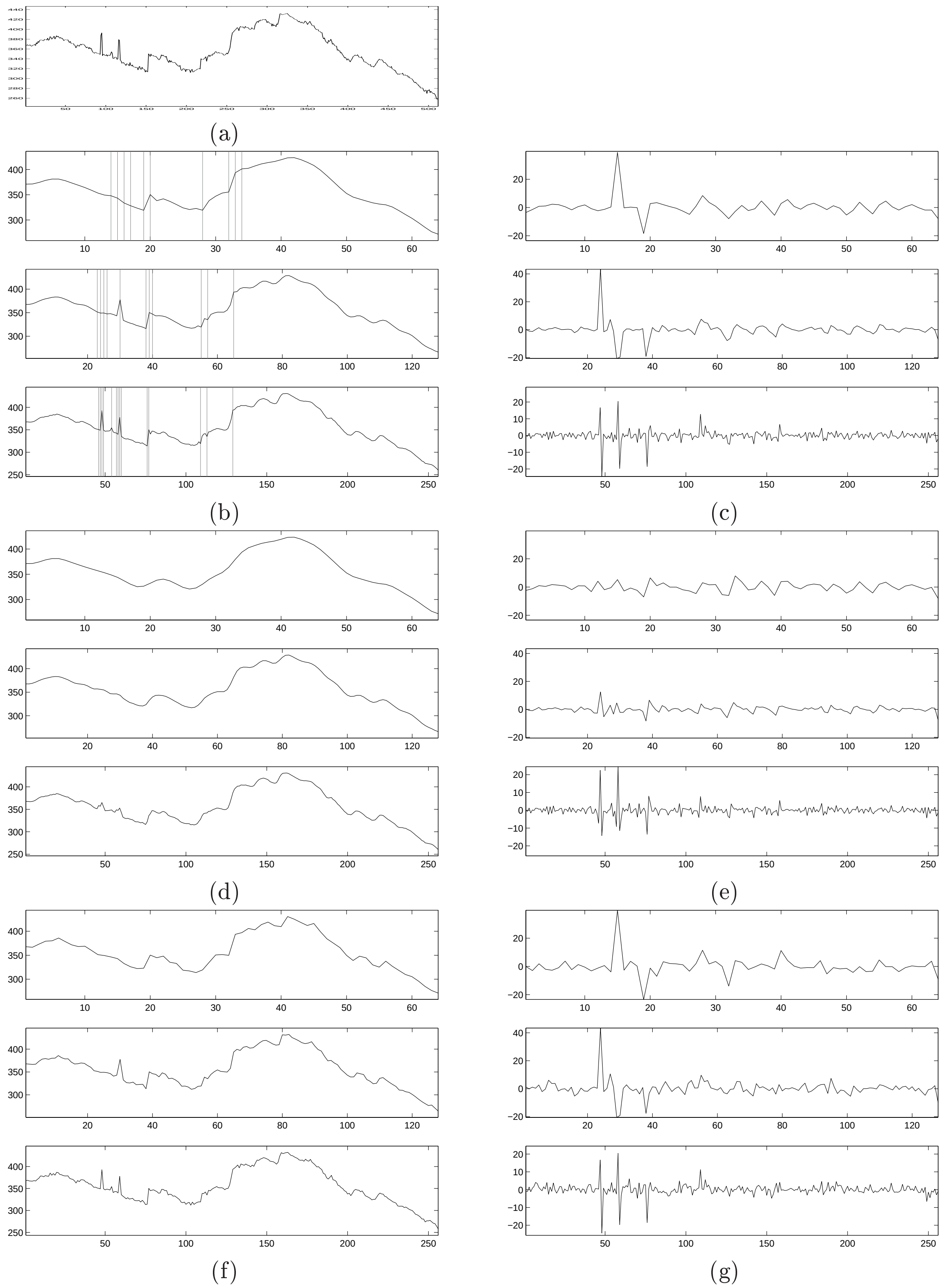

Figure 6: *Decompositions corresponding with Experiment 7.1.2. (a) Original signal; (b)-(c) approximation and detail signals in the adaptive case using a threshold* $T = 18$*; (d)-(e) approximation and detail signals in the non-adaptive case with* $d = 0$*; (f)-(g) approximation and detail signals in the non-adaptive case with* $d = 1$.

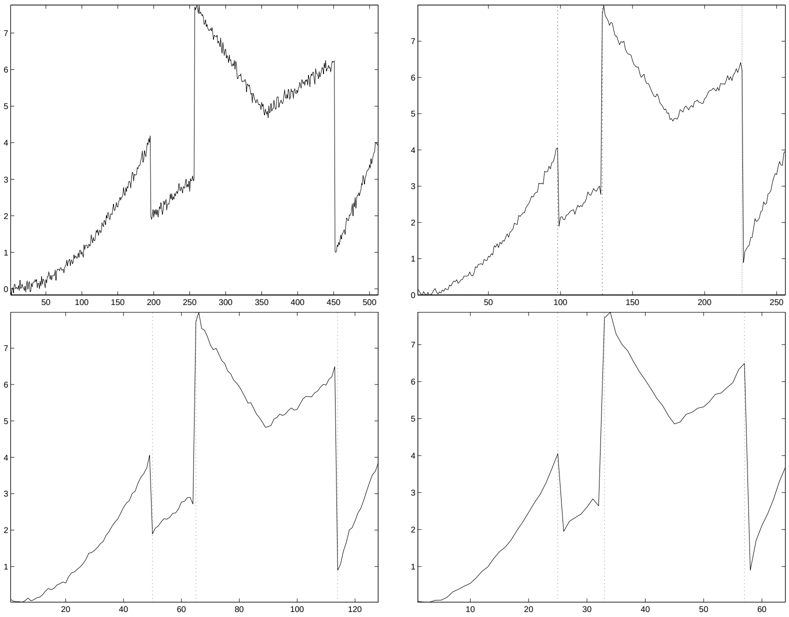

Figure 7: *Adaptive decomposition with polynomial criterion of order* 3*; see Experiment 7.1.3. From left to right and from top to bottom: original signal and approximation signals associated with levels* 1*,* 2 *and* 3*. The vertical dotted lines show the locations where the decision map equals 1 using a threshold* $T = 3.4$*.*

A possible solution is $\beta_0 = 1/5$ and $\beta_1 = 0$. This choice means that at homogeneous areas where $d = 0$, the approximation signal x is averaged with its four neighbours whereas in the vicinity of 'singular' points, where $d = 1$, no filtering is performed.

As input image we choose the 'house' image shown at the top left of Fig. 9. The approximation and detail images obtained after two levels of decomposition are depicted in the middle row. The decision map, depicted at the top right of Fig. 9, shows the high-gradient regions of the approximation image at level 1 (not shown). For comparison, we also show (bottom row) the images for the corresponding non-adaptive case using the fixed update filter U_0.

7.2.2. Experiment (Quadratic seminorm for 2D quincunx, 12 taps - Fig. 10)
In our second experiment we choose update filters with a larger support, namely the samples labeled by $y_1, \ldots, y_{12}$ in Fig. 8. We choose a quadratic norm like in Example 5.6(a) where $M_{jj} = \lambda_j$ is the inverse value of the distance of the corresponding sample to the centre x. Thus $M = \operatorname{diag}\big(1,1,1,1,1/\sqrt{5},\ldots,1/\sqrt{5}\big)$, and now the formula in (5.8) yields $\mu_0 = \frac{5+2\sqrt{5}}{43+16\sqrt{5}}$. Thus

$$\boldsymbol{\beta}_0 = \mu_0\big(1,1,1,1,\frac{1}{\sqrt{5}},\ldots,\frac{1}{\sqrt{5}}\big)^T, \ \alpha_0 = \frac{7\sqrt{5}}{80+43\sqrt{5}} \ \text{ and } \ \alpha_1 = 1\,, \ \boldsymbol{\beta}_1 = \mathbf{0}\,.$$

Again, our input image is the 'house' depicted at the top left of Fig. 10. The approximation and detail images (magnified by a factor 2), after two levels of decomposition, are shown in the middle row.

The corresponding decision map is depicted at the top right. For comparison, we also show (bottom row) the images for the corresponding non-adaptive case using the fixed update filter U_0. Again, we can observe that in the adaptive case the edges are better preserved than in the non-adaptive linear case. We note that the improvement is more visible than in the

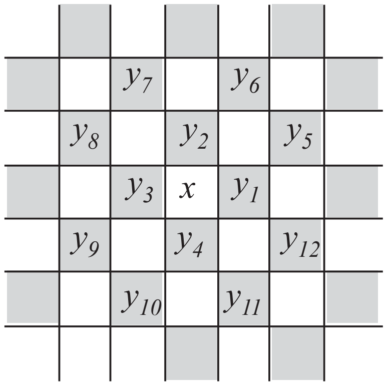

Figure 8: *Filter support for 2D quincunx update. In the first experiment the approximation sample x is to be updated with the detail samples y_1, y_2, y_3 and y_4.*

previous example, which is partly due to the fact that the filter length is larger.

7.2.3. Experiment (Laplacian derivative seminorm for 2D quincunx - Fig. 11)
Let us now consider the case where p models the Laplacian operator, i.e.,

$$p(\boldsymbol{v}) = |\sum_{j=1}^{4} v_j| .$$

As before, the samples $y_1, \ldots, y_4$ correspond with the ones depicted in Fig. 8.

In this case, Proposition 6.1 amounts to $\beta_{d,j} = \beta_d$ for $j = 1, \ldots, 4$ and $|1-4\beta_0| \leq |1-4\beta_1|$. Note that here $|\alpha_1| = |1-4\beta_1|$ does not need to be ≥ 1 (as is required when p is a quadratic seminorm; see Experiment 7.2.1 or 7.2.2).

By choosing $\alpha_1 < 1$, we ensure that, in any case, a low-pass filtering is performed, albeit with a varying degree of smoothness depending on the decision d. We take $\beta_0 = 1/5$, and consider $\beta_1 = 1/20$. For this experiment, we consider as input image the synthetic image depicted at the top left of Fig. 11. As before, we compute two levels of decomposition. We display the decision map at the top right, and the corresponding decomposition images in the middle row. The approximation and detail images obtained in the non-adaptive case ($d = 0$) are shown in the bottom row.

7.3 2D SQUARE CASE
In this and the following subsections we consider a 2D decomposition with 4 bands as depicted in Fig. 12. Note that this decomposition is non-separable. Observe that the decomposition in Fig. 12 has the same structure as the one in Fig. 2. However, we have adopted a new labeling y_v, y_h, y_d of the detail bands, replacing the labeling $y^{(1)}$, etc, in Fig. 2. This reflects the fact that the corresponding outputs y'_h, y'_v, y'_d are sometimes called the *horizontal*, the *vertical*, and the *diagonal detail bands*, respectively.

The input images x, y_h, y_v, y_d are obtained by a polyphase decomposition of an original image x_0, that is: $x(m,n) = x_0(2m, 2n)$, $y_v(m,n) = x_0(2m, 2n+1)$, $y_h(m,n) = x_0(2m+1, 2n)$, $y_d(m,n) = x_0(2m+1, 2n+1)$; see also Fig. 13 below. We label the eight samples surrounding $x(m,n)$ by $y_j(m,n)$, $j = 1, \ldots, 8$. For instance, $y_5(m,n) = y_d(m-1,n)$. Note that this labeling is not one-to-one: for example, $y_8(m-1,n) = y_5(m,n)$.

In the experiments below, we compute the detail signals y'_h, y'_v, y'_d with a prediction scheme as depicted in Fig. 12, with $P_h(x) = P_v(x) = x$ and $P_d(x, y_h, y_v) = x + y_h + y_v$. This yields

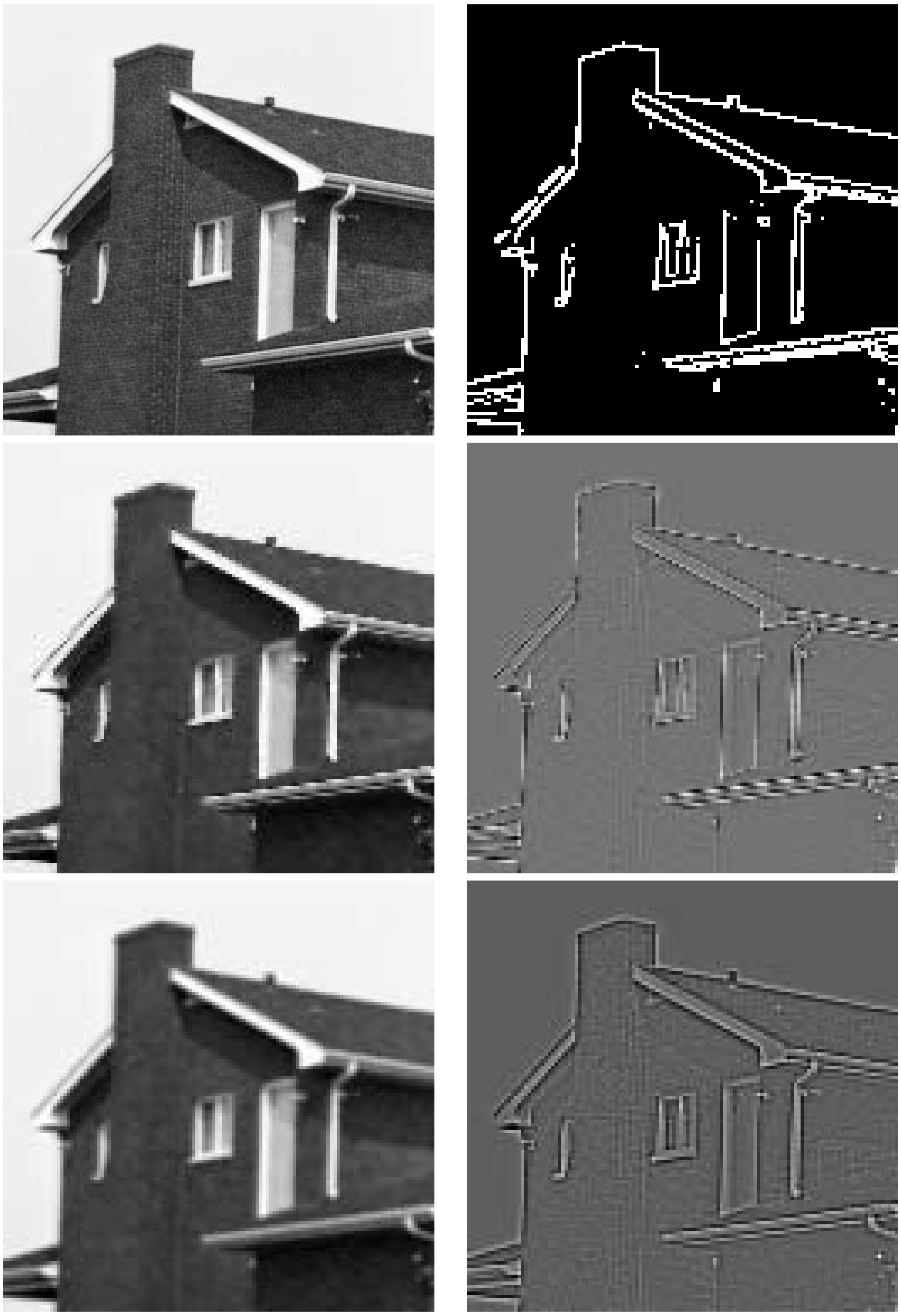

Figure 9: *Decompositions (level* 2*) corresponding with Experiment 7.2.1. Top: input image (left) and decision map (right) using a threshold* $T = 60$*. Middle: approximation (left) and detail (right) images in the adaptive case. Bottom: approximation (left) and detail (right) images in the non-adaptive case (*$d = 0$*).*

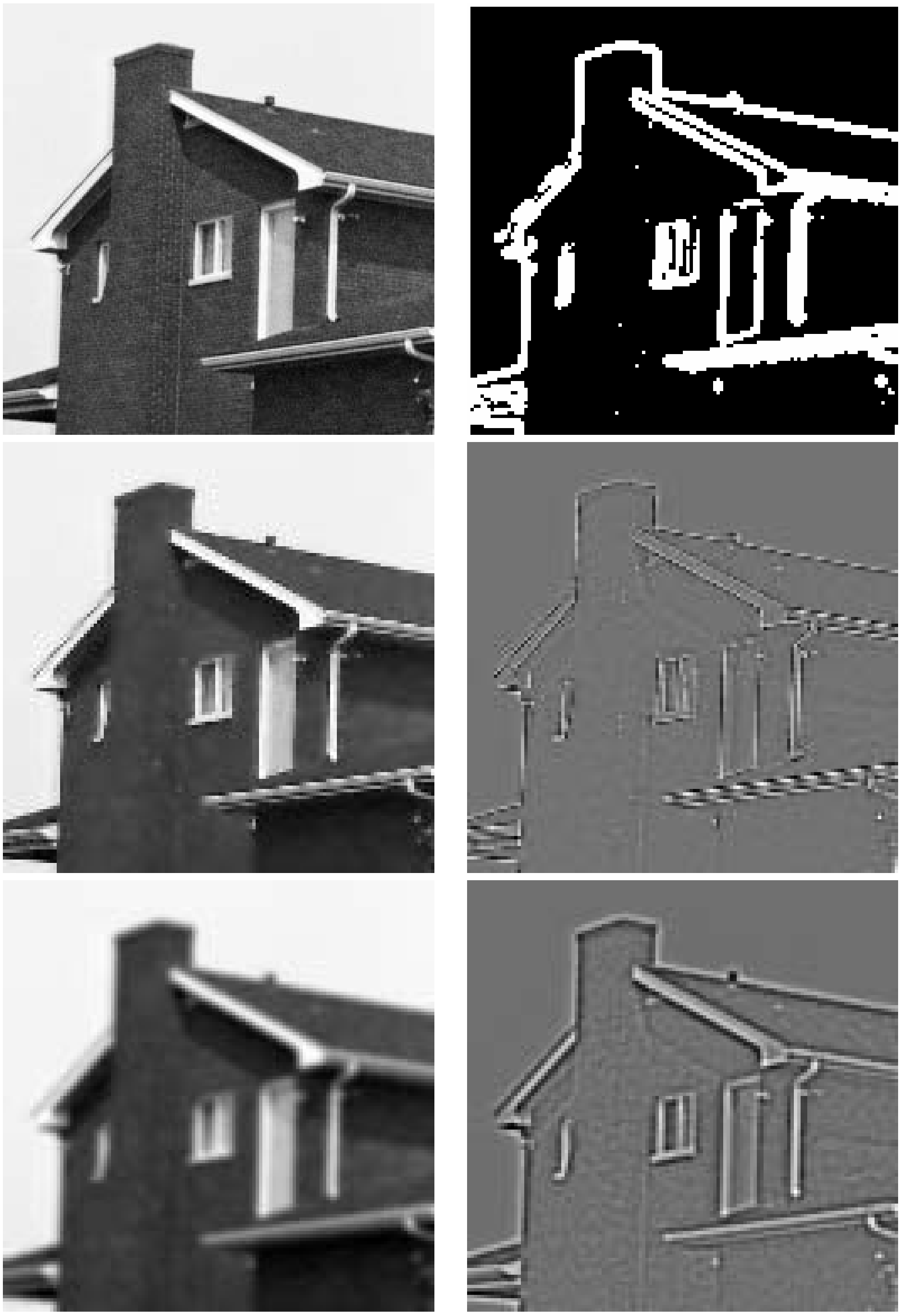

Figure 10: *Decompositions (level* 2*) corresponding with Experiment 7.2.2. Top: input image (left) and decision map (right) using a threshold* $T = 82.5$*. Middle: approximation (left) and detail (right) images in the adaptive case. Bottom: approximation (left) and detail (right) images in the non-adaptive case (*$d = 0$*).*

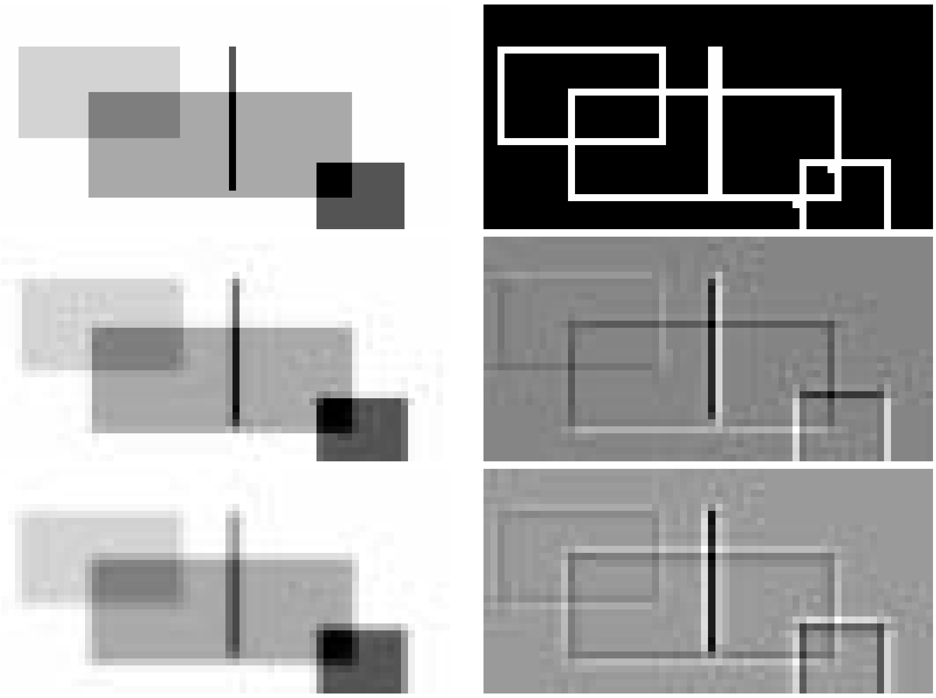

Figure 11: *Decompositions (level 2) corresponding with Experiment 7.2.3. Top: input image (left) and decision map (right) using a threshold $T = 20$. Middle: approximation (left) and detail (right) images in the adaptive case. Bottom: approximation (left) and detail (right) images in the non-adaptive case ($d = 0$).*

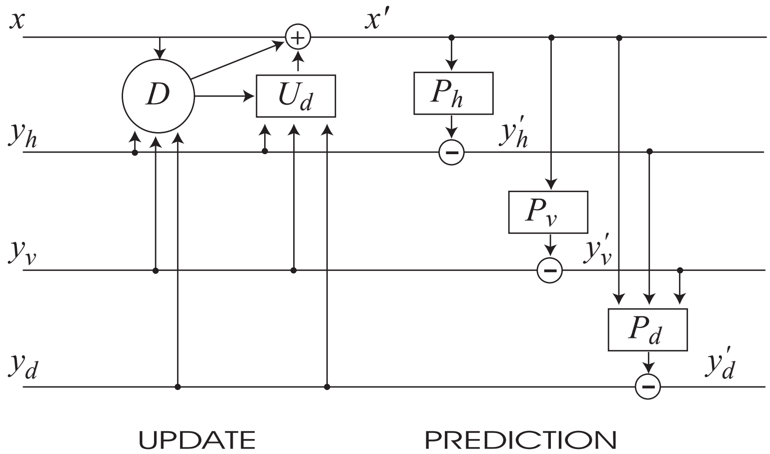

Figure 12: *2D wavelet decomposition comprising an adaptive update lifting step (left) and three consecutive prediction lifting steps (right).*

$x_0(2m-1,2n-1)$ $y_d(m-1,n-1)$ $y_6(m,n)$	$x_0(2m-1,2n)$ $y_h(m-1,n)$ $y_2(m,n)$	$x_0(2m-1,2n+1)$ $y_d(m-1,n)$ $y_5(m,n)$
$x_0(2m,2n-1)$ $y_v(m,n-1)$ $y_3(m,n)$	$x_0(2m,2n)$ $\boldsymbol{x(m,n)}$	$x_0(2m,2n+1)$ $y_v(m,n)$ $y_1(m,n)$
$x_0(2m+1,2n-1)$ $y_d(m,n-1)$ $y_7(m,n)$	$x_0(2m+1,2n)$ $y_h(m,n)$ $y_4(m,n)$	$x_0(2m+1,2n+1)$ $y_d(m,n)$ $y_8(m,n)$

Figure 13: *Labeling of samples in* 3×3 *window centered at* $x_0(2m, 2n)$.

that

$$\begin{aligned} y_h' &= y_h - x' \\ y_v' &= y_v - x' \\ y_d' &= y_d - x' - y_v' - y_h' \,. \end{aligned}$$

Alternatively, $y_d' = y_d + x' - y_v - y_h$.

7.3.1. Experiment (Seminorm $p(\boldsymbol{v}) = |\boldsymbol{a}^T\boldsymbol{v}|$ for 2D square - Fig. 14)

In the first experiment of this subsection we consider the seminorm given by

$$p(\boldsymbol{v}) = \left|\boldsymbol{a}^T\boldsymbol{v}\right| = \left|\sum_{j=1}^{8} a_j v_j\right| \,, \tag{7.1}$$

where $\boldsymbol{a} = (a_1, \ldots, a_8)^T$ with

$$a_1 + a_2 + \cdots + a_8 \neq 0 \,. \tag{7.2}$$

Note that this last condition guarantees that the scheme is truly adaptive: if it is not satisfied, then $p(\boldsymbol{v})$ does not depend on x. In Proposition 6.1 we have derived necessary and sufficient conditions for perfect reconstruction. In fact, we must choose the filter coefficients

$$\boldsymbol{\beta}_d = \gamma_d \boldsymbol{a} \quad \text{with} \quad 0 < |1 - \gamma_0 \sum_j a_j| \leq |1 - \gamma_1 \sum_j a_j| \,.$$

We have also seen in Section 6 that for the 1D case, one can choose the coefficients a_j in such a way that the decision maps 'ignores' polynomials up to a given degree; the seminorm $p(\boldsymbol{v})$ corresponds with a derivative filter in this case. It is easy to see that this can be extended to 2D images. For example, the expression $|x - y_h - y_v + y_d|$ corresponds with a first-order derivative with respect to both horizontal and vertical directions. To obtain this expression, one must choose $a_1 = a_4 = 1$, $a_8 = -1$ and $a_j = 0$ for the other coefficients. For the second-order derivative (with respect to both directions) we can choose $a_1 = a_2 = a_3 = a_4 = 1$ and $a_5 = a_6 = a_7 = a_8 = -1/2$. In this case $\sum_j a_j = 2$ and we must choose γ_0, γ_1 such that $0 \leq |1 - 2\gamma_0| \leq |1 - 2\gamma_1|$. In this experiment we consider this latter choice of $\boldsymbol{a}$, i.e., $\boldsymbol{a} = (1, 1, 1, 1, -\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2})^T$, and we choose $\gamma_0 = 1/4$ and $\gamma_1 = 0$.

As input image we take the synthetic image depicted at the top left of Fig. 14. In the second row we show the approximation and detail image at the first level. For comparison, we show the corresponding decomposition images obtained with a non-adaptive scheme (with $\boldsymbol{\beta}_0$) in the bottom row. Note that the adaptive scheme yields an approximation which preserves well the edges, and a detail image with less oscillatory effects than its non-adaptive counterpart.

Note also from the decision map that the filter $\boldsymbol{a}$ does not 'see' horizontal and vertical edges. Such edges are well preserved in the adaptive as well as in the non-adaptive case.

7.3.2. Experiment (Quadratic seminorm for 2D square - Fig. 15)

Next, we consider the case where the decision criterion is based on the following weighted l^2-norm:

$$p(\boldsymbol{v}) = \left(\sum_{i=1}^{4} |v_i|^2 + \frac{1}{2} \big(\sum_{i=5}^{8} |v_i|^2 \big) \right)^{1/2} .$$

This corresponds with Example 5.6(a) where $M = \operatorname{diag}\big(1, 1, 1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\big)$. Thus the threshold criterion holds if we choose $\boldsymbol{\beta}_d = \mu_d \left(1, 1, 1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)^T$. For smooth regions where $d = 0$, we compute μ_0 from (5.8), which gives $\mu_0 = 6/41$. For $d = 1$, we choose $\mu_1 = 0$, meaning that the approximation image is not modified in high gradient regions.

Fig. 15 illustrates the corresponding images resulting from this experiment.

7.4 Switching between horizontal and vertical filters

So far we have only considered decision maps using one seminorm. In this subsection we build a decision map that uses *two* seminorms, one governing the horizontal gradient and one for the vertical gradient. More specifically, we define

$$p_h(\boldsymbol{v}) = |v_1 + v_3| \quad \text{and} \quad p_v(\boldsymbol{v}) = |v_2 + v_4| \,, \tag{7.3}$$

corresponding, respectively, with a *horizontal* and *vertical* derivative filter of second order. We choose the decision map

$$D(x, y_h, y_v, y_d) = [p_h(\boldsymbol{v}) \leq p_v(\boldsymbol{v})] \,. \tag{7.4}$$

We assume that the update filters U_d have 4 taps corresponding with the detail coefficients labeled by $y_1, \ldots, y_4$. The filter coefficients $\boldsymbol{\beta}_d$ are chosen as follows:

$$\boldsymbol{\beta}_d = (\beta_d, \gamma_d, \beta_d, \gamma_d, 0, 0, 0, 0)^T \quad \text{for } d = 0, 1 \,. \tag{7.5}$$

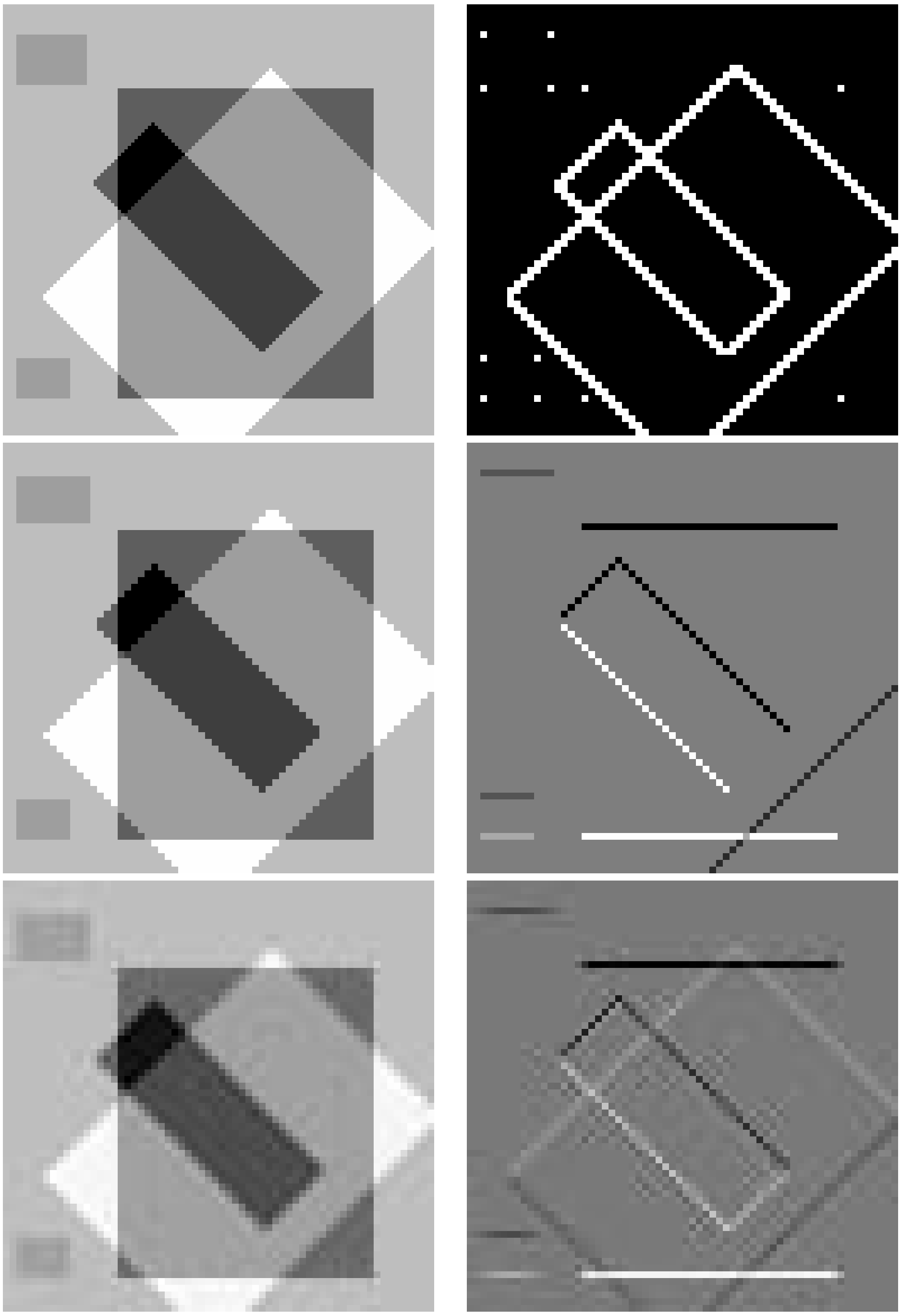

Figure 14: *Decompositions (level* 1*) corresponding with Experiment 7.3.1. Top: input image (left) and decision map (right) using a threshold* $T = 10$*. Middle: approximation (left) and horizontal detail (right) images in the adaptive case. Bottom: approximation (left) and horizontal detail (right) images in the non-adaptive case (*$d = 0$*).*

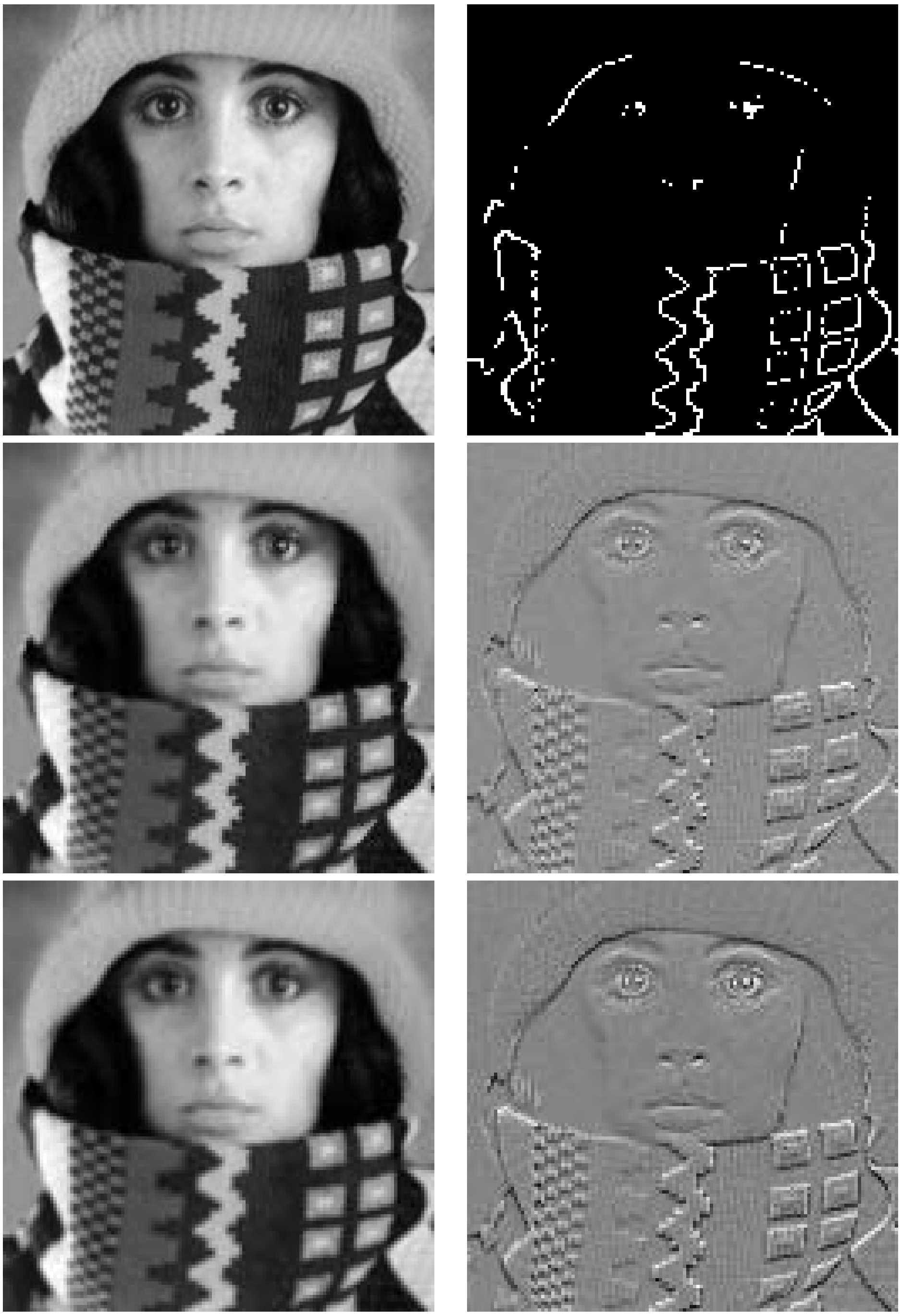

Figure 15: *Decompositions (level* 1*) corresponding with Experiment 7.3.2. Top: original (left) and decision map (right) using a threshold $T = 61$. Middle: approximation (left) and horizontal detail (right) images in the adaptive case. Bottom: approximation (left) and horizontal detail (right) images in the fixed case ($d = 0$).*

This means in particular that only the horizontal and vertical neighbours y_1, y_2, y_3, y_4 are used to update the approximation signal. For example, if $d = 0$, then equation (2.15) reduces to:

$$x' = \alpha_0 x + \beta_0(y_1 + y_3) + \gamma_0(y_2 + y_4)\,. \tag{7.6}$$

Let $\boldsymbol{v}'$ be the gradient vector at synthesis, i.e., $v'_j = x' - y_j$. A straightforward computation shows that

$$\begin{aligned} v'_1 + v'_3 &= (1 - 2\beta_d)(v_1 + v_3) - 2\gamma_d(v_2 + v_4) && (7.7)\\ v'_2 + v'_4 &= -2\beta_d(v_1 + v_3) + (1 - 2\gamma_d)(v_2 + v_4)\,. && (7.8) \end{aligned}$$

If we can choose the coefficients $\beta_0, \gamma_0, \beta_1, \gamma_1$ in such a way that

$$[p_h(\boldsymbol{v}) \le p_v(\boldsymbol{v})] = [p_h(\boldsymbol{v}') \le p_v(\boldsymbol{v}')]\,,$$

then we can recover the original decision in (7.4) from the transformed gradient vectors, and hence perfect reconstruction is possible in this case.

7.1 Proposition. *To have perfect reconstruction it is sufficient that*

$$\begin{aligned} &0 \le \beta_0 < \frac{1}{4} \le \gamma_0 < \frac{1}{2} \quad \textit{and} \quad \beta_0 + \gamma_0 < \frac{1}{2}\\ &0 \le \gamma_1 < \frac{1}{4} \le \beta_1 < \frac{1}{2} \quad \textit{and} \quad \beta_1 + \gamma_1 < \frac{1}{2}\,, \end{aligned}$$

and in this case the decision d can be recovered at synthesis from $d = [p_h(\boldsymbol{v}') \le p_v(\boldsymbol{v}')]$.

Proof. We introduce the following notation for the horizontal and vertical component of the gradient:

$$H = v_1 + v_3 \text{ and } V = v_2 + v_4\,,$$

and the same for H' and V'. From (7.7)-(7.8) we get that

$$\begin{aligned} H' &= (1 - 2\beta_d)H - 2\gamma_d V && (7.9)\\ V' &= -2\beta_d H + (1 - 2\gamma_d)V\,. && (7.10) \end{aligned}$$

We will show that if the decision map D in (7.4) returns $d = 1$, i.e., $|H| \le |V|$, then it also follows that $|H'| \le |V'|$. The proof for the case where the decision map (7.4) returns $d = 0$ is analogous. We distinguish 4 different cases.

(i) $0 \le H < V$: then

$$\begin{aligned} 0 < V' &= 2\beta_1(V - H) + (1 - 2\gamma_1 - 2\beta_1)V\\ H' &= -(1 - 2\beta_1)(V - H) + (1 - 2\gamma_1 - 2\beta_1)V\,. \end{aligned}$$

Since $(1 - 2\gamma_1 - 2\beta_1)V$ is positive and

$$0 < (1 - 2\beta_1)(V - H) \le 2\beta_1(V - H)\,,$$

it follows that $|H'| \le |V'|$.

(ii) $H \ge 0$ and $V < 0$: thus H' in (7.9) comprises two positive terms, whereas V' in (7.10) comprises two negative terms. Since $1 - 2\beta_1 < 2\beta_1$ and $2\gamma_1 < 1 - 2\gamma_1$, it is obvious that $|H'| \le |V'|$.

(iii) $H < 0$ and $V \ge 0$: now H' in (7.9) comprises two negative terms and V' in (7.10) comprises two positive terms. The same reasoning as in (ii) yields that $|H'| \le |V'|$.

(iv) $H < 0$ and $V < 0$: from $|H| \le |V|$ we conclude that $-H \le -V$, hence $H - V \ge 0$. Now

$$\begin{aligned} H' &= (1 - 2\beta_1)(H - V) + (1 - 2\beta_1 - 2\gamma_1)V\\ V' &= -2\beta_1(H - V) + (1 - 2\beta_1 - 2\gamma_1)V\,. \end{aligned}$$

Using a similar argument as in (i) we conclude again that $|H'| \le |V'|$. □

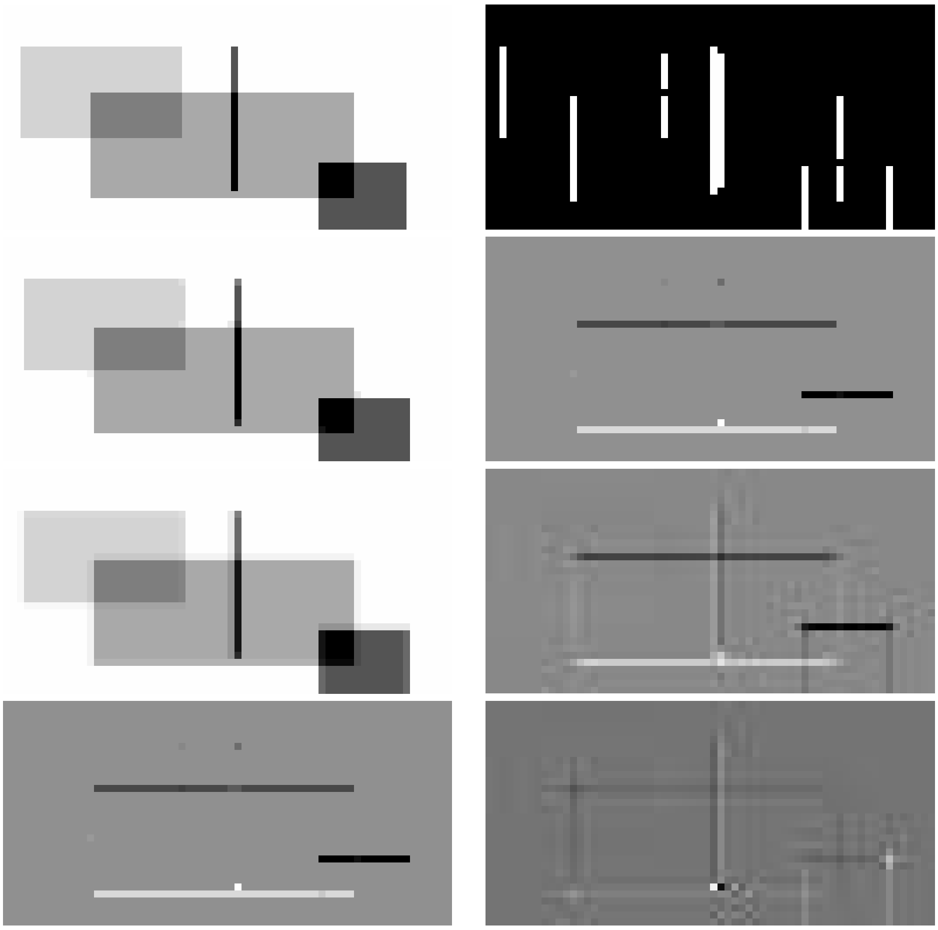

Figure 16: *Decompositions (level* 1*) corresponding with Experiment 7.4.1. Top: input image (left) and decision map (right). Second row: approximation (left) and horizontal detail (right) images in the adaptive case. Third row: approximation (left) and horizontal detail (right) images in the non-adaptive case ($d = 0$). Bottom: diagonal detail images in the adaptive (left) and non-adaptive (right) cases.*

This case has the following geometric interpretation. If $p_h(\boldsymbol{v}) \leq p_v(\boldsymbol{v})$, and hence $d = 1$, then the vertical derivative $2x - y_2 - y_4$ dominates in the absolute value the horizontal derivative $2x - y_1 - y_3$, and in this case we choose the filter in such a way that it causes a stronger smoothing in horizontal than vertical direction, i.e., $\gamma_1 < \beta_1$.

7.4.1. Experiment (Switching between horizontal and vertical filters - Fig. 16)
We choose the filter coefficients like in (7.5) with $\beta_0 = \gamma_1 = 0$ and $\beta_1 = \gamma_0 = 1/4$. Obviously the conditions in Proposition 7.1 are satisfied. We apply this scheme to the original image depicted at the top left of Fig. 16. The decision map is shown at the top right; the approximation and vertical detail images are shown in the second row. The diagonal detail is displayed in the bottom row, on the left. We compare this scheme with the non-adaptive scheme where we perform an isotropic filtering (in the vertical and horizontal directions), i.e, $\beta = \gamma = 1/8$. The corresponding approximation and horizontal images are displayed in the third row of Fig. 16, and the diagonal detail image on the right of the bottom row.

7.5 Entropy computations

To evaluate the potential of these adaptive decompositions for image and video compression, we apply them to some well-known test images such as *house, peppers* and *lena* as well as to the *trui* image depicted in Fig. 15, and compute the entropies of the resulting detail images after a uniform quantisation with 256 bins. We consider the following entropies:

- the *information* or *Shannon entropy* given by

$$h = -\sum_j p_j \log_2 p_j \tag{7.11}$$

  where p_j is the frequency value of bin j;

- the *Coifman-Wickerhauser entropy* [32]:

$$h = -\sum_{j=1}^{J} \frac{|y_j|^2}{\|\boldsymbol{y}\|^2} \log_2 \frac{|y_j|^2}{\|\boldsymbol{y}\|^2}\,; \tag{7.12}$$

  where $\boldsymbol{y} = \{y_j\}_{j=1}^J$ is the detail sequence and $\|\cdot\|$ represents the l^2-norm.

Table 1 and Table 2 correspond to the entropy results obtained with (7.11) and (7.12), respectively. Both tables show that in most cases the entropy values resulting from the adaptive decomposition are lower than in the non-adaptive case.

Experiment		house	peppers	lena	trui
7.2.1	adaptive	4.952	4.640	4.896	4.924
	$d = 0$	5.449	4.883	5.210	5.481
7.2.2	adaptive	4.997	4.883	5.071	5.119
	$d = 0$	5.240	5.020	5.098	5.538
7.2.3	adaptive	5.164	4.872	5.038	5.205
	$d = 0$	5.449	4.883	5.210	5.481
7.3.1	adaptive	4.823	4.529	4.745	5.171
	$d = 0$	4.801	4.573	4.778	5.221
7.3.2	adaptive	4.946	4.676	4.859	5.317
	$d = 0$	5.252	4.584	5.058	5.312
7.4.1	adaptive	4.868	4.566	4.814	5.211
	$d = 0$	4.959	4.920	4.896	5.317

Table 1: *Entropy values for the adaptive and non-adaptive ($d = 0$) decomposition schemes using the information entropy in* (7.11).

8. Conclusions

In this paper, we have generalised the adaptive update lifting scheme that was introduced by two of us in [26]. An important feature of this scheme is that it is neither causal nor that it requires any bookkeeping in order to perform perfect reconstruction. The scheme discussed here comprises two important extensions compared to [26]: (*i*) we are able to deal with update filters of arbitrary length (and arbitrary dimension), whereas the filters in [26] were restricted to length 2; (*ii*) the decision map can be based on an arbitrary seminorm of the gradient vector.

Experiment		house	peppers	lena	trui
7.2.1	adaptive	12.612	12.141	14.807	13.164
	$d = 0$	13.055	12.342	14.975	13.285
7.2.2	adaptive	12.692	12.414	15.009	13.394
	$d = 0$	12.967	12.522	14.928	13.407
7.2.3	adaptive	12.888	12.393	14.928	13.258
	$d = 0$	13.055	12.342	14.975	13.285
7.3.1	adaptive	12.126	11.755	14.619	12.535
	$d = 0$	12.159	11.782	14.664	12.557
7.3.2	adaptive	12.347	11.927	14.763	12.716
	$d = 0$	12.822	11.957	14.863	12.890
7.4.1	adaptive	12.247	11.816	14.738	12.584
	$d = 0$	12.355	12.275	14.753	12.634

Table 2: *Entropy values for the adaptive and non-adaptive ($d = 0$) decomposition schemes using the Coifman-Wickerhauser entropy in* (7.12).

We have treated different seminorms, namely the l^p-norms for $p = 1, 2, \infty$, arbitrary quadratic seminorms, as well as seminorms based on derivative filters. For all these cases we have been able to derive conditions that guarantee perfect reconstruction. In Section 7 we have performed various simulations to show the differences between our adaptive decompositions and the non-adaptive decompositions based on a fixed lifting scheme. The entropy computations in § 7.5 demonstrate the potential usefulness in image and video compression applications.

There are several issues that need to be addressed in the near future. We intend to incorporate our adaptive wavelet decompositions in existing compression schemes both for image and video coding. An important issue in this respect is quantization. We are currently investigating robustness properties of our scheme. Another application area that deserves further exploration is denoising. We believe that the seminorms based on derivative filters could be useful in this context. Again, robustness of the scheme is highly relevant in this case. Finally, we are considering extensions of the current framework with adaptive *prediction* lifting.

References

1. Atkinson, K. E. *An introduction to numerical analysis (2nd ed).* John Wiley & Sons, New York, 1989.
2. Burt, P. J., and Adelson, E. H. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications 31* (1983), 532–540.
3. Candès, E. The curvelet transform for image denoising. In *Proceedings of IEEE International Conference on Image Processing* (Thessaloniki,Greece, October 2001).
4. Cha, H., and Chaparro, L. F. Adaptive morphological representation of signals: Polynomial and wavelet methods. *Multidimensional Systems and Signal Processing 8* (1997), 249–271.
5. Claypoole, R. L., Baraniuk, R. G., and Nowak, R. D. Adaptive wavelet transforms via lifting. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Seattle, Washington, May 12-15, 1998).
6. Claypoole, R. L., Baraniuk, R. G., and Nowak, R. D. Lifting construction of non-linear wavelet transforms. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis* (Pittsburgh, Pennsylvania, October 6-9, 1998), pp. 49–52.
7. Claypoole, R. L., Baraniuk, R. G., and Nowak, R. D. Adaptive wavelet transforms via lifting. Technical Report 9304, Department of Electrical and Computer Engineering, Rice University, Houston, Texas, April 1999.
8. Claypoole, R. L., Davis, G., Sweldens, W., and Baraniuk, R. D. Nonlinear wavelet transforms for image coding. In *Proceedings of the 31st Asilomar Conference on Signals, Systems, and Computers, Volume 1* (1997), pp. 662–667.
9. Combettes, P. L., and Pesquet, J.-C. Convex multiresolution analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20* (1998), 1308–1318.
10. Daubechies, I. *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.
11. de Queiroz, R. L., Florêncio, D. A. F., and Schafer, R. W. Nonexpansive pyramid for image coding using a nonlinear filterbank. *IEEE Transactions on Image Processing 7* (1998), 246–252.
12. Donoho, D. L. Wedgelets: nearly minimax estimation of edges. Tech. rep., Statistics Department, Stanford University, 1997.
13. Donoho, D. L. Orthonormal ridgelets and linear singularities. Tech. rep., Statistics Department, Stanford University, 1998.
14. Egger, O., and Li, W. Very low bit rate image coding using morphological operators and adaptive decompositions. In *Proceedings of the IEEE International Conference on Image Processing* (Austin, Texas, 1994), pp. 326–330.
15. Egger, O., Li, W., and Kunt, M. High compression image coding using an adaptive morphological subband decomposition. *Proceedings of the IEEE 83* (1995), 272–287.

16. Florêncio, D. A. F. *A New Sampling Theory and a Framework for Nonlinear Filter Banks.* PhD thesis, School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia, 1996.

17. Florêncio, D. A. F., and Schafer, R. W. A non-expansive pyramidal morphological image coder. In *Proceedings of the IEEE International Conference on Image Processing* (Austin, Texas, 1994), pp. 331–335.

18. Gerek, O. N., and Çetin, A. E. Adaptive polyphase subband decomposition structures for image compression. *IEEE Transactions on Image Processing 9* (October 2000), 1649–1659.

19. Hampson, F. J. *Méthodes Non Linéaires en Codage d'Images et Estimation de Mouvement.* PhD thesis, l'Université Paris XI Orsay, Paris, France, 1997.

20. Hampson, F. J., and Pesquet, J.-C. A nonlinear subband decomposition with perfect reconstruction. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Atlanta, Georgia, May 7-10, 1996), pp. 1523–1526.

21. Hampson, F. J., and Pesquet, J.-C. M-band nonlinear subband decompositions with perfect reconstruction. *IEEE Transactions on Image Processing 7* (1998), 1547–1560.

22. Heijmans, H. J. A. M., and Goutsias, J. Nonlinear multiresolution signal decomposition schemes: Part II: morphological wavelets. *IEEE Transactions on Image Processing 9*, 11 (2000), 1897–1913.

23. Horn, R. A., and Johnson, C. R. *Matrix Analysis.* Cambridge University Press, New York, 1985.

24. Le Pennec, E., and Mallat, S. Image compression with geometrical wavelets. In *Proceedings of the IEEE International Conference on Image Processing* (Vancouver, September 2000).

25. Mallat, S. *A Wavelet Tour of Signal Processing.* Academic Press, San Diego, California, 1998.

26. Piella, G., and Heijmans, H. J. A. M. Adaptive lifting schemes with perfect reconstruction. *IEEE Transactions on Signal Processing* (to appear).

27. Sweldens, W. The lifting scheme: A new philosophy in biorthogonal wavelet constructions. In *Wavelet Applications in Signal and Image Processing III* (1995), A. F. Lain and M. Unser, Eds., Proceedings of SPIE, vol. 2569, pp. 68–79.

28. Sweldens, W. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computational Harmonic Analysis 3* (1996), 186–200.

29. Sweldens, W. The lifting scheme: A construction of second generation wavelets. *SIAM Journal of Mathematical Analysis 29* (1997), 511–546.

30. Trappe, W., and Liu, K. Adaptivity in the lifting scheme. In *33th Conference on Information Sciences and Systems* (Baltimore, March 1999), pp. 950–955.

31. Vetterli, M., and Kovačević, J. *Wavelets and Subband Coding.* Prentice Hall, Englewood Cliffs, New Jersey, 1995.

32. Wickerhauser, V., and Coifman, R. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory 38* (1992), 713–718.