OPTIMALLY SPARSE 3D APPROXIMATIONS USING SHEARLET REPRESENTATIONS

KANGHUI GUO AND DEMETRIO LABATE

ABSTRACT. This paper introduces a new Parseval frame, based on the 3–D shearlet representation, which is especially designed to capture geometric features such as discontinuous boundaries with very high efficiency. We show that this approach exhibits essentially optimal approximation properties for 3–D functions f which are smooth away from discontinuities along C^2 surfaces. In fact, the N term approximation f_N^S obtained by selecting the N largest coefficients from the shearlet expansion of f satisfies the asymptotic estimate

$$\|f - f_N^S\|_2^2 \asymp N^{-1}(\log N)^2, \qquad \text{as } N \to \infty.$$

Up to the logarithmic factor, this is the optimal behavior for functions in this class and significantly outperforms wavelet approximations, which only yields a $N^{-1/2}$ rate. Indeed, the wavelet approximation rate was the best published nonadaptive result so far and the result presented in this paper is the first nonadaptive construction which is provably optimal (up to a loglike factor) for this class of 3D data.

Our estimate is consistent with the corresponding 2–D (essentially) optimally sparse approximation results obtained by the authors using 2–D shearlets and by Candès and Donoho using curvelets.

1. INTRODUCTION

In a seminal paper published in 2004 [1], Candès and Donoho proved a remarkable result about sparse representations of 2-dimensional data, showing that the curvelet representation, a multiscale system of waveforms defined at various directions and positions at each scale, is essentially as good as an adaptive representation from the point of view of its ability to approximate images containing edges. Specifically, for functions f which are C^2 away from C^2 edges, the N term approximation f_N^C obtained from the N largest coefficients of its curvelet expansion, obeys

$$(1.1) \qquad \|f - f_N^C\|_2^2 \asymp N^{-2}(\log N)^3, \qquad \text{as } N \to \infty.$$

Ignoring the loglike factor, this is the optimal approximation rate for this class of functions while, in comparison, the wavelet and Fourier representations only achieves approximation rate N^{-1} and $N^{-1/2}$, respectively.

The work by Candès and Donoho was motivated by fundamental theoretical questions about the mathematical representations of functions containing edge discontinuities, and has great relevance for a variety of technologies and applications. In fact, the notion of sparsity has implications going far beyond approximation theory, since it entails the intimate understanding of the most essential information contained in data, which is critically important for the development of improved algorithms in areas such as data modeling, feature extraction, image denoising and classification [6]. Inspired in part by this work, the shearlet representation was introduced by the authors of this paper and their collaborators in [13, 9] as an alternative approach which satisfies the same (essentially) optimally sparse approximation rate (1.1) when dealing with the same class of 2–D data. Similarly to the curvelet construction, the elements of the shearlet system form a multiscale pyramid of waveforms defined at various directions and positions and satisfying parabolic scaling. However, the shearlet approach relies on a different mathematical framework, based on the structure of affine systems, so that all elements of the representation system are derived from a single (or finite set of) generators through the action of the affine group. The unique properties of the shearlet approach provide

not only the benefit of greater flexibility and mathematical simplicity, but also ensure that there is a natural transition from the continuum to the discrete setting [17, 8]. This has been exploited in a wide range of very competitive applications such as those presented in [2, 5, 7, 8, 11, 16, 20]. Also notice that, very recently, a novel construction which uses compactly supported shearlet analyzing functions [15], was shown to provide optimally sparse 2D representations as in (1.1).

In this paper, we show that the shearlet approach extends naturally to the 3-dimensional setting where it also provides optimally sparse nonadaptive representations of 3D data. In fact, we construct a Parseval frame of shearlets to represent 3-dimensional functions f which are smooth away from discontinuities along C^2 boundaries, and prove that the N-term approximation f_N^S, obtained from the N largest coefficients of its shearlet representation, satisfies the estimate:

$$(1.2) \qquad \|f - f_N^S\|_2^2 \asymp N^{-1}(\log N)^2, \qquad \text{as } N \to \infty.$$

Up the logarithmic factor, this is the optimal approximation rate for this type of functions [4] in the sense that no orthonormal bases or Parseval frames can yield approximation rates that are better than N^{-1}. Even if one considers finite linear combinations of elements taking from arbitrary dictionaries, there is no depth-limited search dictionary that can achieve a rate better than N^{-1} [4]. In contrast, more traditional methods based on wavelet and Fourier approximations are significantly less efficient since their asymptotic approximation rate only decays as $N^{-1/2}$ and $N^{-1/3}$, respectively [6, 18]. Notice that the result presented in this paper is the first nonadaptive construction which is provably optimal (up to a loglike factor) for a large class of 3D data.
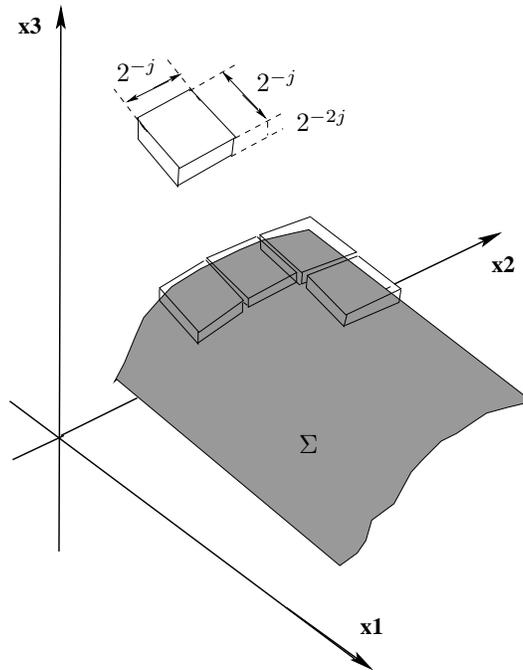
FIGURE 1. *At scale 2^{-2j}, the number of shearlet elements which are tangent to a bounded surface Σ are approximately $2^j \times 2^j = 2^{2j}$.*

1.1. **Significance.** As in the 2–D case, a simple heuristic argument can be used to justify why a 3–D wavelet system cannot yield an approximation rate better than $N^{-1/2}$, in general, when dealing with 3–D data f containing discontinuous surfaces, while a 3-D shearlet system is expected to provide a much sparser representation. Indeed, at scale 2^{-2j}, a wavelet $\phi_{j,k}(x) = 2^{3j}\phi(2^{2j}x - k)$ is essentially supported on a box

of size $2^{-2j} \times 2^{-2j} \times 2^{-2j}$. Hence, there are approximately $O(2^{4j})$ wavelet coefficients $C_{j,k}(f) = \langle f, \phi_{j,k} \rangle$ associated with the surface of discontinuity (while the remaining coefficients are negligible at fine scales). Since

$$\int_{\mathbb{R}^3} |\phi_{j,k}(x)|\, dx = 2^{3j} \int |\psi(2^{2j}x - k)|\, dx = 2^{-3j} \int_{\mathbb{R}^3} |\phi(y)|\, dy,$$

a direct computation shows that, at scale 2^{-2j}, all these wavelet coefficients are controlled by

$$|C_{j,k}(f)| \leq \|f\|_\infty \|\phi_{j,k}\|_{L^1} \leq C\, 2^{-3j}.$$

If follows that the N-th largest wavelet coefficient $|C_N(f)|$ is bounded by $O(N^{-3/4})$ and, thus, if f_N^W is the approximation of f obtained by taking the N largest coefficients of its wavelet expansion, the L^2–error obeys the estimate:

$$\|f - f_N^W\|_{L^2}^2 \leq \sum_{\ell > N} |C_\ell(f)|^2 \leq C\, N^{-1/2}.$$

By contrast, the elements of the shearlet system, denoted by $\psi_{j,\ell,k}$, at scale 2^{-2j}, 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. At fine scales (j "large"), it is reasonable to assume that the only significant shearlet coefficients $\mathcal{S}_{j,\ell,k}(f) = \langle f, \psi_{j,\ell,k}, \rangle$ are those corresponding to the shearlet elements which are tangent to the surface of discontinuity and there are about are $O(2^{2j})$ coefficients of this type (see illustration in Figure 1). Again, a direct computation (see expression (1.3)) gives that

$$\int_{\mathbb{R}^3} |\psi_{j,\ell,k}(x)|\, dx = 2^{-2j} \int_{\mathbb{R}^3} |\psi(y)|\, dy,$$

so that, at scale 2^{-2j}, all these shearlet coefficients are controlled by

$$|\mathcal{S}_{j,\ell k}(f)| \leq \|f\|_\infty \|\psi_{j,\ell,k}\|_{L^1} \leq C\, 2^{-2j}.$$

It follows that the N-th largest shearlet coefficient $|\mathcal{S}_N(f)|$ is bounded by $O(N^{-1})$ and this implies that, if f_N is the approximation of f computed by taking the N largest coefficients of its shearlets expansion, the L^2–error approximately obeys the estimate:

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

The rigorous proof of this estimate, which will be described in this paper, requires a more careful examination of the shearlet coefficients, including the contribution of those elements that have been considered negligible in the heuristic argument. As we will see, this will add an additional logarithmic factor to our estimate, finally yielding (1.2).

1.2. **The shearlet representation.** The shearlet systems considered in this paper will be derived within the framework of wavelets with composite dilations introduced by the authors and their collaborators in [13, 14]. This approach provides a general method for the construction of representation systems made up of functions ranging not only at various scales and locations, as traditional wavelets, but also at various orientations, with the ability to deal very effectively with the type of anisotropic phenomena which are a main feature of the multidimensional data usually found in applications.

Specifically, for $\psi \in L^2(\mathbb{R}^3)$, a 3-D shearlet system is a collection of functions of the form

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

where $A = \left(\begin{smallmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{smallmatrix}\right)$ and, for $\ell = (\ell_1, \ell_2) \in \mathbb{Z}^2$, the shear matrices B_ℓ are defined as $B_\ell = \left(\begin{smallmatrix} 1 & \ell_1 & \ell_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$. Similarly to the 2-D case, we are interested in shearlet systems whose elements are well localized and form a Parseval frame. To achieve this, for $\xi = (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3$, we define ψ by

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

where ψ_1 and ψ_2 satisfy the following assumptions:
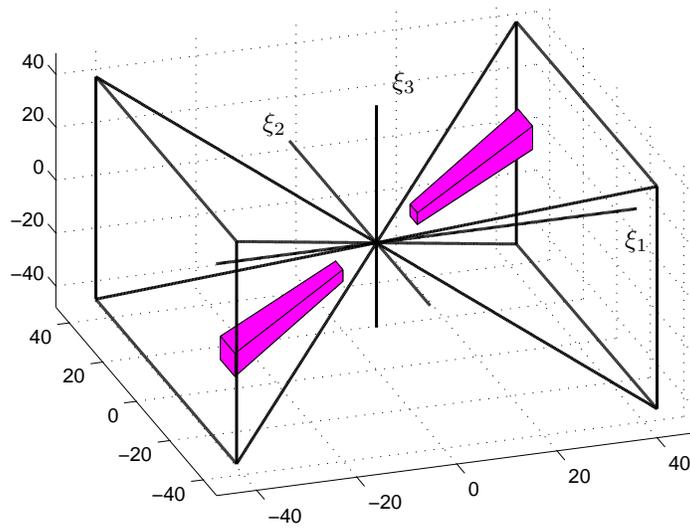
FIGURE 2. *Frequency support of a representative shearlet function $\psi_{j,\ell,k}$, inside the pyramidal region $\mathcal{D}_C$. The orientation of the support region is controlled by $\ell = (\ell_1, \ell_2)$; its shape is becoming more elongated as j increases ($j = 4$ in this plot)*

.

(i) $\hat{\psi}_1 \in C^\infty(\widehat{\mathbb{R}})$, supp $\hat{\psi}_1 \subset [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$ and

$$\text{(1.4)} \qquad \sum_{j \geq 0} |\hat{\psi}_1(2^{-2j}\omega)|^2 = 1 \quad \text{for } |\omega| \geq \frac{1}{8};$$

(ii) $\hat{\psi}_2 \in C^\infty(\widehat{\mathbb{R}})$, supp $\hat{\psi}_2 \subset [-1, 1]$ and

$$\text{(1.5)} \qquad |\hat{\psi}_2(\omega - 1)|^2 + |\hat{\psi}_2(\omega)|^2 + |\hat{\psi}_2(\omega + 1)|^2 = 1 \quad \text{for } |\omega| \leq 1.$$

It was shown in [10] that there are several examples of functions satisfying these properties. It follows from equation (1.5) that, for any $j \geq 0$,

$$\text{(1.6)} \qquad \sum_{m=-2^j}^{2^j} |\hat{\psi}_2(2^j\,\omega + m)|^2 = 1, \quad \text{for } |\omega| \leq 1.$$

Hence, using equations (1.4), (1.6) and the observation that

$$(\xi_1, \xi_2, \xi_3)\, A^{-j} B_{-\ell} = (2^{-2j}\xi_1, -\ell_1 2^{-2j}\xi_1 + 2^{-j}\xi_2, -\ell_2 2^{-2j}\xi_1 + 2^{-j}\xi_3),$$

a direct computation gives that:

$$
\begin{aligned}
\sum_{j \geq 0} \sum_{\ell_1 = -2^j}^{2^j} \sum_{\ell_2 = -2^j}^{2^j} |\hat{\psi}(\xi\, A^{-j} B_{-\ell})|^2
&= \sum_{j \geq 0} \sum_{\ell_1 = -2^j}^{2^j} \sum_{\ell_2 = -2^j}^{2^j} |\hat{\psi}_1(2^{-2j}\,\xi_1)|^2\, |\hat{\psi}_2(2^j\,\tfrac{\xi_2}{\xi_1} - \ell_1)|^2\, |\hat{\psi}_2(2^j\,\tfrac{\xi_3}{\xi_1} - \ell_1)|^2 \\
&= \sum_{j \geq 0} |\hat{\psi}_1(2^{-2j}\,\xi_1)|^2 \sum_{\ell_1 = -2^j}^{2^j} |\hat{\psi}_2(2^j\,\tfrac{\xi_2}{\xi_1} - \ell_1)|^2 \sum_{\ell_2 = -2^j}^{2^j} |\hat{\psi}_2(2^j\,\tfrac{\xi_3}{\xi_1} - \ell_1)|^2 = 1,
\end{aligned}
$$

for $(\xi_1, \xi_2, \xi_3) \in \mathcal{D}_C$, where $\mathcal{D}_C = \{(\xi_1, \xi_2, \xi_3) \in \widehat{\mathbb{R}}^2 : |\xi_1| \geq \frac{1}{8}, |\frac{\xi_2}{\xi_1}| \leq 1, |\frac{\xi_3}{\xi_1}| \leq 1\}$. Using this equation together with the fact that $\hat{\psi}$ is supported inside $[-\frac{1}{2}, \frac{1}{2}]^3$ we have the following result.

Theorem 1.1. *The collection of "horizontal" shearlets:*

$$(1.7) \qquad \mathcal{S}(\psi) = \{\psi_{j,\ell,k}(x) = 2^{2j}\,\psi(B_\ell A^j x - k) : j \geq 0, -2^j \leq \ell_1, \ell_2 \leq 2^j, k \in \mathbb{Z}^3\},$$

is a Parseval frame for $L^2(\mathcal{D}_C)^\vee = \{f \in L^2(\mathbb{R}^3) : \text{supp } \hat{f} \subset \mathcal{D}_C\}.$

Similarly to the corresponding 2–D case [10], the shearlet elements $\psi_{j,\ell,k}$ are well-localized waveforms (in fact, $\hat{\psi}_{j,\ell,k} \in C_0^\infty(\widehat{\mathbb{R}}^2)$), with frequency support on a parallelepiped of approximate size $2^{2j} \times 2^j \times 2^j$, at various scales depending on $j \in \mathbb{Z}$, with orientations controlled by the two–dimensional index $\ell = (\ell_1, \ell_2) \in \mathbb{Z}^2$ and spatial location $k \in \mathbb{Z}^3$. Those supports becomes increasingly more elongated at finer scales (See Figure 2).

Notice that $L^2(\mathcal{D}_C)^\vee$ is a strict subspace of L^2. To obtain a Parseval frame for $L^2(\mathbb{R}^3)$, one can construct a second Parseval frame of shearlets with frequency support in the pyramidal region $\mathcal{D}_{C_2} = \{(\xi_1, \xi_2, \xi_3) \in \widehat{\mathbb{R}}^2 : |\xi_2| \geq \frac{1}{8}, |\frac{\xi_1}{\xi_2}| \leq 1, |\frac{\xi_3}{\xi_2}| \leq 1\}$; similarly one obtains a third Parseval frame of shearlets with frequency support in the pyramidal region $\mathcal{D}_{C_3} = \{(\xi_1, \xi_2, \xi_3) \in \widehat{\mathbb{R}}^2 : |\xi_3| \geq \frac{1}{8}, |\frac{\xi_1}{\xi_3}| \leq 1, |\frac{\xi_2}{\xi_3}| \leq 1\}$; Finally, one can easily construct a Parseval frame (or an orthonormal basis) for $V_0 = L^2([-\frac{1}{8}, \frac{1}{8}]^3)^\vee$. Then any function in $L^2(\mathbb{R}^3)$ can be expressed as a sum $f = P_C f + P_{C_2} f + P_{C_3} f + P_{V_0} f$, where each component corresponds to the orthogonal projection of f into one of the 4 subspaces of $L^2(\mathbb{R}^3)$ described above. Since each one of the shearlet systems defined on a pyramidal region behaves very similarly, in the following it will be sufficient to examine the sparsity properties for the "horizontal" system (1.7).

2. Main Results

Before stating the main theorems, let us define more precisely the class of functions that will be used to model the data we are interested in. We follow [3] and introduce $STAR^2(A)$, a class of indicator functions of sets B with C^2 boundaries ∂B. In polar coordinates, let $\rho(\theta, \phi) : [0, 2\pi) \times [0, \pi) \to [0, 1]^2$ be a radius function and define B by $x \in B$ if and only if $|x| \leq \rho(\theta, \phi)$. In particular, the boundary ∂B can be parametrized as the surface in $\mathbb{R}^3$:

$$(2.8) \qquad \beta(\theta, \phi) = \begin{pmatrix} \rho(\theta, \phi)\,\cos(\theta)\,\sin\phi \\ \rho(\theta, \phi)\,\sin(\theta)\,\sin\phi \\ \rho(\theta, \phi)\,\cos\phi \end{pmatrix}.$$

The class of boundaries of interest to us are defined by

$$(2.9) \qquad \sup_{i+j=2} |\rho_{i,j}(\theta, \phi)| \leq A, \quad \rho \leq \rho_0 < 1.$$

We say that a set $B \in STAR^2(A)$ if $B \subset [0, 1]^3$ and B is a translate of a set obeying (2.8) and (2.9). In addition, we set $C_0^2([0, 1]^3)$ to be the collection of twice differentiable functions supported inside $[0, 1]^3$. Finally, we define the set $\mathcal{E}^2(A)$ of *functions which are C^2 away from a C^2 surface* as the collection of functions of the form

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

where $f_0, f_1 \in C_0^2([0, 1]^3)$, $B \in STAR^2(A)$ and $\|f\|_{C^2} = \sum_{|\alpha| \leq 2} \|D^\alpha f\|_\infty \leq 1.$

We can now state the following main result.

Theorem 2.1. *Let M be the set of indices $\{(j, (\ell_1, \ell_2), k) : j \geq 0, -2^j \leq \ell_1, \ell_2 \leq 2^j, k \in \mathbb{Z}^2\}$ and $\{\psi_\mu\}_{\mu \in M}$ be the Parseval frame of shearlets given by (1.7). 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*

$$(2.10) \qquad \sup_{f \in \mathcal{E}^2(A)} |s(f)|_{(N)} \leq C\,N^{-1}\,(\log N),$$

where $|s(f)|_{(N)}$ denotes the N-th largest entry in this sequence $\{s_\mu(f)\}$.

Using Theorem 2.1, we are just one step away from our main result about shearlet approximations. Indeed, 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 $\{|\langle f, \psi_\mu \rangle|^2 : \mu \in M\}$. Since the approximation error satisfies the estimate

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

from (2.10) we immediately have:

Theorem 2.2. *Let $f \in \mathcal{E}^2(A)$ and f_N^S be the approximation to f defined above. Then*

$$\|f - f_N^S\|_2^2 \leq C \, N^{-1} \, (\log N)^2.$$

2.1. **Arguments and constructions.** For reasons of brevity, it will not be possible to present a complete proof of Theorem 2.1 in this short communication. A complete and detailed proof will appear in a separate work [12]. In the following, we will sketch the main ideas of the proof and emphasize the most significant differences with respect to the 2–D argument. As we will show, the general structure of the proof follows the overall structure of the corresponding 2-dimensional sparsity result in [10]. However, the main technical arguments needed to estimate the effect of the surface discontinuities in the shearlet representation require the introduction of a fundamentally new approach which is significantly different from the 2D case.

As in [10], it will be convenient to introduce the *weak–ℓ^p quasi–norm* $\|\cdot\|_{w\ell^p}$ to measure the sparsity of the shearlet coefficients $\{\langle f, \psi_\mu \rangle : \mu \in M\}$. This is defined as

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

where $|s_\mu|_{(N)}$ is the N-th largest entry in the sequence $\{s_\mu\}$ and it is equivalent (cf. [19, Sec.5.3]) to the expression:

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

To analyze the decay properties of the shearlet coefficients $\{\langle f, \psi_\mu \rangle\}_{\mu \in M}$ at a given scale 2^{-j}, $j \geq 0$, we will smoothly localize the function f near dyadic cubes. Namely, for a scale parameter $j \geq 0$ fixed, let $M_j = \{(j, \ell, k) : -2^j \leq \ell_1, \ell_2 \leq 2^j, k \in \mathbb{Z}^3\}$ and $\mathcal{Q}_j$ be the collection of dyadic cubes of the form $Q = [\frac{k_1}{2^j}, \frac{k_1+1}{2^j}] \times [\frac{k_2}{2^j}, \frac{k_2+1}{2^j}] \times [\frac{k_3}{2^j}, \frac{k_3+1}{2^j}]$, with $k_1, k_2, k_3 \in \mathbb{Z}$. For w a nonnegative C^∞ function with support in $[-1, 1]^3$, we define a smooth partition of unity

$$\sum_{Q \in \mathcal{Q}_j} w_Q(x) = 1, \quad x \in \mathbb{R}^3,$$

where, for each dyadic square $Q \in \mathcal{Q}_j$, $w_Q(x) = w(2^j x_1 - k_1, 2^j x_2 - k_2, 2^j x_3 - k_3)$. We will then examine the shearlet coefficients of the localized function $f_Q = f \, w_Q$, i.e., $\{\langle f_Q, \psi_\mu \rangle : \mu \in M_j\}$.

It turns out that for $f \in \mathcal{E}^2(A)$, the coefficients $\{\langle f_Q, \psi_\mu \rangle : \mu \in M_j\}$ exhibit a very different decay behavior depending on whether the surface 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 surface intersects the support of w_Q. Since each Q has sidelength $2 \cdot 2^{-j}$, then $\mathcal{Q}_j^0$ has cardinality $|\mathcal{Q}_j^0| \leq C_0 \, 2^{2j}$, where C_0 is independent of j. Similarly, since f is compactly supported in $[0, 1]^3$, $|\mathcal{Q}_j^1| \leq 2^{3j} + 6 \cdot 2^{2j}$.

Using this notation, we can now state the basic results that are needed to prove Theorem 2.1.

Theorem 2.3. *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*

$$\|\langle f_Q, \psi_\mu \rangle\|_{w\ell^1} \leq C \, 2^{-2j},$$

for some constant C independent of Q and j.

Theorem 2.4. *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*

$$\|\langle f_Q, \psi_\mu \rangle\|_{w\ell^1} \leq C \, 2^{-3j},$$

for some constant C independent of Q and j.

Before discussing those proofs, let us show how Theorems 2.3 and 2.4 are used to prove Theorem 2.1.

Proof of Theorem 2.1.

Using Theorems 2.3 and 2.4, by the p–triangle inequality for weak ℓ^p spaces, $p \leq 1$, we have

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

Since $|\mathcal{Q}_j^0| \leq C_0 \, 2^{2j}$, where C_0 is independent of j, and $|\mathcal{Q}_j^1| \leq 2^{3j} + 6 \cdot 2^{2j}$, it follows that

$$\|s_j(f)\|_{w\ell^1} \leq C,$$

for some C independent of j. This immediately implies that

(2.11) $$R(j, \epsilon) = \#\{\mu \in M_j : |\langle f, \psi_\mu \rangle| > \epsilon\} \leq C \, \epsilon^{-1}.$$

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

$$
\begin{aligned}
|\langle f, \psi_\mu \rangle| &= \left| \int_{\mathbb{R}^2} f(x) \, 2^{2j} \, \psi(B^\ell A^j x - k) \, dx \right| \\
&\leq 2^{2j} \, \|f\|_\infty \int_{\mathbb{R}^2} |\psi(B^\ell A^j x - k)| \, dx \\
\text{(2.12)} \qquad &= 2^{-2j} \, \|f\|_\infty \int_{\mathbb{R}^2} |\psi(y)| \, dy < C' \, 2^{-2j}.
\end{aligned}
$$

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 (2.12) that $R(j, \epsilon) = 0$ for $j > 2 \, (\log_2(\epsilon^{-1}) + \log_2(C')) > 2 \log_2(\epsilon^{-1})$. Thus, using (2.11), we have that

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

and this implies (2.10). $\square$

The proofs of Theorems 2.3 and 2.4 are rather involved. Theorems 2.4, in particular, follows essentially the same ideas as the 2–D case; this is not surprising since it deals with the situation where the shearlets are away from the discontinuity. In contrast, Theorems 2.3, which deals with shearlet coefficients associated with the discontinuous surface, requires the introduction of new analytical tools which are very different from the 2–D case. In the following section, we will sketch the main ideas of the proof of Theorems 2.3.

2.2. Proof of Theorem 2.3 (sketch). Let us consider a function $f \in \mathcal{E}^2(A)$ which contains a C^2 surface of discontinuity. For $j > j_0$ sufficiently large, over a cube of side 2^{-j}, the surface of discontinuity can be parametrized as $x_1 = E(x_2, x_3)$ or $x_2 = E(x_1, x_3)$ or $x_3 = E(x_1, x_2)$. For simplicity, without loss of generality, we will assume that this surface, denoted by Σ, satisfies the equation

$$x_1 = E(x_2, x_3), \quad -2^{-j} \leq x_2, x_3 \leq 2^j,$$

that

$$E(0, 0) = E_1(0, 0) = E_2(0, 0)$$

and that, for all $m, n \in \mathbb{N}$ such that $m + n = 2$, we have:

$$|E(x_2, x_3)| \leq \frac{1}{2}\, 2^{-2j}\, |E_{mn}(x_2, x_3)| \quad \text{on } [-2^{-j}, 2^{-j}]^2.$$

This ensures that the surface is locally nearly flat near the origin. In the following, we will only discuss the situation $j > j_0$. The situation when $j \leq j_0$ is much simpler and will not be discussed here.

The key step in the following argument is the estimation of the decay of the shearlet coefficient corresponding to the surface of discontinuity, and this requires a more elaborated argument than the 2D case. For this analysis, it is useful to consider the following localized version of f. Let w be a nonnegative C^∞ window function with support in $[-1, 1]^3$, and define a *surface fragment* as a function of the form:

$$(2.13) \qquad\qquad f(x) = w(2^j x)\, g(x)\, \chi_{[x_1 > E(x_2, x_3)]}(x), \quad x \in [-2^{-j}, 2^{-j}]^3,$$

where $g \in C_0^2((-1, 1)^3)$. We have the following fundamental result whose proof requires several delicate steps (involving the computation of the Ray Transform of f) but will omitted for space constraint. Without loss of generality we may assume that $|l_1| \leq |l_2|$ due to the fact that the variables ξ_2 and ξ_3 are symmetric in the construction of the horizontal shearlets.

Lemma 2.5 ([12]). *Let f be the surface fragment given by expression (2.13). Then, for each $\xi \in \mathcal{D}_C$, $j \geq 0$ and $-2^j \leq \ell_2 \leq 2^j$, the following estimate holds:*

$$(2.14) \qquad\qquad \int_{\widehat{\mathbb{R}}^3} |\hat{f}(\xi)|^2\, |\Gamma_{j,\ell}(\xi)|^2\, d\xi \leq C\, 2^{-4j}(1 + |\ell_2|)^{-5},$$

where

$$(2.15) \qquad\qquad \Gamma_{j,\ell}(\xi) = \hat{\psi}_1\big(2^{-2j}\,\xi_1\big)\, \hat{\psi}_2\Big(2^j\,\frac{\xi_2}{\xi_1} - \ell_1\Big)\, \hat{\psi}_3\Big(2^j\,\frac{\xi_3}{\xi_1} - \ell_2\Big).$$

Using Lemma 2.5 and a similar result which is valid for the second order partial derivatives of $\Gamma_{j,\ell}(\xi)$, we now deduce that

$$(2.16) \qquad\qquad \sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} \int_{\widehat{\mathbb{R}}^3} \left| L\left(\hat{f}(\xi)\,\Gamma_{j,\ell}(\xi)\right)\right|^2\, d\xi \leq C\, 2^{-4j}\,(1 + |\ell_2|)^{-5}.$$

where L be the second order differential operator defined by:

$$(2.17) \qquad L = \left(I - \left(\frac{2^{2j}}{2\pi\,(1 + |\ell_2|)}\right)^2 \frac{\partial^2}{\partial \xi_1^2}\right)\left(1 - \left(\frac{2^j}{2\pi}\right)^2 \frac{\partial^2}{\partial \xi_2^2}\right)\left(1 - \left(\frac{2^j}{2\pi}\right)^2 \frac{\partial^2}{\partial \xi_3^2}\right).$$

Observe that

$$\xi A^{-j} B^{-\ell} k = \begin{pmatrix} \xi_1 & \xi_2 & \xi_3 \end{pmatrix} \begin{pmatrix} 2^{-2j} & 0 & 0 \\ 0 & 2^{-j} & 0 \\ 0 & 0 & 2^{-j} \end{pmatrix} \begin{pmatrix} 1 & -\ell_1 & -\ell_2 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$$

$$(2.18) \qquad\qquad = (k_1 - k_2 \ell_1 - k_3 \ell_2)\, 2^{-2j} \xi_1 + k_2 2^{-j} \xi_2 + k_3 2^{-j} \xi_3.$$

Hence, a direct computation gives that

$$(2.19) \quad L\left(e^{2\pi i \xi A^{-j} B^{-\ell} k}\right) = \begin{cases} \left(1 + \big(\frac{|\ell_2|}{(1+|\ell_2|)}\big)^2 \big(\frac{k_1}{|\ell_2|} - \frac{k_2 \ell_1}{|\ell_2|} \pm k_3\big)^2\right)(1 + k_2^2)(1 + k_3^2)\, e^{2\pi i \xi A^{-j} B^{-\ell} k} & \text{if } \ell_2 \neq 0 \\ (1 + k_1^2)(1 + k_2^2)(1 + k_3^2)\, e^{2\pi i \xi A^{-j} B^{-\ell} k} & \text{if } \ell_2 = 0, \end{cases}$$

where we have $+k_3$ or $-k_3$ depending on whether ℓ_2 is positive or negative. Using integration by parts, we have that

$$\langle f, \psi_\mu \rangle = |\det A|^{-j/2} \int_{\widehat{\mathbb{R}}^3} L\left(\hat{f}(\xi)\,\Gamma_{j,\ell}(\xi)\right)\, L^{-1}\left(e^{2\pi i \xi A^{-j} B^{-\ell} k}\right)\, d\xi.$$

We will examine only situation where $\ell_2 \neq 0$. The other case are much simpler and will be omitted. In this case, using (2.19), we have that

$$(2.20) \qquad\qquad L^{-1}\left(e^{2\pi i \xi A^{-j} B^{-\ell} k}\right) = G(k, \ell)^{-1}\, e^{2\pi i \xi A^{-j} B^{-\ell} k},$$

where $G(k, \ell) = \left(1 + (\frac{|\ell_2|}{(1+|\ell_2|)})^2 (\frac{k_1}{|\ell_2|} - \frac{k_2 \ell_1}{|\ell_2|} \pm k_3)^2\right)(1 + k_2^2)(1 + k_3^2)$. Thus, it follows that

$$\langle f, \psi_\mu \rangle = |\det A|^{-j/2} G(k, \ell)^{-1} \int_{\widehat{\mathbb{R}}^3} L\left(\hat{f}(\xi) \Gamma_{j,\ell}(\xi)\right) e^{2\pi i \xi A^{-j} B^{-\ell} k} \, d\xi,$$

or, equivalently, that

$$G(k, \ell) \langle f, \psi_\mu \rangle = |\det A|^{-j/2} \int_{\widehat{\mathbb{R}}^3} L\left(\hat{f}(\xi) \Gamma_{j,\ell}(\xi)\right) e^{2\pi i \xi A^{-j} B^{-\ell} k} \, d\xi.$$

Let $K = (K_1, K_2, K_3) \in \mathbb{Z}^3$ and define $R_K = \{k = (k_1, k_2, k_3) \in \mathbb{Z}^3 : \frac{k_1}{|\ell_2|} \in [K_1, K_1 + 1], -\frac{k_2 \ell_1}{|\ell_2|} \in [K_2, K_2 + 1], k_3 = K_3\}$. 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 $[-\frac{1}{2}, \frac{1}{2}] A^j B^\ell$, and the function $\Gamma_{j,\ell}(\xi)$ is supported on this set, then

$$\sum_{k \in R_K} G(k, \ell)^2 |\langle f, \psi_\mu \rangle|^2 \leq \int_{\widehat{\mathbb{R}}^3} \left| L\left(\hat{f}(\xi) \Gamma_{j,\ell}(\xi)\right) \right|^2 \, d\xi.$$

This implies that

$$\sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} \sum_{k \in R_K} G(k, \ell)^2 |\langle f, \psi_\mu \rangle|^2 \leq \sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} \int_{\widehat{\mathbb{R}}^3} \left| L\left(\hat{f}(\xi) \Gamma_{j,\ell}(\xi)\right) \right|^2 \, d\xi.$$

From the definition of R_K, it follows that

$$\sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} \sum_{k \in R_K} |\langle f, \psi_\mu \rangle|^2 \leq C \left(1 + (K_1 - K_2 \pm K_3)^2\right)^{-2} (1 + K_2^2)^{-2} (1 + K_3^2)^{-2} \sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} \int_{\widehat{\mathbb{R}}^3} \left| L\left(\hat{f}(\xi) \Gamma_{j,\ell}(\xi)\right) \right|^2 \, d\xi.$$

Thus, by (2.16), we have that

$$(2.21) \qquad \sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} \sum_{k \in R_K} |\langle f, \psi_\mu \rangle|^2 \leq C \, L_K^{-2} \, 2^{-4j} (1 + |\ell_2|)^{-5},$$

where $L_K = \left(1 + (K_1 - K_2 \pm K_3)^2\right)(1 + K_2^2)(1 + K_3^2)$.

For j, ℓ fixed, let $N_{j,\ell,K}(\epsilon) = \#\{k \in R_K : |\psi_{j,\ell,k}| > \epsilon\}$. Since $|\ell_1| \leq |\ell_2|$, it is clear that $N_{j,\ell,K}(\epsilon) \leq C(1 + |\ell_2|)^2$ (C is independent of ℓ_1) and, hence, $\sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \leq C(1 + |\ell_2|)^3$. Using the new notation, from (2.21) we have that

$$\sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \leq C \, L_K^{-2} \, 2^{-4j} \, \epsilon^{-2} (1 + |\ell|)^{-5}.$$

This implies that

$$(2.22) \qquad \sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \leq C \min\left((|\ell| + 1)^3, L_K^{-2} \, 2^{-4j} \, \epsilon^{-2} (1 + |\ell|)^{-5}\right).$$

Using (2.22) we will now show that:

$$(2.23) \qquad \sum_{\ell_2 = -2^j}^{2^j} \sum_{\ell_1 = -|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \leq C \, L_K^{-1} \, 2^{-2j} \, \epsilon^{-1}.$$

In fact, let ℓ_2^* be defined by $(\ell_2^* + 1)^3 = L_K^{-2} \, 2^{-4j} \, \epsilon^{-2} (1 + \ell_2^*)^{-5}$. That is, $(\ell_2^* + 1)^4 = L_K^{-1} \, 2^{-2j} \, \epsilon^{-1}$. Then

$$
\begin{aligned}
\sum_{\ell_2=-2^j}^{2^j} \sum_{\ell_1=-|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \;\leq\; & \sum_{|\ell_2|\leq(\ell_2^*+1)} \sum_{\ell_1=-|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) + \sum_{|\ell_2|>(\ell_2^*+1)} \sum_{\ell_1=-|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \\
\leq\; & \sum_{|\ell_2|\leq(\ell_2^*+1)} (|\ell_2|+1)^3 + \sum_{|\ell_2|>(\ell_2^*+1)} L_K^{-2} \, 2^{-4j} \, \epsilon^{-2} (1+|\ell_2|)^{-5} \\
\leq\; & C\ell_2^* + 1)^4 + C L_K^{-2} \, 2^{-4j} \, \epsilon^{-2} (1+\ell_2^*)^{-4} \leq C \, (\ell_2^* + 1)^4,
\end{aligned}
$$

which gives (2.23).

Since $\sum_{K\in\mathbb{Z}^3} L_K^{-1} < \infty$, using (2.23) we then have that

$$
\#\{\mu \in M_j : |\langle f, \psi_\mu \rangle| > \epsilon\} \leq \sum_{K\in\mathbb{Z}^3} \sum_{\ell_2=-2^j}^{2^j} \sum_{\ell_1=-|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \leq C \, 2^{-2j} \, \epsilon^{-1} \sum_{K\in\mathbb{Z}^3} L_K^{-1} \leq C \, 2^{-2j} \, \epsilon^{-1},
$$

and this completes the proof, thanks to the equivalent definition of weak ℓ^1 norm. $\square$

References

[1] E. J. Candès and D. L. Donoho, *New tight frames of curvelets and optimal representations of objects with C^2 singularities*, Comm. Pure Appl. Math., **57** (2004), 219–266.

[2] F. Colonna, G. Easley, K. Guo, and D. Labate, *Radon Transform Inversion using the Shearlet Representation*, Appl. Comput. Harmon. Anal. **29**(2)2 (2010), 232–250.

[3] D. L. Donoho, *Wedgelets: Nearly-minimax estimation of edges*, Annals of Statistics, **27** (1999), 859–897.

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

[5] D. L. Donoho and G. Kutyniok. *Microlocal Analysis of the Geometric Separation Problem*, preprint 2010.

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

[7] G. R. Easley, D. Labate, F. Colonna, *Shearlet-Based Total Variation Diffusion for Denoising*, IEEE Trans. Image Proc. **18** (2) (2009), 260–268.

[8] G. R. Easley, D. Labate, and W. Lim, *Sparse Directional Image Representations using the Discrete Shearlet Transform*, Appl. Comput. Harmon. Anal., **25** (1) (2008), 25–46.

[9] 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.

[10] K. Guo, D. Labate, *Optimally Sparse Multidimensional Representation using Shearlets*, SIAM J. Math. Anal., **9** (2007), 298–318

[11] K. Guo, and D. Labate, *Characterization and analysis of edges using the continuous shearlet transform*, SIAM J. Imag. Sci. **2** (2009), 959–986.

[12] K. Guo, and D. Labate, *Optimally Sparse Representations of 3D Data with C^2 Surface Singularities using Parseval Frames of Shearlets*, Technical Report, University of Houston, 2010.

[13] K. Guo, W.-Q Lim, D. Labate, G. Weiss and E. Wilson, *Wavelets with composite dilations*, Electron. Res. Announc. Amer. Math. Soc., **10** (2004), 78–87

[14] 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), 231–249.

[15] G. Kutyniok and W. Lim, *Compactly Supported Shearlets are Optimally Sparse*, preprint (2010).

[16] G. Kutyniok and T. Sauer. *Adaptive Directional Subdivision Schemes and Shearlet Multiresolution Analysis*, SIAM J. Math. Anal. **41** (2009), 1436–1471.

[17] G. Kutyniok, M. Shahram, and D. L. Donoho. *Development of a Digital Shearlet Transform Based on Pseudo-Polar FFT*. Wavelets XIII (San Diego, CA, 2009), 74460B-1 - 74460B-13, SPIE Proc. 7446, SPIE, Bellingham, WA, 2009.

[18] S. Mallat, *A Wavelet Tour of Signal Processing.Third Edition: The Sparse Way*, Academic Press, San Diego, CA, 2008.

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

[20] S. Yi, D. Labate, G. R. Easley, and H. Krim, *A Shearlet approach to Edge Analysis and Detection*, IEEE Trans. Image Process **18**(5) (2009), 929–941.

Department of Mathematics, Missouri State University, Springfield, Missouri 65804
E-mail address: KanghuiGuo@MissouriState.edu

Department of Mathematics, University of Houston, Houston, Texas 77204
E-mail address: dlabate@math.uh.edu