Compactly Supported Shearlets are Optimally Sparse

Gitta Kutyniok[a,1,*], Wang-Q Lim[a,1]

[a]*Institute of Mathematics, University of Osnabrück, 49069 Osnabrück, Germany*

Abstract

Cartoon-like images, i.e., C^2 functions which are smooth apart from a C^2 discontinuity curve, have by now become a standard model for measuring sparse (non-linear) approximation properties of directional representation systems. It was already shown that curvelets, contourlets, as well as shearlets do exhibit (almost) optimally sparse approximations within this model. However, all those results are only applicable to band-limited generators, whereas, in particular, spatially compactly supported generators are of uttermost importance for applications.

In this paper, we now present the first complete proof of (almost) optimally sparse approximations of cartoon-like images by using a particular class of directional representation systems, which indeed consists of compactly supported elements. This class will be chosen as a subset of shearlet frames – not necessarily required to be tight – with shearlet generators having compact support and satisfying some weak moment conditions.

Keywords: Curvilinear discontinuities, edges, nonlinear approximation, optimal sparsity, shearlets, thresholding, wavelets

1. Introduction

In computer vision, edges were detected as those features governing an image while separating smooth regions in between. About 10 years ago,

---

mathematicians started to design models of images incorporating those findings aiming at designing representation systems which – in such a model – are capable of resolving edges in an optimally sparse way. However, customarily, at that time an image was viewed as an element of a compact subset of L_p characterized by a given Besov regularity with the Kolmogorov entropy of such sets identifying lower bounds for the distortion rates of encoding-decoding pairs in this model. Although wavelets could be shown to behave optimally [2] as an encoding methodology, Besov models are clearly deficient since edges are not adequately captured. This initiated the introduction of a different model, called cartoon-like model (see [5, 19, 1]), which revealed the suboptimal treatment of edges by wavelets.

The introduction of tight curvelet frames in 2004 by Candés and Donoho [1], which provably provide (almost) optimally sparse approximations within such a cartoon-like model can be considered a milestone in applied harmonic analysis. One year later, contourlets were introduced by Do and Vetterli [4] which similarly derived (almost) optimal approximation rates. In the same year, *shearlets* were developed by Labate, Weiss, and the authors in [17] as the first directional representation system with allows a unified treatment of the continuum and digital world similar to wavelets, while also providing (almost) optimally sparse approximations within such a cartoon-like model [11].

In most applications, spatial localization of the analyzing elements of an encoding system is of uttermost importance both for a precise detection of geometric features as well as for a fast decomposition algorithm. However, none of the previously mentioned results cover this situation. In fact, the proofs which were provided do by no means extend to this crucial setting.

In this paper, we now present the first complete proof of (almost) optimally sparse approximations of cartoon-like images by using a particular class of directional representation systems, which indeed consist of compactly supported elements. This class will be chosen as a subset of shearlet frames – not necessarily required to be tight – with shearlet generators having compact support and satisfying some weak moment conditions. Interestingly, our proof is very different from all previous ones caused by the extensive exploration of the compact support of the shearlet generators and the lack of directional vanishing moments.

1.1. A Suitable Model for Images: Cartoon-Like Images

Intuitively, cartoons are smooth image parts separated from other areas by an edge. After a series of initial models [5, 19], the first complete model of cartoons has been introduced in [1], and this is what we intend to use also here. The basic idea is to choose a closed boundary curve and then fill the interior and exterior part with C^2 functions (see Figure 1).
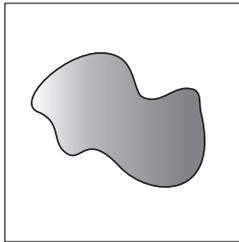

Figure 1: Example of a cartoon-like image.

Let us now be more precise, and introduce $STAR^2(\nu)$, a class of indicator functions of sets B with C^2 boundaries ∂B and curvature bounded by ν, as well as $\mathcal{E}^2(\nu)$, a class of cartoon-like images. For this, in polar coordinates, we let $\rho(\theta) \to [0, 1]$ be a radius function and define the set B by

$$B = \{x \in \mathbb{R}^2 : |x| \leq \rho(\theta), x = (|x|, \theta) \text{ in polar coordinates}\}.$$

In particular, the boundary ∂B of B is given by the curve

$$\beta(\theta) = \begin{pmatrix} \rho(\theta)\cos(\theta) \\ \rho(\theta)\sin(\theta) \end{pmatrix}, \tag{1}$$

and the class of boundaries of interest to us are defined by

$$\sup |\rho''(\theta)| \leq \nu, \quad \rho \leq \rho_0 < 1. \tag{2}$$

The following definition now introduces the notions $STAR^2(\nu)$ and $\mathcal{E}^2(\nu)$ from [1].

Definition 1.1. *For $\nu > 0$, the set $STAR^2(\nu)$ is defined to be the set of all $B \subset [0, 1]^2$ such that B is a translate of a set obeying (1) and (2). Further, $\mathcal{E}^2(\nu)$ denotes the set of functions f on $\mathbb{R}^2$ with compact support in $[0, 1]^2$ of the form*

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

where $f_0, f_1 \in C^2(\mathbb{R}^2)$ with compact support in $[0, 1]^2$, $B \in STAR^2(\nu)$, and $\|f\|_{C^2} = \sum_{|\alpha| \leq 2} \|D^\alpha f\|_\infty \leq 1$.

1.2. Optimal Sparsity of a Directional Representation System

The 'quality' of the performance of a (directional) representation system with respect to cartoon-like images is typically measured by taking a non-linear approximation viewpoint. More precisely, given a cartoon-like image $f \in \mathcal{E}^2(\nu)$ and a (directional) representation system $(\sigma_i)_{i \in I}$ which forms an orthonormal basis, the chosen measure is the asymptotic behavior of the best N-term (non-linear) approximation error in L^2 norm in the number of terms N, i.e.,

$$\|f - f_N\|_2^2 = \left\| f - \sum_{i \in I_N} \langle f, \sigma_i \rangle \sigma_i \right\|_2^2 \quad \text{as } N \to \infty,$$

where $(\langle f, \sigma_i \rangle)_{i \in I_N}$ are the N largest coefficients $\langle f, \sigma_i \rangle$ in magnitude. Wavelet bases exhibit the approximation rate

$$\|f - f_N\|_2^2 \leq C \cdot N^{-1} \quad \text{as } N \to \infty.$$

However, Donoho proved in [6] that the optimal rate which can be achieved under some restrictions on the representation system as well as on the selection procedure of the approximating coefficients is

$$\|f - f_N\|_2^2 \leq C \cdot N^{-2} \quad \text{as } N \to \infty.$$

It was a breakthrough in 2004, when Candés and Donoho introduced the tight curvelet frame in [1] and proved that this system indeed does satisfy

$$\|f - f_N\|_2^2 \leq C \cdot N^{-2} \cdot (\log N)^3 \quad \text{as } N \to \infty,$$

where again the approximation f_N was generated by the N largest coefficients in magnitude. Although the optimal rate is not completely achieved, the log-factor is typically considered negligible compared to the N^{-2}-factor, wherefore the term 'almost optimal' has been adopted into the language. This result is even more surprising taking into account that in case of a tight frame the approximation by the N largest coefficients in magnitude does not even always yield the *best* N-term approximation.

1.3. (Compactly Supported) Shearlet Systems

The directional representation system of *shearlets* has recently emerged and rapidly gained attention due to the fact that – in contrast to other proposed directional representation systems – shearlets provide a unified treatment of the continuum and digital world similar to wavelets. We refer to,

e.g., [9, 15] for the continuum theory, [16, 8, 18] for the digital theory, and [10, 7] for recent applications.

Shearlets are scaled according to a parabolic scaling law encoded in the *parabolic scaling matrices* A_{2^j} or $\tilde{A}_{2^j}$, $j \in \mathbb{Z}$, and exhibit directionality by parameterizing slope encoded in the *shear matrices* S_k, $k \in \mathbb{Z}$, defined by

$$A_{2^j} = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix} \qquad \text{or} \qquad \tilde{A}_{2^j} = \begin{pmatrix} 2^{j/2} & 0 \\ 0 & 2^j \end{pmatrix}$$

and

$$S_k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix},$$

respectively.

To ensure an (almost) equal treatment of the different slopes, which is evidently of significant importance for practical applications, we partition the frequency plane into the following four cones $\mathcal{C}_1 - \mathcal{C}_4$:

$$\mathcal{C}_\iota = \begin{cases} \{(\xi_1, \xi_2) \in \mathbb{R}^2 : \xi_1 \geq 1, |\xi_2/\xi_1| \leq 1\} & : & \iota = 1, \\ \{(\xi_1, \xi_2) \in \mathbb{R}^2 : \xi_2 \geq 1, |\xi_1/\xi_2| \leq 1\} & : & \iota = 2, \\ \{(\xi_1, \xi_2) \in \mathbb{R}^2 : \xi_1 \leq -1, |\xi_2/\xi_1| \leq 1\} & : & \iota = 3, \\ \{(\xi_1, \xi_2) \in \mathbb{R}^2 : \xi_2 \leq -1, |\xi_1/\xi_2| \leq 1\} & : & \iota = 4, \end{cases}$$

and a centered rectangle

$$\mathcal{R} = \{(\xi_1, \xi_2) \in \mathbb{R}^2 : \|(\xi_1, \xi_2)\|_\infty < 1\}.$$

For an illustration, we refer to Figure 2(a).

The rectangle $\mathcal{R}$ corresponds to the low frequency content of a signal and is customarily represented by translations of some scaling function. Anisotropy comes into play when encoding the high frequency content of a signal which corresponds to the cones $\mathcal{C}_1 - \mathcal{C}_4$, where the cones $\mathcal{C}_1$ and $\mathcal{C}_3$ as well as $\mathcal{C}_2$ and $\mathcal{C}_4$ are treated separately as can be seen in the following

Definition 1.2. *For some sampling constant $c > 0$, the* (cone-adapted) *shearlet system $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ generated by a scaling function $\phi \in L^2(\mathbb{R}^2)$ and shearlets $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ is defined by*

$$\mathcal{SH}(c; \phi, \psi, \tilde{\psi}) = \Phi(c; \phi) \cup \Psi(c; \psi) \cup \tilde{\Psi}(c; \tilde{\psi}),$$

where

$$\Phi(c; \phi) = \{\phi_m = \phi(\cdot - cm) : m \in \mathbb{Z}^2\},$$

Figure 2: (a) The cones $\mathcal{C}_1 - \mathcal{C}_4$ and the centered rectangle $\mathcal{R}$ in frequency domain. (b) The tiling of the frequency domain induced by a (cone-adapted) shearlet system.

$$\Psi(c; \psi) = \{\psi_{j,k,m} = 2^{3j/4} \psi(S_k A_{2^j} \cdot - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\},$$

and

$$\tilde{\Psi}(c; \tilde{\psi}) = \{\tilde{\psi}_{j,k,m} = 2^{3j/4} \tilde{\psi}(S_k^T \tilde{A}_{2^j} \cdot - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}.$$

The reader should keep in mind that although not indicated by the notation, the functions ϕ_m, $\psi_{j,k,m}$, and $\tilde{\psi}_{j,k,m}$ all depend on the sampling constant c. For the sake of brevity, we will often write ψ_λ and $\tilde{\psi}_\lambda$, where $\lambda = (j, k, m)$ index scale, shear, and position. For later use, we further let Λ_j be the indexing sets of shearlets in $\Psi(c; \psi)$ and $\tilde{\Psi}(c; \tilde{\psi})$ at scale j, respectively, i.e.,

$$\Psi(c; \psi) = \{\psi_\lambda : \lambda \in \Lambda_j, j = 0, \ldots, \infty\}$$

and

$$\tilde{\Psi}(c; \tilde{\psi}) = \{\tilde{\psi}_\lambda : \lambda \in \Lambda_j, j = 0, \ldots, \infty\}.$$

Finally, we define

$$\Lambda = \bigcup_{j=0}^{\infty} \Lambda_j.$$

The tiling of frequency domain induced by $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ is illustrated in Figure 2(b). From this illustration, the anisotropic footprints of shearlets contained in $\Psi(c; \psi)$ and $\tilde{\Psi}(c; \tilde{\psi})$ can clearly be seen. However, the reader should notice that the tiling indicated here is based on the *essential support* and not the exact support of the analyzing elements, since our focus will be on shearlet systems associated with spatially compactly supported generators.

The corresponding anisotropic footprints of shearlets *in spatial domain* are of size $2^{-j/2}$ times 2^{-j}. A beautiful intuitive extensive explanation of why it is conceivable that such a system – based on parabolic scaling – exhibits optimal sparse approximation of cartoon-like images, is provided in [1], and we would like to refer the reader to this paper. The main idea is to count the number of shearlets intersecting the discontinuity curves, which is 'small' compared to the number of such wavelets, due to their anisotropic footprints.

Certainly, we naturally ask the question when $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ does form a frame for $L^2(\mathbb{R}^2)$. The wavelet literature provides various necessary and sufficient conditions for $\Phi(c; \phi)$ to form a frame for $L^2(\{f \in L^2(\mathbb{R}^2) : \text{supp}(\hat{f}) \subseteq \mathcal{R}\})$, also when ϕ is compactly supported in spatial domain. Although not that well-studied as wavelets yet, several answers are also known for the question when $\Psi(c; \psi)$ forms a frame for

$$L^2(\{f \in L^2(\mathbb{R}^2) : \text{supp}(\hat{f}) \subseteq \mathcal{C}_1 \cup \mathcal{C}_3\}),$$

and we refer to results in [9, 14, 3, 13]. Since $\Psi(c; \phi)$ and $\tilde{\Psi}(c; \tilde{\psi})$ are linked by a simple rotation of 90°, these results immediately provide conditions for $\tilde{\Psi}(c; \tilde{\psi})$ to constitute a frame for

$$L^2(\{f \in L^2(\mathbb{R}^2) : \text{supp}(\hat{f}) \subseteq \mathcal{C}_2 \cup \mathcal{C}_4\}).$$

Very recent results in [12] even focus specifically on the case of spatially compactly supported shearlets – of uttermost importance for applications due to their superior localization. For instance, in [12], the following special class of compactly supported shearlet frames for $L^2(\mathbb{R}^2)$ was constructed: The generating shearlets ψ and $\tilde{\psi}$ were chosen separable, i.e., of the form $\psi_1(x_1) \cdot \psi_2(x_2)$ and $\psi_1(x_2) \cdot \psi_2(x_1)$, respectively, where ψ_1 is a wavelet and ψ_2 is a scaling function both associated with some carefully chosen low pass filter. Intriguingly, our main result in this paper (Theorem 1.3) proves as a corollary that this class of compactly supported shearlet frames provides (almost) optimally sparse approximations of cartoon-like images. We refer to [12] for the precise statement.

Combining those thoughts, we can attest that frame properties of the system $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ including spatially compactly supported generators are already quite well studied. We however wish to mention that there is a trade-off between *compact support* of the shearlet generators, *tightness* of the associated frame, and *separability* of the shearlet generators. The known constructions of tight shearlet frames do not use separable generators, and these

constructions can be shown to *not* be applicable to compactly supported generators. Tightness is difficult to obtain while allowing for compactly supported generators, but we can gain separability, hence fast algorithmic realizations. On the other hand, when allowing non-compactly supported generators, tightness is possible, but separability seems to be out of reach, which makes fast algorithmic realizations very difficult.

1.4. Optimally Sparse Approximation of Cartoon-Like Images by Shearlets

The concept of optimally sparse approximation of cartoon-like images of general (directional) representation systems was already discussed in Section 1.2. However, the attentive reader will have realized that only the situation of tight frames was studied whereas here we need to consider sparse approximations by arbitrary frames. Hence this situation deserves a careful commenting.

Let $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ be a shearlet frame for $L^2(\mathbb{R}^2)$, which for illustrative purposes for a moment we denote by $\mathcal{SH}(c; \phi, \psi, \tilde{\psi}) = (\sigma_i)_{i \in I}$, say. Is it well-known that a frame is associated with a canonical dual frame, which in this case we want to call $(\tilde{\sigma}_i)_{i \in I}$. Then we define the N-term approximation f_N of a cartoon-like image $f \in \mathcal{E}^2(\nu)$ by the frame $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ to be

$$f_N = \sum_{i \in I_N} \langle f, \sigma_i \rangle \tilde{\sigma}_i,$$

where $(\langle f, \sigma_i \rangle)_{i \in I_N}$ are the N largest coefficients $\langle f, \sigma_i \rangle$ in magnitude. As in the tight frame case, this procedure does not always yield the *best* N-term approximation, but surprisingly even with this 'crude' selection procedure – in the situation of spatially compactly supported generators – we can prove an (almost) optimally sparse approximation rate as our main result shows.

Theorem 1.3. *Let $c > 0$, and let $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ be compactly supported. Suppose that, in addition, for all $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$, the shearlet ψ satisfies*

(i) $|\hat{\psi}(\xi)| \leq C_1 \cdot \min(1, |\xi_1|^\alpha) \cdot \min(1, |\xi_1|^{-\gamma}) \cdot \min(1, |\xi_2|^{-\gamma})$ *and*

(ii) $\left| \frac{\partial}{\partial \xi_2} \hat{\psi}(\xi) \right| \leq |h(\xi_1)| \cdot \left(1 + \frac{|\xi_2|}{|\xi_1|} \right)^{-\gamma},$

where $\alpha > 5$, $\gamma \geq 4$, $h \in L^1(\mathbb{R})$, and C_1 is a constant, and suppose that the shearlet $\tilde{\psi}$ satisfies (i) and (ii) with the roles of ξ_1 and ξ_2 reversed. Further, suppose that $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ forms a frame for $L^2(\mathbb{R}^2)$.

Then, for any $\nu > 0$, the shearlet frame $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ provides (almost) optimally sparse approximations of functions $f \in \mathcal{E}^2(\nu)$, i.e., there exists some $C > 0$ such that

$$\|f - f_N\|_2^2 \leq C \cdot N^{-2} \cdot (\log N)^3 \qquad as\ N \to \infty,$$

where f_N is the nonlinear N-term approximation obtained by choosing the N largest shearlet coefficients of f.

Condition (i) can be interpreted as both a condition ensuring (almost) separable behavior as well as a first order moment condition along the horizontal axis, hence enforcing directional selectivity. This condition ensures that the support of shearlets in frequency domain is essentially of the form indicated in Figure 2(b). Condition (ii) (together with (i)) is a weak version of a directional vanishing moment condition[2], which is crucial for having fast decay of the shearlet coefficients when the corresponding shearlet intersects the discontinuity curve. Conditions (i) and (ii) are rather mild conditions on the generators. To compare with the optimality result for band-limited generators we wish to point out that conditions (i) and (ii) are obviously satisfied for band-limited generators.

Notice also that, intriguingly, the – the 'true' optimality destroying – log-factor has the *same* exponent as in the curvelet-, contourlet-, and shearlet-result on (almost) optimally sparse approximation.

1.5. Prior Work and Our Contribution

In 2004, Candés and Donoho [1] achieved a breakthrough when introducing tight curvelet frames, which provide (almost) optimally sparse approximations of cartoon-like images (functions in $\mathcal{E}^2(\nu)$). The main outline of their proof is to break $[0,1]^2$ into smaller cubes and then separately analyze the curvelet coefficients essentially centered in the smooth part of the model and those essentially centered on the discontinuity curve. For both sets of coefficients their weak-$\ell_{2/3}$ norm is estimated; the estimate for the 'non-smooth part' also requiring the usage of the Radon transform.

---

[2]For the precise definition of directional vanishing moments, we refer to [4].

A year later, Do and Vetterli [4] introduced contourlets and proved similar sparsity results for those. However, although their work includes contourlets with compact support, their construction is fully based on discrete filter banks so that directional selectivity is problematic. Because of this fact, infinite directional vanishing moments had to be artificially imposed in order to achieve (almost) optimal sparsity. However, this is impossible for any function with compact support to satisfy. Hence, similar to curvelets, optimal sparsity is only proven for *band-limited* contourlets.

In 2005, shearlets were introduced as the first directional representation system ensuring a unified treatment of the continuum and digital world by Labate, Weiss, and the authors in [17]. One year later, Labate and Guo proved (almost) optimally sparse approximations of cartoon-like images for the at that time customarily utilized shearlet frames [11], which are band-limited such as curvelets. The proof the authors provided follows the proof in [1] very closely step by step.

Concluding, although those pioneering studies deserve all our credit, these results are far from including the important class of directional representation systems consisting of compactly supported functions.

The main contribution of this paper is to provide the first complete proof of (almost) optimally sparse approximations of cartoon-like images using a directional representation system consisting of compactly supported functions. Our proof is indeed very different from all previous ones caused by the necessary extensive exploration of the compact support of the shearlet generators, the only similarity being the breaking of $[0, 1]^2$ into smaller cubes and the separate consideration of shearlet coefficients now being *exactly contained* – in contrast to being *essentially contained* for all other systems – in the smooth part and those which intersect the discontinuity curve. Previous results all require moment conditions along the direction of the discontinuity curve – thereby requiring vanishing moments along infinitely many directions asymptotically in scale –, which is trivially satisfied for band-limited generators. Intriguingly, a weaker version of directional vanishing moments, even only in one direction and the shearing taking care of the remaining directions, is sufficient for our analysis.

1.6. Outline

In Section 2, we present the overall structure of the proof, the results of the analysis of shearlet coefficients being contained in the smooth part and

10

those which intersect the discontinuity curve, and – based on these results – state the proof of Theorem 1.3. The proofs of the results on the behavior of shearlet coefficients in the smooth and non-smooth part are then carried out in Sections 3 and 4, respectively.

2. Architecture of the Proof of Theorem 1.3

We now detail the general structure of the proof of Theorem 1.3, starting by introducing useful notions and explaining the blocking into smaller boxes and splitting into the smooth and non-smooth part. Then the main results concerning the analysis of shearlet coefficients being entirely contained in the smooth part and those intersecting the discontinuity curve will be presented followed by the proof of Theorem 1.3 based on those.

2.1. General Organization

Let now $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ satisfy the hypotheses of Theorem 1.3, and let $f \in \mathcal{E}^2(\nu)$. Further, we let A denote the lower frame bound of $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$.

We first observe that, without loss of generality, we might assume the scaling index j to be sufficiently large, since f as well as all frame elements in the shearlet frame $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$ are compactly supported in spatial domain, hence a finite number does not contribute to the asymptotic estimate we aim for. In particular, this means that we do not need to consider frame elements from $\Phi(c; \phi)$. Also, we are allowed to restrict our analysis to shearlets $\psi_{j,k,m}$, since the frame elements $\tilde{\psi}_{j,k,m}$ can be handled in a similar way.

Our main concern will be to derive appropriate estimates for the shearlet coefficients $\{\langle f, \psi_\lambda \rangle : \lambda \in \Lambda\}$ of f. Letting $|\theta(f)|_n$ denote the nth largest shearlet coefficient $\langle f, \psi_\lambda \rangle$ in absolute value and exploring the frame property of $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$, we conclude that

$$\|f - f_N\|_2^2 \leq \frac{1}{A} \sum_{n > N} |\theta(f)|_n^2,$$

for any positive integer N. Thus, for the proof of Theorem 1.3, it suffices to show that

$$\sum_{n > N} |\theta(f)|_n^2 \leq C \cdot N^{-2} \cdot (\log N)^3 \qquad \text{as } N \to \infty. \tag{3}$$

To derive the anticipated estimate in (3), for any shearlet ψ_λ, we will study two separate cases:

- *Case 1.* The compact support of the shearlet ψ_λ does not intersect the boundary of the set B, i.e., $\mathrm{int}(\mathrm{supp}(\psi_\lambda)) \cap \partial B = \emptyset$.

- *Case 2.* The compact support of the shearlet ψ_λ does intersect the boundary of the set B, i.e., $\mathrm{int}(\mathrm{supp}(\psi_\lambda)) \cap \partial B \neq \emptyset$.

Notice that this exact distinction is only possible due to the spatial compact support of all shearlets in the shearlet frame.

In the sequel – since we are concerned with an asymptotic estimate – for simplicity we will often simply use C as a constant although it might differ for each estimate. Also all the results in the sequel are independent on the sampling constant $c > 0$, wherefore we now fix it once and for all.

2.2. The Smooth and the Non-Smooth Part of a Cartoon-Like Image

To illustrate which conditions on ψ required by Theorem 1.3 are utilized for the decay estimates of the different cases, in this section we do not make any initial assumptions on ψ.

Let us start with the smooth part, which is the easier one to handle. Dealing with this part allows us to consider some $g \in C^2(\mathbb{R}^2)$ with compact support in $[0,1]^2$ and estimate its shearlet coefficients. This is done in the following proposition. Notice that the hypothesis on ψ of the following result is implied by condition (i) in Theorem 1.3.

Proposition 2.1. *Let $g \in C^2(\mathbb{R}^2)$ with compact support in $[0,1]^2$, and let $\psi \in L^2(\mathbb{R}^2)$ be compactly supported and satisfy*

$$|\hat{\psi}(\xi)| \leq C_1 \cdot \min(1, |\xi_1|^\alpha) \cdot \min(1, |\xi_1|^{-\gamma}) \cdot \min(1, |\xi_2|^{-\gamma}) \text{ for all } \xi = (\xi_1, \xi_2) \in \mathbb{R}^2,$$

where $\gamma > 3$, $\alpha > \gamma + 2$, and C_1 is a constant. Then, there exists some $C > 0$ such that

$$\sum_{n>N} |\theta(g)|_n^2 \leq C \cdot N^{-2} \qquad \text{as } N \to \infty.$$

Thus, in this case, optimal sparsity is achieved. The proof of this proposition is given in Section 3.

Next, we turn our attention to the non-smooth part, in particular, to estimating those shearlet coefficients whose spatial support intersects the discontinuity curve. For this, we first need to introduce some new notations. For any scale $j \geq 0$ and any grid point $p \in \mathbb{Z}^2$, we let $\mathcal{Q}_{j,p}$ denote the dyadic cube defined by

$$\mathcal{Q}_{j,p} = [-2^{-j/2}, 2^{-j/2}]^2 + 2^{-j/2}p.$$

Further, let $\mathcal{Q}_j$ be the collection of those dyadic cubes $\mathcal{Q}_{j,p}$ whose interior, in the following denoted by $\mathrm{int}(\mathcal{Q}_{j,p})$, intersects ∂B, i.e.,

$$\mathcal{Q}_j = \{\mathcal{Q}_{j,p} : \mathrm{int}(\mathcal{Q}_{j,p}) \cap \partial B \neq \emptyset, p \in \mathbb{Z}^2\}.$$

Of interest to us is also the set of shearlet indices, which are associated with shearlets intersecting the discontinuity curve inside some $\mathcal{Q}_{j,p} \in \mathcal{Q}_j$, i.e., for $j \geq 0$ and $p \in \mathbb{Z}^2$ with $\mathcal{Q}_{j,p} \in \mathcal{Q}_j$, we will consider the index set

$$\Lambda_{j,p} = \{\lambda \in \Lambda_j : \mathrm{int}(\mathrm{supp}(\psi_\lambda)) \cap \mathrm{int}(\mathcal{Q}_{j,p}) \cap \partial B \neq \emptyset\}.$$

Finally, for $j \geq 0$, $p \in \mathbb{Z}^2$, and $0 < \varepsilon < 1$, we define $\Lambda_{j,p}(\varepsilon)$ to be the index set of shearlets ψ_λ, $\lambda \in \Lambda_{j,p}$, such that the magnitude of the corresponding shearlet coefficient $\langle f, \psi_\lambda \rangle$ is larger than ε and the support of ψ_λ intersects $\mathcal{Q}_{j,p}$ at the jth scale, i.e.,

$$\Lambda_{j,p}(\varepsilon) = \{\lambda \in \Lambda_{j,p} : |\langle f, \psi_\lambda \rangle| > \varepsilon\},$$

and we define $\Lambda(\varepsilon)$ to be the index set for shearlets so that $|\langle f, \psi_\lambda \rangle| > \varepsilon$ across all scales j, i.e.,

$$\Lambda(\varepsilon) = \bigcup_{j,p} \Lambda_{j,p}(\varepsilon).$$

The expert reader will have noticed that in contrast to the proofs in [1] and [11], which also split the domain into smaller scale boxes, we do not apply a weight function to obtain a smooth partition of unity. In our case, this is not necessary due to the spatial compact support of the frame elements.

As mentioned at the beginning of this section, we may assume that j is sufficiently large. Given some scale $j \geq 0$ and position $p \in \mathbb{Z}^2$ for which the associated cube $\mathcal{Q}_{j,p}$ satisfies $\mathcal{Q}_{j,p} \in \mathcal{Q}_j$. Then the set

$$\mathcal{S}_{j,p} = \bigcup_{\lambda \in \Lambda_{j,p}} \mathrm{supp}(\psi_\lambda)$$

is contained in a cubic window of size $C \cdot 2^{-j/2}$ by $C \cdot 2^{-j/2}$, hence is of asymptotically the same size as $\mathcal{Q}_{j,p}$. By smoothness assumption on the discontinuity curve ∂B, the edge curve can be parameterized by either $(x_1, E(x_1))$ or $(x_2, E(x_2))$ with $E \in C^2$ in the interior of $\mathcal{S}_{j,p}$ for sufficiently large j.[3]

Thus, we are facing the following two cases (see also Figure 3):

---

[3]In other words, a part of the edge curve ∂B contained in $\mathcal{S}_{j,p}$ can be described as a C^2 function $x_1 = E(x_2)$ (or $x_2 = E(x_1)$).

- *Case 2a.* The edge curve ∂B can be parameterized by either $(E(x_2), x_2)$ or $(x_1, E(x_1))$ with $E \in C^2$ in the interior of $\mathcal{S}_{j,p}$ such that, for any $\lambda \in \Lambda_{j,p}$, there exists some $\hat{x} = (\hat{x}_1, \hat{x}_2) \in \text{int}(\mathcal{Q}_{j,p}) \cap \text{int}(\text{supp}(\psi_\lambda)) \cap \partial B$ satisfying either $|E'(\hat{x}_2)| \leq 2$ or $|E'(\hat{x}_1)|^{-1} \leq 2$.

- *Case 2b.* The edge curve ∂B can be parameterized by either $(E(x_2), x_2)$ or $(x_1, E(x_1))$ with $E \in C^2$ in the interior of $\mathcal{S}_{j,p}$ such that, for any $\lambda \in \Lambda_{j,p}$, there exists some $\hat{x} = (\hat{x}_1, \hat{x}_2) \in \text{int}(\mathcal{Q}_{j,p}) \cap \text{int}(\text{supp}(\psi_\lambda)) \cap \partial B$ satisfying either $|E'(\hat{x}_2)| > 2$ or $|E'(\hat{x}_1)|^{-1} > 2$. Here, we identify $E'(\hat{x}_1) = 0$ with $|E'(\hat{x}_1)|^{-1} = \infty > 2$.

(a)   (b)

Figure 3: (a) A part of the curve ∂B satisfying Case 2a. (b) A part of the curve ∂B satisfying Case 2b.

For both cases, we will derive the below stated upper estimates (4) and (5) for the absolute value of the associated shearlet coefficients. The proofs of these estimates are contained in Section 4.

Proposition 2.2. *Let $\psi \in L^2(\mathbb{R}^2)$ be compactly supported, and assume that ψ satisfies conditions (i) and (ii) of Theorem 1.3. Further, let $j \geq 0$ and $p \in \mathbb{Z}^2$, and let $\lambda \in \Lambda_{j,p}$. For fixed $\hat{x} = (\hat{x}_1, \hat{x}_2) \in int(\mathcal{Q}_{j,p}) \cap int(supp(\psi_\lambda)) \cap \partial B$, let s be the slope[4] of the tangent to the edge curve ∂B at $(\hat{x}_1, \hat{x}_2)$, more precisely,*

  (i) *if ∂B is parameterized by $(E(x_2), x_2)$ with $E \in C^2$ in the interior of $\mathcal{S}_{j,p}$, then $s = E'(\hat{x}_2)$,*

  (ii) *if ∂B is parameterized by $(x_1, E(x_1))$ with $E \in C^2$ and $E'(\hat{x}_1) \neq 0$ in the interior of $\mathcal{S}_{j,p}$, then $s = (E'(\hat{x}_1))^{-1}$, and*

---

[4]Notice that here we regard the slope of the tangent to a curve $(E(x_2), x_2)$, i.e., we consider s of a curve indexed by the x_2-axis, for instance, by $x_1 = sx_2 + b$. For analyzing shearlets $\tilde{\psi}_{j,k,m}$, the roles of x_1 and x_2 would need to be reversed.

(iii) *if ∂B is parameterized by $(x_1, E(x_1))$ with $E \in C^2$ and $E'(\hat{x}_1) = 0$ in the interior of $\mathcal{S}_{j,p}$, then $s = \infty$.*

Then there exists some $C > 0$ such that

$$|\langle f, \psi_\lambda \rangle| \le C \cdot \frac{2^{-\frac{3}{4}j}}{|k + 2^{j/2}s|^3}, \qquad if\ |s| \le 3. \tag{4}$$

and

$$|\langle f, \psi_\lambda \rangle| \le C \cdot 2^{-\frac{9}{4}j}, \qquad if\ |s| > \frac{3}{2}\ or\ s = \infty, \tag{5}$$

Notice that in Case 2a, condition (i) or (ii) can occur, whereas in Case 2b, all three conditions can occur.

2.3. Proof of Theorem 1.3

Let $f \in \mathcal{E}^2(\nu)$. We first observe that, by Proposition 2.1, we can neglect those shearlet coefficients whose spatial support of the associated shearlet does not intersect the discontinuity curve.

To estimate the remaining shearlet coefficients, we need to analyze their decay properties. For this, let $j \ge 0$ be sufficiently large and let $p \in \mathbb{Z}^2$, be such that the associated cube satisfies $\mathcal{Q}_{j,p} \in \mathcal{Q}_j$. We note that all sets $\text{int}(\text{supp}(\psi_\lambda))$ with $\lambda \in \Lambda_{j,p}$ are contained in the interior of $\mathcal{S}_{j,p}$. Therefore weights as in [1] and [11] are not required here.

Letting $\varepsilon > 0$, our goal will now be to estimate first $|\Lambda_{j,p}(\varepsilon)|$ and then $|\Lambda(\varepsilon)|$. WLOG we might assume $\|\psi\|_1 \le 1$, which implies

$$|\langle f, \psi_\lambda \rangle| \le 2^{-3j/4}.$$

Hence, for estimating $|\Lambda_{j,p}(\varepsilon)|$, it is sufficient to restrict our attention to scales

$$j \le \frac{4}{3} \log_2(\varepsilon^{-1}). \tag{6}$$

We will now deal with Case 2a and Case 2b separately.

Case 2a. First, we let s be the slope of the tangent to the edge curve ∂B at $(\hat{x}_1, \hat{x}_2)$ as defined in Proposition 2.2, i.e., if ∂B is parameterized by $(E(x_2), x_2)$ in the interior of $\mathcal{S}_{j,p}$, then $s = E'(\hat{x}_2)$, and if ∂B is parameterized by $(x_1, E(x_1))$ in the interior of $\mathcal{S}_{j,p}$, then $s = E'(\hat{x}_1)^{-1}$.

By assumption of Case 2a, we have $s \in [-2, 2]$. Now observe that, for each shear index k,

$$|\{\lambda = (j, k, m) : \lambda \in \Lambda_{j,p}\}| \le C \cdot (|k + 2^{j/2}s| + 1). \tag{7}$$

15

Interestingly, this estimate is independent of the choice of the point $\hat{x} \in \text{int}(\mathcal{Q}_{j,p}) \cap \text{int}(\text{supp}(\psi_\lambda)) \cap \partial B$, which can be seen as follows: Let $\hat{x} \in \text{int}(\mathcal{Q}_{j,p}) \cap \text{int}(\text{supp}(\psi_\lambda)) \cap \partial B, \hat{x}' \in \text{int}(\mathcal{Q}_{j,p}) \cap \text{int}(\text{supp}(\psi_{\lambda'})) \cap \partial B$, and let s and s' be the associated slopes of the tangents to the edge curve E in $\hat{x}$ and $\hat{x}'$, respectively. Since $E \in C^2$,

$$|s - s'| \leq C_1 \cdot 2^{-j/2}, \tag{8}$$

and hence

$$|k + 2^{j/2}s'| \leq C \cdot (|k + 2^{j/2}s| + 1).$$

This proves that the estimate (7) remains the same asymptotically, independent of the values of s and s', and hence of $\hat{x}$ and $\hat{x}'$.

We further require the following even stronger observation: Fix some $\hat{x} \in \text{int}(\mathcal{Q}_{j,p}) \cap \text{int}(\text{supp}(\psi_\lambda)) \cap \partial B$, and let s be the associated slope of the tangent to the edge curve E in $\hat{x}$. Now consider another $\hat{x}' \in \text{int}(\mathcal{Q}_{j,p}) \cap \text{int}(\text{supp}(\psi_{\lambda'})) \cap \partial B$, and again let s' be the associated slope of the tangent to the edge curve E in $\hat{x}'$. Then, for sufficiently large scaling index j, by (8), $|s - s'|$ is sufficiently small and we may assume $s' \in [-3, 3]$. Hence the estimate (4) from Proposition 2.2 holds not only for $\hat{x}$ (with s), but also for $\hat{x}'$ (with s'). In fact, it can even be checked that by substituting s' by s in (4) the asymptotic behavior of the estimate for $|\Lambda_{j,p}(\epsilon)|$ does not change. Let us briefly outline the reasoning. First, we observe that WLOG we can assume that $|k + 2^{j/2}s| \geq 2 \cdot C_1$, where C_1 is the constant appearing in (8), since $|\{k \in \mathbb{Z} : |k + 2^{j/2}s| < 2 \cdot C_1\}| \leq C$ with C being independent on j and hence it can be deduced that the hypothesis $|k + 2^{j/2}s| \geq 2 \cdot C_1$ for each $j \geq 0$ does not affect our asymptotic estimate of $|\Lambda_{j,p}(\epsilon)|$. From (8), it then follows that

$$|k + 2^{j/2}s| \leq 2 \cdot |k + 2^{j/2}s'|,$$

which in turn implies

$$\frac{2^{-\frac{3}{4}j}}{|k + 2^{j/2}s'|^3} \leq 8 \frac{2^{-\frac{3}{4}j}}{|k + 2^{j/2}s|^3}.$$

Hence, by substituting s' by s in (4) the asymptotic behavior of the estimate for $|\Lambda_{j,p}(\epsilon)|$ does not change. Thus, it suffices to consider just one fixed $\hat{x} \in \text{int}(\mathcal{Q}_{j,p}) \cap \text{int}(\text{supp}(\psi_\lambda)) \cap \partial B$ with associated slope s in each $\mathcal{Q}_{j,p}$. We now turn to estimating $|\Lambda_{j,p}(\varepsilon)|$ in this case.

For this, by estimate (4) from Proposition 2.2, $|\langle f, \psi_\lambda \rangle| > \varepsilon$ implies

$$|k + 2^{j/2}s| \leq C \cdot \varepsilon^{-1/3} \cdot 2^{-j/4}. \tag{9}$$

From (7) and (9), we then conclude

$$|\Lambda_{j,p}(\varepsilon)| \leq C \cdot \sum_{k \in K_j(\varepsilon)} (|k + 2^{j/2}s| + 1) \leq C \cdot (\varepsilon^{-1/3} \cdot 2^{-j/4} + 1)^2, \tag{10}$$

where $K_j(\varepsilon) = \{k \in \mathbb{Z} : |k + 2^{j/2}s| \leq C \cdot \varepsilon^{-1/3} \cdot 2^{-j/4}\}$.

Case 2b. Exploiting similar arguments as in Case 2a, it also suffices to consider just one fixed $\hat{x} \in \text{int}(\mathcal{Q}_{j,p}) \cap \text{int}(\text{supp}(\psi_\lambda)) \cap \partial B$ with associated slope s in each $\text{int}(\mathcal{Q}_{j,p})$. Again, our goal is now to estimate $|\Lambda_{j,p}(\varepsilon)|$.

By estimate (5) from Proposition 2.2, $|\langle f, \psi_\lambda \rangle| > \varepsilon$ implies

$$C \cdot 2^{-\frac{9}{4}j} \geq \varepsilon,$$

hence

$$\text{and} \quad j \leq \frac{4}{9} \log_2 \left(\varepsilon^{-1} \right) + C. \tag{11}$$

Since there exists some C with

$$|\Lambda_{j,p}| \leq C \cdot 2^j,$$

it then follows that

$$|\Lambda_{j,p}(\varepsilon)| \leq C \cdot 2^j. \tag{12}$$

Notice that this last estimate is exceptionally crude, but will be sufficient for our purposes.

We now combine the estimates for $|\Lambda_{j,p}(\varepsilon)|$ derived in Case 2a and Case 2b. Since

$$\#(\mathcal{Q}_j) \leq C \cdot 2^{j/2}.$$

by (10) (and (6)) and by (11) (and (12)), we have

$$\begin{aligned}
|\Lambda(\varepsilon)| &\leq C \cdot \left[\sum_{j=0}^{\frac{4}{3} \log_2(\varepsilon^{-1})} 2^{j/2} \left(\varepsilon^{-1/3} \cdot 2^{-1/4j} + 1 \right)^2 + \sum_{j=0}^{\frac{4}{9} \log_2(\varepsilon^{-1}) + C} 2^{\frac{3}{2}j} \right] \\
&\leq C \cdot \varepsilon^{-2/3} \cdot \log_2(\varepsilon^{-1}).
\end{aligned} \tag{13}$$

Having estimated $|\Lambda(\varepsilon)|$, we are now ready to prove our main claim. For this, set $N = |\Lambda(\varepsilon)|$, i.e., the total number of shearlets ψ_λ such that the magnitude of the corresponding shearlet coefficient $\langle f, \psi_\lambda \rangle$ is larger than ε. By (13), the value ε can be written as a function of the total number of coefficients N in the following way:

$$\varepsilon(N) \leq C \cdot N^{-3/2} \cdot (\log N)^{3/2}, \quad \text{for sufficiently large } N > 0.$$

This implies that

$$|\theta(f)|_N \leq C \cdot N^{-3/2} \cdot (\log N)^{3/2}.$$

Hence,

$$\sum_{n>N} |\theta(f)|_n^2 \leq C \cdot N^{-2} \cdot (\log N)^3 \quad \text{for sufficiently large } N > 0,$$

which proves (3). The proof of Theorem 1.3 is finished.

3. Analysis of Shearlet Coefficients associated with the Smooth Part of a Cartoon-Like Image

In this section, we will prove Proposition 2.1. For this, we first prove a result which shows that, provided that the shearlet ψ satisfies certain decay conditions even with strong weights such as $(2^{4j})_j$, the system $\Psi(c; \psi)$ forms a Bessel-like sequence for $C^2(\mathbb{R}^2)$ with compact support in $[0,1]^2$.

In the following we will use the notation $r_j \sim s_j$ for $r_j, s_j \in \mathbb{R}$, if $C_1 \cdot r_j \leq s_j \leq C_2 \cdot r_j$ with constants C_1 and C_2 independent on the scale j.

Lemma 3.1. *Let $\psi \in L^2(\mathbb{R}^2)$ satisfy*

$$|\hat{\psi}(\xi)| \leq C_1 \cdot \min(1, |\xi_1|^\alpha) \cdot \min(1, |\xi_1|^{-\gamma}) \cdot \min(1, |\xi_2|^{-\gamma}) \text{ for all } \xi = (\xi_1, \xi_2) \in \mathbb{R}^2,$$

where $\gamma > 3$, $\alpha > \gamma + 2$, and C_1 is some constant. Then, there exists some $C > 0$ such that, for all $g \in C^2(\mathbb{R}^2)$ with compact support in $[0,1]^2$,

$$\sum_{j=0}^{\infty} \sum_{|k| \leq \lceil 2^{j/2} \rceil} \sum_{m \in \mathbb{Z}^2} 2^{4j} |\langle g, \psi_{j,k,m} \rangle|^2 \leq C \cdot \left\| \frac{\partial^2}{\partial x_1^2} g \right\|_2^2.$$

The proof of this lemma will explore the following result from [12], which we state here for the convenience of the reader.

Proposition 3.2. *[12] Let $\psi \in L^2(\mathbb{R}^2)$ satisfy*

$$|\hat{\psi}(\xi)| \leq C_1 \cdot \min(1, |\xi_1|^{\alpha}) \cdot \min(1, |\xi_1|^{-\gamma}) \cdot \min(1, |\xi_2|^{-\gamma}) \text{ for all } \xi = (\xi_1, \xi_2) \in \mathbb{R}^2,$$

where $\alpha > \gamma > 3$ and C_1 is some constant. Then, there exists some $C > 0$ such that, for all $\eta \in L^2(\mathbb{R}^2)$,

$$\sum_{j=0}^{\infty} \sum_{|k| \leq \lceil 2^{j/2} \rceil} \sum_{m \in \mathbb{Z}^2} |\langle \eta, \psi_{j,k,m} \rangle|^2 \leq C \cdot \|\eta\|_2^2.$$

Proof. (Proof of Lemma 3.1). By the assumption on ψ, the parameters α and γ can be chosen such that

$$|\hat{\psi}(\xi)| \leq C_1 \cdot \min(1, |\xi_1|^{\alpha}) \cdot \min(1, |\xi_1|^{-\gamma}) \cdot \min(1, |\xi_2|^{-\gamma}) \text{ for all } \xi = (\xi_1, \xi_2) \in \mathbb{R}^2,$$

where $\alpha > \gamma + 2, \gamma > 3$. Now, let $\eta \in L^2(\mathbb{R}^2)$ be chosen to satisfy

$$\frac{\partial^2}{\partial x_1^2} \eta = \psi.$$

Then a straightforward computation shows that η satisfies the hypotheses of Proposition 3.2. Using integration by parts,

$$\left| \left\langle \frac{\partial^2}{\partial x_1^2} g, \eta_{j,k,m} \right\rangle \right|^2 = 2^{4j} |\langle g, \psi_{j,k,m} \rangle|^2,$$

hence, by Proposition 3.2,

$$\sum_{j=0}^{\infty} \sum_{|k| \leq \lceil 2^{j/2} \rceil} \sum_{m \in \mathbb{Z}^2} 2^{4j} |\langle g, \psi_{j,k,m} \rangle|^2 = \sum_{j=0}^{\infty} \sum_{|k| \leq \lceil 2^{j/2} \rceil} \sum_{m \in \mathbb{Z}^2} \left| \left\langle \frac{\partial^2}{\partial x_1^2} g, \eta_{j,k,m} \right\rangle \right|^2$$

$$< C \cdot \left\| \frac{\partial^2}{\partial x_1^2} g \right\|_2^2.$$

The proof is complete. $\qquad\square$

This now enables us to derive Proposition 2.1 as a corollary.

Proof. (Proof of Proposition 2.1). Set

$$\tilde{\Lambda}_j = \{\lambda \in \Lambda_j : \text{supp}(\psi_\lambda) \cap \text{supp}(g) \neq \emptyset\}, \quad j > 0,$$

i.e., $\tilde{\Lambda}_j$ is the set of indices in Λ_j associated with shearlets whose support intersects the support of g. Then, for each $J > 0$, we have

$$N_J = \left| \bigcup_{j=0}^{J-1} \tilde{\Lambda}_j \right| \sim 2^{2J}. \tag{14}$$

Now, first observe that there exists some $C > 0$ such that

$$\sum_{j=1}^{\infty} 2^{4j} \sum_{n > N_j} |\theta(g)|_n^2 \;\leq\; C \cdot \sum_{j=1}^{\infty} \sum_{j'=j}^{\infty} \sum_{k,m} 2^{4j} |\langle g, \psi_{j',k,m} \rangle|^2$$

$$= \; C \cdot \sum_{j'=1}^{\infty} \sum_{k,m} |\langle g, \psi_{j',k,m} \rangle|^2 \left(\sum_{j=1}^{j'} 2^{4j} \right).$$

By Lemma 3.1, this implies

$$\sum_{j=1}^{\infty} 2^{4j} \sum_{n > N_j} |\theta(g)|_n^2 \leq C \cdot \sum_{j'=1}^{\infty} \sum_{k,m} 2^{4j'} |\langle g, \psi_{j',k,m} \rangle|^2 < \infty$$

and hence, also by (14),

$$\sum_{n > N_j} |\theta(g)|_n^2 \leq C \cdot (2^{2j})^{-2} \leq C \cdot N_j^{-2}.$$

Finally, let $N > 0$. Then there exists a positive integer $j_0 > 0$ satisfying

$$N \sim N_{j_0} \sim 2^{2j_0},$$

and the claim is proved. $\qquad\square$

4. Analysis of Shearlet Coefficients associated with the Discontinuity Curve

4.1. Proof of Proposition 2.2

Let $(j, k, m) \in \Lambda_{j,p}$, and fix $\hat{x} = (\hat{x}_1, \hat{x}_2) \in \operatorname{int}(\mathcal{Q}_{j,p}) \cap \operatorname{int}(\operatorname{supp}(\psi_\lambda)) \cap \partial B$. Let s be the slope of the tangent to the edge curve ∂B at $(\hat{x}_1, \hat{x}_2)$ as defined in Proposition 2.2, i.e., if ∂B is parameterized by $(E(x_2), x_2)$ in the interior of $\mathcal{S}_{j,p}$, then $s = E'(\hat{x}_2)$, and if ∂B is parameterized by $(x_1, E(x_1))$ in the

interior of $\mathcal{S}_{j,p}$, then $s = E'(\hat{x}_1)^{-1}$, where we now assume that $E'(\hat{x}_1) \neq 0$ and consider the case $E'(\hat{x}_1) = 0$ later.

By translation symmetry, WLOG we can assume that the edge curve satisfies $E(0) = 0$ with $(\hat{x}_1, \hat{x}_2) = (0,0)$. Further, since the conditions (i) and (ii) in Theorem 1.3 are independent on the translation parameter m, it does not play a role in our analysis. Hence, WLOG we choose $m = 0$. Also, since ψ is compactly supported, there exists some $L > 0$ such that $\mathrm{supp}(\psi) \subset [-L/2, L/2]^2$. By a rescaling argument, we can might assume $L = 1$. Even more, WLOG we can assume that $\mathrm{supp}(\psi) = [-1/2, 1/2]^2$, which implies

$$
\begin{aligned}
\mathrm{supp}(\psi_{j,k,0}) \ = \ & \{x \in \mathbb{R}^2 : -2^{-j/2}kx_2 - 2^{-j-1} \leq x_1 \leq -2^{-j/2}kx_2 + 2^{-j-1}, \\
& -2^{-\frac{j}{2}} \leq 2x_2 \leq 2^{-\frac{j}{2}}\},
\end{aligned}
\tag{15}
$$

since this will not change our asymptotic estimate for $|\Lambda_{j,p}(\epsilon)|$.

Let now $f \in \mathcal{E}^2(\nu)$, and select $\mathcal{P}$ to be the smallest parallelogram which entirely contains the discontinuity curve parameterized by $(E(x_2), x_2)$ or $(x_1, E(x_1))$ in the interior of $\mathrm{supp}(\psi_{j,k,0})$ and whose two sides are parallel to the tangent $x_1 = sx_2$ to the discontinuity curve at $(\hat{x}_1, \hat{x}_2) = (0,0)$. For an illustration, we refer to Figure 4. We now split the coefficients $|\langle f, \psi_{j,k,0} \rangle|$ into the part 'inside the parallelogram' and 'outside' of it exploiting the shearing property of shearlets for the second part, and obtain

$$
|\langle f, \psi_{j,k,0} \rangle| = |\langle \chi_{\mathcal{P}} f, \psi_{j,k,0} \rangle| + |\langle \chi_{\mathcal{P}^c} f(S_s \cdot), \psi_{j,\hat{k},0} \rangle|
\tag{16}
$$

where $\hat{k} = k + 2^{j/2}s$. From now on, we assume that $\hat{k} < 0$ with $\hat{k} = k + 2^{j/2}s$. The case $k + 2^{j/2}s \geq 0$ can be handled similarly.

Let us start by estimating the first term $|\langle \chi_{\mathcal{P}} f, \psi_{j,k,0} \rangle|$ in (16) stated as **Claim 1.**

$$
|\langle \chi_{\mathcal{P}} f, \psi_{j,k,0} \rangle| \leq C \cdot \frac{(1 + |s|^2)^{3/2}}{2^{3j/4} \cdot |\hat{k}|^3}.
\tag{17}
$$

First, notice that the tangent to the edge curve $(E(x_2), x_2)$ (or $(x_1, E(x_1))$) is given by $x_1 = sx_2$. We now assume that the edge curve is contained in a set $\{x \in \mathbb{R}^2 : x_1 \geq sx_2\}$, and just remark that the general case can be handled similarly. Let now d be the length of that side of $\mathcal{P}$, which is a part of the tangent $x_1 = sx_2$. We observe that d is in fact the distance between two points, in which the tangent $x_1 = sx_2$ intersects the boundary of $\mathrm{supp}(\psi_{j,k,0})$.
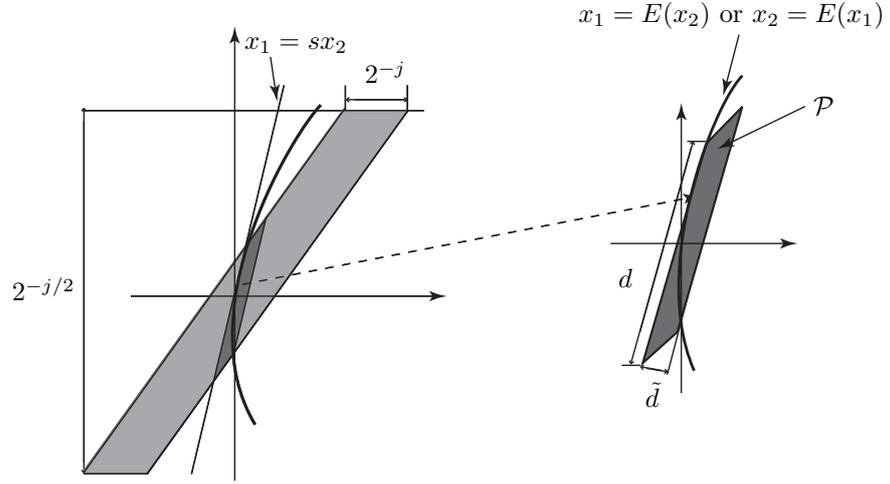
21

Figure 4: A shearlet $\psi_{j,k,0}$ intersecting the edge curve $x_1 = E(x_2)$ or $x_2 = E(x_1)$. The right hand side shows a magnification of the parallelogram $\mathcal{P}$.

For an illustration we wish to refer to Figure 4. From this observation, it follows that

$$d = \frac{2^{-j/2}\sqrt{1 + |s|^2}}{|\hat{k}|}.$$

We now let $\tilde{d}$ be the height of $\mathcal{P}$. Since the edge curve can be parameterized by a C^2 function E with bounded curvature,

$$\tilde{d} \leq C \cdot \Big(\frac{2^{-j/2}\sqrt{1 + |s|^2}}{|\hat{k}|}\Big)^2.$$

Summarizing, the volume of $\mathcal{P}$ can be estimated as

$$|\mathcal{P}| \leq C \cdot \frac{(1 + |s|^2)^{3/2}}{2^{3j/2} \cdot |\hat{k}|^3}.$$

This implies

$$|\langle f\chi_{\mathcal{P}}, \psi_{j,k,0}\rangle| \leq C \cdot 2^{3j/4} \cdot \|f\|_\infty \cdot \|\psi\|_\infty \cdot \frac{(1 + |s|^2)^{3/2}}{2^{3j/2} \cdot |\hat{k}|^3} \leq C \cdot \frac{(1 + |s|^2)^{3/2}}{2^{3j/4} \cdot |\hat{k}|^3},$$

and Claim 1, i.e., estimate (17), is proved.

Next, we estimate the second term, i.e., $|\langle \chi_{\mathcal{P}^c} f(S_s \cdot), \psi_{j,\hat{k},0} \rangle|$, in (16) stated as

Claim 2.

$$|\langle \chi_{\mathcal{P}^c} f(S_s \cdot), \psi_{j,\hat{k},0} \rangle| \leq C \cdot (1 + |s|)^2 \cdot \left(\frac{1}{2^{3j/4} \cdot |\hat{k}|^3} + \frac{1}{2^{7j/4} \cdot |\hat{k}|^2} \right). \qquad (18)$$

Notice that $S_s^{-1} \mathcal{P}$ entirely contains the edge curve of $f(S_s \cdot)$ in the interior of $\mathrm{supp}(\psi_{j,\hat{k},0})$ and that the boundary of the parallelogram $S_s^{-1} \mathcal{P}$ consists of two vertical line segments in the interior of $\mathrm{supp}(\psi_{j,\hat{k},0})$ (see Figure 5). By translation symmetry, this implies that for proving Claim 2, it suffices to estimate

$$\langle f_0(S_s \cdot) \chi_\Omega, \psi_{j,\hat{k},0} \rangle,$$

where $\Omega = \{ (x_1, x_2) \in \mathbb{R}^2 : x_1 > 0 \}$, $f_0 \in C^2(\mathbb{R}^2)$ compactly supported in $[0,1]^2$ and $\sum_{|\alpha| \leq 2} \|D^\alpha f_0\|_\infty \leq 1$. We wish to mention that the consideration of the case $x_1 > 0$ – compare the definition of the set Ω – is by no means restrictive, since the case $x_1 \leq 0$ can be handled in a similar way.

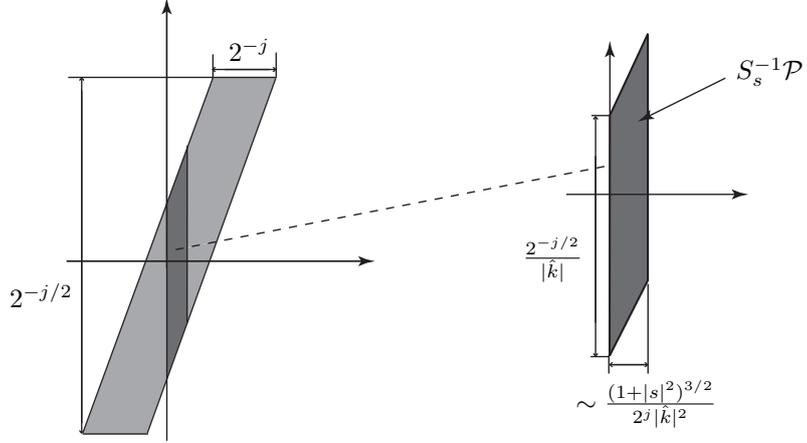

Figure 5: A shearlet $\psi_{j,\hat{k},0}$ intersecting the parallelogram $S_s^{-1} \mathcal{P}$. The right hand side shows a magnification of the parallelogram $S_s^{-1} \mathcal{P}$.

Now, again by translation symmetry, we may assume that $\partial(\mathrm{supp}\,\psi_{j,\hat{k},0})$ intersects the origin. In particular, we may translate $\mathrm{supp}(\psi_{j,\hat{k},0})$ defined in

(15) so that we might now consider

$$\text{supp}(\psi_{j,\hat{k},0}) + (-2^{j-1}, 0)$$

as the support of $\psi_{j,\hat{k},0}$. We refer the reader to Figure 6 for an illustration. This implies that there is one side of the boundary $\partial(\text{supp } \psi_{j,\hat{k},0})$, which is a part of the line

$$\mathcal{L} = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = (-2^{j/2}/\hat{k}) \cdot x_1\}$$

with slope $-2^{j/2}/\hat{k}$, as described in Figure 6. Applying the Taylor expansion for $f_0(S_s \cdot)$ at each point $x = (x_1, x_2) \in \mathcal{L}$, we obtain

$$f_0(S_s x) = a(x_1) + b(x_1)\left(x_2 + \frac{2^{j/2}}{\hat{k}} \cdot x_1\right) + c(x_1, x_2)\left(x_2 + \frac{2^{j/2}}{\hat{k}} \cdot x_1\right)^2,$$

where $a(x_1), b(x_1)$ and $c(x_1, x_2)$ are all bounded in absolute value by $C(1 + |s|)^2$. This implies (compare also an illustration of the area of integration in Figure 6)

$$|\langle f_0(S_s \cdot)\chi_\Omega, \psi_{j,\hat{k},0}\rangle| = \left| \int_0^{-\frac{\hat{k}}{2j}} \int_{-\frac{2^{j/2}}{\hat{k}} \cdot x_1}^{-\frac{2^{j/2}}{\hat{k}} \cdot x_1 - \frac{2^{-j/2}}{\hat{k}}} f_0(S_s x)\psi_{j,\hat{k},0}(x)\, dx_2 dx_1 \right|$$

$$\leq C \cdot (1 + |s|)^2 \cdot \left| \int_0^{-\frac{\hat{k}}{2j}} \sum_{\ell=1}^{3} I_\ell(x_1)\, dx_1 \right|, \qquad (19)$$

where

$$I_1(x_1) = \left| \int_0^{-\frac{2^{-j/2}}{\hat{k}}} T_\beta\left(\psi_{j,\hat{k},0}(x_1, x_2)\right) dx_2 \right|,$$

$$I_2(x_2) = \left| \int_0^{-\frac{2^{-j/2}}{\hat{k}}} x_2 \cdot T_\beta\left(\psi_{j,\hat{k},0}(x_1, x_2)\right) dx_2 \right|,$$

$$I_3(x_2) = \left| \int_0^{-\frac{2^{-j/2}}{\hat{k}}} x_2^2 \cdot T_\beta\left(\psi_{j,\hat{k},0}(x_1, x_2)\right) dx_2 \right|,$$

with T_β being the translation operator defined by $T_\beta(f) = f(\cdot - \beta)$ and $\beta \in \mathbb{R}$ being chosen to be $\beta = (0, (2^{j/2}/\hat{k}) \cdot x_1)$.
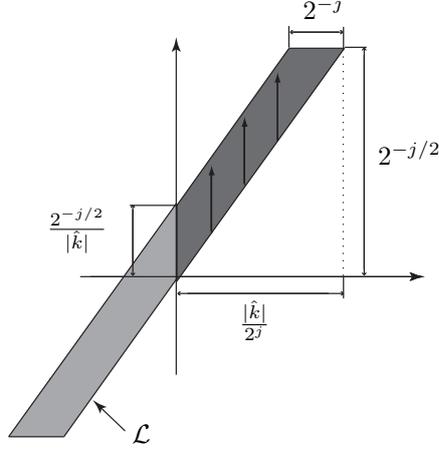
Figure 6: A shearlet $\psi_{j,\hat{k},0}$ intersecting the edge curve $x_1 = 0$ such that $\mathrm{supp}(\psi_{j,\hat{k},0})$ intersects the positive x_2 axis and $\partial(\mathrm{supp}\,\psi_{j,\hat{k},0})$ intersects the origin. The illustration also highlights the integration area for (19).

We first estimate $I_1(x_1)$. We observe that, since

$$\{(x_1,x_2) \in \mathbb{R}^2 : \psi_{j,\hat{k},0}(x_1,x_2) \neq 0\} \subset [0, 2^{-j/2}/|\hat{k}|] \quad \text{for a fixed } x_1 > 0,$$

WLOG, for any $x_1 > 0$, the interval $[0, 2^{-j/2}/|\hat{k}|]$ for the range of the integration in $I_1(x_1)$ can be replaced by $\mathbb{R}$ (see also Figure 6). Therefore, we have

$$I_1(x_1) = \left| \int_{\mathbb{R}} \psi_{j,\hat{k},0}(x_1,x_2)dx_2 \right| = \left| \int_{\mathbb{R}} \hat{\psi}_{j,\hat{k},0}(\xi_1,0) \cdot e^{2\pi i x_1 \xi_1} d\xi_1 \right|. \tag{20}$$

Now

$$|\hat{\psi}_{j,\hat{k},0}(\xi_1,\xi_2)| = 2^{-3j/4} \cdot |\hat{\psi}(2^{-j}\xi_1, 2^{-j/2}\xi_2 - 2^{-j}\hat{k}\xi_1)|.$$

and hence, by hypothesis (i) from Theorem 1.3,

$$|\hat{\psi}_{j,\hat{k},0}(\xi_1,0)| \leq 2^{j/4} \cdot |2^{-j}h(2^{-j}\xi_1)| \cdot |\hat{k}|^{-\gamma}. \tag{21}$$

By (20) and (21), it follows that

$$I_1(x_1) \leq C \cdot \frac{2^{j/4}}{|\hat{k}|^\gamma} \quad \text{for some } C > 0. \tag{22}$$

25

Next, we estimate $I_2(x_1)$. We have

$$
\begin{aligned}
I_2(x_1) &\leq \left| \int_{\mathbb{R}} x_2 \cdot \psi_{j,\hat{k},0}(x_1, x_2) dx_2 \right| + \frac{2^{j/2}}{|\hat{k}|} \cdot |x_1| \cdot \left| \int_{\mathbb{R}} \psi_{j,\hat{k},0}(x_1, x_2) dx_2 \right| \\
&= S_1 + S_2.
\end{aligned}
$$

To estimate S_1, observe that, by hypothesis (ii) from Theorem 1.3,

$$
\begin{aligned}
S_1 &= \frac{1}{2\pi} \left| \int_{\mathbb{R}} \left(\frac{\partial}{\partial \xi_2} \hat{\psi}_{j,\hat{k},0} \right) (\xi_1, 0) e^{2\pi i \xi_1 x_1} d\xi_1 \right| \\
&\leq \frac{1}{2\pi} \int_{\mathbb{R}} (2^{-\frac{j}{4}} \cdot h(2^{-j}\xi_1)) \cdot 2^{-j} \cdot |\hat{k}|^{-\gamma} d\xi_1.
\end{aligned}
\tag{23}
$$

By (22) and the fact that $0 \leq x_1 \leq \frac{|\hat{k}|}{2^j}$, the second term S_2 can be estimated as

$$
S_2 \leq C \cdot \left(\frac{2^{j/2}}{|\hat{k}|} |x_1| \right) \cdot \frac{2^{j/4}}{|\hat{k}|^{\gamma}} \leq C \cdot \left(2^{-j/2} \right) \cdot \frac{2^{j/4}}{|\hat{k}|^{\gamma}} \leq \frac{C}{2^{j/4} \cdot |\hat{k}|^{\gamma}}.
\tag{24}
$$

Concluding from (23) and (24), we obtain

$$
I_2(x_1) \leq S_1 + S_2 \leq \frac{C}{2^{j/4} \cdot |\hat{k}|^{\gamma}}.
\tag{25}
$$

Finally, we estimate $I_3(x_1)$. For this, notice that $2^{-3j/4} T_\beta(\psi_{j,\hat{k},0}(x_1, x_2))$ is bounded, hence

$$
I_3(x_1) \leq 2^{\frac{3}{4}j} \cdot C \cdot \left| \int_0^{\frac{-1}{2^{j/2}\hat{k}}} x_2^2 \, dx_2 \right| \leq \frac{C}{2^{\frac{3}{4}j} \cdot |\hat{k}|^3}.
\tag{26}
$$

Summarizing, by (19), (22), (25), and (26),

$$
\begin{aligned}
|\langle f_0(S_s \cdot) \chi_\Omega, \psi_{j,\hat{k},0} \rangle| &\leq C \cdot (1 + |s|)^2 \cdot \int_0^{\frac{|\hat{k}|}{2^j}} \left(\frac{2^{\frac{1}{4}j}}{|\hat{k}|^{\gamma}} + \frac{1}{2^{\frac{3}{4}j} \cdot |\hat{k}|^3} \right) dx_1 \\
&\leq C \cdot (1 + |s|)^2 \cdot \left(\frac{1}{2^{3j/4} \cdot |\hat{k}|^3} + \frac{1}{2^{\frac{7}{4}j} \cdot |\hat{k}|^2} \right),
\end{aligned}
$$

and Claim 2, i.e., estimate (18), is proved.

From Claim 1 and 2, i.e., from (17) and (18), we conclude that

$$|\langle f, \psi_{j,k,0}\rangle| \le C\Big[(1+|s|)^2\Big(\frac{1}{2^{\frac{3}{4}j}\cdot|\hat{k}|^3} + \frac{1}{2^{\frac{7}{4}j}\cdot|\hat{k}|^2}\Big) + (1+|s|^2)^{3/2}\frac{1}{2^{\frac{3}{4}j}\cdot|\hat{k}|^3}\Big].$$

This implies (4) and (5) except for the case $s = \infty$, which we will study now.

Finally, we consider the case $s = \infty$, i.e., we assume that the edge curve is parameterized by $(x_1, E(x_1))$ in the interior of $\mathcal{S}_{j,p}$ such that $E \in C^2$ and $E'(\hat{x}_1) = 0$. As before, let $f \in \mathcal{E}^2(\nu)$, and select $\mathcal{P}$ to be the smallest parallelogram which entirely contains the discontinuity curve in the interior of $\mathrm{supp}(\psi_{j,k,0})$ and whose two sides are parallel to $x_2 = 0$. Similarly as before, we consider

$$|\langle f, \psi_{j,k,0}\rangle| = |\langle \chi_{\mathcal{P}}f, \psi_{j,k,0}\rangle| + |\langle \chi_{\mathcal{P}^c}f, \psi_{j,k,0}\rangle|. \tag{27}$$

and estimate both terms on the RHS separately.

We first consider the term $|\langle \chi_{\mathcal{P}}f, \psi_{j,k,0}\rangle|$. Using similar arguments as before, which we decided not to include in detail to avoid repetitions, one can prove that

$$|\langle \chi_{\mathcal{P}}f, \psi_{j,k,0}\rangle| \le C \cdot 2^{-\frac{9}{4}j}. \tag{28}$$

Turning our attention to the second term $|\langle \chi_{\mathcal{P}^c}f, \psi_{j,k,0}\rangle|$, we first observe that $\mathcal{P}$ entirely contains the edge curve of f in the interior of $\mathrm{supp}(\psi_{j,k,0})$ and that the boundary of the parallelogram $\mathcal{P}$ consists of two horizontal line segments in the interior of $\mathrm{supp}(\psi_{j,k,0})$ (see Figure 7). By translation symmetry, this implies that for the second term in (27), i.e., for $|\langle \chi_{\mathcal{P}^c}f, \psi_{j,k,0}\rangle|$, it suffices to estimate $|\langle f_0\chi_{\tilde{\Omega}}, \psi_{j,k,0}\rangle|$, where $\tilde{\Omega} = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 > 0\}$ and $f_0 \in C^2(\mathbb{R}^2)$ compactly supported in $[0,1]^2$ with $\sum_{|\alpha|\le 2}\|D^\alpha f_0\|_\infty \le 1$. As before, the consideration of the case $x_2 > 0$ – compare the definition of the set $\tilde{\Omega}$ – is by no means restrictive, since the case $x_2 \le 0$ can be handled in a similar way. Observe that the (horizontal) vanishing moment condition follows from condition (i) in Theorem 1.3. Let us briefly think about this: Letting ξ_2 be fixed, condition (i) immediately implies $\hat{\psi}(0, \xi_2) = 0$. Then, by applying Taylor expansion, it follows that $\hat{\psi}(\xi_1, \xi_2) = \frac{\partial\hat{\psi}}{\partial\xi_1}(0, \xi_2)\cdot\xi_1 + \mathcal{O}(\xi_1^2)$, and, again by condition (i), we have $\frac{\hat{\psi}(\xi_1,\xi_2)}{\xi_1} \to 0$ as $\xi_1 \to 0$, hence $\frac{\partial\hat{\psi}}{\partial\xi_1}(0, \xi_2) = 0$. This procedure can now be continued. Especially, condition (i) from Theorem 1.3 implies

$$\int_{\mathbb{R}} x_1^\ell \cdot \psi(x_1, x_2)dx_1 = 0 \quad \text{for all } x_2 \in \mathbb{R} \text{ and } \ell = 0, 1. \tag{29}$$

Further, we utilize that the shearing operation S_k preserves vanishing moments along the x_1 axis. That is,

$$\int_{\mathbb{R}} x_1^\ell \cdot \psi(S_k(x_1, x_2)^T)dx_1 = 0 \quad \text{for all } k, x_2 \in \mathbb{R} \text{ and } \ell = 0, 1.$$

This can be seen as follows. For $\ell = 0, 1$ and for a fixed $x_2 \in \mathbb{R}$, the function $x_1 \mapsto (x_1 - kx_2)^\ell$ is a polynomial of degree less than or equal to ℓ, hence, by condition (29) on the number of vanishing moments on ψ, we have

$$\int_{\mathbb{R}} x_1^\ell \cdot \psi(S_k(x_1, x_2)^T)dx_1 = \int_{\mathbb{R}} (x_1 - kx_2)^\ell \cdot \psi(x_1, x_2)dx_1 = 0 \quad \text{for all } k \in \mathbb{R}.$$
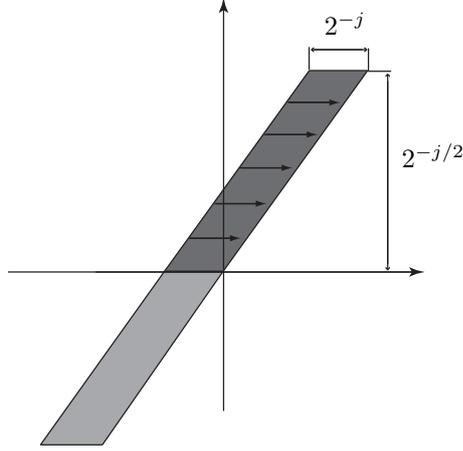

Figure 7: A shearlet $\psi_{j,\hat{k},0}$ intersecting the edge curve $x_2 = 0$ such that $\text{supp}(\psi_{j,\hat{k},0})$ intersects the positive x_1 axis and $\partial(\text{supp}\,\psi_{j,\hat{k},0})$ intersects the origin. The illustration also highlights the integration area for (30).

Employing Taylor expansion and integration (compare Figure 7) similar to the proof of the previous case, we finally obtain

$$|\langle f_0\chi_{\tilde{\Omega}}, \psi_{j,k,0}\rangle| \leq C \cdot 2^{3j/4} \cdot \int_0^{\cdot 2^{-j/2}} \int_{-2^{-j}}^0 x_1^2 dx_1 dx_2 \leq C \cdot 2^{-11j/4} \qquad (30)$$

By (28) and (30), we obtain (5) when $s = \infty$. This completes the proof.

[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** (2004), 219–266.

[2] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, *Tree approximation and optimal encoding,* Appl. Comput. Harmon. Anal. **11** (2001), 192–226.

[3] S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke, *Shearlet Coorbit Spaces and associated Banach Frames*, Appl. Comput. Harmon. Anal. **27** (2009), 195–214.

[4] M. N. Do and M. Vetterli, *The contourlet transform: an efficient directional multiresolution image representation*, IEEE Trans. Image Proc. **14** (2005), 2091-2106.

[5] D. L. Donoho, *Wedgelets: nearly minimax estimation of edges*, Ann. Statist. **27** (1999), 859–897.

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

[7] D. L. Donoho and G. Kutyniok, *Microlocal Analysis of the Geometric Separation Problems*, preprint (2010).

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

[9] K. Guo, G. Kutyniok, and D. Labate, *Sparse Multidimensional Representations using Anisotropic Dilation and Shear Operators*, Wavelets and Splines (Athens, GA, 2005), Nashboro Press, Nashville, TN (2006), 189–201.

[10] K. Guo, D. Labate, and W. Lim, *Edge Analysis and identification using the Continuous Shearlet Transform*, Appl. Comput. Harmon. Anal. **27** (2009), 24–46.

[11] K. Guo and D. Labate, *Optimally Sparse Multidimensional Representation using Shearlets*, SIAM J. Math Anal. **39** (2007), 298–318.

[12] P. Kittipoom, G. Kutyniok, and W.-Q Lim, *Construction of Compactly Supported Shearlets*, preprint (2010).

[13] P. Kittipoom, G. Kutyniok, and W.-Q Lim, *Irregular Shearlet Frames: Geometry and Approximation Properties*, preprint (2010).

[14] G. Kutyniok and D. Labate, *Construction of Regular and Irregular Shearlet Frames*, J. Wavelet Theory and Appl. **1** (2007), 1–10.

[15] G. Kutyniok and D. Labate, *Resolution of the Wavefront Set using Continuous Shearlets*, Trans. Amer. Math. Soc. **361** (2009), 2719–2754.

[16] G. Kutyniok, M. Shahram, and D. L. Donoho, *Development of a Digital Shearlet Transform Based on Pseudo-Polar FFT*, in Wavelets XIII (San Diego, CA, 2009), D. Van De Ville, V. K. Goyal und M. Papadakis, eds., 74460B-1 - 74460B-13, SPIE Proc. **7446**, SPIE, Bellingham, WA, 2009.

[17] D. Labate, W-Q. Lim, G. Kutyniok, and G. Weiss. *Sparse multidimensional representation using shearlets*, Wavelets XI (San Diego, CA, 2005), 254-262, SPIE Proc. 5914, SPIE, Bellingham, WA, 2005.

[18] W. Lim, *Discrete Shearlet Transform: New Multiscale Directional Image Representation*, Proc. SAMPTA09, Marseille 2009.

[19] M. Wakin, J. Romberg, C. Hyeokho, and R. Baraniuk, *Image compression using an efficient edge cartoon + texture model*, In Data Compression Conference, 2002. Proceedings. DCC 2002, **2-4** (2002), 43–52.