Wavelet Transforms that Map Integers to Integers

Robert Calderbank, AT&T Research

Ingrid Daubechies, Princeton University

Wim Sweldens, Bell Laboratories Innovations, Lucent Technologies

Boon-Lock Yeo, IBM T.J. Watson Research Center

1 Introduction

Wavelets and wavelet packets are used in a variety of applications, including image compression [1, 2]. In most applications, the wavelet filters that are used have floating point coefficients. For instance, if one prefers to use orthonormal filters with an assigned number N ($N \geq 2$) of vanishing moments and minimal filter length, then the resulting filter coefficients are real numbers which can be computed with high precision, but for which we don't even have a closed form expression if $N > 3$ [3]. When the input data consist of sequences of integers (as is the case for images), the resulting filtered outputs no longer consist of integers. Yet, for lossless coding it would be of interest to be able to characterize the output completely again with integers. Relaxing the constraint of orthonormality on the wavelet filters makes it possible to obtain filter coefficients that are dyadic rationals [4, 5]; up to scaling these filters can be viewed as mapping integers to integers. This rescaling amplifies the dynamic range of the data considerably, however, so that this does not lead to reasonable lossless compression.

In this paper, we present two approaches that lead to wavelet transforms that map integers to integers, and which can be used for lossless coding.

The first approach, in section 2, is inspired by the precoding invented by Rajiv Laroia [6]. When integer data undergo causal filtering with a leading filter coefficient equal to 1, this precoding makes small (non-integer) adjustments to the input data resulting in integer outputs after filtering. In §2.1, we try to adapt this method to the subband filtering inherent to a wavelet transform. Even though this approach cannot be made to work for other wavelet bases than the Haar basis, it leads to a reformulation of the original problem, by including the possibility of an expansion coefficient. Subsection 2.2 shows how the specific structure of wavelet filters can be exploited to make this expansion coefficient work. In §2.3, we generalize the expansion idea to propose different expansion factors for the low and the high pass filter: for the high pass filter, we can afford a larger expansion factor because the dynamic range of its output tends to be smaller anyway.

A completely different approach (although there are some points of contact) is taken in section 3. An example in [7] of a wavelet transform that maps integers to integers is seen to be an instance of a large family of similar transforms, obtained by combining the lifting constructions proposed in [8] with rounding-off in a reversible way. In §3.1 we review the S transform, and in §3.2 its generalizations, the TS transform, and the S+P transform. In §3.3 we briefly review lifting, and in §4 we show how a decomposition in lifting steps can be turned into an integer transform. In general, this transform exhibits some expansion factors as well. However, the product of the low- and high-pass expansion factors is now always equal to 1, unlike what happened in §2. In §3.5, we show many examples.

Finally, in §4 we show applications of the two approaches to lossless image coding, and we compare our results with other results in the literature.

Except where specified otherwise, our notations conform with the standard notations for filters associated with orthonormal or biorthogonal wavelet bases, as found in e.g., [3].

2 Mapping integers to integers by introducing expansion factors

2.1 Precoding for subband filtering

We start by reviewing briefly Laroia's precoder [6]. Suppose we have to filter a sequence of integers $(a_n)_{n\in\mathbb{Z}}$, with a causal filter (i.e., $h_k = 0$ for $k < 0$) and leading coefficient 1 (i.e., $h_0 = 1$). If the h_k are not themselves integers, then the filtered output $b_n = a_n + \sum_{k=1}^{\infty} h_k a_{n-k}$ will in general not consist of integers either. Laroia's precoder replaces the integers a_n by non-integers a'_n by introducing small shifts r_n, in such a way that the resulting b'_n are integers. More concretely, define

$$a'_n = a_n - r_n$$
$$r_n = \left\{\sum_{k-1}^{\infty} h_k a'_{n-k}\right\} ,$$

where the symbol $\{x\} = x - \lfloor x \rfloor$ stands for the fractional part of x, and $\lfloor x \rfloor$ for the largest integer not exceeding x. In practice, both the a_n and the h_k are non-zero for only finite ranges of the indices n or k, so that the infinite recursion is not a problem: if $a_n = 0$ for $n < N$, then so are a'_n and r_n. If we now compute $b'_n = a'_n + \sum_{k=1}^{\infty} h_k a'_{n-k}$, then we immediately see that

$$b'_n = a_n + \left\lfloor \sum_{k=1}^{\infty} h_k a'_{n-k} \right\rfloor$$

is an integer.

This approach can also be used if $h_0 \neq 1$. In that case, if we have

$$\begin{aligned} a'_n &= a_n - r_n \\ r_n &= \left\{(h_0 - 1)a'_n + \sum_{k=1}^{\infty} h_k a'_{n-k}\right\} , \end{aligned} \tag{2.1}$$

then the

$$\begin{aligned} b'_n &= \sum_{k=0}^{\infty} h_k a'_{n-k} \\ &= a'_n + \left\lfloor (h_0 - 1)a'_n + \sum_{k=1}^{\infty} h_k a'_{n-k} \right\rfloor + r_n \\ &= a_n + \left\lfloor (h_0 - 1)a'_n + \sum_{k=1}^{\infty} h_k a'_{n-k} \right\rfloor \end{aligned}$$

are integers again. The situation is now slightly more complicated, because the r_n are defined implicitly. Since both a_n and the a'_{n-k} for $k \geq 0$ are known when r_n needs to be determined, we can reformulate (2.1) as

$$r_n = \{(1 - h_0)r_n + y_n\}$$

with $y_n = \left\{(h_0 - 1)a_n + \sum_{k=1}^{\infty} h_k a'_{n-k}\right\}$. Whether or not the equation

$$r = \{\alpha r + y\} \tag{2.2}$$

has a solution for arbitrary $y \in [0, 1)$ depends on the size and sign of α. As shown in Figure 1, a solution fails to exist for some choices of y if $0 < \alpha < 2$; for $\alpha \leq 0$ or $\alpha \geq 2$ there always exists at least one solution. This range for α where (2.2) always has a solution corresponds to $|h_0| \geq 1$. When faced with having to use the precoder for cases where $|h_0| < 1$, a simple expedient is to consider a renormalized filter, with $\widetilde{h}_n = \alpha h_n$, so that $|\widetilde{h}_0| \geq 1$. (For instance, $\alpha = h_0^{-1}$ will do.)

In the case of the wavelet transform, we work with several filters, followed by decimation. In this paper, we shall always work with 2-channel filter banks. Assume that the two filters H and G are FIR, with $h_n = g_n = 0$ if $n < 0$ or $n \geq 2N$. Then we have

$$\begin{aligned} s_n^1 &= \sum_k h_{2n-k} s_k^0 \\ d_n^1 &= \sum_k g_{2n-k} s_k^0 \ , \end{aligned} \tag{2.3}$$

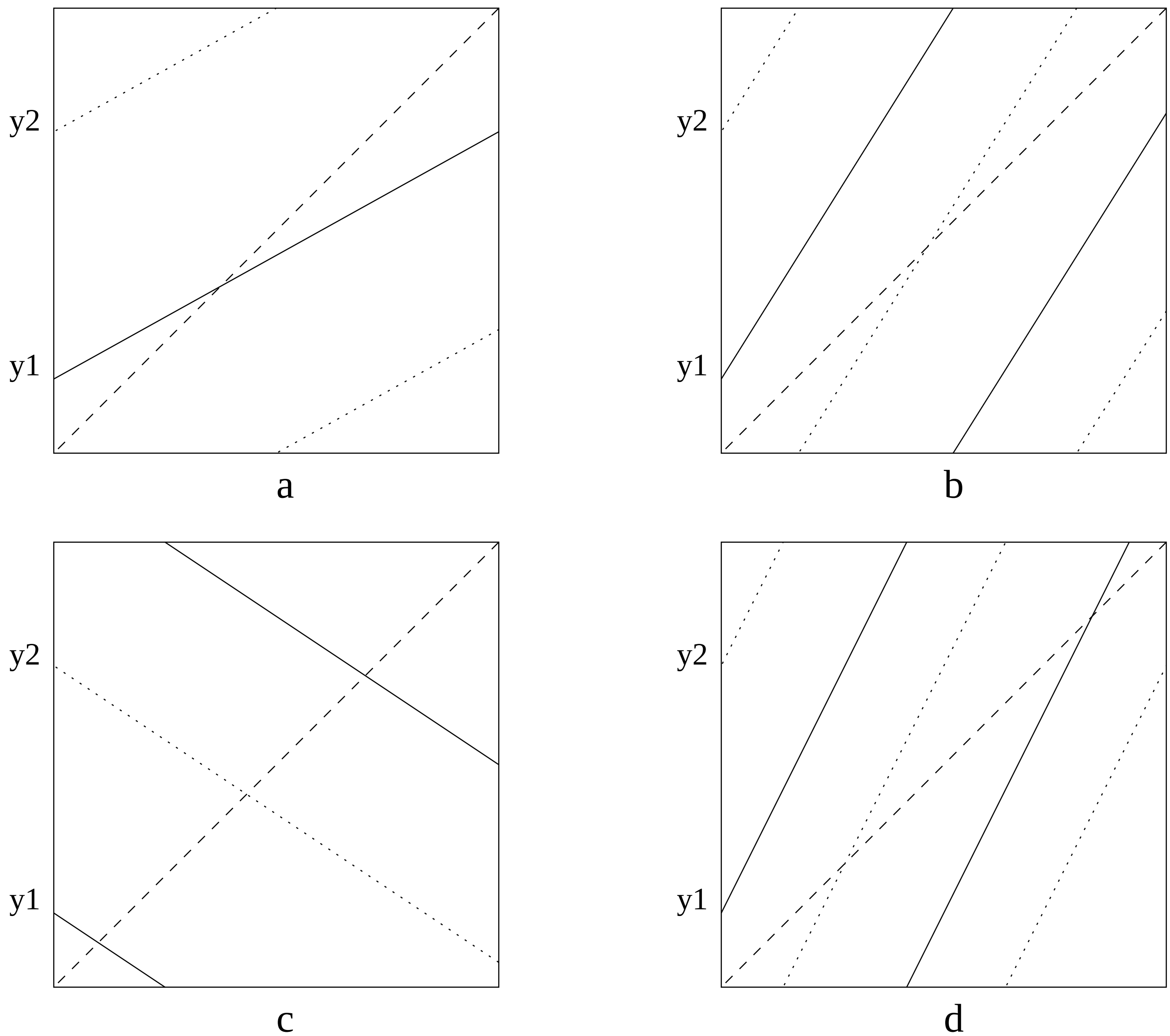

Figure 1:

a. For $0 < \alpha \leq 1$, the equation $r = \{\alpha r + y\}$ has solutions for $y = y_1$ but not for $y = y_2$, since $\{\alpha r + y_1\}$ intersects the diagonal, but $\{\alpha r + y_2\}$ doesn't.

b. For $1 \leq \alpha < 2$, the situation is reversed.

c. For $\alpha \leq 0$ we always have at least one solution.

d. Likewise for $\alpha \geq 2$.

which can be rewritten as

$$\begin{aligned} \begin{pmatrix} s_n^1 \\ \\ d_n^1 \end{pmatrix} &= \sum_k \begin{pmatrix} h_{2(n-k)} & h_{2(n-k)+1} \\ \\ g_{2(n-k)} & g_{2(n-k)+1} \end{pmatrix} \begin{pmatrix} s_{2k}^0 \\ \\ s_{2k-1}^0 \end{pmatrix} \\ &= \sum_{k=0}^N \begin{pmatrix} h_{2k} & h_{2k+1} \\ \\ g_{2k} & g_{2k+1} \end{pmatrix} \begin{pmatrix} s_{2(n-k)}^0 \\ \\ s_{2(n-k)-1}^0 \end{pmatrix} \end{aligned} \tag{2.4}$$

(This corresponds to a polyphase decomposition of the filters; see e.g., [9, 10].) If we define $H_k = \begin{pmatrix} h_{2k} & h_{2k+1} \\ \\ g_{2k} & g_{2k+1} \end{pmatrix}$, with $0 \le k \le N-1$, then (2.4) can be read as a 2D analog to the situation above; now each a_n or b_n is a 2-vector, with $a_n = (s_{2n}^0 s_{2n-1}^0)^T$, and $b_n = (s'_n d'_n)^T$.

For an arbitrary 2-vector $a = (a_1 a_2)^T$ we introduce the notations

$$\{a\} = \begin{pmatrix} \{a_1\} \\ \\ \{a_2\} \end{pmatrix}, \lfloor a \rfloor = \begin{pmatrix} \lfloor a_1 \rfloor \\ \\ \lfloor a_2 \rfloor \end{pmatrix} .$$

We can now generalize Laroia's precoder by defining

$$\begin{aligned} a'_n &= a_n - r_n \\ r_n &= \left\{ (H_0 - Id) a'_n + \textstyle\sum_{k=1}^{N-1} H_k a'_{n-k} \right\} = \{ (Id - H_0) r_n + y_n \} \ , \end{aligned} \tag{2.5}$$

with $y_n = \left\{ (H_0 - Id) a_n + \sum_{k=1}^{N-1} H_k a'_{n-k} \right\}$. If the equations (2.5) for r_n can indeed be solved, then we find, as before, that the

$$b'_n = \sum_{k=0}^{n-1} H_k a'_{n-k}$$

consist of only integer entries. It is not clear, however, that the equation

$$r = \{ (Id - H_0) r + y \} \tag{2.6}$$

always has solutions, for any $y \in [0,1) \times [0,1)$. Indeed, for the Haar case, where

$$H_0 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

one checks that (2.6) has no solution if $y_1 = \frac{1}{4}, y_2 = \frac{1}{2}$.

The simple expedient of "renormalizing" H_0 to αH_0 works here as well. Let us replace the orthonormal filtering (2.3) and its inverse

$$s_k^0 = \sum_n \left[h_{2n-k} s_n^1 + g_{2n-k} d_n^1\right]$$

by scaled versions:

$$\tilde{s}_n^1 = \alpha \textstyle\sum_k h_{2n-k} s_k^0, \;\; \tilde{d}_n^1 = \beta \textstyle\sum_k g_{2n-k} s_k^0$$

$$s_k^0 = \textstyle\sum_n \left[\alpha^{-1} h_{2n-k} \tilde{s}_n^1 + \beta^{-1} g_{2n-k} \tilde{d}_n^1\right] \ .$$

The polyphase regrouping then uses the matrices

$$\widetilde{H}_k = \begin{pmatrix} \alpha h_{2k} & \alpha h_{2k+1} \\ \beta g_{2k} & \beta g_{2k+1} \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} H_k \ ;$$

the corresponding equations for r_1, r_2 become

$$\begin{aligned} r_1 &= \{(1-\alpha h_0) r_1 + (y_1 - \alpha h_1 r_2)\} \\ r_2 &= \{(1-\beta g_1) r_2 + (y_2 - \beta g_0 r_1)\} \ . \end{aligned} \tag{2.7}$$

For the special case where the wavelet basis is the Haar basis, i.e., $h_0 = h_1 = \frac{1}{\sqrt{2}} = g_1 = -g_0$, with all other $h_n, g_n = 0$, we can choose $\alpha = \beta = \sqrt{2}$. The system for r_1, r_2 reduces then to

$$r_1 = \{y_1 - r_2\}$$

$$r_2 = \{y_2 + r_1\}$$

which has the easy solution $r_1 = \left\{\frac{y_1 - y_2}{2}\right\}, r_2 = \left\{\frac{y_1 + y_2}{2}\right\}$ for arbitrary $y_1, y_2 \in [0,1)$. Unfortunately, things are not as easy for orthonormal wavelet filters of higher order. When $N > 1$, orthonormality of the wavelet filters implies

$$h_{2N-2} g_0 + h_{2N-1} g_1 = 0$$

$$h_{2N-2} h_0 + h_{2N-1} h_1 = 0 \ ,$$

where $h_{2N-1} \neq 0$. Hence $h_0 g_1 = h_1 g_0$ or $\det H_0 = 0$.

This implies that

$$r = \{(Id - \alpha H_0) r + y\} \tag{2.8}$$

cannot be solved for arbitrary $y \in [0,1)^2$, regardless of the choice of α, as shown by the following argument. Suppose that for a given y in the interior of $[0,1)^2$, r solves (2.8). Then there exists $n \in \mathbb{Z}^2$ so that

$$r + n = (Id - \alpha H_0) r + y \ ,$$

or $\alpha H_0 r + n = y$. Since both r and y are in $[0,1)^2$, it follows that $\|n\|$ is bounded by some $C > 0$, uniformly in y. Since $\det H_0 = 0$, there exists a 2-vector e in $\mathbb{R}^2$ so that $\alpha H_0 \mathbb{R}^2 \subset \{\lambda e; \lambda \in \mathbb{R}\}$. Take now $f \perp e$, and consider $y' = y + \mu f$. Since y was chosen in the interior of $[0,1)^2$, y' still lies within $[0,1)^2$ for sufficiently small μ. Suppose r' is the corresponding solution of (2.8),

$$\alpha H_0 r' + n' = y' = y + \mu f \ .$$

Taking inner products of both sides with f leads to

$$\langle n', f\rangle = \langle y, f\rangle + \mu \|f\|^2 \ , \text{ or}$$
$$\mu = \langle n' - y, f\rangle / \|f\|^2 \ .$$

Since the right hand side can take on only a finite set of values (since $n' \in \mathbb{Z}^2, \|n'\| \leq C$), this can only hold for finitely many values of μ, and not for the interval around 0 for which $y' = y + \mu f$ is within $[0,1)^2$. (Note that the same no-go argument still works if we replace α by a diagonal matrix A, corresponding to a non-uniform expansion; in that case it suffices to choose $f \perp A H_0 \mathbb{R}^2$.)

There exists a way around this problem, which exploits that the y_n are not really independent of the a_n or r_n, as we have implicitly assumed here. This is the subject of the next subsection.

2.2 Uniform expansion factors

For the Haar filter, we were looking in the last subsection at the equation

$$\begin{pmatrix} y_1 \\ \\ y_2 \end{pmatrix} = \begin{pmatrix} n_1 \\ \\ n_2 \end{pmatrix} + \alpha \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} r_1 \\ \\ r_2 \end{pmatrix}$$

For $\alpha = 1$, $n_1, n_2 \in \mathbb{Z}$ and $r_1, r_2 \in [0,1)$, the possible values taken on by the right hand side range over the union of tilted squares shown in Figure 2a, which clearly do not cover the square $[0,1)^2$. For $\alpha = \sqrt{2}$, the tilted squares are blown up so that their union does not leave any gaps.

In the Haar case, the different pairs (s^0_{2k}, s^0_{2k-1}) remain nicely decoupled in the subband filtering process, which enabled us to reduce the whole analysis to a 2D argument. As we saw above, the most naive way to reduce to 2D does not work for longer filters; a solution to the problem at the end of the last section is possible only by taking into account to some extent the coupling between the pairs (s^0_{2k}, s^0_{2k-1}) in the filtering process. To do this, we shall first consider the full problem, for the complete data sequence, and try to carry out a reduction to fewer dimensions at a later stage. Imagine that the initial data constitute a finite sequence (of length L) of integers, $a \in \mathbb{Z}^L$. The subband filtering operator $\mathbf{H}$ can now

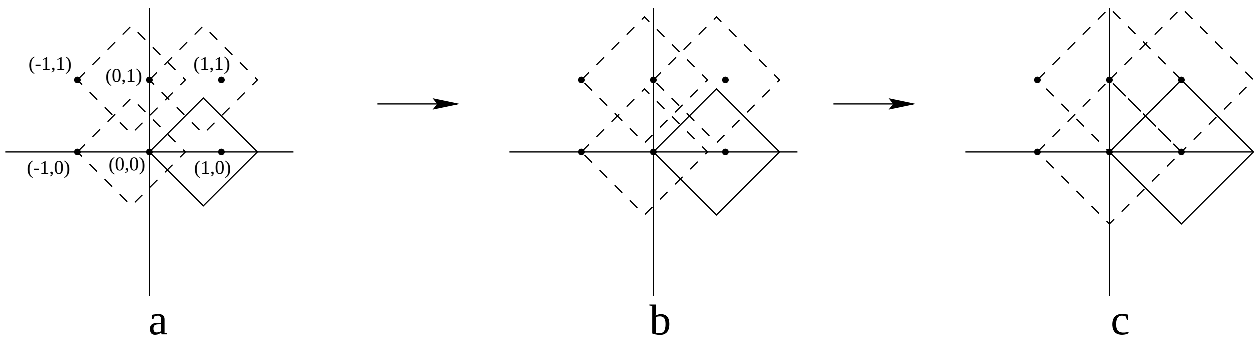

Figure 2:

a. For $H_0 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$, the union $\bigcup_{n\in\mathbb{Z}^2}(H_0[0,1)^2+n)$ does not cover $[0,1)^2$.

b. and c. As $\alpha \geq 1$ increases, the set $[0,1)^2 \setminus \bigcup_{n\in\mathbb{Z}^2}(\alpha H_0[0,1)^2+n)$ becomes smaller; the gap closes for $\alpha = \sqrt{2}$.

be viewed as an $L \times L$-matrix, which decomposes into 2×2 blocks (see §2.1).

$$\mathbf{H} = \begin{pmatrix} H_0 & H_1 & H_2 & \cdots & H_{N-3} & H_{N-2} & H_{N-1} & 0 & 0 & \cdots & 0 \\ 0 & H_0 & H_1 & \cdots & H_{N-4} & H_{N-3} & H_{N-2} & H_{N-1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ H_1 & H_2 & H_3 & \cdots & H_{N-2} & H_{N-1} & 0 & 0 & 0 & \cdots & H_0 \end{pmatrix}$$

(We have here given H a circulant-structure to deal with the finite length of the data sequence, which amounts to periodizing the data. Other ways of extending the data, such as reflections or interpolation techniques, can also be used. This would change only the first and last few rows and columns of the matrix, and it would not affect our argument significantly, although we would, of course, have to take those changes into account near the start and end of the signal. We have also implicitly assumed that L is even here, which is usually the case; odd L can be handled as well, but are slightly more tricky.)

In general, $\mathbf{H}a \notin \mathbb{Z}^L$ if $a \in \mathbb{Z}^L$. An appropriate shift of $\mathbf{H}a$ will bring it back to $\mathbb{Z}^L$. More precisely, if Λ is a fundamental region in $\mathbb{R}^L$ for the lattice $\mathbb{Z}^L$, then we can find $f \in \Lambda$ so that $\mathbf{H}a + f \in \mathbb{Z}^L$. In order for the transform $a \longrightarrow \mathbf{H}a + f$ to be invertible, we must require that if $a \neq a'$, then $\mathbf{H}a + f \neq \mathbf{H}a' + f'$, or since $\mathbf{H}$ itself is invertible, $\mathbf{H}^{-1}f \neq \mathbf{H}^{-1}f' + (a' - a)$. This will be satisfied if $\mathbf{H}^{-1}\Lambda \bigcap (\mathbf{H}^{-1}\Lambda + n) = 0$ for $n \in \mathbb{Z}^L$, $n \neq 0$. The Haar example shows us already that it may not be possible to find such a fundamental region Λ without introducing an expansion factor. If we do use an expansion factor α, then we are dealing with the transform $a \longrightarrow \alpha\mathbf{H}a + f$ instead, and the sufficient condition for invertibility becomes

$$\mathbf{H}^{-1}(\alpha^{-1}\Lambda) \bigcap \mathbf{H}^{-1}(\alpha^{-1}\Lambda + n) = \{0\} \text{ for all } n \in \mathbb{Z}^L \backslash \{0\} . \tag{2.9}$$

(In this section we consider uniform dilation factors only; non-uniform dilations, with different expansion factors for high and low-pass parts, will be considered in the next subsection.) Given any bounded fundamental region Λ, (2.9) will always be satisfied for sufficiently large α; the challenge is, for a given H, to find a "reasonable" Λ for which α is as small as possible. (Since the goal of this approach is to develop a practical lossless coding scheme with wavelets, the region Λ must be tractable for numerical computations. In the Haar case of last section, we had $\Lambda = [0, 1)^L$, and $\alpha = \sqrt{2}$.) It will be useful below, when we try to construct explicit Λ, to relax the condition that Λ be exactly a fundamental region and to require instead that Λ contain a fundamental region. Renaming $\alpha^{-1}\Lambda = \Gamma$, we have therefore the following challenge:

Find $\Gamma \subset \mathbb{R}^N$ such that

- $\Gamma' = \alpha\Gamma$ contains a fundamental region for $\mathbb{Z}^L$ with $\alpha \geq 1$ as small as we can get it)

$$(2.10)$$

- $\mathbf{H}^{-1}\Gamma \bigcap (\mathbf{H}^{-1}\Gamma + n) = \{0\}$ for all $n \in \mathbb{Z}^L, n \neq 0$.

$$(2.11)$$

In order to reduce the problem to a smaller size and make the search for a suitable Γ more tractable, we shall here restrict ourselves to special Γ of the form

$$\Gamma = \Omega \times \Omega \times \ldots \times \Omega \ ,$$

where $\Omega \subset \mathbb{R}^2$ such that $\Omega' = \alpha\Omega$ contains a fundamental region for $\mathbb{Z}^2$. This still enables us to exploit the 2×2-block structure of $\mathbf{H}$.

Let us look at this problem for the simplest example beyond the Haar case, which uses the 4-tap filters

$$h_0 = \tfrac{1+\sqrt{3}}{4\sqrt{2}} \ , \quad h_1 = \tfrac{3+\sqrt{3}}{4\sqrt{2}} \ , \quad h_2 = \tfrac{3-\sqrt{3}}{4\sqrt{2}} \ , \quad h_3 = \tfrac{1-\sqrt{3}}{4\sqrt{2}} \ ,$$

$$g_0 = -h_3 \ , \quad g_1 = h_2 \ , \qquad g_2 = -h_1 \ , \quad h_3 = h_0 \ .$$

Introducing the unitary matrices

$$U = \ \tfrac{1}{2\sqrt{2}} \begin{pmatrix} 1+\sqrt{3} & 1-\sqrt{3} \\ \sqrt{3}-1 & 1+\sqrt{3} \end{pmatrix}$$

$$V = \qquad \tfrac{1}{2} \begin{pmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix} \ ,$$

we can rewrite the matrices H_0 and H_1 (the only non-zero H_k) in this case as

$$H_0 = U \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} V \ , \ H_1 = U \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} V \ .$$

(This is in fact a different way of writing the factorization of [11] for this particular case.)

It follows that $\mathbf{H} = \mathbf{U}\widetilde{\mathbf{H}}\mathbf{V}$, where $\mathbf{U}$ and $\mathbf{V}$ are matrices with the 2×2 blocks U resp. V, on the diagonal, and zeros elsewhere, and where

$$\widetilde{\mathbf{H}} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & 1 & 0 & 0 & 0 & \vdots & \vdots \\ \vdots & \vdots & 0 & 0 & 0 & 1 & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & 1 & 0 \\ 0 & 1 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \end{pmatrix} .$$

Note that $\widetilde{\mathbf{H}}f$ (and therefore $\widetilde{\mathbf{H}}^{-1}f$ as well) is just a permutation of the even entries of f. It follows that if $\widetilde{\Gamma}$ is a product of 2D regions, $\widetilde{\Gamma} = \widetilde{\Omega} \times \ldots \times \widetilde{\Omega}$, with $\widetilde{\Omega} \subset [a_1, a_2) \times [b_1, b_2) = R$ then $\widetilde{\mathbf{H}}\widetilde{\Gamma} \subset R \times \ldots \times R$. In particular, any $\widetilde{\Gamma}$ that is already of the form $R \times \ldots \times R$ is invariant under $\widetilde{\mathbf{H}}$ or $\widetilde{\mathbf{H}}^{-1}$. This suggests the following strategy for finding a suitable Ω that would satisfy all our requirements. Take Ω such that $U^*\Omega$ is a rectangle with its sides parallel to the axes; then $\mathbf{U}^*\Gamma = U^*\Omega \times \ldots \times U^*\Omega$ will be left invariant by $\widetilde{\mathbf{H}}$, so that $\mathbf{H}^{-1}\Gamma = (V^*U^*\Omega \times \ldots \times V^*U^*\Omega)$, and our sufficient condition reduces to a set of 2D requirements. In particular, the last part (2.11) of the sufficient condition reduces to

$$V^*U^*\Omega \bigcap (V^*U^*\Omega + n) = \{0\} \text{ for } n \in \mathbb{Z}^2 \backslash \{0\} .$$

Since $U^*\Omega = [a_1, a_2) \times [b_1, b_2) = R$, this can be rewritten as

$$R \bigcap (R + Vn) = \{0\} \text{ for } n \in \mathbb{Z}^2 \backslash \{0\}$$

or

$$R - R \bigcap V\mathbb{Z}^2 = \{0\} ,$$

where we have used the notation

$$\begin{aligned} R - R &= \{x - y; x, y \in R\} \\ &= \{z = (u, v); |u| < a_2 - a_1, |v| < b_2 - b_1\} \ . \end{aligned}$$

Note that this set is symmetric around the origin, even if R isn't. We can therefore assume, without loss of generality, that R is centered around the origin, so that the condition reduces to

$$2R^\circ \bigcap V\mathbb{Z}^2 = \{0\} \text{ or } 2(V^*R)^\circ \bigcap \mathbb{Z}^2 = \{0\} \ ,$$

where B° denotes the interior of the set B.

In summary, we are thus looking for a rectangle $R_{a,b} = [-a, a) \times [-b, b)$ such that

- $\alpha U R_{a,b}$ contains a fundamental region for $\mathbb{Z}^2$
- $2(V^* R_{a,b})^\circ \bigcap \mathbb{Z}^2 = \{0\}$.

We would like to find a, b so that α is as small as possible. We start by constructing one candidate for $R_{a,b}$, and computing the corresponding α. The following technical lemma will be useful in our construction:

Lemma 2.1 *Start with a parallelogram P that is a fundamental region for $\mathbb{Z}^2$ in $\mathbb{R}^2$. Form a hexagon $\Delta \subset P$ by picking six points on the boundary of P so that opposite vertices are congruent modulo $\mathbb{Z}^2$ and so that no three of the six points are colinear. Then any parallelogram $\widetilde{P}$ that contains Δ also contains the interior of a fundamental region of $\mathbb{Z}^2$.*

Proof 1 See Appendix A.

(Note that as a fundamental region for $\mathbb{Z}^2$, P cannot be open nor closed. Likewise, $\widetilde{P}$ could fail to contain a true fundamental region if $\widetilde{P}$ contained too little of $\partial\widetilde{P}$.)

Corollary 2.2 *Any parallelogram symmetric around 0 that contains the points* $(\frac{1}{2}, 0), (0, \frac{1}{2})$ *and either* $(\frac{1}{2}, \frac{1}{2})$ *or* $(\frac{1}{2}, -\frac{1}{2})$ *contains the interior of a fundamental region for* $\mathbb{Z}^2$.

Let us apply this to our problem. We start with the second requirement, that $R^\circ_{a,b} \bigcap \frac{1}{2} V \mathbb{Z}^2 = \{0\}$. Since V is an orthogonal matrix, $\frac{1}{2} V \mathbb{Z}^2$ is a (tilted) square lattice. It therefore makes sense (because of symmetry considerations) to choose $R_{a,b}$ itself to be a square, i.e., $a = b$. The largest possible value for a is then $\frac{\sqrt{3}}{4}$. (It suffices to check the points $(1, 0), (0, 1)$ and $(1, \pm 1)$ in $\mathbb{Z}^2$.)

Next, we check the other requirement. We have

$$\begin{aligned}
&\alpha U R_{a,b} = \alpha U R_{\frac{\sqrt{3}}{4}, \frac{\sqrt{3}}{4}} \\
&= \{(x, y) \in \mathbb{R}^2; |(1 + \sqrt{3})x + (\sqrt{3} - 1)y| \le \tfrac{\sqrt{3}}{\sqrt{2}}\alpha, |(1 - \sqrt{3})x + (1 + \sqrt{3})y| \le \tfrac{\sqrt{3}}{\sqrt{2}}\alpha\} \ .
\end{aligned}$$

A straightforward application of Corollary 2.2 shows that this contains a fundamental region for $\mathbb{Z}^2$ if $\alpha \ge \sqrt{2}$.

Remark 1 There exist other solutions, in which $R_{a,b}$ is not a square. In fact, we could just as well have replaced our choice $b = \frac{\sqrt{3}}{4}$ with the smaller value $b = \frac{1+\sqrt{3}}{8}$, and obtained the same value for α. This change of b corresponds to trimming the $R_{a,b}$ slightly in a non-essential way, which reduces the overlap of the $\alpha U R_{a,b}$ in the direction where we had room to spare. One can also find completely different $R_{a,b}$, with very different aspect ratios, that avoid $\frac{1}{2} V \mathbb{Z}^2$. These lead to worse α, however.

This approach can also be used for longer filters, and even for biorthogonal pairs. For the general biorthogonal case, the $L \times L$ wavelet matrix $\mathbf{H}$ is of the type

$$\mathbf{H} = \begin{pmatrix} h_0 & h_1 & \cdots & \cdots & h_{2N-2} & h_{2N-1} & 0 & 0 & \cdots & \cdots \\ g_0 & g_1 & \cdots & \cdots & g_{2N-2} & g_{2N-1} & 0 & 0 & \cdots & \cdots \\ 0 & 0 & h_0 & h_1 & \cdots & \cdots & h_{2N-2} & h_{2N-1} & 0 & \cdots \\ 0 & 0 & g_0 & g_1 & \cdots & \cdots & g_{2N-2} & g_{2N-1} & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

with

$$\mathbf{H}^{-1} = \begin{pmatrix} \widetilde{h}_0 & \widetilde{g}_0 & 0 & 0 & \cdots & \cdots \\ \widetilde{h}_1 & \widetilde{g}_1 & 0 & 0 & \cdots & \cdots \\ \widetilde{h}_2 & \widetilde{g}_2 & \widetilde{h}_0 & \widetilde{g}_0 & 0 & \cdots \\ \widetilde{h}_3 & \widetilde{g}_3 & \widetilde{h}_1 & \widetilde{g}_1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

where $\sum_n h_n \widetilde{h}_{n+2k} = \delta_{k,0}$, $\widetilde{g}_n = (-1)^n h_{-n-1+2N}$, $g_n = (-1)^n \widetilde{h}_{-n-1+2N}$. We assume that h_0 and h_1 are not both zero (same for $\widetilde{h}_0, \widetilde{h}_1$), but we place no such restrictions on the final filter taps, so that our notation can accommodate biorthogonal filters of unequal length (which is always the case when the filter lengths are odd, as in the very popular 9-7 filter pair [4]).

Leaving the circulant nature of $\mathbf{H}^{-1}$ aside, $\mathbf{H}^{-1}$ consists of repeats of the two rows

$$\begin{array}{cccccccccccccc} \cdots & 0 & 0 & \widetilde{h}_{2N-2} & \widetilde{g}_{2N-2} & \widetilde{h}_{2N-4} & \widetilde{g}_{2N-4} & \cdots & \widetilde{h}_0 & \widetilde{g}_0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & \widetilde{h}_{2N-1} & \widetilde{g}_{2N-1} & \widetilde{h}_{2N-3} & \widetilde{g}_{2N-3} & \cdots & \widetilde{h}_1 & \widetilde{g}_1 & 0 & 0 & \cdots \end{array}$$

with offsets of 2 at each repeat.

Since $h_0\widetilde{h}_{2N-2} + h_1\widetilde{h}_{2N-1} = 0$ and $h_{2N-2}\widetilde{h}_0 + h_{2N-1}\widetilde{h}_1 = 0$, the two 2×2 matrices at the start and at the end of this repeating block are again singular. Define ratios λ and μ so that

$$\lambda h_1 = \widetilde{h}_{2N-2}, \lambda h_0 = -\widetilde{h}_{2N-1}, \text{ and } \mu\widetilde{h}_1 = h_{2N-2}, \mu\widetilde{h}_0 = -h_{2N-1} \ ,$$

and define the matrices

$$A = \begin{pmatrix} h_1 & \widetilde{h}_0 \\ -h_0 & \widetilde{h}_1 \end{pmatrix} \quad B = \begin{pmatrix} \lambda & 1 \\ 1 & -\mu \end{pmatrix} \ .$$

Then

$$\begin{pmatrix} \widetilde{h}_{2N-2} & \widetilde{g}_{2N-2} \\ \widetilde{h}_{2N-1} & \widetilde{g}_{2N-1} \end{pmatrix} = \begin{pmatrix} \lambda h_1 & h_1 \\ -\lambda h_0 & -h_0 \end{pmatrix} = A \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} B$$

$$\begin{pmatrix} \widetilde{h}_0 & \widetilde{g}_0 \\ \widetilde{h}_1 & \widetilde{g}_1 \end{pmatrix} = \begin{pmatrix} \widetilde{h}_0 & -\mu\widetilde{h}_0 \\ \widetilde{h}_1 & -\mu\widetilde{h}_1 \end{pmatrix} = A \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} B \ .$$

It follows that $\mathbf{H}^{-1}$ has a structure reminiscent of (2.11),

$$\mathbf{H}^{-1} = \mathbf{A} \begin{pmatrix} & \vdots & & \vdots & & \vdots & \\ \cdots & 0 \ 0 \ 1 \ 0 & & & & 0 \ 0 \ 0 \ 0 & \cdots \\ & & & C_1 \ldots C_{N-2} & & & \\ \cdots & 0 \ 0 \ 0 \ 0 & & & & 0 \ 1 \ 0 \ 0 & \cdots \\ & \vdots & & \vdots & & \vdots & \end{pmatrix} \mathbf{B} \ , \qquad (2.12)$$

where $\mathbf{A}$ and $\mathbf{B}$ have the 2×2 blocks A, respectively B, on the diagonal, and zero entries elsewhere. The 2×2 matrices C_k are defined by

$$C_k = A^{-1} \begin{pmatrix} \widetilde{h}_{2N-2-2k} & \widetilde{g}_{2N-2-2k} \\ \widetilde{h}_{2N-1-2k} & \widetilde{g}_{2N-1-2k} \end{pmatrix} B^{-1} \ .$$

Note that since $\mathbf{H}^{-1}$ is invertible, both $\mathbf{A}$ and $\mathbf{B}$, and therefore also A and B must be invertible. (In the orthogonal case, this can be seen immediately because $\det A = h_0^2 + h_1^2, \det B = -1 - \lambda^2$.) We can now follow the same strategy as before: we try to find $\Omega \subset I\!\!R^2$ so that

- $\alpha\Omega$ contains a fundamental region for $\mathbb{Z}^2$

- $\mathbf{H}^{-1}(\Omega \times \ldots \times \Omega) \bigcap [\mathbf{H}^{-1}(\Omega \times \ldots \times \Omega) + n] = 0$ for all $n \in \mathbb{Z}^L$, $n \neq 0$.

$$\tag{2.13}$$

We shall take the ansatz $\Omega = B^{-1} R_{a,b}$. On the other hand, we define $\tilde{\Omega} = A R_{c,d}$; for appropriate c, d, we then have

$$\mathbf{H}^{-1}(\Omega \times \ldots \times \Omega) \subset \tilde{\Omega} \times \ldots \times \tilde{\Omega} \; .$$

Indeed, it suffices that

$$R_{a,b} + C_1 R_{a,b} + \cdots + C_{N-2} R_{a,b} \subset R_{c,d} \; . \tag{2.14}$$

This, in turn, is assured if

$$\begin{pmatrix} 1 + \sum_{\ell=1}^{N-2} |C_{\ell;1,1}| & \sum_{\ell=1}^{N-2} |C_{\ell;1,2}| \\ \sum_{\ell=1}^{N-2} |C_{\ell;2,1}| & 1 + \sum_{\ell=1}^{N-2} |C_{\ell;2,2}| \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix} \; . \tag{2.15}$$

(Of course (2.14) may be satisfied for choices of a, b, c, d where (2.15) is not, but in practice, for the filter pairs we considered, not much is lost.) Finally, if $2AR^\circ_{c,d} \bigcap \mathbb{Z}^2 = \{0\}$ or $R^\circ_{c,d} \bigcap \frac{1}{2} A^{-1} \mathbb{Z}^2 = \{0\}$, then the second part of (2.13) is satisfied.

Let us see how this works on the example of the 6-tap orthogonal filter with three vanishing moments,

$$\begin{array}{lll} h_0 = .332671 & h_1 = .806891 & h_2 = .459878 \\ h_3 = -.135011 & h_4 = -.085441 & h_5 = .035226 \end{array} \; .$$

Because of orthogonality, λ equals μ with $\lambda = \mu = \frac{h_\mu}{h_1} \simeq -.1059$; moreover both A and B are orthogonal matrices (up to an overall constant each),

$$A = (h_0^2 + h_1^2)^{\frac{1}{2}} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \quad B = (1+\lambda^2)^{\frac{1}{2}} \begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix} ,$$

with $\theta = \arctan \frac{h_0}{h_1}$, $\varphi = \arctan \frac{1}{\lambda}$. We now want to determine c, d so that $R^\circ_{c,d} \bigcap \frac{1}{2} A^{-1} \mathbb{Z}^2 = \{0\}$. Since A consists of a rotation combined with an overall scaling, $\frac{1}{2} A^{-1} \mathbb{Z}^2$ is a square lattice, and it makes sense to choose $c = d$, as in the 4-tap case. The largest possible value for $c = d$ is then $\frac{h_1}{2(h_0^2+h_1^2)} \simeq .5296$.

On the other hand, because the filters are so short, we have only one matrix C_1,

$$C_1 = \frac{1}{(h_0^2+h_1^2)(1+\lambda^2)} \begin{pmatrix} h_1 & -h_0 \\ h_0 & h_1 \end{pmatrix} \begin{pmatrix} h_2 & h_3 \\ h_3 & -h_2 \end{pmatrix} \begin{pmatrix} \lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \begin{pmatrix} 0 & .5461 \\ -.5461 & 0 \end{pmatrix}$$

According to (2.15), we then have

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & .5461 \\ .5461 & 1 \end{pmatrix}^{-1} \begin{pmatrix} c \\ d \end{pmatrix} \simeq \begin{pmatrix} .3425 \\ .3425 \end{pmatrix}$$

Finally, it remains to determine α so that $\alpha\Omega = \alpha B^{-1} R_{a,b}$ contains a fundamental region for $\mathbb{Z}^2$. $\alpha B^{-1} R_{a,b}$ is the tilted square delimited by $\lambda x + y = \pm a\alpha(\lambda^2+1), x - \lambda y = \pm b\alpha(\lambda^2+1)$; by Corollary 2.2 this contains a fundamental region if $(\frac{1}{2}, 0), (0, \frac{1}{2})$ and $(-\frac{1}{2}, \frac{1}{2})$ are included, leading to $d \geq \frac{|\lambda|+1}{2a(1+\lambda^2)} \simeq 1.5965$. This expansion factor is significantly worse than the factor $\sqrt{2}$ obtained for the 4-tap orthornormal filter earlier. In fact, applying this same strategy to the family of all 6-tap orthonormal filters with at least one vanishing moment, leads to $\alpha \geq \sqrt{2}$ in all cases, with the minimum $\alpha = \sqrt{2}$ attained only for the 4-tap filter. In the next subsection, we shall see that better results for orthonormal filters can be obtained via non-uniform expansion factors.

The same strategy can also be used for biorthogonal filters. For the simple filter pair

$$h_0 = -\tfrac{1}{8} = h_4, h_1 = \tfrac{1}{4} = h_3, h_2 = \tfrac{3}{4}, \text{ other } h_n = 0$$

$$\widetilde{h}_1 = \widetilde{h}_3 = \tfrac{1}{2}, \widetilde{h}_2 = 1, \text{ other } \widetilde{h}_n = 0 \ ,$$

the matrices A, B, and C are

$$A = \begin{pmatrix} \frac{1}{4} & 0 \\ \frac{1}{8} & \frac{1}{2} \end{pmatrix} \quad B = \begin{pmatrix} 0 & 1 \\ 1 & \frac{1}{4} \end{pmatrix} \quad C = \begin{pmatrix} 0 & 4 \\ -\frac{7}{4} & 0 \end{pmatrix} \ ,$$

leading to the choices $C = 2, d = 1$, and $a = \frac{1}{3}, b = \frac{5}{12}$. Then $\alpha B^{-1} R_{a,b}$ contains $(\frac{1}{2}, 0), (0, \frac{1}{2})$ and $(-\frac{1}{2}, \frac{1}{2})$ (and therefore a fundamental region for $\mathbb{Z}^2$) if $\alpha \geq \frac{3}{2}$. In this case it is not clear, however, what a uniform expansion factor really means, since the two filters h and $\widetilde{h}$ need not have the same normalization in a biorthogonal scheme—as indeed they don't in this example, where $\sum_n h_n = 1, \sum_n \widetilde{h}_n = 2$. One could explore "renormalized" versions, replacing h_n with δh_n and $\widetilde{h}_n$ with $\delta^{-1}\widetilde{h}_n$, and find the corresponding expansion factors. This amounts to the same as introducing nonuniform expansion factors, the subject of the next subsection.

2.3 Different expansion factors for the high- and low-pass channels

Nothing forces us, when we introduce our expansion factor, to choose the *same* expansion for the high pass channel as for the low pass channel; allowing for different factors may lead to better results. In practice, the dynamic ranges of the two channels on (natural) images are very different, and this might be another reason to consider different expansion factors.

The analysis is very similar to what was done in the previous subsection. We replace the uniform factor α by a pair (α_L, α_H), and we denote by $\boldsymbol{\alpha}$ the operation which multiplies the odd-indexed entries a_{2k+1} of $\boldsymbol{\alpha} \in \mathbb{Z}^L$ with α_L, the even-indexed entries a_{2k} by α_H. We again seek a subset $\Gamma \subset I\!\!R^L$ that satisfies (2.11) and such that $\boldsymbol{\alpha}\Gamma$ contains a fundamental region for $Z\!\!\!Z^L$. We can now seek to minimize α_L (which leads to the least increase in dynamical range as we keep iterating the wavelet filtering) or the product $\alpha_L \alpha_H$ (to minimize the impact on the high-low, low-high channels that detect horizontal and vertical edges in $2D$-images).

As a warm-up, let's look at the Haar case again. We can confine our analysis to $I\!\!R^2$ and $Z\!\!\!Z^2$. A pair of expansion factors (α_L, α_H) will work if we can find a set Ω in $I\!\!R^2$ so that

- Ω is a fundamental region for $Z\!\!\!Z^2$
- $H_0^{-1}\alpha^{-1}\Omega \bigcap (H_0^{-1}\alpha^{-1}\Omega + n) = 0$ for $n \in Z\!\!\!Z^2, n \neq 0$,

where α denotes now the 2×2 matrix, $\alpha = \begin{pmatrix} \alpha_L & 0 \\ 0 & \alpha_H \end{pmatrix}$. Let's take $\Omega = [-\frac{1}{2}, \frac{1}{2})^2$, so that Ω° is symmetric with respect to the origin, consistent with our analysis of longer filters in §2.2; this clearly satisfies the first requirement. Then $\alpha^{-1}\Omega$ is the rectangle $[-\frac{\alpha_L}{2}, \frac{\alpha_L}{2}) \times [-\frac{\alpha_H}{2}, \frac{\alpha_H}{2})$. The matrix H_0^{-1} rotates this by 45°, to a tilted rectangle bounded by $y = x \pm \frac{1}{\sqrt{2}}\alpha_H^{-1}, y = -x \pm \frac{1}{\sqrt{2}}\alpha_L^{-1}$ (see Figure 3). In order to satisfy the second requirement, this tilted rectangle can have area at most 1, or $\alpha_H \alpha_L \geq 1$. Can we choose α_L, α_H so as to achieve the extremum $\alpha_H \alpha_L = 1$? In that case $H_0^{-1}\alpha^{-1}\Omega$ must be a fundamental region itself. For this, it is sufficient that $(0, \frac{1}{2}), (\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2}) \in H_0^{-1}\alpha^{-1}\Omega$, or $\frac{1}{2} \leq \frac{1}{\sqrt{2}}\alpha_H^{-1}, 1 \leq \frac{1}{\sqrt{2}}\alpha_L^{-1}$. This leads to the choices $\alpha_H = \sqrt{2}, \alpha_L = \frac{1}{\sqrt{2}}$, which do indeed satisfy $\alpha_L \alpha_H = 1$. For the Haar case, we can therefore find a non-uniform expansion matrix corresponding with a global expansion factor of 1 (i.e., no net expansion factor), a significant gain over the uniform case.

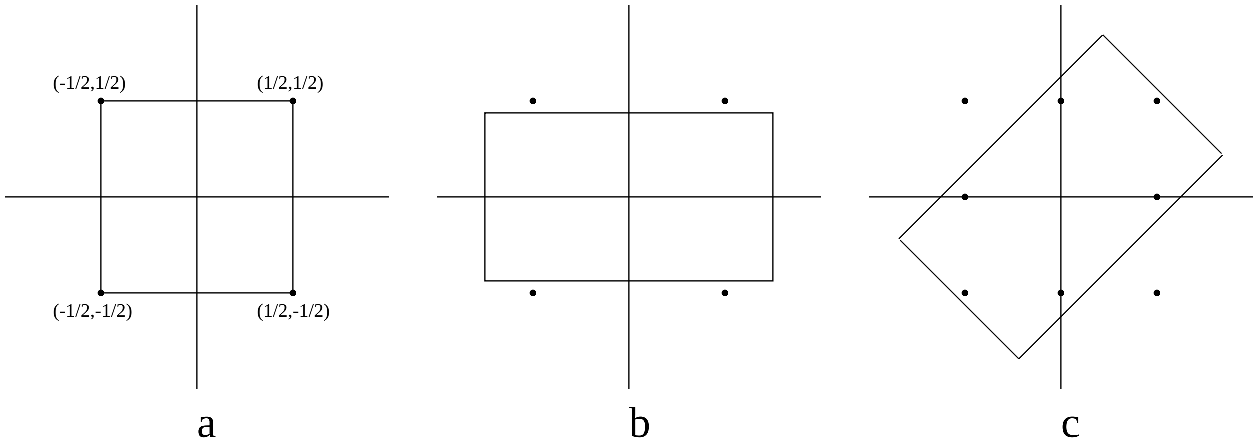

Figure 3:

a. The square $\Omega = \left[-\frac{1}{2}, \frac{1}{2}\right)^2$

b. Applying α^{-1} leads to $\left[-\frac{\alpha_L^{-1}}{2}, \frac{\alpha_L^{-1}}{2}\right) \times \left[-\frac{\alpha_H^{-1}}{2}, \frac{\alpha_H^{-1}}{2}\right)$

c. Applying next H_0^{-1} leads to the rectangle given by $|y - x| \leq \frac{\alpha_H^{-1}}{\sqrt{2}}, |y + x| \leq \frac{\alpha_L^{-1}}{\sqrt{2}}$. This contains $(\frac{1}{2}, 0), (0, \frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{2})$ if $\alpha_H \geq \sqrt{2}, \alpha_L \geq \frac{1}{\sqrt{2}}$.

Note that the total transform αH_0 is then given by

$$\begin{pmatrix} x \\ \\ y \end{pmatrix} \to \begin{pmatrix} \frac{x+y}{2} \\ \\ x - y \end{pmatrix} ,$$

a (non-orthonormal) form of the Haar transform often used in practice, and one for which it has long been known that it could be made into a map of integers to integers, called the S-transform [7]. We shall come back to this later.

After this warm-up, let's consider longer filters. We can re-use much of what was done in the previous section. The factorization (2.12) becomes now

$$\mathbf{H}^{-1}\boldsymbol{\alpha}^{-1} = \mathbf{A} \begin{pmatrix} & \vdots & & \vdots & & \vdots & \\ \cdots & 0\ \ 0\ \ 1\ \ 0 & & & & 0\ \ 0\ \ 0\ \ 0 & \cdots \\ & & & C_1 \ldots C_{N-2} & & & \\ \cdots & 0\ \ 0\ \ 0\ \ 0 & & & & 0\ \ 1\ \ 0\ \ 0 & \cdots \\ & \vdots & & \vdots & & \vdots & \end{pmatrix} \mathbf{B}\boldsymbol{\alpha}^{-1} ,$$

which suggests that we seek $\Omega = \alpha B^{-1} R_{a,b}$ such that $(\boldsymbol{\alpha}\mathbf{H})^{-1}(\Omega \times \ldots \times \Omega) \subset (AR_{c,d} \times \ldots \times AR_{c,d})$ where $R^\circ_{c,d} \bigcap \frac{1}{2} A^{-1} \mathbb{Z}^2 = \{0\}$. Let us check what this leads to in a few examples. We first take the 4-tap orthonormal filter. In that case, there are no C-matrices to worry about, and we have $c = a, d = b$. The matrix A ($= Y^*$ in the notations of the first part of §2.2) is still the same rotation, so we still choose $R_{c,d}$ to be the square given by $c = d = \frac{\sqrt{3}}{4}$. Because $\alpha_L \neq \alpha_H$, the set $\alpha B^{-1} R_{a,b}$ will no longer be a square; it is the parallellogram

$$\begin{aligned} \Omega = \{(x,y); \left|(1+\sqrt{3})\tfrac{x}{\alpha_L} + (1-\sqrt{3})\tfrac{y}{\alpha_H}\right| &\leq \tfrac{\sqrt{6}}{2} \\ \left|(\sqrt{3}-1)\tfrac{x}{\alpha_L} + (\sqrt{3}+1)\tfrac{y}{\alpha_H}\right| &\leq \tfrac{\sqrt{6}}{2}\} \ . \end{aligned}$$

If we assume $\alpha_L \leq \alpha_H$, then the condition $\alpha_L \geq \frac{1+\sqrt{3}}{\sqrt{6}}$ already ensures that $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2}) \in \Omega$. If we pick the minimal $\alpha_L = \frac{1+\sqrt{3}}{\sqrt{6}} \simeq 1.1$, does there exist a solution for α_H and how large does it need to be? With the notation $\alpha_H = \alpha_L \cdot \delta$, with $\delta \geq 1$, the parallellogram Ω contains the point $(\frac{1}{2}, \frac{1}{2})$ (and therefore also a fundamental region for $\mathbb{Z}^2$) if $\left|1 + \frac{1-\sqrt{3}}{1+\sqrt{3}} \frac{1}{\delta}\right| \leq 1$, or $\delta \geq \frac{\sqrt{3}+1}{2}$, leading to the solution $\alpha_H = \frac{2+\sqrt{3}}{\sqrt{6}} \simeq 1.6$. The product $\alpha_L \alpha_H$ is then $\frac{5+3\sqrt{3}}{6} \simeq 1.7$, an improvement over $\alpha^2 = 2$ in the uniform expansion case, although it is not as spectacular as in the Haar case.

Our next example is the 6-tap orthonormal filter also considered in §2.2. We keep again the same $R_{c,d}$ as before, with $c = d \simeq .5296$. The square $R_{a,b}$ is also not changed, with

$a = b \simeq .3425$. It then remains to determine α_L, α_H (minimizing either α_L or $\alpha_L\alpha_H$) so that $\alpha B^{-1}R_{a,b}$ contains a fundamental region. We have

$$\begin{aligned}\alpha B^{-1}R_{a,b} = \{(x,y); \left|\lambda\tfrac{x}{\alpha_L} + \tfrac{y}{\alpha_H}\right| &\leq a(\lambda^2+1)\ ,\\ \left|\tfrac{x}{\alpha_L} - \lambda\tfrac{y}{\alpha_H}\right| &\leq b(\lambda^2+1)\}\ .\end{aligned}$$

This parallellogram contains $(\frac{1}{2},0),(0,\frac{1}{2})$ and $(-\frac{1}{2},\frac{1}{2})$, and therefore a fundamental region for $\mathbb{Z}^2$, if $\alpha_L = \frac{1}{2a(\lambda^2+1)} \simeq 1.4437, \alpha_H = \frac{1}{2a(1-|\lambda|)(1+\lambda^2)} \simeq 1.6147$. Again the product $\alpha_L\alpha_H \simeq 2.3310$ is smaller than the corresponding $(1.5965)^2 \simeq 2.5490$ in the uniform expansion case, but the gain is not as dramatic as in the Haar case; in particular, we still have $\alpha_L\alpha_H > 1$. Moreover, even α_L is still larger than $\sqrt{2}$.

Finally, let's revisit the biorthogonal example at the end of §2.2. Again, the values of c, d, and a, b can be carried over, and we consider

$$\alpha B^{-1}R_{a,b} = \left\{(x,y); \left|\frac{y}{\alpha_H}\right| \leq \frac{1}{3}, \left|\frac{x}{\alpha_L} + \frac{y}{4\alpha_H}\right| \leq \frac{5}{12}\right\}\ .$$

This contains $(\frac{1}{2},0),(0,\frac{1}{2})$ and $(-\frac{1}{2},\frac{1}{2})$ if $\alpha_L = \frac{6}{5}, \alpha_H = \frac{3}{2}$, a gain over uniform expansion, but still a rather large price to pay, when compared to the Haar case (where we obtained $\alpha_L < 1!$).

In the next section, a different approach is explained, which goes beyond the reduction to $2D$ used so far, and which enables us to reduce *any* wavelet filtering to a map from integers to integers without global expansion. In fact, this amounts to constructing appropriate sets Γ without using products of $2D$ regions by using a different representation of the filtering operations.

A Appendix

Proof of Lemma 3.1.

Lemma A.1 *Let P be a parallelogram that is a fundamental region for $\mathbb{Z}^2$, and let A_1, A_2, A_3 be three points on the boundary of P such that*

- *A'_j, the mirror image of A_j with respect to the center of P, is congruent with A_j modulo $\mathbb{Z}^2$ ($j = 1, 2,$ or 3)*
- *no three points among $A_1, A_2, A_3, A'_1, A'_2, A'_3$ are colinear.*

Let Ω be the hexagon with vertices $A_1, A_2, A_3, A'_1, A'_2, A'_3$. Then any closed parallelogram Q that contains Ω also contains a fundamental region for $\mathbb{Z}^2$.

(Note that Ω is a strict subset of a fundamental region, and is therefore *not* a fundamental region itself.)

Proof 2 Without loss of generality, we can assume that P is centered around the origin.

Because of the prohibition of colinearity, the six points must be distributed among the four sides of P, with two points each on two opposite sides, and one point each on the remaining two sides. If none of the points A_j, A'_j lies on a vertex of P, then we can assume (by renumbering if necessary) that $A_1, A_2,$ and A_3 are three consecutive corners of the hexagon Ω, lying on three different sides of P; see Figure 6a. (If one of the A_j, A'_j lies on a vertex, as in Figure 6b, then the rest of this proof needs to be changed in a straightforward way; we leave this as an exercise for the reader.)

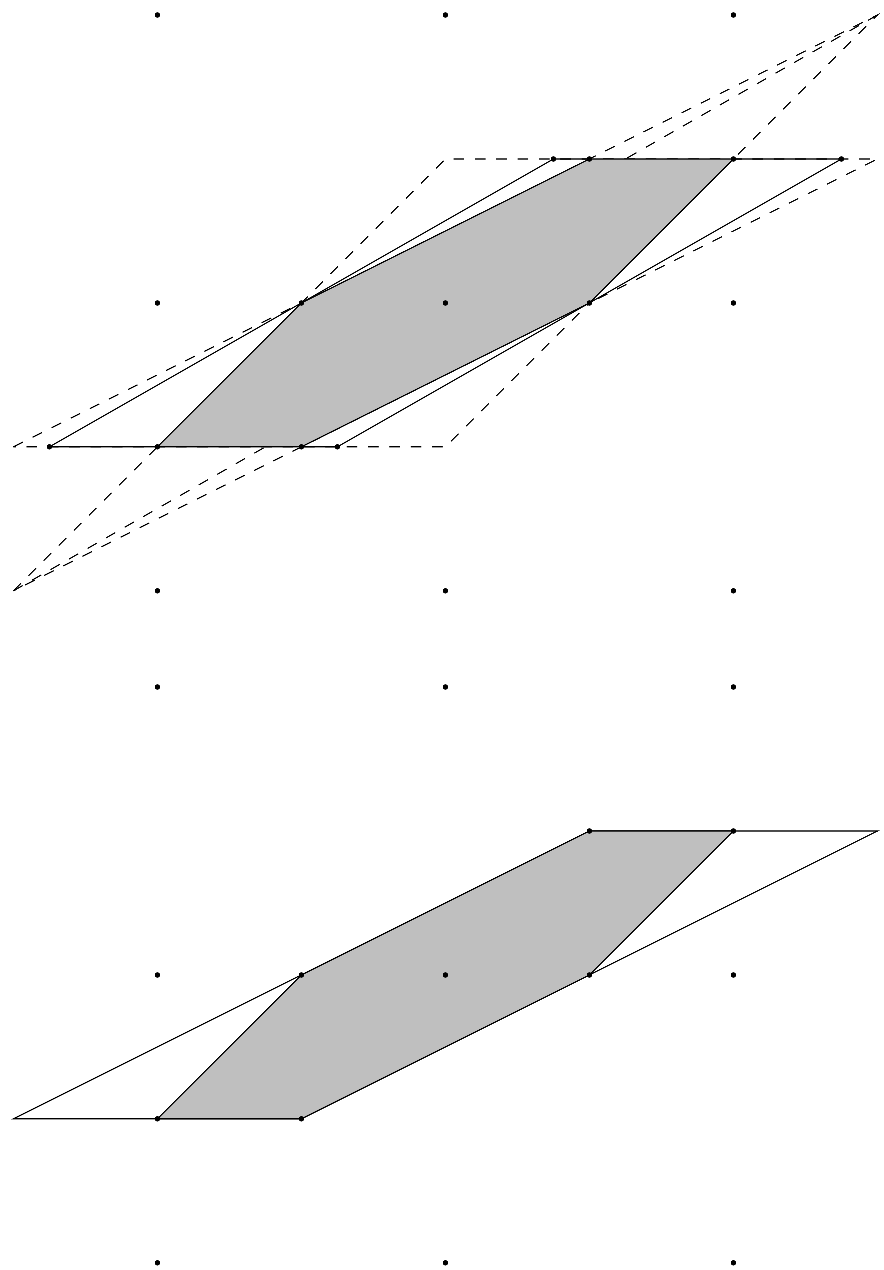

Figure 4:

a. The parallelogram P (with vertices C_{12}, C_{23}, C'_{12}, and C'_{23}) and the inscribed hexagon Ω, with associated triangles $\triangle_{12}, \triangle_{23}, \triangle'_{12}$ and $\triangle'_{23}$. All the $\tilde{}$ and $\hat{}$ versions of the triangles are drawn, but only $\widetilde{\triangle}_{12}$ and $\widehat{\triangle}_{12}$ have been labelled.

b. A special case where some vertices of Ω coincide with vertices of P.

Denote by C_{12} the vertex of P that separates A_1 from A_2, and similarly by C_{23} the vertex that separates A_2 from A_3. We denote their opposites, which separate A'_1 from A'_2, A'_2 from A'_3, respectively, by C'_{12} and C'_{23}.

The original fundamental region P is the union of the hexagon Ω and four triangles $\triangle_{12}$ (with vertices A_1, A_2, C_{12}), $\triangle_{23}$ (vertices A_2, A_3, C_{23}), $\triangle'_{12}$ (vertices A'_1, A'_2, C'_{12}), and $\triangle'_{13}$ (vertices A'_1, A'_3, C'_{13}). For each of the triangles $\triangle_{ij}$ we define the two triangles $\widehat{\triangle}_{ij}$ and $\widetilde{\triangle}_{ij}$ by $\widehat{\triangle}_{ij} = \triangle_{ij} - A_i + A'_i$, $\widetilde{\triangle}_{ij} = \triangle_{ij} - A_j + A'_j$; $\widehat{\triangle}'_{ij}$ and $\widetilde{\triangle}'_{ij}$ are defined analogously. Every one of the four original triangles is congruent $\bmod \mathbb{Z}^2$ with its $\hat{}$ and $\tilde{}$ versions.

Suppose there exists a parallelogram Q that contains Ω and that does not contain a fundamental region for $\mathbb{Z}^2$. If the parallel lines that bound Q do not touch Ω, we may reduce their separation until they do. Let R be the resulting reduced parallelogram, bounded by the parallel lines L, L' and M, M', each of which is incident with Ω. Let x be a point in P so that R does not contain any point congruent to $x \bmod \mathbb{Z}^2$. Then $x \in \triangle_{ij}$ or $\triangle'_{ij}$ for some i, j, and there exist points $\widehat{x}, \widetilde{x}$ in the corresponding $\widehat{\triangle}, \widetilde{\triangle}$ that are congruent to $x \bmod \mathbb{Z}^2$. Observe that if L, L' are parallel lines through opposite vertices j, j' then exactly two of the points $x, \widetilde{x}, \widehat{x}$ lie between L and L'. Similarly for M, M'. Hence one of the points $x, \widetilde{x}, \widehat{x}$ lies in the parallelogram R bounded by L, L' and M, M', which is a contradiction.

References

[1] FBI

[2] Special Issue

[3] I. Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF Regional Conf. Series in Appl. Math., Vol. 61. SIAM Journ. of Math. Anal., Philadelphia, PA, 1992.

[4] A. Cohen, I. Daubechies, and J. Feauveau. Bi-orthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.

[5] M. Vetterli and C. Herley. Wavelets and filter banks: Theory and design. *IEEE Trans. Acoust. Speech Signal Process.*, 40(9): 2207–2232, 1992.

[6] Laroia

[7] A. Zandi, M. Boliek, E.L. Schwartz, and M.J. Gormish. CREW Lossless/lossy medical image compression. Technical Report, RICOH California Research Center, CRC-TR-9526, 1995.

[8] Which one of these two do you want? W. Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Journal of Appl. and Comput. Harmonic Analysis*, 3(2):186–200, 1996. W. Sweldens. The lifting scheme: A construction of second generation wavelets. Technical Report 1995:6, Department of Mathematics, University of South Carolina, 1995.

[9] Is this one correct? P.P. Vaidyanathan and P.-Q. Hoang. Lattice structures for optimal design and robust implementation of two-band perfect reconstruction QMF banks. *IEEE Trans. Acoust. Speech Signal Process.*, 36:626–630, Jan. 1988.

[10] M. Vetterli and J. Kovačević. *Wavelets and Subband Coding.* Prentice Hall, Englewood Cliffs, NJ, 1995.

[11] Which one of these three do you want? P.P. Vaidyanathan. Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having perfect reconstruction property. *IEEE Trans. Acoust. Speech Signal Process.*, 35(2):476–492, 1987. P.P. Vaidyanathan, T.Q. Nguyen, Z. Doğanata, and T. Saramäki. Improved Technique for Design of Perfect Reconstruction FIR QMF Banks with Lossless Polyphase Matrices. *IEEE Trans. Acoust. Speech Signal Process.*, 37(7):1042–1055, 1989. T.Q. Nguyen and P.P. Vaidyanathan. Two-channel perfect-reconstruction FIR QMF structures which yield linear-phase analysis and synthesis filters. *IEEE Trans. Acoust. Speech Signal Process.*, 37:676–690, 1989.