OPTIMALLY SPARSE MULTIDIMENSIONAL REPRESENTATION USING SHEARLETS

KANGHUI GUO* AND DEMETRIO LABATE†

Abstract. In this paper we show that the shearlets, an affine-like system of functions recently introduced by the authors and their collaborators, are essentially optimal in representing 2–dimensional functions f which are C^2 except for discontinuities along C^2 curves. More specifically, if f_N^S is the N–term reconstruction of f obtained by using the N largest coefficients in the shearlet representation, then the asymptotic approximation error decays as

$$\|f - f_N^S\|_2^2 \asymp N^{-2} (\log N)^3, \qquad N \to \infty,$$

which is essentially optimal, and greatly outperforms the corresponding asymptotic approximation rate N^{-1} associated with wavelet approximations.

Unlike the curvelets, that have similar sparsity properties, the shearlets form an affine-like system and have a simpler mathematical structure. In fact, the elements of this system form a Parseval frame and are generated by applying dilations, shear transformations and translations to a single well-localized window function.

Key words. affine systems, curvelets, geometric image processing, shearlets, sparse representation, wavelets

AMS subject classifications. 42C15, 42C40

1. Introduction.

The notion of efficient representation of data plays an increasingly important role in areas across applied mathematics, science and engineering. Over the past few years, there has been a rapidly increasing pressure to handle ever larger and higher dimensional data sets, with the challenge of providing representations of these data that are sparse (that is, "very" few terms of the representation are sufficient to accurately approximate the data) and computationally fast. Sparse representations have implications reaching beyond data compression. Understanding the compression problem for a given data type entails a precise knowledge of the modelling and approximation of that data type. This in turn is useful for many other important tasks, including classification, denoising, interpolation, and segmentation [14].

Multiscale techniques based on wavelets have emerged over the last 2 decades as the most successful methods for the efficient representation of data, as testified, for example, by their use in the new FBI fingerprint database [3] and in JPEG2000, the new standard for image compression [4, 20]. Indeed, wavelets are optimally efficient in representing functions with pointwise singularities [28, Ch.9].

More specifically, consider the wavelet representation (using a "nice" wavelet basis) of a function f of a *single variable* that is smooth apart from a point discontinuity. Because the elements of the wavelet basis are well-localized (i.e., they have very fast decay both in the spatial and in the frequency domain), very few of them interact significantly with the singularity, and, thus, very few elements of the wavelet expansion are sufficient to provide an accurate approximation. This contrasts sharply with the Fourier representation, for which the discontinuity interacts extensively with the elements of the Fourier basis. Denoting by f_N the approximation obtained by using the largest N coefficients in the wavelet expansion, the asymptotic approximation error satisfies

$$\|f - f_N\|_2^2 \asymp N^{-2}, \qquad N \to \infty.$$

This is the optimal approximation rate for this type of functions [11], and outperforms the corresponding Fourier approximation error rate N^{-1} [14, 28]. In addition, the Multiresolution Analysis (MRA) associated with wavelets provides very fast numerical algorithms for computing the wavelet coefficients [10, 28].

*Department of Mathematics, Missouri State University, Springfield,Missouri 65804, USA (kag026f@smsu.edu)

†Department of Mathematics, North Carolina State University, Campus Box 8205, Raleigh, NC 27695, USA (dlabate@math.ncsu.edu)

However, despite their remarkable success in applications, wavelets are far from optimal in dimensions larger than one. Indeed wavelets are very efficient in dealing with pointwise singularities *only*. In higher dimensions other types of singularities are usually present or even dominant, and wavelets are unable to handle them very efficiently. Consider, for example, the wavelet representation of a 2-D function that is smooth away from a discontinuity along a curve of finite length (a reasonable model for an image containing an edge). Because the discontinuity is spatially distributed, it interacts extensively with the elements of the wavelet basis. As a consequence, the wavelet coefficients have a slow decay, and the approximation error $\|f - f_N\|_2^2$ decays at most as fast as $O(N^{-1})$ [28]. This is better than the rate of the Fourier approximation error $N^{-1/2}$, but far from the optimal theoretical approximation rate (cf. [13])

$$\|f - f_N\|_2^2 \asymp N^{-2}, \qquad N \to \infty. \tag{1.1}$$

There is, therefore, large room for improvements, and several attempts have been made in this direction both in the mathematical and engineering communities, in recent years. Those include *contourlets*, *complex wavelets* and other "directional wavelets" in the filter bank literature [1, 2, 12, 23, 27, 29], as well as *brushlets* [8], *ridgelets* [5], *curvelets* [7] and *bandelets* [25].

The most successful approach so far are the curvelets of Candès and Donoho. This is the first and so far the only construction providing an essentially optimal approximation property for 2–D piecewise smooth functions with discontinuities along C^2 curves [7]. The main idea in the curvelet approach is that, in order to approximate functions with edges accurately, one has to exploit their geometric regularity much more efficiently than traditional wavelets. This is achieved by constructing an appropriate tight frame of well-localized functions at various scales, positions and directions. We refer to [6, 7] for more details about this construction.

The main goal of this paper is to show that the shearlets, a construction based on the theory of composite wavelets, also provides an essentially optimal approximation property for 2–D piecewise smooth functions with discontinuities along C^2 curves. We will show that the approximation error associated with the N–term reconstruction f_N^S obtained by taking the N largest coefficients in the shearlet expansion satisfies

$$\|f - f_N^S\|_2^2 \asymp N^{-2}(\log N)^3, \qquad N \to \infty. \tag{1.2}$$

This is exactly the approximation rate obtained using curvelets. The proof of our result adapts several ideas from the corresponding sparsity result of the curvelets [7] and follows the general architecture of that proof, but does not follow directly from the curvelets construction. Indeed, as we will argue in the following, our alternative approach based on shearlets has some mathematical advantages with respect to curvelets, including a simplified construction that provides the framework for a simpler mathematical analysis and fast algorithmic implementation (see also [9, 15]).

The theory of composite wavelets, recently proposed by the authors and their collaborators [17, 18, 19], provides a most general setting for the construction of truly multidimensional, efficient, multiscale representations. Unlike the curvelets, this approach takes full advantage of the theory of affine systems on $\mathbb{R}^n$. Specifically, the **affine systems with composite dilations** are the systems:

$$\mathcal{A}_{AB}(\psi) = \{\psi_{j,\ell,k}(x) = |\det A|^{j/2}\, \psi(B^\ell A^j x - k) : \, j, \ell \in \mathbb{Z}, k \in \mathbb{Z}^n\}, \tag{1.3}$$

where A, B are $n \times n$ invertible matrices and $|\det B| = 1$. The elements of this system are called **composite wavelets** if $\mathcal{A}_{AB}(\psi)$ forms a **Parseval frame** (also called tight frame) for $L^2(\mathbb{R}^n)$; that is,

$$\sum_{j,\ell,k} |\langle f, \psi_{j,\ell,k}\rangle|^2 = \|f\|^2,$$

for all $f \in L^2(\mathbb{R}^n)$. The shearlets, that will be considered in this paper, are a special class of composite wavelets where A is an anisotropic dilation and B is a shear matrix. Details for this construction will be given in Section 1.2. These representations have fully controllable geometrical features, such as orientations, scales and shapes, that set them apart from traditional wavelets as well as complex and directional wavelets. In addition, thanks to their mathematical structure, there is a multiresolution analysis naturally associated with composite wavelets. This is particularly useful for the development of fast algorithmic implementations of these transformations [24, 26].

Observe that curvelets are not of the form (1.3), and, unlike the shearlets, are not generated from the action of a family of operators on a single or finite family of functions.

1.1. Notation. Throughout this paper, we shall consider the points $x \in \mathbb{R}^n$ to be column vectors, i.e., $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, and the points $\xi \in \widehat{\mathbb{R}}^n$ (the frequency domain) to be row vectors, i.e., $\xi = (\xi_1, \dots, \xi_n)$. A vector x multiplying a matrix $a \in GL_n(\mathbb{R})$ on the right, is understood to be a column vector, while a vector ξ multiplying a on the left is a row vector. Thus, $ax \in \mathbb{R}^n$ and $\xi a \in \widehat{\mathbb{R}}^n$. The Fourier transform is defined as

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)\, e^{-2\pi i \xi x}\, dx,$$

where $\xi \in \widehat{\mathbb{R}}^n$, and the inverse Fourier transform is

$$\check{f}(x) = \int_{\widehat{\mathbb{R}}^n} f(\xi)\, e^{2\pi i \xi x}\, d\xi.$$

1.2. Shearlets. The collection of shearlets, that we are going to define in this section, is a special example of composite wavelets in $L^2(\mathbb{R}^2)$, of the form (1.3), where:

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{1.4}$$

and ψ will be defined in the following. It is useful to observe that, by applying the Fourier transform to the elements $\psi_{j,\ell,k}$ in (1.3), we obtain

$$\hat{\psi}_{j,\ell,k}(\xi) = |\det A|^{-j/2}\, \psi(\xi A^{-j} B^{-\ell})\, e^{2\pi i \xi A^{-j} B^{-\ell} k}.$$

For any $\xi = (\xi_1, \xi_2) \in \widehat{\mathbb{R}}^2$, $\xi_1 \neq 0$, let ψ be given by

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\, \hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right), \tag{1.5}$$

where $\hat{\psi}_1, \hat{\psi}_2 \in C^\infty(\widehat{\mathbb{R}})$, $\operatorname{supp} \hat{\psi}_1 \subset [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$ and $\operatorname{supp} \hat{\psi}_2 \subset [-1, 1]$. We assume that

$$\sum_{j \geq 0} |\hat{\psi}_1(2^{-2j}\omega)|^2 = 1 \quad \text{for } |\omega| \geq \frac{1}{8}, \tag{1.6}$$

and

$$|\hat{\psi}_2(\omega - 1)|^2 + |\hat{\psi}_2(\omega)|^2 + |\hat{\psi}_2(\omega + 1)|^2 = 1 \quad \text{for } |\omega| \leq 1. \tag{1.7}$$

It follows from the last equation that, for any $j \geq 0$,

$$\sum_{\ell = -2^j}^{2^j} |\hat{\psi}_2(2^j\, \omega + \ell)|^2 = 1 \quad \text{for } |\omega| \leq 1. \tag{1.8}$$

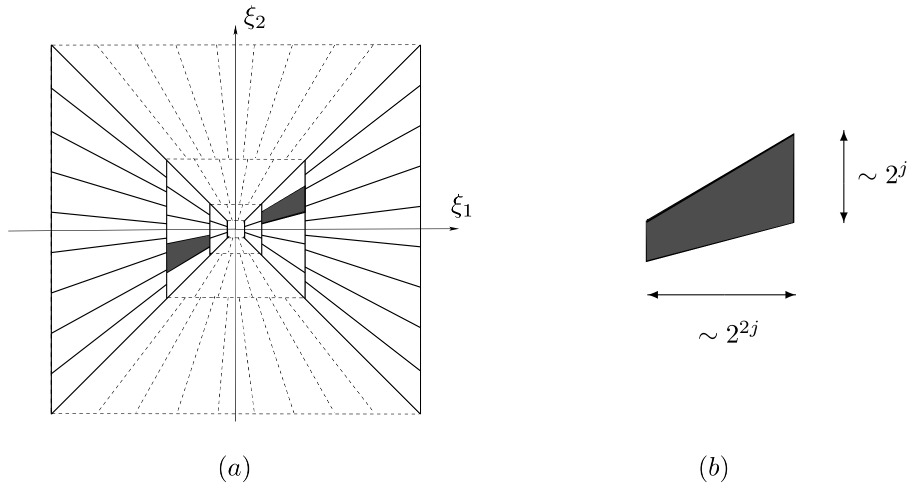


FIG. 1.1. *(a) The tiling of the frequency plane $\widehat{\mathbb{R}}^2$ induced by the shearlets. (b) Frequency support of the shearlet $\psi_{j,\ell,k}$, for $\xi_1 > 0$. The other half of the support, for $\xi_1 < 0$, is symmetrical.*

It also follows from our assumptions that $\hat\psi \in C_0^\infty(\widehat{\mathbb{R}}^2)$, with $\operatorname{supp}\hat\psi \subset [-\frac12,\frac12]^2$. There are several examples of functions ψ_1, ψ_2 satisfying the properties described above (see Appendix A).

Observe that $(\xi_1,\xi_2)\,A^{-j}B^{-\ell} = (2^{-2j}\xi_1, -\ell 2^{-2j}\xi_1 + 2^{-j}\xi_2)$. Using (1.6) and (1.8) it is easy to see that:

$$\sum_{j\ge0}\sum_{\ell=-2^j}^{2^j} |\hat\psi(\xi\, A^{-j}B^{-\ell})|^2 = \sum_{j\ge0}\sum_{\ell=-2^j}^{2^j} |\hat\psi_1(2^{-2j}\,\xi_1)|^2\, |\hat\psi_2(2^j\,\frac{\xi_2}{\xi_1}-\ell)|^2$$

$$= \sum_{j\ge0} |\hat\psi_1(2^{-2j}\,\xi_1)|^2 \sum_{\ell=-2^j}^{2^j} |\hat\psi_2(2^j\,\frac{\xi_2}{\xi_1}-\ell)|^2 = 1,$$

for $(\xi_1,\xi_2)\in\mathcal{D}_C$, where $\mathcal{D}_C = \{(\xi_1,\xi_2)\in\widehat{\mathbb{R}}^2 : |\xi_1|\ge\frac18,\ |\frac{\xi_2}{\xi_1}|\le 1\}$. This equation, together with the fact that $\hat\psi$ is supported inside $[-\frac12,\frac12]^2$, implies that the collection of **shearlets**:

$$\mathcal{SH}(\psi) = \{\psi_{j,\ell,k}(x) = 2^{\frac{3j}{2}}\,\psi(B^\ell A^j x - k) : \ j\ge 0, -2^j \le \ell \le 2^j, k\in\mathbb{Z}^2\}, \qquad (1.9)$$

is a Parseval frame for $L^2(\mathcal{D}_C)^\vee = \{f\in L^2(\mathbb{R}^2) : \operatorname{supp}\hat f\subset\mathcal{D}_C\}$. Details about the argument that this system is a Parseval frame can be found in [19, Sec.5.2.1].

To obtain a Parseval frame for $L^2(\mathbb{R}^2)$, one can construct a second system of shearlets which form a Parseval frame for the functions with frequency support in the vertical cone $\mathcal{D}_{\widetilde C} = \{(\xi_1,\xi_2)\in\widehat{\mathbb{R}}^2 : |\xi_2|\ge\frac18,\ |\frac{\xi_1}{\xi_2}|\le1\}$. Finally, one can easily construct a Parseval frame (or an orthonormal basis) for $L^2([-\frac18,\frac18]^2)^\vee$. Then any function in $L^2(\mathbb{R}^2)$ can be expressed as a sum $f = P_C f + P_{\widetilde C} f + P_0 f$, where each component corresponds to the orthogonal projection of f into one of the 3 subspaces of $L^2(\mathbb{R}^2)$ described above. The tiling of the frequency plane $\widehat{\mathbb{R}}^2$ induced by this system is illustrated in Figure 1.1(a). The above construction was first introduced in [16].

The conditions on the support of $\hat\psi_1$ and $\hat\psi_2$ imply that the functions $\hat\psi_{j,\ell,k}$ have frequency support:

$$\operatorname{supp}\hat\psi_{j,\ell,k} \subset \{(\xi_1,\xi_2) : \xi_1\in[-2^{2j-1},-2^{2j-4}]\cup[2^{2j-4},2^{2j-1}],\ |\tfrac{\xi_2}{\xi_1}-\ell\,2^{-j}|\le 2^{-j}\}.$$

Thus, the system $\mathcal{SH}(\psi)$, given by (1.9), is a Parseval frame exhibiting the following properties.

(i) *Time-Frequency Localization.* Since $\hat{\psi} \in C_0^\infty(\widehat{\mathbb{R}}^2)$, then $|\psi(x)| \le C_N\,(1+|x|)^{-N}$, for any $N \in \mathbb{N}$, and, thus, the elements $\psi_{j,\ell,k}$ are well-localized.
(ii) *Parabolic Scaling.* Each element $\hat{\psi}_{j,\ell,k}$ has support on a pair of trapezoids, each one contained in a box of size approximately $2^{2j} \times 2^j$ (see Figure 1.1(b)). Because the shearlets are well-localized, each $\psi_{j,\ell,k}$ is essentially supported on a box of size $2^{-2j} \times 2^{-j}$. Thus, their supports become increasingly thin as $j \to \infty$.
(iii) *Directional Sensitivity.* The elements $\hat{\psi}_{j,\ell,k}$ are oriented along lines with slope given by $\ell\,2^{-j}$. As a consequence, the corresponding elements $\psi_{j,\ell,k}$ are oriented along lines with slope $-\ell\,2^{-j}$. The number of orientations (approximately) doubles at each finer scale.
(iv) *Spatial Localization.* For any fixed scale and orientation, the shearlets are obtained by translations on the lattice $\mathbb{Z}^2$.
(v) *Oscillatory Behavior.* The shearlets $\psi_{j,\ell,k}$ are non-oscillatory along the orientation axis of slope $-\ell\,2^{-j}$, and oscillatory across this axis.

Observe that the curvelets of Candès and Donoho also satisfy similar properties with the following main differences. Concerning property (iii), the number of orientations in the curvelet constructions doubles at each *other* scale. Concerning property (iv), the curvelets are not associated with a *fixed* translation lattice. However, for a given scale parameter j and orientation θ, they are obtained by translations on a grid that depends on j and θ. In fact, as we mentioned before, unlike the shearlets, the curvelets are not generated from the action of a family of operators on a single or finite family of functions.

1.3. Main results. One major feature of shearlets is that, if f is a compactly supported function which is C^2 away from a C^2 curve, then the sequence of shearlet coefficients $\{\langle f, \psi_{j,\ell,k}\rangle\}$ has (essentially) optimally fast decay. As a consequence, if f_N^S is the N–term approximation of f obtained from the N largest coefficients of its shearlet expansion, then the approximation error is essentially optimal.

Before stating the main theorems, let us define more precisely the class of functions we are interested in. We follow [7] and introduce $STAR^2(A)$, a class of indicator functions of sets B with C^2 boundaries ∂B. In polar coordinates, let $\rho(\theta) : [0, 2\pi) \to [0,1]^2$ be a radius function and define B by $x \in B$ if and only if $|x| \le \rho(\theta)$. In particular, the boundary ∂B is given by the curve in $\mathbb{R}^2$:

$$\beta(\theta) = \begin{pmatrix} \rho(\theta)\,\cos(\theta) \\ \rho(\theta)\,\sin(\theta) \end{pmatrix}. \tag{1.10}$$

The class of boundaries of interest to us are defined by

$$\sup |\rho''(\theta)| \le A, \quad \rho \le \rho_0 < 1. \tag{1.11}$$

We say that a set $B \in STAR^2(A)$ if $B \subset [0,1]^2$ and B is a translate of a set obeying (1.10) and (1.11). In addition, we set $C_0^2([0,1]^2)$ to be the collection of twice differentiable functions supported inside $[0,1]^2$.

Finally, we define the set $\mathcal{E}^2(A)$ of **functions which are C^2 away from a C^2 edge** as the collection of functions of the form

$$f = f_0 + f_1\,\chi_B,$$

where $f_0, f_1 \in C_0^2([0,1]^2)$, $B \in STAR^2(A)$ and $\|f\|_{C^2} = \sum_{|\alpha|\le 2} \|D^\alpha f\|_\infty \le 1$.

Let M be the set of indices $\{(j,\ell,k) : j \ge 0, -2^j \le \ell \le 2^j, k \in \mathbb{Z}^2\}$ and $\{\psi_\mu\}_{\mu\in M}$ be the Parseval frame of shearlets given by (1.9). The **shearlet coefficients** of a given function f are the elements of the sequence $\{s_\mu(f) = \langle f, \psi_\mu\rangle : \mu \in M\}$. We denote by $|s(f)|_{(N)}$ the N-th largest entry in this sequence. We can now state the following results.

Theorem 1.1. *Let $f \in \mathcal{E}^2(A)$, and $\{s_\mu(f) = \langle f, \psi_\mu\rangle : \mu \in M\}$ be the sequence of shearlet coefficients associated with f. Then*

$$\sup_{f\in\mathcal{E}^2(A)} |s(f)|_{(N)} \le C\,N^{-3/2}\,(\log N)^{3/2}. \tag{1.12}$$

Let f_N^S be the N–term approximation of f obtained from the N largest coefficients of its shearlet expansion, namely

$$f_N^S = \sum_{\mu \in I_N} \langle f, \psi_\mu \rangle \, \psi_\mu,$$

where $I_N \subset M$ is the set of indices corresponding to the N largest entries of the sequence $\{|\langle f, \psi_\mu\rangle|^2 : \mu \in M\}$. Then the approximation error satisfies

$$\|f - f_N^S\|_2^2 \leq \sum_{m > N} |s(f)|^2_{(m)}.$$

Therefore, from (1.12) we immediately have:

Theorem 1.2. *Let $f \in \mathcal{E}^2(A)$ and f_N^S be the approximation to f defined above. Then*

$$\|f - f_N^S\|_2^2 \leq C \, N^{-2} \, (\log N)^3.$$

1.4. Analysis of the shearlet coefficients. The argument that will be used to prove Theorem 1.1 follows essentially the architecture of the proofs in [7]. In order to measure the sparsity of the shearlet coefficients $\{\langle f, \psi_\mu\rangle : \mu \in M\}$, we will use the **weak–ℓ^p quasi–norm** $\|\cdot\|_{w\ell^p}$ defined as follows. Let $|s_\mu|_{(N)}$ be the N-th largest entry in the sequence $\{s_\mu\}$. Then

$$\|s_\mu\|_{w\ell^p} = \sup_{N>0} N^{\frac{1}{p}} \, |s_\mu|_{(N)}.$$

One can show (cf. [30, Sec.5.3]) that this definition is equivalent to

$$\|s_\mu\|_{w\ell^p} = \left(\sup_{\epsilon > 0} \#\{\mu : \, |s_\mu| > \epsilon\} \, \epsilon^p \right)^{\frac{1}{p}}.$$

To analyze the decay properties of the shearlet coefficients $\{\langle f, \psi_\mu\rangle\}$ at a given scale 2^{-j}, we will smoothly localize the function f near dyadic squares. Fix the scale parameter $j \geq 0$. For this j fixed, let $M_j = \{(j, \ell, k) : \, -2^j \leq \ell \leq 2^j, k \in \mathbb{Z}^2\}$ and $\mathcal{Q}_j$ be the collection of dyadic cubes of the form $Q = [\frac{k_1}{2^j}, \frac{k_1+1}{2^j}] \times [\frac{k_2}{2^j}, \frac{k_2+1}{2^j}]$, with $k_1, k_2 \in \mathbb{Z}$. For w a nonnegative C^∞ function with support in $[-1,1]^2$, we define a smooth partition of unity

$$\sum_{Q \in \mathcal{Q}_j} w_Q(x) = 1, \quad x \in \mathbb{R}^2,$$

where, for each dyadic square $Q \in \mathcal{Q}_j$, $w_Q(x) = w(2^j x_1 - k_1, 2^j x_2 - k_2)$. We will then examine the shearlet coefficients of the localized function $f_Q = f \, w_Q$, i.e., $\{\langle f_Q, \psi_\mu\rangle : \, \mu \in M_j\}$.

For $f \in \mathcal{E}^2(A)$, the decay properties of the coefficients $\{\langle f_Q, \psi_\mu\rangle : \, \mu \in M_j\}$ will exhibit a very different behavior depending on whether the edge curve intersects the support of w_Q or not. Let $\mathcal{Q}_j = \mathcal{Q}_j^0 \cup \mathcal{Q}_j^1$, where the union is disjoint and $\mathcal{Q}_j^0$ is the collection of those dyadic cubes $Q \in \mathcal{Q}_j$ such that the edge curve intersects the support of w_Q. Since each Q has sidelength $2 \cdot 2^{-j}$, then $\mathcal{Q}_j^0$ has cardinality $|\mathcal{Q}_j^0| \leq C_0 \, 2^j$, where C_0 is independent of j. Similarly, since f is compactly supported in $[0,1]^2$, $|\mathcal{Q}_j^1| \leq 2^{2j} + 4 \cdot 2^j$.

We have the following results, that will be proved in Section 2.

Theorem 1.3. *Let $f \in \mathcal{E}^2(A)$. For $Q \in \mathcal{Q}_j^0$, with $j \geq 0$ fixed, the sequence of shearlet coefficients $\{\langle f_Q, \psi_\mu\rangle : \, \mu \in M_j\}$ obeys*

$$\|\langle f_Q, \psi_\mu\rangle\|_{w\ell^{2/3}} \leq C \, 2^{-\frac{3j}{2}},$$

for some constant C independent of Q and j.

THEOREM 1.4. *Let $f \in \mathcal{E}^2(A)$. For $Q \in \mathcal{Q}_j^1$, with $j \geq 0$ fixed, the sequence of shearlet coefficients $\{\langle f_Q, \psi_\mu \rangle : \mu \in M_j\}$ obeys*

$$\|\langle f_Q, \psi_\mu \rangle\|_{w\ell^{2/3}} \leq C\, 2^{-3j},$$

for some constant C independent of Q and j.

As a consequence of these two theorems, we have the following.

COROLLARY 1.5. *Let $f \in \mathcal{E}^2(A)$ and, for $j \geq 0$, $s_j(f)$ be the sequence $s_j(f) = \{\langle f, \psi_\mu \rangle : \mu \in M_j\}$. Then*

$$\|s_j(f)\|_{w\ell^{2/3}} \leq C,$$

for some C independent of j.

Proof. Using Theorems 1.3 and 1.4, by the p–triangle inequality for weak ℓ^p spaces, $p \leq 1$, we have

$$\begin{aligned}
\|s_j(f)\|_{w\ell^{2/3}}^{2/3} &\leq \sum_{Q \in \mathcal{Q}_j} \|\langle f_Q, \psi_\mu \rangle\|_{w\ell^{2/3}}^{2/3} \\
&= \sum_{Q \in \mathcal{Q}_j^0} \|\langle f_Q, \psi_\mu \rangle\|_{w\ell^{2/3}}^{2/3} + \sum_{Q \in \mathcal{Q}_j^1} \|\langle f_Q, \psi_\mu \rangle\|_{w\ell^{2/3}}^{2/3} \\
&\leq C\, |\mathcal{Q}_j^0|\, 2^{-j} + C\, |\mathcal{Q}_j^1|\, 2^{-2j}.
\end{aligned}$$

The proof is completed by observing that $|\mathcal{Q}_j^0| \leq C_0\, 2^j$, where C_0 is independent of j, and $|\mathcal{Q}_j^1| \leq 2^{2j} + 4 \cdot 2^j$. □

We can now prove Theorem 1.1

Proof of Theorem 1.1. By Corollary 1.5, we have that

$$R(j, \epsilon) = \#\{\mu \in M_j : |\langle f, \psi_\mu \rangle| > \epsilon\} \leq C\, \epsilon^{-2/3}. \tag{1.13}$$

Also, observe that, since $\hat{\psi} \in C_0^\infty(\mathbb{R}^2)$, then

$$\begin{aligned}
|\langle f, \psi_\mu \rangle| &= \left| \int_{\mathbb{R}^2} f(x)\, 2^{3j/2}\, \psi(B^\ell A^j x - k)\, dx \right| \\
&\leq 2^{3j/2}\, \|f\|_\infty \int_{\mathbb{R}^2} |\psi(B^\ell A^j x - k)|\, dx \\
&= 2^{-3j/2}\, \|f\|_\infty \int_{\mathbb{R}^2} |\psi(y)|\, dy < C'\, 2^{-3j/2}.
\end{aligned} \tag{1.14}$$

As a consequence, there is a scale j_ϵ such that $|\langle f, \psi_\mu \rangle| < \epsilon$ for each $j \geq j_\epsilon$. Specifically, it follows from (1.14) that $R(j, \epsilon) = 0$ for $j > \frac{2}{3}\,(\log_2(\epsilon^{-1}) + \log_2(C')) > \frac{2}{3}\, \log_2(\epsilon^{-1})$. Thus, using (1.13), we have that

$$\#\{\mu \in M : |\langle f, \psi_\mu \rangle| > \epsilon\} \leq \sum_{j \geq 0} R(j, \epsilon) \leq C\, \epsilon^{-2/3}\, \log_2(\epsilon^{-1}),$$

and this implies (1.12). □

2. Proofs. This section contains the constructions and proofs needed for the theorems in Section 1.4.

2.1. Proof of Theorem 1.3. Suppose that a function in $\mathcal{E}^2(A)$ contains a C^2 edge. Following the approach in [7], we suppose that, for $j > j_0$, the scale 2^{-j} is small enough so that over the square $-2^{-j} \le x_1, x_2 \le 2^{-j}$ the edge curve may be parametrized as $\begin{pmatrix} E(x_2) \\ x_2 \end{pmatrix}$, or $\begin{pmatrix} x_1 \\ E(x_1) \end{pmatrix}$ (the case where $j \le j_0$ is small requires a much simpler analysis and will be discussed in Section 2.3). Without loss of generality, let us assume that the first parametrization holds. Then an **edge fragment** is a function of the form

$$f(x_1, x_2) = w(2^j x_1, 2^j x_2)\, g(x_1, x_2)\, \chi_{\{x_1 \ge E(x_2)\}},$$

where $g \in C_0^2((0,1)^2)$. For simplicity, let us assume that the edge goes through the origin and, at this point, its tangent is vertical (see Figure 2.1). Then, using the regularity of the edge curve, we have that

(i) $E(0) = 0$, $E'(0) = 0$;

(ii) $\sup_{|x_2| \le 2^{-j}} |E(x_2)| \le \frac{1}{2} \sup_{|x_2| \le 2^{-j}} 2^{-2j} |E''(x_2)|$.

That means that, for $|x_2| \le 2^{-j}$, the edge curve is almost straight. Observe that any arbitrary edge fragment is obtained by rotating and translating the one above. Since the analysis of the edge fragment that will be presented in the following is not affected by these transformations, there is no loss of generality in considering this case only.

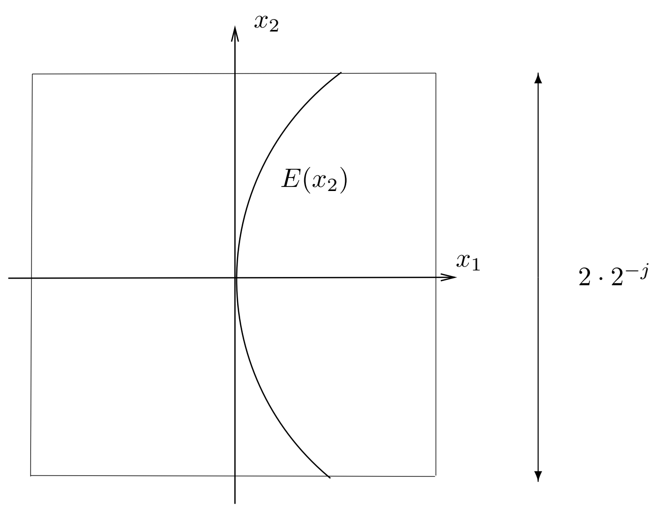


FIG. 2.1. *Representation of an edge fragment.*

In order to quantify the decay properties of the shearlet coefficients, we first need to analyze the decay of the Fourier transform of the edge fragment along radial lines in the region $\mathcal{D}_C \subset \widehat{\mathbb{R}}^2$, defined in Section 1.2. It will be convenient to introduce polar coordinates. Let $\xi = (\xi_1, \xi_2) \in \mathcal{D}_C$. Using polar coordinates, we have

$$\xi_1 = \lambda \cos\theta,\ \xi_2 = \lambda \sin\theta, \quad \text{with } |\theta| \le \frac{\pi}{4},\ \lambda \in \mathbb{R},\ |\xi_1| \ge \frac{1}{8}.$$

Using this notation, the radial lines have the form $(\lambda\cos\theta, \lambda\sin\theta)$, $\lambda \in \mathbb{R}$, $|\theta| \le \frac{\pi}{4}$.

For $\xi = (\xi_1, \xi_2) \in \mathcal{D}_C$, $j \ge 0$, $-2^j \le \ell \le 2^j$, we introduce the notation

$$\Gamma_{j,\ell}(\xi) = \hat\psi_1(2^{-2j}\xi_1)\ \hat\psi_2\left(2^j \frac{\xi_2}{\xi_1} - \ell\right). \tag{2.1}$$

We have that:

PROPOSITION 2.1. *Let f be an edge fragment and $\Gamma_{j,\ell}$ be given by (2.1). Then*

$$\int_{\mathbb{R}^2} |\hat{f}(\xi)|^2 \, |\Gamma_{j,\ell}(\xi)|^2 \, d\xi \le C \, 2^{-3j} \, (1+|\ell|)^{-5}.$$

In order to prove this proposition, we need to recall the following result [7, Theorem 6.1]:

THEOREM 2.2. *Let f be an edge fragment and I_j a dyadic interval $[2^{2j-\alpha}, 2^{2j+\beta}]$ with $\alpha \in \{0,1,2,3,4\}$, $\beta \in \{0,1,2\}$. Then, for all θ:*

$$\int_{|\lambda| \in I_j} |\hat{f}(\lambda\cos\theta, \lambda\sin\theta)|^2 \, d\lambda \le C \, 2^{-3j} \left(1 + 2^j |\sin\theta|\right)^{-5}.$$

Proof of Proposition 2.1. The assumptions on the support of $\hat{\psi}_1$ and $\hat{\psi}_2$ imply that

$$\operatorname{supp} \hat{\psi}_1(2^{-2j}\xi_1) \subset \left\{\xi_1 \in [-2^{2j-1}, -2^{2j-4}] \cup [2^{2j-4}, 2^{2j-1}]\right\}, \tag{2.2}$$

and

$$\operatorname{supp} \hat{\psi}_2(2^j \tfrac{\xi_2}{\xi_1} - \ell) \subset \left\{(\xi_1,\xi_2) : |2^j \frac{\xi_2}{\xi_1} - \ell| \le 1\right\}.$$

Since $\tan\theta = \frac{\xi_2}{\xi_1}$, the last expression can be written as

$$\operatorname{supp} \hat{\psi}_2(2^j \tfrac{\xi_2}{\xi_1} - \ell) \subset \left\{(\lambda,\theta) : \, 2^{-j}(\ell-1) \le \tan\theta \le 2^{-j}(\ell+1)\right\}. \tag{2.3}$$

Since $\lambda^2 = \xi_1^2 + \xi_2^2 = \xi_1^2\,(1+(\tan\theta)^2)$ and $|\ell| \le 2^j$, then, using (2.2) and (2.3), we have:

$$|\lambda| \le 2^{2j-1} \left(1 + 2^{-2j}(1+|\ell|)^2\right)^{\frac{1}{2}} \le 2^{2j-1} \left(1 + 2^{-2j}(1+2^j)^2\right)^{\frac{1}{2}} \le 2^{2j+1};$$

and

$$|\lambda| \ge 2^{2j-4} \left(1 + 2^{-2j}(|\ell|-1)^2\right)^{\frac{1}{2}} \ge 2^{2j-4}.$$

Thus, the support of $\Gamma_{j,\ell}$ is contained in

$$W_{j,\ell} = \{(\lambda,\theta) : \, 2^{2j-4} \le |\lambda| \le 2^{2j+1}, \, \arctan(2^{-j}(\ell-1)) \le \theta \le \arctan(2^{-j}(\ell+1))\}.$$

Observe that, in particular, $|\theta| \le \arctan 2$. Since, for $|\theta| \le 2$, we have that[1] $\tan\theta \approx \sin\theta$, it follows from (2.3) that, on $W_{j,\ell}$

$$2^j \, |\sin\theta| \approx 2^j \, 2^{-j} \, |\ell|) = |\ell|. \tag{2.4}$$

Thus, using (2.4) and Theorem 2.2, we have that

$$\begin{aligned}
\int_{\widehat{\mathbb{R}}^2} |\hat{f}(\xi)|^2 \, |\Gamma_{j,\ell}(\xi)|^2 \, d\xi &\le C \int_{W_{j,\ell}} |\hat{f}(\lambda\cos\theta, \lambda\sin\theta)|^2 \, \lambda \, d\lambda \, d\theta \\
&\le C \int_{\arctan(2^{-j}(\ell-1))}^{\arctan(2^{-j}(\ell+1))} \int_{2^{2j-4}}^{2^{2j+1}} |\hat{f}(\lambda\cos\theta, \lambda\sin\theta)|^2 \, |\lambda| \, d\lambda \, d\theta \\
&\le C \, 2^{2j+1} \int_{\arctan(2^{-j}(\ell-1))}^{\arctan(2^{-j}(\ell+1))} 2^{-4j} \left(1 + 2^j |\sin\theta|\right)^{-5} d\theta \\
&\le C \, 2^{-2j} \, (1+|\ell|)^{-5} \left(\arctan(2^{-j}(\ell-1)) - \arctan(2^{-j}(\ell+1))\right) \\
&= C \, 2^{-3j} \, (1+|\ell|)^{-5}. \quad \square
\end{aligned}$$

[1] We use the notation $f(x) \approx g(x)$, $x \in D$, to mean that there are constants $C_1, C_2 > 0$ such that $C_1 \, g(x) \le f(x) \le C_2 \, g(x)$, for all $x \in D$.

The following proposition provides a similar estimate for the partial derivatives of the Fourier transform of the edge fragment.

PROPOSITION 2.3. *Let f be an edge fragment, $\Gamma_{j,\ell}$ be given by (2.1) and L be the differential operator:*

$$L = \left(I - \left(\frac{2^{2j}}{2\pi\,(1+|\ell|)} \right)^2 \frac{\partial^2}{\partial \xi_1^2} \right) \left(1 - \left(\frac{2^j}{2\pi} \right)^2 \frac{\partial^2}{\partial \xi_2^2} \right)$$

Then

$$\int_{\widehat{\mathbb{R}}^2} \left| L \left(\hat{f}(\xi)\, \Gamma_{j,\ell}(\xi) \right) \right|^2 d\xi \le C\, 2^{-3j}\, (1+|\ell|)^{-5}.$$

In order to prove this proposition, we need to recall the following result [7, Corollary 6.6]:

COROLLARY 2.4. *Let f be an edge fragment and I_j a dyadic interval $[2^{2j-\alpha}, 2^{2j+\beta}]$ with $\alpha \in \{0,1,2,3,4\}$, $\beta \in \{0,1,2\}$. Then, for each $m=(m_1,m_2) \in \overline{\mathbb{N}} \times \overline{\mathbb{N}}$ and for each θ,*

$$\int_{|\lambda| \in I_j} \left| \frac{\partial^{m_1}}{\partial \xi_1^{m_1}} \frac{\partial^{m_2}}{\partial \xi_2^{m_2}} \hat{f}(\lambda \cos\theta, \lambda \sin\theta) \right|^2 d\lambda \le C_m\, 2^{-2j|m|} \left(2^{-(4+2m_1)j}\, (1+2^j |\sin\theta|)^{-5} + 2^{-10j} \right),$$

where C_m is independent of j and ℓ, and $\overline{\mathbb{N}} = \mathbb{N} \cup \{0\}$

We also need the following:

LEMMA 2.5. *Let $\Gamma_{j,\ell}$ be given by (2.1). Then, for each $m=(m_1,m_2) \in \overline{\mathbb{N}} \times \overline{\mathbb{N}}$, $m_1, m_2 \in \{0,1,2\}$,*

$$\left| \frac{\partial^{m_1}}{\partial \xi_1^{m_1}} \frac{\partial^{m_2}}{\partial \xi_2^{m_2}} \Gamma_{j,\ell}(\xi_1,\xi_2) \right| \le C_m\, 2^{-(2m_1+m_2)j}\, (1+|\ell|)^{m_1},$$

where $|m| = m_1 + m_2$ and C_m is independent of j and ℓ.

Proof. We will only check the cases $m = (1,0),(0,1),(2,0),(0,2),(1,1)$. The other cases are similar.

(i) A direct computation gives

$$\frac{\partial}{\partial \xi_1} \Gamma_{j,\ell}(\xi_1,\xi_2) = 2^{-2j}\, \hat{\psi}_1'(2^{-2j}\xi_1)\, \hat{\psi}_2(2^j \tfrac{\xi_2}{\xi_1} - \ell) - 2^j\, \frac{\xi_2}{\xi_1^2}\, \hat{\psi}_1(2^{-2j}\xi_1)\, \hat{\psi}_2'(2^j \tfrac{\xi_2}{\xi_1} - \ell).$$

Since $|2^j \frac{\xi_2}{\xi_1}| \le 1$ and $|xi_1| \ge 2^{2j-4}$, then

$$\left| 2^j\, \frac{\xi_2}{\xi_1^2} \right| \le 2^{-2j} \le 2^{-2j}\, (1+|\ell|). \tag{2.5}$$

Thus, using (2.5), we have:

$$\left| \frac{\partial}{\partial \xi_1}\, \Gamma_{j,\ell}(\xi_1,\xi_2) \right| \le C\, 2^{-2j}\, (1+|\ell|). \tag{2.6}$$

(ii) For the partial derivative with respect to ξ_2 we have:

$$\frac{\partial}{\partial \xi_2} \Gamma_{j,\ell}(\xi_1,\xi_2) = \frac{2^j}{\xi_1}\, \hat{\psi}_1(2^{-2j}\xi_1)\, \hat{\psi}_2'(2^j \tfrac{\xi_2}{\xi_1} - \ell).$$

Thus, using $|\xi_1| \ge 2^{2j-4}$, we have

$$\left| \frac{\partial}{\partial \xi_2}\, \Gamma_{j,\ell}(\xi_1,\xi_2) \right| \le C\, 2^{-j}. \tag{2.7}$$

(iii) For the second partial derivative with respect to ξ_1 we have:

$$\frac{\partial^2}{\partial \xi_1^2}\Gamma_{j,\ell}(\xi_1,\xi_2) = 2^{-2j}\left(2^{-2j}\,\hat\psi_1''(2^{-2j}\xi_1)\,\hat\psi_2(\tfrac{2^j\xi_2}{\xi_1}-\ell) - 2^j\,\frac{\xi_2}{\xi_1^2}\,\hat\psi_1'(2^{-2j}\xi_1)\,\hat\psi_2'(\tfrac{2^j\xi_2}{\xi_1}-\ell)\right)$$
$$- 2^j\left(-\frac{2\xi_2}{\xi_1^3}\,\hat\psi_1(2^{-2j}\xi_1)\,\hat\psi_2'(\tfrac{2^j\xi_2}{\xi_1}-\ell) + \frac{2^{-2j}\xi_2}{\xi_1^2}\,\hat\psi_1'(2^{-2j}\xi_1)\,\hat\psi_2'(\tfrac{2^j\xi_2}{\xi_1}-\ell)\right.$$
$$\left.- 2^{2j}\left(\frac{\xi_2}{\xi_1^2}\right)^2\hat\psi_1(2^{-2j}\xi_1)\,\hat\psi_2''(\tfrac{2^j\xi_2}{\xi_1}-\ell)\right).$$

Using again (2.5) and $|\xi_1| \geq 2^{2j-4}$, we have:

$$\left|\frac{\partial^2}{\partial \xi_1^2}\,\Gamma_{j,\ell}(\xi_1,\xi_2)\right| \leq C\,2^{-4j}\,(1+|\ell|)^2. \tag{2.8}$$

(iv) For the second partial derivative with respect to ξ_2 we have:

$$\frac{\partial^2}{\partial \xi_2^2}\Gamma_{j,\ell}(\xi_1,\xi_2) = \left(\frac{2^j}{\xi_1}\right)^2\hat\psi_1(2^{-2j}\xi_1)\,\hat\psi_2''(2^j\tfrac{\xi_2}{\xi_1}-\ell).$$

Thus:

$$\left|\frac{\partial^2}{\partial \xi_2^2}\,\Gamma_{j,\ell}(\xi_1,\xi_2)\right| \leq C\,2^{-2j}. \tag{2.9}$$

(v) For the mixed second partial derivative we have:

$$\frac{\partial^2}{\partial \xi_2\,\partial \xi_1}\Gamma_{j,\ell}(\xi_1,\xi_2) = \frac{2^{-j}}{\xi_1}\,\hat\psi_1'(2^{-2j}\xi_1)\,\hat\psi_2'(\tfrac{2^j\xi_2}{\xi_1}-\ell) - \frac{2^j}{\xi_1^2}\,\hat\psi_1(2^{-2j}\xi_1)\,\hat\psi_2'(\tfrac{2^j\xi_2}{\xi_1}-\ell)$$
$$+ \frac{2^{2j}\,\xi_2}{\xi_1^3}\,\hat\psi_1(2^{-2j}\xi_1)\,\hat\psi_2''(\tfrac{2^j\xi_2}{\xi_1}-\ell).$$

Thus:

$$\left|\frac{\partial^2}{\partial \xi_2\,\partial \xi_1}\,\Gamma_{j,\ell}(\xi_1,\xi_2)\right| \leq C\,2^{-3j}\,(1+|\ell|). \tag{2.10}$$

This completes the proof. □

We can now prove Proposition 2.3.

Proof of Proposition 2.3. From Corollary 2.4, using (2.4), we have:

$$\int_{2^{2j-4}}^{2^{2j+1}}\left|\frac{\partial^2}{\partial \xi_1^2}\hat f(\lambda\cos\theta,\lambda\sin\theta)\right|^2 d\lambda \leq C\,2^{-4j}\left(2^{-8j}\,(1+|\ell|)^{-5} + 2^{-10j}\right).$$

Thus, using the same idea as in the proof of Proposition 2.1:

$$\int_{\widehat{\mathbb{R}}^2}\left|\left(\frac{\partial^2}{\partial \xi_1^2}\hat f(\xi)\right)\Gamma_{j,\ell}(\xi)\right|^2 d\xi$$
$$\leq C\int_{\arctan(2^{-j}(\ell-1))}^{\arctan(2^{-j}(\ell+1))}\int_{2^{2j-4}}^{2^{2j+1}}\left|\frac{\partial^2}{\partial \xi_1^2}\hat f(\lambda\cos\theta,\lambda\sin\theta)\right|^2 |\lambda|\,d\lambda\,d\theta$$
$$\leq C\,2^{-3j}\left(2^{-8j}\,(1+|\ell|)^{-5} + 2^{-10j}\right). \tag{2.11}$$

Similarly, using Corollary 2.4 and Lemma 2.5, we have

$$\int_{\widehat{\mathbb{R}}^2} \left| \Big(\frac{\partial}{\partial \xi_1} \hat f(\xi)\Big) \Big(\frac{\partial}{\partial \xi_1} \Gamma_{j,\ell}(\xi)\Big)\right|^2 d\xi$$
$$\le C\, 2^{-4j} (1+|\ell|)^2 \int_{\arctan(2^{-j}(\ell-1))}^{\arctan(2^{-j}(\ell+1))} \int_{2^{2j-4}}^{2^{2j+1}} \left| \frac{\partial}{\partial \xi_1} \hat f(\lambda\cos\theta, \lambda\sin\theta)\right|^2 |\lambda|\, d\lambda\, d\theta$$
$$\le C\, 2^{-4j} (1+|\ell|)^2\, 2^{-j} \Big(2^{-6j}\,(1+|\ell|)^{-5} + 2^{-10j}\Big)$$
$$= C\ 2^{-5j} (1+|\ell|)^2 \Big(2^{-6j}\,(1+|\ell|)^{-5} + 2^{-10j}\Big); \tag{2.12}$$

and

$$\int_{\widehat{\mathbb{R}}^2} \left| \hat f(\xi) \Big(\frac{\partial^2}{\partial \xi_1^2} \Gamma_{j,\ell}(\xi)\Big)\right|^2 d\xi$$
$$\le C\, 2^{-8j} (1+|\ell|)^4 \int_{\arctan(2^{-j}(\ell-1))}^{\arctan(2^{-j}(\ell+1))} \int_{2^{2j-4}}^{2^{2j+1}} \left| \hat f(\lambda\cos\theta, \lambda\sin\theta)\right|^2 |\lambda|\, d\lambda\, d\theta$$
$$\le C\, 2^{-8j} (1+|\ell|)^4\, 2^{-3j}\,(1+|\ell|)^{-5} = C\, 2^{-11j}\,(1+|\ell|)^{-1}. \tag{2.13}$$

Finally, combining (2.11), (2.12), (2.13), and using the fact that $|\ell| \le 2^j$, we have that

$$\int_{\widehat{\mathbb{R}}^2} \left| \Big(\frac{2^{2j}}{2\pi(1+|\ell|)}\Big)^2 \frac{\partial^2}{\partial \xi_1^2} \Big(\hat f(\xi)\, \Gamma_{j,\ell}(\xi)\Big)\right|^2 d\xi \le C\, 2^{-3j}\,(1+|\ell|)^{-5}. \tag{2.14}$$

Similarly for the derivatives with respect to ξ_2, we have

$$\int_{\widehat{\mathbb{R}}^2} \left| \Big(\frac{\partial^2}{\partial \xi_2^2} \hat f(\xi)\Big)\, \Gamma_{j,\ell}(\xi)\right|^2 d\xi \le C\, 2^{-3j} \Big(2^{-4j}\,(1+|\ell|)^{-5} + 2^{-10j}\Big); \tag{2.15}$$

$$\int_{\widehat{\mathbb{R}}^2} \left| \Big(\frac{\partial}{\partial \xi_2} \hat f(\xi)\Big) \Big(\frac{\partial}{\partial \xi_2} \Gamma_{j,\ell}(\xi)\Big)\right|^2 d\xi \le C\, 2^{-3j} \Big(2^{-4j}\,(1+|\ell|)^{-5} + 2^{-10j}\Big); \tag{2.16}$$

$$\int_{\widehat{\mathbb{R}}^2} \left| \hat f(\xi) \Big(\frac{\partial^2}{\partial \xi_2^2} \Gamma_{j,\ell}(\xi)\Big)\right|^2 d\xi \le C\, 2^{-7j}\,(1+|\ell|)^{-5}. \tag{2.17}$$

Combining (2.15), (2.16), (2.17), and using again the fact that $|\ell| \le 2^j$, we have that

$$\int_{\widehat{\mathbb{R}}^2} \left| \Big(\frac{2^{j}}{2\pi}\Big)^2 \frac{\partial^2}{\partial \xi_2^2} \Big(\hat f(\xi)\, \Gamma_{j,\ell}(\xi)\Big)\right|^2 d\xi \le C\, 2^{-3j}\,(1+|\ell|)^{-5}. \tag{2.18}$$

Similarly, one can show that

$$\int_{\widehat{\mathbb{R}}^2} \left| \frac{2^{3j}}{(1+|\ell|)(2\pi)^2} \frac{\partial^2}{\partial \xi_2^2} \frac{\partial^2}{\partial \xi_1^2} \Big(\hat f(\xi)\, \Gamma_{j,\ell}(\xi)\Big)\right|^2 d\xi \le C\, 2^{-3j}\,(1+|\ell|)^{-5}. \tag{2.19}$$

The proof is completed using (2.14), (2.18), (2.19) and Lemma 2.5. □

We can now prove Theorem 1.3. The following proof adapts some ideas from [7].

Proof of Theorem 1.3. Fix $j \ge 0$ and, for simplicity of notation, let $f = f_Q$. For $\mu \in M_j$, the shearlet coefficient of f is

$$\langle f, \psi_\mu\rangle = \langle f, \psi_{j,\ell,k}\rangle = |\det A|^{-j/2} \int_{\widehat{\mathbb{R}}^2} \hat f(\xi)\, \Gamma_{j,\ell}(\xi)\, e^{2\pi i \xi A^{-j} B^{-\ell} k}\, d\xi,$$

where $\Gamma_{j,\ell}(\xi)$ is given by (2.1) and A, B are given by (1.4). Observe that

$$\begin{aligned} 2\pi i\xi A^{-j}B^{-\ell}k &= 2\pi i \begin{pmatrix}\xi_1 & \xi_2\end{pmatrix} \begin{pmatrix}2^{-2j} & 0\\ 0 & 2^{-j}\end{pmatrix} \begin{pmatrix}1 & -\ell\\ 0 & 1\end{pmatrix} \begin{pmatrix}k_1\\ k_2\end{pmatrix} \\ &= 2\pi i \left((k_1-k_2\ell)2^{-2j}\xi_1 + k_2 2^{-j}\xi_2\right). \end{aligned} \tag{2.20}$$

Using (2.20), a direct computation shows that

$$\begin{aligned} \frac{\partial^2}{\partial\xi_1^2}\left(2\pi\xi A^{-j}B^{-\ell}k\right) &= -(2\pi)^2\,2^{-4j}(k_1-k_2\ell)^2 = \begin{cases} -(2\pi)^2\,\ell^2\,2^{-4j}(\frac{k_1}{\ell}-k_2)^2 & \text{if } \ell\neq 0\\ -(2\pi)^2\,2^{-4j}k_1^2 & \text{if } \ell = 0\end{cases} \\ \frac{\partial^2}{\partial\xi_2^2}\left(2\pi\xi A^{-j}B^{-\ell}k\right) &= -(2\pi)^2\,2^{-2j}k_2^2. \end{aligned} \tag{2.21}$$

By the equivalent definition of weak ℓ^p norm, the theorem is proved provided we show that

$$\#\{\mu\in M_j : |\langle f,\psi_\mu\rangle| > \epsilon\} \le C\,2^{-j}\,\epsilon^{-\frac{2}{3}}. \tag{2.22}$$

Let L be the second order differential operator defined in Proposition 2.3. Using (2.20) and (2.21), it follows that

$$L\left(e^{2\pi i\xi A^{-j}B^{-\ell}k}\right) = \begin{cases} \left(1+(\frac{\ell}{(1+|\ell|)})^2(\frac{k_1}{\ell}-k_2)^2\right)(1+k_2^2)\,e^{2\pi i\xi A^{-j}B^{-\ell}k} & \text{if } \ell\neq 0\\ (1+k_1^2)(1+k_2^2)\,e^{2\pi i\xi A^{-j}B^{-\ell}k} & \text{if } \ell = 0.\end{cases} \tag{2.23}$$

Integration by parts gives

$$\langle f,\psi_\mu\rangle = |\det A|^{-j/2}\int_{\widehat{\mathbb{R}}^2} L\left(\hat f(\xi)\,\Gamma_{j,\ell}(\xi)\right)\,L^{-1}\left(e^{2\pi i\xi A^{-j}B^{-\ell}k}\right)\,d\xi.$$

Let us consider first the case $\ell\neq 0$. In this case, from (2.23) we have that

$$L^{-1}\left(e^{2\pi i\xi A^{-j}B^{-\ell}k}\right) = G(k,\ell)^{-1}\,e^{2\pi i\xi A^{-j}B^{-\ell}k}. \tag{2.24}$$

where $G(k,\ell) = \left(1+\left(\frac{\ell}{(1+|\ell|)}\right)^2\left(\frac{k_1}{\ell}-k_2\right)^2\right)(1+k_2^2)$. Thus:

$$\langle f,\psi_\mu\rangle = |\det A|^{-j/2}\,G(k,\ell)^{-1}\int_{\widehat{\mathbb{R}}^2} L\left(\hat f(\xi)\,\Gamma_{j,\ell}(\xi)\right)\,e^{2\pi i\xi A^{-j}B^{-\ell}k}\,d\xi,$$

or, equivalently:

$$G(k,\ell)\,\langle f,\psi_\mu\rangle = |\det A|^{-j/2}\int_{\widehat{\mathbb{R}}^2} L\left(\hat f(\xi)\,\Gamma_{j,\ell}(\xi)\right)\,e^{2\pi i\xi A^{-j}B^{-\ell}k}\,d\xi.$$

Let $K=(K_1,K_2)\in\mathbb{Z}^2$ and define $R_K = \{k=(k_1,k_2)\in\mathbb{Z}^2 : \frac{k_1}{\ell}\in[K_1,K_1+1],\ k_2=K_2\}$. Since, for j,ℓ fixed, the set $\{|\det A|^{-j/2}\,e^{2\pi i\xi A^{-j}B^{-\ell}k} : k\in\mathbb{Z}^2\}$ is an orthonormal basis for the L^2 functions on $[-\frac{1}{2},\frac{1}{2}]A^jB^\ell$, and the function $\Gamma_{j,\ell}(\xi)$ is supported on this set, then

$$\sum_{k\in R_K} |\langle G(k,\ell)\,f,\psi_\mu\rangle|^2 \le \int_{\widehat{\mathbb{R}}^2}\left|L\left(\hat f(\xi)\,\Gamma_{j,\ell}(\xi)\right)\right|^2 d\xi.$$

From the definition of R_K it follows that

$$\sum_{k\in R_K} |\langle f,\psi_\mu\rangle|^2 \le C\,\left(1+(K_1-K_2)^2\right)^{-2}(1+K_2)^{-2}\int_{\widehat{\mathbb{R}}^2}\left|L\left(\hat f(\xi)\,\Gamma_{j,\ell}(\xi)\right)\right|^2 d\xi.$$

By Proposition 2.3,

$$\sum_{k\in R_K} |\langle f,\psi_\mu\rangle|^2 \le C\, L_K^{-2}\, 2^{-3j}(1+|\ell|)^{-5}, \tag{2.25}$$

where $L_K = \left(1+(K_1-K_2)^2\right)\left(1+K_2^2\right)$. For j,ℓ fixed, let $N_{j,\ell,K}(\epsilon) = \#\{k\in R_K : |\psi_{j,\ell,k}| > \epsilon\}$. Then it is clear that $N_{j,\ell,K}(\epsilon) \le C\,(|\ell|+1)$ and (2.25) implies that

$$N_{j,\ell,K}(\epsilon) \le C\, L_K^{-2}\, 2^{-3j}\, \epsilon^{-2}(1+|\ell|)^{-5}.$$

Thus

$$N_{j,\ell,K}(\epsilon) \le C\, \min\left(|\ell|+1, L_K^{-2}\, 2^{-3j}\, \epsilon^{-2}(1+|\ell|)^{-5}\right). \tag{2.26}$$

Using (2.26) we will now show that:

$$\sum_{\ell=-2^j}^{2^j} N_{j,\ell,K}(\epsilon) \le C\, L_K^{-\frac{2}{3}}\, 2^{-j}\, \epsilon^{-\frac{2}{3}}. \tag{2.27}$$

In fact, let ℓ^* be defined by $(\ell^*+1) = L_K^{-2}\, 2^{-3j}\, \epsilon^{-2}(1+\ell^*)^{-5}$. That is, $\ell^*+1 = L_K^{-1/3}\, 2^{-j/2}\, \epsilon^{-1/3}$. Then

$$\begin{aligned}\sum_{\ell=-2^j}^{2^j} N_{j,\ell,K}(\epsilon) &\le \sum_{|\ell|\le(\ell^*+1)} N_{j,\ell,K}(\epsilon) + \sum_{|\ell|>(\ell^*+1)} N_{j,\ell,K}(\epsilon)\\ &\le \sum_{|\ell|\le(\ell^*+1)} (|\ell|+1) + \sum_{|\ell|>(\ell^*+1)} L_K^{-2}\, 2^{-3j}\, \epsilon^{-2}(1+|\ell|)^{-5}\\ &\le (\ell^*+1)^2 + C\, L_K^{-2}\, 2^{-3j}\, \epsilon^{-2}(1+\ell^*)^{-4} \le C\,(\ell^*+1)^2,\end{aligned}$$

which gives (2.27).

Since $\sum_{K\in\mathbb{Z}^2} L_K^{-\frac{2}{3}} < \infty$, using (2.27) we then have that

$$\#\{\mu\in M_j : |\langle f,\psi_\mu\rangle| > \epsilon\} \le \sum_{K\in\mathbb{Z}^2}\sum_{\ell=-2^j}^{2^j} N_{j,\ell,K}(\epsilon) \le C\, 2^{-j}\, \epsilon^{-\frac{2}{3}} \sum_{K\in\mathbb{Z}^2} L_K^{-\frac{2}{3}} \le C\, 2^{-j}\, \epsilon^{-\frac{2}{3}},$$

and, thus, (2.22) holds.

The case $\ell = 0$ is similar. Indeed, in this case

$$L^{-1}\left(e^{2\pi i\xi A^{-j}B^{-\ell}k}\right) = (1+k_1^2)^{-1}(1+k_2^2)^{-1}\, e^{2\pi i\xi A^{-j}B^{-\ell}k},$$

and we can proceed as in the case $\ell\neq 0$, with $L_K = (1+K_1^2)\,(1+K_2^2)$. It is clear that also in this case $\sum_{K\in\mathbb{Z}^2} L_K^{-\frac{2}{3}} < \infty$. This completes the proof of the theorem. □

2.2. Proof of Theorem 1.4. In order to prove Theorem 1.4, the following lemmata will be useful.

Lemma 2.6. *Let $f = g\, w_Q$, where $g\in\mathcal{E}^2(A)$ and $Q\in\mathcal{Q}_j^1$. Then*

$$\int_{W_{j,\ell}} |\hat f(\xi)|^2\, d\xi \le C\, 2^{-10j}. \tag{2.28}$$

Proof. The following proof follows [7, Lemma 8.1] and is reported here for completeness. The function f belongs to $C_0^2(\mathbb{R}^2)$ and its second partial derivative with respect to x_1 is

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{\partial^2 g}{\partial x_1^2}\, w_Q + 2\,\frac{\partial\, g}{\partial x_1}\,\frac{\partial\, w_Q}{\partial x_1} + f\,\frac{\partial^2 w_Q}{\partial x_1^2} = h_1 + h_2 + h_3.$$

Using the fact that w_Q is supported in a square of sidelength $2 \cdot 2^{-j}$, we have

$$\int_{\widehat{\mathbb{R}}^2} |\hat{h}_1(\xi)|^2\, d\xi = \int_{\mathbb{R}^2} |h_1(x)|^2 \; dx \le C\, 2^{-2j}.$$

Next, observe that $\|\frac{\partial}{\partial x_1} h_2\|_\infty \le C\, 2^{2j}$. Using again the condition on the support of w_Q it follows that

$$\int_{\widehat{\mathbb{R}}^2} |2\pi\xi_1\, \hat{h}_2(\xi)|^2\, d\xi = \int_{\mathbb{R}^2} \left|\frac{\partial}{\partial x_1} h_2(x)\right|^2 \; dx \le C\, 2^{2j},$$

and thus, for $\xi \in W_{j,\ell}$ (hence $\xi_1 \approx 2^{2j}$),

$$\int_{W_{j,\ell}} |\hat{h}_2(\xi)|^2\, d\xi \le C\, 2^{-2j}.$$

Finally, observing that $\|\frac{\partial^2}{\partial x_1^2} h_3\|_\infty \le C\, 2^{4j}$, a similar computation to the one above shows that

$$\int_{W_{j,\ell}} |\hat{h}_3(\xi)|^2\, d\xi \le C\, 2^{-2j}.$$

Since $-(2\pi)^2\, \xi_1^2\, \hat{f}(\xi) = \hat{h}_1(\xi) + \hat{h}_2(\xi) + \hat{h}_3(\xi)$, it follows from the estimates above that

$$\int_{W_{j,\ell}} |\hat{f}(\xi)|^2\, d\xi \le C\, 2^{-10j}.$$

This completes the proof. ∎

Lemma 2.7. *Let $m = (m_1, m_2) \in \overline{\mathbb{N}} \times \overline{\mathbb{N}}$, $\xi = (\xi_1, \xi_2) \in \widehat{\mathbb{R}}^2$ and $\Gamma_{j,\ell}$ be given by (2.1). Then*

$$\sum_{\ell=-2^j}^{2^j} \left| \frac{\partial^{m_1}}{\partial \xi_1^{m_1}}\, \frac{\partial^{m_2}}{\partial \xi_2^{m_2}}\, \Gamma_{j,\ell}(\xi) \right|^2 \le C_m\, 2^{-2|m|j},$$

where C_m is independent of j and ξ and $|m| = m_1 + m_2$.

Proof. Observe that $W_{j,\ell} \cap W_{j,\ell+\ell'} = \emptyset$, whenever $|\ell'| \ge 3$. Since $|\ell| \le 2^j$, the lemma then follows from Lemma 2.5. ∎

Lemma 2.8. *Let $f = g\, w_Q$, where $g \in \mathcal{E}^2(A)$ and $Q \in \mathcal{Q}_j^1$. Define*

$$T = \left(I - \frac{2^j}{(2\pi)^2}\, \Delta \right), \tag{2.29}$$

where $\Delta = \frac{\partial^2}{\partial \xi_1^2} + \frac{\partial^2}{\partial \xi_2^2}$. Then

$$\int_{\widehat{\mathbb{R}}^2} \sum_{\ell=-2^j}^{2^j} \left| T^2 \left(\hat{f}\, \Gamma_{j,\ell} \right)(\xi) \right|^2 \, d\xi \le C\, 2^{-10j}.$$

Proof. Observe that, for $N \in \overline{\mathbb{N}}$,

$$\Delta^N \left(\hat{f}\,\Gamma_{j,\ell}\right) = \sum_{|\alpha|+|\beta|=2N} C_{\alpha,\beta} \left(\frac{\partial^{\alpha_1}}{\partial\xi_1^{\alpha_1}}\,\frac{\partial^{\alpha_2}}{\partial\xi_2^{\alpha_2}}\,\hat{f}\right) \left(\frac{\partial^{\beta_1}}{\partial\xi_1^{\beta_1}}\,\frac{\partial^{\beta_2}}{\partial\xi_2^{\beta_2}}\,\Gamma_{j,\ell}\right).$$

Then, using Lemma 2.7, we have that

$$\int_{\widehat{\mathbb{R}}^2} \sum_{\ell=-2^j}^{2^j} \left|\frac{\partial^{\alpha_1}}{\partial\xi_1^{\alpha_1}}\,\frac{\partial^{\alpha_2}}{\partial\xi_2^{\alpha_2}}\,\hat{f}(\xi)\right|^2 \left|\frac{\partial^{\beta_1}}{\partial\xi_1^{\beta_1}}\,\frac{\partial^{\beta_2}}{\partial\xi_2^{\beta_2}}\,\Gamma_{j,\ell}(\xi)\right|^2 d\xi \le C_\beta\, 2^{-2|\beta| j} \int_{W_{j,\ell}} \left|\frac{\partial^{\alpha_1}}{\partial\xi_1^{\alpha_1}}\,\frac{\partial^{\alpha_2}}{\partial\xi_2^{\alpha_2}}\,\hat{f}(\xi)\right|^2 d\xi$$

Recall that $f(x)$ is of the form $g(x)\,w(2^j x)$. It follows that $x^\alpha\, f(x) = 2^{-j|\alpha|}\, g(x)\, w_\alpha(2^j x)$, where $w_\alpha(x) = x^\alpha w(x)$. By Lemma 2.6, $g(x)\, w_\alpha(2^j x)$ obeys the estimate (2.28). Thus, observing that $\frac{\partial^{\alpha_1}}{\partial\xi_1^{\alpha_1}}\,\frac{\partial^{\alpha_2}}{\partial\xi_2^{\alpha_2}}\,\hat{f}(\xi)$ is the Fourier transform of $(-2\pi i x)^\alpha f(x)$, we have that

$$\int_{W_{j,\ell}} \left|\frac{\partial^{\alpha_1}}{\partial\xi_1^{\alpha_1}}\,\frac{\partial^{\alpha_2}}{\partial\xi_2^{\alpha_2}}\,\hat{f}(\xi)\right|^2 d\xi \le C_\alpha\, 2^{-2j|\alpha|}\, 2^{-10j}.$$

Combining the estimates above we have that, for each α, β with $|\alpha| + |\beta| = 2N$,

$$\int_{\widehat{\mathbb{R}}^2} \sum_{\ell=-2^j}^{2^j} \left|\frac{\partial^{\alpha_1}}{\partial\xi_1^{\alpha_1}}\,\frac{\partial^{\alpha_2}}{\partial\xi_2^{\alpha_2}}\,\hat{f}(\xi)\right|^2 \left|\frac{\partial^{\beta_1}}{\partial\xi_1^{\beta_1}}\,\frac{\partial^{\beta_2}}{\partial\xi_2^{\beta_2}}\,\Gamma_{j,\ell}(\xi)\right|^2 d\xi \le C_{\alpha,\beta}\, 2^{-10j}\, 2^{-4jN}. \tag{2.30}$$

Since $T^2 = 1 - 2\,\frac{2^j}{(2\pi)^2}\,\Delta + \frac{2^{2j}}{(2\pi)^4}\,\Delta^2$, the lemma now follows from (2.30) and Lemma 2.7. □

We can now prove Theorem 1.4.

Proof of Theorem 1.4.

Using (2.21), for T given by (2.29), we have that

$$T\left(e^{2\pi i \xi A^{-j} B^{-\ell} k}\right) = \left(1 + 2^{-2j}(k_1 - k_2\,\ell)^2 + k_2^2\right)\, e^{2\pi i \xi A^{-j} B^{-\ell} k} \tag{2.31}$$

Fix $j \ge 0$ and let $f = f_Q$. Then, using integration by parts as in the proof of Theorem 1.3, from (2.31) it follows that

$$\langle f, \psi_\mu\rangle = |\det A|^{-j}\left(1 + 2^{-2j}(k_1 - k_2\,\ell)^2 + k_2^2\right)^{-2} \int_{\widehat{\mathbb{R}}^2} T^2\left(\hat{f}\,\Gamma_{j,\ell}\right)(\xi)\, e^{2\pi i \xi A^{-j} B^{-\ell} k}\, d\xi.$$

Let $K = (K_1, K_2) \in \mathbb{Z}^2$ and R_K be the set $\{(k_1, k_2) \in \mathbb{Z}^2 : k_2 = K_2, 2^{-j}(k_1 - K_2\ell) \in [K_1, K_1 + 1]\}$. Observing that, for each K, there are only $1 + 2^j$ choices for k_1 in R_K, it follows that the number of terms in R_K is bounded by $1 + 2^j$. Thus, arguing again as in the proof of Theorem 1.3, we have that

$$\sum_{k\in R_K} |\langle f, \psi_\mu\rangle|^2 \le C\, \left(1 + K_1^2 + K_2^2\right)^{-4} \int_{\widehat{\mathbb{R}}^2} \left|T^2\left(\hat{f}\,\Gamma_{j,\ell}\right)(\xi)\right|^2 d\xi.$$

From this inequality, using Lemma 2.8, we have that

$$\begin{aligned} \sum_{\ell=-2^j}^{2^j} \sum_{k\in R_K} |\langle f, \psi_\mu\rangle|^2 &\le C\,(1 + K^2)^{-4} \int_{\widehat{\mathbb{R}}^2} \sum_{\ell=-2^j}^{2^j} \left|T^2\left(\hat{f}\,\Gamma_{j,\ell}\right)(\xi)\right|^2 d\xi \\ &\le C\,(1 + K^2)^{-4}\, 2^{-10j}. \end{aligned} \tag{2.32}$$

For any $N \in \mathbb{N}$, provided $\frac{1}{2} < p < 2$, the Hölder inequality yields:

$$\sum_{m=1}^{N} |a_m|^p \le \left(\sum_{m=1}^{N} |a_m|^2 \right)^{\frac{p}{2}} N^{(1-\frac{p}{2})}. \tag{2.33}$$

Since the cardinality of R_K is bounded by $1+2^j$, it follows from (2.32) and (2.33) that, for $\frac{1}{2} < p < 2$,

$$\sum_{\ell=-2^j}^{2^j} \sum_{k \in R_K} |\langle f, \psi_\mu \rangle|^p \le C \, \left(2^{2j}\right)^{(1-\frac{p}{2})} (1+K^2)^{-2p} \, 2^{-5pj}.$$

Thus, since $p > \frac{1}{2}$,

$$\sum_{\mu \in M_j} |\langle f, \psi_\mu \rangle|^p \le C \, 2^{\left(2j(1-\frac{p}{2})-5pj\right)} \sum_{K \in \mathbb{Z}^2} (1+K^2)^{-2p} \le C \, 2^{(2-3p)j},$$

and, in particular

$$\|\langle f, \psi_\mu \rangle\|_{\ell^{2/3}} \le C \, 2^{-3j}. \qquad \square$$

2.3. Coarse Scale Analysis. In Section 2.1, we assumed that the scale parameter j is large enough. The situation where j is small can be treated in a much simpler way. In fact, if f_Q is an edge fragment, then a trivial estimate shows that

$$\|f_Q\|_2 = \left(\int_{\operatorname{supp} w_Q} |f_Q(x)|^2 \, dx \right)^{1/2} \le C \, |\operatorname{supp} w_Q| = C \, 2^{-j}.$$

It follows that $\|\langle f_Q, \psi_\mu \rangle\|_{\ell^2} \le C \, 2^{-j}$ and, thus, by Hölder inequality,

$$\|\langle f_Q, \psi_\mu \rangle\|_{\ell^{2/3}} \le C \, 2^{j}.$$

This satisfies Theorem 1.3 for j small.

2.4. Additional Remarks.

- In order to define the collection of shearlets, in Section 1.2 we have constructed a function $\hat{\psi} \in C_0^\infty$. This property allows us to obtained a collection of elements that are well localized. Observe, however, that we only need $\hat{\psi} \in C_0^2$ in order to prove all the results presented in this paper.
- In this paper, we have considered the representation of functions containing a discontinuity along a C^2 curve. More generally, we can consider the situation where a function f contains many edge curves of this type, exhibiting finitely many junctions or corners between them. In this setting, the discontinuity curve is not globally C^2 but only piecewice C^2. The results reported in this paper, namely Theorems 1.1 and 1.2, extend to this setting as well. We refer to [7] for a similar discussion in the case of curvelets.
- The assumption we made about the regularity of the discontinuity curve plays a critical role in our construction. If the discontinuity curve is in C^α, with $\alpha > 2$, then our argument still works and we can still prove Theorem 1.2. This result, however, is not (essentially) optimal as in the case $\alpha = 2$. On the other hand, if the discontinuity curve is in C^α, with $\alpha < 2$, then the estimate given by Theorem 1.2 does not hold and the estimate could be worse, in general. We refer to [25] for additional observations about this fact, and for an alternative approach, based on an adaptive construction, to the sparse representation of functions with edges.

- There are natural ways of extending the shearlets to dimensions larger than 2. We refer to [19] for a discussion of these extensions, as well as the extensions of the shear transformations to the general multidimensional setting. For example, in dimension 3, let $A = \begin{pmatrix} 4 & 0 \\ 0 & 2\,I_2 \end{pmatrix}$, define the shear matrices $\{S_k = \begin{pmatrix} 1 & k \\ 0 & I_2 \end{pmatrix} : k \in \mathbb{Z}^2\}$, where I_2 is the 2×2 identity matrix, $0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, and, for $\xi = (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3$, define ψ by

$$\hat{\psi}(\xi) = \hat{\psi}_1(\xi_1)\, \hat{\psi}_2\Big(\frac{\xi_2}{\xi_1}\Big)\, \hat{\psi}_2\Big(\frac{\xi_3}{\xi_1}\Big),$$

where ψ_1 and ψ_2 are given as in the 2–D case. Then, similarly to their 2–D counterpart, one can construct a Parseval frame of well-localized 3–D shearlets

$$\{\psi_{j,\ell,k} = |\det A|^{-j/2}\, \psi(S_\ell\, A^{-j} x - k) :\, j \in \mathbb{Z}, \ell \in \mathbb{Z}^2, k \in \mathbb{Z}^2\},$$

with frequency support on a parallelepiped of approximate size $2^{2j} \times 2^j \times 2^j$, at various scales j, with orientations controlled by the two–dimensional index ℓ and spatial location k. Then, using an heuristic argument, one can argue that these systems provide sparse representations for 3–dimensional functions f that are smooth away from 'nice' surface discontinuities of finite area. In fact, thanks to their frequency support and their localization properties, the elements $\psi_{j,\ell,k}$, at scale j, are essentially supported on a parallelepiped of size $2^{-2j} \times 2^{-j} \times 2^{-j}$, with location controlled by k, and orientation controlled by ℓ. Thus, there are at most $O(2^{2j})$ significant shearlet coefficients $\mathcal{SH}_{j,\ell,k}(f) = \langle f, \psi_{j,\ell,k}\rangle$, and they are bounded by $C\, 2^{-2j}$. This implies that the N–th largest 3–D shearlet coefficient $|\mathcal{SH}_N(f)|$ is bounded by $O(N^{-1})$ and, thus, if f is approximated by taking the N largest coefficients in the 3–D shearlets expansion, the L^2–error would approximately obey:

$$\|f - f_N^S\|_{L^2}^2 \le \sum_{\ell > N} |\mathcal{SH}_\ell(f)|^2 \le C\, N^{-1},$$

up to lower order factors. A rigorous proof of this fact will be presented elsewhere.

Appendix A. Construction of ψ_1, ψ_2. In this section we show how to construct examples of functions ψ_1, ψ_2 satisfying the properties described in Section 1.2.

In order to construct ψ_1, let $h(t)$ be an even C_0^∞ function, with support in $(-\frac{1}{6}, \frac{1}{6})$, satisfying $\int_{\mathbb{R}} h(t)\, dt = \frac{\pi}{2}$, and define $\theta(\omega) = \int_{-\infty}^{\omega} h(t)\, dt$. Then one can construct a smooth bell function as

$$b(\omega) = \begin{cases} \sin\left(\theta\left(|\omega| - \frac{1}{2}\right)\right) & \text{if } \frac{1}{3} \le |\omega| \le \frac{2}{3}, \\ \sin\left(\frac{\pi}{2} - \theta\left(\frac{|\omega|}{2} - \frac{1}{2}\right)\right) & \text{if } \frac{2}{3} < |\omega| \le \frac{4}{3}, \\ 0 & \text{otherwise.} \end{cases}$$

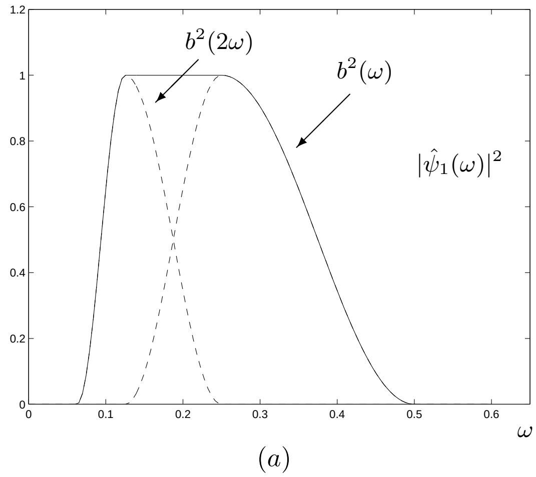
It follows from the assumptions we made (cf. [21, Sec.1.4]) that

$$\sum_{j=-1}^{\infty} b^2(2^{-j}\omega) = 1 \quad \text{for } |\omega| \ge \frac{1}{3}.$$

Now letting $u^2(\omega) = b^2(2\omega) + b^2(\omega)$, it follows that

$$\sum_{j \ge 0}^{\infty} u^2(2^{-2j}\omega) = \sum_{j=-1}^{\infty} b^2(2^{-j}\omega) = 1 \quad \text{for } |\omega| \ge \frac{1}{3}.$$

Finally, let ψ_1 be defined by $\hat{\psi}_1(\omega) = u(\frac{8}{3}\omega)$. Then $\operatorname{supp} \hat{\psi}_1 \subset [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$ and equation (1.6) is satisfied. This construction is illustrated in Figure A.1(a).



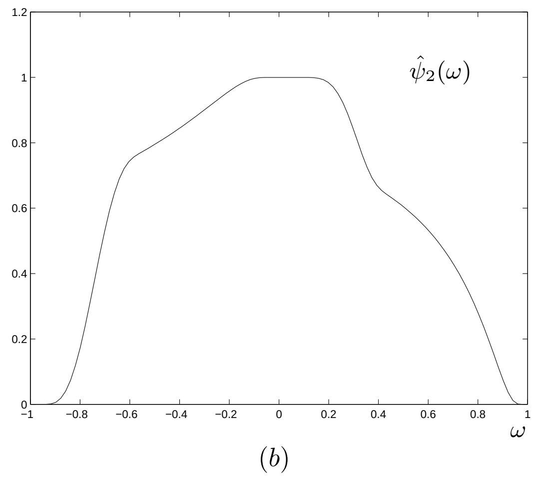


FIG. A.1. *(a) The function $|\hat{\psi}_1(\omega)|^2$ (solid line), for $\omega > 0$; the negative side is symmetrical. This function is obtained, after rescaling, from the sum of the window functions $b^2(\omega) + b^2(2\omega)$ (dashed lines). (b) The function $\hat{\psi}_2(\omega)$.*

For the construction of ψ_2, we start by considering a smooth bump function $f_1 \in C_0^\infty(-1, \ 1)$ such that $0 \le f_1 \le 1$ on $(-1, 1)$ and $f_1 = 1$ on $[-\frac{1}{2}, \frac{1}{2}]$ (cf. [22, Sec. 1.4]). Next, let $f_2(t) = \sqrt{1 - e^{\frac{1}{t}}}$. Then (in the left limit sense) $f_2(0) = 1$, $f_2^{(k)}(0) = 0$, for $k \ge 1$ and $0 < f_2 < 1$ on $(-1, 0)$. Define $f(t) = f_1(t) f_2(t)$, for $t \in [-1, 0]$. It is then easy to see that $f^{(k)}(-1) = 0$, for $k \ge 0$, and $f(0) = 1$, $f^{(k)}(0) = 0$, for $k \ge 1$. Since $g(t) = e^{\frac{1}{2(t-1)}}$, for $t \in (\frac{1}{2}, 1)$, it follows that $\lim_{t \to 1-} g^{(k)}(t) = 0$, for $k \ge 0$. Finally, we define

$$\hat{\psi}_2(\omega) = \begin{cases} f(\omega) & \text{if } \omega \in [-1, 0), \\ g(\omega) & \text{if } \omega \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

Then $\hat{\psi}_2 \in C_0^\infty(\mathbb{R})$, with $\operatorname{supp} \hat{\psi}_2 \subset [-1, 1]$, and

$$\hat{\psi}_2^{\,2}(\omega) + \hat{\psi}_2^{\,2}(\omega - 1) = 1, \quad \omega \in [0, 1].$$

The last equality implies (1.7). The function $\hat{\psi}_2$ is illustrated in Figure A.1(b).

Acknowledgments. The authors thank E. Candès, P. Gremaud, G. Kutyniok, W. Lim, G. Weiss and E. Wilson for useful discussions. DL acknowledges support from a FR&PD grant from NCSU.

REFERENCES

[1] J. P. Antoine, R. Murenzi, and P. Vandergheynst, Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns, Appl. Computat. Harmon. Anal. **6**(1999), pp. 314-345.

[2] R. H. Bamberger, and M. J. T. Smith, A filter bank for directional decomposition of images: theory and design, IEEE Trans. Signal Process. **40**(1992), pp. 882–893.

[3] C. M. Brislawn, Fingerprints go digital, Notices Amer. Math. Soc. **42**(1995), pp. 1278-1283.

[4] C. M. Brislawn, and M. D. Quirk, Image compression with the JPEG-2000 standard, in *Encyclopedia of Optical Engineering*, R. G. Driggers (Ed.), Marcel Dekker, New York, 2003, pp. 780-785.

[5] E. J. Candès and D. L. Donoho, Ridgelets: a key to higher–dimensional intermittency?, Phil. Trans. Royal Soc. London A **357**(1999), pp. 2495–2509.

[6] E. J. Candès and D.L. Donoho, Curvelets – a surprising effective nonadaptive representation for objects with edges, in: Curves and Surfaces, C. Rabut, A. Cohen and L.L. Schumaker (Eds.), Vanderbilt University Press, Nashville, 2000.

[7] E. J. Candès and D. L. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities, Comm. Pure and Appl. Math. **56**(2004), pp. 216–266.

[8] R. R. Coifman, and F. G. Meyer, Brushlets: a tool for directional image analysis and image compression, Appl. Comp. Harmonic Anal. **5**(1997), pp. 147–187.

[9] S. Dahlke, G. Kutyniok, P. Maass, C. Sagiv, and H.-G. Stark, The Uncertainty Principle Associated with the Continuous Shearlet Transform, preprint (2006).

[10] I. Daubechies, Ten Lectures on Wavelets, CBMS–NSF Regional Conference Series in Applied mathematics, Vol.61, SIAM, Philadelphia, 1992.

[11] R. A. DeVore, Nonlinear approximation, in: Acta Numerica, A. Iserles (Ed.), Cambridge University Press, Cambridge, 1998, pp. 51-150.

[12] M. N. Do and M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, to appear in: IEEE Transactions Image on Processing, 2006.

[13] D. L. Donoho, Sparse components of images and optimal atomic decomposition, Constr. Approx. **17**(2001), pp. 353–382.

[14] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, Data compression and harmonic analysis, IEEE Trans. Inform. Th. **44**(1998), pp. 2435–2476.

[15] G. Easley, D. Labate, and W-Q. Lim Sparse Directional Image Representations using the Discrete Shearlet Transform, preprint (2006).

[16] K. Guo, G. Kutyniok, and D. Labate, *Sparse Multidimensional Representations using Anisotropic Dilation and Shear Operators,* in: Wavelets and Splines, G. Chen and M. Lai (eds.), Nashboro Press, Nashville, TN (2006), pp. 189–201.

[17] K. Guo, W-Q. Lim, D. Labate, G. Weiss and E. Wilson, Wavelets with composite dilations, Electr. Res. Announc. of AMS **10**(2004), pp. 78–87 .

[18] K. Guo, W-Q. Lim, D. Labate, G. Weiss and E. Wilson, The theory of wavelets with composite dilations, in: Harmonic Analysis and Applications, C. Heil (ed.), Birkhäuser, Boston, MA, 2006.

[19] K. Guo, W-Q. Lim, D. Labate, G. Weiss and E. Wilson, Wavelets with composite dilations and their MRA properties, Appl. Computat. Harmon. Anal. **20**(2006), pp. 231–249.

[20] JPEG 2000 Image Coding System, Part 1, ser. ISO/IEC Intl. Standard 15444-1, ITU-T Rec. T.800. Intl. Org. for Standardization, Dec. 2000.

[21] E. Hernández, and G. Weiss, A first course on wavelets, CRC Press, Boca Raton, FL, 1996.

[22] L. Hörmander, The analysis of linear partial differential operators. I. Distribution theory and Fourier analysis. Springer-Verlag, Berlin, 2003.

[23] N. Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, Appl. Computat. Harmon. Anal. **10**(2001), pp. 234-253.

[24] D. Labate, W.-Q. Lim, G. Kutyniok and G. Weiss, Sparse multidimensional representation using shearlets, Wavelets XI (San Diego, CA, 2005), SPIE Proc., 5914, SPIE, Bellingham, WA, 2005.

[25] E. Le Pennec, and S. Mallat, Sparse geometric image representations with bandelets, IEEE Trans. Image Process. **14**(2005), pp. 423–438.

[26] W-Q. Lim, Wavelets with Composite Dilations, Ph.D. Thesis, Washington University in St. Louis, St. Louis, MO, 2006.

[27] Y. Lu, and M. N. Do, A directional extension for multidimensional wavelet transforms, to appear in IEEE Transactions on Image Processing (2006).

[28] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, CA, 1998.

[29] D. D. Po and M. N. Do, Directional multiscale modeling of images using the contourlet transform, to appear in: IEEE Transactions Image on Processing, 2006.

[30] E.M. Stein, G. Weiss, Introduction to Fourier Analysis on Euclidean Spaces, Princeton University Press, Princeton, NJ, 1970.