# Geometric Separation of Singularities using Combined Multiscale Dictionaries

Kanghui Guo[1] and Demetrio Labate[2]

November 11, 2014

**Abstract**

Several empirical results appeared in the literature during the last decade have shown that it is often possible to separate images and other multidimensional data into geometrically distinct constituents. A rigorous mathematical analysis of the *geometric separation* problem in the two-dimensional setting was recently introduced by Donoho and Kutyniok [6], who proposed a mathematical framework to separate point and smooth curve singularities in 2D images using a combined dictionary consisting of curvelets and wavelets. In this paper, we adapt their approach and introduce a novel argument to extend geometric separation to the three-dimensional setting. We show that it is possible to separate point and piecewise linear singularities in 3D using a combined dictionary consisting of shearlets and wavelets. Our new approach takes advantage of the microlocal properties of the shearlet transform and has the ability to handle singularities containing vertices and corner points, which cannot be handled using the original arguments.

*Key words and phrases*: cluster coherence, geometric separation, $\ell^1$ minimization, shearlets, sparse representations, wavelets.
*AMS Mathematics Subject Classification*: 42C15, 42C40.

## 1 Introduction

Data found in applications ranging from astronomy through remote sensing and biomedical imaging can be frequently modeled as superpositions of several distinct geometric components. Starck et al. [26, 27], in particular, proposed a very effective algorithmic approach, called Morphological Component Analysis (MCA), which assumes that a signal is the linear mixture of several constituents, the so-called *morphological components*, each one endowed with specific geometric properties. Under the assumption that the various morphological components are sufficiently distinct and that each one is sparsely represented in a specific basis but not in the other ones, MCA algorithms are very successful in separating the various components, as illustrated by several numerical applications.

The idea of using combined-basis representations and taking advantage of their sparsity properties has a long history in applied harmonic analysis and image processing. Some pioneering work about combined-basis representations can be found in the papers of Coifman and Wickerhauser [4] and Mallat and Zhang [24]. Another fundamental contribution to formalize these ideas was the introduction of Basis Pursuit [3], which established $\ell^1$-norm minimization as an effective method to promote sparse representations from multiple bases. In more recent years, several other mostly empirical papers have further exploited this point of view and provided remarkable applications to problems from image processing. In addition to the work cited above, we recall, for example, the work in [8, 25, 29, 31][1].

[1]Department of Mathematics, Missouri State University, Springfield, Missouri 65804, USA. E-mail: KanghuiGuo@MissouriState.edu
[2]Department of Mathematics, University of Houston, Houston, Texas 77204, USA. E-mail: dlabate@math.uh.edu

[1]Note: as also remarked in [6], the geometric separation we consider here is rather different from the *separation of texture from smooth structure* where sparsity in the representation does not play a relevant role

In order to provide a more rigorous mathematical formalization of the problem of data separation into geometrically distinct components, Donoho and Kutyniok [6] recently proposed a theoretical framework for the geometric separation of point and curve singularities in 2D (cf. also related work in [19]). The underlying ansatz is that the success of MCA "stems from an interplay between geometric properties of objects to be separated and the harmonic analysis for singularities of various geometric types" (cf. [6]). Hence, as a mathematical idealization of a large class of two-dimensional objects, they consider distributions of the form $f = \mathcal{P} + \mathcal{T}$, where $\mathcal{P}$ is a collection of point-like singularities and $\mathcal{T}$ is a cartoon-like image, that is, a planar region enclosed by a smooth closed curve, hence producing an 'edge' singularity along the curve. Their goal is to find a highly sparse representation of $f$, that is, an expansion into a basis (or frame) that can accurately represent $f$ using a relatively small number of terms. It is well-known that the type of basis which best sparsifies a function or distribution depends on the geometry of its singularities. Wavelets, in particular, provide very sparse representations of point-like singularities; curvelets [1] and sherlets [10, 23] provide very sparse representations of curve-like singularities. However, neither wavelets nor curvelets (or sherlets) *alone* can provide a very sparse representation of $f = \mathcal{P} + \mathcal{T}$, where different types of singularities appear jointly. The natural alternative is to look for an appropriate representation of $f$ in terms of a combined basis of wavelets and curvelets.

Donoho and Kutyniok developed an ingenious machinery to obtain a sparse representation of $f$ with respect to a joint wavelet-curvelet dictionary, where sparsity is enforced via a procedure of minimization of the expansion coefficients in the $\ell^1$-norm. Their viewpoint for the analysis of the singularities is derived from microlocal analysis, and, specifically, the observation that while points and curves may overlap spatially, they are separated microlocally. In order to reveal this separation, the authors observe that the elements of the wavelet system are 'incoherent' to the curvelet system, meaning that they have limited overlap in phase space. This is used to derive an asymptotic estimate showing that, at very fine scales, the pointlike structure of $f$ is essentially captured by the wavelet basis while curvelike structure of $f$ is essentially captured by the curvelet basis. The use of the $\ell^1$-norm is critical in this context to enforce the desired sparsity of the expansion of $f$ with respect to the combined dictionary.

In this paper, we extend the framework of geometric separation of Donoho and Kutyniok to the three-dimensional setting by considering distributions with domain in $\mathbb{R}^3$ containing two different types of singularities: point singularities and singularities along polyhedral surfaces. We adapt the general approach from [6] consisting of using a combined dictionary of incoherent bases and, in particular, we adopt the important notion of *cluster coherence*. However, to prove the geometric separation of the two types of singularities, we do not use the argument in [6] which maps the singularities in phase space, since this argument does not extend to the 3D setting. Instead, we introduce a novel method and more streamlined approach which is based on techniques previously developed by the authors and their collaborators for the geometric characterization of edge singularities in terms of the shearlet transform [11, 16, 20]. Because of this, we choose a shearlet system, rather than a curvelet system, to sparsely represent singularities along surfaces; we use a standard wavelet system to handle point-like singularities.

In addition to solving the 3D geometric separation problem, an advantage of our new approach is the ability to deal with singularities including 3D edges and vertices, which pose an additional difficulty in the geometric separation problem and cannot be handled using the original arguments from [6].

The rest of the paper is organized as follows. After setting some useful notation, we formulate the geometric separation problem and state our main theorem in Section 2. We present the proof of the main theorem in Section 3.

## 1.1 Notation.

In the following, we adopt the convention that $x \in \mathbb{R}^3$ is a column vector, i.e., $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, and that $\xi \in \widehat{\mathbb{R}}^3$ (in the frequency domain) is a row vector, i.e., $\xi = (\xi_1, \xi_2, \xi_3)$. A vector $x$ multiplying a matrix $A \in GL_3(\mathbb{R})$ on the right is understood to be a column vector, while a vector $\xi$ multiplying $A$ on the left is a row vector.

Thus, $Ax \in \mathbb{R}^3$ and $\xi A \in \widehat{\mathbb{R}}^3$. The *Fourier transform* of $f \in L^1(\mathbb{R}^3)$ is defined as

$$\hat{f}(\xi) = \int_{\mathbb{R}^3} f(x)\, e^{-2\pi i \xi x}\, dx,$$

where $\xi \in \widehat{\mathbb{R}}^3$, and the inverse Fourier transform is

$$\check{f}(x) = \int_{\widehat{\mathbb{R}}^3} f(\xi)\, e^{2\pi i \xi x}\, d\xi.$$

We will use the notation $f \simeq g$ if there exist constants $0 < C_1 \leq C_2 < \infty$, independent of $x$, such that $C_1\, g(x) \leq f(x) \leq C_2\, g(x)$.

## 2 The geometric separation problem

As a mathematical model of images and other multidimensional data, it is useful to consider functions or distributions that contain singularities with different types of geometry, such as points and curves if the domain is $\mathbb{R}^2$ or points and surfaces if the domain is $\mathbb{R}^3$. In this paper, we consider idealized three-dimensional objects of the form $f = \mathcal{P} + \mathcal{T}$, where $\mathcal{P}$ is a collection of pointwise singularities in $\mathbb{R}^3$ and $\mathcal{T}$ is a cartoon-like functions in $\mathbb{R}^3$, that we use to model singularities along a class of surface boundaries.

We are interested in finding a representation that is able to decompose $f$ into its distinct geometric components. As mentioned above, we can find bases that are ideally suited to specific types of singularities. Wavelets, in particular, offer optimally sparse representations, in a certain sense, for functions with point singularities, while shearlets were shown to provide optimally sparse representations for functions with discontinuities along piecewise smooth edges and surface boundaries [13, 21]. However, neither wavelets nor shearlets alone (and no other single basis or traditional linear representation methods) are very efficient at representing $f = \mathcal{P} + \mathcal{T}$. This observation leads naturally to consider a multiple-basis dictionary comprising both wavelets and shearlets. Among all possible representations of $f$ within this dictionary, we look for an 'ideally' sparse representation where wavelets are used to sparsely represent $\mathcal{P}$ and shearlets to sparsely represent $\mathcal{T}$.

Let us be more precise about the statement of the problem and the singularities we consider. Following the general idea from [6], we take $\mathcal{P}$ to be of the form

$$\mathcal{P} = \sum_{i=1}^{I} |x - x_i|^{-1}. \tag{2.1}$$

This defines a function that is smooth away from the singular points $\{x_i : 1 \leq i \leq I\} \subset \mathbb{R}^3$. For $\mathcal{T}$, we consider special cartoon-like images

$$\mathcal{T} = \sum_{j=1}^{J} \chi_{B_j}, \tag{2.2}$$

where each $B_j$ is a polyhedron, that is, a compact region in $\mathbb{R}^3$ with polygonal boundaries. The reason for choosing the exponent $-1$ in $\mathcal{P}$ is that we want to match the energies of $\mathcal{P}$ and $\mathcal{T}$ at each scale $2^{-2j}$, $j \in \mathbb{Z}$. That is, we want to make the two types of singularities comparable at each scale, so that the separation is challenging at every scale and it is not possible to trivially separate the two components of $f$ at different scales.

To justify our observation about the matching energies, note in fact that $\widehat{\mathcal{P}}(\xi) \simeq |\xi|^{-2}$ (cf. [30, Ch.4]), and this easily implies that $\int_{2^{2j}}^{2^{2j+2}} |\widehat{\mathcal{P}}(\xi)|^2 d\xi \simeq 2^{-2j}$. For the cartoon-like component, we will show below in Lemma 3.3 that $\mathcal{T}$ satisfies the same type of estimate so that $\int_{2^{2j}}^{2^{2j+2}} |\widehat{\mathcal{T}}(\xi)|^2 d\xi \simeq 2^{-2j}$.

Following the language in [6], the *geometric separation problem* can be stated as follows. *From the observation of $f = \mathcal{P} + \mathcal{T}$, we want to recover the unknown components $\mathcal{P}$ and $T$ of $f$ where we only know that they are of the form* (2.1) *and* (2.2).

To solve this problem, we will adopt the successful strategy of Donoho and Kutyniok [6] based on $\ell^1$ minimization. More precisely, we will expand $f$ with respect to a representation consisting of the union of a Parseval frame of wavelets in $L^2(\mathbb{R}^3)$ and a Parseval frame of shearlets in $L^2(\mathbb{R}^3)$ and enforce sparsity by minimizing the representation coefficients in the $\ell^1$-norm. As mentioned above, the sparsity-inducing properties of the $\ell^1$-norm are well known in applied harmonic analysis. We recall that these properties play a fundamental role in the celebrated theory of *compressed sensing* (cf. [2, 5]).

Let us now define the wavelet and shearlet systems that we will use to represent $f = \mathcal{P} + \mathcal{T}$.

## 2.1 A Parseval frame of 3D wavelets

For the wavelet system, we will consider a *Parseval frame of Lemariè-Meyer wavelets* (cf. [18] for more details about this class of wavelets) in $L^2(\mathbb{R}^3)$ that we denote as $\Phi = \{\phi_\lambda : \lambda \in \Lambda\}$, for $\Lambda = \{\lambda = (j,k), j \geq -1, k \in \mathbb{Z}^3\}$. Here the functions $\phi_\lambda = \phi_{j,k} \in L^2(\mathbb{R}^3)$ are defined in the Fourier domain by

$$\widehat{\phi}_{j,k}(\xi) = \begin{cases} 2^{-3j}\, W(2^{-2j}\xi)\, e^{2\pi i 2^{-2j}\xi k}, & \text{for } j \geq 0, \\ \widetilde{W}(\xi)\, e^{2\pi i \xi k}, & \text{for } j = -1, \end{cases}$$

where $W, \widetilde{W} \in C_0^\infty(\mathbb{R}^3)$ satisfy the condition

$$\widetilde{W}^2(\xi) + \sum_{j\geq 0} |W(2^{-2j}\xi)|^2 = 1, \quad \text{for a.e. } \xi \in \widehat{\mathbb{R}}^3.$$

In particular, we assume that the window function $W$ has support $\operatorname{supp}(W) \subset [-\frac{1}{2}, \frac{1}{2}]^3 \setminus [-\frac{1}{16}, \frac{1}{16}]^3$ so that the functions $W_j = W(2^{-2j}\,\cdot)$ have supports inside the Cartesian coronae

$$[-2^{-2j-1}, 2^{-2j-1}]^3 \setminus [-2^{-2j-4}, 2^{-2j-4}]^3 \subset \widehat{\mathbb{R}}^3, \tag{2.3}$$

and the collection of window functions $\widetilde{W}^2, W_j^2$, $j \geq 0$, produce a smooth tiling of the frequency space into concentric Cartesian coronae associated with various frequency bands indexed by $j$.

Recall that the *Parseval frame condition* implies that, for any $f \in L^2(\mathbb{R}^3)$, we have the reproducing formula:

$$f = \sum_{\lambda\in\Lambda} \langle f, \phi_\lambda\rangle\, \phi_\lambda,$$

with convergence in $L^2$-norm.

## 2.2 A Parseval frame of 3D shearlets

The shearlet representation is one of the multiscale methods introduced during the last decade to overcome the limitations of conventional wavelets in the analysis of multivariate functions. Similar to the curvelets of Candès and Donoho [1], the shearlets form a collection of well localized functions defined not only across several scales and locations, as the conventional wavelets, but also across several orientations and with highly anisotropic shapes, so that they can more efficiently represent functions containing distributed singularities, e.g., edges in images. Thanks to their ability to combine multiscale anlysis and high directional sensitivity, shearlets are useful to provide a precise characterization of the geometry of singularities of functions and distributions of several variables [11, 16, 12] and enable optimally sparse representations, in a precise sense, for a large class of multivariate functions where traditional wavelets are suboptimal [10, 13].

With respect to curvelets, shearlets offer a combination of useful features: their mathematical structure is derived from the theory of affine systems and the directionality is controlled by shear matrices rather than rotations. This last property enables a unified framework for both continuum and discrete settings since shear transformations preserve the rectangular lattice and this is an advantage in deriving faithful digital implementations [7, 22]. Furthermore, there is a well-developed shearlet-based theory for the analysis of singularities (cf. [9] in addition to the references cited above). This theory sets the foundation for the main ideas that we will employ for the analysis of the surface singularities and it is the main reason for selecting this representation in our approach to the geometric separation problem.

Roughly speaking, our shearlet system is obtained by introducing an angular partition in the multiscale decomposition defined by the window functions $\widetilde{W}^2, W_j^2$ used above for the construction of the wavelet system. Since we will not use rotations but shear transformations (defined below), we need first to split the Fourier space $\widehat{\mathbb{R}}^3$ into the following 3 pyramidal regions in $\widehat{\mathbb{R}}^3$:

$$\mathcal{P}_1 = \left\{(\xi_1,\xi_2,\xi_3) \in \widehat{\mathbb{R}}^3 : |\frac{\xi_2}{\xi_1}| \le 1, |\frac{\xi_3}{\xi_1}| \le 1\right\},$$

$$\mathcal{P}_2 = \left\{(\xi_1,\xi_2,\xi_3) \in \widehat{\mathbb{R}}^3 : |\frac{\xi_1}{\xi_2}| < 1, |\frac{\xi_3}{\xi_2}| \le 1\right\},$$

$$\mathcal{P}_3 = \left\{(\xi_1,\xi_2,\xi_3) \in \widehat{\mathbb{R}}^3 : |\frac{\xi_1}{\xi_3}| < 1, |\frac{\xi_2}{\xi_3}| < 1\right\}.$$

Now, we let $W \in C_0^\infty(\mathbb{R}^3)$ be the *same* window defined above and let $v \in C^\infty(\mathbb{R})$ be an appropriate 'bump function' satisfying $\operatorname{supp} v \subset [-1,1]$ and

$$|v(u-1)|^2 + |v(u)|^2 + |v(u+1)|^2 = 1 \quad \text{for } |u| \le 1. \tag{2.4}$$

For $d = 1,2,3$, $\ell = (\ell_1,\ell_2) \in \mathbb{Z}^2$, a 3D *shearlet systems associated with the pyramidal regions* $\mathcal{P}_d$ is a collection

$$\{\psi^{(d)}_{j,\ell,k} : j \ge 0, -2^j \le \ell_1,\ell_2 \le 2^j, k \in \mathbb{Z}^3\}, \tag{2.5}$$

where

$$\hat{\psi}^{(d)}_{j,\ell,k}(\xi) = |\det A_{(d)}|^{-j/2}\, W(2^{-2j}\xi)\, V_{(d)}(\xi A_{(d)}^{-j} B_{(d)}^{[-\ell]})\, e^{2\pi i \xi A_{(d)}^{-j} B_{(d)}^{[-\ell]} k}, \tag{2.6}$$

$V_{(1)}(\xi_1,\xi_2,\xi_3) = v(\frac{\xi_2}{\xi_1})v(\frac{\xi_3}{\xi_1})$, $V_{(2)}(\xi_1,\xi_2,\xi_3) = v(\frac{\xi_1}{\xi_2})v(\frac{\xi_3}{\xi_2})$, and $V_{(3)}(\xi_1,\xi_2,\xi_3) = v(\frac{\xi_1}{\xi_3})v(\frac{\xi_2}{\xi_3})$; the matrices $A_{(d)}$ are given by

$$A_{(1)} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix},\ A_{(2)} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix},\ A_{(3)} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix},$$

and the matrices $B_{(d)}$, called *shear matrices*, are defined by

$$B^{[\ell]}_{(1)} = \begin{pmatrix} 1 & \ell_1 & \ell_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ B^{[\ell]}_{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ \ell_1 & 1 & \ell_2 \\ 0 & 0 & 1 \end{pmatrix},\ B^{[\ell]}_{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \ell_1 & \ell_2 & 1 \end{pmatrix}.$$

Notice that $(B^{[\ell]}_{(d)})^{-1} = B^{[-\ell]}_{(d)}$. Let us make a few observations about the properties of these systems.

Due to the support conditions on $W$ and $v$, the elements of the system of shearlets (2.5) have compact support in Fourier domain. For example, for $d = 1$, the shearlets $\hat{\psi}^{(1)}_{j,\ell,k}(\xi)$ can be written explicitly as

$$\hat{\psi}^{(1)}_{j,\ell_1,\ell_2,k}(\xi) = 2^{-2j}\, W(2^{-2j}\xi)\, v\Big(2^j \frac{\xi_2}{\xi_1} - \ell_1\Big)\, v\Big(2^j \frac{\xi_3}{\xi_1} - \ell_2\Big)\, e^{2\pi i \xi A_{(1)}^{-j} B_{(1)}^{[-\ell_1,-\ell_2]} k}, \tag{2.7}$$

showing that their supports are contained inside the regions

$$\begin{aligned} &U_{j,\ell} = U_{j,\ell_1,\ell_2} \\ &= \{(\xi_1,\xi_2,\xi_3) : \xi_1 \in [-2^{2j-1}, -2^{2j-4}] \cup [2^{2j-4}, 2^{2j-1}], |\tfrac{\xi_2}{\xi_1} - \ell_1 2^{-j}| \le 2^{-j}, |\tfrac{\xi_3}{\xi_1} - \ell_2 2^{-j}| \le 2^{-j}\}. \end{aligned} \tag{2.8}$$

That is, the shearlets $\hat{\psi}^{(1)}_{j,\ell,k}$ have supports contained in *trapezoidal regions* defined at *various scales*, controlled by $j > 0$, and *various orientations*, controlled by the shear parameters $\ell_1, \ell_2$. Hence, the elements of the

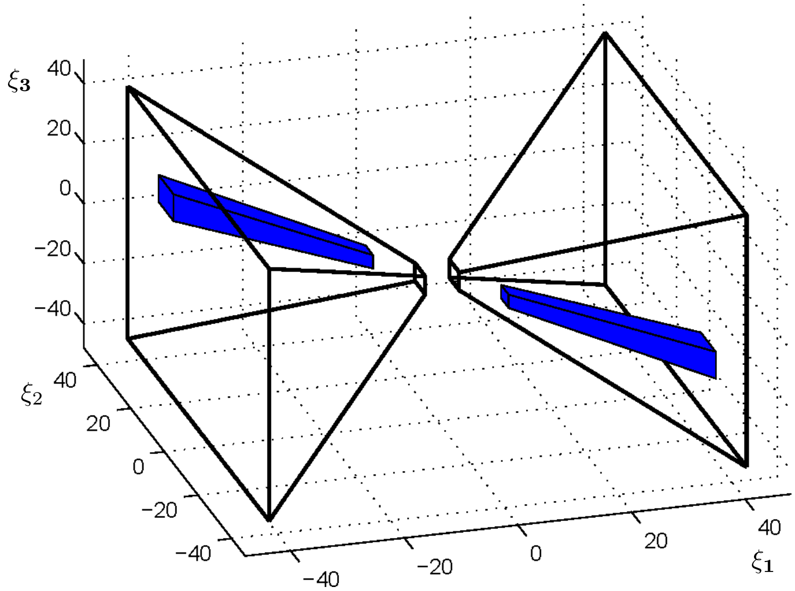


Figure 1: *Frequency support of a representative shearlet $\psi^{(1)}_{j,\ell,k}$, inside the pyramidal region $\mathcal{P}_1$. The orientation of the support region is controlled by the shear index $\ell = (\ell_1, \ell_2)$.*

shearlet system (2.5) are well-localized functions, defined over a range of locations, scales and orientations, controlled by the indices $j, \ell = (\ell_1, \ell_2)$ and $k$, respectively. The support region of a representative shearlet $\hat{\psi}^{(1)}_{j,\ell,k}$ is illustrated in Fig. 1.

A *Parseval frame of shearlets* for $L^2(\mathbb{R}^3)$ is obtained by combining the shearlet systems (2.5) associated with the cone-shaped regions $\mathcal{P}_d$ together with the *coarse scale* system $\{\phi_{-1,k} : k \in \mathbb{Z}^3\}$. Note that this is the same coarse scale system of the Lemeriè-Meyer wavelet system defined above. For brevity, in the following we will denote the Parseval frame of 3D shearlets as $\Psi = \{\psi_\eta : \eta \in M\} \subset L^2(\mathbb{R}^3)$, where the index set is $M = M_C \cup M_F$, $M_C = \{k \in \mathbb{Z}^3\}$ is the set of indices associated with coarse-scale shearlets and $M_F = \{\eta = (j, \ell, k, d) : j \geq 0, |\ell_1| \leq 2^j, |\ell_2| \leq 2^j, k \in \mathbb{Z}^2, d = 1, 2, 3\}$ is the set of indices associated with fine-scale shearlets. As above, the Parseval frame condition implies that, for any $f \in L^2(\mathbb{R}^3)$, we have the reproducing formula:

$$f = \sum_{\eta \in M} \langle f, \psi_\eta \rangle \, \psi_\eta,$$

with convergence in $L^2$-norm.

**Remark.** To simplify the presentation, our construction above omits some technical details. To ensure that the frame is tight when the shearlet functions from the 3 pyramidal regions are combined and still guarantee that all elements of the combined shearlet system are $C_0^\infty$ in the Fourier domain, it is convenient to slightly modify the functions $\psi^{(d)}_{j,\ell_1,\ell_2,k}$, for $\ell_1, \ell_2 = \pm 2^j$ (these are the functions whose support overlap the boundaries of the regions $\mathcal{P}_d$). This modification consists, essentially, in merging shearlet elements from contiguous pyramidal regions. The construction of these *boundary shearlets* is rather technical and plays no role in the paper. We refer the interested reader to [13, 15].

## 2.3 Main theorem

Our main theorem shows that it is possible to separate the two geometrically distinct components of $f$ taking advantage of the sparsity properties of the Parseval frames of wavelets and shearlets. Similar to [6], this separation result is true only *asymptotically* in scale, that is, we show that we separate point singularities and singularities along polyhedral surfaces only as a limiting process, when the scale tends to zero (i.e., $j \to \infty$).

Therefore, it will useful to derive an appropriate multiscale decomposition of $f = \mathcal{P} + \mathcal{T}$.

For this, we recall that the window functions $W_j$ used in the construction of the wavelet and shearlet systems produce a multiscale decomposition of the Fourier space $L^2(\mathbb{R}^3)$ into the Cartesian coronae (2.3). Hence, we can define a family of band-pass filters $F_j$, $j \geq -1$, by setting $\widehat{F_j}(\xi) = W(2^{-2j}\xi)$, for $j \geq 0$, $\widehat{F_{-1}}(\xi) = \widetilde{W}(\xi)$. By applying these filters to $f$, $\mathcal{P}$ and $\mathcal{T}$ we define

$$\mathcal{P}_j = \mathcal{P} * F_j, \quad \mathcal{T}_j = \mathcal{T} * F_j, \quad f_j = f * F_j, \tag{2.9}$$

where, as observed above, we have that $\|\mathcal{P}_j\|_2 \simeq 2^{-j}$ and $\|\mathcal{T}_j\|_2 \simeq 2^{-j}$. It follows that the functions $\hat{f}_j$ are band-limited, with frequency support contained in the Cartesian coronae $[-2^{2j-1}, 2^{2j-1}]^3 \setminus [-2^{2j-4}, 2^{2j-4}]^3 \subset \widehat{\mathbb{R}}^3$. Furthermore, for $f \in L^2(\mathbb{R}^3)$, it follows from the tiling properties of the function $W(2^{-2j}\cdot)$ that

$$f = \sum_j F_j * f_j, \tag{2.10}$$

with convergence in the $L^2$-norm.

Let $\mathcal{F}_j$ denote the range of the operator of convolution with $F_j$. It is a simple calculation to verify that the shearlets and wavelets at level $j'$ are orthogonal to $\mathcal{F}_j$ unless $|j'-j| \leq 1$, that is, unless $j' = j-1, j, j+1$. It is useful to introduce the notation

$$\Lambda_j = \{\lambda = (j', k) : \, |j'-j| \leq 1, \, k \in \mathbb{Z}^3\} \subset \Lambda \tag{2.11}$$

and

$$M_j = \{\eta = (j', \ell, k, d) : \, |j'-j| \leq 1, |\ell_1| \leq 2^j, \, |\ell_2| \leq 2^j, \, k \in \mathbb{Z}^3, d = 1,2,3\} \subset M. \tag{2.12}$$

Due to the Parseval frame property and the observation above, we have that a function $f_j \in \mathcal{F}_j$ can be expanded using *only* the elements of the wavelet system in $\Lambda_j$ but also using *only* the elements of the shearlet system in $M_j$. In other words, at the level $j$, we can use the wavelet system to represent $f_j$ as

$$f_j = \sum_{j'=j-1}^{j'=j+1} \sum_{k' \in \mathbb{Z}^2} \langle f_j, \phi_{j',k'} \rangle \, \phi_{j',k'} = \sum_{\lambda \in \Lambda_j} \langle f_j, \phi_\lambda \rangle \, \phi_\lambda;$$

or we can use the shearlet system to represent $f_j$ as

$$f_j = \sum_{d=1}^{3} \sum_{j'=j-1}^{j'=j+1} \sum_{|\ell_1| \leq 2^{j'}} \sum_{|\ell_2| \leq 2^{j'}} \sum_{k \in \mathbb{Z}^2} \langle f_j, \psi^{(d)}_{j',\ell_1,\ell_2,k} \rangle \, \psi^{(d)}_{j',\ell_1,\ell_2,k} = \sum_{\eta \in M_j} \langle f_j, \psi_\eta \rangle \, \psi_\eta.$$

Clearly, we can also consider a combined representation of the form

$$f_j = \sum_{\lambda \in \Lambda_j} u_\lambda \, \phi_\lambda + \sum_{\eta \in M_j} t_\eta \, \psi_\eta,$$

for an appropriate choice of coefficients $u = (u_\lambda)$ and $t = (t_\eta)$. Since, in this last expression, we are considering an overcomplete dictionary, there are many possible choices of coefficients $u$ and $t$, some of which may provide sparser representations than either one of the two expansions above. Similar to [6], we seek a solution providing a *geometric separation*, that is, we consider the following dual-frame *component separation problem* based on $\ell_1$ minimization:

$$(U_j^*, T_j^*) = \arg\min(\|u\|_1 + \|t\|_1), \quad \text{subject to } f_j = U_j + T_j, \tag{2.13}$$

where $u_\lambda = \langle U_j, \phi_\lambda \rangle$, $\lambda \in \Lambda_j$ and $t_\eta = \langle T_j, \psi_\eta \rangle$, $\eta \in M_j$. It follows from (2.10) that, if we let $\tilde{P} = \sum_j F_j * U_j$, $\tilde{T} = \sum_j F_j * T_j$, then we can express $f$ as the superposition $f = \tilde{P} + \tilde{T}$.

The main result of our paper is the following theorem, stating that we achieve the *separation* of the distinct geometric objects $\mathcal{P}$ and $\mathcal{T}$, *asymptotically at fine scales*, by applying $\ell_1$ minimization over the expansion coefficients of $f$ with respect to our combined wavelet-shearlet dictionary.

**Theorem 2.1.** *Let $\Phi$ and $\Psi$ be the Parseval frames of wavelets and shearlets, respectively, defined above and denote $\|g\|_{1,\Phi} = \sum_{\lambda \in \Lambda_j} |\langle g, \phi_\lambda \rangle|$ and $\|g\|_{1,\Psi} = \sum_{\eta \in M_j} |\langle g, \psi_\eta \rangle|$, where $\Lambda_j$ is given by* (2.11) *and $M_j$ is given by* (2.12)*. Let $f_j = U_j + T_j$ be given as above and $\mathcal{P}_j$, $\mathcal{T}_j$ be given by* (2.9)*. We have that*

$$\lim_{j \to \infty} \frac{\|U_j - \mathcal{P}_j\|_{1,\Phi} + \|T_j - \mathcal{T}_j\|_{1,\Psi}}{\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi}} = 0.$$

That is, asymptotically as the scale tends to zero, the pointlike component of $f$ is captured by the Parseval frame of wavelets and the piecewise linear component of $f$ is captured by the Parseval frame of shearlets.

**Remark 2.1.** *The statement of Theorem 2.1 is different from the corresponding Theorem 1.1 in [6], valid in the two-dimensional case, since it is formulated using the $\ell^1$–norm rather than the $\ell^2$–norm used in [6]. This difference with respect to the original result is crucial to be able to handle singularities containing edges and vertices.*

*We recall that the geometric separation result originally obtained in [6] deals with 2D images containing point-like and smooth curve-like singularities, but does not handle linear or curve-like singularities containing corner points. The approach presented in our paper can be easily adapted to the two-dimensional case so that our Theorem 2.1 is indeed valid also in the case of 2D images containing point-like and piecewise linear singularities including corner points. We remark that the handling of corner points is not trivial and cannot be derived from the arguments in [6].*

*On the other hand, our result does not cover general surface and curve singularities (e.g., an arc of a circle or a a section of a sphere). After presenting our proof, we will make additional comments to show where our arguments break down in this situation.*

The rest of the paper is devoted mostly to the proof of Theorem 2.1.

# 3 Proof of main theorem

As we will show below, our proof follows the general architecture of the proof from [6], which is centered around the notion of cluster coherence. However, the most critical parts of the proof, that is, the actual proof that the cluster coherence satisfies the desired estimates for the three-dimensional objects considered in this paper, do not follow from [6] and are completely new. These arguments are contained in the proofs of Lemma 3.5 and Lemma 3.6 and are presented in the next sections. The proof of Lemma 3.5, in particular, uses a self-improving process that is a valuable idea in itself. Lemma 3.6 gives the decay estimate for edges and corner points and is significantly more involved than the corresponding result used in [6] to analyze the smooth segment case (also compare our Lemma 3.7 in Section 3.2). We will make further comments about the significance of these lemmata below, after introducing some definitions.

Let $\Phi = \{\phi_\lambda : \lambda \in \Lambda\}$ and $\Psi = \{\psi_\mu : \mu \in M\}$ be the Parseval frames of 3D wavelets and 3D shearlets introduced above, respectively. For each level $j \in \mathbb{Z}$, we will identify certain subsets of the indices $\Lambda$ and $M$ that we denote as $S_{1,j} \subset \Lambda_j$ and $S_{2,j} \subset M_j$. Following the terminology in [6], we refer to them as *indices of significant wavelet coefficients* and *indices of significant shearlet coefficients*, respectively. These index sets will identify, essentially, those wavelet and shearlet coefficients whose magnitude is above a certain scale-dependent threshold (hence, the name 'significant'). Their explicit definition, when the expansion coefficients are computed on $f = \mathcal{P} + \mathcal{T}$, will be determined in Sec. 3.1 (for $S_{1,j}$) and Sec. 3.2, after Lemma 3.7 (for $S_{2,j}$).

Corresponding to the sets $S_{1,j}$ and $S_{2,j}$, we define the *wavelet approximation error* and the *shearlet approximation error* at the level $j$ as

$$\delta_{1,j} = \sum_{\lambda \in S_{1,j}^c} |\langle P_j, \phi_\lambda \rangle|, \quad \delta_{2,j} = \sum_{\eta \in S_{2,j}^c} |\langle T_j, \psi_\eta \rangle|$$

respectively. As we will see below, it will be possible to determine the indices of significant wavelet and shearlet coefficients $S_{1,j}$ and $S_{2,j}$ in such a way that the wavelet and shearlet approximation errors are small, meaning that the $\ell^1$-norm of the wavelet and shearlet coefficients is negligible (asymptotically, at fine scales), when the indices are *outside* the sets $S_{1,j}$ and $S_{2,j}$.

We define the *cluster coherences* as

$$\mu_c(S_{1,j}, \Phi; \Psi) = \max_\eta \sum_{\lambda \in S_{1,j}} |\langle \phi_\lambda, \psi_\eta \rangle|, \quad \mu_c(S_{2,j}, \Psi; \Phi) = \max_\lambda \sum_{\eta \in S_{2,j}} |\langle \phi_\lambda, \psi_\eta \rangle|.$$

The notion of cluster coherence was originally proposed in [6]. Unlike the more standard definition of *coherence*, given by $\mu(\Phi, \Psi) = \max_{\lambda,\eta} |\langle \phi_\lambda, \psi_\eta \rangle|$, the cluster coherence bounds coherence between a single member of a frame and a *cluster* of members of another frame.

Let $\Phi$ be the matrix representation of the Parseval frame of wavelets and $\Psi$ the matrix representation of our Parseval frame of shearlets. For a $g_j \in L^2(\mathbb{R}^3) \cap L^1(\mathbb{R}^3)$ such that $\operatorname{supp}(\hat{g}_j) \subset \mathcal{F}_j$, let

$$\|1_{S_{1,j}} \Phi^T g_j\|_1 = \sum_{\lambda \in S_{1,j}} |\langle g_j, \phi_\lambda \rangle|, \quad \|1_{S_{2,j}} \Psi^T g_j\|_1 = \sum_{\eta \in S_{2,j}} |\langle g_j, \psi_\eta \rangle|.$$

We define the *joint concentration* by

$$\kappa = \kappa(S_{1,j}, S_{2,j}) = \sup_{g_j} \frac{\|1_{S_{1,j}} \Phi^T g_j\|_1 + \|1_{S_{2,j}} \Psi^T g_j\|_1}{\|\Phi^T g_j\|_{1,\Phi} + \|\Psi^T g_j\|_{1,\Psi}}.$$

The following observation from [6] illustrates the relationship between joint concentration and data separation.

**Proposition 3.1. ([6, Prop. 2.1])** *Suppose that, for $j \in \mathbb{Z}$, $f_j = U_j + T_j$ so that each component of $f_j$ is relatively sparse in $\Phi$ or $\Psi$, that is,*

$$\|1_{S_{1,j}^C} \Phi^T U_j\|_1 \le \delta_{1,j}, \quad \|1_{S_{2,j}^C} \Psi^T T_j\|_1 \le \delta_{2,j}.$$

*If $(U_j^*, T_j^*)$ solves (2.13), then*

$$\|U_j^* - \mathcal{P}_j\|_{1,\Phi} + \|T_j^* - \mathcal{T}_j\|_{1,\Psi} \le \frac{2(\delta_{1,j} + \delta_{2,j})}{1 - 2\kappa}.$$

Another observation from [6] is that the joint concentration is bounded above by the maximum of the cluster coherences:

**Lemma 3.2. ([6, Lemma. 2.1])**

$$\kappa(S_{1,j}, S_{2,j}) \le \max\{\mu_c(S_{1,j}, \Phi; \Psi),\ \mu_c(S_{2,j}, \Psi; \Phi)\}$$

It follows from Proposition 3.1 and Lemma 3.2 that Theorem 2.1 is proved if we can construct appropriate sets of significant wavelet and shearlet coefficients $S_{1,j}$ and $S_{2,j}$ such that $\delta_{1,j} = o(\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi})$, $\delta_{2,j} = o(\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi})$ and

$$\mu_c(S_{1,j}, \Phi; \Psi) \to 0, \quad \mu_c(S_{2,j}, \Psi; \Phi) \to 0, \quad \text{as } j \to \infty.$$

The rest of the proof is organized as follows. In Section 3.1, we will select an appropriate set $S_{1,j}$ and show that $\mu_c(S_{1,j}, \Phi; \Psi) \to 0$ and $\delta_{1,j} = o(\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi})$. This is the easy part of the argument and it follows from an idea similar to [6]. For the 'hard' part of the proof, concerning the analysis of the singularities along the polyhedral surfaces, it is not possible to apply the original argument from [6] and we will derive a novel approach. Namely, in Section 3.2, we will construct an appropriate index set of significant shearlet coefficients $S_{2,j}$ and show that $\mu_c(S_{2,j}, \Psi; \Phi) \to 0$ and $\delta_{2,j} = o(\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi})$. Our most delicate estimates are contained in the proof of Lemma 3.6 in Section 3.2, where it is highly nontrivial to control the size of the translation variable $k$ corresponding to 3D edges and vertices. Our novel argument is based on techniques for the analysis of singularities using the shearlet transform that we originally developed for the characterization of piecewise smooth boundaries of multivariate functions in [12, 14].

In the following, for all our arguments, it will be sufficient to consider the shearlet system associated with the cone-shaped regions $\mathcal{P}_1 \subset \widehat{\mathbb{R}}^3$ only, since the properties of the similar system in $\mathcal{P}_2$ and $\mathcal{P}_3$ are the same. The elements (2.7) of such shearlet system can be written[2] as

$$\hat{\psi}^{(1)}_{j,\ell_1,\ell_2,k}(\xi) = 2^{-2j}\Gamma_{j,\ell_1,\ell_2}(\xi)\, e^{2\pi i \xi A_{(1)}^{-j} B_{(1)}^{[-\ell_1,-\ell_2]} k},$$

where

$$\Gamma_{j,\ell_1,\ell_2}(\xi) = W(2^{-2j}\xi)\, v\Big(2^j \frac{\xi_2}{\xi_1} - \ell_1\Big)\, v\Big(2^j \frac{\xi_3}{\xi_1} - \ell_2\Big).$$

Note that $A_{(1)}^{-j} B_{(1)}^{[-\ell_1,-\ell_2]} k = (2^{-2j}(k_1 - \ell_1 k_2 - \ell_2 k_3), 2^{-j}k_2 + 2^{-j}k_3)$. Each function $\Gamma_{j,\ell_1,\ell_2}$ is supported inside the set $U_{j,\ell_1,\ell_2}$, given by (2.8). It is easy to verify that its measure satisfies $|U_{j,\ell_1,\ell_2}| \le C\, 2^{4j}$.

### 3.1 Estimate for the point singularities

In this section, we will select the set $S_{1,j}$ and prove that $\mu_c(S_{1,j}, \Phi; \Psi) \to 0$ and $\delta_{1,j} = o(\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi})$, asymptotically as $j \to \infty$. As indicated above, the rather simple argument that we use is similar to [6].

Let $\phi_{j',k'}$ and $\psi_{j,\ell_1,\ell_2,k}$ be generic elements from the Parseval frames of wavelets and shearlets, respectively. Due to the frequency support of $W$, for any $\ell_1, \ell_2, k$ and $k'$ we have that $\langle \widehat{\psi_{j,\ell,k}}, \widehat{\phi_{j',k'}} \rangle = 0$ if $|j - j'| > 1$. Thus for all large $j'$ and $j = j'-1, j', j'+1$, a direct computation shows that

$$\begin{aligned}
|\langle \widehat{\psi_{j,\ell_1,\ell_2,k}}, \widehat{\phi_{j',k'}} \rangle| &= \left| \int_{\mathbb{R}^2} \Big(2^{-2j}\Gamma_{j,\ell_1,\ell_2}(\xi)\, e^{-2\pi i \xi A_{(1)}^{-j} B_{(1)}^{[-\ell_1,-\ell_2]} k}\Big) \Big(2^{-3j'} W(2^{-2j'}\xi) e^{2\pi i 2^{-3j'}\xi\cdot k'}\Big)\, d\xi \right| \\
&\le 2^{-2j} 2^{-2j'} \int_{\mathbb{R}^3} |\Gamma_{j,\ell_1,\ell_2}(\xi)\, W(2^{-2j'}\xi)|\, d\xi \\
&\le C\, 2^{-2j} 2^{-3j'} \int_{\Omega_{j,\ell_1,\ell_2}} d\xi \le C\, 2^{-2j} 2^{-3j'} 2^{4j} \le C\, 2^{-j},
\end{aligned}$$

where $C$ is independent of $\ell_1, \ell_2, k, k'$ and $j$.

For a fixed $0 < \epsilon < 1$, we set $S_{1,j} = \{(j', k') : j' = j-1, j, j+1;\ |k'| \le 2^{\epsilon j'}\}$. Then, using the calculation above, we have that

$$\mu_c(S_{1,j}, \Phi; \Psi) \le C \max_{\ell_1,\ell_2,k} \sum_{j'=j-1}^{j+1} \sum_{|k'| \le 2^{\epsilon j'}} |\langle \widehat{\psi_{j,\ell_1,\ell_2,k}}, \widehat{\phi_{j',k'}} \rangle| \le C 2^{(-1+\epsilon)j}.$$

It follows that $\mu_c(S_{1,j}, \Phi; \Psi) \to 0$, as $j \to \infty$.

[2]Here we ignore the fact that the boundary elements corresponding to $\ell = \pm 2^j$ are slightly modified, since this is irrelevant for all our arguments.

We also observe that $\langle \widehat{\phi_{j',k'}}, \widehat{\mathcal{P}_j}\rangle = 0$ for all $k'$ if $|j'-j|>1$. For $|j'-j|\le 1$, we have that

$$\begin{aligned}
\langle \widehat{\phi_{j',k'}}, \widehat{\mathcal{P}_j}\rangle &= 2^{-3j'}\, C \int_{\mathbb{R}^3} W(2^{-2j'}\xi)\, e^{2\pi i 2^{-2j'}\xi\cdot k'} W(2^{-2j}\xi)|\xi|^{-2}\, d\xi \\
&= C2^{-j'} \int_{\mathbb{R}^3} W(\xi) W(2^{2(j'-j)}\xi)\, e^{2\pi i \xi\cdot k'} |\xi|^{-2}\, d\xi.
\end{aligned}$$

Hence, for $|k'|\ge 2^{\epsilon j}$, integration by parts gives that

$$|\langle \widehat{\phi_{j',k'}}, \widehat{\mathcal{P}_j}\rangle| \le C_N\, 2^{-j}(1+|k'|)^{-N} \le C_N\, 2^{-(1+N\epsilon)j}.$$

It follows that, by choosing $N$ sufficiently large, we have:

$$\delta_{1,j} = \sum_{\lambda\in S_{1,j}^c} |\langle \phi_\lambda, P_j\rangle| \le C\, 2^{-2j} = o(2^{-j}) = o\left(\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi}\right).$$

## 3.2 Estimate for the piecewise linear singularities

We first recall the Divergence Theorem in $\mathbb{R}^3$. Let $\vec{F}$ be a smooth vector field in $\mathbb{R}^2$ and $S$ be a compact region in the plane with a piecewise smooth simple boundary $\partial S$. Then

$$\int_S \bigtriangledown\cdot \vec{F}\, dA = \int_{\partial S} \vec{F}\cdot\vec{n}\, ds,$$

where $\vec{n}(x)$ is the outer normal direction at $x\in\partial S$.

Let $\mathcal{T}$ be the characteristic function of a polyhedron of $M$ faces in $\mathbb{R}^3$. Without loss of generality, we may assume that the polyhedron is contained inside the cube $[-1,1]^3$ (if not, the polyhedron can be rescaled by dilation on the space variables). Let $S=\cup_{m=1}^M S_m$ be the boundary of the polyhedron, where for each $1\le m\le M$, $S_m$ is a polygon in $\mathbb{R}^2$. For $\xi\in\widehat{\mathbb{R}}^3$, Using the divergence theorem we can express the Fourier transform of $\mathcal{T}$ as follows

$$\begin{aligned}
\widehat{\mathcal{T}}(\xi) &= -\frac{1}{2\pi i|\xi|}\int_S e^{-2\pi i\xi\cdot x}\,\frac{\xi}{|\xi|}\cdot\vec{n}(x)\, d\sigma(x) \\
&= \sum_1^M\left(-\frac{1}{2\pi i|\xi|}\right)\int_{S_m} e^{-2\pi i\xi\cdot x}\,\frac{\xi}{|\xi|}\cdot\vec{n}(x)\, d\sigma(x) \\
&= \sum_1^M \widehat{\mathcal{T}}^{(m)}(\xi),
\end{aligned}$$

where $\widehat{\mathcal{T}}^{(m)}(\xi) = -\frac{1}{2\pi i|\xi|}\int_{S_m} e^{-2\pi i\xi\cdot x}\,\frac{\xi}{|\xi|}\cdot\vec{n}(x)\, d\sigma(x)$ for each $1\le m\le M$.

In order to estimate $\widehat{\mathcal{T}}^{(m)}(\xi)$, we divide $S_m$ into finitely many sub-surfaces so that each sub-surface is the graph of a linear function on a triangle domain $D_m\in\mathbb{R}^2$. Without loss of generality, we may assume that each surface $S_m$ can be written as $S_m=\{(Ax_2+Bx_3,x_2,x_3):(x_2,x_3)\in D_m\}$, where $D_m=\{(x_2,x_3): 0\le x_3\le bx_2,\ a_1\le x_2\le a_2\}$ and $A,B,b,a_1,a_2$ are appropriate constants. Let $\eta=(\eta_1,\eta_2)=(A\xi_1+\xi_2, B\xi_1+\xi_3)$ so that, for $x\in S_m$, we have $\xi\cdot x=\xi_1(Ax_2+Bx_2)+\xi_2x_2+\xi_3x_3=(A\xi_1+\xi_2)x_2+(B\xi_1+\xi_3)x_3=\eta\cdot(x_2,x_3)$.

A direction calculation shows that

$$\begin{aligned}
\int_{S_m} e^{-2\pi i\xi\cdot x}\,\frac{\xi}{|\xi|}\cdot\vec{n}(x)\, d\sigma(x) &= \int_{D_m} e^{-2\pi i\,\eta\cdot(x_2,x_3)}\,\frac{\xi}{|\xi|}\cdot(-1,A,B)\, dx_3dx_2 \\
&\simeq \frac{1}{1+|B\xi_1+\xi_3|}\,\frac{1}{1+|A\xi_1+\xi_2+b(B\xi_1+\xi_3)|}.
\end{aligned}$$

It follows that

$$\widehat{\mathcal{T}}^{(m)}(\xi) \simeq \frac{1}{|\xi|}\, \frac{1}{1+|B\xi_1+\xi_3|}\, \frac{1}{1+|A\xi_1+\xi_2+b(B\xi_1+\xi_3)|}. \tag{3.14}$$

Let $K$ be the characteristic function of the unit ball in $\mathbb{R}^3$. It is known that $\widehat{K}(\xi) = |\xi|^{-\frac{3}{2}} J_{\frac{3}{2}}(|\xi|)$, where $J_{\frac{3}{2}}$ is the Bessel function of order 3/2. Using the observation that the asymptotic decay of $J_{\frac{3}{2}}(|\xi|)$, for large $|\xi|$, is of the order $|\xi|^{-\frac{1}{2}}$ (cf. [28, Ch.8]), it follows that $\|K_j\|_2 = \|K * F_j\|_2 \simeq 2^{-j}$. The following lemma shows that the same estimate holds for $\mathcal{T}_j$.

**Lemma 3.3.** *For $j \in \mathbb{N}$ and $\mathcal{T}_j$ defined above, the following estimate holds:*

$$\|\mathcal{T}_j\|_2 \simeq 2^{-j}.$$

*Proof.* Using the notation introduced above, let $\widehat{\mathcal{T}}_j^{(m)}(\xi) = \widehat{\mathcal{T}}^{(m)}(\xi)\, W(2^{-2j}\xi)$. Since $\widehat{\mathcal{T}}_j = \sum_1^M \widehat{\mathcal{T}}_j^{(m)}$ and $2^{2j} \le |\xi| \le 2^{2j+2}$ inside the support of $W(2^{-2j}\xi)$, an easy calculation using spherical coordinates gives that

$$\begin{aligned} &\|\mathcal{T}_j^{(m)}\|_2^2 \\ \simeq\ & \int_{2^{2j}\le|\xi|\le 2^{2j+2}} \frac{1}{|\xi|^2}\, \frac{1}{1+|B\xi_1+\xi_3|^2}\, \frac{1}{1+|A\xi_1+\xi_2+b(B\xi_1+\xi_3)|^2} d\xi \\ \simeq\ & 2^{-2j}. \end{aligned}$$

Hence, $\|\mathcal{T}_j\|_2 = \|\sum_1^M \mathcal{T}_j^{(m)}\|_2 \simeq 2^{-j}$. $\square$

The following lemma is a special case of Proposition 4.7 in [17].

**Lemma 3.4.** *Let $\beta = (\beta_i)$ be a sequence of non-negative numbers and let $|\beta|_{(N)}$ be the Nth-largest element in the decreasing rearrangement of $\beta$. Then*

$$\left(\sum_{\beta_i \le |\beta|_{(N)}} \beta_i^2\right)^{\frac{1}{2}} \le C\, N^{-\frac{1}{2}}\, \|\beta\|_1,$$

*where $C$ is independent of $N$.*

Using the Lemma 3.4, we derive the following observation.

**Lemma 3.5.** *For a fixed large $j$, let $\beta_j = \{\beta_j(j',\ell_1,\ell_2,k) = \langle \mathcal{T}_j, \psi_{j',\ell_1,\ell_2,k}\rangle : |\ell_1| \le 2^{j'},\ |\ell_2| \le 2^{j'}, j-1 \le j' \le j+1, k \in \mathbb{Z}^3\}$ and, for $v = 1,2$, define the norms*

$$\|\beta_j\|_v = \left(\sum_{j'=j-1}^{j+1} \sum_{|\ell_1|\le 2^{j'}} \sum_{|\ell_2|\le 2^{j'}} \sum_{k\in\mathbb{Z}^3} |\langle \mathcal{T}_j, \psi_{j',\ell_1,\ell_2,k}\rangle|^v\right)^{\frac{1}{v}}.$$

*Then there is a constant $C > 0$ such that $\|\beta_j\|_1 \ge C\, 2^{-\frac{1}{4}j}$.*

*Proof.* Let

$$\widehat{\alpha_{j',j,\ell}}(\xi A_{(1)}^{-j'} B_{(1)}^{[-\ell]}) = 2^{-2j'} W(2^{-2j'}\xi)\, \Gamma_{j',\ell}(\xi)\, e^{2\pi i \xi A_{(1)}^{-j'} B_{(1)}^{[-\ell]} k} W(2^{-2j}\xi)$$

so that

$$\beta_j(j',\ell_1,\ell_2,k) = 2^{2j'} \int_{[-1,1]^3} \mathcal{T}(x)\, \alpha_{j',j,\ell_1,\ell_2}(B_{(1)}^{[\ell]} A_{(1)}^{j'} x - k)\, dx.$$

Using the change of variable $y = B_{(1)}^{[\ell]} A_{(1)}^{j'} x$, it is easy to verify that there is a constact $C > 0$ such that $|\beta_j(j', \ell, k)| \le C\, 2^{-2j'}$ since $\int_{\mathbb{R}^3} |\alpha_{j',j,\ell_1,\ell_2}(y-k)|dy \le C$ uniformly for all $j', j, \ell_1, \ell_2$.

For a fixed $j \ge 0$, as in Lemma 3.4, we denote by $|\beta|_{(n)}$ the nth-largest element for $|\beta_j(j', \ell_1, \ell_2, k)|$. If we let $N = 2^j$, then we have the following estimate for the sum of the squares of the first $2^j$ largest terms:

$$(\sum_{n=1}^{2^j} |\beta|_{(n)}^2)^{\frac{1}{2}} \le C\, 2^{\frac{1}{2}j}\, 2^{-2j'} = C\, 2^{-\frac{1}{2}j}\, 2^{-j} \le C\, 2^{-\frac{1}{2}j} \|\beta_j\|_2,$$

where we used the fact that $\|\beta_j\|_2 \simeq 2^{-j}$ since the shearlet system is a tight frame.

The last expression is controlled by $2^{-\frac{1}{2}j}\|\beta_j\|_1$ since $\|\beta_j\|_2 \le \|\beta_j\|_1$. Thus, combining this fact with the estimate of Lemma 3.4 for $N = 2^j$, we have that

$$\|\beta_j\|_1 \ge 2^{\frac{1}{2}j} \|\beta_j\|_2 \ge 2^{-\frac{1}{2}j}.$$

Now we apply Lemma 3.4 again with $N = 2^{\frac{3}{2}j}$ to get that there is constant $C > 0$ such that

$$\left( \sum_{\beta_i \le |\beta|_{(2^{3j/2})}} \beta_i^2 \right)^{\frac{1}{2}} \le C\, 2^{-\frac{3}{4}j}\, \|\beta_j\|_1.$$

Again we observe that

$$(\sum_{n=1}^{2^{\frac{3}{2}j}} |\beta|_{(n)}^2)^{\frac{1}{2}} \le C\, 2^{\frac{3}{4}j}\, 2^{-2j'} = C\, 2^{-\frac{5}{4}j} = C\, 2^{-\frac{3}{4}j} 2^{-\frac{1}{2}j} \le C\, 2^{-\frac{3}{4}j} \|\beta_j\|_1.$$

Combining the above two estimates, we conclude that there is constant $C > 0$ such that $\|\beta_j\|_1 \ge C 2^{\frac{3}{4}j} \|\beta_j\|_2 \ge C 2^{-\frac{1}{4}j}$. $\square$

Let $\beta_j^{(m)}(j', \ell_1, \ell_2, k) = \langle \mathcal{T}_j^{(m)}, \psi_{j',\ell_1,\ell_2,k} \rangle$, where (using the same notation as in the proof of Lemma 3.3) $\widehat{\mathcal{T}}_j^{(m)}(\xi) = \widehat{\mathcal{T}}^{(m)}(\xi)\, W(2^{-2j}\xi)$ and $\widehat{\mathcal{T}}^{(m)}(\xi)$ is given by (3.14). Since

$$\langle \mathcal{T}_j, \psi_{j',\ell_1,\ell_2,k} \rangle = \langle \widehat{\mathcal{T}}_j, \widehat{\psi_{j',\ell_1,\ell_2,k}} \rangle = \sum_{m=1}^{M} \langle \widehat{\mathcal{T}}_j^{(m)}, \widehat{\psi_{j',\ell_1,\ell_2,k}} \rangle,$$

we can write

$$\beta_j(j', \ell_1, \ell_2, k) = \sum_{m=1}^{M} \langle \mathcal{T}_j^{(m)}, \psi_{j',\ell_1,\ell_2,k} \rangle = \sum_{m=1}^{M} \beta_j^{(m)}(j', \ell_1, \ell_2, k).$$

A direct calculation gives that

$$\begin{aligned} \beta_j^{(m)}(j', \ell_1, \ell_2, k) &= \langle \widehat{\mathcal{T}}_j^{(m)}, \widehat{\psi_{j',\ell_l,\ell_2,k}} \rangle \\ &= 2^{-2j'} \int_{U_{j',\ell_1,\ell_2}} \widehat{\mathcal{T}}_j^{(m)}(\xi)\, \Gamma_{j',\ell_1,\ell_2}(\xi)\, e^{2\pi i \xi A_{(1)}^{-j'} B_{(1)}^{[-\ell]} k}\, d\xi. \end{aligned}$$

In order to deal with the edges and vertices in the singularity set and control the size of $k$, we will now decompose the functions $\widehat{\mathcal{T}}_j^{(m)}$ into their 'directional' components.

Let $g(t) = |V(t)|^2$, where $V$ is the function introduced in Section 2 for the construction of the shearlets. Clearly, $g \in C_0^\infty(-1, 1)$ and, by (2.4), it follows that for all $j \ge 0$ it satisfies the equation

$$\sum_{p=-2^j}^{2^j} g(2^j t - p) = 1, \quad \text{for} \quad |t| \le 1.$$

Hence, we obtain the following expansion of $\widehat{\mathcal{T}}_j^{(m)}$ into its the 'directional' components:

$$\widehat{\mathcal{T}_j^{(m)}}(\xi) = \sum_{p=-2^j}^{2^j} \sum_{q=-2^j}^{2^j} \widehat{\mathcal{T}_j^{(m,p,q)}}(\xi),$$

where

$$\widehat{\mathcal{T}}_j^{(m,p,q)}(\xi) = -\frac{W(2^{-2j}\xi)}{(2\pi)^2|\xi|}\frac{\xi}{|\xi|} \cdot (-1, A, B) \int_D g(2^j x_2 - p) g(2^j x_3 - q) e^{-2\pi i \eta \cdot (x_2, x_3)} dx_3 dx_2.$$

Let $\beta_j^{(m,p,q)}(j', \ell_1, \ell_2, k) = \langle \widehat{\mathcal{T}}_j^{(m,p,q)}, \widehat{\psi_{j',\ell_1,\ell_2,k}} \rangle$. For simplicity of notations, in the following we will assume $j' = j$. If $j' \neq j$, the only difference is that one needs to add up the terms corresponding to $j' = j-1, j, j+1$ and, as we have seen above, the effect of adding up $j' = j-1, j, j+1$ is irrelevant for the argument since it only yields a different uniform constants in the final estimates. Thus, in the following, we will ignore the sum over $j'$ and only consider the terms

$$\beta_j^{(m,p,q)}(\ell_1, \ell_2, k) := \langle \widehat{\mathcal{T}}_j^{(m,p,q)}, \widehat{\psi_{j,\ell_1,\ell_2,k}} \rangle.$$

Let $\eta = (A\xi_1 + \xi_2, B\xi_1 + \xi_3) = \xi_1(A + \frac{\xi_2}{\xi_1}, B + \frac{\xi_3}{\xi_1})$. Due to the assumptions on the support of $\hat{\psi}$, we see that, on the support $U_{j,\ell_1,\ell_2}$ of $\Gamma_{j,\ell_1,\ell_2}$, we have that $|2^j \frac{\xi_2}{\xi_1} - \ell_1| \leq 1$, $|2^j \frac{\xi_3}{\xi_1} - \ell_2| \leq 1$ and $2^{2j} \leq |\xi_1| \leq 2^{2j+2}$. Let $\ell_{1,m} = -A2^j, ; \ell_{2,m} = -B2^j$. It follows that

$$\begin{aligned} |\eta|^2 &= |\xi_1|^2 \left( (A + \frac{\xi_2}{\xi_1})^2 + (B + \frac{\xi_3}{\xi_1})^2 \right) \\ &\simeq 2^j \left( (2^j \frac{\xi_2}{\xi_1} + A2^j)^2 + (2^j \frac{\xi_3}{\xi_1} + B2^j)^2 \right) \\ &\simeq 2^j \left( (\ell_1 - \ell_{1,m})^2 + (\ell_2 - \ell_{2,m})^2 \right). \end{aligned}$$

For each pair of $(p, q) \in \mathbb{Z} \times \mathbb{Z}$, it is easy to see that the support of the function $g(2^j x_2 - p) g(2^j x_3 - q)$ is contained inside the set $I_{p,q} = (2^{-j}p - 2^{-j}, 2^{-j}p + 2^{-j}) \times (2^{-j}q - 2^{-j}, 2^{-j}q + 2^{-j})$ and that $\bigcup_{p,q=-2^j}^{2j} I_{p,q}$ is an open cover of $[-1, 1]^2$. We wil consider the following two types of integer pairs $(p, q)$:

$$J_{p,q}^{(1)} = \{(p, q) \in \mathbb{Z} \times \mathbb{Z} : I_{p,q} \bigcap \partial D \neq \emptyset\}, \quad J_{p,q}^{(2)} = \{(p, q) \in \mathbb{Z} \times \mathbb{Z} : I_{p,q} \bigcap \partial D = \emptyset\}.$$

We observe that there are at most $C2^j$ pairs of $(p, q)$ in $J_{p,q}^{(1)}$ and at most $C2^{2j}$ pairs of $(p, q)$ in $J_{p,q}^{(1)}$. The following two lemmata are the key estimates for the proof of Theorem 2.1. The ideas used below are based on the techniques developed by the authors for the shearlet-based analysis of singularities in [11, 13].

**Lemma 3.6.** *For a given $\epsilon > 0$ and for $(p, q)$ in $J_{p,q}^{(1)}$, we have that*

$$\sum_{(p,q) \in J_{p,q}^{(1)}} \sum_{m=1}^{M} \sum_{|\ell_1| \leq 2^j} \sum_{|\ell_2| \leq 2^j} \sum_{k \in \mathbb{Z}^3} |\beta_j^{(m,p,q)}(\ell_1, \ell_2, k)| \leq C\, 2^{2\epsilon j} 2^{-\frac{1}{2}j}.$$

*Proof.* Let $L$ be the differential operator:

$$L = \left( I - (\frac{2^{2j}}{2\pi})^2 \frac{\partial^2}{\partial \xi_1^2} \right) \left( 1 - (\frac{2^j}{2\pi})^2 \frac{\partial^2}{\partial \xi_2^2} \right)^2 \left( 1 - (\frac{2^j}{2\pi})^2 \frac{\partial^2}{\partial \xi_3^2} \right).$$

A direct computation shows that

$$
\begin{aligned}
&\beta_j^{(m,p,q)}(\ell_1,\ell_2,k) = \langle \widehat{\mathcal{T}}_j^{(m,p,q)}, \widehat{\psi_{j,\ell_1,\ell_2,k}}\rangle \\
&= \frac{2^{-2j} i}{2\pi} \int_D g(2^j x_2 - p) g(2^j x_3 - p) \int_{U_{j,\ell}} \frac{\xi\cdot(-1,A,B)}{|\xi|} W(2^{-2j}\xi)\,\Gamma_{j,\ell_1,\ell_2}(\xi) \\
&\times e^{2\pi i \xi\cdot(2^{-2j}(k_1-\ell_1 k_2-\ell_2 k_3-2^{2j}(Ax_2+Bx_3),2^{-j}(k_2-2^j x_2)),2^{-j}(k_3-2^j x_2))}\, d\xi\, dx_3\, dx_2 \\
&= \frac{2^{-2j} i}{2\pi} \int_D g(2^j x_2 - p) g(2^j x_3 - p) \int_{U_{j,\ell}} L^N\Big(\frac{\xi\cdot(-1,A,B)}{|\xi|}\, W(2^{-2j}\xi)\,\Gamma_{j,\ell_1,\ell_2}(\xi)\Big) \\
&\times L^{-N}\Big(e^{2\pi i \xi\cdot(2^{-2j}(k_1-\ell_1 k_2-\ell_2 k_3-2^{2j}(Ax_2+Bx_3),2^{-j}(k_2-2^j x_2)),2^{-j}(k_3-2^j x_2))}\Big)\, d\xi\, dx_3\, dx_2.
\end{aligned}
$$

It is easy to verify that, for $\xi \in U_{j,\ell} = U_{j,\ell_1,\ell_2}$, we have:

$$
|L^N\left(\frac{\xi\cdot(-1,A,B)}{|\xi|}\, W(2^{-2j}\xi)\,\Gamma_{j,\ell_1,\ell_2}(\xi)\right)| \le C_N\, 2^{-2j}.
$$

Since $|U_{j,\ell}| \le C\, 2^{4j}$, it follows that

$$
\begin{aligned}
|\beta_j^{(m,p,q)}(\ell_1,\ell_2,k)| \;\le\; & C_N \int_D \big((1+(k_1-\ell_1 k_2-\ell_2 k_3-2^{2j}(Ax_2+Bx_3))^2) \\
& \times\; \big(1+(k_2-2^j x_2)^2\big)\big(1+(k_3-2^j x_3)^2\big)\big)^{-N} dx_3 dx_2.
\end{aligned}
$$

For any given $\epsilon > 0$, if $|k_2 - 2^j x_2| > 2^{\epsilon j}$ or $|k_3 - 2^j x_3| > 2^{\epsilon j}$ or $|k_1 - k_2\ell_1 - k_3\ell_2 - 2^{2j}(Ax_2+Bx_3)| > 2^{\epsilon j}$ for all $(x_2,x_3) \in I_{p,q}$, then the above inequality with sufficient large $N$ yields the lemma. Thus, let us instead examine the situation where, for some $(x_2,x_3) \in I_{p,q}$, $k = (k_1,k_2,k_3)$ satisfies the following set of conditions:

$$
|k_2 - 2^j x_2| \le 2^{\epsilon j}, \quad |k_3 - 2^j x_3| \le 2^{\epsilon j}, \tag{3.15}
$$

$$
|k_1 - k_2\ell_1 - k_3\ell_2 - 2^{2j}(Ax_2+Bx_3)| \le 2^{\epsilon j}. \tag{3.16}
$$

To complete the proof, we want to obtain an $\ell^1$ estimate for the sequence $\{\beta_j^{(m,p,q)}(\ell_1,\ell_2,k),\ |\ell_1| \le 2^j,\ |\ell_2| \le 2^j,\ k \in \mathbb{Z}^3\}$. By the Parseval's Theorem on Fourier series, it is easy to derive an $\ell^2$ estimate for this sequence with respect to the index $k$. In order to apply Hölder's inequality and obtain the desired $\ell^1$ estimate from the $\ell^2$ estimate, we need to use the inequalities (3.15) and (3.16) to estimate the cardinality of the set of those $k$ that are really involved in the sequence $\{\beta_j^{(m,p,q)}(\ell_1,\ell_2,k)\}$. From the above two inequalities, we obtain the norm estimate $|k| \le C 2^{4j}$ for those $k$ in the set and this gives an upper bound for the cardinality of the set. Unfortunately, this estimate is too rough to be useful for the proof of the result of this lemma. In fact, we need a bound for the cardinality of the set of order $O(e^{(1+3\epsilon)j})$. To this end, we observe that a translation of the set will change the norm estimate of the set, but will not change the cardinality of the set. It turns out that, on each open interval $I_{p,q} \subset [-1,1]^2$, we can find the appropriate translation so that the estimate $O(e^{(1+3\epsilon)j})$ for the cardinality of the set is achieved.

Recalling that $I_{p,q} = (2^{-j}p - 2^{-j}, 2^{-j}p + 2^{-j}) \times (2^{-j}q - 2^{-j}, 2^{-j}q + 2^{-j})$, it follows that $2^j x_2 = p + \alpha(x_2)$, $2^j x_3 = q + \beta(x_3)$, and $2^{2j}(Ax_2 + Bx_3) = 2^j(Ap + Bq) + \gamma(x_2,x_3)$ with $|\alpha(x_2)| \le 1$, $|\beta(x_3)| \le 1$ and $|\gamma(x_2,x_3)| \le C2^j$ for all $(x_2,x_3) \in I_{p,q}$. Let

$$
\begin{aligned}
S_1 &= \{k_2 :\ |k_2 - 2^j x_2| \le 2^{\epsilon}\}, \quad T_1 = \{k_3 :\ |k_3 - 2^j x_3| \le 2^{\epsilon}\}, \ \text{ for some } (x_2,x_3) \in I_{p,q}, \\
S_2 &= \{k_2 :\ |k_2 - \alpha(x_2)| \le 2^{\epsilon}\}, \quad T_2 = \{k_3 :\ |k_3 - \beta(x_3)| \le 2^{\epsilon}\}, \quad \text{for some } (x_2,x_3) \in I_{p,q}, \\
J_1 &= \{k_1 :\ |k_1 - k_2\ell_1 - k_3\ell_2 - 2^{2j}(Ax_2+Bx_3)| \le 2^{\epsilon j},\ k_2 \in S_1,\ k_3 \in T_1\}, \quad \text{for some } (x_2,x_3) \in I_{p,q}, \\
J_2 &= \{k_1 :\ |k_1 - k_2'\ell_1 - k_3'\ell_2 - 2^{2j} - \gamma(x_2,x_3)| \le 2^{\epsilon j},\ k_2 \in S_2,\ k_3 \in T_2\}, \quad \text{for some } (x_2,x_3) \in I_{p,q}, \\
K_1 &= S_1 \times J_1 \times T_1, \quad K_2 = S_1 \times J_1 \times T_1.
\end{aligned}
$$

From inequalities (3.15) and (3.16), we see that we need to estimate card$(K_1)$. In the following we will show that, on $I_{p,q}$, $K_2$ is a translation of $K_1$ and that card$(K_2) \le C2^{(1+3\epsilon)j}$, which will yield the desired estimate card$(K_1) \le C2^{(1+3\epsilon)j}$.

Since $2^j x_2 = p + \alpha(x_2)$ and $2^j x_3 = q + \beta(x_3)$, it follows that card$(S_1) =$ card$(S_2)$ and card$(T_1) =$ card$(T_2)$. For any $k_2 \in S_2,\ k_3 \in T_2$, we have that $|k_2| \le 2^{\epsilon j} + |\alpha(x_2)| \le C2^{\epsilon j}$ and $|k_3| \le 2^{\epsilon j} + |\beta(x_3)| \le C2^{\epsilon j}$. Thus, card$(S_1) =$ card$(S_2) \le C2^{\epsilon j}$ and card$(T_1) =$ card$(T_2) \le C2^{\epsilon j}$ uniformly for $(x_2, x_3) \in I_{p,q}$. Note that, for $k_2 \in S_1,\ k_3 \in T_1$, we have that $k_2 = k_2 - p + p = k_2' + p$ and $k_3 = k_3 - q + q = k_3' + q$. Since $k_2' \in S_2$, $k_3' \in T_2$, we have that $|k_2'| \le C2^{\epsilon j}$ and $|k_3'| \le C2^{\epsilon j}$.

Similarly, for fixed $k_2, k_3$, let $J_1 = \{k_1 :\ |k_1 - k_2\ell_1 - k_3\ell_2 - 2^{2j}(Ax_2 + Bx_3)| \le 2^{\epsilon j}\}$ and $J_2 = \{k_1 : |k_1 - k_2'\ell_1 - k_3'\ell_2 - 2^{2j} - \gamma(x_2, x_3)| \le 2^{\epsilon j}\}$, for some $(x_2, x_3) \in I_{p,q}$ so that card$(J_1) =$ card$(J_2)$. For $k_1 \in J_2$, we have that $|k_1 - k_2'\ell_1 - k_3'\ell_2 - \gamma(x_2, x_3)| \le 2^{\epsilon j}$. It follows that

$$|k_1| \le |k_2'\ell_1| + |k_3'\ell_2| + \gamma(x_2, x_3)| + 2^{\epsilon j} \le C2^{(1+\epsilon)j},$$

for all $x_2, x_3 \in I_{p,q}$ such that $|\ell_1| \le 2^j$, $|\ell_2| \le 2^j$. We have shown that card$(J_1) =$ card$(J_2) \le C2^{(1+\epsilon)j}$, uniformly for all $(x_2, x_3) \in I_{p,q}$ and all $|\ell_1| \le 2^j$, $|\ell_2| \le 2^j$. Thus, for the rest of the proof, we can assume that card$(K_1) \le C\, 2^{(1+3\epsilon)j}$.

We recall that

$$\begin{aligned}\widehat{\mathcal{T}}_j^{(m,p,q)}(\xi) &= \frac{-W(^{-2j}\xi)}{(2\pi)^2|\xi|}\frac{\xi}{|\xi|}\cdot(-1, A, B)\int_D g(2^j x_2 - p)g(2^j x_3 - q)e^{-2\pi i\eta\cdot(x_2,x_3)}\,dx_3\,dx_2 \\ &= \frac{-2^{-2j}W(^{-2j}\xi)}{(2\pi)^2|\xi|}\frac{\xi}{|\xi|}\cdot(-1, A, B)\int_{2^j D} g(x_2 - p)g(x_3 - q)e^{-2\pi i 2^{-j}\eta\cdot(x_2,x_3)}\,dx_3\,dx_2.\end{aligned}$$

Similar to (3.14), we have the following estimate:

$$|\widehat{\mathcal{T}_j^{(m,p,q)}}(\xi)| \le C\,\frac{2^{-2j}}{|\xi|}\;\frac{1}{1 + 2^{-j}|B\xi_1 + \xi_3|}\;\frac{1}{1 + 2^{-j}|A\xi_1 + \xi_2 + b(B\xi_1 + \xi_3)|}.$$

It follows that, for $\xi \in U_{j,\ell_1,\ell_2}$, we have that

$$|\widehat{\mathcal{T}}_j^{(m,p,q)}(\xi)| \le C\,2^{-4j}(1 + |\ell_2 - \ell_{2,m}|)^{-1}\,(1 + |(\ell_1 - \ell_{1,m}) + b(\ell_2 - \ell_{2,0})|)^{-1}\,.$$

Using this last observation, together with the fact that the family

$$\{2^{-2j}e^{2\pi i\xi A_{(1)}^{-j}B_{(1)}^{-\ell_1,\ell_2}k}\,:\,k \in \mathbb{Z}^3\}$$

is an orthonormal basis for $L^2(U_{j,\ell_1,\ell_2})$, it follows that

$$\begin{aligned}\sum_{k\in\mathbb{Z}^3}|\beta_j^{(m,p,q)}(\ell_1, \ell_2, k)|^2 &= \int_{U_{j,\ell_1,\ell_2}}|W(2^{-2j}\xi)\,\Gamma_{j,\ell_1,\ell_2}(\xi)|^2\,|\widehat{\mathcal{T}}_j^{(m,p,q)}(\xi)|^2\,d\xi \\ &\le 2^{4j}2^{-8j}\,(1 + |\ell_2 - \ell_{2,m}|)^{-2}\,(1 + |(\ell_1 - \ell_{1,m}) + b(\ell_2 - \ell_{2,m})|)^{-2}\,.\end{aligned}$$

Since card$(K_1)$ involved in the above sum is of order $O(2^{1+3\epsilon)j})$, Hölder's inequality yields that

$$\sum_{k\in\mathbb{Z}^3}|\beta_j^{(m,p,q)}(\ell_1, \ell_2, k)| \le C\,2^{\frac{1+3\epsilon}{2}j}2^{-2j}\,(1 + |\ell_2 - \ell_{2,m}|)^{-1}\,(1 + |(\ell_1 - \ell_{1,m}) + b(\ell_2 - \ell_{2,m})|)^{-1}\,.$$

It follows that

$$\sum_{|\ell_1|\le 2^j}\sum_{|\ell_2|\le 2^j}\sum_{k\in\mathbb{Z}^3}|\beta_j^{(m,p,q)}(\ell_1, \ell_2, k)| \le C\,j^2\,2^{-\frac{3}{2}j}2^{\frac{3}{2}j} \le C\,2^{2\epsilon j}2^{-\frac{3}{2}j}.$$

Hence, computing also the sum over the indices $m$ and the indices $(p,q)$ in $J^{(1)}_{p,q}$, we have that

$$\sum_{(p,q)\in J^{(1)}_{p,q}} \sum_{m=1}^{M} \sum_{|\ell_1|\le 2^j} \sum_{|\ell_2|\le 2^j} \sum_{k\in\mathbb{Z}^3} |\beta_j^{(m,p,q)}(\ell_1,\ell_2,k)| \le C\, 2^{2\epsilon j} 2^{-\frac{1}{2}j}.$$

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

**Lemma 3.7.** *Using the notation introduced above, for a given $\epsilon > 0$ and for $(p,q)$ in $J^{(2)}_{p,q}$, we have that:*

$$\sum_{(p,q)\in J^{(2)}_{p,q}} \sum_{m=1}^{M} \sum_{|\ell_2-\ell_{2,0}|\ge 2^{\epsilon j}} \sum_{|\ell_2|\le 2^j} \sum_{k\in\mathbb{Z}^3} |\beta_j^{(m,p,q)}(\ell_1,\ell_2,k)| \le C\, 2^{2\epsilon j} 2^{-\frac{1}{2}j}$$

$$\sum_{(p,q)\in J^{(2)}_{p,q}} \sum_{m=1}^{M} \sum_{|\ell_1|\le 2^j} \sum_{|(\ell_1-\ell_{1,0})+b(\ell_2-\ell_{2,0})|\ge 2^{\epsilon j}} \sum_{k\in\mathbb{Z}^3} |\beta_j^{(m,p,q)}(\ell_1,\ell_2,k)| \le C 2^{2\epsilon j} 2^{-\frac{1}{2}j}$$

*Proof.* We will only prove the first of the two estimates since the second one can be proved using a similar argument.

As in the proof of Lemma 3.6, we use the following identity:

$$\widehat{\mathcal{T}}_j^{(m,p,q)}(\xi) = -\frac{2^{-2j}W(2^{-2j}\xi)}{(2\pi)^2|\xi|}\frac{\xi}{|\xi|}\cdot(-1,A,B)\int_{2^jD} g(x_2-p)g(x_3-q)e^{-2\pi i 2^{-j}\eta\cdot(x_2,x_3)}\,dx_3\,dx_2.$$

Since $I_{p,q}\bigcap\partial D=\emptyset$, we see that $\operatorname{supp}\left(g(x_2-p)g(x_3-q)\right)\bigcap\partial 2^jD=\emptyset$. We also have that $|\xi|\simeq 2^{2j}$ on $U_{j,\ell_1,\ell_2}$. Therefore, using integration by parts $N$ times with respect to the variable $x_2$ and $x_3$ in the integral above, we have that

$$|\widehat{\mathcal{T}}_j^{(m,p,q)}(\xi)| \le C\,2^{-4j}(1+|\ell_2-\ell_{2,m}|)^{-N}\left(1+|(\ell_1-\ell_{1,m})+b(\ell_2-\ell_{2,m})|\right)^{-N}.$$

It follows that

$$\begin{aligned}\sum_{k\in\mathbb{Z}^3}|\beta_j^{(m,p,q)}(\ell_1,\ell_2,k)|^2 &\le \int_{\Omega_{j,\ell_1,\ell_2}} |\Gamma_{j,\ell_1,\ell_2}(\xi)\,W(2^{-2j}\xi)|^2\,|\widehat{\mathcal{T}}_j^{(m,p,q)}(\xi)|^2\,d\xi\\ &\le 2^{4j}2^{-8j}(1+|\ell_2-\ell_{2,m}|)^{-2N}\left(1+|(\ell_1-\ell_{1,m})+b(\ell_2-\ell_{2,m})|\right)^{-2N}.\end{aligned}$$

As in the proof of Lemma 3.6, we can assume that $k\in K_1$ with $\operatorname{card}(K_1)=O(2^{(1+3\epsilon)j})$ uniformly for all $(p,q)$ (the case $k\in\mathbb{Z}^3\setminus K_1$ yields fast decay for the sequence). Hence, from the last inequality, we have that

$$\sum_{k\in K_1}|\beta_j^{(m,p,q)}(\ell_1,\ell_2,k)| \le C_N\,2^{\frac{1+3\epsilon}{2}l}2^{-2j}(1+|\ell_2-\ell_{2,m}|)^{-N}\left(1+|(\ell_1-\ell_{1,m})+b(\ell_2-\ell_{2,m})|\right)^{-N}.$$

This implies that

$$\sum_{|\ell_2-\ell_{2,m}|\ge 2^{\epsilon j}} \sum_{|\ell_1|\le 2^j} \sum_{k\in\mathbb{Z}^3} |\beta_j^{(m,p,q)}(\ell_1,\ell_2,k)| \le C_N\,2^{\frac{1+3\epsilon}{2}j}2^{-2j}2^{(N-1)\epsilon j}.$$

It follows that, for $N$ large enough, we have the estimate:

$$\begin{aligned}\sum_{(p,q)\in J^{(2)}_{p,q}} \sum_{m=1}^{M} \sum_{|\ell_2-\ell_{2,m}|\ge 2^{\epsilon j}} \sum_{|\ell_1|\le 2^j} \sum_{k\in\mathbb{Z}^3} |\beta_j^{(m,p,q)}(\ell_1,\ell_2,k)| &\le C_N\,2^{2j}2^{\frac{1+3\epsilon}{2}j}2^{-2j}2^{(N-1)\epsilon j}\\ &\le C\,2^{2\epsilon j}2^{-\frac{1}{2}j}. \quad \square\end{aligned}$$

We are now ready to prove Theorem 2.1.

We choose any $0 < \epsilon < \frac{1}{8}$ and we introduce the notation

$$S_{2,j}^{(m)} = \{(j, \ell_1, \ell_2, m, k) : |\ell_1 - \ell_{1,m}| \le C2^{\epsilon j},\ |\ell_2 - \ell_{2,m}| \le 2^{\epsilon j},\ k \in \mathbb{Z}^3\}.$$

We have that $S_{2,j} = \bigcup_{m=1}^{M} S_{2,j}^{(m)}$. Since $|\ell_2 - \ell_{2,m}| \le 2^{\epsilon j}$ and $|(\ell_1 - \ell_{1,m}) + b(\ell_2 - \ell_{2,m})| \le 2^{\epsilon j}$ yield $|\ell_1 - \ell_{1,m}| \le C2^{\epsilon j}$, it follows from Lemma 3.6 and Lemma 3.7 that

$$\delta_{2,j} = \sum_{\eta \in S_{2,j}^c} |\langle \mathcal{T}_j, \psi_\eta \rangle| \le C\, M\, 2^{2\epsilon j}\, 2^{-\frac{1}{2}j}.$$

Thus, since $0 < \epsilon < \frac{1}{8}$, Lemma 3.5 implies that $\delta_{2,j} = o(\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi})$.

As we pointed out above, to complete the proof of Theorem 2.1, it remains to show that $\mu_c(S_{2,j}, \Psi; \Phi) \to 0$, as $j \to \infty$. Since

$$\mu_c(S_{2,j}, \Psi; \Phi) = \max_\lambda \sum_{\eta \in S_{2,j}} |\langle \phi_\lambda, \psi_\eta \rangle| \le \max_\lambda \sum_{m=1}^{M} \sum_{\eta \in S_{2,j}^{(m)}} |\langle \phi_\lambda, \psi_\eta \rangle|,$$

it is sufficient to show that

$$\max_\lambda \sum_{\eta \in S_{2,j}^{(m)}} |\langle \phi_\lambda, \psi_\eta \rangle| \to 0, \quad \text{as } j \to \infty.$$

This means that the proof of Theorem 2.1 is completed once we show the following result.

**Proposition 3.8.** *Using the notation from above, for a given $\epsilon > 0$, we have that*

$$\max_{k' \in \mathbb{Z}^3} \sum_{|\ell_1 - \ell_{1,m}| \le C2^{\epsilon j}} \sum_{|\ell_2 - \ell_{2,m}| \le 2^{\epsilon j}} \sum_{k \in \mathbb{Z}^3} |\langle \phi_{j,k'}, \psi_{j,\ell_1,\ell_2,k} \rangle| \to 0, \quad \text{as } j \to \infty.$$

*Proof.* Let $L$ be the differential operator

$$L_1 = \left(I - \frac{1}{(2\pi)^2} \frac{\partial^2}{\partial \xi_1^2}\right) \left(I - \frac{1}{(2\pi)^2} \frac{\partial^2}{\partial \xi_2^2}\right) \left(I - \frac{1}{(2\pi)^2} \frac{\partial^2}{\partial \xi_3^2}\right).$$

For brevity, let

$$\alpha = (\alpha_1, \alpha_2, \alpha_3) = B_{(1)}^{[\ell_1,\ell_2]} A_{(1)}^{j} (2^{-2j} k') = (k_1' + 2^{-j}\ell_1 k_2' + 2^{-j}\ell_2 k_3', 2^{-j}k_2', 2^{-j}k_3').$$

By direct calculation, we have that

$$\begin{aligned}
\langle \widehat{\psi_{j,\ell,k}}, \widehat{\phi_{j,k'}} \rangle &= \int_{\mathbb{R}^3} \left(2^{-2j}\, \Gamma_{j,\ell_1,\ell_2}(\xi)\, e^{2\pi i \xi A_{(1)}^{-j} B_{(1)}^{[-\ell]} k}\right) \left(2^{-3j} W(2^{-2j}\xi)\, e^{-2\pi i 2^{-2j} \xi \cdot k'}\right) d\xi \\
&= 2^{-5j} \int_{R^3} \Gamma_{j,\ell_1,\ell_2}(\xi)\, W(2^{-2j}\xi)\, e^{2\pi i \xi [A_{(1)}^{-j} B_{(1)}^{[-\ell]}(k-\alpha)]}\, d\xi \\
&= 2^{-j} \int_{\mathbb{R}^3} \hat{\psi}_2(\tfrac{\eta_2}{\eta_1}) \hat{\psi}_2(\tfrac{\eta_3}{\eta_1})\, W^2(\eta_1, 2^{-j}(\ell_1\eta_1 + \eta_2), 2^{-j}(\ell_2\eta_1 + \eta_3))\, e^{2\pi i \eta \cdot (k-\alpha)}\, d\eta \\
&= 2^{-j} \int_{\mathbb{R}^3} L_1 \left(\hat{\psi}_2(\tfrac{\eta_2}{\eta_1}) \hat{\psi}_2(\tfrac{\eta_3}{\eta_1})\, W^2(\eta_1, 2^{-j}(\ell_1\eta_1 + \eta_2), 2^{-j}(\ell_2\eta_1 + \eta_3))\right) L_1^{-1}\Big(e^{2\pi i \eta \cdot (k-\alpha)}\Big)\, d\eta.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sum_{|\ell_1 - \ell_{1,m}| \le C2^{\epsilon j}} \sum_{|\ell_2 - \ell_{2,m}| \le 2^{\epsilon j}} \sum_{k \in \mathbb{Z}^3} |\langle \phi_{j,k'}, \psi_{j,\ell_1,\ell_2,k} \rangle| \\
&\le C\, 2^{2\epsilon j} 2^{-j} \sum_{k \in \mathbb{Z}^3} (1 + (k_1 - \alpha_1)^2)^{-1} (1 + (k_2 - \alpha_2)^2)^{-1} (1 + (k_3 - \alpha_3)^2)^{-1} \\
&\le C\, 2^{(-1+2\epsilon)j},
\end{aligned}$$

where the constant $C > 0$ is independent of $k'$. Since $\epsilon < \frac{1}{8}$, it follows that $-1 + 2\epsilon < 0$, and this implies that

$$\mu_c(S_{2,j}, \Psi; \Phi) \to 0 \quad \text{as } j \to \infty.$$

This completes the proof of the proposition. □

This also completes the proof of Theorem 2.1.

**Remark 3.1.** *Concerning the extension of our 3D result to the situation of singularities along general piece-wise smooth surfaces, we remark that our arguments do not work for general surfaces, e.g., a section of a sphere. In fact, the proofs of Lemma 3.6 and Lemma 3.7 rely on the crucial observation that all values of $p$ and $q$ from $2^{-j}$ to $2^j$ correspond to the same $\ell_{1,m}$ and $\ell_{2,m}$, and this requires the assumption that a singularity along a surface is linear or piece-wise linear. As a consequence, a different argument is needed to deal with singularities along more general surfaces.*

**Acknowledgments.** The authors acknowledge support from NSF grant DMS 1008900/1008907; the second author also acknowledges support from NSF grant DMS 1005799 and DMS 1320910.

# 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.*, 56:219–266, 2004.

[2] E. J. Candès, J. K. Romberg and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.

[3] S. S. Chen, D. L. Donoho and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43:129–159, 2001.

[4] R. R. Coifman and M. V. Wickerhauser. Wavelets and adapted waveform analysis. A toolkit for signal processing and numerical analysis, Different perspectives on wavelets (San Antonio, TX, 1993), 119–153. *Proc. Sympos. Appl. Math.*, 47, Amer. Math. Soc., Providence, RI, 1993.

[5] D. L. Donoho. For most large underdetermined systems of linear equations the minimal $\ell^1$-norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59: 797–829, 2006.

[6] D. Donoho and G. Kutyniok. Microlocal analysis of the geometric separation problem. *Comm. Pure Appl. Math.*, 66:1–47, 2013.

[7] G. R. Easley, D. Labate and W. Lim. Sparse directional image representations using the discrete shearlet transform. *Appl. Comput. Harmon. Anal.*, 25:25–46, 2008.

[8] R. Gribonval and E. Bacry. Harmonic decomposition of audio signals with matching pursuit. *IEEE Trans. Signal Proc.*, 51:1001-111, 2003.

[9] K. Guo, R. Houska and D. Labate. Microlocal analysis of singularities from directional multiscale representations. Approximation Theory XIV (San Antonio 2013), 173–196. *Springer Proceedings in Mathematics & Statistics*, 83, 2014.

[10] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM J Math. Anal.*, 39:298–318, 2007.

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

[12] K. Guo and D. Labate. Analysis and Detection of Surface Discontinuities using the 3D Continuous Shearlet Transform. *Appl. Comput. Harmon. Anal.*, 30:231–242, 2010.

[13] K. Guo and D. Labate. Optimally sparse representations of 3D data with $C^2$ surface singularities using Parseval frames of shearlets. *SIAM J Math. Anal.*, 44:851–886, 2012.

[14] K. Guo and D. Labate. Characterization of Piecewise Smooth Surfaces using the 3D Continuous Shearlet Transform. *J. Fourier Anal. Appl.*, 18:488–516, 2012.

[15] K. Guo and D. Labate. The construction of smooth Parseval frames of shearlets. *Math. Model. Nat. Phenom.*, 8:82-105, 2013.

[16] K. Guo, D. Labate and W. Lim. Edge analysis and identification using the continuous shearlet transform. *Appl. Comput. Harmon. Anal.*, 27:24–46, 2009.

[17] G. Garrigòs and E. Hernandez. Sharp Jackson and Bernstein inequalities for N-term approximation in sequence spaces with applications. *Indiana Univ. Math. J.*, 53:1739–1762, 2004.

[18] E. Hernandez, and G. Weiss. *A first course on wavelets*. CRC Press, Boca Raton, FL, 1996.

[19] G. Kutyniok, Geometric Separation by Single-Pass Alternating Thresholding. *Appl. Comput. Harmon. Anal.*, 36:23–50, 2014.

[20] G. Kutyniok and D. Labate. Resolution of the wavefront set using continuous shearlets. *Trans. Amer. Math. Soc.*, 361:2719–2754, 2009.

[21] G. Kutyniok, J. Lemvig and W.-Q Lim. Optimally Sparse Approximations of 3D Functions by Compactly Supported Shearlet Frames. *SIAM J. Math. Anal.*, 44: 2962–3017, 2012.

[22] G. Kutyniok, M. Shahram and X. Zhuang. ShearLab: A Rational Design of a Digital Parabolic Scaling. *SIAM J. Imaging Sciences*, 5:1291–1332, 2012.

[23] D. Labate, W. Lim, G. Kutyniok and G. Weiss. Sparse multidimensional representation using shearlets. *SPIE Proc.*, 5914: 254–262, 2005.

[24] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Proc.*, 41:3397–3415, 1993.

[25] F. G. Meyer, A. Averbuch and R. R. Coifman. Multi-layered Image Representation: Application to Image Compression. *IEEE Trans. Image Process.*, 11:1072–1080, 2002.

[26] J.-L. Starck, M. Elad and D.L. Donoho. Redundant Multiscale Transforms and their Application for Morphological Component Analysis. *Adv. Imag. Elect. Phys.*, 132:287–348, 2004.

[27] J.-L. Starck, M. Elad and D.L. Donoho. Image Decomposition Via the Combination of Sparse Representation and a Variational Approach. *IEEE Trans. Image Process.*, 14:1570–1582, 2005.

[28] E. M. Stein. *Harmonic Analysis: real-variable methods, orthogonality, and oscillatory integrals*. Princeton University Press, Princeton, 1993

[29] J.-L. Starck, M. Nguyen and F. Murtagh, Wavelets and curvelets for image deconvolution: A combined approach. *Signal Process.*, 83:2279–2283, 2003.

[30] R. S. Strichartz. *A guide to distribution theory and Fourier transforms*. World Scientific Pub., 2003.

[31] G. Teschke. Multi-frame representations in linear inverse problems with mixed multi-constraints. *Appl. Comput. Harmon. Anal.*, 22:43–60, 2007.