

Geometric Separation in $\mathbb{R}^3$

Kanghui Guo¹ and Demetrio Labate²

July 6, 2017

Abstract

The geometric separation problem, initially posed by Donoho and Kutyniok [7], aims to separate a distribution containing a non-trivial superposition of point and curvilinear singularities into its distinct geometric constituents. The solution proposed in [7] considers expansions with respect to a combined wavelet-curvelet dictionary and applies an ℓ^1 -norm minimization over the expansion coefficients to achieve separation asymptotically at fine scales. However, the original proof of this result uses a heavy machinery relying on sparse representations of Fourier integral operators which does not extend directly to the 3D setting. In this paper, we extend the geometric separation result to the 3D setting using a novel and simpler argument which relies in part on techniques developed by the authors for the shearlet-based analysis of curvilinear edges. Our new result also yields a significantly simpler proof of the original 2D geometric separation problem and extends a prior result by the authors which was limited to piecewise linear singularities.

Key words and phrases: analysis of singularities, cluster coherence, geometric separation, ℓ^1 minimization, shearlets, sparse representations, wavelets.

AMS Mathematics Subject Classification: 42C15, 42C40.

1 Introduction

Combining multiple basis representations and taking advantage of their sparsity properties to build a more efficient representation of complex data has a rather long history in applied harmonic analysis and signal processing. Among the earliest formulations of combined-basis representations, we recall the seminal papers of Coifman and Wickerhauser [5] and Mallat and Zhang [28]. Perhaps the first attempt to introduce a rigorous formalization of these ideas was the method of Basis Pursuit [4] by Chen, Donoho and Saunders which established ℓ^1 -norm minimization as a method to promote sparse representations from multiple bases. In more recent years, other mostly empirical papers have further exploited this point of view and provided remarkable applications to problems from signal and image processing. Starck et al. [30, 31], for instance, proposed an algorithmic approach, called Morphological Component Analysis (MCA), which assumes that a signal is the linear mixture of several morphological components, each one endowed with specific geometric properties. Under the assumption that such components are sufficiently distinct and that each one is sparsely represented in a specific basis, MCA algorithms (using ℓ^1 -norm minimization) are able to effectively separate the various signal components in many numerical applications. In addition to such work, we also recall the contributions in [9, 29, 33, 35].

More recently, Donoho and Kutyniok [7, 21] introduced a rigorous mathematical formalization of the problem of separating data into geometrically distinct components. Their motivation is the observation that the success of many successful numerical algorithms based on multiple basis representations “stem from an interplay between geometric properties of objects to be separated and the harmonic analysis for singularities of various geometric types” (cf. [7]). As a mathematical idealization of two-dimensional data containing

¹Department of Mathematics, Missouri State University, Springfield, Missouri 65804, USA. E-mail: KanghuiGuo@MissouriState.edu

²Department of Mathematics, University of Houston, Houston, Texas 77204, USA. E-mail: dlabate@math.uh.edu

distinct geometric constituents, they consider distributions on $\mathbb{R}^2$ 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. The question is: how to separate f into its components $\mathcal{P}$ and $\mathcal{T}$?

To address this question, they observe that while points and curves may overlap spatially, they are separated microlocally. Therefore, they construct 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 ℓ^1 -norm. The choice of such dictionary is due to the fact that wavelets provide very sparse representations of point-like singularities, while curvelets [2] or shearlets [11, 27] provide very sparse representations of curve-like singularities. By applying an ℓ^1 -norm minimization over the expansion coefficients of the combined dictionary, they prove that f can be separated into its components $\mathcal{P}$ and $\mathcal{T}$ asymptotically at fine scales.

The proof of this separation result in [7] relies on the heavy machinery of a sparse matrix representation of Fourier integral operators in $\mathbb{R}^2$ and does not extend directly to the 3-dimensional setting. In [17], we introduced a different and simpler argument to deal with the 3-dimensional setting which is based on techniques we previously developed for the geometric characterization of edge singularities in terms of the shearlet transform [12, 19, 23]. However, this result was limited to the separation of point-wise and polyhedral singularities in $\mathbb{R}^3$, as our techniques could not be extended to the more difficult situation of curvilinear singularities in $\mathbb{R}^3$.

In this paper, we finally introduce a new and more powerful argument that allows us to handle the geometric separation problem in the case of 3-dimensional curvilinear singularities. Similar to the general approach in [7, 17], we consider a combined dictionary of wavelets and shearlets and adopt the important notion of *cluster coherence* as a main tool to prove geometric separation. The most critical and difficult part of this proof is the derivation of appropriate estimates on the cluster coherence of the wavelet and shearlet bases. As we will further explain below, this part of the proof is new and relies in part on techniques developed by the authors for the shearlet-based analysis of curvilinear edges [13, 15, 18].

In addition to solving the 3D geometric separation problem, our new approach also yields a much simpler and more streamlined argument for the corresponding 2D problem, without the need of the machinery based on Fourier integral operators of the original arguments in [7]. We also recall that the geometric separation result has important implications to the solution of the inpainting problem, as shown by King et al [22].

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 this theorem in Section 3.

Notation

In this paper, 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 ξ 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

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

Given the functions f and g , we use the notation $f \simeq g$ if there exist constants $0 < C_1 \leq C_2 < \infty$, independent of x , such that $C_1 g(x) \leq f(x) \leq C_2 g(x)$. Similarly, given the index set J , we use the notation $f(j) \simeq g(j)$ if there exist $0 < C_1 \leq C_2 < \infty$, independent of j , such that $C_1 g(j) \leq f(j) \leq C_2 g(j)$ for all $j \in J$.

2 The geometric separation problem

As a model of multidimensional data found in many applications, it is often useful to consider functions or distributions containing different types of singularities; for instance, singularities supported at single points or surface boundaries 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 point singularities and $\mathcal{T}$ is a section of a paraboloid.

Our goal is to find a highly sparse representation of f , that is, to derive a representation method such that f can be accurately approximated using a ‘very small’ number of representation coefficients. 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 precise sense, for functions with point singularities, while shearlets were shown to provide optimally sparse representations for functions with discontinuities along piecewise smooth surfaces [14, 24]. 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 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 [7], we take $\mathcal{P}$ to be of the form

$$\mathcal{P} = \sum_{i=1}^N |x - y^{(i)}|^{-2}, \quad (2.1)$$

which defines a distribution being smooth away from the singularity points $y^{(i)} \in \mathbb{R}^3$, $i = 1, \dots, N$. Next, to define a singularity supported on a surface boundary, we let B to be a section of a paraboloid in $\mathbb{R}^3$ with graph $z(u, v)$ for $(u, v) \in U \subset \mathbb{R}^2$. For $\alpha(u) \in C_0^\infty(U)$ and $\phi \in \mathcal{S}(\mathbb{R}^3)$, we define a distribution $\mathcal{T}$ concentrated on B as

$$\langle \mathcal{T}, \phi \rangle = \int_U \phi(z(u, v), u, v) \alpha(u, v) du dv.$$

The reason for choosing the exponent -2 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 singularities comparable *at each scale*. Without this assumption, it would be possible to trivially separate the two components of f at different scales, as the energy of each singularity would dominate at a certain scale. By contrast, the model we adopt (as in [7]) makes the separation problem challenging at every scale.

To justify our observation about the matching energies, we remark that $\widehat{\mathcal{P}}(\xi) \simeq |\xi|^{-2}$ (cf. [34, Ch.4]) and this implies that $\int_{2^{2j}}^{2^{2j+2}} |\widehat{\mathcal{P}}(\xi)|^2 d\xi \simeq 2^{2j}$. In Section 3.2, we will show that $\mathcal{T}$ satisfies a similar estimate so that also in this case $\int_{2^{2j}}^{2^{2j+2}} |\widehat{\mathcal{T}}(\xi)|^2 d\xi \simeq 2^{2j}$.

Following the language introduced in [7], we hence state *geometric separation problem* as follows.

Geometric separation problem. *Given the observation $f = \mathcal{P} + \mathcal{T}$, we want to recover the unknown components $\mathcal{P}$ and $\mathcal{T}$ of f based only on the knowledge that they are of the form (2.1) and (2.2).*

To solve this problem, we will adapt the strategy based on ℓ^1 minimization proposed by Donoho and Kutyniok [7]. That is, 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 we will enforce sparsity by minimizing the representation coefficients in the ℓ^1 -norm. As mentioned above, the sparsity-inducing properties of the ℓ^1 -norm are well known in applied harmonic analysis and play a critical role, for instance, in the celebrated theory of *compressed sensing* (cf. [3, 6]).

In the following, to simplify notation and avoid unnecessary calculations we will assume that $\mathcal{P}$ contains only one singularity point. In addition we will assume that the point singularity is centered at the origin, that is

$$\mathcal{P} = |x|^{-2}.$$

For the singularity supported on a surface boundary, we let $B = \{(\frac{1}{2}(x_2^2 + x_3^2), x_2, x_3) : (x_2, x_3) \in U\}$, where $U = \{(x_2, x_3) : x_2^2 + x_3^2 \leq 1\}$. Hence, for $\alpha(u) \in C_0^\infty(U)$ and $\phi \in \mathcal{S}(\mathbb{R}^3)$, we define $\mathcal{T}$ as

$$\langle \mathcal{T}, \phi \rangle = \int_U \phi(\frac{1}{2}(x_2^2 + x_3^2), x_2, x_3) \alpha(x_2, x_3) dx_2 dx_3. \quad (2.2)$$

Let us next define the wavelets and shearlets systems that we will use to represent $f = \mathcal{P} + \mathcal{T}$.

2.1 A Parseval frame of 3D wavelets

As our wavelet system, we will choose a *Parseval frame of Lemarié-Meyer wavelets* (cf. [20]) in $L^2(\mathbb{R}^3)$. This system will be denoted as $\Phi = \{\phi_\lambda : \lambda \in \Lambda\}$, for $\Lambda = \{\lambda = (j, k), j \geq -1, k \in \mathbb{Z}^3\}$, where the functions $\phi_\lambda = \phi_{j,k} \in L^2(\mathbb{R}^3)$ are defined in the Fourier domain by

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

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

$$|\widetilde{W}(\xi)|^2 + \sum_{j \geq 0} |W(2^{-2j}\xi)|^2 = 1, \quad \text{for a.e. } \xi \in \widehat{\mathbb{R}}^3. \quad (2.3)$$

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

$$[-2^{-2j-1}, 2^{-2j-1}]^3 \setminus [-2^{-2j-4}, 2^{-2j-4}]^3 \subset \widehat{\mathbb{R}}^3, \quad (2.4)$$

and the resulting 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 frequency bands indexed by $j \geq 0$.

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

Shearlet were introduced to overcome certain limitations of conventional wavelets in the analysis of multivariate functions [27]. Similar to the curvelets of Candès and Donoho [2], they 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. By combining multiscale analysis and high directional sensitivity, shearlets are able to precisely characterize of the geometry of singularities of functions and distributions of several variables [12, 13, 19, 25] and enable optimally sparse representations, in a precise sense, for a large class of multivariate functions where traditional wavelets are suboptimal [11, 14].

With respect to curvelets, shearlets have some distinctive 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 [8, 26]. Furthermore, there is a well-developed shearlet-based theory for the analysis of singularities (cf. [10] in addition to the references cited above). This theory sets the foundation for the main

ideas that we employ in this paper for the analysis of surface singularities and it is the main reason for selecting this representation in our approach to the geometric separation problem.

Our shearlet system is defined by introducing an angular subdivision within the multiscale decomposition associated with the window functions $\widehat{W}^2, W_j^2$ used above for the construction of the wavelet system. For this construction, we start by first partitioning the Fourier space $\widehat{\mathbb{R}}^3$ into the following 3 pyramidal regions in $\widehat{\mathbb{R}}^3$:

$$\begin{aligned} \mathcal{C}_1 &= \left\{ (\xi_1, \xi_2, \xi_3) \in \widehat{\mathbb{R}}^3 : \left| \frac{\xi_2}{\xi_1} \right| \leq 1, \left| \frac{\xi_3}{\xi_1} \right| \leq 1 \right\}, \\ \mathcal{C}_2 &= \left\{ (\xi_1, \xi_2, \xi_3) \in \widehat{\mathbb{R}}^3 : \left| \frac{\xi_1}{\xi_2} \right| < 1, \left| \frac{\xi_3}{\xi_2} \right| \leq 1 \right\}, \\ \mathcal{C}_3 &= \left\{ (\xi_1, \xi_2, \xi_3) \in \widehat{\mathbb{R}}^3 : \left| \frac{\xi_1}{\xi_3} \right| < 1, \left| \frac{\xi_2}{\xi_3} \right| < 1 \right\}. \end{aligned}$$

We let $W \in C_0^\infty(\mathbb{R}^3)$ be the *same* window as the one defined in Section 2.1 and let $v \in C^\infty(\mathbb{R})$ be an appropriate ‘bump function’ satisfying $\text{supp } v \subset [-1, 1]$ and

$$|v(u-1)|^2 + |v(u)|^2 + |v(u+1)|^2 = 1 \quad \text{for } |u| \leq 1. \quad (2.5)$$

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

$$\{\psi_{j,\ell,k}^{(d)} : j \geq 0, -2^j \leq \ell_1, \ell_2 \leq 2^j, k \in \mathbb{Z}^3\}, \quad (2.6)$$

where

$$\hat{\psi}_{j,\ell,k}^{(d)}(\xi) = |\det A_{(d)}|^{-j/2} W(2^{-2j}\xi) V_{(d)}(\xi A_{(d)}^{-j} B_{(d)}^{[-\ell]}) e^{2\pi i \xi A_{(d)}^{-j} B_{(d)}^{[-\ell]} k}, \quad (2.7)$$

$V_{(1)}(\xi_1, \xi_2, \xi_3) = v(\frac{\xi_2}{\xi_1})v(\frac{\xi_3}{\xi_1})$, $V_{(2)}(\xi_1, \xi_2, \xi_3) = v(\frac{\xi_1}{\xi_2})v(\frac{\xi_3}{\xi_2})$, 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

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

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

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

Notice that $(B_{(d)}^{[\ell]})^{-1} = B_{(d)}^{[-\ell]}$. 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 (2.6) have compact support in Fourier domain. In particular, for $d = 1$, the shearlets $\hat{\psi}_{j,\ell,k}^{(1)}(\xi)$ can be written explicitly as

$$\hat{\psi}_{j,\ell_1,\ell_2,k}^{(1)}(\xi) = 2^{-2j} W(2^{-2j}\xi) v\left(2^j \frac{\xi_2}{\xi_1} - \ell_1\right) v\left(2^j \frac{\xi_3}{\xi_1} - \ell_2\right) e^{2\pi i \xi A_{(1)}^{-j} B_{(1)}^{[-\ell_1, -\ell_2]} k}, \quad (2.8)$$

showing that their supports are contained inside the regions

$$\begin{aligned} U_{j,\ell} &= U_{j,\ell_1,\ell_2} \\ &= \{(\xi_1, \xi_2, \xi_3) : \xi_1 \in [-2^{2j-1}, -2^{2j-4}] \cup [2^{2j-4}, 2^{2j-1}], |\frac{\xi_2}{\xi_1} - \ell_1 2^{-j}| \leq 2^{-j}, |\frac{\xi_3}{\xi_1} - \ell_2 2^{-j}| \leq 2^{-j}\}. \end{aligned} \quad (2.9)$$

That is, the shearlets $\hat{\psi}_{j,\ell,k}^{(1)}$ have supports contained in *trapezoidal regions* defined at *various scales*, controlled by $j > 0$, and *various orientations*, controlled by the shear parameters ℓ_1, ℓ_2 . This shows that elements of the

shearlet system (2.6) the 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.

A *Parseval frame of shearlets* for $L^2(\mathbb{R}^3)$ is obtained by combining the shearlet systems (2.6) associated with the cone-shaped regions $\mathcal{C}_d$ 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 Lemerìe-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 a technical detail. To ensure that the frame of shearlets obtained by combining the elements from the different pyramidal systems is tight while guaranteeing that all such elements are C_0^∞ in the Fourier domain, one has to slightly modify the functions $\psi_{j,\ell_1,\ell_2,k}^{(d)}$, for $\ell_1, \ell_2 = \pm 2^j$ (these are the functions whose support overlap the boundaries of the regions $\mathcal{P}_d$) by 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 [14, 16].

2.3 Main theorem

Our main theorem below shows that it is possible to separate geometrically distinct components of a distribution $f = \mathcal{P} + \mathcal{T}$ by taking advantage of the sparsity properties of the Parseval frames of wavelets and shearlets. Similar to the result in [7, 17], the separation result holds *asymptotically* in scale, that is, we can separate point singularities and singularities along a parabolic surface only as a limiting process, when the scale tends to zero (i.e., $j \rightarrow \infty$). To formulate this result, we start by deriving an appropriate multiscale decomposition of f .

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 (2.4). Consistently with this decomposition, we define a family of band-pass filters F_j , $j \geq -1$, by $\widehat{F_j}(\xi) = W(2^{-2j}\xi)$, for $j \geq 0$, $\widehat{F_{-1}}(\xi) = \widehat{W}(\xi)$. By applying these filters to f , $\mathcal{P}$ and $\mathcal{T}$ we obtain

$$\mathcal{P}_j = \mathcal{P} * F_j, \quad \mathcal{T}_j = \mathcal{T} * F_j, \quad f_j = f * F_j, \quad (2.10)$$

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 $[-2^{2j-1}, 2^{2j-1}]^3 \setminus [-2^{2j-4}, 2^{2j-4}]^3 \subset \widehat{\mathbb{R}}^3$. In addition, for $f \in L^2(\mathbb{R}^3)$, it follows from (2.3) that

$$f = \sum_j F_j * f_j, \quad (2.11)$$

with convergence in the L^2 -norm.

Let $\mathcal{F}_j$ denote the range of the operator of convolution with F_j . Using a simple calculation one can verify that 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

$$\Lambda_j = \{\lambda = (j', k) : |j' - j| \leq 1, k \in \mathbb{Z}^3\} \subset \Lambda \quad (2.12)$$

and

$$M_j = \{\eta = (j', \ell, k, d) : |j' - j| \leq 1, |\ell_1| \leq 2^j, |\ell_2| \leq 2^j, k \in \mathbb{Z}^3, d = 1, 2, 3\} \subset M. \quad (2.13)$$

Due to the Parseval frame property and the observation above, any function $f_j \in \mathcal{F}_j$ can be expanded using *only* the elements of the wavelet system in Λ_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

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

or we can use the shearlet system to represent f_j as

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

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

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

for an appropriate choice of coefficients $p = (p_\lambda)$ and $t = (t_\eta)$. Since the last expression involves an overcomplete dictionary, there are many possible choices of coefficients p and t , some of which may provide sparser representations than either one of the two expansions above. Similar to [7, 17], we seek a solution enabling a *geometric separation*, that is, we consider the following dual-frame *component separation problem* based on ℓ_1 minimization:

$$(P_j^*, T_j^*) = \arg \min(\|p\|_1 + \|t\|_1), \quad \text{subject to } f_j = P_j + T_j, \quad (2.14)$$

where $p_\lambda = \langle P_j, \phi_\lambda \rangle$, $\lambda \in \Lambda_j$ and $t_\eta = \langle T_j, \psi_\eta \rangle$, $\eta \in M_j$. It follows from (2.11) that, if we let $\tilde{P} = \sum_j F_j * P_j$, $\tilde{T} = \sum_j F_j * T_j$, then we can express f as the superposition $f = \tilde{P} + \tilde{T}$.

The main result of our paper is the following theorem, showing that, by applying ℓ_1 minimization over the expansion coefficients of f with respect to our combined wavelet-shearlet dictionary, we achieve the *separation* of the distinct geometric objects $\mathcal{P}$ and $\mathcal{T}$, *asymptotically at fine scales*. That is, asymptotically as the scale tends to zero, the pointlike component of f is captured by the Parseval frame of wavelets and the curvilinear component of f is captured by the Parseval frame of shearlets.

Theorem 2.1. *Let Φ and Ψ be the Parseval frames of wavelets and shearlets, respectively, defined above, with Λ_j , M_j given by (2.12)-(2.13). Let $f_j = P_j + T_j$ be given as above and $\mathcal{P}_j$, $\mathcal{T}_j$ be given by (2.10). Then*

$$\lim_{j \rightarrow \infty} \frac{\|P_j - \mathcal{P}_j\|_{1,\Phi} + \|T_j - \mathcal{T}_j\|_{1,\Psi}}{\|\mathcal{P}_j\|_2 + \|\mathcal{T}_j\|_2} = 0.$$

This theorem extends our previous result in [17], where $\mathcal{T}$ was a piecewise linear singularity. We also recall that the geometric separation result originally obtained in [7] deals with 2D images containing point-like and smooth curve-like singularities. The approach presented in this paper can be easily adapted to the two-dimensional case yielding a much simpler and more direct proof of the result in [7].

As already observed in [7], the method of geometric separation presented in this work can be generalized to other situations, for example we can allow small perturbations and consider $f = (\mathcal{P} + \mathcal{T} + g)h$, where g, h are smooth function of rapid decay. Then we can set $\|f_j\|_2$ in the denominator of the expression in Theorem 2.1. One could also potentially consider the situation where $f = (\mathcal{P} + \mathcal{T} + \mathcal{L})$, where $\mathcal{L}$ is a line singularities. In this case, one should introduce a third dictionary in addition to wavelets and shearlets that sparsely represents $\mathcal{L}$ and has some incoherence with respect to the other two dictionaries. One reasonable choice could a dictionary of ridgelets [1], as it is optimally suited to deal with line singularities.

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

3 Proof of main theorem

Our proof follows the general architecture of the proof in [7, 17], which relies on the notion of cluster coherence. The most critical and difficult part of the proof is about showing that the cluster coherence goes asymptotically to zero for the three-dimensional curvilinear discontinuities considered in this paper. The way we address this part of the proof is completely original and does not follow from the arguments in [7, 17] which in fact cannot handle this situation. The new technical elements of our new arguments are contained in the proofs of Lemmata 3.5, 3.6 and in the proof of Theorem 2.1 in Section 3.2.

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 Λ and M that we denote as $S_{1,j} \subset \Lambda_j$ and $S_{2,j} \subset M_j$. Following the terminology in [7], 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, in the proof of Theorem 2.1 (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 ℓ^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 [7]. 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 Φ be the matrix representation of the Parseval frame of wavelets and Ψ 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 $\text{supp}(\hat{g}_j) \subset \mathcal{F}_j$, 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 known observation (Proposition 2.1 in [7]) illustrates the relationship between joint concentration and data separation.

Proposition 3.1. *Suppose that, for $j \in \mathbb{Z}$, $f_j = P_j + T_j$ so that each component of f_j is relatively sparse in Φ or Ψ , that is,*

$$\|1_{S_{1,j}^c} \Phi^T P_j\|_1 \leq \delta_{1,j}, \quad \|1_{S_{2,j}^c} \Psi^T T_j\|_1 \leq \delta_{2,j}.$$

If (P_j^, T_j^*) solves (2.14), then*

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

A related observation from [7] is that the joint concentration is bounded above by the maximum of the cluster coherences. We have (cf. [7, Lemma 2.1]):

Lemma 3.2.

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

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

$$\mu_c(S_{1,j}, \Phi; \Psi) \rightarrow 0, \quad \mu_c(S_{2,j}, \Psi; \Phi) \rightarrow 0, \quad \text{as } j \rightarrow \infty. \quad (3.15)$$

The rest of the proof is hence devoted to construct such sets $S_{1,j}$, $S_{2,j}$ and prove estimates (3.15). In Section 3.1, we will select an appropriate set $S_{1,j}$ and show that $\mu_c(S_{1,j}, \Phi; \Psi) \rightarrow 0$ and $\delta_{1,j} = o(\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi})$. This part of the proof is rather simple and follows from an idea similar to [7]. For the difficult part of the proof, concerning the analysis of curvilinear, it is not possible to apply the argument from [7] or [17] and we introduce a novel approach which is derived in Section 3.2. Our new argument is inspired in part from techniques for the analysis of singularities developed for the characterization of piecewise smooth boundaries of multivariate functions in [13, 15, 18].

In the following, for all our arguments, it will be sufficient to consider the shearlet system associated with the pyramidal shaped regions $\mathcal{C}_1 \subset \mathbb{R}^3$ only since the systems in $\mathcal{C}_2$ and $\mathcal{C}_3$ behave essentially in the same way. The elements (2.8) of such shearlet system can be written¹ as

$$\hat{\psi}_{j,\ell_1,\ell_2,k}^{(1)}(\xi) = 2^{-2j} \Gamma_{j,\ell_1,\ell_2}(\xi) e^{2\pi i \xi A_{(1)}^{-j} B_{(1)}^{[-\ell_1, -\ell_2]} k}, \quad (3.16)$$

where

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

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 Γ_{j,ℓ_1,ℓ_2} is supported inside the set U_{j,ℓ_1,ℓ_2} , given by (2.9). It is easy to verify that its measure satisfies $|U_{j,\ell_1,\ell_2}| \leq C 2^{4j}$.

3.1 Analysis of point singularities

This section is very similar to the corresponding section in [17] and is reported below for completeness.

In the following, we will select a set $S_{1,j}$ and prove that $\mu_c(S_{1,j}, \Phi; \Psi) \rightarrow 0$ and $\delta_{1,j} = o(\|\mathcal{P}_j\|_{1,\Phi} + \|\mathcal{T}_j\|_{1,\Psi})$, asymptotically as $j \rightarrow \infty$.

Let $\phi_{j',k'}$ and $\psi_{j,\ell_1,\ell_2,k}$ be generic elements from a Parseval frame of wavelets and shearlets, respectively. Due to our assumptions on the Fourier support of W , for any ℓ_1, ℓ_2, k and k' we have that $\langle \widehat{\psi_{j,\ell_1,\ell_2,k}}, \widehat{\phi_{j',k'}} \rangle = 0$ if $|j - j'| > 1$. Hence, for all large j' and $j = j' - 1, j', j' + 1$, we observe

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

where C is independent of ℓ_1, ℓ_2, k, k' and j .

¹Here we ignore that boundary elements corresponding to $\ell = \pm 2^j$ are slightly modified as it is irrelevant for our arguments.

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

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

and, thus, $\mu_c(S_{1,j}, \Phi; \Psi) \rightarrow 0$, as $j \rightarrow \infty$.

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

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

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

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

By choosing N sufficiently large, we conclude that

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

3.2 Analysis of curvilinear singularities

We start by proving the following estimate for the functions $\mathcal{T}_j$, $j \geq 0$.

Proposition 3.3. *For $j \geq 0$ we have:*

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

To prove Proposition 3.3, we need the following lemma which is a special case of the classical method of stationary phase (cf. Proposition 8.6 in [32]).

Lemma 3.4. *Let U_ϵ be the ball in $\mathbb{R}^2$ with center at the origin and radius $\epsilon > 0$ and $\psi \in C_0^\infty(U_\epsilon)$. Let $J(\lambda) = \int_{\mathbb{R}^2} e^{i\lambda H(u)} \psi(u) du$. For $u_0 \in \text{supp } \psi \subset U_\epsilon$, if $\Delta\phi(u_0) = 0$ and the determinant of the Hessian matrix of H at u_0 is not zero, then*

$$J(\lambda) = e^{i\lambda H(u_0)} \lambda^{-1} [a(u(u_0)) + O(\lambda^{-\frac{1}{2}})],$$

as $\lambda \rightarrow \infty$, where $a(u_0)$ is a smooth function of u_0 .

Proof of Proposition 3.3. The Fourier transform of $\mathcal{T}$, given by (2.2), is

$$\widehat{\mathcal{T}}(\xi) = \widehat{\chi_B}(\xi) = \int_U e^{-2\pi i \xi \cdot (\frac{1}{2}(u_1^2 + u_2^2), u)} \alpha(u) du,$$

where $\xi \in \mathbb{R}^3$ and the sets B, U are defined in Sec. 2. By converting to polar coordinates with $\xi = \rho \Theta(\theta, \phi)$, where $\Theta(\theta, \phi) = (\cos \theta \sin \phi, \sin \theta \sin \phi, \cos \phi)$, $\theta \in [0, 2\pi]$ and $\phi \in [0, \pi]$, we have

$$\widehat{\mathcal{T}}(\rho, \theta, \phi) = \int_U e^{-2\pi i \rho \Theta(\theta, \phi) \cdot (\frac{1}{2}(u_1^2 + u_2^2), u)} \alpha(u) du.$$

Let $H(u) = \Theta(\theta, \phi) \cdot (\frac{1}{2}u_1^2 + \frac{1}{2}u_2^2, u)$. Then $\Delta H(u) = (\Theta(\theta, \phi) \cdot (u_1, 1, 0), \Theta(\theta, \phi) \cdot (u_2, 0, 1))$. It is easy to see that, for the given θ and ϕ , the solution $u_{\theta, \phi}$ of $\Delta H(u) = (0, 0)$ is $u_{\theta, \phi} = -(\tan \theta, \sec \theta \cot \phi)$. Without loss of generality we may assume that the solution $u_{\theta, \phi} \in U$ for all θ and ϕ .

Now applying Lemma 3.4 and omitting the higher order decay term, for $\rho \rightarrow \infty$, we have

$$\widehat{\mathcal{T}}(\rho, \theta, \phi) = \frac{1}{\rho} a(u_{\theta, \phi}) e^{-2\pi i \rho H(u_{\theta, \phi})} \quad (3.17)$$

and, thus, we have

$$\|\mathcal{T}_j\|_2^2 \simeq \int_{2^{2j}}^{2^{2j+2}} \int_0^{2\pi} \int_0^\pi \rho^{-2} a^2(u_{\theta, \phi}) \rho^2 \sin \phi \, d\theta \, d\phi \, d\rho \simeq 2^{2j}.$$

It follows that $\|\mathcal{T}_j\|_2 \simeq 2^j$. $\square$

We also need the following lemma

Lemma 3.5. *Let $k = (k_1, k_2, k_3) \in \mathbb{Z}^3$, $k' = (k'_1, k'_2, k'_3) \in \mathbb{Z}^3$ and $\ell = (\ell_1, \ell_2) \in \mathbb{Z}^2$ with $|\ell_1| \leq 2^j$, $|\ell_2| \leq 2^j$, let $Q_{k, k'} = \{\ell = (\ell_1, \ell_2) : |\frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_2 + k_3 + k'_1 + 2^{-j}\ell_1 k'_2 + 2^{-j}\ell_2 k'_3| \leq 2^{\frac{1}{2}j}\}$. Then the cardinality of the set $Q_{k, k'}$ satisfies the inequality $\#(Q_{k, k'}) \leq C 2^{\frac{1}{2}j}$, with C independent of k and k' .*

Proof: We have that

$$\begin{aligned} & \frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_2 + k_3 + k'_1 + 2^{-j}\ell_1 k'_2 + 2^{-j}\ell_2 k'_3 \\ &= \frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_2 + k_3 + k'_1 + 2^{-j}k'_2(1 + \ell_1) - 2^{-j}k'_2 + 2^{-j}k'_3(1 + \ell_2) - 2^{-j}k'_3 \\ &= \frac{1}{2}(2^{-j}k'_2 + 1 + \ell_1)^2 + \frac{1}{2}(2^{-j}k'_3 + 1 + \ell_2)^2 + k_2 + k_3 + k'_1 - 2^{-j}k'_2 - 2^{-j}k'_3. \end{aligned}$$

Let $\alpha = 2(k_2 + k_3 + k'_1 - 2^{-j}k'_2 - 2^{-j}k'_3)$. To prove the Lemma, we consider two separate cases dependent on the value of α .

Case 1: $\alpha > -2^{\frac{1}{2}j+1}$. In this case, we have $Q_{k, k'} \subset \{\ell = (\ell_1, \ell_2) : (2^{-j}k'_2 + 1 + \ell_1)^2 + (2^{-j}k'_3 + 1 + \ell_2)^2 \leq 2^{\frac{1}{2}j+2}\}$. It follows that there is constant C independent of k and k' such that

$$\begin{aligned} \#(Q_{k, k'}) &\leq C \int_{(x+1+2^{-j}k'_2)^2 + (y+1+2^{-j}k'_3)^2 \leq 2^{\frac{1}{2}j+4}} dx \, dy \\ &= C 2\pi \int_0^{2^{\frac{1}{4}j+2}} r \, dr \\ &\leq C 2^{\frac{1}{2}j}. \end{aligned}$$

Case 2: $\alpha \leq -2^{\frac{1}{2}j+1}$. In this case, we have

$$Q_{k, k'} = \{\ell = (\ell_1, \ell_2) : -\alpha - 2^{\frac{1}{2}j+1} \leq (2^{-j}k'_2 + 1 + \ell_1)^2 + (2^{-j}k'_3 + 1 + \ell_2)^2 \leq -\alpha + 2^{\frac{1}{2}j+1}\}.$$

It follows that there is constant C independent of k and k' such that

$$\begin{aligned} \#(Q_{k, k'}) &\leq C \int_{-\alpha - 2^{\frac{1}{2}j+1} \leq (x+1+2^{-j}k'_2)^2 + (y+1+2^{-j}k'_3)^2 \leq -\alpha + 2^{\frac{1}{2}j+1}} dx \, dy \\ &= C 2\pi \int_{\sqrt{-\alpha - 2^{\frac{1}{2}j+1}}}^{\sqrt{-\alpha + 2^{\frac{1}{2}j+1}}} r \, dr \\ &\leq C 2^{\frac{1}{2}j}. \end{aligned}$$

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

We can now complete the proof of our main theorem.

Proof of Theorem 2.1. From (3.17), we have that

$$\widehat{\mathcal{T}}_j(\rho, \theta, \phi) = \frac{1}{\rho} W(2^{-2j} \rho \Theta(\theta, \phi)) a(u_{\theta, \phi}) e^{-2\pi i \rho H(u_{\theta, \phi})}.$$

Let $\beta_{j, \ell, k} = \langle \mathcal{T}_j, \psi_{j, \ell, k}^{(1)} \rangle$. Using the expression of $\widehat{\mathcal{T}}_j$ above and of $\widehat{\psi}^{(1)}$ in (3.16), we have

$$\begin{aligned} \beta_{j, \ell, k} &= \int_{\mathbb{R}^3} \overline{\widehat{\psi}_{j, \ell, k}^{(1)}(\xi)} \widehat{\mathcal{T}}_j(\xi) d\xi \\ &= 2^{-2j} \int_{\mathbb{R}^3} W^2(2^{-2j} \xi) v\left(2^j \frac{\xi_2}{\xi_1} - \ell_1\right) v\left(2^j \frac{\xi_3}{\xi_1} - \ell_2\right) e^{2\pi i \xi \cdot A_{(1)}^{-j} B_{(1)}^{[-\ell_1, -\ell_2]} k} \widehat{\mathcal{T}}(\xi) d\xi \\ &= 2^{-2j} \int_{\mathbb{R}^3} W^2(2^{-2j} \xi) v\left(2^j \frac{\xi_2}{\xi_1} - \ell_1\right) v\left(2^j \frac{\xi_3}{\xi_1} - \ell_2\right) e^{2\pi i \xi \cdot (2^{-2j}(k_1 - \ell_1 k_2 - \ell_2 k_3), 2^{-j} k_2, 2^{-j} k_3)} \widehat{\mathcal{T}}(\xi) d\xi \\ &= 2^{-2j} \int_0^\infty \int_0^\pi \int_0^{2\pi} W^2(2^{-2j} \rho \Theta(\theta, \phi)) v\left(2^j \tan \theta - \ell_1\right) v\left(2^j \sec \theta \cot \phi - \ell_2\right) \times \\ &\times e^{2\pi i \rho \Theta(\theta, \phi) \cdot (2^{-2j}(k_1 - \ell_1 k_2 - \ell_2 k_3), 2^{-j} k_2, 2^{-j} k_3)} e^{-2\pi i \rho H(u_{\theta, \phi})} \rho a(u_{\theta, \phi}) \sin \phi d\phi d\theta d\rho \\ &= 2^{-2j} \int_0^\infty \int_0^\pi \int_0^{2\pi} W^2(2^{-2j} \rho \Theta(\theta, \phi)) v\left(2^j \tan \theta - \ell_1\right) v\left(2^j \sec \theta \cot \phi - \ell_2\right) \times \\ &\times e^{2\pi i \rho \Theta(\theta, \phi) \cdot \left((2^{-2j}(k_1 - \ell_1 k_2 - \ell_2 k_3), 2^{-j} k_2, 2^{-j} k_3) - \left(\frac{1}{2}(\tan \theta)^2 + \frac{1}{2}(\sec \theta \cot \phi)^2, -\tan \theta, -\sec \theta \cot \phi\right)\right)} \rho a(u_{\theta, \phi}) \sin \phi d\phi d\theta d\rho \\ &= 2^{2j} \int_0^\infty \int_0^\pi \int_0^{2\pi} W^2(\rho \Theta(\theta, \phi)) v\left(2^j \tan \theta - \ell_1\right) v\left(2^j \sec \theta \cot \phi - \ell_2\right) \times \\ &\times e^{2\pi i \rho \Theta(\theta, \phi) \cdot (k_1 - \ell_1 k_2 - \ell_2 k_3 - \frac{1}{2}(\tan \theta)^2 - \frac{1}{2}(\sec \theta \cot \phi)^2, 2^j k_2 + 2^{2j} \tan \theta, 2^j k_3 + 2^{2j} \sec \theta \cot \phi)} \rho a(u_{\theta, \phi}) \sin \phi d\phi d\theta d\rho. \end{aligned} \tag{3.18}$$

Let L be the differential operator:

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

From (3.18), we have that for any $N \in \mathbb{N}$

$$\begin{aligned} \beta_{j, \ell, k} &= 2^{-2j} \int_{\mathbb{R}^3} L^N \left(W^2(2^{-2j} \xi) \widehat{\mathcal{T}}(\xi) v\left(2^j \frac{\xi_2}{\xi_1} - \ell_1\right) v\left(2^j \frac{\xi_3}{\xi_1} - \ell_2\right) \right) \times \\ &\times L^{-N} \left(e^{2\pi i \xi \cdot (2^{-2j}(k_1 - \ell_1 k_2 - \ell_2 k_3), 2^{-j} k_2, 2^{-j} k_3)} \right) d\xi. \end{aligned}$$

Hence, using the fact that W , v and $\mathcal{T}$ are continuously differentiable N times, for any N , a direct computation shows that there exists a constant C_N such that

$$\begin{aligned} |\beta_{j, \ell, k}| &\leq C_N \int_U \left(1 + (k_1 - \ell_1 k_2 - \ell_2 k_3 - 2^{2j}(\frac{1}{2}u_1^2 + \frac{1}{2}u_2^2))^2 \right)^{-N} \\ &\times (1 + (k_2 - 2^j u_1)^2)^{-N} (1 + (k_3 - 2^j u_2)^2)^{-N} du. \end{aligned}$$

Let

$$J = \{k = (k_1, k_2, k_3), |k_2| > 2 \cdot 2^j \text{ or } |k_3| > 2 \cdot 2^j \text{ or } |k_1| > 3 \cdot 2^{2j}\}.$$

Then $|k_2 - 2^j u_1| > 2^j$ or $|k_3 - 2^j u_2| > 2^j$ or $|k_1 - k_2 \ell_1 - k_3 \ell_2 - 2^{2j}(\frac{1}{2}u_1^2 + \frac{1}{2}u_2^2)| > 2^{2j}$ for $k \in J$ and for all $(u_1, u_2) \in U$. It follows that

$$\sum_{k \in J} |\beta_{j, \ell, k}| \leq C_N 2^{-(2N-1)j}. \tag{3.19}$$

Also we have

$$\begin{aligned}
& \beta_{j,\ell,k} \\
&= 2^{2j} \int_0^\infty \int_0^{2\pi} \int_0^\pi W^2(\rho\Theta(\theta, \phi)) v(2^j \tan \theta - \ell_1) v(2^j \sec \theta \cot \phi - \ell_2) \\
&\times e^{2\pi i \rho \Theta(\theta, \phi) \cdot ((k_1 - \ell_1 k_2 - \ell_2 k_3 - \frac{2^{2j}}{2} \tan^2 \theta - \frac{2^{2j}}{2} (\sec \theta \cot \phi)^2, 2^j k_2 + 2^{2j} \tan \theta, 2^j k_3 + 2^{2j} \sec \theta \cot \phi))} \rho a(u_{\theta, \phi}) \sin \phi d\phi d\theta d\rho \\
&= P_1(j, \ell, k) + P_2(j, \ell, k),
\end{aligned}$$

where $P_1(j, \ell, k)$ is obtained by restricting the integration with respect to the variable θ to the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $P_2(j, \ell, k)$ by restricting the integration with respect to the variable θ to the interval $[\frac{\pi}{2}, \frac{3\pi}{2}]$.

For $P_1(j, \ell, k)$, we use the change of variables $t_1 = 2^j \tan \theta - \ell_1$, $t_2 = 2^j \sec \theta \cot \phi - \ell_2$ so that $\theta(t_1) = \tan^{-1}(2^{-j}(t_1 + \ell_1))$, $\phi(t_1, t_2) = 2^{-j} \cos(\theta(t_1))(t_2 + \ell_2)$ and $d\phi d\theta = 2^{-2j} \cos^3(\theta(t_1)) \sin^2(\phi(t_1, t_2)) dt_2 dt_1$. It follows that

$$\begin{aligned}
P_1(j, \ell, k) &= \int_0^\infty \int_{-1}^1 \int_{-1}^1 W^2(\rho\Theta(\theta(t_1), \phi(t_1, t_2))) v(t_1) v(t_2) \rho a(u_{\theta(t_1), \phi(t_1, t_2)}) \cos^3(\theta(t_1)) \sin^3(\phi(t_1, t_2)) \\
&\times e^{2\pi i \rho \cos(\theta(t_1)) \sin(\phi(t_1, t_2)) ((k_1 - \ell_1 k_2 - \ell_2 k_3 + \frac{1}{2}(t_1 + \ell_1)^2 + \frac{1}{2}(t_2 + \ell_2)^2 + [(t_1 + \ell_1)k_2 + (t_2 + \ell_2)k_3])} dt_2 dt_1 d\rho \\
&= \int_0^\infty \int_{-1}^1 \int_{-1}^1 W^2(\rho\Theta(\theta(t_1), \phi(t_1, t_2))) v(t_1) v(t_2) \rho a(u_{\theta(t_1), \phi(t_1, t_2)}) \cos^3(\theta(t_1)) \sin^3(\phi(t_1, t_2)) \\
&\times e^{2\pi i \rho \cos(\theta(t_1)) \sin(\phi(t_1, t_2)) (\frac{1}{2}(t_1 + \ell_1)^2 + \frac{1}{2}(t_2 + \ell_2)^2 + k_1 + t_1 k_2 + t_2 k_3)} dt_2 dt_1 d\rho.
\end{aligned}$$

For $P_2(j, \ell, k)$, we first let $\theta' = \theta - \pi$ which implies that $\cos \theta = -\cos \theta'$, $\sin \theta = -\sin \theta'$, $\tan \theta = \tan \theta'$. Next, using the change of variables $t_1 = 2^j \tan \theta' - \ell_1$, $t_2 = -2^j \sec \theta' \cot \phi - \ell_2$ we have

$$\begin{aligned}
P_2(j, \ell, k) &= \int_0^\infty \int_{-1}^1 \int_{-1}^1 W^2[\rho(-\cos(\theta'(t_1)) \sin(\phi(t_1, t_2), -\sin(\theta'(t_1)) \sin(\phi(t_1, t_2), \cos(\phi(t_1, t_2))) v(t_1) v(t_2) \\
&\times \rho a(u_{\theta'(t_1), \phi(t_1, t_2)}) \cos^3(\theta'(t_1)) \sin^3(\phi(t_1, t_2)) \\
&\times e^{-2\pi i \rho \cos(\theta(t_1)) \sin(\phi(t_1, t_2)) (\frac{1}{2}(t_1 + \ell_1)^2 + \frac{1}{2}(t_2 + \ell_2)^2 + k_1 + t_1 k_2 + t_2 k_3)} dt_2 dt_1 d\rho.
\end{aligned}$$

Next, we define the set $S_{2,j}$ as $S_{2,j} = \bigcup_{i=0}^4 F_{i,j}$, where

$$\begin{aligned}
F_{0,j} &= \{(\ell, k) : |\frac{1}{2}(\ell_1^2 + \ell_2)^2 + k_1| \leq 2^{\frac{1}{4}j}, \} \\
F_{1,j} &= \{(\ell, k) : |\frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_1 + k_2 + k_3| \leq 2^{\frac{1}{4}j}, \} \\
F_{2,j} &= \{(\ell, k) : |\frac{1}{2}(-1 + \ell_1)^2 + \frac{1}{2}(-1 + \ell_2)^2 + k_1 - k_2 - k_3| \leq 2^{\frac{1}{4}j}, \} \\
F_{3,j} &= \{(\ell, k) : |\frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(-1 + \ell_2)^2 + k_1 + k_2 - k_3| \leq 2^{\frac{1}{4}j}, \} \\
F_{4,j} &= \{(\ell, k) : |\frac{1}{2}(-1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_1 - k_2 + k_3| \leq 2^{\frac{1}{4}j}, \}.
\end{aligned}$$

Let $g(t_1, t_2) = \frac{1}{2}(t_1 + \ell_1)^2 + \frac{1}{2}(t_2 + \ell_2)^2 + k_1 + t_1 k_2 + t_2 k_3$. Then $\frac{\partial g}{\partial t_1}(t) = \ell_1 + t_1 + k_2$ and $\frac{\partial g}{\partial t_2}(t) = \ell_2 + t_2 + k_3$. If $\ell_1 + k_2 \geq 1$, then $\frac{\partial g}{\partial t_1}(t) \geq 0$ on $[-1, 1]$, while if $\ell_1 + k_2 \leq -1$, then $\frac{\partial g}{\partial t_1}(t) \leq 0$ on $[-1, 1]$. It follows that for $\ell_1 \neq -k_2$ and for each fixed t_2 , the function $g(t_1, t_2)$ is monotone for $t_1 \in [-1, 1]$. Similarly for $\ell_2 \neq -k_3$ and for each fixed t_1 , the function $g(t_1, t_2)$ is monotone for $t_2 \in [-1, 1]$. When $\ell_1 = -k_2$ and $\ell_2 = -k_3$, the function $g(t_1, t_2)$ has only the critical point $(0, 0)$. Therefore the function $g(t)$ can attain its extreme values only at the points $(1, 1)$, $(-1, -1)$, $(1, -1)$, $(-1, 1)$, $(0, 0)$.

Thus, since $S_{2,j}^c = \bigcap_{p=1}^5 F_{2,j}^c$, we see that $|g(t)| \geq 2^{\frac{1}{4}j}$ for all $t \in [-1, 1]^2$ and all $(\ell, k) \in S_{2,j}^c$. Due to the assumptions on support of the functions W and v , we have that $\frac{1}{16} < \rho < \frac{1}{2}$, $|\theta| \leq \frac{\pi}{4}$, $|\phi - \frac{\pi}{2}| \leq \frac{\pi}{4}$

so that $\rho \cos(\theta(t_1)) \sin(\phi(t_1, t_2)) \geq c > 0$ for $\rho \in \text{supp } W$ and for $|\theta| \leq \frac{\pi}{4}$, $|\phi - \frac{\pi}{2}| \leq \frac{\pi}{4}$. It follows that if, $(\ell, k) \in S_{2,j}^c$, then integration by parts N times on the variable ρ of the integral $P_1(j, \ell, k)$ yields that there is a constant C_N independent of ℓ, k such that $|P_1(j, \ell, k)| \leq C_N 2^{-\frac{1}{4}Nj}$. Similarly we have the estimate $|P_2(j, \ell, k)| \leq C_N 2^{-\frac{1}{4}Nj}$. This implies that there is a constant C_N independent of ℓ, k such that

$$|\beta_{j,\ell,k}| \leq 2C_N 2^{-\frac{1}{4}Nj}.$$

Combining (3.19) and the above estimate, we have

$$\begin{aligned} \delta_{2,j} &= \sum_{(\ell,k) \in S_{2,j}^c} |\beta_{j,\ell,k}| \\ &= \sum_{(\ell,k) \in S_{2,j}^c \cap J} |\beta_{j,\ell,k}| + \sum_{(\ell,k) \in S_{2,j}^c \cap J^c} |\beta_{j,\ell,k}| \\ &\leq C_N \left(2^{-(2N-1)j} + 24 \cdot 2^{-\frac{1}{4}Nj} 2^{4j} \right), \end{aligned}$$

which is valid for any $N \in \mathbb{N}$. By taking $N = 13$, we have that $\delta_{2,j} = o(2^j)$ as $j \rightarrow \infty$.

As we pointed out above, to complete the proof of Theorem 2.1 it remains to show that $\mu_c(S_{2,j}, \Psi; \Phi) = \max_{k' \in \mathbb{Z}^3} \sum_{(\ell,k) \in S_{2,j}} |\langle \phi_{j,k'}, \psi_{j,\ell,k} \rangle| \rightarrow 0$, as $j \rightarrow \infty$. By the construction of $S_{2,j}$, it follows that the proof of Theorem 2.1 is completed once we show the following lemma.

Lemma 3.6. *Let Φ and Ψ be defined as in Theorem 2.1. For any $p = 1, 2, 3, 4, 5$, there exists a constant $C > 0$ independent of j, ℓ, k, k' such that*

$$\sum_{(\ell,k) \in F_{p,j}} |\langle \psi_{j,\ell,k}, \phi_{j,k'} \rangle| \leq C 2^{-\frac{1}{2}j}.$$

Proof. We will only verify the above inequality for $F_{1,j}$ as all other cases can be handled using a very similar argument.

For fixed k_2, k_3 and ℓ , let $K_{k_2, k_3, \ell} = \{k_1, (\ell, k) \in F_{1,j}\}$. Since $|\ell_1| \leq 2^j$, $|\ell_2| \leq 2^j$, for $\ell = (\ell_1, \ell_2)$, we have that $\|\ell\| \leq 2^{j+\frac{1}{2}}$. We need to show that there is a constant C independent of j, ℓ, k, k' such that

$$\sum_{\|\ell\| \leq 2^{j+\frac{1}{2}}} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \sum_{k_1 \in K_{k_2, k_3, \ell}} |\langle \psi_{j,\ell_1, \ell_2, k}, \phi_{j,k'} \rangle| \leq C 2^{-\frac{1}{4}j}.$$

Let L_1 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

$$\alpha = \alpha(j, \ell, k) = B_{(1)}^\ell A_{(1)}^j (2^{-2j} k') = (k'_1 + 2^{-j} \ell_1 k'_2 + 2^{-j} \ell_2 k'_3, 2^{-j} k'_2, 2^{-j} k'_3),$$

and $\alpha = (\alpha_1, \alpha_2, \alpha_3)$. By direct calculation, for any positive integer N we have that

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

It follows that

$$\begin{aligned}
& \sum_{\|\ell\| \leq 2^{j+\frac{1}{2}}} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \sum_{k_1 \in K_{k_2, k_3, \ell}} |\langle \psi_{j, \ell_1, \ell_2, k}, \phi_{j, k'} \rangle| \\
& \leq C 2^{-j} \sum_{\|\ell\| \leq 2^{j+\frac{1}{2}}} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \sum_{k_1 \in K_{k_2, k_3, \ell}} (1 + (k_1 - \alpha_1)^2)^{-N} (1 + (k_2 - \alpha_2)^2)^{-N} (1 + (k_3 - \alpha_3)^2)^{-N} \\
& \leq C 2^{-j} \sum_{\|\ell\| \leq 2^{j+\frac{1}{2}}} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \sum_{k_1 \in K_{k_2, k_3, \ell}} (1 + (k_1 - k'_1 - 2^{-j} \ell_1 k'_2 - 2^{-j} \ell_2 k'_3)^2)^{-N} \\
& \quad \times (1 + (k_2 - 2^{-j} k'_2)^2)^{-N} (1 + (k_3 - 2^{-j} k'_3)^2)^{-N}.
\end{aligned}$$

As in Lemma 3.5, we consider the sets

$$\begin{aligned}
Q_{k, k'} &= \{ \ell = (\ell_1, \ell_2) : \left| \frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_2 + k_3 + \alpha_1 \right| \leq 2^{\frac{1}{2}j} \} \\
&= \{ \ell = (\ell_1, \ell_2) : \left| \frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_2 + k_3 + k'_1 + 2^{-j} \ell_1 k'_2 + 2^{-j} \ell_2 k'_3 \right| \leq 2^{\frac{1}{2}j} \}
\end{aligned}$$

By Lemma 3.5, we have that there is a constant C independent of k and k' such that $\#(Q_{k, k'}) \leq C 2^{\frac{1}{2}j}$. It follows that

$$\begin{aligned}
& \sum_{\ell \in Q_{k, k'}} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \sum_{k_1 \in K_{k_2, k_3, \ell}} |\langle \psi_{j, \ell_1, \ell_2, k}, \phi_{j, k'} \rangle| \\
& \leq C_N \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \sum_{\ell \in Q_{k_2, k_3, \alpha_1}} \sum_{k_1 \in K_{k_2, k_3, \ell}} [1 + (k_1 - \alpha_1)^2]^{-N} [1 + (k_2 - \alpha_2)^2]^{-N} [1 + (k_3 - \alpha_3)^2]^{-N} \\
& \leq C_N \sum_{\ell \in Q_{k, k'}} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \left(\sum_{k_1 \in \mathbb{Z}} (1 + (k_1 - k'_1 - 2^{-j} \ell_1 k'_2 - 2^{-j} \ell_2 k'_3)^2)^{-N} \right) (1 + (k_2 - 2^{-j} k'_2)^2)^{-N} \\
& \quad \times (1 + (k_3 - 2^{-j} k'_3)^2)^{-N} \\
& \leq C_N 2^{\frac{1}{2}j}.
\end{aligned}$$

For $\ell \in Q_{k, k'}^c$, we have $|\frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_2 + k_3 + \alpha_1| \geq 2^{\frac{1}{2}j}$. Since k_1 satisfies the inequality $|k_1 + \frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_2 + k_3| \leq 2^{\frac{1}{4}j}$, for all $j \geq 4$, we have

$$\begin{aligned}
|k_1 - \alpha_1| &= |k_1 + \frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_2 + k_3 - (\frac{1}{2}(1 + \ell_1)^2 + \frac{1}{2}(1 + \ell_2)^2 + k_2 + k_3 + \alpha_1)| \\
&\geq 2^{\frac{1}{2}j} - 2^{\frac{1}{4}j} \\
&\geq \frac{1}{2} 2^{\frac{1}{2}j}.
\end{aligned}$$

It follows that

$$\begin{aligned}
& \sum_{\ell \in Q_{k, k'}^c} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \sum_{k_1 \in K_{k_2, k_3, \ell}} |\langle \psi_{j, \ell_1, \ell_2, k}, \phi_{j, k'} \rangle| \\
& \leq C_N \sum_{\ell \in Q_{k, k'}^c} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \sum_{k_1 \in K_{k_2, k_3, \ell}} [1 + (k_1 - \alpha_1)^2]^{-N} [1 + (k_2 - \alpha_2)^2]^{-N} [1 + (k_3 - \alpha_3)^2]^{-N} \\
& \leq C_N \sum_{|\ell| \leq 2^{j+\frac{1}{2}}} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} \sum_{|k_1 - \alpha_1| \geq \frac{1}{2} 2^{\frac{1}{2}j}} [1 + (k_1 - \alpha_1)^2]^{-N} [1 + (k_2 - \alpha_2)^2]^{-N} [1 + (k_3 - \alpha_3)^2]^{-N} \\
& \leq C_N 2^{2j} 2^{-(2N-1)\frac{1}{2}j} \\
& \leq C 2^{-\frac{1}{2}j},
\end{aligned}$$

where, in the last inequality, we took $N = 3$. $\square$

Combining the above estimates, we have proved

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

This completes the proof of Theorem 2.1. $\square$

References

- [1] E. J. Candès and D. L. Donoho. Ridgelets: the key to high dimensional intermittency? *Philosophical Transactions of the Royal Society of London A*, 357:2495–2509, 1999.
- [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.*, 56:219–266, 2004.
- [3] 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.
- [4] S. S. Chen, D. L. Donoho and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43:129–159, 2001.
- [5] 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.
- [6] D. L. Donoho. For most large underdetermined systems of linear equations the minimal ℓ^1 -norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59:97–829, 2006.
- [7] D. Donoho and G. Kutyniok. Microlocal analysis of the geometric separation problem. *Comm. Pure Appl. Math.*, 66:1–47, 2013.
- [8] 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.
- [9] R. Gribonval and E. Bacry. Harmonic decomposition of audio signals with matching pursuit. *IEEE Trans. Signal Proc.*, 51:1001–111, 2003.
- [10] 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.
- [11] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM J Math. Anal.*, 39:298–318, 2007.
- [12] K. Guo and D. Labate. Characterization and analysis of edges using the continuous shearlet transform. *SIAM J. Imaging Sci.*, 2:959–986, 2009.
- [13] 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.
- [14] 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.
- [15] K. Guo and D. Labate. Characterization of Piecewise Smooth Surfaces using the 3D Continuous Shearlet Transform. *J. Fourier Anal. Appl.*, 18:488–516, 2012.

- [16] K. Guo and D. Labate. The construction of smooth Parseval frames of shearlets. *Math. Model. Nat. Phenom.*, 8:82-105, 2013.
- [17] K. Guo, and D. Labate. Geometric separation of singularities using combined multiscale dictionaries. *J. Fourier Anal. Appl.*, 21:667–693, 2015.
- [18] K. Guo, and D. Labate. Microlocal analysis of edge flatness through directional multiscale representations. *Advances in Computational Mathematics* (in press) (doi: 10.1007/s10444-016-9486-8), 2017.
- [19] K. Guo, D. Labate and W. Lim. Edge analysis and identification using the continuous shearlet transform. *Appl. Comput. Harmon. Anal.*, 27:24–46, 2009.
- [20] E. Hernandez, and G. Weiss. *A first course on wavelets*. CRC Press, Boca Raton, FL, 1996.
- [21] G. Kutyniok, Geometric Separation by Single-Pass Alternating Thresholding. *Appl. Comput. Harmon. Anal.*, 36:23–50, 2014.
- [22] E. King, G. Kutyniok, G. and X. Zhuang. Analysis of inpainting via clustered sparsity and microlocal analysis. *Journal of Mathematical Imaging and Vision*, 48(2): 205-234, 2014.
- [23] G. Kutyniok and D. Labate. Resolution of the wavefront set using continuous shearlets. *Trans. Amer. Math. Soc.*, 361:2719–2754, 2009.
- [24] 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.
- [25] G. Kutyniok, P. Petersen. Classification of edges using compactly supported shearlets. *Appl. Comput. Harmon. Anal.*, 42:245-293, 2017.
- [26] G. Kutyniok, M. Shahram and X. Zhuang. ShearLab: A Rational Design of a Digital Parabolic Scaling. *SIAM J. Imaging Sciences*, 5:1291–1332, 2012.
- [27] D. Labate, W. Lim, G. Kutyniok and G. Weiss. Sparse multidimensional representation using shearlets. *SPIE Proc.*, 5914:254–262, 2005.
- [28] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Proc.*, 41:3397–3415, 1993.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] E. M. Stein. *Harmonic Analysis: real-variable methods, orthogonality, and oscillatory integrals*. Princeton University Press, Princeton, 1993
- [33] J.-L. Starck, M. Nguyen and F. Murtagh, Wavelets and curvelets for image deconvolution: A combined approach. *Signal Process.*, 83:2279–2283, 2003.
- [34] R. S. Strichartz. *A guide to distribution theory and Fourier transforms*. World Scientific Pub., 2003.
- [35] G. Teschke. Multi-frame representations in linear inverse problems with mixed multi-constraints. *Appl. Comput. Harmon. Anal.*, 22:43–60, 2007.