OPTIMALLY SPARSE REPRESENTATIONS OF 3D DATA WITH C^2 SURFACE SINGULARITIES USING PARSEVAL FRAMES OF SHEARLETS

KANGHUI GUO* AND DEMETRIO LABATE†

Abstract. This paper introduces a new Parseval frame of shearlets for the representation of 3–D data, which is especially designed to handle geometric features such as discontinuous boundaries with very high efficiency. This new system of shearlets forms a multiscale pyramid of well-localized waveforms at various locations and orientations, which become increasingly wafer-like at fine scales. We prove that the new 3-D shearlet representation exhibits essentially optimal approximation properties for tri-variate functions f which are smooth away from discontinuities along C^2 surfaces. Specifically, the N-term approximation f_N^S obtained by selecting the N largest coefficients of 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. This result extends to the 3D setting the (essentially) optimally sparse approximation results obtained by the authors using 2–D shearlets and by Candès and Donoho using curvelets. The result presented in this paper is the first nonadaptive construction to provide provably optimal approximation properties (up to a loglike factor) for a large class of 3-dimensional data.

Key words. Affine systems, curvelets, shearlets, sparsity, wavelets.

AMS subject classifications. 42C15, 42C40

1. Introduction.

Sparse representations of multidimensional data have gained more and more prominence in recent years as a variety of applied problems require to process massive and multi-dimensional data sets in a timely and effective manner. This is a major challenge in applications such as remote sensing, satellite imagery, scientific simulations and electronic surveillance. Sparse representations allow not only to accurately and reliably compress data and expedite their transmission and storage, but also to developed more effective algorithms for tasks such as feature extraction and pattern recognition. In fact, constructing sparse representations for data in a certain class entails the intimate understanding of their true nature and structure [6].

Wavelets and other traditional multiscale methods have been extremely successful during the past 20 years thanks to their ability to provide optimally sparse representations for data with point singularities. This was exploited to develop a number of impressive applications in signal and image processing. Wavelets however fail to be equally efficient when dealing with distributed discontinuities, and this is a major limitation in multidimensional applications where edges and boundaries of discontinuity are frequently the dominant features of the objects to be analyzed. This inefficiency of wavelets in dealing with distributed singularities is due to their isotropic nature, which makes them not very adapted to capture edges and other essential geometric features of multidimensional data. To overcome these limitations, a new generation of multiscale systems was introduced in recent years, most notably the *curvelets* [2] and the *shearlets* [9, 10], which are especially designed to represent efficiently anisotropic features in images. The intuitive idea behind their construction is that, in order to deal efficiently with the edges and the other geometric features which are prominent in most images of practical interest, the analyzing elements must be defined not only at various locations and scales, as traditional wavelets, but also at various orientations and with highly anisotropic shapes. Thanks to their geometrical properties, the curvelet and shearlet representations turn out to be essentially as good as an adaptive representation from the point of view of their 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^S obtained from the N largest coefficients in its curvelet or shearlet expansion, obeys

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

*Department of Mathematics, Missouri State University, Springfield, Missouri 65804 (KanghuiGuo@MissouriState.edu).
†Department of Mathematics, University of Houston, Houston, Texas 77204 (dlabate@math.uh.edu)

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

The goal of this paper is to extend to the 3D setting the remarkable optimal approximation result achieved for images with edges. This extension is highly nontrivial since, as it will be apparent from the description below, the proof of the (almost) optimal sparsity does not follow directly from the arguments used in the bivariate case. Notice that how to *construct* different versions of curvelets and shearlets in the 3D setting has been already discussed in the literature (e.g., [10, 12]) and some useful properties of 3D shearlets have been recently analyzed by the authors in [12]. In addition, a discrete implementation of 3D curvelets, which extends the corresponding 2D implementation, was introduced in [1]. However, no rigorous analysis of the sparsity properties of curvelets or shearlets (or any other related system) in the 3D setting has been published so far. The main contribution of this paper is to construct a new Parseval frame of shearlets and prove that this new shearlet representation exhibits essentially optimal approximation properties for tri-variate smooth functions with discontinuities along C^2 boundaries. This is the first published result of this type. Specifically, denoting by f_N^S the shearlet approximation of f which is obtained from the N largest coefficients of its shearlet representation, we will show that the approximation error satisfies

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

Up the logarithmic factor, this is the optimal approximation rate for this type of functions [5] in the sense that no orthonormal bases or Parseval frames can yield approximation rates than are better than N^{-1}. Indeed, 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} [5]. In particular, it significantly outperforms wavelet and Fourier approximations, whose asymptotic approximation rates are of the order of $N^{-1/2}$ and $N^{-1/3}$, respectively.

Finally, it is important to emphasize that the approach presented in this paper is purely non-adaptive. A different approach, which uses adaptive constructions, was recently proposed by Le Pennec, Mallat and Peyre [19, 20, 23, 24]. Remarkably, for the class of functions considered in this paper, the shearlet approach is as effective as an adaptive representation with respect to its ability to approximate 3D data with discontinuous boundaries.

Remark. During the final editing of this paper, we found that a similar (essentially) optimal sparsity result was recently announced by Kutyniok, Lemvig and Lim, based on a new remarkable construction of compactly supported shearlet frames [18] (see also [15] for the corresponding 2D case).

1.1. Outline. The paper is organized as follows. The construction of the new 3D Parseval frame of shearlets is presented in Section 2. The main results of the paper are given in Section 3. The technical constructions needed for the proofs are collected in Section 4. Finally, Section 5 contains additional remarks about the extension of our sparsity results to the situation where the boundary surface is piecewise smooth.

2. The shearlet representation.

The shearlet representation, which is derived within the framework of wavelets with composite dilations introduced by the authors and their collaborators in [13, 14], 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. Thanks to the ability of the shearlet system to deal with directionality and anisotropy, the geometric content of multivariate functions and data is captured much more efficiently than using wavelets and other more traditional methods. In addition, thanks to its affine structure, the elements of the representation systems are obtained from the action of the affine group on a single or finite collection of generators. This property provides not only greater flexibility and mathematical simplicity with respect to other directional representations, but it also ensures that there is a natural transition from the continuum to the discrete setting. These unique features have been exploited in several imaging applications such as those described in [3, 7, 8, 22, 26].

In dimension $d = 3$, a *shearlet system* is defined as an affine system associated with the action of the affine group $\mathcal{A} = \{(M_{j,\ell}, k) \in GL_3(\mathbb{Z}) \times \mathbb{Z}^3\}$, where the matrices $M_{j,k}$ are obtained from the composition of anisotropic dilation and shear matrices. Namely, for $\psi \in L^2(\mathbb{R}^3)$, a shearlet system is a collection of

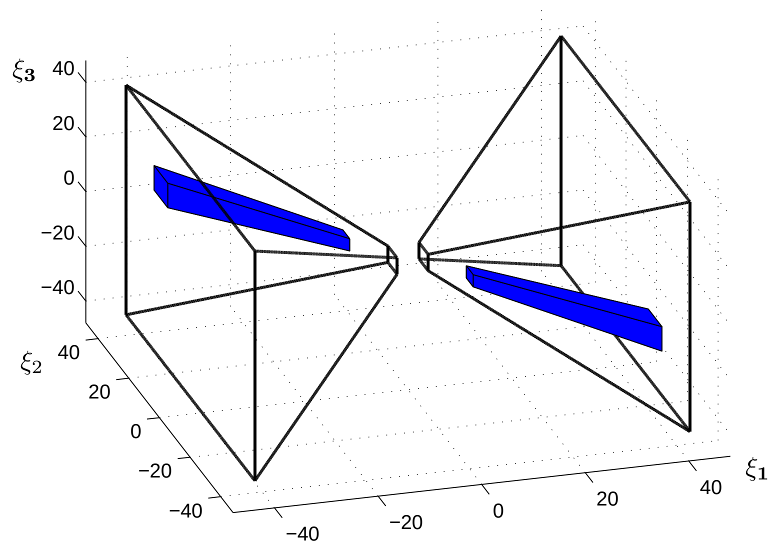

FIG. 2.1. *Frequency support of a representative shearlet function $\psi_{j,\ell,k}$, inside the pyramidal region $\mathcal{D}_C$. The orientation of the support region is controlled by $\ell = (\ell_1, \ell_2)$; its shape is becoming more elongated as j increases ($j = 4$ in this plot)*

functions of the form

$$\{\psi_{j,\ell,k} = |\det A|^{j/2}\, \psi(B_\ell\, A^j x - k) : j \in \mathbb{Z}, \ell \in L \subset \mathbb{Z}^2, k \in \mathbb{Z}^3\}, \tag{2.1}$$

where

$$A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad B_\ell = \begin{pmatrix} 1 & \ell_1 & \ell_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \text{and } \ell = (\ell_1, \ell_2) \in \mathbb{Z}^2.$$

As in the 2-D case, we are interested in 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

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

where ψ_1 and ψ_2 satisfy the following assumptions:

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

$$\sum_{j\geq 0} |\hat{\psi}_1(2^{-2j}\omega)|^2 = 1 \quad \text{ for } |\omega| \geq \frac{1}{8}; \tag{2.2}$$

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

$$|\hat{\psi}_2(\omega-1)|^2 + |\hat{\psi}_2(\omega)|^2 + |\hat{\psi}_2(\omega+1)|^2 = 1 \quad \text{ for } |\omega| \leq 1. \tag{2.3}$$

It was shown in [10] that there are several examples of functions satisfying these properties. It follows from equation (2.3) that, for any $j \geq 0$,

$$\sum_{m=-2^j}^{2^j} |\hat{\psi}_2(2^j\, \omega + m)|^2 = 1, \quad \text{ for } |\omega| \leq 1. \tag{2.4}$$

Notice that, in the frequency domain, the elements $\psi_{j,\ell,k}$, given by (2.1), have the form

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

Hence, using equations (2.2), (2.4) 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:

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

for $(\xi_1,\xi_2,\xi_3)\in\mathcal{D}_C$, where $\mathcal{D}_C = \{(\xi_1,\xi_2,\xi_3)\in\widehat{\mathbb{R}}^2 : |\xi_1|\ge\frac{1}{8}, |\frac{\xi_2}{\xi_1}|\le 1, |\frac{\xi_3}{\xi_1}|\le 1\}$. This equation, together with the fact that $\hat{\psi}$ is supported inside $[-\frac{1}{2},\frac{1}{2}]^3$, implies that the collection of *"horizontal" shearlets*

$$\mathcal{S}(\psi) = \{\psi_{j,\ell,k}(x) = 2^{2j}\,\psi(B_\ell A^j x - k) : j\ge 0, -2^j\le \ell_1,\ell_2\le 2^j, k\in\mathbb{Z}^3\} \tag{2.5}$$

is a Parseval frame for $L^2(\mathcal{D}_C)^\vee = \{f\in L^2(\mathbb{R}^3) : \operatorname{supp}\hat{f}\subset\mathcal{D}_C\}$. Similar to the corresponding 2–D case [10], the shearlet elements $\psi_{j,\ell,k}$ are well-localized waveforms (in fact, $\hat{\psi}_{j,\ell,k}\in C_0^\infty(\widehat{\mathbb{R}}^2)$), at various scales depending on $j\in\mathbb{Z}$, with frequency support contained on parallelepipeds of approximate size $2^{2j}\times 2^j\times 2^j$, defined at various orientations controlled by the two–dimensional index $\ell=(\ell_1,\ell_2)\in\mathbb{Z}^2$ and spatial location $k\in\mathbb{Z}^3$. These support regions become increasingly more elongated at finer scales (See Figure 2.1).

Our construction, so far, only provides a Parseval for the subspace $L^2(\mathcal{D}_C)^\vee$ of functions in L^2 whose frequency support is contained in the pyramidal region $\mathcal{D}_C$. 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 $\mathcal{D}_{C_2} = \{(\xi_1,\xi_2,\xi_3)\in\widehat{\mathbb{R}}^2 : |\xi_2|\ge\frac{1}{8}, |\frac{\xi_1}{\xi_2}|\le 1, |\frac{\xi_3}{\xi_2}|\le 1\}$; similarly, a third Parseval frame of shearlets can be constructed with frequency support in the pyramidal region $\mathcal{D}_{C_3} = \{(\xi_1,\xi_2,\xi_3)\in\widehat{\mathbb{R}}^2 : |\xi_3|\ge\frac{1}{8}, |\frac{\xi_1}{\xi_3}|\le 1, |\frac{\xi_2}{\xi_3}|\le 1\}$; Finally, one can easily define 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.

Concerning the comparison of shearlet and curvelet representations, it is important to recall that, while both methods extend the classical multiscale analysis by introducing a notion of directionality, the shearlets have a fundamentally different mathematical structure. In fact, they are an affine system, where all elements are generated by the action of translations, dilations and shear transformations on a *single* or finite set of generators. This not true for curvelets. An important consequence is that there is Multiresolution Analysis associated with shearlets. In addition, thanks to the use of shear matrices rather than rotations, the shearlet approach ensures a natural transitions from the continuum to the discrete and digital settings [11, 14, 16, 17].

2.1. Significance. Before presenting the proof of the main sparsity result, it is useful to describe a simple heuristic argument to justify why a 3-D shearlet system like the one constructed above should be effective in providing very sparse representations for functions of 3 variables with discontinuous boundaries. In fact, let us consider a bounded function f, defined on a bounded domain, which is smooth away from a discontinuity along a smooth surface. We will examine the behavior of the *shearlet coefficients* of f, which are given by $\mathcal{S}_{j,\ell,k}(f) = \langle f, \psi_{j,\ell,k,}\rangle$, where the shearlet elements $\psi_{j,\ell,k}$ are defined by (2.1). The first observation is that, thanks to their localization properties, at scale 2^{-2j}, the elements $\psi_{j,\ell,k}$, 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. Also, notice that, since

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

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

$$|\mathcal{S}_{j,\ell k}(f)| \leq \|f\|_\infty \, \|\psi_{j,\ell,k}\|_{L^1} \leq C\, 2^{-2j}. \tag{2.6}$$

At fine scales (j "large"), it is reasonable to assume that the only significant coefficients are those corresponding to the shearlet elements which are tangent to the surface of discontinuity. Since there are $O(2^{2j})$ coefficients of this type and they are bounded by (2.6), it follows that the N-th largest shearlet coefficient $|\mathcal{S}_N(f)|$ is bounded by $O(N^{-1})$. This implies that, if f is approximated by taking the N largest coefficients in the shearlets expansion, the L^2–error approximately obeys the estimate:

$$\|f - f_N\|_{L^2}^2 \leq \sum_{\ell > N} |\mathcal{S}_\ell(f)|^2 \leq C\, N^{-1}. \tag{2.7}$$

A rigorous analysis of the behavior of the shearlet coefficients is the main goal of this paper and will be presented below. This requires a careful examinations of the terms which were neglected in our heuristic argument and – as this analysis will show – this produces an additional logarithmic factor to our estimate, finally yielding estimate (1.2).

3. Main Results. Before stating our main results, let us define the class of functions that will be considered in this paper. Fix a constant $A > 0$. We will consider a class $\mathcal{M}(A)$ of indicator functions of sets $B \subset [0,1]^3$ whose boundary $\Sigma = \partial B$ is a C^2 2-manifold which can be written as $\bigcup_\alpha \Sigma_\alpha$, where α ranges over a finite index set and $\Sigma_\alpha = \{(v,\ E_\alpha(v)),\ \ v \in V_\alpha \subset \mathbb{R}^2\}$, such that $\|E_\alpha\|_{C^2(V_\alpha)} \leq A$ for all α. Also, let $C_0^2([0,1]^3)$ be the collection of twice differentiable functions supported inside $[0,1]^3$. Hence, 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 $f_0, f_1 \in C_0^2([0,1]^3)$, $B \in \mathcal{M}(A)$ and $\|f\|_{C^2} = \sum_{|\alpha| \leq 2} \|D^\alpha f\|_\infty \leq 1$. Notice that the set $\mathcal{E}^2(A)$ contains the class of "cartoon-like" images introduced by Donoho [4].

For simplicity of notation, let $\{\psi_\mu\}_{\mu \in M}$ denote our Parseval frame of shearlets, described in Section 2, where M is the set of indices $\{(j,(\ell_1,\ell_2),k) : j \geq 0, -2^j \leq \ell_1, \ell_2 \leq 2^j, k \in \mathbb{Z}^2\}$. 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 3.1. *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*

$$\sup_{f \in \mathcal{E}^2(A)} |s(f)|_{(N)} \leq C\, N^{-1} (\log N). \tag{3.1}$$

Using Theorem 3.1, 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 $\{|\langle f, \psi_\mu \rangle|^2 : \mu \in M\}$. The approximation error satisfies the estimate:

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

Therefore, from (3.1) we immediately have:

THEOREM 3.2. *Let $f \in \mathcal{E}^2(A)$ and f_N^S be the approximation to f defined above. Then*

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

3.1. Arguments and constructions. The general structure of the proof of Theorem 3.1 follows the overall structure of the corresponding 2-dimensional sparsity result in [10]. However, as it will be clear below, the core of the proof requires 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 by

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

wheret $|s_\mu|_{(N)}$ is the N-th largest entry in the sequence $\{s_\mu\}$. One can show (cf. [25, 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 $\{\langle f, \psi_\mu\rangle\}_{\mu\in M}$ at a given scale 2^{-j}, $j \ge 0$, we will smoothly localize the function f near dyadic cubes. Namely, for a scale parameter $j \ge 0$ fixed, let $M_j = \{(j, \ell, k) : -2^j \le \ell_1, \ell_2 \le 2^j, k \in \mathbb{Z}^3\}$ 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., $\{\langle f_Q, \psi_\mu\rangle : \mu \in M_j\}$.

As it will be shown below, for $f \in \mathcal{E}^2(A)$, the coefficients $\{\langle f_Q, \psi_\mu\rangle : \mu \in M_j\}$ exhibit a 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 $|\mathcal{Q}_j^0| \le C_0\, 2^{2j}$, where C_0 is independent of j. Similarly, since f is compactly supported in $[0,1]^3$, $|\mathcal{Q}_j^1| \le 2^{3j} + 6 \cdot 2^{2j}$.

Using this notation, we can now state the basic results that are needed to prove Theorem 3.1. For simplicity, in the following, we will use the same letter C to denote different uniform constants.

THEOREM 3.3. *Let $f \in \mathcal{E}^2(A)$. For $Q \in \mathcal{Q}_j^0$, with $j \ge 0$ fixed, the sequence of shearlet coefficients $\{\langle f_Q, \psi_\mu\rangle : \mu \in M_j\}$ obeys*

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

for some constant C independent of Q and j.

THEOREM 3.4. *Let $f \in \mathcal{E}^2(A)$. For $Q \in \mathcal{Q}_j^1$, with $j \ge 0$ fixed, the sequence of shearlet coefficients $\{\langle f_Q, \psi_\mu\rangle : \mu \in M_j\}$ obeys*

$$\|\langle f_Q, \psi_\mu\rangle\|_{\ell^1} \le C\, 2^{-4j},$$

for some constant C independent of Q and j.

The proofs of Theorems 3.3 and 3.4 are rather involved. Theorems 3.3, in particular, is the "hardest" part of the new sparsity result, and requires a fundamentally new approach with respect to the 2-dimensional case.

Concerning Theorem 3.4, it also shows that 3–D shearlets are as effective as traditional isotropic wavelets in dealing with smooth functions. [1] Before presenting the proofs of Theorems 3.3 and 3.4, we show how these two theorems are used to prove Theorem 3.1. Indeed, we have the following simple corollary.

COROLLARY 3.5. *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 there is a constant C independent of j such that:*

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

Proof. Using Theorems 3.3 and 3.4, by the triangle inequality for weak ℓ^1 spaces, we have

$$\begin{aligned}
\|s_j(f)\|_{w\ell^1} &\leq \sum_{Q \in \mathcal{Q}_j} \|\partial f_Q \psi_\mu\|_{w\ell^1} \\
&\leq \sum_{Q \in \mathcal{Q}_j^0} \|\langle f_Q, \psi_\mu \rangle\|_{w\ell^1} + \sum_{Q \in \mathcal{Q}_j^1} \|\langle f_Q, \psi_\mu \rangle\|_{\ell^1} \\
&\leq C\,|\mathcal{Q}_j^0|\, 2^{-2j} + C\,|\mathcal{Q}_j^1|\, 2^{-4j} \\
&\leq C(2^{2j}\, 2^{-2j} + 2^{3j}\, 2^{-4j}) \leq C.
\end{aligned}$$

Here we used the facts that $|\mathcal{Q}_j^0| \leq C\, 2^{2j}$, where C is independent of j, and $|\mathcal{Q}_j^1| \leq 2^{3j} + 6 \cdot 2^{2j}$. ∎

We can now prove Theorem 3.1

Proof of Theorem 3.1. By Corollary 3.5, we have that

$$R(j,\epsilon) = \#\{\mu \in M_j : |\langle f, \psi_\mu \rangle| > \epsilon\} \leq C\,\epsilon^{-1}. \tag{3.2}$$

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

$$\begin{aligned}
|\langle f, \psi_\mu \rangle| &= \left| \int_{\mathbb{R}^2} f(x)\, 2^{2j}\, \psi(B^\ell A^j x - k)\, dx \right| \\
&\leq 2^{2j}\, \|f\|_\infty \int_{\mathbb{R}^2} |\psi(B^\ell A^j x - k)|\, dx \\
&= 2^{-2j}\, \|f\|_\infty \int_{\mathbb{R}^2} |\psi(y)|\, dy < C'\, 2^{-2j}.
\end{aligned} \tag{3.3}$$

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 (3.3) that $R(j,\epsilon) = 0$ for $j > 2\,(\log_2(\epsilon^{-1}) + \log_2(C')) > 2\,\log_2(\epsilon^{-1})$. Thus, using (3.2), 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 (3.1). □

4. Proofs of Main Theorems.

4.1. Proof of Theorem 3.3. 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, the scale 2^{-j} is small enough that, 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, we will assume that this surface, denoted by Σ, satisfies the equation

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

[1] Furthermore, an argument similar to Theorem 8.2 in [2] can be used to analyze the estimate the Sobolev norm of a smooth function using shearlet coefficients.

Also we assume that the surface contains the origin $(0,0,0)$ and the normal direction of the surface at $(0,0,0)$ is $(1,0,0)$, which is equivalent to assuming that $E(0,0) = E_{x_2}(0,0) = E_{x_3}(0,0) = 0$. As we will show in Section 4.5, there is no loss in generality in analyzing this case only, since the situation where the surface does not contain the origin or has a different normal direction can be easily converted into the case where $E(0,0) = E_{x_2}(0,0) = E_{x_3}(0,0) = 0$. To further simplify the notation, throughout the remainder of the paper, for a function $g(x)$ with $x \in \mathbb{R}^2$ and $m = (m_1, m_2)$ with $0 \le |m| = m_1 + m_2 \le 2$, we will write $\frac{\partial^m}{\partial x^m} g$ as g_m.

From Taylor's Theorem we have that $E(x_2, x_3) = \frac{1}{2}(E_{(2,0)}(c)x_2^2 + 2E_{(1,1)}(c)x_2x_3 + E_{(0,2)}(c)x_3^2)$, where $c = (c_2, c_3)$ is some point in $[-2^{-j}, 2^{-j}]^2$. It follows that

$$|E(x_2, x_3)| \le 2^{-2j}(\|E_{(2,0)}\|_\infty + \|E_{(1,1)}\|_\infty + \|E_{(0,2)}\|_\infty).$$

Thus, the surface is locally nearly flat near the origin. Notice that this only holds for $j > j_0$. The situation when $j \le j_0$ is much simpler and will be handled separately in Section 4.6.

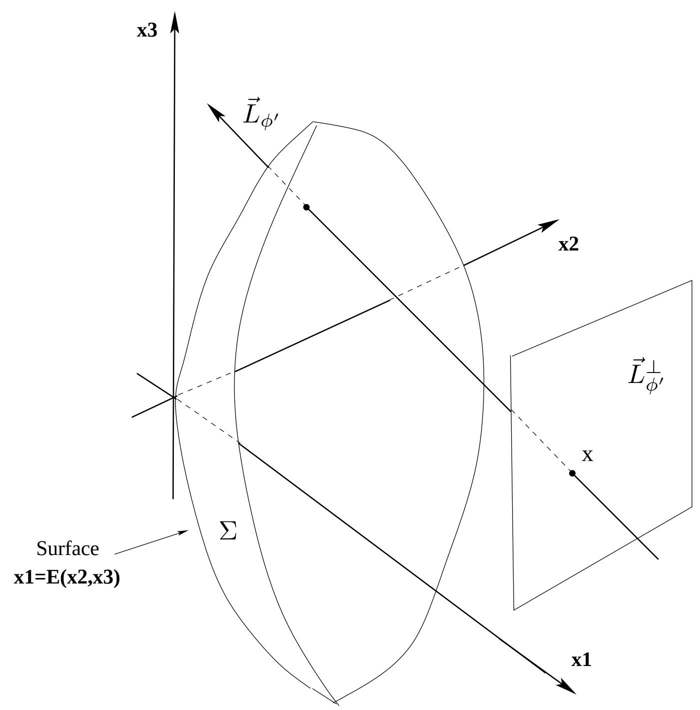

FIG. 4.1. *The surface of discontinuity Σ of equation $x_1 = E(x_2, x_3)$. A line with direction $\vec{L}_{\phi'}$ through the point x intersects the surface at most in one point.*

The key step in the following argument is based on the estimate of the decay of the function f near the surface of discontinuity. In order to define this localized version of f, let w_0 be a nonnegative C^∞ window function with support in $[-1,1]^3$. Hence, for $j \in \mathbb{Z}$, a *surface fragment* is a function of the form:

$$f(x) = w_0(2^j x)\, g(x)\, \chi_{[x_1 > E(x_2,x_3)]}(x), \quad x \in [-2^{-j}, 2^{-j}]^3, \tag{4.1}$$

where $g \in C_0^2((-1,1)^3)$. After re-scaling, we have

$$F(x) = f(2^{-j}x) = w_0(x)\, g(2^{-j}x)\, \chi_{[x_1 > E^{(j)}(x_2,x_3)]}(x), \quad x \in [-1,1]^3, \tag{4.2}$$

where $E^{(j)}(x_2, x_3) = 2^j\, E(2^{-j}x_2, 2^{-j}x_3)$. In particular, we have that $\widehat{F}(\xi) = 2^{3j}\, \hat{f}(2^j \xi)$, and, thus,

$$\int_{|\lambda| \in I_j} |\hat{f}(\lambda\Theta)|^2\, d\lambda = 2^{-5j} \int_{|\lambda| \in 2^{-j} I_j} |\hat{F}(\lambda\Theta)|^2\, d\lambda. \tag{4.3}$$

For simplicity of notation, without loss of generality we may assume that $(\|E_{(2,0)}\|_\infty+\|E_{(1,1)}\|_\infty+\|E_{(0,2)}\|_\infty) = 1$, which yields that $|E(x_2,x_3)| \le 2^{-2j}$ and $|E_m(x_2,x_3)| \le 2^{-j}$ for $|m| \le 2$ for all $(x_2,x_3) \in [-1,1]$.

4.2. Analysis of the Surface Fragment. The main goal of this section is to obtain an estimate for the integral of the surface fragment (4.3). In this section, as well as in the following, we will only consider the analysis on the frequency region defined in the pyramidal region $\mathcal{D}_C$, where the shearlet system (2.5) is employed. Since the other regions can be handled in exactly the same way, there will be no need to consider the shearlet systems defined in the other pyramidal regions.

It will be convenient to express $\xi \in \mathbb{R}^3$, using spherical coordinates, as $\xi = (\rho\cos\theta\sin\phi, \rho\cos\theta\sin\phi, \rho\cos\phi)$, where $\rho > 0$, $\theta \in [0, 2\pi)$ and $\phi \in [0,\pi]$. Since we are only dealing with the frequency region contained in $\mathcal{D}_C$, we will assume that $\phi \in [\frac{\pi}{4}, \frac{3\pi}{4}]$ and $\theta \in [-\frac{\pi}{4}, \frac{\pi}{4}]$. Also notice that, since the variables ξ_2, ξ_3 are symmetric in the construction of the shearlets in $\mathcal{D}_C$, we may assume that $|\ell_1| \le |\ell_2|$.

For $\xi = (\xi_1, \xi_2, \xi_3) \in \mathcal{D}_C$, $j \ge 0$, $|\ell_1| \le |\ell_2| \le 2^j$, let

$$\Gamma_{j,\ell}(\xi) = \hat\psi_1\big(2^{-2j}\,\xi_1\big)\;\hat\psi_2\left(2^j\,\frac{\xi_2}{\xi_1} - \ell_1\right)\;\hat\psi_3\left(2^j\,\frac{\xi_3}{\xi_1} - \ell_2\right). \tag{4.4}$$

We have the following important result:

THEOREM 4.1. *Let f be the surface fragment given by expression (4.1). Then, for each $\xi \in \mathcal{D}_C$, $j \ge 0$ and $-2^j \le \ell_2 \le 2^j$, the following estimate holds:*

$$\int_{\widehat{\mathbb{R}}^3} |\hat f(\xi)|^2\,|\Gamma_{j,\ell}(\xi)|^2\,d\xi \le C\,2^{-4j}(1+|\ell_2|)^{-5}. \tag{4.5}$$

The proof of these results is based on the computation of the Ray Transform of the surface fragment f which is presented below.

4.3. Ray Transform And Fourier Slice Theorem. While the Radon and Ray transforms of bivariate functions are equivalent, this is not true in the three-dimensional setting [21]. Namely, the 3-dimensional *Ray Transform* maps a function on $\mathbb{R}^3$ into the sets of its line integrals; this is different from the *Radon transform* which maps a function on $\mathbb{R}^3$ into the sets of its integrals over planes in $\mathbb{R}^3$. More precisely, if $\Theta \in S^2$ and $x \in \mathbb{R}^3$, then the Ray Transform of $g \in \mathcal{S}(\mathbb{R}^3)$ is defined by

$$Pg(\Theta, x) = \int_{\mathbb{R}} g(t\Theta + x)\,dt.$$

This is the integral of g over the straight line through x with direction Θ (see Figure 4.2). Notice that $Pg(\Theta, x)$ does not change if x is moved in the direction Θ. Hence, x is normally restricted to $\Theta^\perp$ so that Pf is a function on the tangent bundle $\{(\Theta, x) : \Theta \in S^2, x \in \Theta^\perp\}$. It is useful to recall the *Fourier Slice Theorem* which establishes that following relationship between the Ray Transform of g and its Fourier transform:

$$\mathcal{F}_2[Pg](\Theta, \eta) = \int_{\Theta^\perp} Pg(\Theta, x)\,e^{-2\pi i\eta x}\,dx = \hat g(\eta), \quad \eta \in \Theta^\perp,$$

where $\mathcal{F}_2$ denotes the Fourier transform over the second variable. We refer the reader to [21] for this and additional properties of the Ray Transform.

In order to deduce an estimate for the integral of the surface fragment given by the expression (4.3), we will analyze the Ray Transform of the surface fragment F, given by (4.2). Let $\phi' \in [-\frac{\pi}{4}, \frac{\pi}{4}]$. The Ray transform of F in the direction $\vec{L}_{\phi'} = (\sin\phi', 0, \cos\phi')$ is given by

$$PF(\phi', x) = \int_{\mathbb{R}} F(tL_{\phi'} + x)\,dt \tag{4.6}$$

where $x \in \mathbb{R}^3$. This is the integral of F over the straight line through x with direction $L_{\phi'}$. Notice that $PF(\phi', x)$ does not change if x moves along the direction $\vec{L}_{\phi'}$. Hence, x is effectively restricted to $\vec{L}_{\phi'}^\perp$ so

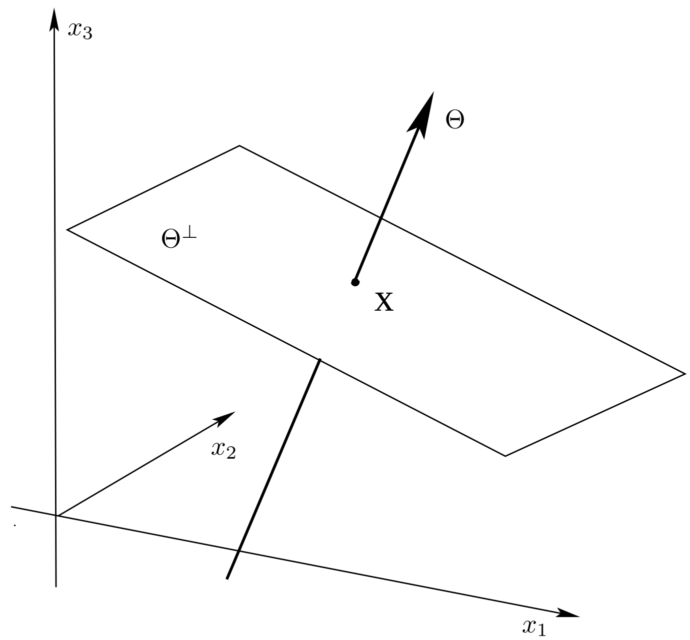

FIG. 4.2. *The Ray transform is defined by integration over the lines through the point x with direction Θ.*

that PF is a function on the tangent bundle $\{(\vec{L}_{\phi'}, x) :\ \vec{L}_{\phi'} \in S^2, x \in \vec{L}_{\phi'}^{\perp}\}$. By introducing the vectors $\vec{L}_1 = (0, -1, 0)$ and $\vec{L}_2 = (\cos\phi', 0, -\sin\phi')$, we can express $x \in L_{\phi'}^{\perp}$ as

$$\{x \in \vec{L}_{\phi'}^{\perp}\} = \{s\vec{L}_1 + w\vec{L}_2 :\ s, w \in \mathbb{R}\}. \tag{4.7}$$

It follows that

$$PF(\phi', s, w) = \int_{\mathbb{R}} F\left(\rho_{\phi'} \begin{pmatrix} t \\ s \\ w \end{pmatrix}\right) dt, \tag{4.8}$$

where $\rho_{\phi'} = \begin{pmatrix} \sin\phi' & 0 & \cos\phi' \\ 0 & -1 & 0 \\ \cos\phi' & 0 & -\sin\phi' \end{pmatrix}$.

By the Fourier Slice Theorem, we have that

$$\mathcal{F}_2[PF](\phi', \eta) = \int_{\vec{L}_{\phi'}^{\perp}} PF(\phi', s, w)\, e^{-2\pi i \eta\cdot(s,w)}\, ds\, dw = \hat{F}(\eta, \phi'), \quad \eta \in \vec{L}_{\phi'}^{\perp}.$$

Hence, by the properties of the Fourier transform (Plancherel and differentiation theorems), we obtain the following identity:

$$\|(PF)_{ss}\|^2 + 2\|(PF)_{sw}\|^2 + \|(PF)_{ww}\|^2 = (2\pi)^4 \int_{\mathbb{R}^2} |\eta|^4\, |\hat{F}(\eta, \phi')|^2\, d\eta, \tag{4.9}$$

where $\eta = \eta_1\, \vec{L}_1 + \eta_2\, \vec{L}_2$.

4.3.1. Ray Transform of the Surface Fragment. For brevity, let us introduce the following notation:

$$F^{\phi'}(t, s, w) = F\left(\rho_{\phi'} \begin{pmatrix} t \\ s \\ w \end{pmatrix}\right), \quad g^{\phi'}(t, s, w) = g\left(2^{-j}\rho_{\phi'} \begin{pmatrix} t \\ s \\ w \end{pmatrix}\right), \quad w^{\phi'}(t, s, w) = w\left(\rho_{\phi'} \begin{pmatrix} t \\ s \\ w \end{pmatrix}\right).$$

Using this notation, we will rewrite the Ray transform of the surface fragment, given by expression (4.8), as

$$PF(\phi', s, w) = \int_{\mathbb{R}} F^{\phi'}(t, s, w)\, dt. \tag{4.10}$$

As described above, this is an integral over the lines $\Lambda_{s,w,\phi'} = \{y \in \mathbb{R}^3 : \, y \cdot \vec{L}_1 = s \quad \& \quad y \cdot \vec{L}_2 = w\}$, where $\vec{L}_1$ and $\vec{L}_2$, given by (4.7), depend on ϕ'. Depending on the values of (s, w, ϕ'), the lines $\Lambda_{s,w,\phi'}$ may or may not intersect the surface $\Sigma = \{(E^{(j)}(u, v), u, v) : \, |u|, |v| \leq 1\}$. In the following, we will analyze the two situations separately.

Case 1: No Intersection.
When the line $\Lambda_{s,w,\phi'}$ does not intersect the surface Σ, the Ray transform of F takes the form:

$$PF(\phi', s, w) = \int_{\mathbb{R}} g^{\phi'}(t, s, w)\, w^{\phi'}(t, s, w)\, dt. \tag{4.11}$$

In this case we have the following result.

Proposition 4.2. *The function PF is twice differentiable as a function of s and w and admits the decomposition*

$$(PF(\phi', s, w))_{ss}\,(\phi', s, w) + (PF(\phi', s, w))_{sw}\,(\phi', s, w) + (PF(\phi', s, w))_{ww}\,(\phi', s, w) = F^0(\phi', s, w) + F^1(\phi', s, w),$$

where

$$\|F^0(\phi', s, w)\|^2 \leq C\, 2^{-2j},$$

$$\|\big(F^1(\phi', s, w)\big)_s\|^2 + \|\big(F^1(\phi', s, w)\big)_w\|^2 \leq C.$$

Proof. With an abuse of notation, in the following we will write g for $g^{\phi'}$ and w_0 for w_0^{ϕ}. By direct computation we have:

$$(PF)_{ss}(\phi, s, w) = \int_{\mathbb{R}} \frac{\partial^2}{\partial s^2} \left(g(t, s, w) w(t, s, w)\right)\, dt = F^0(\phi', s, w) + F^1(\phi', s, w),$$

where $F^0(\phi', s, w) = \int_{\mathbb{R}} (g_{ss}\, w_0 + 2 g_s\, w_{0s})\, dt$ and $F^1(\phi', s, w) = \int_{\mathbb{R}} g\, w_{0ss}\, dt$.

Recalling that $g(t, s, w) = g^{\phi'}(t, s, w) = g\left(2^{-j} \rho_{\phi'} \left(\begin{smallmatrix} t \\ s \\ w \end{smallmatrix}\right)\right)$, a direct computation yields that $|g_s| \leq C\, 2^{-j}$ and $|g_{ss}| \leq C\, 2^{-2j}$. It follows that $|g_s w_{0s}| \leq C\, 2^{-j}$ and $|g_{ss} w_0| \leq C\, 2^{-2j}$. Since w_0 (and hence PF) has compact support, it follows that $\int_{\mathbb{R}} |g_{ss}\, w_0| dt \leq C\, 2^{-2j}$, and $\int_{\mathbb{R}} |g_s\, w_{0s}| dt \leq C\, 2^{-j}$. This implies that

$$\|F^0(\phi', s, w)\|^2 \leq C\, 2^{-2j}.$$

For $F^1(\phi', s, w)$, we have

$$\frac{\partial}{\partial s}(F^1(\phi', s, w)) = \int_{\mathbb{R}} \frac{\partial}{\partial s}(g\, w_{0ss})\, dt = \int_{\mathbb{R}} (g_s\, w_{ss} + g\, w_{sss})\, dt.$$

Using the same argument as the one used for $F^0(\phi', s, w)$, it follows that $\|\big(F^1(\phi', s, w)\big)_s\|^2 \leq C$. Similarly it follows that $\|\big(F^1(\phi', s, w)\big)_w\|^2 \leq C$. The proof is completed by repeating the same argument for $(PF)_{sw}(\phi, s, w)$ and $(PF)_{ww}(\phi, s, w)$. $\square$

From Proposition 4.2, using the Fourier Slice Theorem for the Ray transform and the Plancherel theorem, it follows that

$$\int_0^\infty \int_0^{2\pi} |\hat{F}(r, \theta', \phi')|^2\, r^5\, d\theta'\, dr \leq C\, 2^{-2j}$$

and, hence, that

$$\int_{2^{j-2}}^{2^{j+1}} \int_0^{2\pi} |\hat{F}(r,\theta',\phi')|^2 \, d\theta' \, dr \le C\, 2^{-7j}.$$

Since $F(x) = f(2^{-j}x)$, we have $\hat{F}(\xi) = 2^{3j}\hat{f}(2^j\xi)$. Thus, the above inequality implies the following one:

$$\int_{2^{2j-4}}^{2^{2j+2}} \int_0^{2\pi} |\hat{f}(r,\theta',\phi')|^2 \, d\theta' \, dr \le C\, 2^{-12j}. \tag{4.12}$$

This completes the analysis in the case where there is no intersection.

Case 2: Intersection.

In order to find the intersection of the line $\Lambda_{s,w,\phi'}$ and the surface Σ, one has to solve the equation

$$\rho'_\phi \begin{pmatrix} t \\ s \\ w \end{pmatrix} = \begin{pmatrix} E^{(j)}(u) \\ u \\ v \end{pmatrix},$$

which leads to the system:

$$t = E^{(j)}(u,v)\sin\phi' + v\cos\phi', \tag{4.13}$$
$$s = -u, \tag{4.14}$$
$$w = E^{(j)}(u,v)\cos\phi' - v\sin\phi'. \tag{4.15}$$

To compute the solution of this system, we will use the Implicit Function Theorem to express t as a function of s and w. In order to do that, we first check that the conditions of the Implicit Function Theorem are satisfied. A direct computation gives:

$$s_u = -1, s_v = 0,$$
$$w_u = E_u^{(j)}(u,v)\cos\phi', w_v = E_v^{(j)}(u,v)\cos\phi' - \sin\phi',$$

and

$$\Delta(\phi') = \det \begin{pmatrix} s_u & s_v \\ w_u & w_v \end{pmatrix} = \sin\phi' - E_v^{(j)}\cos\phi' \tag{4.16}$$

The following proposition deals with the case when $|\sin\phi'| \le 2^{1-j}$.

Proposition 4.3. *Assume that $|\sin\phi'| \le 2^{1-j}$. Then, for each fixed j and ϕ', we have that*

$$\int_{2^{2j-4}}^{2^{2j+2}} \int_0^{2\pi} |\hat{f}(r,\theta',\phi')|^2 \, d\theta' \, dr \le C\, 2^{-7j},$$

where where C is independent of j and ϕ'.

Proof. Since $|E_v^{(j)}| \le 2^{-j}$ (from the assumption that $\|E''\|_{L^\infty} = 1$), it follows that $|\Delta(\phi')| \le C2^{-j}$ with C independent of j and ϕ'. Let A be the region defined by $\{(s(u,v), w(u,v)) : (u,v) \in [-1,1]^2\}$. Since $\int_A ds\,dw = \int_{-1}^1\int_{-1}^1 |\Delta(\phi')|\,du\,dv \le C|\sin\phi'|$ and F is bounded (and hence PF is bounded), it follows from a direct calculation that $\|(PF)\|_{L^2}^2 \le C\, \int_{-1}^1\int_{-1}^1 |\Delta(\phi')|\,du\,dv \le C2^{-j}$. Using the Fourier Slice Theorem for the Ray transform and the Plancherel theorem, we have that

$$\int_0^\infty \int_0^{2\pi} |\hat{F}(r,\theta',\phi')|^2 \, r \, d\theta' \, dr \le C\, 2^{-j}$$

and, hence, that

$$\int_{2^{j-2}}^{2^{j+1}} \int_0^{2\pi} |\hat{F}(r,\theta',\phi')|^2 \, d\theta' \, dr \le C\, 2^{-2j}.$$

Since $F(x) = f(2^{-j}x)$, we have $\hat{F}(\xi) = 2^{3j}\hat{f}(2^j\xi)$. Thus the above inequality gives

$$\int_{2^{2j-4}}^{2^{2j+2}} \int_0^{2\pi} |\hat{f}(r,\theta',\phi')|^2 \, d\theta' \, dr \le C\, 2^{-7j}.$$

This finishes the proof of Proposition 4.3. □

For the case when $|\sin\phi'| \ge 2^{1-j}$, we have that $2^{-j} \le \frac{1}{2}|\sin\phi'| \le |\Delta(\phi')| \le 2|\sin\phi'|$. Thus, we can apply the Inverse Function Theorem and use equations (4.14) and (4.15) to derive the functions $u = u(s,w)$ and $v = v(s,w)$. Inserting these functions into (4.13), we obtain the intersection point in terms of t as

$$t_0(s,w,\phi') = E^{(j)}(u(s,w),v(s,w))\sin\phi' + v(s,w)\cos\phi'. \tag{4.17}$$

This shows that there is at most one point of intersection for each fixed (s,w) and ϕ'.

We can write $\eta \in \vec{L}^{\perp}_{\phi'}$ as $\eta = (\eta_2\cos\phi', -\eta_1, -\eta_2\sin\phi') = (r\sin\theta'\cos\phi', -r\cos\theta', -r\sin\theta'\sin\phi')$, where $\eta_1 = r\cos\theta'$, $\eta_2 = r\sin\theta'$. Then (4.9) can be rewritten as

$$\|(PF)_{ss}\|^2 + 2\|(PF)_{sw}\|^2 + \|(PF)_{ww}\|^2 = \int_0^{\infty} \int_0^{2\pi} r^5 |\hat{F}(r,\theta',\phi')|^2 \, d\theta' \, dr. \tag{4.18}$$

Since the same η can also be expressed in spherical coordinates as $\eta = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \cos\phi)$, it follows that we must have $\rho = r$ and

$$\begin{aligned} \sin\theta'\cos\phi' &= \cos\theta\sin\phi, \\ \cos\theta' &= \sin\theta\sin\phi, \\ -\sin\theta'\sin\phi' &= \cos\phi. \end{aligned}$$

From the first and the third identities, we have $\tan\phi' = -\cot\phi\sec\theta$, which implies that ϕ' is *equivalent* to $\phi - \frac{\pi}{2}$, that is, there is are constants $0 < C_1(\theta) \le C_2(\theta) < \infty$ such that $C_1(\theta)\,\phi' \le \phi - \frac{\pi}{2} \le C_2(\theta)\,\phi'$. Also since $|\phi - \frac{\pi}{2}| \le \frac{\pi}{4}$ and $|\theta| \le \frac{\pi}{4}$, we see that $|\frac{\partial\phi'}{\partial\phi}| \le C$ and $|\frac{\partial\phi'}{\partial\theta}| \le C$ and hence

$$|\phi_1' - \phi_2'| \le C\,(|\phi_1 - \phi_2| + |\theta_1 - \theta_2|). \tag{4.19}$$

Also, we have

$$u_s = \frac{E_v^{(j)}(u,v)\cos\phi - \sin\phi}{\Delta(\phi')}, u_w = 0, \tag{4.20}$$

$$v_s = -\frac{E_u^{(j)}(u,v)\cos\phi}{\Delta(\phi')}, v_w = -\frac{1}{\Delta(\phi')}. \tag{4.21}$$

From (4.20) and (4.21), it is easy to verify the following proposition.

Proposition 4.4.

$$|u_s| \le C\frac{1}{|\sin\phi'|},\ |u_{s^2}| \le C\frac{2^{-j}}{|\sin\phi'|^3},\ |u_{sw}| \le C\frac{2^{-j}}{|\sin\phi'|^3},\ |u_{w^2}| \le C\frac{2^{-j}}{|\sin\phi'|^3}$$

$$|v_s| \le C\frac{1}{|\sin\phi'|},\ |v_{s^2}| \le C\frac{2^{-j}}{|\sin\phi'|^3},\ |v_{sw}| \le C\frac{2^{-j}}{|\sin\phi'|^3},\ |v_{w^2}| \le C\frac{2^{-j}}{|\sin\phi'|^3},$$

where the constant C is independent of $(u,v) \in [-1,1]^2$, $\phi' \in [-\frac{\pi}{4}, \frac{\pi}{4}]$ *with* $|\sin\phi'| \geq 2^{1-j}$.

Using the expression (4.17) that was found for the intersection point, from (4.6) and (4.8) we obtain the following formulation of the Ray transform $PF(\phi', s, w)$:

$$PF(\phi', s, w) = \int_{-\infty}^{t_0(s,w,\phi')} F\left(\rho_{\phi'}\begin{pmatrix} t \\ s \\ w \end{pmatrix}\right) dt. \tag{4.22}$$

From Proposition 4.4, one can use essentially the same argument as the 2-dimensional case (see Lemma 6.2 in [2]) to prove the following proposition. For completeness, a sketch of its proof will be provided.

Proposition 4.5. *The Ray Transform of F is twice differentiable as a function of s and w and admits the decomposition*

$$(PF(\phi',s,w))_{ss}(\phi',s,w)+(PF(\phi',s,w))_{sw}(\phi',s,w)+(PF(\phi',s,w))_{ww}(\phi',s,w) = F^0(\phi',s,w)+F^1(\phi',s,w),$$

where

$$\|F^0(\phi',s,w)\|^2 \leq C\,2^{-2j}|\sin\phi'|^{-5},$$

$$\|\left(F^1(\phi',s,w)\right)_s\|^2 + \|\left(F^1(\phi',s,w)\right)_w\|^2 \leq C|\sin\phi'|^{-5}.$$

Proof (Sketch). We will adopt the same notations as in Proposition 4.2.

From (4.22), we have that

$$PF(\phi', s, w) = \int_{-\infty}^{t_0(s,w,\phi')} F\left(\rho_{\phi'}\begin{pmatrix} t \\ s \\ w \end{pmatrix}\right) dt = \int_{-\infty}^{t_0(s,w,\phi')} g(t,s,w)w_0(t,s,w)\,dt.$$

This implies that

$$(PF)_s(\phi',s,w) = g(t_0,s,w)\,w_0(t_0,s,w)\,t_{0s} + \int_{-\infty}^{t_0(s,w,\phi')} \left(g_s(t,s,w)\,w_0(t,s,w) + g(t,s,w)\,w_{0s}(t,s,w)\right) dt$$
$$(PF)_{ss}(\phi',s,w) = T_1 + T_2 + T_3 + T_4,$$

where $T_1 = g_t\,w_0\,(t_{0s})^2 + g_s\,w_0\,t_{0s} + g\,w_0\,t_{0ss}$, $T_2 = gw_{0t}\,(t_{0s})^2 + g\,w_{0s}\,t_{0s}$, $T_3 = \int_{-\infty}^{t_0(s,w,\phi')}(g_{ss}\,w_0 + 2g_s\,w_{0s})\,dt$, $T_4 = \int_{-\infty}^{t_0(s,w,\phi')} g\,w_{0ss}\,dt$.

From $t_0(s,w,\phi') = E^{(j)}\,(u(s,w),v(s,w))\,\sin\phi' + v(s,w)\,\cos\phi'$, using Proposition 4.4, it is easy to verify that $|t_{0s}| \leq C\,\frac{1}{|\sin\phi'|}$, $|t_{0ss}| \leq C\,\frac{2^{-j}}{|\sin\phi'|^3}$. It follows that $|T_1| \leq C\frac{2^{-j}}{|\sin\phi'|^3}$ and, hence, $\|T_1\|^2 \leq C\,\frac{2^{-2j}}{|\sin\phi'|^6}\int_A ds dw \leq \frac{2^{-2j}}{|\sin\phi'|^5}$ since $\int_A ds dw \leq C|\sin\phi'|$. Using the assumption that $|\sin\phi'| \geq 2^{2-j}$, one can verify that $|(T_2)_s| \leq C\frac{1}{|\sin\phi'|^3}$. Similarly one can verify that $|T_3| \leq C2^{-j}$, and $(T_4)_s| \leq C$. Thus, it follows that $\|T_3\|^2 \leq C\,\frac{2^{-2j}}{|\sin\phi'|^5}$, and $\|(T_4)s\|^2 \leq C\,|\sin\phi'|^5$ since $|\sin\phi'| \leq 1$.

Now the argument is completed by letting $F^0(\phi',s,w) = T_1 + T_3$ and $F^1(\phi',s,w) = T_2 + T_4$. □

As a direct corollary of Proposition 4.5, it follows that

$$\int_0^{\infty}\int_0^{2\pi} r^5|\hat{F}(r,\theta',\phi')|^2\,d\theta'\,dr \leq C\,2^{-2j}|\sin\phi'|^{-5},$$

which implies that

$$\int_{2^{j-2}}^{2^{j+1}}\int_0^{2\pi} |\hat{F}(r,\theta',\phi')|^2\,d\theta'\,dr \leq C\,2^{-7j}|\sin\phi'|^{-5}. \tag{4.23}$$

Using again the identity $\hat F(\xi) = 2^{3j} f(2^j \xi)$, from (4.23) it follows that

$$\int_{2^{2j-4}}^{2^{2j+2}} \int_0^{2\pi} |\hat f(r,\theta',\phi')|^2 \, d\theta' \, dr \le C\, 2^{-12j} |\sin\phi'|^{-5}. \tag{4.24}$$

We can now prove Theorem 4.1

Proof of Theorem 4.1.

Recall that we are assuming that $\xi = (\xi_1,\xi_2,\xi_3) \in \mathcal{D}_C$, $j \ge 0$, $|\ell_1| \le |\ell_2| \le 2^j$. In addition, the assumptions on the support of $\hat\psi_1$ and $\hat\psi_2$ imply that

$$\begin{aligned}
\operatorname{supp} \hat\psi_1(2^{-2j}\xi_1) &\subset \big\{\xi_1 \in [-2^{2j-1}, -2^{2j-4}] \cup [2^{2j-4}, 2^{2j-1}]\big\},\\
\operatorname{supp} \hat\psi_2(2^j \tfrac{\xi_2}{\xi_1} - \ell) &\subset \big\{(\xi_1,\xi_2,\xi_3) : |2^j \frac{\xi_2}{\xi_1} - \ell_1| \le 1\big\},\\
\operatorname{supp} \hat\psi_2(2^j \tfrac{\xi_3}{\xi_1} - \ell) &\subset \big\{(\xi_1,\xi_2,\xi_3) : |2^j \frac{\xi_3}{\xi_1} - \ell_2| \le 1\big\}.
\end{aligned}$$

By representing (ξ_1,ξ_2,ξ_3) using spherical coordinates as $(\lambda\cos\theta\sin\phi, \lambda\sin\theta\sin\phi, \lambda\cos\phi)$, we can write the last two expressions as

$$\begin{aligned}
\operatorname{supp} \hat\psi_2(2^j \tfrac{\xi_2}{\xi_1} - \ell) &\subset \big\{(\lambda,\theta,\phi) :\, 2^{-j}(\ell_1 - 1) \le \tan\theta \le 2^{-j}(\ell_1+1)\big\},\\
\operatorname{supp} \hat\psi_2(2^j \tfrac{\xi_3}{\xi_1} - \ell) &\subset \big\{(\lambda,\theta,\phi) :\, 2^{-j}(\ell_2 - 1) \le \frac{\cot\phi}{\cos\theta} \le 2^{-j}(\ell_2+1)\big\}.
\end{aligned}$$

Notice that $|\theta| \le \frac{\pi}{4}$, so that $1 \le |\cos\theta| \le \frac{\sqrt2}{2}$.

Since $\lambda^2 = \xi_1^2 + \xi_2^2 + \xi_3^2 = \xi_1^2\,(1 + (\tan\theta)^2 + \frac{(\cot\phi)^2}{(\cos\theta)^2})$ and $|\ell_1| \le |\ell_2| \le 2^j$, it is easy to verify that

$$2^{2j-4} \le |\lambda| \le 2^{2j+2}.$$

Thus, using the fact that $\tan\phi \ge 2^{-j}\,\cos\theta\,(\ell_2 - 1)$, it follows that the support of $\Gamma_{j,\ell}$ is contained in the set:

$$\begin{aligned}
W_{j,\ell} = \{(\lambda,\theta,\phi) :\, & 2^{2j-4} \le |\lambda| \le 2^{2j+2},\ \tan^{-1}(2^{-j}(\ell_1-1)) \le \theta \le \tan^{-1}(2^{-j}(\ell_1+1)),\\
& \cot^{-1}(2^{-j}(\ell_2-1)) \le \phi \le \cot^{-1}(2^{-j}(\ell_2+1))\}.
\end{aligned} \tag{4.25}$$

When (λ,θ,ϕ) is contained in the set $W_{j,\ell}$, the variables θ and ϕ are contained in intervals of length $C\,2^{-j}$, which, in the following, will be denoted by I_θ and I_ϕ, respectively. Hence, from (4.19), it follows that ϕ' is contained in an interval $I_{\phi'}$ of length $C\,2^{-j}$. Furthermore, if $(\lambda,\theta,\phi) \in W_{j,\ell}$ and $|\sin\phi'| \ge 2^{1-j}$, then $2^j|\sin\phi'|$ is equivalent to $|\ell_2|$, so that $\ell_2 \ne 0$.

Let $\xi_1 = r\sin\theta'\cos\phi'$, $\xi_2 = -r\cos\theta'$, $\xi_3 = -r\sin\theta'\sin\phi'$. A direct computation shows that the Jacobian of (ξ_1,ξ_2,ξ_3) with respect to (r,θ',ϕ') is $-r^2\sin^2\theta'$. It follows that

$$\begin{aligned}
\int_{\widehat{\mathbb{R}}^3} |\hat f(\xi)|^2\, |\Gamma_{j,\ell}(\xi)|^2\, d\xi &\le \int_{W_{j,\ell}} |\hat f(\xi)|^2\, d\xi\\
&\le \int_{I_{\phi'}} \int_{2^{2j-2}}^{2^{2j+4}} \int_0^{2\pi} |\hat f(r,\theta',\phi')|^2 r^2 \sin^2\theta'\, dr\, d\phi'\\
&\le C\, 2^{4j} \int_{I_{\phi'}} \int_{2^{2j-2}}^{2^{2j+4}} \int_0^{2\pi} |\hat f(r,\theta',\phi')\xi)|^2\, d\theta'\, dr\, d\phi'
\end{aligned} \tag{4.26}$$

We can now use the estimates from Propositions 4.2, 4.3 and 4.5 to complete the proof. Namely, in the non-intersection case, inequality (4.12) gives that

$$\int_{\widehat{\mathbb{R}}^3} |\hat f(\xi)|^2\, |\Gamma_{j,\ell}(\xi)|^2\, d\xi \le C\, 2^{-9j}. \tag{4.27}$$

For the intersection case, with the assumption that $|\sin\phi'| \le 2^{1-j}$, Proposition 4.3 gives that

$$\int_{\widehat{\mathbb{R}}^3} |\hat{f}(\xi)|^2\, |\Gamma_{j,\ell}(\xi)|^2\, d\xi \le C\, 2^{-4j}$$

Finally, for the intersection case, with the assumption that $|\sin\phi'| \ge 2^{1-j}$, inequality (4.24) yields

$$\begin{aligned}\int_{\widehat{\mathbb{R}}^3} |\hat{f}(\xi)|^2\, |\Gamma_{j,\ell}(\xi)|^2\, d\xi &\le C\, 2^{-8j} \int_{I_{\phi'}} |\sin\phi'|^{-5}\, d\phi' \\ &\le C\, 2^{-4j}\, |\ell_2|^{-5}.\end{aligned}$$

Since $|\ell_2| \le 2^j$, the proof of Theorem 4.1 is completed by combining the three inequalities given above. □

Before proving Theorem 3.3, we need some additional estimates involving the derivatives of the surface fragment.

Let $m = (m_1, m_2, m_3)$ and, let us adopt the usual multi-index notation where $|m| = m_1 + m_2 + m_3$, $x^m = x_1^{m_1} x_2^{m_2} x_3^{m_3}$ and $\frac{\partial^m}{\partial\xi^m}\hat{f} = \frac{\partial^{m_1}}{\partial\xi_1^{m_1}} \frac{\partial^{m_2}}{\partial\xi_2^{m_2}} \frac{\partial^{m_3}}{\partial\xi_3^{m_3}} \hat{f}(\xi)$. For a surface fragment f, we may rewrite $x^m f(x)$ as

$$x^m f(x) = 2^{-j|m|} f_m(x),$$

where $f_m(x) = g(x)(2^j x)^m\, w(2^j x)\, \chi_{[x_1 \ge E(x_2,x_3)]}(x)$ is another surface fragment. Since the Fourier transform of $x^m f(x)$ is $i^m \frac{\partial^m}{\partial\xi^m}\hat{f}$, the inequalities (4.5) and (4.27) imply the following estimates:

$$\begin{aligned}\int_{\widehat{\mathbb{R}}^3} \Big|\frac{\partial^m}{\partial\xi^m}\hat{f}(\xi)\Big|^2 |\Gamma_{j,\ell}(\xi)|^2\, d\xi &\le C\, 2^{-j|m|}\, 2^{-4j}(1+|\ell_2|)^{-5}, \quad \text{if there is an intersection,} \\ \int_{\widehat{\mathbb{R}}^3} \Big|\frac{\partial^m}{\partial\xi^m}\hat{f}(\xi)\Big|^2 |\Gamma_{j,\ell}(\xi)|^2\, d\xi &\le C\, 2^{-j|m|}\, 2^{-9j}, \quad \text{if there is no intersection.}\end{aligned}$$

Notice that, for the non-intersection case, the estimate $2^{-j|m|}2^{-9j}$ is the best possible one. However, for the intersection case, the estimate $2^{-j|m|}2^{-4j}(1+|\ell_2|)^{-5}$ can be improved if $m_1 > 0$. The reason is that, on the surface, $|x_1| = |E^j(x_2,x_3)| \le C\, 2^{-j}$. Indeed, using the argument of Proposition 4.5 for the surface fragment $F_m(x)$ (recall that $F_m(x) = f_m(2^{-j}x)$), if the derivatives don't involve x_1, then one obtains the additional factor 2^{-jm_1}. On the other hand, when one takes one derivative with respect to $x_1^{m_1}$, this only produces a factor $2^{-j(m_1-1)}$. However, in this last case, one can compute one additional derivative with respect to the the remaining function in the expression of $F_m(x)$ so that the missing factor 2^{-j} can be compensated, thanks to Plancherel theorem and the observation that, in the Fourier domain, the domain is restricted to the region where $2^{j-1} \le |\xi| \le 2^{j+2}$. Indeed this is the key idea in the proof of Lemma 6.2 in [2] (and hence in the proof of Proposition 4.5).

Using these observations, we obtain the following refinement of Proposition 4.5 valid for $F_m(x)$, in the case where $m_1 = 2$. The behavior for other values of m_1 is similar.

Proposition 4.6. *The Ray Transform of F_m is twice differentiable as a function of s and w and admits the decomposition*

$$\begin{aligned}&(PF(\phi',s,w))_{ss}(\phi',s,w) + (PF(\phi',s,w))_{sw}(\phi',s,w) + (PF(\phi',s,w))_{ww}(\phi',s,w) \\ &= F^0(\phi',s,w) + F^1(\phi',s,w) + F^2(\phi',s,w) + F^3(\phi',s,w),\end{aligned}$$

where, for $q = (q_1, q_2)$ and $|q| = q_1 + q_2$, we have that

$$\begin{aligned}&\|F^0(\phi',s,w)\|^2 \le C\, 2^{-2jm_1} 2^{-2j} |\sin\phi'|^{-5}, \\ &\|\big(F^1(\phi',s,w)\big)_s\|^2 + \|\big(F^1(\phi',s,w)\big)_w\|^2 \le C 2^{-2jm_1} |\sin\phi'|^{-5}, \\ &\sum_{|q|=2} \|\big(F^2(\phi',s,w)\big)_{s^{q_1} w^{q_2}}\|^2 \le C 2^{-2j(m_1-1)} |\sin\phi'|^{-5}, \\ &\sum_{|q|=3} \|\big(F^3(\phi',s,w)\big)_{s^{q_1} w^{q_2}}\|^2 \le C 2^{-2j(m_1-2)} |\sin\phi'|^{-5}.\end{aligned}$$

From the assumption on the support of $\hat{\psi}_1$ and $\hat{\psi}_2$ and the assumption that $|\ell_1| \leq |\ell_2|$, one can easily verify the following inequality (see the proof of Lemma 2.5 in [10] for a similar argument):

$$|\frac{\partial^m}{\xi^m}\Gamma_{j,\ell}(\xi)| \leq C_m\, 2^{-m_1 j} 2^{-|m|j}(1+|\ell_2|)^{m_1}.$$

Since the sets W_{j,ℓ_1,ℓ_2} and $W_{j,\ell_{1'},\ell_2}$ are essentially disjoint for $\ell_1 \neq \ell_{1'}$ (that is, each point lies in a finite number of sets W_{j,ℓ_1,ℓ_2}), using the last inequality we obtain that

$$\sum_{\ell_1=-|\ell_2|}^{|\ell_2|} |\frac{\partial^m}{\partial\xi^m}\Gamma_{j,\ell}(\xi)| \leq C_m\, 2^{-m_1 j} 2^{-|m|j}(1+|\ell_2|)^{m_1}. \tag{4.28}$$

Finally, letting $m_f = (m_{f1}, m_{f2}, m_{f3})$, $m_\gamma = (m_{\gamma 1}, m_{\gamma 2}, m_{\gamma 3})$, using Proposition 4.6 and inequality (4.28) we obtain:

$$\begin{aligned}&\sum_{\ell_1=-|\ell_2|}^{|\ell_2|} \int_{\widehat{\mathbb{R}}^3} |\frac{\partial^{m_f}}{\partial\xi^{m_f}}\hat{f}(\xi)|^2\, |\frac{\partial^{m_\gamma}}{\partial\xi^{m_\gamma}}\Gamma_{j,\ell}(\xi)|^2\, d\xi \\ &\leq C\, 2^{-2j|m_f|}\, \left(2^{-2jm_{f1}} 2^{-4j}\,(1+|\ell_2|)^{-5} + 2^{-9j}\right)\, 2^{-m_{\gamma 1} j} 2^{-|m_\gamma|j}\,(1+|\ell_2|)^{m_{\gamma 1}}.\end{aligned} \tag{4.29}$$

Let L be the second order differential operator defined by:

$$L = \left(I - \left(\frac{2^{2j}}{2\pi\,(1+|\ell_2|)}\right)^2 \frac{\partial^2}{\partial\xi_1^2}\right) \left(1 - \left(\frac{2^j}{2\pi}\right)^2 \frac{\partial^2}{\partial\xi_2^2}\right) \left(1 - \left(\frac{2^j}{2\pi}\right)^2 \frac{\partial^2}{\partial\xi_3^2}\right). \tag{4.30}$$

From inequality (4.29), a routine calculation gives the following theorem which extends the result in Theorem 4.1 (again using the fact that $|\ell_2| \leq 2^j$).

THEOREM 4.7. *Let f be the surface fragment given by expression (4.1) and $\Gamma_{j,\ell}$ be given by (4.4). Then, for each $\xi \in \mathcal{D}_C$, $j \geq 0$ and $-2^j \leq \ell_2 \leq 2^j$, the following estimate holds:*

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

4.4. Proof of Theorem 3.3. Using the preparatory work from the previous sections, we are now ready to provide the proof of Theorem 3.3.

Fix $j \geq 0$ and, for simplicity of notation, let $f = f_Q$. For $\mu \in M_j$, the shearlet coefficients of f can be expressed as

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

where A and B are given after equation (2.1). By the equivalent definition of weak ℓ^1 norm, the theorem is proved provided we show that

$$\#\{\mu \in M_j : |\langle f, \psi_\mu\rangle| > \epsilon\} \leq C\, 2^{-2j}\, \epsilon^{-1}. \tag{4.31}$$

Observe that

$$\begin{aligned}\xi A^{-j} B^{-\ell} k &= \begin{pmatrix}\xi_1 & \xi_2 & \xi_3\end{pmatrix} \begin{pmatrix}2^{-2j} & 0 & 0\\ 0 & 2^{-j} & 0\\ 0 & 0 & 2^{-j}\end{pmatrix} \begin{pmatrix}1 & -\ell_1 & -\ell_2\\ 0 & 1 & 0\\ 0 & 1 & 1\end{pmatrix} \begin{pmatrix}k_1\\ k_2\\ k_3\end{pmatrix}\\ &= (k_1 - k_2\ell_1 - k_3\ell_2)\, 2^{-2j}\xi_1 + k_2 2^{-j}\xi_2 + k_3 2^{-j}\xi_3.\end{aligned} \tag{4.32}$$

Let L be the second order differential operator defined by (4.30). It is easy to check that

$$L\left(e^{2\pi i\xi A^{-j}B^{-\ell}k}\right) = \begin{cases} \left(1+(\frac{|\ell_2|}{(1+|\ell_2|)})^2(\frac{k_1}{|\ell_2|}-\frac{k_2\ell_1}{|\ell_2|}\pm k_3)^2\right)(1+k_2^2)(1+k_3^2)\,e^{2\pi i\xi A^{-j}B^{-\ell}k} & \text{if } \ell_2\neq 0 \\ (1+k_1^2)(1+k_2^2)(1+k_3^2)\,e^{2\pi i\xi A^{-j}B^{-\ell}k} & \text{if } \ell_2=0, \end{cases} \tag{4.33}$$

where we have $\pm k_3$ depending on whether ℓ_2 is positive or negative. Using integration by parts, we have:

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

To analyze this quantity, we will consider separately the case $\ell\neq 0$ and $\ell=0$.

Case 1: $\ell_2\neq 0$**.** In this case, using (4.33), we have that

$$L^{-1}\left(e^{2\pi i\xi A^{-j}B^{-\ell}k}\right) = G(k,\ell)^{-1}\,e^{2\pi i\xi A^{-j}B^{-\ell}k}, \tag{4.34}$$

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, we have 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 $[-\frac12,\frac12]A^jB^\ell$, and the function $\Gamma_{j,\ell}(\xi)$ is supported on this set, then

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

This implies that

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

From the definition of R_K, it follows that

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

Thus, by Theorem 4.7, we have that

$$\sum_{\ell_1=-|\ell_2|}^{|\ell_2|}\sum_{k\in R_K}|\langle f,\psi_\mu\rangle|^2 \le C\,L_K^{-2}\,2^{-4j}(1+|\ell_2|)^{-5}, \tag{4.35}$$

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 $|\ell_1| \le |\ell_2|$, it is clear that $N_{j,\ell,K}(\epsilon) \le C\,(1+|\ell_2|)^2$ (C is independent of ℓ_1) and, hence, $\sum_{\ell_1=-|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \le C\,(1+|\ell_2|)^3$. Using the new notation, from (4.35) we have that

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

This implies that

$$\sum_{\ell_1=-|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \le C\, \min\left((|\ell|+1)^3, L_K^{-2}\, 2^{-4j}\, \epsilon^{-2}(1+|\ell|)^{-5}\right). \tag{4.36}$$

Using (4.36) we will now show that:

$$\sum_{\ell_2=-2^j}^{2^j} \sum_{\ell_1=-|\ell_2|}^{|\ell_2|} N_{j,\ell,K}(\epsilon) \le C\, L_K^{-1}\, 2^{-2j}\, \epsilon^{-1}. \tag{4.37}$$

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

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

which gives (4.37).

Since $\sum_{K\in\mathbb{Z}^3} L_K^{-1} < \infty$, using (4.37) we then have that

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

and, thus, (4.31) holds.

Case 2: $\ell_2 = 0$**.** In this case, we also have $\ell_1 = 0$. It follows that

$$L^{-1}\left(e^{2\pi i \xi A^{-j}k}\right) = (1+k_1^2)^{-1}(1+k_2^2)^{-1}(1+k_3^2)^{-1}\, e^{2\pi i \xi A^{-j}k}.$$

Let $L_k = (1+k_1^2)\,(1+k_2^2)\,(1+k_3^2)$. It is clear that also in this case $\sum_{k\in\mathbb{Z}^3} L_k^{-1} < \infty$. We have

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

or, equivalently, that

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

It follows that

$$\sum_{k\in\mathbb{Z}^3} L_k^2\, |\langle f, \psi_{j,k}\rangle|^2 = \int_{\widehat{\mathbb{R}}^3} \left|L\left(\hat f(\xi)\, \Gamma_{j,\ell}(\xi)\right)\right|^2\, d\xi \le C 2^{-4j}.$$

In particular, for each $k \in \mathbb{Z}^3$, we have $|\langle f, \psi_{j,k}\rangle| \le C\, L_k^{-1}\, 2^{-2j}$ and hence $\sum k \in \mathbb{Z}^3 |\langle f, \psi_{j,k}\rangle| \le C\, 2^{-2j}$, or $\|\langle f, \psi_{j,k}\rangle\|_{l^1} \le C\, 2^{-2j}$ which implies $\|\langle f, \psi_{j,k}\rangle\|_{wl^1} \le C\, 2^{-2j}$.

This completes the proof of the theorem. $\square$

4.5. Remark on the proof of Theorem 3.3. In the proof of Theorem 3.3, it was assumed that the boundary surface contains the origin and has normal direction $(1,0,0)$ at the origin. In general, if this is not the case, one can "transform" any given surface into one which satisfies the above assumptions by using a combination of translation and rotation. Obviously the translation has no impact on the proof which was given above. It only remains to explain the effect of rotations, since the shearlet system is not invariant with respect to rotations.

As in the proof of Theorem 4.1, for $\xi = (\xi_1, \xi_2, \xi_3) \in \mathcal{D}_C$, let

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

Recall that the support of $\Gamma_{j,\ell}$ is contained in a set $W_{j,\ell}$ which, using spherical coordinates, is given by:

$$W_{j,\ell} = \{(\lambda,\theta,\phi):\, 2^{2j-4} \le |\lambda| \le 2^{2j+2},\, \tan^{-1}(2^{-j}(\ell_1-1)) \le \theta \le \tan^{-1}(2^{-j}(\ell_1+1)),\\ \cot^{-1}(2^{-j}(\ell_2-1)) \le \phi \le \cot^{-1}(2^{-j}(\ell_2+1))\}.$$

Hence, using spherical coordinates, a rotation can be realized by the mapping $(\lambda,\theta,\phi) \to (\lambda, \theta-\theta_0, \phi-\phi_0)$, where θ_0 and ϕ_0 are the two rotation angles. Also recall that, inside $W_{j,\ell}$, we have $2^j \sin\theta \approx \ell_1$ and $2^j\cos\phi \approx \ell_2$, which implies that there are $\overline{\ell}_1, \overline{\ell}_2$ such that $2^j\sin\theta_0 \approx \overline{\ell}_1$, $2^j\cos\phi_0 \approx \overline{\ell}_2$. Let $\Gamma^0_{j,\ell}(\xi)$, $W^0_{j,\ell}$ and $\hat{f}^0$ be the images of $\Gamma_{j,\ell}(\xi)$, $W_{j,\ell}$ and $\hat{f}$ under the rotation by θ_0 and ϕ_0. Then it is easy to see that

$$W^0_{j,\ell} \approx \{(\lambda,\theta,\phi):\, 2^{2j-4} \le |\lambda| \le 2^{2j+2},\, \tan^{-1}(2^{-j}(\ell_1+\overline{\ell}_1-1)) \le \theta \le \tan^{-1}(2^{-j}(\ell_1+\overline{\ell}_1+1)),\\ \cot^{-1}(2^{-j}(\ell_2+\overline{\ell}_2-1)) \le \phi \le \cot^{-1}(2^{-j}(\ell_2+\overline{\ell}_2+1))\}$$

From these observations, using the same argument as in the proof of Theorem 4.1, one can show that

$$\int_{\widehat{\mathbb{R}}^3} |\hat{f}^0(\xi)|^2\,|\Gamma^0_{j,\ell}(\xi)|^2\,d\xi \le C\,2^{-4j}(1+|\ell_2+\overline{\ell}_2|)^{-5}.$$

The rest of the argument is exactly the same as in the proof of Theorem 3.3, where ℓ_2 is replaced by $\ell_2+\overline{\ell}_2$.

4.6. Analysis of the coarse scale. At the beginning of Section 4.1, we assumed that the scale parameter j is large enough, i.e., $j > j_0$ for some $j_0 > 0$. The situation where $j \le j_0$ is much simpler. In fact, if f_Q is an edge fragment, then a trivial estimate shows that

$$\|f_Q\|_2 = \left(\int_{\operatorname{supp} w_Q} |f_Q(x)|^2\,dx\right)^{1/2} \le C\,|\operatorname{supp} w_Q|^{1/2} = C\,2^{-\frac{3}{2}j}.$$

It follows that $\|\langle f_Q,\psi_\mu\rangle\|_{\ell^2} \le C\,\|f_Q\|_2 \le C\,2^{-\frac{3}{2}j}$. To deduce an ℓ^1 type estimate, we notice that

$$\|\langle f_Q,\psi_\mu\rangle\|_{\ell^p} \le N^{\frac{1}{p}-\frac{1}{2}}\,\|\langle f_Q,\psi_\mu\rangle\|_{\ell^2},$$

is valid for any sequence $\{\langle f_Q,\psi_\mu\rangle\}$ of N elements. Since, at scale 2^{-j}, there are about 2^{2j} shearlet elements in $\mathcal{Q}^0_j$, it follows that

$$\|\langle f_Q,\psi_\mu\rangle\|_{\ell^1} \le C\,2^j\,2^{-\frac{3}{2}j} = C\,2^{-\frac{1}{2}j}.$$

This satisfies Theorem 3.3 for $j \le j_0$.

4.7. Proof of Theorem 3.4. The proof of Theorem 3.4 follows essentially the idea from the 2-dimensional case in [10]. We start by proving the following lemmata which will be useful in the following.

LEMMA 4.8. *Let $f = g\,w_Q$, where $g \in \mathcal{E}^2(A)$ and $Q \in \mathcal{Q}^1_j$ and $W_{j,\ell}$ be given by (4.25). Then*

$$\int_{W_{j,\ell}} |\hat{f}(\xi)|^2\,d\xi \le C\,2^{-11j}. \tag{4.38}$$

Proof. The following proof adapts [10, Lemma 2.6].

The function f belongs to $C_0^2(\mathbb{R}^3)$ 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

$$\int_{\widehat{\mathbb{R}}^3} |\hat{h}_1(\xi)|^2\, d\xi = \int_{\mathbb{R}^3} |h_1(x)|^2\; dx \le C\, 2^{-3j}.$$

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

$$\int_{\widehat{\mathbb{R}}^3} |2\pi\xi_1\, \hat{h}_2(\xi)|^2\, d\xi = \int_{\mathbb{R}^3} \left|\frac{\partial}{\partial x_1} h_2(x)\right|^2\; dx \le C\, 2^{j},$$

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

$$\int_{W_{j,\ell}} |\hat{h}_2(\xi)|^2\, d\xi \le C\, 2^{-3j}.$$

Finally, observing that $\|\frac{\partial^2}{\partial x_1^2} h_3\|_\infty \le C\, 2^{4j}$, it follows that $\int_{\widehat{\mathbb{R}}^3} |\hat{h}_3(\xi)|^2\, d\xi \le C\, 2^{5j}$ and, thus,

$$\int_{W_{j,\ell}} |\hat{h}_3(\xi)|^2\, d\xi \le C\, 2^{-3j}.$$

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

$$\int_{W_{j,\ell}} |\hat{f}(\xi)|^2\, d\xi \le C\, 2^{-11j}.$$

This completes the proof. ∎

LEMMA 4.9. *Let $m = (m_1, m_2, m_3) \in \overline{\mathbb{N}} \times \overline{\mathbb{N}} \times \overline{\mathbb{N}}$, $\xi = (\xi_1, \xi_2, \xi_3) \in \widehat{\mathbb{R}}^3$ and $\Gamma_{j,\ell}$ be given by (4.4), where $\ell = (\ell_1, \ell_2)$. Then*

$$\sum_{\ell_1=-2^j}^{2^j} \sum_{\ell_2=-2^j}^{2^j} \left|\frac{\partial^m}{\partial \xi^m}\, \Gamma_{j,\ell_1,\ell_2}(\xi)\right|^2 \le C_m\, 2^{-2|m|j},$$

where C_m is independent of j and ξ and $|m| = m_1 + m_2 + m_3$.

Proof. Observe that $W_{j,\ell} \cap W_{j,\ell+\ell'} = \emptyset$, whenever $|\ell_1'| \ge 3$ or $|\ell_2'| \ge 3$. Since $|\ell_1|, |\ell_2| \le 2^j$, the lemma then follows from (4.28). ∎

LEMMA 4.10. *Let $f = g\, w_Q$, where $g \in \mathcal{E}^2(A)$ and $Q \in \mathcal{Q}_j^1$ and set*

$$T = \left(I - \frac{2^j}{(2\pi)^2}\, \Delta\right), \tag{4.39}$$

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

$$\int_{\widehat{\mathbb{R}}^3} \sum_{\ell_1=-2^j}^{2^j} \sum_{\ell_2=-2^j}^{2^j} \left|T^2\left(\hat{f}\, \Gamma_{j,\ell_1,\ell_2}\right)(\xi)\right|^2\, d\xi \le C\, 2^{-11j}.$$

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

$$\Delta^N \left(\hat f \, \Gamma_{j,\ell}\right) = \sum_{|\alpha|+|\beta|=2N} C_{\alpha,\beta} \left(\frac{\partial^\alpha}{\partial \xi^\alpha}\, \hat f\right) \left(\frac{\partial^\beta}{\partial \xi^\beta}\, \Gamma_{j,\ell}\right),$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, $\beta = (\beta_1, \beta_2, \beta_3)$, and $\alpha_i, \beta_i \in \mathbb{N}$. Also notice that, by Lemma 4.9, we have that

$$\int_{\widehat{\mathbb{R}}^3} \sum_{\ell_1,\ell_2=-2^j}^{2^j} \left|\frac{\partial^\alpha}{\partial \xi^\alpha}\, \hat f(\xi)\right|^2 \left|\frac{\partial^\beta}{\partial \xi^\beta}\, \Gamma_{j,\ell}(\xi)\right|^2 d\xi \le C_\beta\, 2^{-2|\beta| j} \int_{W_{j,\ell}} \left|\frac{\partial^\alpha}{\partial \xi^\alpha}\, \hat f(\xi)\right|^2 d\xi.$$

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 4.8, $g(x)\, w_\alpha(2^j x)$ obeys the estimate (4.38). Thus, observing that $\frac{\partial^\alpha}{\partial \xi^\alpha}\, \hat f(\xi)$ is the Fourier transform of $(-2\pi i x)^\alpha f(x)$, we have that

$$\int_{W_{j,\ell}} \left|\frac{\partial^\alpha}{\partial \xi^\alpha}\, \hat f(\xi)\right|^2 d\xi \le C_\alpha\, 2^{-2j|\alpha|}\, 2^{-11j}.$$

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

$$\int_{\widehat{\mathbb{R}}^3} \sum_{\ell_2=-2^j}^{2^j} \sum_{\ell_1=-2^j}^{2^j} \left|\frac{\partial^\alpha}{\partial \xi^\alpha}\, \hat f(\xi)\right|^2 \left|\frac{\partial^\beta}{\partial \xi^\beta}\, \Gamma_{j,\ell}(\xi)\right|^2 d\xi \le C_{\alpha,\beta}\, 2^{-11j}\, 2^{-4jN}. \tag{4.40}$$

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 (4.40) and Lemma 4.9. □

We can now prove Theorem 3.4.

Proof of Theorem 3.4.

For T given by (4.39) and $\ell = (\ell_1, \ell_2)$, a direct computation gives that

$$T\left(e^{2\pi i \xi A^{-j} B^{-\ell} k}\right) = \left(1 + 2^{-2j}(k_1 - k_2\, \ell_1 - k_3\, \ell_2)^2 + k_2^2 + k_3^2\right)\, e^{2\pi i \xi A^{-j} B^{-\ell} k}. \tag{4.41}$$

Hence,

$$T^2\left(e^{2\pi i \xi A^{-j} B^{-\ell} k}\right) = \left(1 + 2^{-2j}(k_1 - k_2\, \ell_1 - k_3\, \ell_2)^2 + k_2^2 + k_3^2\right)^2\, e^{2\pi i \xi A^{-j} B^{-\ell} k}. \tag{4.42}$$

Fix $j \ge 0$ and let $f = f_Q$, with $Q \in \mathcal{Q}_j^1$. Then, using integration by parts as in the proof of Theorem 3.3, from (4.42) it follows that

$$\langle f, \psi_\mu \rangle = |\det A|^{-j}\left(1 + 2^{-2j}(k_1 - k_2\, \ell_1 - k_3\, \ell_2)^2 + k_2^2 + k_3^2\right)^{-2} \int_{\widehat{\mathbb{R}}^2} T^2\left(\hat f(\xi)\, \Gamma_{j,\ell}(\xi)\right)\, e^{2\pi i \xi A^{-j} B^{-\ell} k}\, d\xi.$$

Let $K = (K_1, K_2, K_3) \in \mathbb{Z}^3$ and R_K be the set

$$R_K = \{(k_1, k_2, k_2) \in \mathbb{Z}^3 : k_2 = K_2, k_3 = K_3, 2^{-j}(k_1 - K_2\ell_1 - K_3\ell_2) \in [K_1, K_1 + 1]\}.$$

Observe that, for each K and each fixed ℓ, there are only $1 + 2^j$ choices for k_1 in R_K. In fact, $R_K = \{k_1 : 2^j K_1 \le k_1 - K_2\ell_1 - K_3\ell_2 \le 2^j(K_1 + 1)\}$. Hence the number of terms in R_K is bounded by $1 + 2^j$. Also notice that, as in the proof of Theorem 3.3, we can take advantage of the fact that, for j, ℓ fixed, the set $\{|\det A|^{-j/2} e^{2\pi i A^{"-j} B^{ell} k} : k \in \mathbb{Z}^3\}$ is an orthonormal basis for the L^2 functions supported in the set $[-\frac{1}{2},\frac{1}{2}]^3 A^j B^\ell$. Thus, using this observation and the fact that the function $\Gamma_{j,\ell}$ is supported on the set $[-\frac{1}{2},\frac{1}{2}]^3 A^j B^\ell$, we have that

$$\sum_{k\in R_K} |\langle f,\psi_\mu\rangle|^2 \le C\,\left(1+K_1^2+K_2^2+K_3^2\right)^{-4}\int_{\widehat{\mathbb{R}}^3}\left|T^2\left(\hat f\,\Gamma_{j,\ell}\right)(\xi)\right|^2\,d\xi.$$

From this inequality, using Lemma 4.10, we have that

$$\begin{aligned}\sum_{\ell_2=-2^j}^{2^j}\sum_{\ell_1=-2^j}^{2^j}\sum_{k\in R_K}|\langle f,\psi_\mu\rangle|^2 &\le C\,(1+K^2)^{-4}\int_{\widehat{\mathbb{R}}^2}\sum_{\ell_2=-2^j}^{2^j}\sum_{\ell_1=-2^j}^{2^j}\left|T^2\left(\hat f\,\Gamma_{j,\ell}\right)(\xi)\right|^2\,d\xi\\ &\le C\,(1+K^2)^{-4}\,2^{-11j}.\end{aligned}\tag{4.43}$$

For any $N\in\mathbb{N}$, the Hölder inequality yields:

$$\sum_{m=1}^N |a_m| \le \left(\sum_{m=1}^N |a_m|^2\right)^{\frac{1}{2}}\,N^{\frac{1}{2}}.\tag{4.44}$$

Since the cardinality of R_K is bounded by $1+2^j$, it follows from (4.43) and (4.44) that

$$\sum_{\ell_2=-2^j}^{2^j}\sum_{\ell_1=-2^j}^{2^j}\sum_{k\in R_K}|\langle f,\psi_\mu\rangle| \le C\,\left(2^{3j}\right)^{\frac{1}{2}}(1+K^2)^{-2}\,2^{-\frac{11}{2}j}\le C\,2^{-4j}.$$

Thus, for $f=f_Q$, with $Q\in\mathcal{Q}_j^1$, we have that:

$$\sum_{\mu\in M_j}|\langle f,\psi_\mu\rangle|\le C\,2^{-4j}.$$

5. Extension to the Piecewise C^2 Setting. We briefly outline how to extend Theorem 3.2 to the situation where the surface boundary is not C^2 but piecewise C^2. For reason of space, only a brief sketch of the argument can be presented in this paper. In particular, for simplicity, we will only describe the argument in the bivariate case: the same ideas can be carried over from the bivariate to the trivariate case without significant changes.

Let S be the boundary curve and let us assume that S has a corner point centered at the origin. Assume that, near the origin, the curve is parametrized as $S=\{(E(x_2),x_2),\ x_2\in(-1,1)\}$ and that $(0,0)$ is the only corner point of S. Let $x_1=L_1\,x_2$, where $x_2\in[0,1)$, and $x_1=L_2\,x_2$, where $x_2\in(-1,0]$, be the two tangent lines of S at $(0,0)$. The region $G=\{(x_1,x_2),x_1\ge E(x_2),\ x_2\in(-1,1)\}\bigcap[-1,1]^2$ can be expressed as the union of three subregions $G_1,\ G_2$ and G_3, where

$$\begin{aligned}G_1&=\{(x_1,x_2),E(x_2)\le x_1\le L_1x_2,\ x_2\in[0,1)\}\bigcap[-1,1]^2,\\ G_2&=\{(x_1,x_2),E(x_2)\le x_1\le L_2x_2,\ x_2\in(-1,0]\}\bigcap[-1,1]^2,\\ G_3&=\{(x_1,x_2),x_1\ge L_1x_2,\ x_2\in[0,1);\ \ x_1\ge L_2x_2,\ x_2\in(-1,0]\}\bigcap[-1,1]^2\end{aligned}$$

For G_1, an argument similar to Section 4.5 can be used to translate and rotate the domain so that, in the new domain, the tangent line is given by the equation $x_1=0$. Next, one can use the same argument valid for the smooth surface to prove the desired result. Similarly one can deal with G_2. For G_3 a different argument is needed since it contains a corner edged by two straight lines with (possible) different slopes. In this case, however, one can follow the argument in [2, Sec. 9] to complete the proof.

Acknowledgments. The authors acknowledge support from NSF grant DMS 1008900/1008907; the second author also acknowledges support from NSF grant DMS 0746778.

REFERENCES

[1] E. J. Candès, L. Demanet, D. L. Donoho and L. Ying, *Fast discrete curvelet transforms*, Multiscale Model. Simul., **5** 861-899, 2005 .

[2] 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), pp. 219–266.

[3] 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.

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

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

[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), pp. 298–318

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

[12] K. Guo, and D. Labate, *Analysis and Detection of Surface Discontinuities using the 3D Continuous Shearlet Transform*, to appear in Appl. Comput. Harmon. Anal., (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), pp. 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), pp. 231–249.

[15] G. Kutyniok and W.-Q 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] J. Lemvig, *Optimally Sparse Approximations of Functions in $L^2(\mathbb{R}^3)$ with C^α Singularities using Shearlet*, Oberwolfach Report **44** (2010)

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

[20] S. Mallat,]*Geometrical grouplets* Applied and Computational Harmonic Analysis, **26**(2) (2009), 161–180.

[21] F. Natterer, and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM Monographs on Mathematical Modeling and Computation, Philadelphia, 2001.

[22] V. M. Patel, G. R. Easley and D. M. Healy, Jr., *Shearlet-Based Deconvolution*, IEEE Transactions on Image Processing **18** (2009), 26732685.

[23] G. Peyre and S. Mallat, Surface compression with geometric bandelets, ACM Transactions on Graphics **24**(3) (2005), 601–608.

[24] G. Peyre and S. Mallat, Orthogonal Bandlet Bases for Geometric Images Approximation, Communications on Pure and Applied Mathematics **61**(9) (2008), 1173–1212.

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

[26] 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.