

# Orthogonal Bandlet Bases for Geometric Images Approximation

Stéphane Mallat and Gabriel Peyré

December 4, 2006

## Abstract

This paper introduces a construction of bandlet orthogonal bases to approximate images having sharp transitions along regular geometric curves. Bandlet coefficients are calculated by applying orthogonal Alpert transforms over orthogonal wavelet coefficients. These Alpert transforms are calculated along a geometrical flow that is optimized for each image. It is proved that images that are  $C^\alpha$  with singularities along  $C^\alpha$  curves are approximated in a best bandlet basis with an optimal asymptotic error decay. A fast discrete orthogonal bandlet transform is described for images, with an application to compression.

## I Introduction

An orthogonal wavelet transform is well adapted to approximate functions whose local regularity is mostly isotropic but can not take advantage of the anisotropic regularity produced by geometric structures in images. Section II describes a model of geometrically regular images with functions that are piecewise  $C^\alpha$  with edge discontinuities along piecewise  $C^\alpha$  curves. Such functions are said to be  $C^\alpha$  geometrically regular. Because of the discontinuities, in a wavelet basis the best  $M$ -terms approximation error of such functions decays asymptotically like  $M^{-1}$ . Several constructions have been proposed to improve this asymptotic decay by taking advantage of geometric regularity, among which the curvelet frame of Candès and Donoho [3] and the bandlet frames of Le Pennec and Mallat [10].

This paper introduces new bandlet orthogonal bases constructed with parameterized orthogonal operators applied to orthogonal wavelet bases. The resulting bandlets are both orthogonal and regular. The approximation of a geometrically regular function is performed in a “best” bandlet bases.

The construction of bandlet orthogonal bases over wavelet bases is motivated by the regularity of wavelet coefficients for geometrically regular functions. Section IV proves that wavelet coefficients of  $C^\alpha$  geometrically regular functions are samples of a function which has an anisotropic regularity along a geometric flow. An approximation of such coefficients is obtained with polynomials over bands that are parallel to this geometrical flow. Section IV proves that if  $f$  is  $C^\alpha$  geometrically

regular then its wavelet coefficients can be approximated with  $M$  parameters while reconstructing an approximation whose error decays like  $M^{-\alpha}$ , up to a logarithmic term.

To build adaptive polynomial approximations of wavelet coefficients, Section V introduces polynomial Alpert transforms over wavelet coefficients. Orthogonal bandlets are functions resulting from a transformation of orthogonal wavelets with an orthogonal Alpert transform along a geometric flow. A fast “best basis” search algorithm is introduced. It finds a “best” geometric flow corresponding to a “best” bandlet basis having nearly optimal properties to approximate a given image. The main theorem proves that if  $f$  is  $C^\alpha$  geometrically regular then an approximation  $f_M$  in a best orthogonal bandlet basis satisfies

$$\|f - f_M\|_2^2 = O(M^{-\alpha}),$$

where  $M$  is the total number of parameters specifying  $f_M$ . This approximation scheme is adaptive with respect to the regularity exponent  $\alpha$  which is a priori unknown.

For discrete images with  $N$  pixels, this bandlet construction also defines orthogonal bases of  $\mathbb{R}^N$ . Section VII describes a fast algorithm that decomposes an image in a best bandlet basis with  $O(N)$  operations.

For compression applications, images are decomposed in a best bandlet basis and the resulting coefficients are quantized and entropy coded. For images that are discretizations of  $C^\alpha$  geometrically regular functions  $f$ , Section VIII proves that the error introduced by this compression scheme decays like  $\log^\alpha(R) R^{-\alpha}$ , where  $R$  is the number of bits of the compressed code. The resulting distortion-rate curve approaches the Kolmogorov asymptotic lower bound up to a logarithmic factor.

## II Wavelet Approximation of Geometrically Regularity Images

Donoho [7] introduced a geometric image cartoon model where  $f(x)$  for  $x \in [0, 1]^2$  is  $C^\alpha$  over regions whose boundaries are piecewise  $C^\alpha$  curves. To incorporate the diffraction blur produced by the optics of a camera this model is refined in [10] with a convolution by an unknown regular kernel. An example is shown in figure 2, left. The following definition formalizes this model.

**Definition 1.** *A function  $f \in L^2([0, 1]^2)$  is said to be  $C^\alpha$ -geometrically regular with a scale  $s > 0$  if  $f = \tilde{f} * h$  where  $\tilde{f} \in C^\alpha(\Lambda)$  for  $\Lambda = [0, 1]^2 - \{\gamma_i\}_{1 \leq i \leq G}$ . The blurring kernel  $h$  is  $C^\alpha$ , supported in  $[-s, s]^2$  with  $\|h\|_{C^\alpha} \leq s^{-(2+\alpha)}$ . The edge curves  $\gamma_i$  are  $C^\alpha$  and do not intersect tangentially. For  $s = 0$ , the same definition is valid for  $f = \tilde{f}$ .*

The image  $\tilde{f}$  is typically discontinuous along the edge curves  $\gamma_i$ , that may correspond to boundary of objects in the observed scene. The convolution with the blurring kernel  $h$  accounts for the diffraction phenomenon. The scale parameter  $s$  of  $h$  may be arbitrarily small. For an open set  $\Omega$  in  $\mathbb{R}^2$ ,  $C^\alpha(\Omega)$  is the space of  $\alpha$ -holderian functions, and  $\|f\|_{C^\alpha(\Omega)}$  is the usual norm on this space.

**Wavelet approximation.** Let us review why wavelet bases are suboptimal to approximate geometrically regular functions. An isotropic wavelet orthogonal basis of  $L^2([0, 1]^2)$  is obtained by translating and dilating three mother wavelets  $\{\psi^H, \psi^V, \psi^D\}$  (for the horizontal, vertical and diagonal directions) [11]. We consider compactly supported wavelets having a support in  $[-K, K]^2$ . Let  $p$  be the number of vanishing moments of these wavelets. Inner products in this basis of  $L^2([0, 1]^2)$  are written

$$f_j^k[n] \stackrel{\text{def.}}{=} \langle f, \psi_{jn}^k \rangle \quad \text{with} \quad \begin{cases} k \in \{H, V, D\}, & j < 0, \\ n = (n_1, n_2) \in \{0, \dots, 2^{-j} - 1\}^2, \\ \psi_{jn}^k(x) = 2^{-j} \psi^k(2^{-j}x_1 - n_1, 2^{-j}x_2 - n_2), \end{cases} \quad (1)$$

with appropriate modifications of  $\psi_{jn}^k$  to maintain their support in  $[0, 1]^2$ , as explained by Cohen et al. in [5]. In the following we shall drop the orientation index  $k \in \{H, V, D\}$  to simplify notations.

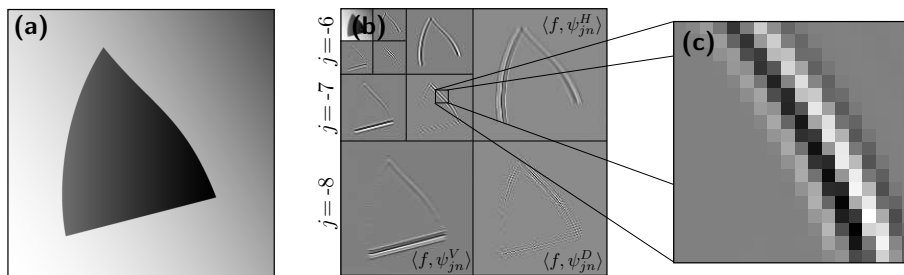


Figure 1: (a) Example of  $C^\alpha$  regular function  $f$  for  $\alpha = 3$ . (b) Orthogonal wavelet coefficients at different scales  $2^j$ . (c) Zoom over wavelet coefficients located in a square including a singularity curve.

Let  $M$  be the number of wavelet coefficients above a threshold  $T$ . The  $M$ -term wavelet approximation of  $f$  is

$$f_M = \sum_{|f_j[n]| \geq T} f_j[n] \psi_{jn}, \quad (2)$$

and

$$\|f - f_M\|_2^2 = \sum_{|f_j[n]| < T} |f_j[n]|^2.$$

If  $f \in C^\alpha([0, 1]^2)$  with  $\alpha < p$  then wavelet coefficients are small at fine scales and one can prove [13] that  $|f_j[n]| = |\langle f, \psi_{jn} \rangle| = O(2^{j(\alpha+1)})$ . Since there are  $2^{-2j}$  wavelet coefficients  $f_j[n]$  at each scale  $2^j$ , a direct calculation proves that

$$\sum_{|f_j[n]| > T} |f_j[n]|^2 = \|f - f_M\|_2^2 = O(M^{-\alpha}). \quad (3)$$

An image that is  $C^\alpha$  geometrically regular is typically not uniformly  $C^\alpha$  over  $[0, 1]^2$  because of the discontinuities across edges. To simplify explanations, we shall suppose that  $s = 0$ , which means that  $f$  is not regularized across edges. In the following, we briefly explain why the resulting  $M$ -term wavelet approximation is suboptimal with  $\|f - f_M\|_2^2 = O(M^{-1})$ .

The wavelet coefficients  $f_j[n] = \langle f, \psi_{jn} \rangle$  can be divided in 3 categories.

- Regular coefficients  $(n, j) \in \mathcal{R}$  correspond to wavelets  $\psi_{jn}$  whose support does not intersect an edge.
- Edge coefficients  $(n, j) \in \mathcal{E}$  correspond to wavelets  $\psi_{jn}$  whose support intersects only one edge curve.
- Corner coefficients  $(n, j) \in \mathcal{C}$  correspond to wavelets  $\psi_{jn}$  whose support intersects more than one edge curve.

Keeping  $M$  wavelet coefficients above a threshold  $T$  yields an approximation  $f_M$  with an error that can be decomposed in three parts:

$$\|f - \tilde{f}\|_2^2 = \sum_{\substack{(j,n) \in \mathcal{R} \\ |f_j[n]| < T}} |f_j[n]|^2 + \sum_{\substack{(j,n) \in \mathcal{C} \\ |f_j[n]| < T}} |f_j[n]|^2 + \sum_{\substack{(j,n) \in \mathcal{E} \\ |f_j[n]| < T}} |f_j[n]|^2. \quad (4)$$

If  $f$  is a  $C^\alpha$  geometrically regular then  $f$  is uniformly  $C^\alpha$  over the support of all wavelets corresponding to regular coefficients and hence  $|f_j[n]| = O(2^{j(\alpha+1)})$ . If  $M_r$  is the number of regular coefficients above  $T$ , it results from (3) that

$$\sum_{\substack{(j,n) \in \mathcal{R} \\ |f_j[n]| < T}} |f_j[n]|^2 = O(M_r^{-\alpha}). \quad (5)$$

At corner wavelet coefficients,  $f$  is not regular but is bounded and hence  $|f_j[n]| = O(2^j)$ . At each scale  $2^j$ , one verifies that the number of corner wavelet coefficients is proportional to the number of corners of  $f$ , which is bounded. If  $M_c$  is the number of corner coefficients above  $T$  it results that the approximation error on corner coefficients has an exponential decays in  $M_C$  and hence

$$\sum_{\substack{(j,n) \in \mathcal{C} \\ |f_j[n]| < T}} |f_j[n]|^2 = O(M_c^{-\alpha}). \quad (6)$$

At each scale  $2^j$  there are  $O(2^{-j})$  edge wavelet coefficients because edges have a finite length and wavelets have a support of size  $2^j$ . At fine scale  $2^j$  there are many such edge coefficients. These coefficients satisfy  $|f_j[n]| = O(2^j)$  because  $f$  is bounded. If  $M_e$  is the number of edge coefficients above  $T$  then a direct calculation shows that

$$\sum_{\substack{(j,n) \in \mathcal{E} \\ |f_j[n]| < T}} |f_j[n]|^2 = O(M_e^{-1}). \quad (7)$$

As a result, for an approximation with  $M = M_r + M_c + M_e$  coefficients, the error satisfies  $\|f - f_M\|^2 = O(M^{-1})$  and is dominated by the error introduced by edge coefficients.

### III Review of Geometric Image Approximations

For  $C^\alpha$  geometrically regular functions, we want to find an approximation scheme from  $M$  parameters which yields an error that decays like  $M^{-\alpha}$ , as in a wavelet approximation of uniformly  $C^\alpha$  function. Indeed, although these functions may be discontinuous, one can take advantage of the regularity of the geometry of their edge curves.

**Curvelet frame.** In order to exploit the geometric image regularity along edge curves, the image is decomposed over functions having vanishing moments and a support that is elongated. The curvelets Candès and Donoho [3] are such elongated functions that define a frame. Candès and Donoho proved that for  $\alpha = 2$  the best  $M$ -term approximations  $f_M$  of a  $C^\alpha$ -geometrically regular function  $f$  satisfies

$$\|f - f_M\|_2^2 = O(\log^3(M) M^{-2}). \quad (8)$$

Curvelet approximations are nearly optimal for  $\alpha = 2$ , but one do not reach the  $M^{-\alpha}$  optimal bound for  $\alpha > 2$  To define an orthonormal basis of  $\mathbb{R}^N$ , Do and Vetterli [6] have introduced a modified construction with contourlets implemented with a multiscale and directional filter bank. However, contourlets do not satisfy the asymptotic decay error property (8) of curvelets.

**Adaptive schemes.** Instead of decomposing the image in a fixed basis or frame, adaptive schemes adapt the approximation procedure to an estimated geometry calculated from the image.

The wedgelet scheme of Donoho [7] divides the image support into adapted dyadic squares as in figure 7. Over each square the image is approximated with a “wedge” that is constant on each side of a straight line that approximate the image edge in the square. This approach is generalized by Shukla et al. [15] with polynomials separated by polynomial curves. A Classification and Regression Tree (CART) [2] algorithm is used to optimize the dyadic image segmentation.

These approximation schemes can reach the same error decay (8) as curvelets only if the edges are discontinuities with no blurring, and the image segmentation in dyadic squares introduces blocking artifacts in the approximation which are discontinuities at the squares boundaries.

**Bandlet approximation.** The warped bandlet transform, introduced by Le Pennec and Mallat [10] also uses a dyadic square segmentation to approximate functions. In each square the geometry is not defined by finding an edge location but an orientation along which the image has regular variations. This orientation is defined by a vector field called geometric flow, that is nearly parallel to the edge as shown in figure 2, (a). To take advantage of the image regularity along the flow, a larger band parallel to the flow is warped into a rectangle a rectangle as in figure 2, (b). The image in the band is warped into a rectangular image whose flow is either horizontal or vertical. Decomposing this warped image over an anisotropic separable orthogonal wavelet basis is equivalent to decomposing the original image in orthogonal bandlets obtained by warping these wavelets [10]. Figure 2 illustrates this process. The union of these bandlet basis over all the bands defines a bandlet frame of  $L^2([0, 1]^2)$ . Le Pennec and Mallat proved that decomposing a  $C^\alpha$ -geometrically regular image over a best bandlet frame yields a bandlet approximation that satisfies  $\|f - f_M\|_2^2 = O(M^{-\alpha})$  where  $M$  is the total number of bandlet coefficients and parameters that specifies the geometric flow and segmentation.

Different image warping are used on different squares which introduces boundary discontinuities in bandlets. It is a source of mathematical and algorithmic difficulties. The resulting bandlet family defined over all squares is a frame that is not an

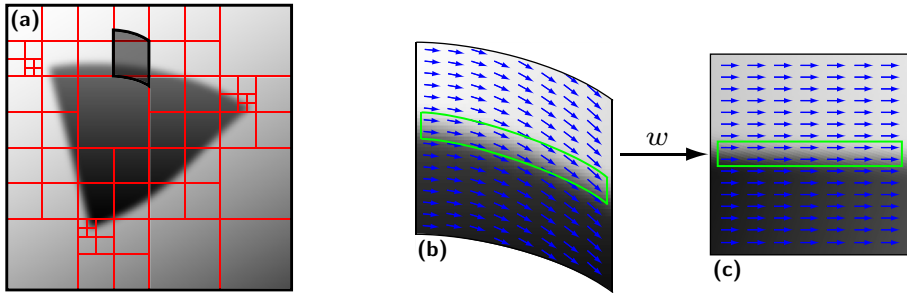


Figure 2: (a) Image segmentation in dyadic squares, inside each of which there is a single edge. (b) Zoom on a band shown in (a), with its geometric flow nearly parallel to the edge. (c) This band is warped into a square where the flow becomes horizontal. A separable anisotropic wavelet with a rectangular support in this square corresponds to a bandlet parallel to the flow in the original band.

orthogonal basis. Finding a scheme that avoids these issues is at the core of this paper.

**Capturing wavelet regularity along edges** Section II shows that wavelet approximations of geometrically regular functions are inefficient because edges create many large amplitude wavelet coefficients. To improve wavelet representations, several approaches have been proposed to further transform wavelet coefficients along edges. Wakin et al. [16] and Dragotti and Vetterli [9] perform a vector quantization of wavelet coefficients. Following the work of Matei and Cohen on adaptive ENO lifting [12], new lifting schemes have also been introduced to predict wavelet coefficients from their neighbors [4]. These approaches give good numerical results but are mostly algorithmic. They require to detect edge curves, which makes it difficult to obtain optimal approximation results when the image is blurred at any scale  $s > 0$ .

This paper constructs approximation schemes over orthogonal wavelet coefficients, which are adapted to obtain an optimal error decay  $\|f - f_M\|_2^2 = O(M^{-\alpha})$  for functions  $f$  that are  $C^\alpha$  geometrically regular at a scale  $s$ . The regularity exponent  $\alpha$  is a priori unknown and the error decay bound does not depend upon  $s$ .

## IV Polynomial Approximations of Wavelet Coefficients

To improve wavelet approximations, this section studies the regularity of wavelet coefficients located along edges. This regularity suggests approximating these wavelet coefficients with piecewise polynomials in the wavelet coefficient domain. For  $C^\alpha$  geometrically regular functions, the resulting approximation scheme yields a nearly optimal approximation error and gives the basic principles of bandlet approximations constructed over wavelet bases.

## IV.1 Anisotropic Regularity of Wavelets Coefficients

Orthogonal wavelet coefficients of  $f$  are calculated through convolutions with scaled wavelets

$$f_j \stackrel{\text{def.}}{=} \langle f, \psi_{jn}^k \rangle = f * \psi_j(n2^j) \quad \text{where} \quad \psi_j(x) = \frac{1}{2^j} \psi(-2^{-j}x).$$

The convolution guarantees that  $f_j$  is at least as regular as  $\psi_j$ . The function  $f_j$  also inherits the regularity of  $f$ .

If  $f$  is  $C^\alpha$  geometrically regular with a scale  $s$  in the sense of Definition 1 then it can be written  $f = \tilde{f} * h$ , where  $\tilde{f}$  is piecewise  $C^\alpha$  with singularities along piecewise  $C^\alpha$  curves and where  $h$  is a regularization kernel that is supported in  $[-s, s]$ . As a consequence

$$f_j = \tilde{f} * h * \psi_j = 2^j \tilde{f} * h_j \quad (9)$$

where  $h_j = 2^{-j} h * \psi_j$  is a new regularization kernel. Since  $\psi_j$  has a support in  $[-K2^j, K2^j]^2$  the new kernel  $h_j$  has a support in  $[-s_j, s_j]^2$  with  $s_j = s + K2^j$ .

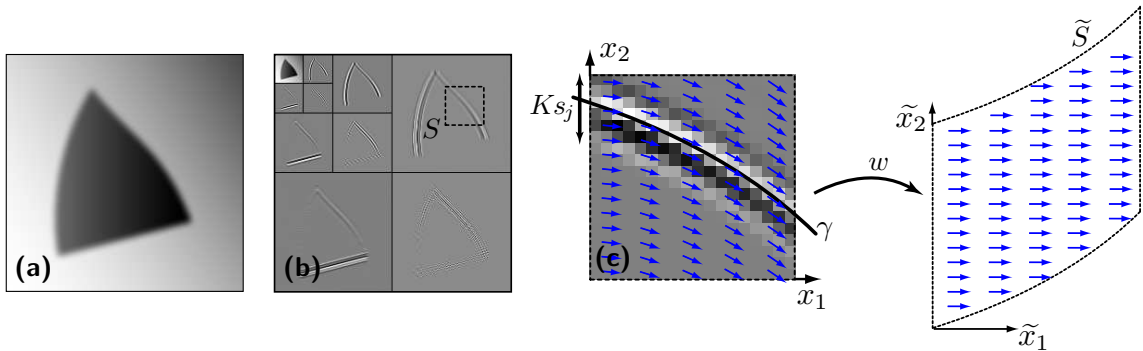


Figure 3: (a) Original image  $f$ . (b) Wavelet coefficients of  $f$ . (c) Zoom on wavelet coefficients in a square  $S$  including an edge. A geometric flow is a vector field nearly parallel to the edge curve. The warping  $w$  aligns the flow horizontally or vertically.

**Local warping** In the neighborhood of an edge, the values of the wavelet coefficients  $f_j(x)$  are regular when moving nearly parallel to the edge. Such displacement directions are specified by a geometric flow which is a vector field. Following the original bandlet approach [10], the geometric regularity of  $f_j$  is characterized after a warping that transforms the geometric flow in a horizontal or vertical vector field.

Locally an edge can be parameterized horizontally or vertically. The geometric flow is a vector field which is also parameterized horizontally or vertically and which is constant in the other direction. Figure 3 (c) shows an example in a square  $S \subset [0, 1]^2$  of length  $\lambda > 0$ , where the edge curve is parameterized horizontally by  $x_2 = \gamma(x_1)$ . For  $(x_1, x_2) \in S$  the coordinates of a flow vector can be written  $(1, \tilde{\gamma}'(x_1))$ , where  $\tilde{\gamma}'$  specifies the flow direction. If the edge is parameterized vertically, then the geometric flow is parallel horizontally and hence can locally be written  $(\tilde{\gamma}'(x_2), 1)$ .

Let us consider the warping operator

$$(\tilde{x}_1, \tilde{x}_2) = w(x_1, x_2) \stackrel{\text{def.}}{=} (x_1, x_2 - \tilde{\gamma}(x_1)), \quad (10)$$

where  $\tilde{\gamma}'$  is the derivative of  $\tilde{\gamma}$ . The curve  $x_2 = \tilde{\gamma}(x_1)$  is an integral curve of the flow. This warping transforms the geometric flow in a horizontal flow as illustrated in figure 3(c). If the flow direction  $\tilde{\gamma}'$  is sufficiently close to the edge direction  $\gamma'$  then Proposition 1 gives upper bounds on the partial derivatives of the warped wavelet coefficients  $f_{jw}(\tilde{x}) \stackrel{\text{def.}}{=} f_j(w^{-1}(\tilde{x}))$ . A similar result is proved in [10].

**Proposition 1.** *Let  $f$  be a  $C^\alpha$  geometrically regular function with a scale  $s$ . Suppose that  $f$  has only one singularity curve parameterized horizontally by  $x_2 = \gamma(x_1)$ . Let  $s_j = s + K2^j$ . There exists a constant  $C > 0$  such that for any  $2^j$  if, over a square  $S$  of length  $\lambda \leq s_j^{1/\alpha}$  a polynomial flow direction  $\gamma'(x)$  satisfies*

$$\forall (x_1, x_2) \in S = S_1 \times S_2, \quad |\gamma'(x_1) - \tilde{\gamma}'(x_1)| \leq (1 + \|\gamma\|_{C^\alpha}) \lambda^{\alpha-1}, \quad (11)$$

then the resulting warped wavelet coefficients  $f_{jw}(\tilde{x}) \stackrel{\text{def.}}{=} f_j(w^{-1}(\tilde{x}))$  satisfy

$$\forall i_1 \leq \alpha, \forall i_2 \leq p, \forall \tilde{x} \in w(S), \quad \left| \frac{\partial^{i_1+i_2} f_{jw}}{\partial x_1^{i_1} \partial x_2^{i_2}}(\tilde{x}) \right| \leq C 2^j (1 + \|\gamma\|_{C^\alpha}^\alpha) s_j^{-i_1/\alpha - i_2}. \quad (12)$$

**Proof.** Condition (11) implies that there exists an integral curve  $\tilde{\gamma}$  of the flow  $\tilde{\gamma}'$  that satisfies  $\|\gamma - \tilde{\gamma}\|_\infty \leq (1 + \|\gamma\|_{C^\alpha}) \lambda^\alpha$ .

The derivatives of  $\gamma - \tilde{\gamma}$  are first bounded using condition (11). Let  $\gamma_1$  be a Taylor expansion of degree  $\alpha - 1$  of  $\gamma$  inside  $S_1$ . It satisfies  $\|\gamma^{(i)} - \gamma_1^{(i)}\|_\infty \leq \|\gamma\|_{C^\alpha} \lambda^{\alpha-i}$ . The derivatives of  $\gamma - \tilde{\gamma}$  can be bounded using

$$\|\gamma^{(i)} - \tilde{\gamma}^{(i)}\|_\infty \leq \|\gamma^{(i)} - \gamma_1^{(i)}\|_\infty + \|\gamma_1^{(i)} - \tilde{\gamma}^{(i)}\|_\infty.$$

The second term is bounded using an expansion in  $\{\theta_m\}_{m=0}^{\alpha-1}$  the orthogonal family of Lagrange polynomial on  $S_1$

$$\gamma_1^{(i)} - \tilde{\gamma}^{(i)} = \sum_{m=0}^{\alpha-1} \langle \gamma_1 - \tilde{\gamma}, \theta_m \rangle \theta_m^{(i)}$$

and thus

$$\|\gamma_1^{(i)} - \tilde{\gamma}^{(i)}\|_\infty \leq \alpha \|\gamma_1 - \tilde{\gamma}\|_{L^\infty} \max_m \left( \|\theta_m\|_{L^1} \|\theta_m^{(i)}\|_\infty \right)$$

There exists a constant  $C_\theta$  independent of  $\lambda$  such that  $\|\theta_m\|_1 \leq C_\theta \lambda^{1/2}$  and  $\|\theta_m^{(i)}\|_\infty \leq C_\theta \lambda^{-i-1/2}$ . Using the fact that  $\|\gamma_1^{(i)} - \tilde{\gamma}^{(i)}\|_\infty \leq \|\gamma_1^{(i)} - \gamma^{(i)}\|_\infty + \|\gamma^{(i)} - \tilde{\gamma}^{(i)}\|_\infty \leq (1 + 2\|\gamma\|_{C^\alpha}) \lambda^{\alpha-i}$ , one has

$$\forall i \leq \alpha, \quad \|\gamma^{(i)} - \tilde{\gamma}^{(i)}\|_\infty \leq C_1 \lambda^{\alpha-i} \leq C_1 s_j^{1-i/\alpha}, \quad (13)$$

for a constant  $C_1$ .

Equation (9) shows that  $f_j = 2^j \tilde{f} * h_j$  where  $h_j = 2^{-j} h * \psi_j$  has a support in  $[-s_j, s_j]^2$ . The proof of inequality (12) is performed by expanding the derivatives of the convolution

$$\begin{aligned} f_{jw}(x) &= 2^j \int \tilde{f}(x_1 - u_1, x_2 + \tilde{\gamma}(x_1) - u_2) h_j(u) du \\ &= 2^j \int \tilde{f}(A(x, u)) h_j(B(x, u)) du \end{aligned}$$



where

$$A(x, u) \stackrel{\text{def.}}{=} (x - u_1, x_2 + \gamma(x_1 - u_1) - u_2) \quad \text{and} \quad B(x, u) \stackrel{\text{def.}}{=} (u_1, u_2 + \tilde{\gamma}(x_1) - \gamma(x_1 - u_1)).$$

Taking derivatives leads to

$$\frac{\partial^{i_1+i_2} f_{j\text{w}}}{\partial x_1^{i_1} \partial x_2^{i_2}}(x) = 2^j \int \sum_{d=0}^{i_1} \binom{d}{i_1} \frac{\partial^{i_1-d}}{\partial x_1^{i_1-d}} \left[ \tilde{f}(A(x, u)) \right] \frac{\partial^d}{\partial x_1^d} \left[ \left( \frac{\partial^{i_2} h_j}{\partial x_2^{i_2}} \right) (B(x, u)) \right] du. \quad (14)$$

By hypothesis the function  $x_1 \mapsto \tilde{f}(A(x, u))$  is regular and there exists a constant  $C$  such that

$$\left| \frac{\partial^{i_1-d}}{\partial x_1^{i_1-d}} \left[ \tilde{f}(A(x, u)) \right] \right| \leq C \|f\|_{C^\alpha(\Lambda)} \max(\|\gamma\|_{C^\alpha}^\alpha, 1).$$

The second term of equation (14) is bounded with the Faa di Bruno formula for the derivatives of a composition

$$\frac{\partial^d}{\partial x_1^d} \left[ \left( \frac{\partial^{i_2} h_j}{\partial x_2^{i_2}} \right) (B(x, u)) \right] \quad (15)$$

$$= \sum_{(k_s)_s} \frac{d!}{k_1! \dots k_d!} \frac{\partial^{k+i_2} h_j}{\partial^k x_1 \partial^{i_2} x_2} (B(x, u)) \prod_{s=1}^d \left( \frac{\tilde{\gamma}^{(s)}(x_1) - \gamma^{(s)}(x_1 - u_1)}{s!} \right)^{k_s}, \quad (16)$$

where the sum is on all  $d$ -tupel  $(k_s)_s$  such that  $\sum s k_s = d$  and where  $k \stackrel{\text{def.}}{=} \sum k_s$ . The deviation on the geometry is bounded using

$$|\tilde{\gamma}^{(s)}(x_1) - \gamma^{(s)}(x_1 - u_1)| \leq \underbrace{|\tilde{\gamma}^{(s)}(x_1) - \gamma^{(s)}(x_1)|}_{\leq C_1 s_j^{1-s/\alpha}} + \underbrace{|\gamma^{(s)}(x_1) - \gamma^{(s)}(x_1 - u_1)|}_{\leq \|\gamma\|_{C^\alpha} s_j} \quad (17)$$

$$\leq C_1 \max(1, \|\gamma\|_{C^\alpha}^\alpha) s_j^{1-s/\alpha}. \quad (18)$$

Using the fact that

$$\left| \frac{\partial^{k+i_2} h_j}{\partial^k x_1 \partial^{i_2} x_2} (B(x, u)) \right| \leq C_\psi s_j^{-2-i_2-k}, \quad (19)$$

where  $C_\psi$  is a constant that depends only on  $\psi$ , expression (15) together with bounds (18) and (19) leads to

$$\begin{aligned} \left| \frac{\partial^d}{\partial x_1^d} \left[ \frac{\partial^{i_2} h_j}{\partial x_2^{i_2}} (B(x, u)) \right] \right| &\leq C \sum_{(k_s)_s} s_j^{-2-i_2-k} \prod_{s=1}^d s_j^{k_s(1-i/\alpha)} \\ &\leq C \max(1, \|\gamma\|_{C^\alpha}^\alpha) s_j^{-2-i_2-\alpha/d}. \end{aligned}$$

The term corresponding to  $d = i_1$  thus dominates in equation (14) and one concludes to the bound (12) using the fact that the size of the support of  $h_j$  is  $s_j^2$ .  $\square$

## IV.2 Polynomial Regression Over Band-shaped Domains

The regularity of warped wavelet coefficients along edges implies that these coefficients can be approximated with piecewise polynomials. This section explains how to construct these polynomials to approximate edge wavelet coefficients at each scale  $2^j$ . We suppose that  $f$  is  $C^\alpha$  geometrically regular, with a regularization scale  $s$ . A regularization by a scale  $s \neq 0$  is essentially equivalent to translating the wavelet scale  $2^j$  by  $s$  and replacing  $2^j$  by  $s_j = s + K2^j$ . We suppose that  $s = 0$  to simplify explanations.

This section proves that appropriate piecewise polynomial approximation of the wavelet coefficients of  $f$  leads to an approximation  $f_M$  of  $f$  that satisfies

$$\|f - f_M\|_2^2 \leq C \log(M)^\alpha M^{-\alpha}, \quad (20)$$

where  $M$  is the total number of parameters needed to specify  $f_M$ . Technical details are omitted to carry the main ideas. A more precise theorem is proved in section V.

The set  $\mathcal{E}$  of edge wavelet coefficients  $f_j[n]$  corresponds to wavelets  $\psi_{jn}$  whose support intersects a single edge of  $f$ . Since the support of  $\psi_{jn}$  is included in  $[2^j n - K2^j, 2^j n + K2^j]$ , it corresponds to the set of points  $2^j n$  that are at a distance smaller than  $K2^j$  from an edge. This set  $\mathcal{E}$  of edge wavelet coefficients is first segmented in squares of length  $\lambda$  such that in any such square  $S$ , the corresponding edge is parameterized either horizontally or vertically. Figure 4 (c) shows an example of such a segmentation. In the following, a horizontal parameterization  $x_2 = \gamma(x_1)$  is assumed.

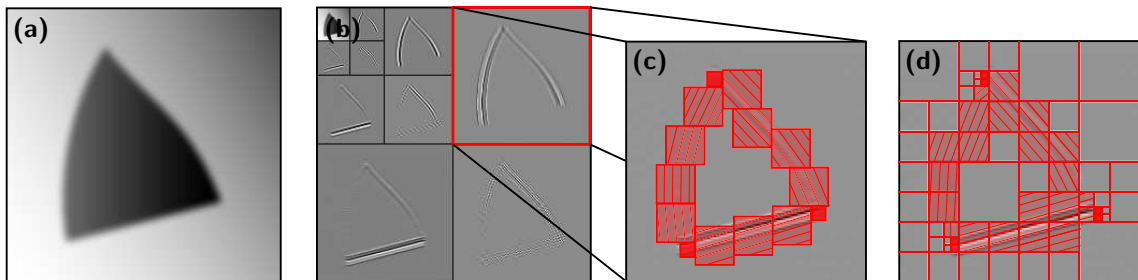


Figure 4: (a) A geometrically regular image. (b) Wavelet coefficients. (c) A non-dyadic segmentation into square together with bands of size  $\lambda \times \mu$  over each square crossing the singularities. (d) A dyadic subdivision of the coefficients together with a dyadic subdivision into bands.

The width  $\lambda$  of the squares is chosen in order to match the precision of the approximation. Over each square  $S$ , an approximated flow  $\tilde{\gamma}'$  is defined as the Taylor expansion of degree  $\alpha - 2$  of  $\gamma$  inside  $S$ . It satisfies inside  $S$

$$\|\gamma' - \tilde{\gamma}'\|_\infty \leq \|\gamma\|_{C^\alpha} \lambda^{\alpha-1}. \quad (21)$$

The warping  $w(x_1, x_2) = (x_1, x_2 - \tilde{\gamma}(x_1))$  maps the flow  $\tilde{\gamma}'$  onto a horizontal flow. As shown on Figure 5 (d), the warped domain  $w(S)$  is subdivided into horizontal bands of length  $\lambda$  and width  $\mu$  that will be adjusted. This set of warped band defines a segmentation of the square  $S$  into bands that follow the approximated flow  $\tilde{\gamma}'$ , see Figure 5 (e).

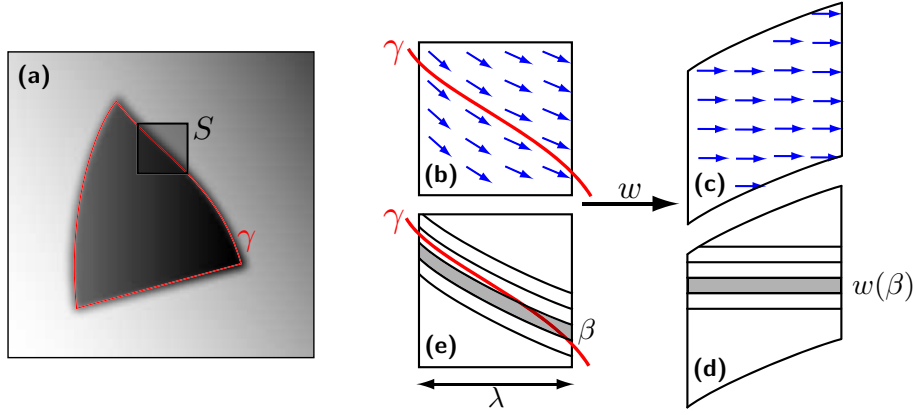


Figure 5: (a) A geometrically regular image with a square  $S$  of width  $\lambda$  on which the geometry  $\gamma$  is parameterized horizontally. (b) Approximated geometric flow  $\tilde{\gamma}$ . (c) Horizontal flow over the warped domain  $w(S)$ . (d) A horizontal band  $w(\beta)$  of size  $\lambda \times \mu$  in the warped domain. (e) The corresponding band  $\beta$ .

**Polynomial approximation.** The size  $\lambda \times \mu$  of the bands is set by analyzing the polynomial approximation error over each band  $\beta \subset S$ . Wavelet coefficients  $f_j[n]$ , for  $2^j n \in \beta$  are approximated by a polynomial  $P_\beta$  of degree  $p - 1$  defined over the warped domain  $w(\beta)$ . Choosing  $P_\beta$  as a Taylor expansion of the warped function  $f_{jw}$  inside  $w(\beta)$  leads to an error

$$\forall x \in \beta, \quad |f_j(x) - P_\beta(w(x))| = |f_{jw}(w(x)) - P_\beta(w(x))| \quad (22)$$

$$\leq \sum_{i_1+i_2=\alpha} \left\| \frac{\partial^\alpha f_{jw}}{\partial x_1^{i_1} \partial x_2^{i_2}} \right\|_\infty \lambda^{i_1} \mu^{i_2}. \quad (23)$$

Condition (21) allows to use Lemma 1 to bound the derivatives of  $f_{jw}$  and leads to

$$\begin{aligned} \forall x \in \beta, \quad |f_j(x) - P_\beta(w(x))| &\leq C (1 + \|\gamma\|_{C^\alpha}^\alpha) 2^j \sum_{i_1+i_2=\alpha} s_j^{-i_1/\alpha - i_2} \lambda^{i_1} \mu^{i_2} \\ &\leq C' (2^j s_j^{-1} \lambda^\alpha + 2^j s_j^{-\alpha} \mu^\alpha). \end{aligned}$$

To minimize this approximation error bound, the ratio width/length =  $\mu/\lambda$  of the bands is chosen so that

$$s_j^{-1} \lambda^\alpha = s_j^{-\alpha} \mu^\alpha \implies \forall x \in \beta, \quad |f_j(x) - P_\beta(w(x))| \leq 2 C' 2^j s_j^{-1} \lambda^\alpha. \quad (24)$$

This defines the width  $\mu$  of the bands as a function of their length  $\lambda$ .

Since the number of bands is

$$M_j = \frac{\text{area of all bands}}{\text{area of a band}} = L K 2^j / (\mu \lambda),$$

where  $L$  is the total length of the edges curves of  $f$ , the chosen length  $\lambda$  can be expressed as a function of the number of bands  $M_j$

$$\lambda = \mu s_j^{1/\alpha - 1} \implies \lambda = (CK)^{-1} s_j^{-1} M_j^{-1}. \quad (25)$$

**Approximation error.** Let  $\tilde{f}_j[n] = P_\beta(w(2^j n))$  be the approximated wavelet coefficients. This construction is carried over each band  $\beta$ , which defines a discrete piecewise polynomial approximation  $\tilde{f}_j$  over the set of coefficients located near edges. Let  $M_0 > 0$  be some fixed integer. For locations  $(2^j n)$  outside the set of bands, the approximated coefficients are set to  $\tilde{f}_j[n] = f_j[n]$  if  $f_j[n]$  is one of the  $M_0$  highest coefficients and  $\tilde{f}_j[n] = 0$  otherwise.

Following equation (4),  $\mathcal{E}$  denotes the set of coefficients located near edge curves and  $\mathcal{R} \cup \mathcal{C}$  are the remaining coefficients. For each scale  $2^j \leq 2^{j_0} \stackrel{\text{def.}}{=} M_0^{-\alpha}$ , the number of band is fixed to  $M_j = M_0$ . Using equations (24) and (25), the approximation error can be computed for the set of edge coefficients

$$\sum_{j \geq j_0} \|f_j - \tilde{f}_j\|_{\ell^2(\mathcal{E})}^2 \leq \sum_{j \geq j_0} (\text{nbr. coef. in bands}) \max_{\beta, (2^j n) \in \beta} |f_j[n] - P_\beta(w(2^j n))|^2 \quad (26)$$

$$\leq \sum_{j \geq j_0} L K 2^{-j} (2 C' 2^j s_j^{-1} \lambda^\alpha)^2 \leq \sum_{j \geq j_0} C_1 s_j^{-1} \lambda^{2\alpha} \quad (27)$$

$$\leq \sum_{j \geq j_0} C_1 M_j^{-\alpha} \leq C M_0^{-\alpha} \log^\alpha(M_0). \quad (28)$$

The scale  $2^{j_0} = M_0^{-\alpha}$  ensures that the error on the remaining fine scales  $2^j < 2^{j_0}$  is bounded by  $O(M_0^{-\alpha})$ . As shown in equations (5) and (6), the error for the remaining coefficients in region  $\mathcal{R} \cup \mathcal{C}$  satisfies

$$\sum_{j \geq j_0} \|f_j - \tilde{f}_j\|_{\ell^2(\mathcal{R} \cup \mathcal{C})}^2 = O(M_0^{-\alpha}). \quad (29)$$

Let

$$f_M \stackrel{\text{def.}}{=} \sum_{j,n} \tilde{f}_j[n] \psi_{jn}.$$

It results from equations (28) and (29) that

$$\|f - f_M\|_2^2 = \sum_j \|f_j - \tilde{f}_j\|_2^2 = O(M_0^{-\alpha}). \quad (30)$$

**Number of parameters.** Let  $M$  be the total number of parameters that specifies  $f_M$ . One has  $M = M_B + M_G$  where  $M_B$  is the number of polynomial coefficients to specify the approximated coefficients  $\tilde{f}_j[n]$  and  $M_G$  is the number of geometric coefficients to specify the polynomial bands.

For each relevant scale  $2^j \geq 2^{j_0}$ , the number of polynomial coefficients is  $\frac{p(p+1)}{2} M_j = \frac{p(p+1)}{2} M_0$  so that

$$M_B = M_0 + \sum_{j \geq j_0} \frac{p(p+1)}{2} M_j = C \log(M_0) M_0.$$

To specify each band  $\beta$  one need to record the coordinates of the segmentation square  $S$  containing  $\beta$  and the coefficients of the adapted polynomial flow  $\tilde{\gamma}'$ . Since  $\tilde{\gamma}'$  is a 1D polynomial of degree  $p - 1$ , the number of geometric coefficients at each scale  $2^j \geq 2^{j_0}$  is proportional to  $M_j$  and thus  $M_G$  is proportional to  $M_0 \log(M_0)$ .

Summing the number of polynomial and geometric parameters leads to

$$M = M_B + M_G = C \log(M_0) M_0.$$

Combining this result with the error bound (30) leads to the global approximation bound (20)

$$\|f - f_M\|_2^2 \leq C \log(M)^\alpha M^{-\alpha}.$$

The technical details are skipped since the bandlet basis construction presented in section V gives a constructive proof of this results, without the sub-optimal  $\log(M)^\alpha$  factor.

## V Orthogonal Bandlet Bases

The scheme presented in the previous section does not provide an effective algorithm to compute the adapted geometric flows  $\tilde{\gamma}'$  since the location of the edges  $\gamma$  is unknown. It does not either describe a way to estimate the optimal size  $\lambda \times \mu$  of the bands since the Holder exponent  $\alpha$  is unknown.

In this section, the adapted flow and the segmentation are computed using a best basis search and the bands size is indirectly by a thresholding in an Alpert wavelet basis. For fast computations, the set of potential segmentations is obtained by dyadic subdivisions of the wavelet image coefficients. Figure 4 (d), shows an example of such dyadic subdivisions. Orthogonal bandlet bases are defined with a polynomial multiwavelet Alpert transform over the wavelet image coefficients.

### V.1 Alpert Wavelets Bases

Wavelet coefficients located in a square  $S \subset [0, 1]^2$  of length  $\lambda$  at a given scale  $2^j$  are approximated with an adaptive scheme. A multiresolution is defined over the space  $\ell^2(S)$  of wavelet coefficients in  $S$  with piecewise polynomials over bands of dyadic widths. A multiwavelet Alpert basis [1] is derived from this multiresolution. A thresholding in this basis is equivalent to automatically adjusting the optimal width  $\mu = s_j^{1/\alpha-1} \lambda$  without any knowledge of  $\alpha$ .

**Alpert polynomial multiresolution.** A geometric flow direction  $\tilde{\gamma}'$  is assumed to be known over  $S$ . The warping operator  $w$  in (10) warps  $S$  into  $\tilde{S}$  as shown in Figure 6 (a). Points  $x_n \stackrel{\text{def.}}{=} 2^j n \in S$  are warped onto  $\tilde{x}_n \stackrel{\text{def.}}{=} w(2^j n)$ . The space  $\ell^2(S)$  is the set of sampled functions  $\{g(x_n)\}_{2^j n \in S}$ . Similarly  $\ell^2(\tilde{S})$  denotes functions sampled in the warped domain  $\{\tilde{g}(\tilde{x}_n)\}_{2^j n \in S}$ . A set of wavelets coefficients  $\{f_j[n]\}_{2^j n \in S}$  can be considered as samples from the function  $f_j(x)$  at points  $x_n$  or from the function  $\tilde{f}_j(\tilde{x})$  at points  $\tilde{x}_n$ .

To define a multiresolution, for each scale  $2^\ell$ , the warped square  $\tilde{S}$  is recursively subdivided into  $2^{-\ell}$  horizontal bands  $\tilde{S} = \bigcup_{i=0}^{2^{-\ell}-1} \tilde{\beta}_{\ell,i}$ . This process is illustrated in figure 6. The scheme divides  $\tilde{\beta}_{\ell,i} = \tilde{\beta}_{\ell-1,2i} \cup \tilde{\beta}_{\ell-1,2i+1}$  by looking for a horizontal cut ensuring that  $\tilde{\beta}_{\ell-1,2i}$  and  $\tilde{\beta}_{\ell-1,2i+1}$  contain the same number of points. The recursive subdivision is stopped at the scale  $\ell = L$  such that  $2^L (\lambda 2^{-j})^2 \leq p(p+1)/2$ .

Each band  $\beta_{\ell,i} \stackrel{\text{def.}}{=} w^{-1}(\tilde{\beta}_{\ell,i})$  in the original square  $S$  has a width roughly equal to  $\lambda 2^\ell$  and contains  $2^\ell(\lambda 2^{-j})^2$  sampling points. Note that some bands near the boundary of  $S$  might be disconnected.

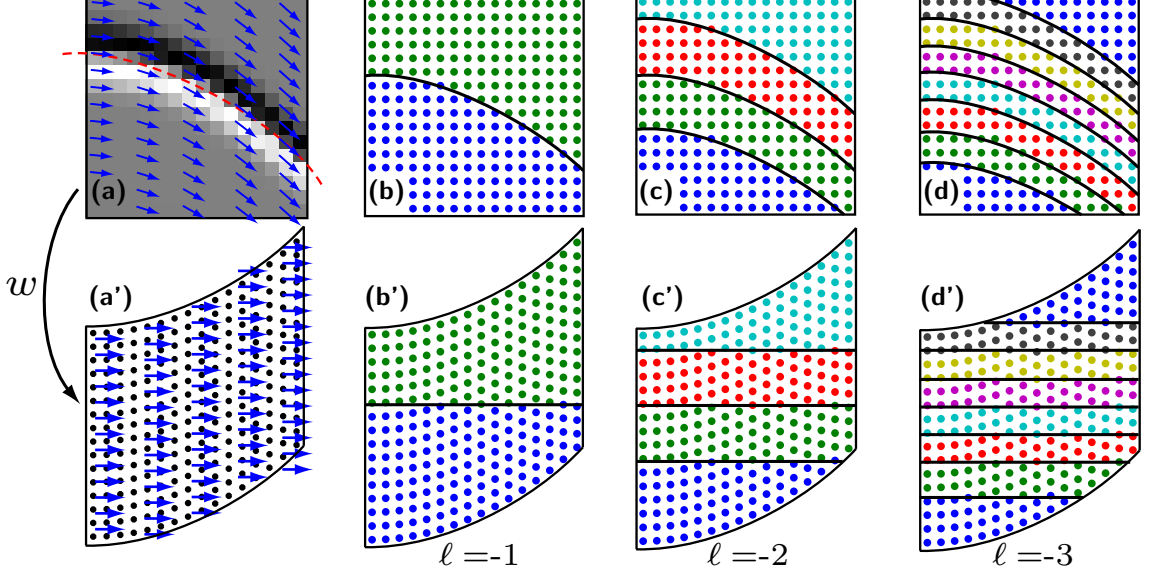


Figure 6: (a) Wavelet coefficients over a square  $S$  near an edge curve. The adapted flow  $\tilde{\gamma}'$  is depicted as arrows. (a') The warping maps to flow onto a horizontal flow. (c-d) Further refinements of the segmentation in bands.

Alpert multiresolution spaces  $\tilde{V}_\ell \subset \ell^2(\tilde{S})$  are defined for each scale  $2^\ell$ ,  $L \leq \ell \leq 0$  by

$$\tilde{V}_\ell \stackrel{\text{def.}}{=} \left\{ \tilde{g} \in \ell^2(\tilde{S}) \mid \forall (2^j n) \in \beta_{\ell,i}, \quad \tilde{g}(\tilde{x}_n) = P_i(\tilde{x}_n), \right. \\ \left. \text{with } P_i \text{ polynomial and } \deg(P_i) < p. \right\}.$$

These spaces are embedded since  $\tilde{V}_\ell \subset \tilde{V}_{\ell-1}$ . An orthogonal basis  $\{\tilde{h}_{\ell,i,k}\}_{i,k}$  of each space  $\tilde{V}_\ell$  is defined using discrete Legendre polynomials, where  $k = (k_1, k_2)$  with  $k_1 + k_2 < p$  indexes the polynomial degree and  $0 \leq i < 2^{-\ell}$  indexes the position. Basis vectors  $\{\tilde{h}_{\ell,i,k}\}_k$  are obtained by Gram-Schmidt orthogonalization of the set of monomials  $\{\tilde{P}_k\}_k$  vectors defined by

$$\forall \tilde{x}_n \in \tilde{\beta}_{\ell,i}, \quad \tilde{P}_k(\tilde{x}_n) = (\tilde{x}_n^1)^{k_1} (\tilde{x}_n^2)^{k_2}. \quad \text{where} \quad \tilde{x}_n = (\tilde{x}_n^1, \tilde{x}_n^2).$$

Alpert wavelets  $\{\tilde{a}_{\ell,i,k}\}_{i,k}$  are an orthogonal basis of the orthogonal complement  $\tilde{W}_\ell$  of  $\tilde{V}_\ell$  in  $\tilde{V}_{\ell-1}$  that satisfies  $\tilde{V}_{\ell-1} = \tilde{V}_\ell \oplus^\perp \tilde{W}_\ell$ . The Alpert wavelet vectors  $\{\tilde{a}_{\ell,i,k}\}_k$  are computed by Gram-Schmidt orthogonalization of the family

$$\{\tilde{h}_{\ell-1,2i,k} - \tilde{h}_{\ell-1,2i+1,k}\}_{k_1+k_2 < p} \subset \tilde{V}_{\ell-1}$$

against the family  $\{\tilde{h}_{\ell,i,k}\}_{k_1+k_2 < p} \subset \tilde{V}_\ell$ .

The resulting multi-wavelets vectors  $\tilde{a}_{\ell,i,k}$  are sampled from piecewise polynomial functions that are discontinuous at the middle of the band  $\tilde{\beta}_{\ell,i,k}$ . Each vector  $\tilde{a}_{\ell,i,k}$  has vanishing moments over the warped domain since is orthogonal to  $\tilde{V}_\ell$

$$\forall k_1 + k_2 < p, \quad \sum_n \tilde{a}_{\ell,i,k}(\tilde{x}_n) (\tilde{x}_n)^k = 0$$

where  $(\tilde{x}_n)^k \stackrel{\text{def.}}{=} (\tilde{x}^1)^{k_1} (\tilde{x}^2)^{k_2}$  for each point  $\tilde{x}_n = (\tilde{x}^1, \tilde{x}^2)$  in the warped domain. The orthogonal basis  $\{\tilde{a}_{\ell,i,k}\}_{\ell,i,k}$  of  $\ell^2(\tilde{S})$  defines an orthogonal Alpert basis of  $\ell^2(S)$  by

$$a_{\ell,i,k}(x_n) \stackrel{\text{def.}}{=} \tilde{a}_{\ell,i,k}(\tilde{x}_n).$$

Note that this definition over the original domain  $S$  does not involve any interpolation.

In the following  $m = (i, k)$  indexes the  $p(p+1)2^{\ell-1}$  Alpert wavelets  $\{a_{\ell,m}\}$  at a scale  $2^\ell$  and we consider  $m$  as an integer. The orthogonal Alpert basis  $\mathcal{B}(S, \tilde{\gamma}')$  of  $\ell^2(S)$  is defined by

$$\mathcal{B}(S, \tilde{\gamma}') \stackrel{\text{def.}}{=} \{a_{\ell,m} \mid L \leq \ell \leq 0 \text{ and } 0 \leq m < p(p+1)2^{\ell-1}\}.$$

In the following we write  $a_{\ell,m}[n] = a_{\ell,m}(x_n)$  the coordinates of the discrete Alpert vector.

The following proposition gives the normalization of the Alpert basis vectors, which is used to find upper bounds for Alpert coefficients.

**Proposition 2.** *There exists a constant  $C_a$  such that for any flow  $\tilde{\gamma}'$  defined over  $S$  the Alpert basis  $\mathcal{B}(S, \tilde{\gamma}') = \{a_{\ell,m}\}_{\ell,m}$  satisfies*

$$\|a_{\ell,m}\|_2 = 1, \quad \text{and} \quad \|a_{\ell,m}\|_\infty \leq C_a \lambda^{-1} 2^{-\ell/2} 2^j, \quad (31)$$

$$\text{and} \quad \|a_{\ell,m}\|_1 \leq C_a \lambda 2^{\ell/2} 2^{-j}. \quad (32)$$

**Proof.** The support of an Alpert vector  $a_{\ell,m}$  is  $\beta_{\ell,m}$  and its cardinal is

$$\text{Card}(\beta_{\ell,m}) = \frac{\#\{\text{points in } S\}}{\#\{\text{nbr. bands}\}} = \lambda^2 2^\ell 2^{-2j}.$$

Let

$$I_{\ell,m} = \left\{ n \mid |a_{\ell,m}[n]| \geq \frac{1}{2} \|a_{\ell,m}\|_\infty \right\}. \quad (33)$$

One has the following bound

$$1 = \|a_{\ell,m}\|_2 \geq \sum_{n \in I_{\ell,m}} |a_{\ell,m}[n]|^2 \geq \frac{1}{2} \text{Card}(I_{\ell,m}) \|a_{\ell,m}\|_\infty^2. \quad (34)$$

Asymptotically, the coefficients  $a_{\ell,m}[n]$  are samples from one of  $p(p+1)/2$  piecewise polynomial scaled by a factor  $2^\ell$ . One thus has

$$\forall \ell, m, \quad \text{Card}(I_{\ell,m}) \geq C_a \text{Card}(\beta_{\ell,m}),$$

where  $C_b$  is a constant that does not depend on  $\ell$  and  $m$ . It results that

$$\|a_{\ell,m}\|_\infty \leq \sqrt{2} C_a \text{Card}(\beta_{\ell,m})^{-1/2}, \quad (35)$$

which implies (31) by inserting (35) and (33) in (34). The bound on the  $\ell^1$  norm is obtained using

$$\|a_{\ell,m}\|_1 \leq \text{Card}(\beta_{\ell,m}) \|a_{\ell,m}\|_\infty.$$

□

## V.2 Bandletization of Wavelet Coefficients

**Segmentation of wavelets coefficients.** For each scale  $2^j$ , a segmentation  $\mathcal{S}_j = \{S\}_{S \in \mathcal{S}_j}$  subdivides  $[0, 1]^2$  into non-overlapping squares  $S$  of width bigger than  $2^j$ . Following the ideas of section IV.2, the length  $\lambda$  of the squares must be adapted to the approximation precision.

In order to do so, the segmentation  $\mathcal{S}_j$  is enforced to contain only squares of dyadic lengths  $2^{-k}$ . A segmentation of  $[0, 1]^2$  using dyadic squares is obtained by a recursive subdivision of the original square into four squares of equal sizes. On figure 7 one can see the subdivision steps leading to the construction of a dyadic subdivision  $\mathcal{S}_j$ , together with the tree representing the subdivision process. Each square  $S \in \mathcal{S}_j$  of size  $2^{-k} \times 2^{-k}$  corresponds to a leaf of the tree at depth  $k$ .

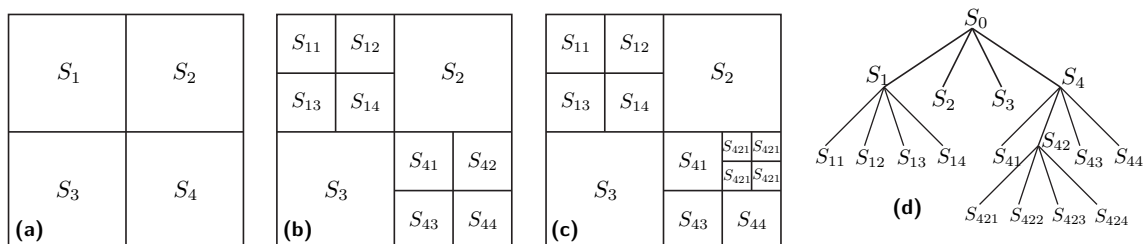


Figure 7: (a-c) Construction of a dyadic segmentation by successive subdivisions. (d) Quadtree representation of the segmentation. Each leaf of the tree indicates a square in the segmentation and can be represented with a binary string whose length is the depth of the tree.

For a geometrically regular image  $f$ , an adapted segmentation  $\mathcal{S}_j$  should encapsulate the singularity curves in a set of squares whose size  $\lambda$  matches the approximation precision. Junctions between singularity curves should be covered by small squares as in section IV.2 and the remaining domain should be covered by the largest possible squares.

The squares of the dyadic segmentation  $\mathcal{S}_j$ , are partitioned into several groups:

- The set  $\mathcal{E}(\mathcal{S}_j) = \mathcal{E}^H(\mathcal{S}_j) \cup \mathcal{E}^V(\mathcal{S}_j)$  of edge squares. By definition, an horizontal (resp. vertical) edge square  $S \in \mathcal{E}^H$  (resp.  $S \in \mathcal{E}^V$ ) is a square smaller than  $\lambda(T)$  that is at a distance less than  $s_j \stackrel{\text{def.}}{=} s + K2^j$  from one and only one edge curve. This curve is supposed to be parameterized horizontally (resp. vertically) by  $x_2 = \gamma(x_1)$  (resp.  $x_1 = \gamma(x_2)$ ) with

$$|\gamma'| \leq 2. \quad (36)$$

Over such a square, the adapted flow  $\tilde{\gamma}'$  should be designed to approximate  $\gamma'$ .

- The set  $\mathcal{C}(\mathcal{S}_j)$  of corner squares. By definition, a corner square is a square that contains the junction of two curves.
- The set  $\mathcal{R}(\mathcal{S}_j)$  of regular squares: these are the squares that do not contain any edge curves.

Figure 8 shows an example of such an adapted dyadic segmentation.



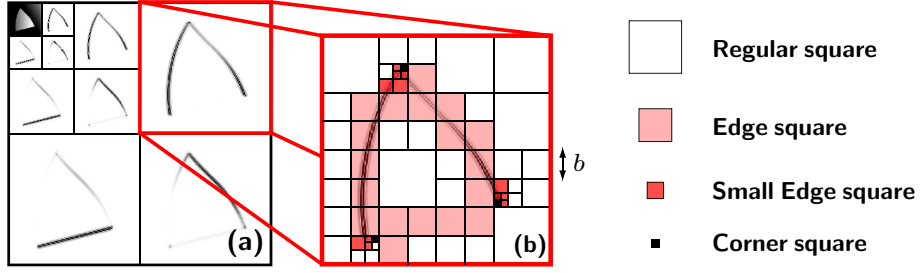


Figure 8: (a) Wavelet transform of a geometrically regular function. (b) Example of a segmentation  $\mathcal{S}_j$  adapted to the geometry of the function.

**Bandlet basis.** Inside each square  $S \in \mathcal{S}_j$ , the choice of discrete Alpert basis  $\mathcal{B}(S, \tilde{\gamma}'_S)$  constructed in section V.1 depends on the choice of a geometric flow  $\tilde{\gamma}'_S$ .

Over regular square  $S \in \mathcal{R}(\mathcal{S}_j)$ , a basis adapted to the geometrically image  $f$  should not transform the wavelets coefficients. For such a square  $S$ , the flow  $\tilde{\gamma}'_S$  is undefined and the projection onto  $\mathcal{B}(S, \tilde{\gamma}'_S)$  leaves the wavelet coefficients in  $S$  unchanged.

A dyadic segmentation together with the adapted flows  $\Gamma_j = (\mathcal{S}_j, \{\tilde{\gamma}'_S\}_{S \in \mathcal{S}_j})$  specifies a bandeletization basis  $\mathcal{B}(\Gamma_j)$  of the whole space of wavelet coefficients at a scale  $2^j$

$$\mathcal{B}(\Gamma_j) \stackrel{\text{def.}}{=} \bigcup_{S \in \mathcal{S}_j} \mathcal{B}(S, \tilde{\gamma}'_S).$$

A discrete Alpert vector  $a_\nu \in \mathcal{B}(\Gamma_j)$  is thus specified by  $\nu = (j, S, \tilde{\gamma}'_S, \ell, m)$  where

- $2^j$  is a scale of the 2D wavelet transform,
- $S \in \mathcal{S}_j$  is a square of width  $\lambda = 2^{-L/2} 2^j$ ,
- $\tilde{\gamma}'_S$  is a geometric flow,
- $\ell \in \{L, \dots, 0\}$  and  $m \in \{0, \dots, p(p+1)2^{-\ell-1} - 1\}$  are the scale and index of a discrete Alpert vector  $a_{\ell, m} \in \mathcal{B}(S, \tilde{\gamma}'_S)$  and

$$\forall (2^j n) \in S \in \mathcal{S}_j, \quad a_\nu[n] = a_{\ell, m}[n] \quad \text{where} \quad \mathcal{B}(S, \tilde{\gamma}'_S) = \{a_{\ell, m}\}_{\ell, m}.$$

The coefficients  $a_\nu[n]$  are the coordinates of a bandlet function  $b_\nu \in L^2([0, 1]^2)$  in the wavelet basis. This function is defined by

$$b_\nu(x) = \sum_n a_\nu[n] \psi_{jn}(x). \quad (37)$$

This function is called a bandlet. Indeed it is a combination of wavelets along a band and its support is thus also along a band as illustrated in figure 9. Bandlets are obtained from an orthogonal wavelet basis with an orthogonal transformation that we call a “bandletization”. It results that if we apply this transformation to each scale  $2^j$ ,

$$\mathcal{B}(\Gamma) \stackrel{\text{def.}}{=} \bigcup_{j \leq 0} \{b_\nu \setminus a_\nu \in \mathcal{B}(\Gamma_j)\}, \quad \text{where} \quad \Gamma \stackrel{\text{def.}}{=} \bigcup_{j \leq 0} \Gamma_j,$$

is an orthogonal basis of  $L^2([0, 1]^2)$ .

Bandlets are regular functions because the underlying wavelets are regular. The support of bandlets functions overlap because the support of wavelets overlap. Approximations in bandlet basis creates no blocking artifact.

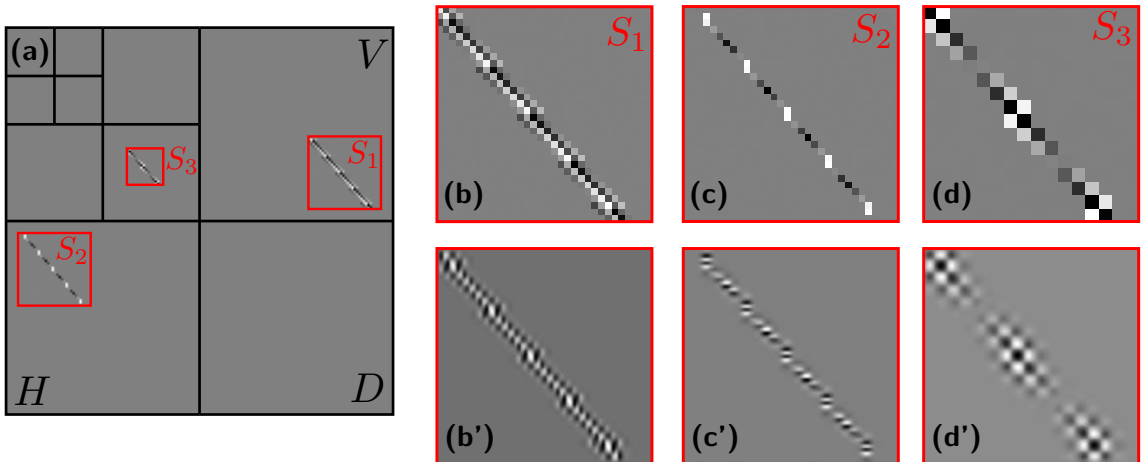


Figure 9: (a) Localization on the wavelet domain of the squares  $S_i$  on which each Alpert wavelet vector is defined. (b-d) Discrete Alpert vectors  $a_{li}$  for various scale  $2^l$ . (b'-d') Corresponding bandlet functions  $b_{li}$ .

## VI Best Bandlet Basis Approximation

The set of bandlet bases is a dictionary of orthogonal basis  $\mathcal{D} = \{\mathcal{B}(\Gamma)\}_{\Gamma \in G}$ . One would like to approximate a function  $f$  by a thresholding in a bandlet basis  $\mathcal{B}(\Gamma^*) = \{b_\nu\}_\nu$  that provides the lowest approximation error for a given  $M$  number of parameters. The thresholded approximation of  $f$  in  $\mathcal{B}(\Gamma^*)$

$$f_M \stackrel{\text{def.}}{=} \sum_{|\langle f, b_\nu \rangle| > T} \langle f, b_\nu \rangle b_\nu \quad (38)$$

is described with  $M = M_B + M_D$  coefficients, where

$$M_B \stackrel{\text{def.}}{=} \text{Card} \{ \nu \mid |\langle b_\nu, f \rangle| \geq T \}, \quad (39)$$

is the number of bandlet coefficients above the threshold and  $M_D$  is the number of coefficients needed to specify  $\Gamma^*$  in  $G$ . Our goal is to choose  $\mathcal{B}(\Gamma^*)$  such that

$$\|f - f_M\|_2 = O(M^{-\alpha}).$$

### VI.1 Bandlet Basis Dictionary

The number of bandlet bases in the dictionary is reduced to a finite size by parameterizing the flows  $\tilde{\gamma}'_S$  up to a precision  $T^2$  and by limiting the bandletization to the scales  $2^j > T^2$ .

**Parameterizing the geometry.** Inside a square  $S$  of length  $\lambda$ , a geometric flow direction  $\tilde{\gamma}'$  is parameterized by a polynomial of degree  $p - 2$  whose coefficients are quantized at a precision  $\tau$

$$\tilde{\gamma}'(x) = \sum_{i=0}^{p-2} \frac{a_i \tau}{\lambda^{i+1}} x^i \quad \text{with} \quad a_i \in \mathbb{Z}. \quad (40)$$

In the following, the precision is set to

$$\tau = \tau(\lambda) \stackrel{\text{def.}}{=} \lambda^p / (p - 1), \quad (41)$$

which ensures that a flow  $\gamma'$  can be approximated by a flow  $\tilde{\gamma}'$  quantized as in equation (40) with  $\|\gamma' - \tilde{\gamma}'\|_\infty = O(\lambda^{\alpha-1})$ .

The adapted geometric flow  $\tilde{\gamma}'$  is thus chosen in the finite set

$$\mathcal{G}(S) \stackrel{\text{def.}}{=} \left\{ \tilde{\gamma}'(x) = \sum_{i=0}^{p-2} \frac{a_i \tau(\lambda)}{\lambda^{i+1}} x^i \mid a_i \in \mathbb{Z} \text{ and } |a_i| \leq 2\bar{C} \lambda / \tau(\lambda) \right\}, \quad (42)$$

where the constant  $\bar{C}$  is set so that so that the following bound on polynomial expansion holds

$$\forall x \in [0, 1], \quad |a_0 + a_1 x + \dots + a_{\alpha-1} x^{\alpha-1}| \leq 1 \implies \forall i, \quad |a_i| \leq \bar{C}. \quad (43)$$

**Construction of the dictionary.** The dictionary  $\mathcal{D}_j$  of bandletization bases at a given scale  $2^j$  is composed of the bases  $\mathcal{B}(\Gamma_j)$ , for all possible dyadic segmentations  $\mathcal{S}_j$  and geometric flow directions  $\tilde{\gamma}'_S \in \mathcal{G}(S)$  inside the squares  $S$  of the segmentation.

$$\mathcal{D}_j \stackrel{\text{def.}}{=} \{ \mathcal{B}(\Gamma_j) \mid \Gamma_j = (\mathcal{S}_j, \{ \tilde{\gamma}'_S \}) \text{ and } \forall S \in \mathcal{S}_j, \tilde{\gamma}'_S \in \mathcal{G}(S) \}.$$

Let  $T > 0$  be the non-linear approximation threshold. A finite bandlet dictionary  $\mathcal{D}_{T^2}$  of bases of  $L^2([0, 1]^2)$  is constructed by using bandletization bases for the first wavelet scales  $2^j \geq T^2$  and using the wavelet basis functions for the remaining scales.

$$\mathcal{D}_{T^2} \stackrel{\text{def.}}{=} \{ \mathcal{B}(\Gamma) \mid \forall 2^j \geq T^2, \mathcal{B}(\Gamma_j) \in \mathcal{D}_j \}.$$

## VI.2 Best Basis Approximation

A Lagrangian minimization computes a bandlet basis  $\mathcal{B}(\Gamma^*)$  whose segmentation  $\{\mathcal{S}_j\}_j$  and geometric flows  $\{\tilde{\gamma}'_S\}_{S \in \mathcal{S}_j}$  are adapted to  $f$ .

**Number of coefficients.** Let  $T > 0$  be some approximation threshold and

$$\mathcal{B}(\Gamma) = \{b_\nu\}_\nu \in \mathcal{D}_{T^2}, \quad \text{where} \quad \Gamma = \bigcup_j \Gamma_j, \quad (44)$$

be a bandlet basis, where  $\Gamma_j = (\mathcal{S}_j, \{ \tilde{\gamma}'_S \}_S)$  are the parameters that describe the basis at each scale. The thresholded approximation at  $T$  in this basis is defined by equation (38). The number of parameters  $M$  needed to describe  $f_M$  is decomposed as

$$M = M_B + M_{\mathcal{D}} \quad \text{where} \quad M_{\mathcal{D}} = M_G + M_S = \sum_{j,k} M_{Gj} + M_{Sj}. \quad (45)$$

The number of bandlets coefficients  $M_B$  is defined by equation (39) and for each scale  $2^j$ ,

- $M_{S_j}$  is the number of parameters needed to specify the segmentation  $\mathcal{S}_j$ . The dyadic segmentation is described using a quadtree structure, as shown on figure 7. The quadtree is coded using one coefficient per node to specify whether it is an interior node or if it corresponds to a square  $S$  that is either horizontal edge  $S \in \mathcal{E}^H(\mathcal{S}_j)$ , vertical edge  $S \in \mathcal{E}^V(\mathcal{S}_j)$  or regular  $S \in \mathcal{R}(\mathcal{S}_j)$ . Corner squares  $S \in \mathcal{C}(\mathcal{S}_j)$  are treated as regular squares since no bandletization is performed on the wavelets coefficients of these squares. One has

$$M_{S_j} \leq \text{Card}(\mathcal{S}_j). \quad (46)$$

- $M_{G_j}$  is the number of parameters needed to specify the geometric flows  $\tilde{\gamma}'_S$  over all the dyadic squares  $S \in \mathcal{S}_j$ . The adapted flow  $\tilde{\gamma}'_S$  is parameterized with  $p - 1$  polynomial coefficients. The number of geometric coefficients is thus

$$M_{G_j} \leq (p - 1) \text{Card}(\mathcal{S}_j). \quad (47)$$

**Lagrangian minimization.** A best basis is computed by minimizing the Lagrangian

$$\mathcal{L}(f, \mathcal{B}(\Gamma), T) \stackrel{\text{def.}}{=} \sum_{|\langle b_\nu, f \rangle| < T} |\langle b_\nu, f \rangle|^2 + T^2 M, \quad (48)$$

as previously done by Donoho in [7] and Le Pennec and Mallat in [10]. The best bandlet basis  $\mathcal{B}(\Gamma^*)$  adapted to  $f$  is defined by

$$\mathcal{B}(\Gamma^*) \stackrel{\text{def.}}{=} \underset{\mathcal{B}(\Gamma) \in \mathcal{D}_{T^2}}{\text{argmin}} \mathcal{L}(f, \mathcal{B}(\Gamma), T). \quad (49)$$

This best bandlet basis can be computed with a fast algorithm described in section VII. The following theorem gives the approximation rate of a geometrically regular function  $f$  in this adapted bandlet basis  $\mathcal{B}(\Gamma^*)$ .

**Theorem 1.** *Let  $f$  be a  $C^\alpha$ -geometrically regular function. There exists  $C$  such that for any  $T > 0$  the  $M$  parameters approximation  $f_M$  in the best bandlet basis  $\mathcal{B}(\Gamma^*)$  satisfies*

$$\|f - f_M\|_2^2 \leq C M^{-\alpha}. \quad (50)$$

This theorem states that the approximation of geometrically regular functions in a best bandlet basis recovers the asymptotic  $M^{-\alpha}$  decay of the approximation of uniformly regular functions in a wavelet basis. This asymptotic decay rate is thus optimal.

**Proof.** The proof relies on Lemma 1 that constructs an optimized bandlet basis within the dictionary which yields approximation error from  $M$  parameters that decays in  $O(M^{-\alpha})$ .

**Lemma 1.** *Let  $f$  be a  $C^\alpha$ -geometrically regular function of Definition 1. There exists  $C$  such that for all  $T > 0$  there exists a bandlet basis  $\mathcal{B}(\Gamma) \in \mathcal{D}_{T^2}$  in which the thresholded approximation  $f_M$  of  $f$  at  $T$  in this basis satisfies*

$$\|f - f_M\|_2^2 \leq C T^{\frac{2\alpha}{\alpha+1}} \quad \text{and} \quad M \leq C T^{-\frac{2}{\alpha+1}}. \quad (51)$$

The proof of Lemma 1 is in appendix A. Theorem 1 is derived by showing that the best basis that minimizes the Lagrangian is nearly as good as the optimized bandlet basis provided by lemma 1.

Lemma 1 provides an adapted bandlet basis  $\mathcal{B}(\Gamma) \in \mathcal{D}_{T^2}$  such that

$$\mathcal{L}(f, \mathcal{B}(\Gamma^*), T) \leq \mathcal{L}(f, \mathcal{B}(\Gamma), T) \leq CT^{\frac{2\alpha}{\alpha+1}},$$

$$\text{so} \quad \|f - f_M\|_2^2 \leq \mathcal{L}(f, \mathcal{B}(\Gamma^*), T) \leq CT^{\frac{2\alpha}{\alpha+1}} \quad (52)$$

$$\text{and} \quad MT^2 \leq \mathcal{L}(f, \mathcal{B}(\Gamma^*), T) \leq CT^{\frac{2\alpha}{\alpha+1}}. \quad (53)$$

Combining these equations proved equation (50) in theorem 1.

□

## VII Fast Bandlet Approximation

This section describes the fast transform of a discretized image in a best bandlet basis. A Matlab implementation of this transform is available [14].

A discretized image  $\bar{f}$  of  $N \times N$  pixels is obtained by projecting a function  $f \in L^2([0, 1]^2)$  onto a set of orthogonal scaling functions  $\{\varphi_{Jn}\}_n$  at a resolution  $2^J = N^{-1}$

$$\forall n \in \{0, \dots, N-1\}^2, \quad \bar{f}[n] \stackrel{\text{def.}}{=} \langle f, \varphi_{Jn} \rangle \quad \text{where} \quad \varphi_{Jn}(x) \stackrel{\text{def.}}{=} 2^{-J} \varphi(2^{-J}x - n).$$

In the following,  $\varphi$  is assumed to be the scaling function associated with the 2D wavelet functions used for the construction of the bandlet bases. The general case of compactly supported scaling functions is treated in [10].

The forward fast discrete bandlet transform decomposes the discrete image  $\bar{f}$  in a best bandlet basis of  $\mathbb{R}^{N \times N}$ , which is equivalent to the decomposition of the underlying function  $f$  on a best bandlet basis of  $L^2([0, 1]^2)$ .

**(1) 2D wavelet transform.** The wavelet coefficients  $\{f_j[n]\}_{j,n}$  are computed using a discrete wavelet transform of the image  $\bar{f}$ . The complexity of the fast wavelet transform is  $O(N^2)$  [11].

**(2) Fast Alpert transform.** For each scale  $2^j \geq T^2$ , for each dyadic square  $S$  of length  $\lambda$  larger than  $2^j$ , for each geometric flow  $\tilde{\gamma}' \in \mathcal{G}(S)$ , let  $\mathcal{B}(S, \tilde{\gamma}') = \{a_{\ell,m}\}_{\ell,m}$ . The Alpert coefficients  $\langle f_j, a_{\ell,m} \rangle$  of the wavelet coefficients  $f_j$  of  $f$  inside  $S$  are computed using the fast Alpert transform described in section B. The complexity of the Alpert transform for each square is  $O(m^2)$  where  $m^2 = (2^{-j}\lambda)^2$  is the number of coefficients in the square  $S$ .

**(3) Lagrangian minimization.** For each dyadic square  $S$  the best adapted flow  $\tilde{\gamma}'_S$  is computed by minimizing the Lagrangian

$$\tilde{\gamma}'_S = \operatorname{argmin}_{\tilde{\gamma}' \in \mathcal{G}(S)} \mathcal{L}(f_j, \mathcal{B}(S, \tilde{\gamma}'), T) \quad \text{and} \quad \mathcal{L}(S) \stackrel{\text{def.}}{=} \mathcal{L}(f_j, \mathcal{B}(S, \tilde{\gamma}'_S), T).$$

The Lagrangian of the untransformed coefficients  $f_j$  is also included in this minimization in order to take into account regular squares without geometry. For each scale  $2^j$ , the dyadic segmentation  $\mathcal{S}_j$  is computed using a fast bottom-up algorithm similar to the CART regression procedure [2]:

- Initialization of the segmentation: each square  $S$  of width  $\lambda = 2^j$  is a leaf of the quadtree representing  $\mathcal{S}_j$ . The corresponding optimal geometry  $\tilde{\gamma}'_S$  is recorded and the cumulative Lagrangian  $\mathcal{L}_0$  of the sub-tree is initialized to  $\mathcal{L}_0(S) = \mathcal{L}(S)$ . The width  $\lambda$  of the squares is initialized to  $\lambda = 2 \times 2^j$ .
- For each square  $S$ , let  $(S_1, S_2, S_3, S_4)$  be its 4 sub-squares. Observe that

$$\mathcal{L}'(S) \stackrel{\text{def.}}{=} \mathcal{L}_0(S_1) + \mathcal{L}_0(S_2) + \mathcal{L}_0(S_3) + \mathcal{L}_0(S_4) + T^2$$

is the Lagrangian of the sub-tree (the additional  $T^2$  is due to the split cost of one coefficient). The sub-squares are merged if  $\mathcal{L}(S) < \mathcal{L}'(S)$ . If so,  $S$  is declared as a leaf and the optimal geometry  $\tilde{\gamma}'_S$  is recorded. The lagrangian is updated using  $\mathcal{L}_0(S) = \min(\mathcal{L}(S), \mathcal{L}'(S))$ .

- While  $\lambda < 1$ , the current length is updated using  $\lambda \leftarrow 2 \times \lambda$  and the previous step is repeated.

**Numerical Complexity** The lagrangian minimization requires to compute the Alpert transform over each square  $S$  of dyadic length for each geometry of  $\mathcal{G}(S)$ . The bottom-up procedure that builds the quadtree has a negligible complexity. For each scale  $2^j$ , the complexity of computing the Alpert transform over the set of all squares of length  $\lambda$  and all geometries is

$$\underbrace{(1/\lambda)^2}_{\text{nbr.squares}} \times \underbrace{C_{\mathcal{A}}(2^{-j} \lambda)^2}_{\text{complexity Alpert tr.}} \times \underbrace{C_{\mathcal{G}} \lambda^{-(p-1)^2}}_{\text{nbr.geometries}} = C' 2^{-2j} T^{-2(p-1)^2},$$

since equation (42) shows that the cardinal of  $\mathcal{G}(S)$  is proportional to  $\lambda^{-(p-1)^2}$  and that  $\lambda \geq 2^j \geq T^2$ . As the number of such widths  $\lambda$  and scales  $2^j$  is proportional to  $|\log_2(T)|$ , the overall complexity of the bandlet transform is  $O(N^2 T^{-2(p-1)^2})$ .

## VIII Best Bandlet Basis Compression

An image  $f$  is compressed in a bandlet basis  $\mathcal{B}(\Gamma) = \{b_\nu\}_\nu \in \mathcal{D}_{T^2}$  by quantizing and coding its transformed coefficients and by coding the geometric parameters  $\Gamma = \{\Gamma_j\}_j$  that describe the basis, where  $\Gamma_j = (\mathcal{S}_j, \{\tilde{\gamma}'_S\}_{S \in \mathcal{S}_j})$ . The restored image from the compressed code is

$$f_R \stackrel{\text{def.}}{=} \sum_{\nu} Q_T(\langle f, b_\nu \rangle) b_\nu, \quad (54)$$

where  $Q_T$  is a uniform quantizer defined by

$$Q_T(x) = qT, \quad \text{if} \quad (q - 1/2)T \leq x \leq (q + 1/2)T. \quad (55)$$

The coding distortion is

$$D(R) \stackrel{\text{def.}}{=} \|f - f_R\|_2^2.$$

Since  $|x - Q_T(x)| \leq T/2$  and  $Q_T(x) = 0$  if  $|x| \leq T/2$ , one has

$$D(R) = \|f - f_R\|_2^2 = \sum_{\nu} |\langle f, b_{\nu} \rangle - Q_T(\langle f, b_{\nu} \rangle)|^2 \quad (56)$$

$$\leq \sum_{|\langle f, b_{\nu} \rangle| < T/2} |\langle f, b_{\nu} \rangle|^2 + \frac{1}{4} M_B T^2 \quad (57)$$

$$\leq \|f - f_M\|_2^2 + \frac{1}{4} M_B T^2, \quad (58)$$

which links the distortion  $D(R)$  with the non-linear approximation  $f_M$  obtained with a thresholding at  $T/2$  as defined in (38). The number of coefficients  $M = M_B + M_S + M_G$  is computed following section VI.2.

The bit budget of this transformed code is

$$R \stackrel{\text{def.}}{=} R_B + R_S + R_G = \sum_j (R_{Bj} + R_{Sj} + R_{Gj})$$

where

- $R_{Bj}$  is the number of bits needed to code the bandlet coefficients  $\langle f, b_{\nu} \rangle = \langle f_j, a_{\nu} \rangle$  for a single scale  $2^j$ . Since there are  $2^{-2j}$  bandlets coefficients at a scale  $2^j$ , the index of each of the  $M_{Bj}$  non-zero quantized coefficients is coded using  $\log_2(2^{-2j})$  bits per coefficient. For a bounded image  $f$ , one has

$$|\langle f, b_{\nu} \rangle| \leq \|f_j\|_{\infty} \|a_{\nu}\|_1 \leq 2^j \|f\|_{\infty} \|\psi\|_1 C_b \leq \|f\|_{\infty} \|\psi\|_1 C_b,$$

so there exists a constant  $C_1$  such that the quantized amplitudes  $Q_T(\langle f, b_{\nu} \rangle)$  are coded using less than  $\log_2(C_1/T)$  bits per coefficient. The number of bits to code the bandlet coefficients is thus bounded by

$$R_{Bj} \leq M_{Bj} (\log_2(2^{-2j}) + \log_2(C_1/T)).$$

The usual scale restriction  $2^j > T^2$  implies that  $R_{Bj} \leq C M_{Bj} |\log_2(T)|$ .

- $R_{Sj}$  the number of bits needed to code the quadtree segmentation  $\mathcal{S}_j$  for a single scale  $2^j$ . One needs to differentiate between interior, vertical edge and horizontal edge nodes so 2 bits per segmentation coefficient is needed, and thus  $R_{Sj} = 2M_{Sj}$  where  $M_{Sj}$  is the number of coefficients needed to specify the dyadic segmentation  $\mathcal{S}_j$ , as described in section VI.2.
- $R_{Gj}$  the number of bits needed to code the adapted geometric flow  $\tilde{\gamma}'_S \in \mathcal{G}(S)$  inside each square  $S$  of each quadtree  $\mathcal{S}_j$ . Each geometric coefficient  $M_{Gj}$  is quantized and equation (42) shows that there are  $\text{Card}(\mathcal{G}(S)) = C_G \lambda^{-(1-p)^2}$  possible quantized geometries where  $C_G$  is a constant. As  $\lambda \geq 2^j$ , the condition  $2^j \geq T^2$  implies

$$R_{Gj} \leq M_{Gj} \log_2(C_G \lambda^{-(1-p)^2}) \leq C M_{Gj} |\log_2(T)|.$$

Using this coding scheme, the total bit budget is thus

$$R = \sum_{2^j \geq T^2} (R_{Bj} + R_{Sj} + R_{Gj}) \leq C M |\log(T)|,$$

where  $M = \sum_j (M_{Bj} + M_{Sj} + M_{Gj})$  is the total number of coefficients needed to specify  $f_M$  as described in section VI.2.

In order to minimize  $D(R)$ , equation (58) shows that one should use the Lagrangian minimization of (49) with a Lagrange multiplier equal to  $T/2$ . An adaptive compression of the image is thus performed by using the best bandlet basis defined by

$$\mathcal{B}(\Gamma^*) \stackrel{\text{def.}}{=} \underset{\mathcal{B}(\Gamma) \in \mathcal{D}_{T^2}}{\text{argmin}} \mathcal{L}(f, \mathcal{B}(\Gamma), T/2).$$

One needs to compute the distortion  $D(R)$  in this basis and link this distortion with the number of bits  $R$ .

The bounds of equation (52) shows that the thresholding approximation  $f_M$  at  $T/2$  in the bandlet basis  $\mathcal{B}(\Gamma^*)$  satisfies

$$\|f - f_M\|_2^2 \leq C (T/2)^{\frac{2\alpha}{\alpha+1}}, \quad \text{with} \quad M \leq C (T/2)^{-\frac{2}{\alpha+1}}. \quad (59)$$

The scheme used to code the bandlet coefficients and the geometric parameters ensures that  $R \leq C M |\log(T)|$ . Combining this result with the bounds of equations (58) and (59) proves the following theorem.

**Theorem 2.** *Let  $f$  be a  $C^\alpha$ -geometrically regular function. There exists  $C > 0$  such that for any  $T > 0$ , the compressed image  $f_R$  in the best bandlet basis  $\mathcal{B}(\Gamma^*)$  satisfies*

$$\|f - f_R\|_2^2 \leq C \log_2(R)^\alpha R^{-\alpha}.$$

The class of geometrically regular functions contains the class of uniformly  $C^\alpha$  functions, for which the Kolmogorov bound decreases like  $R^{-\alpha}$ , see [8]. This theorem thus proves that the asymptotic coding error decay in an adapted bandlet basis reaches the Kolmogorov lower bound for geometrically regular functions up to a  $|\log(R)|^\alpha$  factor.

## A Proof of lemma 1

**Proof.** A bandlet basis  $\mathcal{B}(\Gamma) \in \mathcal{D}_{T^2}$  is built by choosing, for each scale  $2^j$ , a discrete Alpert basis  $\mathcal{B}(\Gamma_j)$ . At each scale  $2^j$  the thresholded approximation  $f_{jM_j}$  of the wavelets coefficients  $f_j$  at  $T$  in  $\mathcal{B}(\Gamma_j) = \{a_\nu\}_\nu$  is defined by

$$f_{jM_j} = \sum_{|\langle f_j, a_\nu \rangle| \geq T} \langle f_j, a_\nu \rangle a_\nu.$$

The number of parameters is defined following equation (45) by  $M_j = M_{Gj} + M_{Sj} + M_{Bj}$ . Using the fact that  $\langle f, a_\nu \rangle = \langle f_j, b_{\ell,m} \rangle$  for some index  $(\ell, m)$ , one can decompose the approximation error as

$$\|f - f_M\|_2^2 = \sum_j \|f_j - f_{jM_j}\|_2^2 \quad \text{where} \quad M \stackrel{\text{def.}}{=} \sum_j M_j.$$



In order to build a bandlet basis adapted to the function  $f$ , one has to consider three approximation modes depending on the scale.

**For fine scales:**  $2^j < 2^{j_0} \stackrel{\text{def.}}{=} T^{\frac{2\alpha}{\alpha+1}}$ , a standard result, already used to prove equation (7), states that  $\|f_j\|_2^2 \leq C 2^j$  for a constant  $C$  that depends only on  $f$ . As  $|f_j[n]| \leq \|f\|_\infty \|\psi\|_1 2^j$ , one has, for  $T$  small enough,  $f_j[n] = 0$ , so  $M_j = 0$  and

$$\sum_{j < j_0} \|f_j - f_{jM_j}\|_2^2 \leq \sum_{j < j_0} C 2^j \leq 2C 2^{j_0} = 2C T^{\frac{2\alpha}{\alpha+1}}, \quad \text{with} \quad \sum_{j < j_0} M_j = 0.$$

**For coarse scales:**  $2^j > 2^{j_1} \stackrel{\text{def.}}{=} T^{\frac{1}{\alpha+1}}$ , there is less than  $2^{-2j_1}$  coefficients so

$$\sum_{j > j_1} \|f_j - f_{jM_j}\|_2^2 \leq 2^{-2j_1} T^2 \leq T^{\frac{2\alpha}{\alpha+1}}.$$

Since there is no segmentation and no flow  $M_{S_j} = M_{G_j} = 0$  and hence

$$\sum_{j > j_1} M_j \leq 2^{-2j_1} \leq T^{\frac{-2}{\alpha+1}}.$$

**For intermediate scales:**  $2^{j_0} \leq 2^j \leq 2^{j_1}$ , a bandletization basis  $\mathcal{B}(\Gamma_j)$  adapted to  $f$  is used to approximate the coefficients. Such a basis is provided by Lemma 3. This lemma proves that a thresholding in this basis gives the following error

$$\begin{aligned} \sum_{j=j_0}^{j_1} \|f_j - f_{jM_j}\|_2^2 &\leq \sum_{j=j_0}^{j_1} C \left( s_j^{\frac{\alpha-p}{\delta}} T^{\frac{4\alpha p}{\delta}} + 2^{2j\alpha} \right) \quad \text{where} \quad \delta \stackrel{\text{def.}}{=} 2\alpha p + \alpha + p \\ &\leq 2C \left( 2^{j_0 \frac{\alpha-p}{\delta}} T^{\frac{4\alpha p}{\delta}} + 2^{2j_1\alpha} \right) \leq 4C T^{\frac{2\alpha}{\alpha+1}}. \end{aligned}$$

Combining all these bounds proves (51) in lemma 1.  $\square$

**Adapted segmentation of wavelet coefficients.** In order to match the approximation precision given by  $T$ , the optimal width of the square in an adapted segmentation  $\mathcal{S}_j$  is set to

$$\lambda(T) \stackrel{\text{def.}}{=} C_0 T^{\frac{2p}{\delta}} s_j^{\frac{p+1}{\delta}} \quad \text{and} \quad \delta \stackrel{\text{def.}}{=} 2\alpha p + \alpha + p, \quad (60)$$

where  $C_0$  is a constant that depends only on  $f$  and whose value is tuned during the proof of Lemma 4. This definition of the optimal width is similar to the one given in equation (25) but it is parameterized by the threshold value  $T$ .

Section V.2 has introduced the partition of squares  $S \in \mathcal{S}_j$  into edge squares  $S \in \mathcal{E}(\mathcal{S}_j)$ , corner squares  $S \in \mathcal{C}(\mathcal{S}_j)$  and regular squares  $S \in \mathcal{R}(\mathcal{S}_j)$ . Only edges squares  $S \in \mathcal{E}$  contain an adapted flow  $\tilde{\gamma}$ , since in the other squares the original coefficients are not modified. Ideally, one would like to have only regular squares and squares of width  $\lambda$ . But topological constraints (such as edges crossings and corners) and curvature variations (horizontal edges that become vertical) forces to subdivide some squares of width  $\lambda$  into smaller ones. The set of edge squares is thus subdivided as  $\mathcal{E} = \mathcal{E}_\lambda \cup \tilde{\mathcal{E}}_\lambda$  where  $\mathcal{E}_\lambda$  is the set of squares length  $\lambda(T)$  and  $\tilde{\mathcal{E}}_\lambda$  contains squares of smaller lengths. The following lemma shows that one can build a dyadic segmentation with a small number of these subdivided squares  $S \in \tilde{\mathcal{E}}_\lambda$ .

**Lemma 2.** *Let  $f$  be a geometrically regular function. There exists a constant  $C$  such that for all  $\lambda > 0$ , there exists a dyadic segmentation  $\mathcal{S}_j$  of  $[0, 1]^2$  into squares of width larger than  $2^j$ , that has the following properties:*

$$\begin{aligned} \text{Card}(\mathcal{S}_j) &\leq C (\lambda^{-1} + |\log_2(\lambda/2^j)|), & \text{Card}(\mathcal{E}_\lambda(\mathcal{S}_j)) &\leq C \lambda^{-1}, \\ \text{Card}(\tilde{\mathcal{E}}_\lambda(\mathcal{S}_j)) &\leq C |\log_2(\lambda/2^j)|, & \text{Card}(\mathcal{C}(\mathcal{S}_j)) &\leq C. \end{aligned}$$

**Proof.** Following [10], a dyadic image segmentation is performed by iteratively labeling edge squares  $S \in \mathcal{E}$  and corner squares  $S \in \mathcal{C}$  while removing temporary unlabeled squares  $S \in \mathcal{T}$ . The algorithm proceeds as follow:

- Initialization: label the square  $S = [0, 1]^2$  as temporary  $S \in \mathcal{T}$ .
- Step 1: Split in four every temporary square  $S \in \mathcal{T}$  and remove  $S$  from  $\mathcal{T}$ .
- Step 2: Label, in the following order, each new subdivided square  $S$  as a:
  - *regular square*  $S \in \mathcal{R}(\mathcal{S}_j)$  if it is at a distance larger than  $s_j$  from all edges,
  - *corner square*  $S \in \mathcal{C}(\mathcal{S}_j)$  if its size is smaller than  $2^j$ ,
  - *horizontal edge square*  $S \in \mathcal{E}^H(\mathcal{S}_j)$  if a single horizontal edge component  $\gamma$  is closer than  $s_j$  from  $S$ . Following the edge square definition of section V.2, the curve is supposed to be parameterized horizontally with  $|\gamma'| \leq 2$ .
  - *vertical edge square*  $S \in \mathcal{E}^V(\mathcal{S}_j)$  if a single vertical edge component  $\gamma$  is closer than  $s_j$  from  $S$ ,
  - *temporary square*  $S \in \mathcal{T}$  otherwise.
- Step 3: Go to step 1 if there remains temporary squares.

Figure 10 illustrate this process.

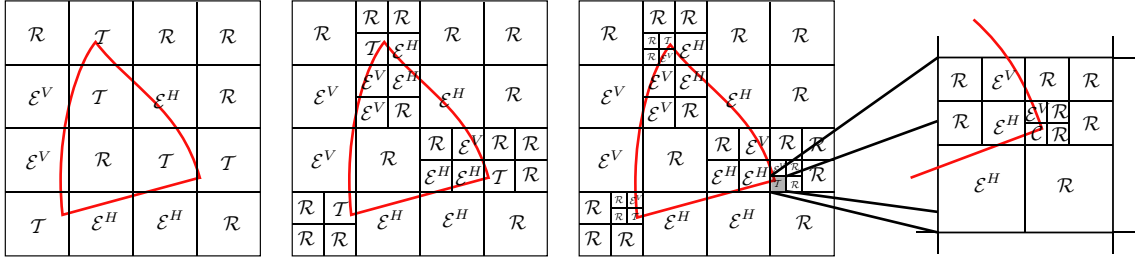


Figure 10: *Recursive subdivision of dyadic squares with the corresponding labeling  $S \in \{\mathcal{R}, \mathcal{E}^V, \mathcal{E}^H, \mathcal{C}\}$ .*

The first set of subdivision steps subdivides squares that contain edges until the square length reaches  $\lambda(T)$ . Due to the regularity of the curves, the number of resulting edge squares is of order  $O(\lambda^{-1})$ . The non-tangency condition of the curves also ensures that after a constant number of steps, only a constant number of squares near junctions are labeled as temporary.

In the neighborhood of the junctions, the recursive splitting continues during  $\log(\lambda/2^j)$  steps until the length of the squares is  $2^j$ . As there is only a finite number of such junctions, the number of the small edge squares is of order  $O(\log(\lambda/2^j))$ .  $\square$

The following lemma constructs an adapted bandletization basis for a given scale  $2^j$ . It uses the construction of an adapted dyadic segmentation together with an adapted quantized geometric flow that closely matches the real geometry.

**Lemma 3.** *Let  $f$  be a  $C^\alpha$ -geometrically regular function. There exists  $C$  such that for any  $T > 0$ , if  $2^j$  satisfies*

$$T^{\frac{2\alpha}{\alpha+1}} \stackrel{\text{def.}}{=} 2^{j_0} \leq 2^j \leq 2^{j_1} \stackrel{\text{def.}}{=} T^{\frac{1}{\alpha+1}},$$

*then there exists a bandletization basis  $\mathcal{B}(\Gamma_j) \in \mathcal{D}_j$ , such that the thresholding approximation  $f_{jM_j}$  at  $T$  of  $f_j$  in this basis satisfies*

$$\begin{cases} \|f_j - f_{jM_j}\|_2^2 \leq C \max\left(s_j^{\frac{\alpha-p}{\delta}} T^{\frac{4\alpha p}{\delta}}, 2^{2j\alpha}\right), \\ M_j \leq C s_j^{\frac{\alpha-p}{\delta}} T^{-2\frac{\alpha+p}{\delta}}, \quad \text{with} \quad \delta \stackrel{\text{def.}}{=} 2\alpha p + \alpha + p. \end{cases} \quad (61)$$

**Proof.** The optimal width  $\lambda = \lambda(T)$  of the edge squares is defined in equation (60). Lemma 2 provides a dyadic segmentation of  $[0, 1]^2$  with edge squares conforming as much as possible to this optimal length  $\lambda$ . A bandletization basis  $\mathcal{B}(\Gamma_j)$  is defined by choosing the bandlet basis  $\mathcal{B}(S, \tilde{\gamma}'_S)$  provided by lemma 4 over edge squares  $S \in \mathcal{E}_\lambda(\mathcal{S}_j) \cup \tilde{\mathcal{E}}_\lambda(\mathcal{S}_j)$  and by keeping the original wavelet coefficients over the remaining squares. Wavelet coefficients inside a square  $S$  are denoted  $f_j[n]$  and the corresponding thresholded approximation with  $M(S)$  coefficients is denoted  $f_{jM(S)}$ . One has the following error

$$\begin{aligned} \|f_j - f_{jM_j}\|_2^2 &= \sum_{S \in \mathcal{R}(\mathcal{S}_j)} \|f_j - f_{jM(S)}\|_{\ell^2(S)}^2 + \sum_{S \in \mathcal{C}(\mathcal{S}_j)} \|f_j - f_{jM(S)}\|_{\ell^2(S)}^2 \\ &\quad + \sum_{S \in \mathcal{E}_\lambda(\mathcal{S}_j)} \|f_j - f_{jM(S)}\|_{\ell^2(S)}^2 + \sum_{S \in \tilde{\mathcal{E}}_\lambda(\mathcal{S}_j)} \|f_j - f_{jM(S)}\|_{\ell^2(S)}^2 \end{aligned}$$

with

$$M_{B_j} = \sum_{S \in \mathcal{R}(\mathcal{S}_j)} M(S) + \sum_{S \in \mathcal{C}(\mathcal{S}_j)} M(S) + \sum_{S \in \mathcal{E}_\lambda(\mathcal{S}_j)} M(S) + \sum_{S \in \tilde{\mathcal{E}}_\lambda(\mathcal{S}_j)} M(S)$$

In regular squares  $S \in \mathcal{R}(\mathcal{S}_j)$ , we saw in equation (3) that

$$\forall (2^j n) \in S, \quad |\langle f, \psi_{jn} \rangle| \leq C_f 2^{j(1+\alpha)},$$

where  $C_f$  is proportional to  $\|f\|_{C^\alpha(\Lambda)}$ . Let  $2^{j_\star}$  be the cutoff scale defined by

$$2^{j_\star} \stackrel{\text{def.}}{=} (C_f)^{-\frac{1}{\alpha+1}} T^{\frac{1}{\alpha+1}}$$

If  $j < j_\star$ , one has  $|f_j[n]| < T$ , thus  $\sum_{S \in \mathcal{R}(\mathcal{S}_j)} M(S) = 0$ . If  $j \geq j_\star$ , one has

$$\sum_{S \in \mathcal{R}(\mathcal{S}_j)} M(S) \leq 2^{-2j_\star} \leq (C_f)^{\frac{2}{\alpha+1}} T^{-\frac{2}{\alpha+1}} \leq (C_f)^{\frac{2}{\alpha+1}} s_j^{\frac{\alpha-p}{\delta}} T^{-2\frac{\alpha+p}{\delta}}.$$

In both cases one has

$$\sum_{S \in \mathcal{R}(\mathcal{S}_j)} \|f_j - f_{jM(S)}\|_{\ell^2(S)}^2 \leq 2^{-2j} \left( \max_{S \in \mathcal{R}(\mathcal{S}_j)} |f_j[n]|^2 \right) \leq (C_f)^2 2^{2j\alpha}.$$

**In corner squares**  $S \in \mathcal{C}(\mathcal{S}_j)$ , there is a constant number  $C$  of coefficients and thus

$$\sum_{S \in \mathcal{C}(\mathcal{S}_j)} \|f_{jM(S)} - f_j\|_{\ell^2(S)}^2 \leq C T^2 \leq C s_j^{\frac{\alpha-p}{\delta}} T^{\frac{4\alpha p}{\delta}} \quad \text{and} \quad \sum_{S \in \mathcal{C}(\mathcal{S}_j)} M(S) \leq C,$$

which satisfies bounds (61).

**In edge squares**  $S \in \mathcal{E}_\lambda(\mathcal{S}_j)$  of size  $\lambda$ , lemma 4 bounds the bandlet approximation error and the number of coefficients needed. Since there are less than  $C \lambda^{-1}$  such squares, one has

$$\sum_{S \in \mathcal{E}_\lambda(\mathcal{S}_j)} \|f_j - f_{jM(S)}\|_{\ell^2(S)}^2 \leq (C \lambda^{-1}) \lambda s_j^{\frac{\alpha-p}{\delta}} T^{\frac{4\alpha p}{\delta}} \quad (62)$$

together with

$$\sum_{S \in \mathcal{E}_\lambda(\mathcal{S}_j)} M(S) \leq (C \lambda^{-1}) \lambda s_j^{\frac{\alpha-p}{\delta}} T^{-2\frac{\alpha+p}{\delta}}, \quad (63)$$

and since

$$\lambda(T) \stackrel{\text{def.}}{=} C_0 T^{\frac{2p}{\delta}} s_j^{\frac{p+1}{\delta}} \quad \text{and} \quad \delta \stackrel{\text{def.}}{=} 2\alpha p + \alpha + p, \quad (64)$$

one get (61) by inserting (64) into (62) and (63).

**In small edges squares**  $S \in \tilde{\mathcal{E}}(\mathcal{S}_j)$  of size less than  $\lambda$ , lemma 4 still applies. The number of bandlets coefficients  $M(S)$  is bounded by the number of coefficients needed for squares of optimal size  $\lambda$ . Since there is less than  $|\log_2(\lambda/2^j)| \leq C_1 |\log_2(T)|$  such squares, bounds (61) still holds.

**Geometric coefficients.** By combining these bounds together one get the error bound of equation (61). Equations (46) and (47) show that

$$\begin{aligned} M_{S_j} + M_{G_j} &\leq p \text{Card}(\mathcal{S}_j) \leq C \lambda^{-1} \leq C C_0^{-1} T^{-2\frac{p}{\delta}} s_j^{-\frac{p+1}{\delta}} \\ &\leq C C_0^{-1} s_j^{\frac{\alpha-p}{\delta}} T^{-2\frac{\alpha+p}{\delta}} \underbrace{\left( T^{-2\frac{\alpha}{\delta}} s_j^{-\frac{1+\alpha}{\delta}} \right)}_{\leq 1 \text{ since } s_j \geq T^{\frac{2\alpha}{\alpha+1}}} \leq (C C_0^{-1}) s_j^{\frac{\alpha-p}{\delta}} T^{-2\frac{\alpha+p}{\delta}}, \end{aligned}$$

which gives the bound of equation (61) for the number of coefficients.  $\square$

The technical analysis of the bandlet approximation in an edge square  $S$  is detailed in Lemma 4. In this Lemma,  $f_j$  is the set of wavelet coefficients inside  $S$ ,  $f_{jM(S)}$  is the thresholded approximation of  $f_j$  at  $T$  in basis  $\mathcal{B}(S, \tilde{\gamma}')$  and  $M(S)$  is the number of needed bandlet coefficients.

**Lemma 4.** *Let  $f$  be a  $C^\alpha$ -geometrically regular function. There exists  $C > 0$  such that for any  $T > 0$  and  $2^j$  satisfying*

$$T^{\frac{2\alpha}{\alpha+1}} \stackrel{\text{def.}}{=} 2^{j_0} \leq 2^j \leq 2^{j_1} \stackrel{\text{def.}}{=} T^{\frac{1}{\alpha+1}}, \quad (65)$$

for any edge square  $S$  of width  $\lambda \stackrel{\text{def.}}{=} \lambda(T)$  there exists an adapted geometric flow  $\tilde{\gamma}' \in \mathcal{G}(S)$  such that

$$\begin{cases} \|f_j - f_{jM(S)}\|_{\ell^2(S)}^2 \leq C \lambda s_j^{\frac{\alpha-p}{\delta}} T^{\frac{4\alpha p}{\delta}} \\ M(S) \leq C \lambda s_j^{\frac{\alpha-p}{\delta}} T^{-2\frac{\alpha+p}{\delta}}, \end{cases} \quad \text{with} \quad \delta \stackrel{\text{def.}}{=} 2\alpha p + \alpha + p. \quad (66)$$

**Proof.** The edge curve is parameterized in  $S$  by  $x_2 = \gamma(x_1)$ . Let  $\tilde{\gamma}'_0$  be a Taylor polynomial expansion of degree  $\alpha - 2$  of  $\gamma'$  inside  $S$ . An adapted polynomial flow  $\tilde{\gamma}' \in \mathcal{G}(S)$  quantized with a precision  $\tau = \lambda^\alpha / (p - 1)$  is defined by

$$\tilde{\gamma}'_0(x) = \sum_{i=0}^{p-2} \frac{\alpha_i}{\lambda^{i+1}} x^i \quad \text{and} \quad \tilde{\gamma}'(x) = \sum_{i=0}^{p-2} \frac{a_i \tau}{\lambda^{i+1}} x^i \quad \text{with} \quad a_i = Q_\tau(\alpha_i),$$

where the uniform quantizer  $Q_\tau$  is defined by

$$Q_\tau(x) = q\tau, \quad \text{if} \quad (q - 1/2)\tau \leq x \leq (q + 1/2)\tau.$$

Since the edge square definition of equation (36) enforces that  $|\gamma'| \leq 2$ , one has

$$\forall x \in [0, 1], \quad \left| \sum_{i=0}^{p-2} (\alpha_i / \lambda) x^i \right| \leq 2 \quad \implies \quad |a_i| \leq 2\bar{C}\lambda / \tau,$$

where the constant  $\bar{C}$  is defined by equation (43) and thus  $\tilde{\gamma}' \in \mathcal{G}(S)$ . Furthermore, this adapted flow satisfies

$$\|\tilde{\gamma}' - \gamma'\|_\infty \leq \|\tilde{\gamma}' - \tilde{\gamma}'_0\|_\infty + \|\tilde{\gamma}'_0 - \gamma'\|_\infty \quad (67)$$

$$\leq (p - 1)\tau / \lambda + \|\gamma\|_{C^\alpha} \lambda^{\alpha-1} \leq (1 + \|\gamma\|_{C^\alpha}) \lambda^{\alpha-1}. \quad (68)$$

The set of points  $x$  such that  $|\gamma(x_1) - x_2| \leq K s_j$  is denoted  $\mathcal{T}$ . The support of an Alpert vector is  $\beta_{\ell,m}$ .

**Bounding bandlets coefficients.** There are two kinds of Alpert vectors  $a_{\ell,m} \in \mathcal{B}(S, \tilde{\gamma}')$ :

– If  $\beta_{\ell,m}$  does not intersect  $\mathcal{T}$ , the fact that  $f$  is  $C^\alpha$ -regular away from edges implies

$$\forall (2^j n) \notin \mathcal{T}, \quad |f_j[n]| = |\langle f, \psi_{jn} \rangle| \leq C_f 2^{j(1+\alpha)}, \quad (69)$$

where  $C_f$  is proportional to  $\|f\|_{C^\alpha(\Lambda)}$ , which leads to

$$|\langle f_j, a_{\ell,m} \rangle| \leq C_f \|f_j\|_\infty \|a_{\ell,m}\|_1 \leq (C_f C_a) 2^{j\alpha} 2^{\ell/2} \lambda$$

where  $\|a_{\ell,m}\|_1$  is given by equation (31). A bandlet cutoff scale  $2^{L_0}$  is defined by

$$(C_f C_a) 2^{j\alpha} 2^{L_0/2} \lambda = T \quad \implies \quad 2^{L_0} \stackrel{\text{def.}}{=} (C_f C_a)^{-2} 2^{-2j\alpha} T^2 \lambda^{-2},$$

hence

$$2^\ell \leq 2^{L_0} \quad \implies \quad |\langle f, a_{\ell,m} \rangle| \leq T. \quad (70)$$

– If  $\beta_{\ell,m}$  intersects  $\mathcal{T}$ , one needs to use the regularity of the warped function  $f_{jw}$ . Let  $P(x)$  be a Taylor polynomial expansion of degree  $p$  of the function  $f_{jw}$  inside  $w(\beta_{\ell,m})$ , where  $w$  is the warping operator defined by equation (10). The definition (60) of  $\lambda(\tau)$  together with the bound  $s_j \geq 2^j \geq T^{\frac{2\alpha}{\alpha+1}}$  implies that  $\lambda(\tau) \leq C_0 s_j^{1/\alpha}$ . Equation (68) then allows to apply Proposition 1 which provides bounds on the derivatives of  $f_{jw}$ . The construction of the band detailed in section V.1 enforces

that  $w(\beta_{\ell,m})$  is of size  $\lambda \times \mu$  where  $\mu \stackrel{\text{def.}}{=} \lambda 2^\ell$ . Similarly to equation (23), the error is bounded by

$$\begin{aligned} \forall \tilde{x} \in w(\beta_{\ell,m}), \quad |f_{jw}(\tilde{x}) - P(\tilde{x})| &\leq \sum_{(i_1, i_2) \in I_\alpha^p} \left\| \frac{\partial^\alpha f_{jw}}{\partial x_1^{i_1} \partial x_2^{i_2}} \right\|_{L^\infty} \lambda^{i_1} \mu^{i_2} \\ &\leq C (1 + \|\gamma\|_{C^\alpha}^\alpha) \sum_{(i_1, i_2) \in I_\alpha^p} 2^j s_j^{-i_1/\alpha - i_2} \lambda^{i_1} \mu^{i_2} \\ &\leq C_w 2^j (s_j^{-1} \lambda^\alpha + s_j^{-p} (2^\ell \lambda)^p), \end{aligned}$$

where the constant  $C_w$  depends on  $\|\gamma\|_{C^\alpha}$  and  $\|f\|_{C^\alpha(\Lambda)}$  and where the set of indices is

$$I_\alpha^p \stackrel{\text{def.}}{=} \left\{ (i_1, i_2) \mid \begin{array}{l} i_1 + i_2 = p \quad \text{and} \quad i_1 < \alpha \\ \alpha + i_2 \leq p \quad \text{and} \quad i_1 = \alpha \end{array} \right\}.$$

It results that for all  $\tilde{x} = w(x) \in w(\beta_{\ell,m})$

$$f_{jw}(\tilde{x}) = P(\tilde{x}) + \varepsilon(\tilde{x}) \quad \text{with} \quad \|\varepsilon\|_\infty \leq C_w 2^j (s_j^{-1} \lambda^\alpha + s_j^{-p} (2^\ell \lambda)^p). \quad (71)$$

The bandlets inner products are bounded using the fact that  $a_{\ell,m}$  is orthogonal to the space of discrete warped polynomials of degree  $p-1$  on  $w(\beta_{\ell,m})$

$$\begin{aligned} \langle f, a_{\ell,m} \rangle &= \sum_{2^j n \in \beta_{\ell,m}} f_j[n] a_{\ell,m}[n] = \sum_{2^j n \in \beta_{\ell,m}} f_{jw}(w(2^j n)) a_{\ell,m}[n] \\ &= \sum_{2^j n \in \beta_{\ell,m}} \varepsilon(w(2^j n)) a_{\ell,m}[n], \end{aligned}$$

which leads to the following bound

$$|\langle f, a_{\ell,m} \rangle| \leq \|\varepsilon\|_\infty \|a_{\ell,m}\|_1 \leq (C_w C_a) \lambda^{p+1} s_j^{-p} 2^{\ell/2} \max(s_j^{p-1} \lambda^{\alpha-p}, 2^{\ell p}), \quad (72)$$

where  $\|a_{\ell,m}\|_1$  is bounded by equation (31). A bandlet cutoff scale  $2^{L_1}$  is defined by

$$s_j^{p-1} \lambda^{\alpha-p} = 2^{L_1 p} \implies 2^{L_1} \stackrel{\text{def.}}{=} s_j^{1-1/p} \lambda^{\alpha/p-1}.$$

The constant  $C_0$  that defines the optimal length  $\lambda(T)$  in equation (60) is set to

$$C_0 \stackrel{\text{def.}}{=} (C_w C_a)^{-\delta/(2p)}$$

which implies

$$2^\ell \leq 2^{L_1} \implies |\langle f, a_{\ell,m} \rangle| \leq (C_w C_a) \lambda^{\frac{\delta}{2p}} s_j^{-\frac{1+p}{2}} \leq T. \quad (73)$$

**Bounding  $M(S)$ .** The number of coefficients above the threshold

$$M(S) \stackrel{\text{def.}}{=} \text{Card}(J_T) \quad \text{where} \quad J_T \stackrel{\text{def.}}{=} \{(\ell, m) \mid |\langle f, a_{\ell,m} \rangle| \geq T\}$$

can be bounded using the sets

$$\begin{aligned} J_0 &\stackrel{\text{def.}}{=} \{(\ell, m) \mid \text{Supp}(a_{\ell,m}) \cap \mathcal{T} = \emptyset \quad \text{and} \quad 2^\ell \geq 2^{L_0}\} \\ \text{and} \quad J_1 &\stackrel{\text{def.}}{=} \{(\ell, m) \mid \text{Supp}(a_{\ell,m}) \cap \mathcal{T} \neq \emptyset \quad \text{and} \quad 2^\ell \geq 2^{L_1}\} \end{aligned}$$

since equations (70) and (73) implies that  $J_T \subset (J_0 \cup J_1)$ .

One has

$$\begin{aligned} \text{Card}(J_0) &\leq \sum_{\ell \geq L_0} 2^{-\ell} \leq 2 \cdot 2^{-L_0} \leq 2(C_f C_a)^2 \lambda \underbrace{2^{2j\alpha}}_{\leq T^{\frac{2\alpha}{\alpha+1}}} \underbrace{\lambda}_{\leq T^{\frac{2p}{\delta}}} T^{-2} \\ &\leq 2(C_f C_a)^2 \lambda T^{-2 \frac{\alpha}{\alpha+1} \frac{p+1}{\delta}} \leq 2(C_f C_a)^2 \lambda s_j^{\frac{\alpha-p}{\delta}} T^{-2 \frac{\alpha+p}{\delta}}. \end{aligned}$$

For each scale  $2^\ell$ , there are less than  $K 2^j / (\lambda 2^\ell)$  bandlets in  $\{a_{\ell,m}\}_m$  that intersect  $\mathcal{T}$ , so

$$\begin{aligned} \text{Card}(J_1) &\leq \sum_{\ell \geq L_1} K 2^j / (\lambda 2^\ell) \leq 2 K 2^j \lambda^{-1} 2^{-L_1} \\ &\leq 2 K 2^j s_j^{1/p-1} \lambda^{-\alpha/p} \leq 2 K s_j^{1/p} \lambda^{-\alpha/p} \leq 2 K \lambda s_j^{\frac{\alpha-p}{\delta}} T^{-2 \frac{\alpha+p}{\delta}} \end{aligned}$$

The number of coefficients is thus bounded by

$$M(S) \leq \text{Card}(J_0) + \text{Card}(J_1) \leq C_2 \lambda s_j^{\frac{\alpha-p}{\delta}} T^{-2 \frac{\alpha+p}{\delta}}.$$

**Bounding  $\|f_j - f_{jM(S)}\|_2^2$ .** One has

$$\|f_j - f_{jM(S)}\|_2^2 = \sum_{(\ell,m) \notin J_T} |\langle f_j, a_{\ell,m} \rangle|^2 \quad (74)$$

$$= \sum_{(\ell,m) \in (J_0 \cup J_1) \setminus J_T} |\langle f_j, a_{\ell,m} \rangle|^2 + \sum_{(\ell,m) \in \tilde{J}_0 \cup \tilde{J}_1} |\langle f_j, a_{\ell,m} \rangle|^2, \quad (75)$$

where the set of coefficients are split according to

$$\begin{aligned} \tilde{J}_0 &\stackrel{\text{def.}}{=} \{(\ell, m) \setminus \text{Supp}(a_{\ell,m}) \cap \mathcal{T} = \emptyset \quad \text{and} \quad 2^\ell < 2^{L_0}\} \\ \text{and} \quad \tilde{J}_1 &\stackrel{\text{def.}}{=} \{(\ell, m) \setminus \text{Supp}(a_{\ell,m}) \cap \mathcal{T} \neq \emptyset \quad \text{and} \quad 2^\ell < 2^{L_1}\}. \end{aligned}$$

The first part of the error in equation (75) is bounded using

$$\begin{aligned} \sum_{(\ell,m) \in (J_0 \cup J_1) \setminus J_T} |\langle f_j, a_{\ell,m} \rangle|^2 &\leq (\text{Card}((J_0 \cup J_1) \setminus J_T)) T^2 \\ &\leq (\text{Card}(J_0 \cup J_1)) T^2 \leq C_2 \lambda s_j^{\frac{\alpha-p}{\delta}} T^{\frac{4\alpha p}{\delta}}. \end{aligned}$$

Bounding the contribution of  $\tilde{J}_0 \cup \tilde{J}_1$  to the error using directly equation (72) leads to a sub-optimal result. The linear space

$$V \stackrel{\text{def.}}{=} \text{Span} \{a_{\ell,m} \setminus (\ell, m) \in J_0 \cup J_1\}$$

allows us to write the second part of the error in equation (75) as a projection

$$\sum_{(\ell,m) \in \tilde{J}_0 \cup \tilde{J}_1} |\langle f_j, a_{\ell,m} \rangle|^2 = \|f_j - P_V(f_j)\|_2^2.$$

The following decomposition

$$f_j = f_j^0 + f_j^1 \quad \text{where} \quad \begin{cases} \forall (2^j n) \in \mathcal{T}, & f_j^0[n] = 0, & \text{and} & f_j^1[n] = f_j[n], \\ \forall (2^j n) \notin \mathcal{T}, & f_j^0[n] = f_j[n], & \text{and} & f_j^1[n] = 0, \end{cases}$$

leads to

$$\begin{aligned} \|f_j - P_V(f_j)\|_2 &\leq \|f_j^1 - P_V(f_j^1)\|_2 + \|f_j^0\|_2 + \|P_V(f_j^0)\|_2 \\ &\leq \|f_j^1 - P_V(f_j^1)\|_2 + 2\|f_j^0\|_2. \end{aligned}$$

The norm of  $f_j^0$  is bounded using the fact that the wavelet coefficients are bounded by equation (69) outside  $\mathcal{T}$

$$\begin{aligned} \|f_j^0\|_2^2 &\leq \underbrace{(\lambda^2 2^{-2j})}_{\text{nbr. pts. in } S} (C_f)^2 2^{2j(\alpha+1)} \leq (C_f)^2 \lambda \underbrace{2^{2j\alpha}}_{\leq T^{\frac{2\alpha}{\alpha+1}}} \underbrace{\lambda}_{\leq T^{\frac{2p}{\delta}}} \\ &\leq (C_f)^2 \lambda T^{\frac{2\alpha}{\alpha+1} + \frac{2p}{\delta}} s_j^{\frac{p+1}{\delta}} \leq (C_f)^2 \lambda T^{\frac{4\alpha p}{\delta}} s_j^{\frac{\alpha-p}{\delta}}. \end{aligned}$$

Equation (71) allows to decompose  $f_{j_w} = P + \varepsilon$  inside each warped domain  $w(\beta_{L_0, m})$ . The fact that warped discret polynomials  $P$  belongs to  $V$  implies that the pointwise error is bounded by

$$\forall (2^j n) \in \beta_{L_0, m}, \quad |f_j^1[n] - P_V(f_j^1)[n]| \leq 2\|\varepsilon\|_\infty \leq 4C_w 2^j s_j^{-1} \lambda^\alpha.$$

Since the number of points  $(2^j n) \in \mathcal{T}$  where  $f_j^1[n] \neq 0$  is  $(K s_j / \lambda)(b 2^{-j})^2$ , one gets the estimate

$$\|f_j - P_V(f_j)\|_2^2 \leq (K s_j / \lambda)(\lambda 2^{-j})^2 (2\|\varepsilon\|_\infty)^2 \leq C s_j^{-1} \lambda^{2\alpha+1} \leq C \lambda T^{\frac{4\alpha p}{\delta}} s_j^{\frac{\alpha-p}{\delta}}.$$

□

## B Fast Alpert Transform

Let  $\mathcal{B}(S, \tilde{\gamma}') = \{a_{\ell, m}\}_{\ell, m}$  be an Alpert bandletization basis over a square  $S$  containing wavelets coefficients  $f_j[n]$  for  $(2^j n) \in S$ . An adapted Alpert transform [1] computes with a fast algorithm the coefficients  $\{\langle f_j, a_{\ell, m} \rangle\}_{\ell, m}$  of  $f_j$  with a complexity of  $O(N)$  where  $N$  is the number of input wavelet coefficients  $f_j[n]$ .

For  $x \in \mathbb{R}^2$ , let  $x^t = x_1^{t_1} x_2^{t_2}$  and  $|t| = t_1 + t_2$ . The 2D monomials are indexed using

$$\forall t \text{ such that } |t| < p, \quad \tilde{t} \stackrel{\text{def.}}{=} \frac{1}{2}|t|(|t| + 1) + t_2 \in \{0, \dots, p(p+1)/2 - 1\}.$$

In the following, the sampling locations of wavelet coefficients are written as  $x_n \stackrel{\text{def.}}{=} 2^j n \in [0, 1]^2$  and the warped points  $\tilde{x}_n \stackrel{\text{def.}}{=} w(x_n)$  where the warping is defined by (10). If  $A$  and  $B$  are two matrices,  $[A; B]$  denotes the concatenation along columns,  $[A, B]$  the concatenation along rows and  $\text{diag}(A, B)$  the concatenation along the diagonal.



**Polynomial inner-product.** The dot product of two polynomials  $P$  and  $Q$  defined over a band  $\tilde{\beta}_{\ell,m}$  is

$$\langle P, Q \rangle_{\ell,m} \stackrel{\text{def.}}{=} \sum_{x_n \in \beta_{\ell,m}} P(\tilde{x}_n) Q(\tilde{x}_n) = \bar{P}^T A_{\ell,m} \bar{Q}.$$

where  $\bar{P}$  is the column vector of coefficients of  $P$  in the basis of monomials

$$P(x) = \sum_{|t| < p} \bar{P}_t x^t.$$

For all  $m$ , the symmetric matrices  $A_{\ell,m}$  of size  $p(p+1)/2 \times p(p+1)/2$  are the matrices of the dot products over  $\tilde{\beta}_{\ell,m}$  for discrete polynomial vectors expressed in the basis of monomials, defined by

$$\forall m = 0, \dots, 2^\ell - 1, \quad (A_{\ell,m})_{\tilde{s}, \tilde{t}} \stackrel{\text{def.}}{=} \sum_{x_n \in \beta_{\ell,m}} (\tilde{x}_n)^{s+t}. \quad (76)$$

These matrices are computed iteratively during the Alpert transform.

**Initialization.** The finest partition  $S = \bigcup_{m=0}^{2^{-L}-1} \beta_{Lm}$  is computed by recursive splits as explained in section V.1. For all  $m \in \{0, \dots, 2^{-\ell} - 1\}$ , one needs to compute the polynomial  $P_{Lm}$  of degree  $p$  that interpolates  $f_j$  on  $\beta_{Lm}$ ,

$$\forall x_n \in \beta_{Lm}, \quad P_{Lm}(\tilde{x}_n) = f_j[n].$$

These polynomials correspond to the finest scale representation of  $f_j$  in the basis of local monomials. The matrices  $A_{Lm}$  must also be calculated with (76) for  $\ell = L$ .

**Computation of Alpert coefficients.** For each scale  $\ell = L, \dots, -1$  and each  $m$ , the pair of polynomials  $(P_0, P_1) \stackrel{\text{def.}}{=} (P_{\ell,2m}, P_{\ell,2m+1})$  is decomposed into a sum of a lower scale polynomial  $P \stackrel{\text{def.}}{=} P_{\ell+1,m}$  and a residual  $(Q_0, Q_1)$  which is orthogonal to polynomials of degree less than  $p-1$ . The polynomials  $P_0$  and  $Q_0$  are restricted to  $\tilde{\beta}_{\ell,2m}$ , the polynomials  $P_1$  and  $Q_1$  are restricted to  $\tilde{\beta}_{\ell,2m+1}$  whereas the polynomial  $P = P_0 - Q_0 = P_1 - Q_1$  is restricted to  $\tilde{\beta}_{\ell+1,m}$ . In the following we denote  $A_0 = A_{\ell,2m}$ ,  $A_1 = A_{\ell,2m+1}$  and  $A \stackrel{\text{def.}}{=} \text{diag}(A_0, A_1)$ .

The residual polynomials  $(Q_0, Q_1)$  is computed as the projection of  $(P_0, P_1)$  on a set of orthogonal couples of polynomials  $\{(h_t^0, h_t^1)\}_{0 \leq |t| < p}$ . Each couple  $h_t \stackrel{\text{def.}}{=} (h_t^0, h_t^1)$  is the piecewise polynomial that interpolates the bandlet vector  $a_{\tilde{i}}$  for  $\tilde{i} \stackrel{\text{def.}}{=} p(p+1)i/2 + \tilde{t}$  over the set of locations  $\tilde{\beta}_{\ell+1,m} = \tilde{\beta}_{\ell,2m} \cup \tilde{\beta}_{\ell,2m+1}$ ,

$$\forall \tilde{x}_n \in \tilde{\beta}_{\ell,2m+\varepsilon}, \quad b_{\tilde{i}, \tilde{x}_n} \stackrel{\text{def.}}{=} h_t^\varepsilon(\tilde{x}_n), \quad \text{for } \varepsilon \in \{0, 1\}.$$

The bandlet piecewise-polynomials  $h_t$  are computed by satisfying two criteria.

– **Orthogonality:** For each  $0 \leq |t|, |t'| < p$ , one should have

$$\delta_{t,t'} = \sum_{x_n \in \beta_{\ell+1,i}} h_t(\tilde{x}_n) h_{t'}(\tilde{x}_n) = \sum_{x_n \in \beta_{\ell,2i}} h_t^0(\tilde{x}_n) h_{t'}^0(\tilde{x}_n) + \sum_{x_n \in \beta_{\ell,2i+1}} h_{t'}^1(\tilde{x}_n) h_t^1(\tilde{x}_n) \quad (77)$$

where  $\delta$  is the Kronecker symbol.

– **Vanishing moments:** For each  $0 \leq |s|, |t| < p$ ,  $h_t$  is orthogonal to the monomial  $x^s$  defined on  $\tilde{\beta}_{\ell, m}$

$$0 = \sum_{x_n \in \beta_{\ell+1, i}} h_t(\tilde{x}_n)(\tilde{x}_n)^s = \sum_{x_n \in \beta_{\ell, 2i}} h_t^0(\tilde{x}_n)(\tilde{x}_n)^s + \sum_{x_n \in \beta_{\ell, 2i+1}} h_t^1(\tilde{x}_n)(\tilde{x}_n)^s. \quad (78)$$

One can express conditions (77) and (78) in matrix form in the basis of the monomials as

$$((77) \iff H^T A H = \text{Id}) \quad \text{and} \quad ((78) \iff [A_0, A_1] H = 0), \quad (79)$$

where the column of index  $\tilde{t}$  of the matrix  $H$  of size  $p(p+1) \times p(p+1)/2$  is  $[\bar{h}_t^0; \bar{h}_t^1]$ . Conditions (79) mean that

$$A^{1/2} H \text{ is an orthogonal basis of the kernel of } [A_0, A_1] A^{-1/2}.$$

Matrix  $H$  can thus be computed with a constant number of operations proportional to  $(p(p+1)/2)^3$ . The residual is computed by  $[\bar{Q}_0; \bar{Q}_1] = H [\bar{P}_0; \bar{P}_1]$  and the low scale polynomial is  $P = P_0 - Q_0$ . The bandlet coefficients are the projections of  $(Q_0, Q_1)$  onto the computed basis using the dot product defined by  $A$

$$\langle f_j, a_{\ell \tilde{m}} \rangle = \langle h_t^0, Q_0 \rangle_{\ell, 2m} + \langle h_t^1, Q_1 \rangle_{\ell, 2m+1} = (\bar{h}_t)^T A [\bar{Q}_0; \bar{Q}_1]$$

for  $\tilde{m} \stackrel{\text{def}}{=} p(p+1)i/2 + \tilde{t}$ . The dot product matrix for the next scale is computed using  $A_{\ell+1, m} = A_{\ell, 2m} + A_{\ell, 2m+1}$ .

A constant number of operations is needed to compute each bandelet matrix  $H$  and to update matrices  $A_{\ell, m}$  for each scale  $2^\ell$ . Since this process is repeated  $N/2^\ell$  times for each scale, the overall complexity of the algorithm is  $O(N)$  to transform  $N$  wavelet coefficients  $f_j[n]$ .

## References

- [1] B. Alpert. A class of bases in  $L^2$  for the sparse representation of integral operators. *SIAM J. Math. Anal.*, (24), 1993.
- [2] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen. *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.
- [3] E. Candes and D. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise  $c_2$  singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.
- [4] R. L. Claypoole, G. M. Davis, W. Sweldens, and R. G. Baraniuk. Nonlinear wavelet transforms for image coding via lifting. *IEEE Trans. Image Processing*, 12(12):1449–1459, December 2003.
- [5] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.*, 1(1):54–81, 1993.

- [6] M. N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions Image on Processing*, To appear, 2005.
- [7] D. Donoho. Wedgelets: Nearly-minimax estimation of edges. *Ann. Statist*, 27:353–382, 1999.
- [8] D. Donoho. Counting bits with kolmogorov and shannon. Technical report, Stanford, 2000.
- [9] P.L. Dragotti and M. Vetterli. Wavelet footprints: Theory, algorithms and applications. *IEEE Trans. on Signal Processing*, 51(5):1306–1323, Mat 2003.
- [10] E. Le Pennec and S. Mallat. Bandelet Image Approximation and Compression. *SIAM Multiscale Modeling and Simulation*, page to appear, 2005.
- [11] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998.
- [12] B. Matei and A. Cohen. *Nonlinear Subdivison Schemes : Applications to Image processing*, in *Tutorials on Multiresolution in Geometric Modelling*, pages 93–97. Springer Verlag, 2002.
- [13] Y Meyer. *Wavelets and Operators*. Cambridge University Press, 1993.
- [14] G. Peyré and S. Mallat. Bandelets toolbox, available on Matlab Central. <http://www.mathworks.com/matlabcentral/>, 2005.
- [15] R. Shukla, P.L. Dragotti, M. Do, and M. Vetterli. Rate distortion optimized tree structured compression algorithms for piecewise smooth images. *IEEE Transactions on Image Processing*, April 2004.
- [16] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk. Wavelet-domain Approximation and Compression of Piecewise Smooth Images. *IEEE Transactions on Image Processing*, To appear, 2005.