Dualizable Shearlet Frames and Sparse Approximation

Gitta Kutyniok1 · Wang-Q Lim2

Received: 9 November 2014 / Revised: 17 October 2015 / Accepted: 22 January 2016 / Published online: 10 March 2016 © Springer Science+Business Media New York 2016

Abstract Shearlet systems have been introduced as directional representation systems, which provide optimally sparse approximations of a certain model class of functions governed by anisotropic features while allowing faithful numerical realizations by a unified treatment of the continuum and digital realm. They are redundant systems, and their frame properties have been extensively studied. In contrast to certain band-limited shearlets, compactly supported shearlets provide high spatial localization but do not constitute Parseval frames. Thus reconstruction of a signal from shearlet coefficients requires knowledge of a dual frame. However, no closed and easily computable form of any dual frame is known. In this paper, we introduce the class of dualizable shearlet systems, which consist of compactly supported elements and can...
be proved to form frames for $L^2(\mathbb{R}^2)$. For each such dualizable shearlet system, we then provide an explicit construction of an associated dual frame, which can be stated in closed form and is efficiently computed. We also show that dualizable shearlet frames still provide near optimal sparse approximations of anisotropic features.

**Keywords** Anisotropic features · Dual frames · Frames · Shearlets · Sparse approximation

**Mathematics Subject Classification** Primary 42C40; Secondary 42C15 · 65T60 · 94A08

1 Introduction

In recent years, methodologies utilizing sparse approximations have had a tremendous impact on data science. This is primarily due to the method of compressed sensing (see [2,10] or [8]), which played a major role in the initiation of today’s paradigm that typical data admits a sparse representation within a suitably chosen orthonormal basis, or, more generally, a frame [4]. In fact, frames—redundant, yet stable systems—are typically preferable due to the added flexibility the redundancy provides. However, although a frame might provide even optimally sparse approximations within a model situation, in the end, one still needs to reconstruct the data from the respective frame coefficients. For an orthonormal basis, this can be achieved by the classical decomposition formula. In the situation of a frame though, a so-called dual frame is required.

In this paper, we will consider this problem in the situation of imaging sciences. Since it is typically assumed that images are governed by edge-like structures, a common model situation is that of cartoon-like functions, which are—coarsely speaking—compactly supported piecewise smooth functions. Shearlet systems [13], which might be among the most widely used directional representation systems today, have been shown to deliver optimally sparse approximations of this class. However, their compactly supported version, though superior due to high spatial localization, forms a (nontight) frame for $L^2(\mathbb{R}^2)$; but the construction of a dual having a closed and easily computable form is an open problem.

1.1 Data Processing by Frames

Frames have a long history of providing decompositions and expansions for data processing, and the reader might consult [3] for applications in audio processing or communication theory. A frame for a Hilbert space $\mathcal{H}$ is a sequence $(\varphi_i)_{i \in I}$ satisfying $A\|f\|^2 \leq \sum_{i \in I} |\langle f, \varphi_i \rangle|^2 \leq B\|f\|^2$ for all $f \in \mathcal{H}$ with $0 < A \leq B < \infty$. If the frame bounds $A$ and $B$ can chosen to be equal, it is typically called a tight frame, and in the case of $A = B = 1$, a Parseval frame.

Analysis of an element $f \in \mathcal{H}$ by a frame $(\varphi_i)_{i \in I}$ is typically achieved by application of the analysis operator $T$ given by

$$T : \mathcal{H} \to \ell^2(I), \quad f \mapsto (\langle f, \varphi_i \rangle)_{i \in I}.$$
Reconstruction of $f$ from the sequence of frame coefficients $(\langle f, \varphi_i \rangle)_{i \in I}$ is possible by utilizing the adjoint operator $T^*$, since it can be shown that

$$f = \sum_{i \in I} \langle f, \varphi_i \rangle (T^* T)^{-1} \varphi_i \quad \text{for all } f \in \mathcal{H}. \quad (1)$$

Unless $(\varphi_i)_{i \in I}$ forms a tight frame—in which case $T^* T$ is a multiple of the identity, we face the difficulty of having to invert the operator $T^* T$ in order to compute the canonical dual frame $((T^* T)^{-1} \varphi_i)_{i \in I}$.

In fact, certainly, the canonical dual frame is not the only choice for deriving a reconstruction formula such as (1). In general, one calls $(\tilde{\varphi}_i)_{i \in I}$ an associated dual frame if the following is true:

$$f = \sum_{i \in I} \langle f, \varphi_i \rangle \tilde{\varphi}_i \quad \text{for all } f \in \mathcal{H}. \quad (2)$$

### 1.2 Sparse Approximation Using Frames

One key feature of frames, which is in particular beneficial for deriving sparse approximations, is their redundancy. Sparse approximation by a frame $(\varphi_i)_{i \in I}$ can be regarded from two sides: On the one hand, we might consider expansions in terms of the frame such as

$$f = \sum_{i \in I} c_i \varphi_i, \quad (3)$$

expecting the existence of some coefficient sequence $(c_i)_{i \in I}$, which is sparse in the sense of, for instance, $\| (c_i)_{i \in I} \|_{\ell^1(I)} < \infty$ or at least $\| (c_i)_{i \in I} \|_{\ell^p(I)} < \infty$ for some $p < 2$.

This is however not the approach normally taken in data science, in particular as related to compressed sensing. Instead, we expect that the sequence of frame coefficients $(\langle f, \varphi_i \rangle)_{i \in I}$ is sparse. In [18], this situation is termed co-sparsity, and in fact sparsity within a frame is typically exploited in this way. For instance, reconstruction from highly undersampled data is then achieved by placing the $\ell_1$-norm on such coefficient sequences and minimizing over all $f \in \mathcal{H}$.

Thus, instead of expansions of the form (3), we consider (2) in the sense of a reconstruction procedure. This certainly requires having access to some dual frame associated with $(\varphi_i)_{i \in I}$. One can circumvent this problem by using iterative methods such as conjugate gradients whose efficiency depends heavily on the ratio of the frame bounds. But such methods deliver only approximate solutions and are rather slow.

### 1.3 Imaging Science and Anisotropic Features

Images play a key part in data science, as a significant percentage of data today are in fact images. Following the program discussed before, it is illusory to assume
that reasonable results can be derived for the whole Hilbert space $L^2(\mathbb{R}^2)$. Hence we restrict to an appropriate subset that we assume models the essential features of the images. For such a class, so-called cartoon-like functions introduced in [9] are typically taken. These are basically compactly supported functions that are $C^2$ apart from a closed $C^2$ discontinuity curve with bounded curvature (Definition 4.1). The intuition is that edge-like structures are typically prominent in images and, in addition, the neurons in the visual cortex of humans also react very strongly to those features. It should be emphasized that certainly such structures appear in other situations as well, such as in solutions of transport dominated equations [5,6].

Donoho then proved in [9] that the $L^2$-error of best $N$-term approximation $f_N$ of such a cartoon-like function $f$ by any frame for $L^2(\mathbb{R}^2)$ behaves as

$$\|f - f_N\|_2 \gtrsim N^{-1} \text{ as } N \to \infty.$$  

This result provides a notion of optimality, and frames satisfying this approximation rate up to a log-factor are customarily referred to as systems delivering optimal sparse approximations within the class of cartoon-like functions.

1.4 Shearlet Systems

Shearlet systems were originally introduced in [12] as a directional representation system that meets this benchmark result, but which—in contrast to the previously advocated system of curvelets [1]—fit into the framework of affine systems and also allow a faithful implementation by a unified treatment of the continuum and digital realm.

Shearlet systems are based on three operations: parabolic scaling $A_j$, $j \in \mathbb{Z}$ to provide different resolution levels, shearing $S_k$, $k \in \mathbb{Z}$ to provide different orientations, both given by

$$A_j = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{\lfloor \frac{j}{2} \rfloor} \end{pmatrix} \quad \text{and} \quad S_k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix},$$

as well as translation to provide different positions. The definition of a (cone-adapted) shearlet system is then as follows. We wish to mention that the term “cone-adapted” is due to the fact that the different systems $\Psi(\psi; c)$ and $\tilde{\Psi}(\tilde{\psi}; c)$ are responsible for the horizontal and vertical cone in the Fourier domain, respectively; thereby, together with $\Phi(\phi; c_1)$ achieving a complete system with a finite set of shears for each sale $j$.

**Definition 1.1** For $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ and $c = (c_1, c_2) \in (\mathbb{R}^+)^2$, the (cone-adapted) shearlet system $S\mathcal{H}(\phi, \psi, \tilde{\psi}; c)$ is defined by

$$S\mathcal{H}(\phi, \psi, \tilde{\psi}; c) = \Phi(\phi; c_1) \cup \Psi(\psi; c) \cup \tilde{\Psi}(\tilde{\psi}; c),$$

\(\square\) Springer
where
\[
\Phi(\phi; c_1) = \left\{ \psi_m := \phi(\cdot - c_1 m) : m \in \mathbb{Z}^2 \right\},
\]
\[
\Psi(\psi; c) = \left\{ \psi_{j,k,m} := |\det(A_j)|^{1/2} \psi(S_k A_j \cdot - \text{diag}(c_1, c_2)m) : j \geq 0, |k| \leq 2^{[j/2]}, m \in \mathbb{Z}^2 \right\},
\]
\[
\tilde{\Psi}(\tilde{\psi}; c) = \left\{ \tilde{\psi}_{j,k,m} := |\det(\tilde{A}_j)|^{1/2} \tilde{\psi}(S_k ^T \tilde{A}_j \cdot - \text{diag}(c_2, c_1)m) : j \geq 0, |k| \leq 2^{[j/2]}, m \in \mathbb{Z}^2 \right\},
\]

with \( \tilde{A}_j = \text{diag}(2^{j/2}, 2^j) \).

1.5 Problems with Shearlet Frames

For high spatial localization, compactly supported shearlet systems are considered, which are also implemented in ShearLab (see www.ShearLab.org) [17]. As shown in [15], compactly supported generators \( \phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2) \) can be constructed such that the associated shearlet system \( \mathcal{SH}(\phi, \psi, \tilde{\psi}; c) \) constitutes a frame—not a tight frame—for \( L^2(\mathbb{R}^2) \) with controllable frame bounds dependent on \( c \). Under slightly stronger conditions, it was proved in [16] that such systems also deliver optimally sparse approximations of cartoon-like functions.

In the situation of bandlimited shearlet frames (i.e., the Fourier transform is compactly supported), Grohs derived results on the existence of “nice” duals [11]. However, for compactly supported shearlet frames, no closed, easily computable form of an associated dual frame is known, even when allowing small modifications of the shearlet system.

1.6 Our Contribution

In this paper, we present a solution to this problem. We construct a shearlet system that can be regarded as being of the form \( \mathcal{SH}(\phi, \psi, \tilde{\psi}; c) \) and that satisfies the following properties:

- The system is compactly supported and forms a frame for \( L^2(\mathbb{R}^2) \).
- The system delivers near optimal sparse approximations of cartoon-like functions.
- An associated dual frame can be stated in closed form and is efficiently computed.
- It is composed of orthonormal bases, which provides it with a distinct, accessible structure.

In addition, the novel proof technique which we use for proving the approximation properties of dualizable shearlet frames along the way allow an improvement of previous approximation results from [16] for the class of compactly supported shearlet frames introduced in [15] with respect to the exponent of the additional log-term (see Theorem 4.4). It should be mentioned that with this result, this exponent in the
log-term is the smallest known for any directional representation system, in particular, curvelets [1].

1.7 Outline

The paper is organized as follows. The construction of dualizable shearlet systems is presented and discussed in Sect. 2; the definition is stated in Definition 2.5. Section 3 is devoted to the analysis of frame properties of dualizable shearlet systems, namely showing (in Theorem 3.1) that these systems do form frames for \( L^2(\mathbb{R}^2) \) and that an associated dual frame can be explicitly given in closed form. The statement that dualizable shearlet systems do provide near optimal sparse approximations of cartoon-like functions is presented in Sect. 4 as Theorem 4.3. Since the proof is very involved, the key steps and the core part are presented in Sect. 5 with the proofs of several preliminary lemmata being outsourced to Sect. 6.

2 Construction of Dualizable Shearlet Frames

This section is devoted to the construction of dualizable shearlet frames. One key ingredient is a family of orthonormal bases for each shearing direction, which will be discussed in Sect. 2.1. Since those elements lack directionality in the sense of wedge-like shape elements, they are subsequently filtered, yielding the desired dualizable shearlet systems (see Sect. 2.2). We emphasize that we will only present the construction for the horizontal Fourier cone in detail—compare the cone-based definition of shearlet systems from Sect. 1.4; the vertically aligned system will be derived by switching the variables.

Since the construction is rather technical in nature, it is not initially clear that the term “shearlets” is justified; we argue in Sect. 2.3 in favor of this expression by comparing the novel systems to customarily defined shearlets (cf. Sect. 1.4).

2.1 Basic Ingredients

To construct a family of orthonormal bases for \( L^2(\mathbb{R}^2) \) for each shearing direction, we first let \( \varphi_1, \psi_1 \in L^2(\mathbb{R}) \) be compactly supported functions that satisfy the support condition

\[
\delta_{\varphi_1} = \inf_{\xi \in \left[-\frac{1}{2}, \frac{1}{2}\right]} |\hat{\varphi}_1(\xi)| > 0, \tag{4}
\]

as well as, for some \( \rho \in (0, \frac{1}{12}), \alpha \geq \frac{6}{\rho} + 1, \) and \( \beta > \alpha + 1, \) the decay conditions

\[
\left| \left( \frac{d}{d\xi} \right)^\ell \psi_1(\xi) \right| \lesssim \min\{1, |\xi|^{\alpha}\} \left( 1 + |\xi| \right)^{-\beta} \quad \text{and} \quad \left| \left( \frac{d}{d\xi} \right)^\ell \hat{\varphi}_1(\xi) \right| \lesssim \frac{1}{\left( 1 + |\xi| \right)^\beta} \quad \text{for } \ell = 0, 1. \tag{5}
\]
In addition, we require the system
\[ \{ \varphi_1 (\cdot - m) : m \in \mathbb{Z} \} \cup \{ 2^{j/2} \psi_1 (2^j \cdot -m) : j \geq 0, m \in \mathbb{Z} \} \]
to form an orthonormal basis for \( L^2(\mathbb{R}) \). For the existence of such functions, we refer to [7].

Second, we utilize this univariate system to construct the desired family of orthonormal bases. We now follow the following strategy: We lift the system to \( L^2(\mathbb{R}^2) \) in such a way that we achieve a tiling of Fourier domain according to Fig. 1a and then apply shearing operators.

To generate the anticipated tiling, for \( x = (x_1, x_2) \in \mathbb{R}^2 \), we set
\[ \psi^0(x) := \psi_1(x_1) \varphi_1(x_2) \quad \text{and} \quad \psi^p(x) := 2^{(p-1)/2} \psi_1(x_1) \varphi_1(2^p x_2) \quad \text{for} \quad p > 0, \]
as well as
\[ \varphi^0(x) := \varphi_1(x_1) \varphi_1(x_2) \quad \text{and} \quad \varphi^p(x) := 2^{(p-1)/2} \varphi_1(x_1) \varphi_1(2^p x_2) \quad \text{for} \quad p > 0. \]

Notice that the parameter \( p \in \mathbb{N}_0 \) will be utilized to derive the dyadic substructure in the vertical direction.

For a fixed integer \( j_0 \geq 0 \), we next consider the system given by
\[ \{ |\det(A_{j_0})|^{1/2} \varphi^p(A_{j_0} \cdot -D_p m) \cdot D_p m) : j \geq j_0, p \geq 0 \}, \]
where \( D_p = \text{diag}(1, 2^{-\max\{p-1,0\}}) \), which achieves the tiling of Fourier domain as depicted in Fig. 1a. It will be shown in Lemma 2.2 that this system forms an orthonormal basis for \( L^2(\mathbb{R}^2) \).

We now carefully insert the shearing operator. Drawing from the definition of “standard” shearlets, the shear parameter should equal \( \frac{k}{2^{j/2}} \) with \( |k| \leq 2^{j/2} \). Since we later need to parameterize by those quotients, we require a unique representation without ambiguity. For this, we define the map
\[ s : (j, q) : j = 0, q = 0 \cup \{ j : j \geq 0 \} \times \{ q : |q| \leq 2^j, q \in 2\mathbb{Z} + 1 \} \rightarrow [-1, 1], \]
\[ s(j, q) := \frac{q}{2^j}, \]
which is obviously injective. Thus from now on, we consider the set of shear parameters given by
\[ \mathcal{S} = \{ s(j, q) = 0 : j = 0, q = 0 \} \cup \{ s(j, q) : j \geq 0, |q| \leq 2^j, q \in 2\mathbb{Z} + 1 \}. \]

Armed with this definition, we can now define what we call shearlet-type wavelet systems.

**Definition 2.1** Let \( \varphi_1, \psi_1 \in L^2(\mathbb{R}) \), and let \( \varphi^p, \psi^p \in L^2(\mathbb{R}^2) \), \( p \geq 0 \), be defined as before. Further, set \( D_p := \text{diag}(1, 2^{-d_p}) \) with \( d_p := \max\{p - 1, 0\} \). Then, for each
shear parameter \(s := s(\lceil j_0/2 \rceil, q_0) \in \mathbb{S}\), where \(j_0\) is the smallest nonnegative integer such that \(s = \frac{q_0}{2^{j_0/2}}\), we define the shearlet-type wavelet system \(\Psi_s(\varphi_1, \psi_1)\) by
\[
\Psi_s(\varphi_1, \psi_1) := \{\varphi_{j_0, s, m, p}, \psi_{j_0, s, m, p} : j \geq j_0, m \in \mathbb{Z}^2, p \geq 0\},
\]
where
\[
\varphi_{j, s, m, p} := |\det(A_j)|^{1/2} \varphi^p(A_j S_s \cdot -D_p m) \quad \text{and} \quad \psi_{j, s, m, p} := |\det(A_j)|^{1/2} \psi^p(A_j S_s \cdot -D_p m).
\]

The tiling of the Fourier domain by shearlet-type wavelet systems is depicted in Fig. 1.

The next result shows that, for each shear parameter, the associated shearlet-type wavelet system indeed constitutes an orthonormal basis.

**Lemma 2.2** For each \(s \in \mathbb{S}\), the shearlet-type wavelet system \(\Psi_s(\varphi_1, \psi_1)\) is an orthonormal basis for \(L^2(\mathbb{R}^2)\).

**Proof** Without loss of generality, we consider \(\Psi_0(\varphi_1, \psi_1)\), where \(s = 0\) (and hence \(j_0 = 0\)) in Definition 2.1. Then, for \(x = (x_1, x_2) \in \mathbb{R}^2\), by definition,
\[
\varphi_{0, 0, m, 0}(x) = \varphi_1(x_1 - m_1)\varphi_1(x_2 - m_2) \quad \text{and} \quad \varphi_{0, 0, m, p}(x) = \varphi_1(x_1 - m_1)2^{d_p} \varphi_1 \left(2^{d_p} x_2 - m_2\right),
\]
as well as, in addition for \(j \geq 0\) and \(p > 0\),
\[
\psi_{j, 0, m, 0}(x) = 2^{j} \psi_1 \left(2^j x_1 - m_1\right)2^{j+1} \varphi_1 \left(2^{j+1} x_2 - m_2\right)
\]
and
\[
\psi_{j, 0, m, p}(x) = 2^{j} \psi_1 \left(2^j x_1 - m_1\right)2^{d_p} 2^{j+1} \varphi_1 \left(2^{d_p} 2^{j+1} x_2 - m_2\right).
\]
Next, for each \(j \geq 0\), let \(V_j\) and \(W_j\) be the subspaces of \(L^2(\mathbb{R}^2)\) defined by
\[
V_j = \text{span}\{2^j \varphi_1(2^j \cdot -m) : m \in \mathbb{Z}\} \quad \text{and} \quad W_j = \text{span}\{2^j \psi_1(2^j \cdot -m) : m \in \mathbb{Z}\}.
\]
By construction, for each \( p > 0 \), the systems \( \{ \varphi_{0,0,m,0} : m \in \mathbb{Z}^2 \} \) and \( \{ \varphi_{0,0,m,p} : m \in \mathbb{Z}^2 \} \) form orthonormal bases for \( V_0 \otimes V_0 \) and \( V_0 \otimes W_{dp} \), respectively. Similarly, for \( j \geq 0 \) and \( p > 0 \), \( \{ \psi_{j,0,m,0} : m \in \mathbb{Z}^2 \} \) and \( \{ \psi_{j,0,m,p} : m \in \mathbb{Z}^2 \} \) form orthonormal bases for \( W_{j} \otimes V_{\lfloor \frac{j}{2} \rfloor} \) and \( W_{j} \otimes W_{dp+\lfloor \frac{j}{2} \rfloor} \), respectively.

Since \( V_0 \perp W_{j} \) and \( W_{j} \perp W_{j'} \) for \( j, j' \geq 0 \), \( j \neq j' \), and \( dp = p - 1 \) for \( p > 0 \), those subspaces are mutually orthogonal, and, for each \( j \geq 0 \),

\[
(V_0 \otimes V_0) \oplus \bigoplus_{p=1}^{\infty} (V_0 \otimes W_{dp}) \oplus (W_{j} \otimes V_{\lfloor \frac{j}{2} \rfloor}) \oplus \bigoplus_{p=1}^{\infty} (W_{j} \otimes W_{dp+\lfloor \frac{j}{2} \rfloor}) = (V_0 \otimes L^2(\mathbb{R})) \oplus (W_{j} \otimes L^2(\mathbb{R})).
\]

From this, we finally obtain

\[
(V_0 \otimes L^2(\mathbb{R})) \oplus \bigoplus_{j=0}^{\infty} W_{j} \otimes L^2(\mathbb{R}) = L^2(\mathbb{R}^2).
\]

This proves our claim. \( \square \)

We wish to mention that the definition of a dualizable shearlet system in Definition 2.5 will also require the systems derived by a permutation matrix \( R := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \) switching the variables in \( \Psi_s(\varphi_1, \psi_1) \).

### 2.2 Definition

The next step consists of a filtering procedure. To define the filters, let \( g \in L^2(\mathbb{R}^2) \) be a compactly supported function satisfying the conic support condition

\[
\delta_g = \inf_{\xi \in \Omega_g} |\hat{g}(\xi)| > 0, \quad \text{where} \ \Omega_g = \left\{ \xi \in \mathbb{R}^2 : |\xi_2| < |\xi_1|, \frac{1}{2} < |\xi_1| < 1 \right\},
\]

as well as the decay condition

\[
\left| \left( \frac{\partial}{\partial \xi_2} \right)^{\ell} \hat{g}(\xi) \right| \lesssim \frac{\min\{1, |\xi_1|^{\alpha}\}}{(1 + |\xi_1|)^{\beta}(1 + |\xi_2|)^{\beta}} \quad \text{for} \ \ell = 0, 1,
\]

with \( \alpha \) and \( \beta \) chosen as before (i.e., \( \rho \in (0, \frac{1}{12}) \), \( \alpha \geq \frac{6}{\rho} + 1 \), and \( \beta > \alpha + 1 \)).

At this point, we pause in the description of the construction, and first observe the following frame-type equation, which follows from our choices. Notice that this result already combines systems for the horizontal and vertical cones. For the proof, we refer to [15].
Fig. 2 Frequency tiling with $\hat{G}_s$ for $s = 0, \frac{1}{3}, \frac{2}{3}, \frac{3}{3}$ (for $\xi_1 > 0$)

Lemma 2.3 ([15]) Letting $\varphi_1, \psi_1 \in L^2(\mathbb{R})$ and $g \in L^2(\mathbb{R}^2)$ be defined as before, we have

$$0 < A \leq |\hat{\varphi}^0(\xi)|^2 + \sum_{j \geq 0} \sum_{|k| \leq 2^{j/2}} |\hat{g}(S_{-k}A_j^{-1}\xi)|^2 \leq B < \infty \text{ for a.e. } \xi \in \mathbb{R}^2,$$

where $\varphi^0(x) = \varphi_1(x_1)\varphi_1(x_2)$.

The filters $G_s, s = s([j_0/2], q_0) \in \mathbb{S}$ are then defined by

$$\hat{G}_0(\xi) = |\hat{\varphi}^0(\xi)|^2 + \sum_{j=0}^{\infty} |\hat{g}(A_j^{-1}\xi)|^2 \quad \text{and} \quad \hat{G}_s(\xi) = \sum_{j=j_0}^{\infty} |\hat{g}(A_j^{-1}S_{-s}^T\xi)|^2 \text{ for } s \neq 0. \tag{8}$$

Figure 2 illustrates the frequency tiling by the essential supports of $\hat{G}_s$, showing the wedgelike shape geometry.

The following result provides an identity that will be a key ingredient to prove the frame property of the dualizable shearlet systems in Theorem 3.1.

Lemma 2.4 Let $G_s, s \in \mathbb{S}$ be defined as in (8). Then

$$\sum_{s \in \mathbb{S}} \hat{G}_s(\xi) = |\hat{\varphi}^0(\xi)|^2 + \sum_{j \geq 0} \sum_{|k| \leq 2^{j/2}} |\hat{g}(S_{-k}A_j^{-1}\xi)|^2 \text{ for a.e. } \xi \in \mathbb{R}^2.$$

Proof This follows directly from the definition of the filters and the set $\mathbb{S}$. \qed

Finally, after this preparation, we can now formally define dualizable shearlet systems by filtering the shearlet-type wavelet systems defined in Definition 2.1 using the filters $G_s, s \in \mathbb{S}$.

Definition 2.5 For any $s \in \mathbb{S}$, let $\Psi_s(\varphi_1, \psi_1)$ be the shearlet-type wavelet system, and let $G_s$ be the filter generated by $g \in L^2(\mathbb{R}^2)$ as defined in (8). Then the dualizable shearlet system $\mathcal{SH}(\varphi_1, \psi_1; g)$ is defined by

\[\text{Springer}\]
\[ S\mathcal{H}(\varphi_1, \psi_1; g) = \left\{ \psi^\ell_\lambda : \lambda \in \Lambda_s, s \in \mathbb{S}, \ell = 0, 1 \right\}, \]

with index set
\[ \Lambda_s = \{ (j, s, m, p) : j \in \{-1\} \cup \{j_0, j_0 + 1, \ldots\}, m \in \mathbb{Z}^2, p \geq 0 \} \quad \text{for} \quad s = s(\lceil j_0/2 \rceil, q_0), \]

where
\[ \psi^0_\lambda = \begin{cases} G_s \ast \varphi_{j_0, s, m, p} & : \lambda = (-1, s, m, p) \in \Lambda_s, \\ G_s \ast \psi_{j, s, m, p} & : \lambda = (j, s, m, p) \in \Lambda_s, \end{cases} \]

and \( \psi^1_\lambda = \psi^0_\lambda \circ R. \)

We immediately observe that the constructed system is compactly supported.

**Lemma 2.6** Each dualizable shearlet system is compactly supported.

**Proof** Let \( S\mathcal{H}(\varphi_1, \psi_1; g) \) be a dualizable shearlet system. Then by construction, there exist \( C_1, C_2 > 0 \) such that, for all \( s = s(\lceil j_0/2 \rceil, q_0) \in \mathbb{S} \) and \( j \geq j_0, m \in \mathbb{Z}^2, p \geq 0 \), the filters \( G_s \) are compactly supported with
\[ \text{supp}(G_s) \subset S_s^{-1} A^{-1}_{j_0} (\{-C_1, C_1\}^2) \]

and the elements of the shearlet-type wavelet systems satisfy
\[ \text{supp}(\varphi_{j_0, s, m, p}) \subset S_s^{-1} A^{-1}_{j_0} (\{-C_1, C_1\}^2), \quad \text{supp}(\psi_{j, s, m, p}) \subset S_s^{-1} A^{-1}_{j} (\{-C_1, C_1\}^2). \]

Hence, there exists some \( C > 0 \) such that
\[ \text{supp}(G_s \ast \varphi_{j_0, s, m, p}), \text{supp}(G_s \ast \psi_{j, s, m, p}) \subset S_s^{-1} A^{-1}_{j_0} (\{-C_1, C_1\}^2), \]

which proves the claim. \( \square \)

### 2.3 Comparison with Customarily Defined Shearlet Systems

We now aim to justify the term “shearlets” by rewriting the elements of \( S\mathcal{H}(\varphi_1, \psi_1; g) \) such that the resemblance with cone-adapted shearlet systems (cf. Sect. 1.4) is revealed. We will observe that the dualizable shearlet system consists of functions of the form contained in the original shearlet system except for the oversampling matrix \( D_p \) for \( p \geq 0 \). As already mentioned before, this ingredient ensures that a dualizable shearlet system is composed of subsystems that are filtered versions of orthonormal bases. This structure will be key to have a closed form for an associated dual frame (see Theorem 3.1).

**Proposition 2.7** Let \( S\mathcal{H}(\varphi_1, \psi_1; g) \) be a dualizable shearlet system. Define
\[ \hat{\Theta}(\xi) = |\hat{\varphi}^0(\xi)|^2 + \sum_{j=0}^{\infty} |\hat{g} \left( A^{-1}_j \xi \right)|^2 \quad \text{and} \quad \hat{\Theta}_\ell(\xi) = \sum_{j=-\ell}^{\infty} |\hat{g} \left( A^{-1}_j \xi \right)|^2, \quad \ell \geq 0. \]
Then, for the elements of $\mathcal{SH}(\varphi_1, \psi_1; g)$, the following hold:

(i) For all $m \in \mathbb{Z}^2$ and $p \geq 0$,

$$
\psi_\lambda^0 = (\Theta \ast \varphi^p)(-D_pm), \quad \text{with } \lambda = (-1, 0, m, p) \in \Lambda_0,
$$

and for all $j \geq 0$, $m \in \mathbb{Z}^2$, and $p \geq 0$,

$$
\psi_\lambda^0 = |\det(A_j)|^{1/2}(\Theta \ast \varphi^p)(A_j \cdot -D_pm), \quad \text{with } \lambda = (j, 0, m, p) \in \Lambda_0.
$$

(ii) Letting $s = s(\lceil j_0/2 \rceil, q_0) \in \mathbb{S} \setminus \{0\}$, for all $k \in \mathbb{Z}$ with $s = \frac{k}{2^{\lceil j_0/2 \rceil}}$ and for all $m \in \mathbb{Z}^2$, $p \geq 0$,

$$
\psi_\lambda^0 = |\det(A_{j_0})|^{1/2}(\Theta_0 \ast \varphi^p)(S_k A_{j_0} \cdot -D_pm), \quad \text{with } \lambda = (-1, s, m, p) \in \Lambda_s,
$$

and for all $j \geq j_0$ and $k \in \mathbb{Z}$ with $s = \frac{k}{2^{\lceil j/2 \rceil}}$, and for all $m \in \mathbb{Z}^2$, $p \geq 0$,

$$
\psi_\lambda^0 = |\det(A_j)|^{1/2}(\Theta_{j-j_0} \ast \varphi^p)(S_k A_j \cdot -D_pm), \quad \text{with } \lambda = (j, s, m, p) \in \Lambda_s.
$$

Proof: We will only consider the last equation in (ii) for $\psi_\lambda^0$. The other cases can be derived similarly with minor modifications for notation. First note that for each $(j, k) \in \mathbb{N}_0 \times \{-2^{\lceil j/2 \rceil}, \ldots, 2^{\lceil j/2 \rceil}\} \setminus \{0\}$, there exists a unique shear parameter $s(\lceil j_0/2 \rceil, q_0) \in \mathbb{S}$ with $j \geq j_0$ and $k = (2^{\lceil j/2 \rceil} - \lceil j_0/2 \rceil)q_0$. This ensures that

$$
A_j^{-1} S_{-s}^T = S_{-k} A_j^{-1}.
$$

Using this relation, we obtain

$$
\begin{align*}
G_s \ast \psi_{j,s,m,p}(\xi) &= \sum_{j'=j_0}^{\infty} |\hat{g}(A_{j'}^{-1} S_{-s}^T \xi)|^2 |\det(A_{j'})|^{-1/2} \hat{\psi}^p (A_{j'}^{-1} S_{-s}^T \xi) e^{-2\pi i \langle m, D_p A_{j'}^{-1} S_{-s}^T \xi \rangle} \\
&= |\det(A_{j_0})|^{-1/2} \hat{\Theta}_{j-j_0} (S_{-k} A_j^{-1} \xi) \hat{\psi}^p (S_{-k} A_j^{-1} \xi) e^{-2\pi i \langle m, D_p S_{-k} A_j^{-1} \xi \rangle}.
\end{align*}
$$

Application of the inverse Fourier transform and careful consideration of the different cases yield the claim.

3 A Dual of Dualizable Shearlet Frames

Dualizable shearlet systems are foremost designed to provide a closed, easily computable form of a dual frame while still delivering near optimal sparse approximation of cartoon-like functions. The first item will now be formally stated and proved.
Theorem 3.1 Let $\mathcal{SH}(\varphi_1, \psi_1; g) = \{\psi_\lambda^\ell : \lambda \in \Lambda_s, s \in S, \ell = 0, 1\}$ be a dualizable shearlet system that constitutes a frame for $L^2(\mathbb{R}^2)$. Then

$$\tilde{\mathcal{SH}}(\varphi_1, \psi_1; g) = \left\{ \tilde{\psi}_\lambda^\ell : \lambda \in \Lambda_s, s \in S, \ell = 0, 1 \right\}$$

is a dual frame for $\mathcal{SH}(\varphi_1, \psi_1; g)$, where, for $\lambda \in \Lambda_s$,

$$\hat{\tilde{\psi}}_0 = \sum_{s' \in S} |\hat{G}_{s'}|^2 + |\hat{G}_{s'} \circ R|^2$$

and

$$\tilde{\psi}_1 = \hat{\tilde{\psi}}_0 \circ R.$$  \hfill (9)

Proof For the proof, we use the convention that $g(\cdot) = g(-\cdot)$. We first observe that the structure of a dualizable shearlet system allows a decomposition as

$$\sum_{s \in S} \sum_{\lambda \in \Lambda_s} |\langle f, \psi_\lambda^0 \rangle|^2 = \int_{\mathbb{R}^2} \left( \sum_{s \in S} |\hat{G}_s|^2 \right) |\hat{f}(\xi)|^2 d\xi.$$  \hfill (10)

Similarly, we can show that

$$\sum_{s \in S} \sum_{\lambda \in \Lambda_s} |\langle f, \psi_\lambda^1 \rangle|^2 = \int_{\mathbb{R}^2} \left( \sum_{s \in S} |\hat{G}_s(R\xi)|^2 \right) |\hat{f}(\xi)|^2 d\xi.$$  \hfill (11)

By Lemmata 2.3 and 2.4, it follows that

$$\sum_{s \in S} \left( |\hat{G}_s(\xi)|^2 + |\hat{G}_s(R\xi)|^2 \right) \leq 2B^2 < \infty \quad \text{for a.e. } \xi.$$ 

Combining this inequality with (10) and (11) implies the existence of an upper frame bound for $\mathcal{SH}(\varphi_1, \psi_1; g)$.

To derive a lower frame bound, we use the support conditions on $\varphi_1$ and $g$, namely (4) and (6), which imply

$$\sum_{s \in S} \left( |\hat{G}_s(\xi)|^2 + |\hat{G}_s(R\xi)|^2 \right) \geq \left( \chi_{[-1/2,1/2]} + \sum_{j \geq 0, |k| \leq 2^{j/2}} \left( \chi_{k/2^j} \hat{\chi}_{k/2^j A_j \Omega_g} + \chi_{k/2^j} \hat{\chi}_{k/2^j A_j \Omega_g} \circ R \right) \right) \cdot (\min\{\delta_{\varphi_1}, \delta_g\})^2$$

$$\geq (\min\{\delta_{\varphi_1}, \delta_g\})^2 > 0 \quad \text{for a.e. } \xi.$$
The frame property of $\mathcal{SH}(\varphi_1, \psi_1; g)$ can be shown by similar arguments.

It remains to prove that $\mathcal{SH}(\varphi_1, \psi_1; g)$ indeed forms a dual frame of $\mathcal{SH}(\varphi_1, \psi_1; g)$. For this, we use the structure of the system $\mathcal{SH}(\varphi_1, \psi_1; g)$ to obtain

\[
\sum_{\lambda \in \Lambda_s} \sum_{s \in S} \left\langle \hat{f}, \hat{\psi}_{\lambda}^0 \right\rangle \hat{\psi}_{\lambda}^0 = \sum_{s = (j_0, q_0) \in \mathbb{S}} \frac{\hat{G}_s}{\sum_{s' \in S} \left| \hat{G}_{s'} \right|^2 + \left| \hat{G}_{s'} \circ R \right|^2} \times \left( \sum_{m \in \mathbb{Z}^2} \sum_{p \in \mathbb{N}_0} \left\langle \hat{G}_s \hat{f}, \hat{\varphi}_{j_0, s, m, p} \right\rangle \hat{\varphi}_{j_0, s, m, p} + \sum_{j = j_0}^{\infty} \sum_{m \in \mathbb{Z}^2} \sum_{p \in \mathbb{N}_0} \left\langle \hat{G}_s \hat{f}, \hat{\psi}_{j, s, m, p} \right\rangle \hat{\psi}_{j, s, m, p} \right).
\]

Lemma 2.2 again implies

\[
\sum_{\lambda \in \Lambda_s} \sum_{s \in S} \left\langle \hat{f}, \hat{\psi}_{\lambda}^0 \right\rangle \hat{\psi}_{\lambda}^0 = \sum_{s \in S} \frac{\left| \hat{G}_s \right|^2 \cdot \hat{f}}{\sum_{s' \in S} \left( \left| \hat{G}_{s'} \right|^2 + \left| \hat{G}_{s'} \circ R \right|^2 \right)}.
\]

Similarly,

\[
\sum_{\lambda \in \Lambda_s} \sum_{s \in S} \left\langle \hat{f}, \hat{\psi}_{\lambda}^1 \right\rangle \hat{\psi}_{\lambda}^1 = \sum_{s \in S} \frac{\left| \hat{G}_s \circ R \right|^2 \cdot \hat{f}}{\sum_{s' \in S} \left( \left| \hat{G}_{s'} \right|^2 + \left| \hat{G}_{s'} \circ R \right|^2 \right)}.
\]

Using the filter properties as well as (12) and (13) finally yields

\[
\sum_{\ell = 0}^{1} \sum_{s \in S} \sum_{\lambda \in \Lambda_s} \left\langle \hat{f}, \hat{\psi}_{\lambda}^{\ell} \right\rangle \hat{\psi}_{\lambda}^{\ell} = \hat{f}.
\]

The theorem is proved.

We remark that the dual frame does not form a (dualizable) shearlet system. However, each dual shearlet frame element $\hat{\psi}_{\lambda}^{\ell}$ defined in (9) with $\ell = 0, 1$ can be efficiently implemented for $N \times N$ digital images in the frequency domain without inverting the frame operator implicitly. In fact, the term $\hat{g}(A_j^{-1} S_{-s} T^T \xi)$ in (8) can be faithfully digitized by taking the discrete Fourier transform of the digital shearlet filters developed in [17], which allows for faithfully discretizing $\hat{G}_s$ for $N \times N$ digital images in the frequency domain. For this, we can first take the digital shearlet filters up to a finite scale the $j = J$ for computing $\hat{g}(A_j^{-1} S_{-s} T^T \xi)$ in the digital domain and then take the finite sum over the scales $j = j_0 \ldots J$ in (8). The same procedure can be applied for $\hat{G}_s \circ R$ in (9) except for switching the variables. This allows us to then explicitly discretize

\[
\hat{G} = \sum_{s \in S} \left( \left| \hat{G}_s \right|^2 + \left| \hat{G}_s \circ R \right|^2 \right).
\]
and $1/\hat{G}$ for $N \times N$ images in the frequency domain. We remark that one can ensure $\inf_\xi \hat{G}(\xi) > 0$ with suitable digital shearlet filters—see [17].

Then the $N \times N$ digital dual shearlet filters for $\tilde{\psi}_\lambda^\ell$ are obtained by taking the inverse discrete Fourier transform of $\hat{\psi}_\lambda^\ell / \hat{G}$. Moreover, reconstruction from the shearlet coefficients $\langle f, \psi_\lambda^\ell \rangle$ can be obtained by

$$f = \mathcal{F}^{-1} \left( \frac{1}{G} \cdot \mathcal{F} \left( \sum_{\ell=0}^{1} \sum_{\lambda} \langle f, \psi_\lambda^\ell \rangle \psi_\lambda^\ell \right) \right),$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier transform and its inverse, respectively.

### 4 Sparse Approximation Properties

We now turn to analyzing sparse approximation properties of dualizable shearlet systems with respect to anisotropic features that are modeled by the class of cartoon-like functions. We start by formally introducing this class, which was first defined in [9]. We remark that the superscript 2 in $\mathcal{E}^2(\mathbb{R}^2)$ is due to the fact that the discontinuity curve is assumed to be $C^2$. Generalizations of cartoon-like functions with different types of regularity can, for instance, be found in [14].

**Definition 4.1** The set of *cartoon-like functions* $\mathcal{E}^2(\mathbb{R}^2)$ is defined by

$$\mathcal{E}^2(\mathbb{R}^2) = \{ f \in L^2(\mathbb{R}^2) : f = f_0 + f_1 \cdot \chi_B \},$$

where $B \subset [0, 1]^2$ is a nonempty, simply connected set with $C^2$-boundary, $\partial B$ has bounded curvature, and $f_i \in C^2(\mathbb{R}^2)$ satisfies $\text{supp} f_i \subseteq [0, 1]^2$ and $\|f_i\|_{C^2} \leq 1$ for $i = 0, 1$.

We now let

$$\Lambda := \{0, 1\} \times \bigcup_{s \in \mathcal{S}} \Lambda_s,$$

which is the index set for $\mathcal{S} \mathcal{H}(\varphi_1, \psi_1; g)$. Given a dualizable shearlet system

$$\mathcal{S} \mathcal{H}(\varphi_1, \psi_1; g) = \left\{ \psi_\lambda^\ell : \lambda \in \Lambda_s, s \in \mathcal{S}, \ell = 0, 1 \right\},$$

with associated dual frame

$$\mathcal{S} \mathcal{H}(\varphi_1, \psi_1; g) = \left\{ \tilde{\psi}_\lambda^\ell : \lambda \in \Lambda_s, s \in \mathcal{S}, \ell = 0, 1 \right\}$$

as defined in Theorem 3.1, we are interested in $N$-term approximations of $f \in \mathcal{E}^2(\mathbb{R}^2)$ of the form

$$f_N = \sum_{(\ell, \lambda) \in \Lambda_N} \langle f, \psi_\lambda^\ell \rangle \tilde{\psi}_\lambda^\ell.$$
where $\Lambda_N \subseteq \Lambda$, $\# \Lambda_N = N$. Let us remind the reader that we choose expansions in terms of the dual frame, since applications usually require reconstruction from the frame coefficients $(\langle f, \psi^\ell_\lambda \rangle)_{\ell,\lambda}$ (cf. Sect. 1).

Without loss of generality, we will only consider shearlet elements $\psi^0_\lambda \in SH(\phi_1, \psi_1; g)$ associated with one frequency cone. Since the elements $\psi^1_\lambda$ just arise when switching the variables, they can be dealt with similarly. Hence, for the sake of brevity, from now on we omit the superscript “0”; i.e., we write

$$
\psi_\lambda := \psi^0_\lambda \text{ and } \tilde{\psi}_\lambda := \tilde{\psi}^0_\lambda.
$$

Next we recall that the optimally achievable approximation rate, i.e., a benchmark for any frame for $L^2(\mathbb{R}^2)$, was also derived in [9].

**Theorem 4.2** ([9]) Let $(h_i)_{i \in I} \subseteq L^2(\mathbb{R}^2)$ be a frame for $L^2(\mathbb{R}^2)$. Then, for any $f \in E^2(\mathbb{R}^2)$, the $L^2$-error of best $N$-term approximation by $f_N$ with respect to $(h_i)_{i \in I}$ satisfies

$$
\| f - f_N \|_2 \gtrsim N^{-1} \text{ as } N \to \infty.
$$

The following result shows that the approximation rate of dualizable shearlets for cartoon-like functions can be arbitrarily close to the optimal rate as the smoothness of the generators is increased, i.e., as $\rho \to 0$.

**Theorem 4.3** Let $SH(\phi_1, \psi_1; g) = \{ \psi^\ell_\lambda : \lambda \in \Lambda_3, s \in S, \ell = 0, 1 \}$ be a dualizable shearlet system, and let $\rho$ be the smoothness parameter defined in (5) and (7) associated with $\phi_1, \psi_1, g$. Further, let $f \in E^2(\mathbb{R}^2)$. Then

$$
\| f - f_N \|_2 \lesssim N^{-1+\rho} \cdot \log(N) \text{ as } N \to \infty,
$$

where $f_N = \sum_{(\ell, \lambda) \in \Lambda_N} (\langle f, \psi^\ell_\lambda \rangle \tilde{\psi}^\ell_\lambda$ with $\Lambda_N \subseteq \Lambda$, $\# \Lambda_N = N$ is the $N$-term approximation using the $N$ largest coefficients $(\langle f, \psi^\ell_\lambda \rangle)_{\ell,\lambda}$.

Before we discuss the overall structure and details of the proof, we would like to highlight that using this new proof technique, even previous results can be improved. In fact, we can lower the exponent of the log-factor in the decay rate of the compactly supported shearlet system defined in [16] from $\log(N)^{3/2}$ to $\log(N)$.

**Theorem 4.4** For each $f \in E^2(\mathbb{R}^2)$, the compactly supported shearlet system defined in [16] provides an approximation rate of

$$
\| f - f_N \|_2 \lesssim N^{-1} \cdot \log(N) \text{ as } N \to \infty
$$

with $f_N$ being the $N$-term approximation consisting of the $N$ largest shearlet coefficients.

The proof of this theorem follows the proof of Theorem 4.3 quite closely except for slight modifications that we describe in Sect. 6.6.
5 Proof of Theorem 4.3

Since the proof is rather technical and complex, we start by discussing its overall architecture. We recall from Proposition 2.7 that for all $\lambda = (j, s, m, p) \in \Lambda_s$, $s = s([j_0/2], q_0) \in \mathbb{S}$, $j \geq j_0$ and $k \in \mathbb{Z}$ with $s = \frac{k}{2^{j_0+1}}$,

$$\psi_\lambda = |\det(A_j)|^{1/2}(\Theta_{j-j_0} \ast \psi^p)(S_k A_j \cdot -D_pm).$$

(14)

We mention that without loss of generality, we only need to consider shearlet elements of this form. Nearly identical arguments can be applied for the elements $\psi_\lambda$ with $\lambda = (-1, s, m, p) \in \Lambda_s$ with minor modifications for notation.

One might think that due to the fact that dualizable shearlets have this strong structural similarity with “standard” shearlets, the steps of the proof of the (optimal sparse) approximation result from [16] could be directly applied. This is however not the case. Although, in the end, we will be able to utilize some of those steps, careful preparation for this is required. Moreover, it will turn out that we will eventually even improve the approach from [16] in the sense of Theorem 4.4, i.e., by reducing the number of log-factors.

In a first step, we prove two basic estimates for the shearlet coefficients, namely for an $L^\infty$ and for an $L^2$ function. This is made precise in the following lemma, whose technically natured proof can be found in Sect. 6.2.

Lemma 5.1

(i) For $f \in L^\infty(\mathbb{R}^2)$, we have

$$|\langle f, \psi_\lambda \rangle| \lesssim 2^{-\frac{j}{2}} \cdot \|f\|_\infty$$

for all $\lambda = (j, s, m, p) \in \Lambda$.

(ii) For $f \in L^2(\mathbb{R}^2)$, we have

$$|\langle f, \psi_\lambda \rangle| \lesssim 2^{-\frac{js}{2}} \cdot \|f\|_2$$

for all $\lambda = (j, s, p, m) \in \Lambda$.

One main difficulty in proving this result is the analysis of the function $\Theta_{j-j_0} \ast \psi^p$, which is now the generator of the shearlet element in (14). In fact, we require a universal upper bound for this function, which is given by the following result. For its proof, we refer to Sect. 6.1.

Lemma 5.2 There exists a universal constant $C$ such that

$$\|\Theta_{j-j_0} \ast \psi^p\|_1 \leq C \quad \text{for all} \quad j \geq j_0, p \geq 0.$$
for scale \( j = j_0 \). However, \( \text{supp}(\psi_\lambda) \) is much larger when \( j \gg j_0 \). To resolve this issue for our newly defined dualizable shearlets \( \psi_\lambda \), we will approximate \( \psi_\lambda \) by more suitable functions of smaller supports comparable to the size of the supports of “standard” shearlets with controllable error bound. This is the essence of the following result, whose proof is outsourced to Sect. 6.3. For the following lemmata, the parameter \( \rho \) defined in (5) will be used.

**Lemma 5.3** Let \( \lambda = (j, s, m, p) \in \Lambda_s \) with \( s = s(\lceil j_0/2\rceil, q_0) \in \mathbb{S} \), \( j \geq j_0 \), and set

\[
\hat{\psi}_\lambda^\rho := \sum_{j' = \max(\lceil j(1-\rho)\rceil, j_0)}^\infty \abs{\hat{g}(A_{j'}^{-1}S_{-s}^T\xi)}^2 \hat{\psi}_{j,s,m,p}.
\]

Then the following hold:

(i) There exists some \( C > 0 \) such that

\[
\text{supp}(\psi_\lambda^\rho) \subset S_{-s}^{-1}A_j^{-1}(A_{j_0}([-C, C]^2) + D \rho m).
\]

(ii) We have

\[
\abs{(f, (\psi_\lambda - \psi_\lambda^\rho))} \lesssim 2^{-j\rho a}2^{-\rho a} \| f \|_2 \text{ for all } f \in L^2(\mathbb{R}^2).
\]

The second key condition is a directional vanishing moment condition, which can be shown to be fulfilled by the generators \( \Theta_{j-\max(\lceil j(1-\rho)\rceil, j_0)} * \psi^\rho \) for \( \psi_\lambda^\rho \). In fact, following the same argument as in the proof of Proposition 2.7(ii), we see that

\[
\psi_\lambda^\rho = \det(A_j)^{1/2}(\Theta_{j-\max(\lceil j(1-\rho)\rceil, j_0)} * \psi^\rho)(S_k A_j \cdot -D \rho m)
\]

for \( \lambda = (j, s, m, p) \in \Lambda_s \), \( s = s(\lceil j_0/2\rceil, q_0) \in \mathbb{S} \), \( j \geq j_0 \) and \( k \in \mathbb{Z} \) with \( s = k2^{-j+j_0} \).

The proof of the following result is provided in Sect. 6.4.

**Lemma 5.4** For all \( j \geq j_0 \), \( p \geq 0 \), \( \ell = 0, 1 \), and \( \gamma \in [1, \alpha) \),

\[
\abs{\left( \frac{\partial}{\partial \xi_2} \right)^\ell (\Theta_{j-\max(\lceil j(1-\rho)\rceil, j_0)} * \psi^\rho)(\xi)} \lesssim \frac{\min\{1, |\xi_1|^\alpha-1\}}{(1 + |\xi_1|)\rho - \gamma (1 + |\xi_2|)^\gamma}.
\]

One main last ingredient, which we state as a lemma before providing the complete proof of Theorem 4.3, are decay rates of the shearlet coefficients \( (f, \psi_\lambda) \) for cartoon-like functions, where we now carefully insert conditions related to the functions \( \psi_\lambda^\rho \).

Again, the proof can be found in Sect. 6.5.

**Lemma 5.5** Assume that \( f \in E^2(\mathbb{R}^2) \) with \( C^2 \) discontinuity curve given by \( x_1 = E(x_2) \). For \( \psi_\lambda \in S^H(\phi_1, \psi_1; g) \) with \( \lambda = (j, s, m, p) \in \Lambda_s \), \( s = s(\lceil j_0/2\rceil, q_0) \in \mathbb{S} \) and \( j \geq j_0 \), let \( \psi_\lambda^\rho \in L^2(\mathbb{R}^2) \) be defined as in Lemma 5.4. Let \( \hat{x}_2 \in \mathbb{R} \) so that \( (E(\hat{x}_2), \hat{x}_2) \in \text{supp}(\psi_\lambda^\rho) \) and \( \hat{s} = E' (\hat{x}_2) \). Also let \( k_3 \in \mathbb{Z} \) so that \( s = k_32^{-j+j_0} \). Then the following hold:
(i) If $|\hat{s}| \leq 3$, then
\[
|\langle f, \psi_\lambda \rangle| \lesssim \min \left\{ 2^{-\frac{3}{4}j}, \frac{2^{-\frac{3}{4}j} 2^{3j}}{|k_x + 2^{j/2} |\hat{s}|^3} \right\},
\]

(ii) If $|\hat{s}| > \frac{3}{2}$, then
\[
|\langle f, \psi_\lambda \rangle| \lesssim 2^{3j} 2^{-\frac{9}{4}j}.
\]

After these strategic discussions, we are now ready to present the proof of Theorem 4.3.

**Proof of Theorem 4.3** We start by defining dyadic cubes $Q_{j,\ell} \subseteq [0, 1]^2$ for $j \geq 0$ and $\ell \in \mathbb{Z}^2$ by setting
\[
Q_{j,\ell} := 2^{-\lfloor j/2 \rfloor} [0, 1]^2 + 2^{-\lfloor j/2 \rfloor} \ell.
\]

The set of dyadic cubes intersecting the discontinuity curve $\Gamma$ of $f \in \mathcal{E}^2(\mathbb{R}^2)$ is then given by
\[
Q_j = \{ Q_{j,\ell} : \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset \},
\]
where int$(Q_{j,\ell})$ is the interior set of $Q_{j,\ell}$.

Next, without loss of generality, we assume that the discontinuity curve $\Gamma$ is given by $x_1 = E(x_2)$ with $E \in C^2([0, 1])$. In fact, for sufficiently large $j$, the discontinuity $\Gamma$ can be expressed as either $x_1 = E(x_2)$ or $x_2 = \tilde{E}(x_1)$ within $Q_{j,\ell} \in Q_j$. Hence the same arguments can be applied for $x_2 = \tilde{E}(x_1)$ except for switching the order of variables. For each $Q_{j,\ell} \in Q_j$, now let $E_{j,\ell}$ be a $C^2$ function such that
\[
\Gamma \cap \text{int}(Q_{j,\ell}) = \{(x_1, x_2) \in \text{int}(Q_{j,\ell}) : x_1 = E_{j,\ell}(x_2)\}.
\]

This allows us to define
\[
Q_j^0 := \{ Q_{j,\ell} \in Q_j : \|E'_{j,\ell}\|_\infty \leq 3 \}
\]
and
\[
Q_j^1 := Q_j \cap (Q_j^0)^c.
\]

Notice that, for all $E_{j,\ell}$ associated with $Q_{j,\ell} \in Q_j^1$, we may assume
\[
\inf_{(x_1, x_2) \in \text{int}(Q_{j,\ell})} |E'_{j,\ell}(x_2)| > 3/2
\]
for sufficiently large $j$. 
We further define the orientation of the discontinuity curve $\Gamma$ in each dyadic cube $Q_{j,\ell}$ by

$$\hat{s}_{j,\ell} = E'_{j,\ell}(\hat{x}_2)$$

for some $(E_{j,\ell}(\hat{x}_2), \hat{x}_2) \in \text{int}(Q_{j,\ell}) \cap \Gamma$.

Moreover, for any $J > 0$, we define $S_{J/2}$ as a finite subset of $S$ by

$$S_{J/2} = \{s(j, q) = 0 : j = 0, q = 0\} \cup \{s(\lfloor j/2\rfloor, q) : 0 \leq j \leq J, |q| \leq 2^{\lfloor j/2\rfloor}, q \in 2\mathbb{Z} + 1\}.$$ 

Finally, let $k_{j,s} \in \mathbb{Z}$ be chosen so that

$$s = \frac{k_{j,s}}{2^{\lfloor j/2\rfloor}}$$

for $s = s(\lfloor j_0/2\rfloor, q_0) \in S_{J/2}$ and $j \geq j_0$.

We will now separately consider the two cases, namely when the supports of approximated shearlets $\psi^\sharp_\lambda$ intersect the discontinuity curve of $f$ or not. For this, we define subsets $\Lambda^0$ and $\Lambda^1$ of the general index set $\Lambda = \{0, 1\} \times \bigcup_{s \in S} \Lambda_s$ by

$$\Lambda^0 = \{\lambda \in \Lambda : \text{int}(\text{supp}(\psi^\sharp_\lambda)) \cap \Gamma \neq \emptyset\} \quad \text{and} \quad \Lambda^1 = \Lambda \cap (\Lambda^0)^c.$$ 

For the case when $\text{supp}(\psi^\sharp_\lambda)$ does not intersect the discontinuity of $f$, we define subsets $\Lambda^{1,0}$ and $\Lambda^{1,1}$ of $\Lambda^1$ so that $\Lambda^1 = \Lambda^{1,0} \cup \Lambda^{1,1}$ by

$$\Lambda^{1,0} = \bigcup_{j,p \geq 0} \Lambda^{1,0}_{j,p} \quad \text{and} \quad \Lambda^{1,1} = \bigcup_{j,p \geq 0} \Lambda^{1,1}_{j,p},$$

where

$$\Lambda^{1,0}_{j,p} = \{\lambda = (j', s, m, p') \in \Lambda^1 : \text{int}(\psi^\sharp_\lambda) \cap \text{int}(\text{supp}(f)) = \emptyset, j' = j, p' = p\}$$

and

$$\Lambda^{1,1}_{j,p} = \{\lambda = (j', s, m, p') \in \Lambda^1 : \text{int}(\psi^\sharp_\lambda) \cap \text{int}(\text{supp}(f)) \neq \emptyset, j' = j, p' = p\}.$$ 

The smooth part, i.e., shearlet coefficients not intersecting the discontinuity curve $\Gamma$, can now be handled as follows. For this, we will estimate

$$\sum_{\lambda \in (\Lambda_N)^c \cap \Lambda^{1,r}} |\langle f, \psi_\lambda \rangle|^2$$

as $N \to \infty$ (16)

for $r = 0, 1$, where $\Lambda_N$ is the index set for the $N$ largest shearlet coefficients in magnitude. We may assume that $\Lambda$ only consists of indices $\lambda \in \Lambda$ satisfying $\text{int}(\text{supp}(\psi_\lambda)) \cap \text{int}(\text{supp}(f)) \neq \emptyset$. This assumption is reasonable, since $(f, \psi_\lambda) = 0$ if $\text{int}(\text{supp}(\psi_\lambda)) \cap \text{int}(\text{supp}(f)) = \emptyset$. Note that there exist about $2^{2j+p}$ shearlets $\psi_\lambda$ with $\lambda = (j, s, m, p) \in \Lambda$ for each fixed $j$ and $p \geq 0$, since $|s| \lesssim 2^{j/2}$. Moreover,
there exist about \(|A_j| \cdot |D_p^{-1}| = 2^{\frac{s}{2}} 2^p\) translates for each shear parameter \(s\). This implies
\[
\sharp(A_{j,p}^{1,r}) \lesssim 2^{2j+p} \quad \text{for} \quad r = 0, 1. \tag{17}
\]

We now define
\[
\tilde{\Lambda}_N^0 := \{\lambda = (j, s, m, p) \in \Lambda^{1,0} : j \leq J/8, p \leq J/4\},
\]
with \(N\) chosen as \(N = \sharp(\tilde{\Lambda}_N^0)\). Notice that, by (17), we have \(N \sim 2^{J/2}\).

Next, we show that
\[
\sum_{\lambda \in (\tilde{\Lambda}_N \cap \Lambda^{1,0})} |\langle f, \psi_{\lambda} \rangle|^2 \lesssim \sum_{\lambda \in (\tilde{\Lambda}_N^0 \cap \Lambda^{1,0})} |\langle f, \psi_{\lambda} \rangle|^2 \lesssim 2^{-j} \quad \text{as} \quad J \to \infty. \tag{18}
\]

Since the first inequality is obvious, it remains to prove the second inequality in (18). For this, using Lemma 5.3 (ii) and (17), we obtain
\[
\sum_{\lambda \in (\tilde{\Lambda}_N^0 \cap \Lambda^{1,0})} |\langle f, \psi_{\lambda} \rangle|^2 \leq \sum_{j > J/8} \sum_{p > J/4} \sum_{\lambda \in \Lambda^{1,0}} |\langle f, \psi_{\lambda} \rangle|^2 + \sum_{j \leq J/8} \sum_{p > J/4} \sum_{\lambda \in \Lambda^{1,0}} |\langle f, \psi_{\lambda} \rangle|^2
\]
\[
+ \sum_{j > J/8} \sum_{p \leq J/4} \sum_{\lambda \in \Lambda^{1,0}} |\langle f, \psi_{\lambda} \rangle|^2
\]
\[
\lesssim \sum_{j > J/8} \sum_{p > J/4} 2^{2j+p} 2^{-2\alpha j p} 2^{-2\alpha p}
\]
\[
+ \sum_{j \leq J/8} \sum_{p > J/4} 2^{2j+p} 2^{-2\alpha j p} 2^{-2\alpha p}
\]
\[
+ \sum_{p \leq J/4} \sum_{j > J/8} 2^{2j+p} 2^{-2\alpha j p} 2^{-p}
\]
\[
\lesssim \sum_{j > J/8} \sum_{p > J/4} 2^{-2(\alpha - 1)p} 2^{-j(2\alpha p - 2)}
\]
\[
+ \sum_{j \leq J/8} \sum_{p > J/4} 2^{-2(\alpha - 1)p} 2^{-j(2\alpha p - 2)}
\]
\[
+ \sum_{p \leq J/4} \sum_{j > J/8} 2^{-2(\alpha - 1)p} 2^{-j(2\alpha p - 2)} \lesssim 2^{-j}.
\]

For the last inequality, we used the condition \(\alpha > \frac{5}{2}\) and \(\alpha > \frac{5}{2}\).

We can now estimate (16) with \(r = 1\). For this, we may assume that \(f\) is a compactly supported \(C^2\) function. We first note that \(\psi_{\lambda} = G_s \ast \psi_{j,s,m,p}\) (or \(G_s \ast \varphi_{j,s,m,p}\)) with \(\lambda = (j, s, m, p) \in \Lambda^{1,1}\) from Definition 2.5. All functions \(\psi_{j,s,m,p}\) (or \(\varphi_{j,s,m,p}\)) are generated by \(\psi^p\) (or \(\varphi^p\)) with sufficient vanishing moments and fast decay in frequency such as standard shearlets considered in [16] for each \(p \in \mathbb{N}_0\). Further, we have
\[ \langle f, \psi_{\lambda} \rangle = \langle G_s \ast f, \psi_{j,s,m,p} \rangle \quad (\text{or} \: \langle G_s \ast f, \varphi_{j_0,s,m,p} \rangle), \]

where \( G_s(\cdot) = G_s(\cdot - \cdot) \) and \( G_s \ast f \) belongs to a Sobolev space \( H^2(\mathbb{R}^2) \). Following the proof of Proposition 2.1 in [16], we obtain

\[
\sum_{\lambda \in (\tilde{\Lambda}_N)^c \cap \Lambda^{1,1}} |\langle f, \psi_{\lambda} \rangle|^2 \lesssim 2^{-J + p}, \tag{19}
\]

where \( \tilde{\Lambda}_N \) is the set of indices \( \lambda \in \bigcup_{j \geq 0} \Lambda^{1,1}_{j,p} \) corresponding to the \( 2^{J/2 + p} \) largest shearlet coefficients \( \langle f, \psi_{\lambda} \rangle \) for \( \lambda \in \bigcup_{j \geq 0} \Lambda^{1,1}_{j,p} \) with \( \#(\tilde{\Lambda}_N) \sim 2^{J/2 + p} \) for each \( p \in \mathbb{N}_0 \). Also, one can choose some function \( \eta^p \in L^2(\mathbb{R}^2) \) such that \( \frac{\partial^2}{\partial \lambda_1^2} \eta^p = \psi^p (\text{or} \: \varphi^p) \).

Using integration by parts,

\[
\left\| \frac{\partial^2}{\partial \lambda_1^2} G_s \ast f, \eta_{j,s,m,p} \right\|^2 = 2^{4J} |\langle f, \psi_{\lambda} \rangle|, \]

where \( \eta_{j,s,m,p} \) is defined as \( \psi_{j,s,m,p} \) except for replacing the generator \( \psi^p \) by \( \eta^p \). This implies

\[
|\langle f, \psi_{\lambda} \rangle|^2 \lesssim 2^{-4J}, \tag{20}
\]

and the same argument should be applied for \( \psi_{\lambda} = G_s \ast \varphi_{j_0,s,m,p} \). Also, from Lemma 5.1 (ii), we have

\[
|\langle f, \psi_{\lambda} \rangle|^2 \lesssim 2^{-ap}. \tag{21}
\]

Now let

\[
\tilde{\Lambda}_N^1 = \bigcup_{p=0}^{[Jp]} \tilde{\Lambda}_N
\]

and \( N = \#(\tilde{\Lambda}_N^1) \sim 2^{J(J+1)/2} \). We note that \( \#(\Lambda_{j,p}^{1,1}) \lesssim 2^{2J + p} \) by the same argument as for (17). Using (19)–(21), we obtain

\[
\sum_{\lambda \in (\Lambda_N)^c \cap \Lambda^{1,1}} |\langle f, \psi_{\lambda} \rangle|^2 \lesssim \sum_{\lambda \in (\tilde{\Lambda}_N^1)^c \cap \Lambda^{1,1}} |\langle f, \psi_{\lambda} \rangle|^2 \\
\sum_{\lambda \in (\tilde{\Lambda}_N^1)^c \cap \Lambda^{1,1}} \sum_{p=0}^{[Jp]} |\langle f, \psi_{\lambda} \rangle|^2 \\
\sum_{j \leq J} \sum_{p= [Jp]} |\langle f, \psi_{\lambda} \rangle|^2 \\
\sum_{j \geq J} \sum_{p= [Jp]} |\langle f, \psi_{\lambda} \rangle|^2 + \sum_{j \geq J} \sum_{p \geq j} |\langle f, \psi_{\lambda} \rangle|^2 \\
\sum_{j \geq J} \sum_{p= [Jp]} |\langle f, \psi_{\lambda} \rangle|^2 + \sum_{j \geq J} \sum_{p \geq j} |\langle f, \psi_{\lambda} \rangle|^2
\]
\[ \lesssim 2^{-J(1-\rho)} + \sum_{j \leq J} \sum_{p \geq \lceil J\rho \rceil} 2^{2j+p}2^{-ap} + \sum_{j \geq J} \sum_{p = \lceil J\rho \rceil} 2^{2j+p}2^{-4j} \]
\[ \lesssim \sum_{j \geq J} \sum_{p \geq J} 2^{2j+p}2^{-ap} \lesssim 2^{-J(1-\rho)}. \]

This implies that, for \( N \sim 2^{J(1+2\rho)/2} \),
\[ \sum_{\lambda \in (\Lambda_N)^c \cap \Lambda_{1,1}} |\langle f, \psi_{\lambda} \rangle|^2 \lesssim 2^{-(1-\rho)J} \quad \text{as} \quad J \to \infty. \quad (22) \]

We now turn to analyzing shearlets corresponding to \( \Lambda^0 \), aiming to prove that, for \( N \sim J2^{J(1+4\rho)/2} \),
\[ \sum_{\lambda \in (\Lambda_N)^c \cap \Lambda^0} |\langle f, \psi_{\lambda} \rangle|^2 \lesssim 2^{-(1-8\rho)J} \quad \text{as} \quad J \to \infty. \quad (23) \]

For this, we fix some \( J \geq 0 \). Then we define subsets \( \Lambda^0_j, j \geq 0 \), of \( \Lambda^0 \) by
\[ \Lambda^0_j := \{ \lambda = (j', s, m, p) \in \Lambda^0 : j' = j \}. \quad (24) \]

Notice that \( \Lambda^0 = \bigcup_{j=0}^{\infty} \Lambda^0_j \). Further, for \( r = 0, 1 \), we define again subsets of those sets by
\[ \Lambda^0_{j,r} := \left\{ \lambda = (j, s, m, p) \in \Lambda^0_j : \text{int}(\text{supp}(\psi^r_{\lambda})) \cap \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset, \right. \]
\[ \left. \text{for} \quad p \leq \max \left( \frac{j'p}{2}, \frac{j'\rho}{2} \right), \quad Q_{j,\ell} \in Q^r_j \right\}, \quad (25) \]
corresponding to areas in which the discontinuity curve has a certain slope.

We further aim to collect all indices from the sets \( \Lambda^0_j \) that correspond to significant shearlet coefficients. We might overestimate at this point in the sense of also collecting indices corresponding to small shearlet coefficients, but it will turn out in the end that this more or less crude collection is sufficient for deriving the anticipated sparse approximation behavior. The first set for this purpose extracts such indices, which are related to the set \( Q^0_j \), by choosing
\[ \tilde{\Lambda}_j^0 := \left\{ \lambda = (j, s, m, p) \in \Lambda^0_j : p \leq \frac{j\rho}{2} \right\} \quad \text{for} \quad j = 0, 1, \ldots, \lceil J/4 \rceil - 1 \]
and
\[ \tilde{\Lambda}_j^0 := \left\{ \lambda = (j, s, m, p) \in \Lambda^0_j : \text{int}(\text{supp}(\psi^r_{\lambda})) \cap \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset \right. \]
\[ \left. \text{and} \quad |k_{j,s} + 2^{j/2} \delta_{j,\ell}| \leq 2^{j/4} \quad \text{for} \quad Q_{j,\ell} \in Q^0_j \quad \text{and} \quad p \leq \frac{j\rho}{2} \right\} \quad \text{for} \quad j \geq \lceil J/4 \rceil. \]
Similar considerations lead to the following selection related to the set \( Q^1_j \):

\[
\tilde{\Lambda}^1_j := \left\{ \lambda = (j, s, m, p) \in \Lambda^0_j : \text{int} (\text{supp}(\psi^\lambda_j)) \cap \text{int}(Q_{j, \ell}) \cap \Gamma \\
\neq \emptyset \text{ for } Q_{j, \ell} \in Q^1_j \text{ and } p \leq \frac{J_0}{2} \right\}.
\]

Finally, we define \( \tilde{\Lambda}^0 \) as a set of indices \( \lambda \in \Lambda \) containing all significant shearlet coefficients \( (f, \psi_\lambda) \) (and presumably also others) as follows:

\[
\tilde{\Lambda}^0 := \left( \bigcup_{j=0}^{J} \tilde{\Lambda}^0_j \right) \bigcup \left( \bigcup_{j=\lceil J/4 \rceil}^{\lceil J/3 \rceil} \tilde{\Lambda}^1_j \right).
\]  

(26)

We now turn to estimating \( \#(\tilde{\Lambda}^0) \). Using the same argument as in [16] (page 19), for \( s \in \mathbb{S}, j \geq 0 \) fixed and each \( Q_{j, \ell} \in Q^0_j \), we obtain

\[
\# \left( \left\{ \lambda = (j, s, m, p) \in \Lambda^0_j : \text{int} (\text{supp}(\psi^\lambda_j)) \cap \text{int}(Q_{j, \ell}) \cap \Gamma \\
\neq \emptyset \text{ for } p \leq \frac{J_0}{2} \right\} \right) 
\leq 2^{\left( \frac{j}{2} + \frac{3}{2} \right) \rho} \left( 1 + |\hat{k}_{j, \ell}(s)| \right),
\]

(27)

where \( \hat{k}_{j, \ell}(s) = k_{j, s} + 2^{[j/2]} \xi_{j, \ell} \) and the additional factor \( 2^{\left( \frac{j}{2} + \frac{3}{2} \right) \rho} \) comes from the oversampling parameter \( p \) associated with the sampling matrix \( D_\rho \) and \( A_{j, \rho} \) in (15) for \( \lambda = (j, s, m, p) \). Also, for \( p \in \mathbb{N}_0 \) and \( j \geq 0 \) fixed, it is immediate that

\[
\#(\{\lambda = (j', s, m, p') : j' = j \text{ and } p' = p\}) \leq 2^{2j/2^p} \text{ for } j = 0, \ldots, \lceil J/4 \rceil - 1.
\]

(28)

Finally, we obtain

\[
\#((\lambda = (j', s, m, p') \in \Lambda^0_{j, r} : j' = j \text{ and } p' = p)) \leq 2^{3j/2^p} \text{ for } j \geq 0, r = 0, 1,
\]

(29)

from arguing as follows: There exist at most about \( 2^{j+p} \) shearlets \( \psi_\lambda \) whose approximated part \( \psi^\lambda_j \) intersects \( \Gamma \) for \( \lambda = (j, s, m, p) \) with fixed \( j, s, \) and \( p \). Also, if \( \lambda = (j, s, m, p) \in \Lambda^0_j \), then \( s \in \mathbb{S}_{j/2} \) and \( \#(\mathbb{S}_{j/2}) \leq 2^{j/2} \). Thus, in this case, there are about \( 2^{j+p} \) translates with respect to \( m \) and \( 2^{j/2} \) shearings with respect to \( s \) yielding the estimate in (29).

By (27)–(29), we now derive an estimate for \( \#(\tilde{\Lambda}^0) \) as follows:

\[
\#(\tilde{\Lambda}^0) \leq \sum_{j=0}^{\lfloor J/4 \rfloor - 1} \#(\tilde{\Lambda}^0_j) + \sum_{j=\lceil J/4 \rceil}^{\lceil J/3 \rceil} \#(\tilde{\Lambda}^1_j) + \sum_{j=\lceil J/4 \rceil}^{J} \#(\tilde{\Lambda}^0_j)
\leq \sum_{j=0}^{\lfloor J/4 \rfloor - 1} \left( 2^{2j/2^p} \right) + \sum_{j=\lceil J/4 \rceil}^{\lceil J/3 \rceil} \left( 2^{3j/2^p} \right) 2^{J_0/2}.
\]
Without loss of generality, we may assume that \( N \). For the second inequality above, we used the fact that \( \tilde{\lambda} \). Now let \( N > 0 \) be given. Then we choose \( J > 0 \) such that \( N \sim J^{2(1 + 4\rho)} \). Without loss of generality, we may assume that \( N \geq \#(\tilde{\Lambda}^0) \). Then we have

\[
\sum_{\lambda \in \Lambda^0 \cap (\Lambda_N)^c} |(f, \psi_\lambda)|^2 \leq \sum_{\lambda \in \Lambda^0 \cap (\tilde{\Lambda}^0)^c} |(f, \psi_\lambda)|^2
\]

\[
\lesssim \sum_{j=[J/4]}^{J} \sum_{\lambda \in \Lambda^0_{j,0} \cap (\tilde{\Lambda}^0)^c} |(f, \psi_\lambda)|^2 + \sum_{j=[J/4]}^{[J/3]} \sum_{\lambda \in \Lambda^0_{j,1} \cap (\tilde{\Lambda}^0)^c} |(f, \psi_\lambda)|^2
\]

\[
+ \sum_{j=[J/3]+1}^{J} \sum_{\lambda \in \Lambda^0_{j,0} \cap (\tilde{\Lambda}^0)^c} |(f, \psi_\lambda)|^2 + \sum_{j=0}^{J} \sum_{\lambda \in \Lambda^0_{j} \cap (\tilde{\Lambda}^0)^c} |(f, \psi_\lambda)|^2
\]

\[
= (I) + (II) + (III) + (IV) + (V).
\]  

(31)

For the second inequality above, we used the fact that \( \tilde{\Lambda}^0_j = \Lambda^0_{j,0} \cup \Lambda^0_{j,1} \) for \( j < [J/4] \).

We now estimate (I)-(V). For this, for each \( s \in S_{j/2} \) and \( Q_{j,\ell} \in Q_j^0 \), let

\[ \hat{k}_{j,\ell}(s) = k_{j,s} + 2^{j/2} \hat{s}_{j,\ell}. \]

We start with (I). Using Lemma 5.5(i) and (27), we obtain

\[
(I) \lesssim \sum_{j=[J/4]}^{J} \sum_{\ell: Q_{j,\ell} \in Q_j^0} \sum_{s \in S_{j/2}: |\hat{k}_{j,\ell}(s)| > 2^{-j/4}} 2^{2J\rho} \left( 1 + |\hat{k}_{j,\ell}(s)| \right)^2 \left( \frac{2^{3\rho} 2^{-\frac{3}{2}j}}{|\hat{k}_{j,\ell}(s)|^3} \right)^2
\]

\[
\lesssim \sum_{j=[J/4]}^{J} \sum_{\ell: Q_{j,\ell} \in Q_j^0} 2^{2J\rho} 2^{6\rho} j 2^{-\frac{3}{2}j} \left( 2^{-j/4} \right)^4
\]

\[
\lesssim \sum_{j=[J/4]}^{J} (2^{j/2})2^{2J\rho} 2^{6\rho} j 2^{-\frac{3}{2}j} \left( 2^{-j/4} \right)^4 \lesssim 2^{-J(1-8\rho)}. \]

(32)
Second we turn to (II). For this, we notice that \( \Lambda_{j,1}^0 = \tilde{\Lambda}_j^1 \) for \( j \leq \lceil J/3 \rceil \). But this immediately implies (II) = 0.

To estimate (III), we use Lemma 5.5(ii) and (29) to obtain

\[
(III) \lesssim \sum_{j=\lceil J/3 \rceil}^J 2^{\frac{jp}{2}} \left( 2^{3p} 2^{-9j/4} \right)^2 \lesssim \sum_{j=\lceil J/3 \rceil}^J 2^{\frac{jp}{2}} 2^{-3j(1-2\rho)} \lesssim 2^{\frac{jp}{2}} 2^{-J(1-2\rho)} = 2^{-J(1-5/2\rho)}.
\]

(33)

For the third inequality, we used that \( \rho < \frac{1}{2} \).

Term (IV) is estimated by using Lemma 5.1(ii) and (28). Using also \( \alpha \geq \frac{6}{\rho} + 1 \), we have

\[
(IV) \lesssim \sum_{j=0}^\infty \sum_{p=0}^\infty \left( \frac{2^{3p}}{2} 2^{j \rho} \right) \left( 2^{-\alpha(\frac{J}{2}+\frac{p}{2})} \right)^2 \\
\lesssim \sum_{j=0}^J \left( \sum_{p=0}^\infty 2^{-p(\alpha-1)} \right) 2^{j \rho} 2^{2j} 2^{-a \frac{j}{2}} \\
\lesssim \sum_{j=0}^J 2^{2J-\alpha a \frac{j}{2} + J} \lesssim 2^{-J}.
\]

(34)

Finally, the last term can be estimated by

\[
(V) \leq \sum_{j \geq J} \sum_{\lambda \in \Lambda_{j,0}^0} \sum_{\ell} |\langle f, \psi_{\lambda,\ell} \rangle|^2 + \sum_{j \geq J} \sum_{\lambda \in \Lambda_{j,1}^0} |\langle f, \psi_{\lambda} \rangle|^2 + \sum_{j \geq J} \sum_{\lambda \in \Lambda_{j,0}^0 \cap (\Lambda_{j,0}^0 \cup \Lambda_{j,1}^0)^c} |\langle f, \psi_{\lambda} \rangle|^2 \\
= (A) + (B) + (C).
\]

(35)

It remains to analyze the terms (A)–(C). We start with (A). By Lemmata 5.1(i) and 5.5(i) as well as (27), we obtain

\[
(A) \lesssim \sum_{j \geq J} \sum_{\ell} \sum_{Q_j,\ell \in Q_j^0} 2^{2jp} \sum_{s \in \mathcal{S}_j/2} \left( 1 + |\hat{k}_{j,\ell}(s)| \right) \min \left\{ 2^{-\frac{3}{4}j}, \frac{2^{3p} 2^{-\frac{3}{4}j}}{|\hat{k}_{j,\ell}(s)|^3} \right\}^2 \\
\lesssim \sum_{j \geq J} 2^{2jp} \left( \#(Q_j^0) 2^{-\frac{3}{4}j} 2^{6p} \right) \\
\lesssim \sum_{j \geq J} 2^{-j} 2^{8jp} \lesssim 2^{-J(1-8\rho)}.
\]

The terms (B) and (C) can be estimated by using Lemmata 5.5(ii) and 5.1(ii) as well as equations (28) and (29) to obtain

\[
(B) \lesssim \sum_{j \geq J} \left( 2^{\frac{j}{4}} \left( 2^{\frac{j}{2}} \right) \left( 2^{-\frac{9}{4}j} 2^{3p} \right) \right)^2 = \sum_{j \geq J} 2^{-3j} 2^{6p} 2^{\frac{jp}{2}} \lesssim 2^{-3J + \frac{13}{2} \rho J}
\]

\( \nabla \) Springer
and
\[
(C) \lesssim \sum_{j \geq J} \sum_{p=0}^{\infty} \left( 2^{j} \rho \right) \left( 2^{2j} \rho \right) \left( 2^{\alpha(j+\frac{\rho}{2})} \right)^{2} = \sum_{j \geq J} \left( \sum_{p=0}^{\infty} 2^{-p(\alpha-1)} \right) 2^{2j} \rho + j^{2} \rho^{2} \lesssim 2^{-J}.\]

Thus, continuing (35),
\[
(V) \lesssim 2^{-J(1-8\rho)}. \tag{36}
\]

Summarizing, (32)–(36) imply (23).

Finally, using (18), (22), (23), and the frame property of the shearlet system $\mathcal{S}_{\mathcal{H}}(\varphi_{1}, \psi_{1}; g)$, we can conclude that
\[
\|f - f_{N}\|_{2}^{2} \lesssim \sum_{\lambda \notin \Lambda_{N}} |\langle f, \psi_{\lambda} \rangle|^{2} \lesssim 2^{-J(1-8\rho)}
\]
with $N \sim J^{2 \frac{J+4j}{2}}$, which implies our claim. \qed

6 Proofs of Preliminary Lemmata and Theorem 4.4

6.1 Proof of Lemma 5.2

We first observe that
\[
(Q_{j-j_{0}} \ast \psi p)(\xi) = \sum_{j' = j_{0} - j}^{\infty} \hat{h}_{j', p}(\xi), \quad \text{where } \hat{h}_{j', p}(\xi) := |\hat{g}(A_{j}^{-1}\xi)|^{2} \hat{\psi}(\xi),
\]
with
\[
\text{supp}(h_{j', p}) \subset A_{\text{max}[-j', 0]}[-C, C]^{2}
\]
for some $C > 0$. We can then estimate $\|Q_{j-j_{0}} \ast \psi p\|_{1}$ by
\[
\|Q_{j-j_{0}} \ast \psi p\|_{1} = \int_{\mathbb{R}^{2}} \left| \int_{\mathbb{R}^{2}} (Q_{j-j_{0}} \ast \psi p)(\xi) e^{2\pi i \langle \xi, x \rangle} d\xi \right| dx \lesssim \sum_{j' = 0}^{J-3} \int_{[-C, C]^{2}} \int_{\mathbb{R}^{2}} |\hat{h}_{j', p}(\xi)| d\xi dx + \sum_{j' = 0}^{J-3} \int_{A_{j}[-C, C]^{2}} \int_{\mathbb{R}^{2}} |\hat{h}_{j', p}(\xi)| d\xi dx = (I) + (II).
\]
We will use the following inequality to estimate (I) and (II). Assume that \( j_2 \geq j_1 \) and \( \beta > \alpha + 1 \) with \( \alpha > 0 \). Then
\[
\int_{\mathbb{R}} \frac{2^{-j_1} \min\{1, |2^{-j_2}x|\}^\alpha}{(1 + |2^{-j_1}x|)^\beta} \, dx \lesssim 2^{-\alpha(j_2-j_1)}.
\] (37)

We are now ready to estimate (I) and (II). First, by (5)–(7), we have

\[
\begin{align*}
(I) & \lesssim \sum_{j=0}^{\infty} \int_{[-C, C]^2} \int_{\mathbb{R}^2} \frac{\min\{1, |2^{-j} \xi_1|\}^\alpha \min\{1, |\xi_1|\}^\alpha \min\{1, |2^{-p} \xi_2|\}^\alpha}{(1+|2^{-j} \xi_1|)^\beta (1+|2^{-j/2} \xi_2|)^\beta (1+|\xi_1|)^\beta (1+|2^{-p} \xi_2|)^\beta} \, d\xi \, dx \\
& \lesssim \sum_{j=0}^{2^{p-1}} \int_{[-C, C]^2} \int_{\mathbb{R}^2} \frac{\min\{1, |2^{-j} \xi_1|\}^\alpha}{(1+|\xi_1|)^\beta} \, d\xi_1 \int_{\mathbb{R}^2} \frac{\min\{1, |2^{-p} \xi_2|\}^\alpha}{(1+|2^{-j/2} \xi_2|)^\beta} \, d\xi_2 \, dx \\
& \quad + \sum_{j=2^p}^{\infty} \int_{[-C, C]^2} \int_{\mathbb{R}^2} \frac{\min\{1, |2^{-j} \xi_1|\}^\alpha}{(1+|\xi_1|)^\beta} \, d\xi_1 \int_{\mathbb{R}^2} \frac{\min\{1, |2^{-p} \xi_2|\}^\alpha}{(1+|2^{-j/2} \xi_2|)^\beta} \, d\xi_2 \, dx \\
& \lesssim \sum_{j=0}^{2^{p-1}} 2^{-j(\alpha-1)} 2^{-\alpha(p-j/2) 2^p} + \sum_{j=2^p}^{\infty} 2^{-j(\alpha-1) 2^p} \lesssim 2^{-p(\alpha-1)} + 2^{-p(2\alpha-3)} \leq 2.
\end{align*}
\]

Second,

\[
\begin{align*}
(II) & \lesssim \sum_{j=0}^{\infty} \int_{A_j([-C, C]^2)} \int_{\mathbb{R}^2} \frac{\min\{1, |2^{j} \xi_1|\}^\alpha \min\{1, |\xi_1|\}^\alpha \min\{1, |2^{-p} \xi_2|\}^\alpha}{(1+|2^{j/2} \xi_2|)^\beta (1+|2^j \xi_1|)^\beta (1+|2^{-p} \xi_2|)^\beta} \, d\xi \, dx \\
& \lesssim \sum_{j=0}^{\infty} \int_{A_j([-C, C]^2)} \int_{\mathbb{R}^2} \frac{\min\{1, |\xi_1|\}^\alpha}{(1+|2^{j/2} \xi_2|)^\beta} \, d\xi_1 \int_{\mathbb{R}^2} \frac{\min\{1, |2^{-p} \xi_2|\}^\alpha}{(1+|2^j \xi_1|)^\beta (1+|2^{-p} \xi_2|)^\beta} \, d\xi_2 \, dx \\
& \lesssim \sum_{j=0}^{\infty} 2^{-j} 2^{-\alpha j} \int_{\mathbb{R}^2} \frac{\min\{1, |2^{-p} \xi_2|\}^2}{(1+|2^{j/2} \xi_2|)^2 (1+|2^{-p} \xi_2|)^2} \, d\xi_2 \, dx \\
& \lesssim \sum_{j=0}^{\infty} 2^{-j} 2^{-\alpha j} \int_{\mathbb{R}^2} \frac{2^{-j/2}}{(1+|\xi_2|)^2} \, d\xi_2 \, dx \\
& \lesssim \sum_{j=0}^{\infty} 2^{-j} 2^{-\alpha j} 2^{-j/2} \leq 2.
\end{align*}
\]

Therefore, (I) and (II) are uniformly bounded, which implies the uniform boundedness of \( \| Q_{j-j_0} * \psi^p \|_1 \).
6.2 Proof of Lemma 5.1

We start by proving (i). First, note that Proposition 2.7(ii) implies the form

\[ \psi_\lambda = |\det(A_j)|^{1/2}(\Theta_{j_0} \ast \psi') (S_k A_j \cdot -D_p m), \]

where \( \lambda = (j, s, m, p) \in \Lambda \) with \( s = s([j_0/2], q_0) \in S \), \( j \geq j_0 \) and \( k \in \mathbb{Z} \) with \( s = \frac{k}{2^{j_0/2}} \). This allows us to estimate \(|\langle f, \psi_\lambda \rangle|\) as follows:

\[
|\langle f, \psi_\lambda \rangle| \leq |\det(A_j)|^{1/2} \int_{\mathbb{R}^2} |(\Theta_{j_0} \ast \psi') (S_k A_j x - D_p m) f(x)| \, dx \\
\leq 2^{-3j} \int_{\mathbb{R}^2} |f(A_j^{-1} S_k^{-1} (y + D_p m))||(\Theta_{j_0} \ast \psi')'(y)| \, dy \\
\leq 2^{-3j} \|f\|_\infty \|\Theta_{j_0} \ast \psi'\|_1. 
\]

The claim in (i) now follows from Lemma 5.2.

We next turn to proving (ii). Since by definition, \( \psi_\lambda = G_s \ast \psi_{j,s,m,p} \), we have

\[
|\langle f, \psi_\lambda \rangle|^2 \leq \|\hat{f}\|_2^2 \|\hat{G}_s \cdot \hat{\psi}_{j,s,m,p}\|_2^2. 
\]

Now we estimate \( \|\hat{G}_s \cdot \hat{\psi}_{j,s,m,p}\|_2^2 \) as follows. By Lemma 2.3,

\[
\|\hat{G}_s \cdot \hat{\psi}_{j,s,m,p}\|_2^2 \lesssim \int_{\mathbb{R}^2} \left| \sum_{j'=0}^{\infty} \hat{\psi}(A_{-j'} \xi) \right|^2 \|\hat{\psi}_{j,0,m,p}(\xi)\|^2 d\xi \\
\lesssim \int_{\mathbb{R}^2} \sum_{j'=0}^{\infty} |\hat{\psi}(A_{-j'} \xi)|^2 |\hat{\psi}_{j,0,m,p}(\xi)|^2 d\xi \\
\lesssim \sum_{j'=0}^{j+p} \int_{\mathbb{R}^2} \cdots + \sum_{j'=j+p}^{\infty} \int_{\mathbb{R}^2} \cdots = (I) + (II). 
\]

Note that (5) and (7) imply that

\[
|\hat{\psi}_{j,0,m,p}(\xi)|^2 \lesssim 2^{-j/p} \cdot \frac{\min\{1, |2^{-j} \xi|^{2\alpha}\}}{(1 + |2^{-j} \xi_1|)^{2\beta}} \cdot \frac{\min\{1, |2^{-j/2-p} \xi_2|^{2\alpha}\}}{(1 + |2^{-j/2-p} \xi_2|)^{2\beta}}. 
\]

Hence, using (7), (37), and (38), we can estimate (I) by

\[
(I) \lesssim \sum_{j'=0}^{j+p} 2^{-j/2-p} \int_{\mathbb{R}} \frac{\min\{1, |2^{-j/2-p} \xi_2|^{2\alpha}\}}{(1 + |2^{-j/2-p} \xi_2|)^{2\beta}} \, d\xi_2 \int_{\mathbb{R}} 2^{-j} \frac{1}{(1 + |2^{-j} \xi_1|)^{2\beta}} \, d\xi_1 \\
\lesssim \sum_{j'=0}^{j+p} 2^{-j/2-p} \int_{\mathbb{R}} \frac{\min\{1, |2^{-j/2-p} \xi_2|^{2\alpha}\}}{(1 + |2^{-j/2-p} \xi_2|)^{2\beta}} \, d\xi_2 \lesssim 2^{-\alpha p}. 
\]
Similarly, we can estimate (II) as follows:

\[
(II) \lesssim \sum_{j' = j + p}^{\infty} 2^{-j} \int_{\mathbb{R}} \min\{1, |2^{-j} \xi_1|^{2\alpha}\} \frac{d\xi_1}{(1 + |2^{-j} \xi_1|^{2\beta})} \int_{\mathbb{R}} \frac{2^{-j/2 - p}}{(1 + |2^{-j/2 - p} \xi_2|^{2\beta})} d\xi_2
\]

\[
\lesssim \sum_{j' = j + p}^{\infty} 2^{-j} \int_{\mathbb{R}} \min\{1, |2^{-j} \xi_1|^{2\alpha}\} \frac{d\xi_1}{(1 + |2^{-j} \xi_1|^{2\beta})}
\]

\[
\lesssim 2^{-2\alpha p}.
\]

This proves (ii).

### 6.3 Proof of Lemma 5.3

First, note that (i) is obvious from the definition of \(\psi^\#_\lambda\) and

\[
|\langle f, (\psi_\lambda - \psi^\#_\lambda)\rangle|^2 \leq \|f\|_2^2 \|\psi_\lambda - \psi^\#_\lambda\|_2^2.
\]

Next, we estimate \(\|\psi_\lambda - \psi^\#_\lambda\|_2^2\) to show (ii). By using (5), (7), and (37), we obtain

\[
\|\psi_\lambda - \psi^\#_\lambda\|_2^2 \lesssim \int_{\mathbb{R}^2} \left| \sum_{j' = 0}^{j(1 - \rho)} \hat{g}(A_{j'}^{1-\xi_1})^2 \hat{\psi}_{j,0,0,p}(\xi) \right|^2 d\xi
\]

\[
\lesssim \int_{\mathbb{R}^2} \sum_{j' = 0}^{j(1 - \rho)} \left| \hat{g}(A_{j'}^{1-\xi_1}) \right|^2 \left| \hat{\psi}_{j,0,0,p}(\xi) \right|^2 d\xi
\]

\[
\lesssim \sum_{j' = 0}^{j(1 - \rho)} \int_{\mathbb{R}^2} \frac{2^{-j'} \min\{1, |2^{-j'} \xi_1|^{2\alpha}\}}{(1 + |2^{-j'} \xi_1|^{\beta})} \cdot \frac{2^{-p - j/2} \min\{1, |2^{-j'/2 - p} \xi_2|^{2\alpha}\}}{(1 + |2^{-j'/2 - p} \xi_2|^{\beta})} d\xi
\]

\[
\lesssim \sum_{j' = 0}^{j(1 - \rho)} 2^{-2\alpha(j - j') - (j - j' + 2p)\alpha} \lesssim 2^{-2\alpha j\rho - 2^{-2\alpha}}.
\]

This proves our claim.

### 6.4 Proof of Lemma 5.4

We only consider case \(\ell = 1\), since the other case can be shown similarly. By (5) and (7), we have

\[
\left| \left( \frac{\partial}{\partial \xi_2} \right)^{j - \max\{\lfloor j(1 - \rho)\rfloor, j_0\}} \Theta^\rho P(\xi) \right|
\]

\[
\lesssim \sum_{j = 0}^{\infty} \left| \left( \frac{\partial}{\partial \xi_2} \right)^{j} \hat{g}(A_j \xi) \right|^2 \hat{\psi} P(\xi) + \sum_{j = 0}^{\infty} \left| \hat{g}(A_j \xi) \right|^2 \left( \frac{\partial}{\partial \xi_2} \right)^{j} \hat{\psi} P(\xi)
\]

\(\copyright\) Springer
These estimates for (I) and (II) then prove the lemma.

Then we have

\[
(I) \lesssim \sum_{j=0}^{\infty} 2^{j/2} |\xi| \min\{1, |2^j \xi_1|^{\alpha} \} \min\{1, |\xi_1|^{\alpha-1} \} \min\{1, |2^{-p} \xi_2|^{\alpha} \} \\
\frac{|\xi_1|(1 + |2^j \xi_2|)^{\beta-1}(1 + |2^{j/2} \xi_2|)^\beta(1 + |\xi_1|)^\beta(1 + |2^{-p} \xi_2|)^\beta}{(1 + |\xi_2|)^{\beta-1}(1 + |\xi_1|)^\beta}
\]

and

\[
(II) \lesssim \sum_{j=1}^{\infty} \frac{\min\{1, |\xi_1|^{\alpha} \} 2^{-p^j} |\xi_2|^2 |2^{j/2} \xi_1|^\gamma}{(1 + |\xi_1|)^\beta(1 + |2^{-p} \xi_2|)^\gamma |\xi_2|^\gamma} \\
\lesssim \frac{\min\{1, |\xi_1|^{\alpha} \}}{(1 + |\xi_1|)^{\beta-\gamma}(1 + |\xi_2|)^\gamma}.
\]

These estimates for (I) and (II) then prove the lemma.

### 6.5 Proof of Lemma 5.5

For some \( q \in L^2(\mathbb{R}^2) \), set

\[
q_{j,k,s,m} := |\det(A_j)|^{1/2} q(S_k, A_j \cdot m)
\]

for \( j \geq 0, k, s \in \mathbb{Z} \) and \( m \in \mathbb{Z}^2 \). Further, assume that

\[
\text{supp}(q_{j,k,s,m}) \subset A_j^{-1} S_{k,s}^{-1} \left([-2^{j/2} L, 2^{j/2} L] \times [-2^{j/2} L, 2^{j/2} L] + m\right)
\]

for some \( L > 0 \). Provided that in addition, for \( \alpha_1 \geq 5, \alpha_2 \geq 4 \), and \( h \in L^1(\mathbb{R}) \), we have

\[
|\hat{q}(\xi)| \lesssim \frac{\min\{1, |\xi_1|^{\alpha_1} \}}{(1 + |\xi_1|)^{\alpha_2}(1 + |\xi_2|)^{\alpha_2}} \quad \text{and} \quad \left| \frac{\partial}{\partial \xi_2} \hat{q}(\xi) \right| \leq |h(\xi_1)| \cdot \left(1 + \frac{|\xi_2|}{|\xi_1|}\right)^{\alpha_2},
\]
by following the proof of Proposition 2.2 in [16], we can show that
\[ |\langle f, q_{j,k,m} \rangle| \lesssim \frac{2^{-\frac{3}{4}j}2^{3\rho j}}{|k_s + 2^{j/2}|\hat{s}|^3} \text{ if } |\hat{s}| \leq 3 \] (39)
and
\[ |\langle f, q_{j,k,m} \rangle| \lesssim 2^{3\rho j/2 - \frac{9}{4}} \text{ if } |\hat{s}| > 3/2. \] (40)

We next choose \( q_{j,k,m} := \psi^\#_\lambda \) for \( \lambda = (j, s, m, p) \in \Lambda_s \) with \( s = s(j_0/2), q_0 \in \mathbb{S}, s = \frac{k_s}{2^{j/2}}, \) and \( j \geq j_0. \) Hence, in particular, \( q = \Theta_{j - \max\{|j_0-1|, j_0\}}^\# \psi^p. \) By Lemmas 5.3(i) and 5.4, we derive (39) and (40) for this choice. Thus, by Lemma 5.3(ii),
\[ |\langle f, \psi_\lambda \rangle| \lesssim \frac{2^{-\frac{3}{4}j}2^{3\rho j}}{|k_s + 2^{j/2}|\hat{s}|^3} + 2^{-\frac{1}{4}j}2^\rho 2^j \text{ if } |\hat{s}| \leq 3 \] (41)
and
\[ |\langle f, \psi_\lambda \rangle| \lesssim 2^{3\rho j/2 - \frac{9}{4}j} + 2^{-\frac{1}{4}j}2^\rho 2^j \text{ if } |\hat{s}| > 3/2. \] (42)
Finally, Lemma 5.1(i), (41), and (42) imply (i) and (ii) for \( \alpha \geq \frac{6}{\rho}. \)

### 6.6 Proof of Theorem 4.4

We will retain all notation used in the proof of Theorem 4.3; otherwise we will specify them. First, note that the case when the supports of standard shearlets \( \psi_\lambda \) considered in [16] do not intersect the discontinuity of \( f \) can be handled by Proposition 2.1 in [16].

Thus, it suffices to only consider the case when the supports of standard shearlets intersect the discontinuity curve of \( f. \) From now on, we denote by \( \psi_\lambda \) standard shearlets considered in [16]. We first observe that shearlets \( \psi_\lambda \) are a special case of Definition 2.5, since each \( \psi_\lambda \) can be given as
\[ \psi_\lambda = G_s * \psi_{j,s,m,p} \text{ with } p = 0 \text{ and } \hat{G}_s = 1, \] (43)
see also Proposition 2.7. From (43), we obtain standard shearlets \( \psi_\lambda \) generated by the shearlet generator \( \psi, \) namely \( \psi_1(x_1)\psi_1(x_2). \) Here, we do not need to require \( \psi \) to be a separable function \( \psi_1(x_1)\psi_1(x_2); \) it could also be a nonseparable function. In fact, separability for the dualizable shearlet generator in Definition 2.5 is never used in the proof of Theorem 4.3. It is only used for constructing orthonormal systems \( \Psi_s(\varphi_1, \psi_1) \) to obtain frame properties.

Having shearlets \( \psi_\lambda \) of the form (43), we can follow the proof of Theorem 4.3 to show (23) with \( \rho = 0 \) and \( N \sim J^{2^j/2}. \) For this, we first consider index sets \( \Lambda^0_j, \Lambda^0_{j,r}, \Lambda^0_{j,r}, \Lambda^0_j \) as defined in (24)–(26) for \( r = 0, 1. \) These are index sets used to estimate (23) in the proof of Theorem 4.3. Note that the additional parameter \( p \in \mathbb{N}_0 \) associated with \( D_p \) is not needed for standard shearlets \( \psi_\lambda. \) Hence one can remove this index \( p \in \mathbb{N}_0 \) in those index sets. Thus,
\[ \Lambda_j^0 = \{ \lambda = (j', s, m) \in \Lambda^0 : j' = j \} \]

as well as

\[ \Lambda_{j,r}^0 = \{ \lambda = (j, s, m) \in \Lambda_j^0 : \text{int}(\text{supp}(\psi_\lambda)) \cap \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset, \text{ for } Q_{j,\ell} \in Q_j^r \} \]

for \( r = 0, 1 \). Moreover, for \( j < \lceil J/4 \rceil \), we have \( \tilde{\Lambda}_j^0 = \Lambda_j^0 \) and, for \( j \geq \lceil J/4 \rceil \),

\[ \tilde{\Lambda}_j^0 = \{ \lambda = (j, s, m) \in \Lambda_j^0 : \text{int}(\text{supp}(\psi_\lambda)) \cap \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset, |k_j, s + 2^{\lceil j/2 \rceil} \delta_{j,\ell}| \leq 2^{J-j}, Q_{j,\ell} \in Q_j^0 \} \].

Also,

\[ \tilde{\Lambda}_j^1 = \{ \lambda = (j, s, m) \in \Lambda_j^0 : \text{int}(\text{supp}(\psi_\lambda)) \cap \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset \text{ for } Q_{j,\ell} \in Q_j^1 \} \].

Further, we note that the size of the support of \( \psi_\lambda \) is given as in Lemma 5.3 (i) with \( \rho = 0 \) and \( p = 0 \). In other words, we obtain

\[ \text{supp}(\psi_\lambda) \subset S_j^{-1} A_j^{-1}([-C, C]^2 + m). \] \hspace{0.5cm} (44)

Now, (44) and the fact that the additional index \( p \in \mathbb{N}_0 \) is not needed imply (27)–(29) with \( \rho = 0 \) and \( p = 0 \). Therefore, we can apply the same estimates as the ones in (30) now with \( \rho = 0 \) to show

\[ \sharp(\tilde{\Lambda}^0) \lesssim J^{2^{J/2}} \]. \hspace{0.5cm} (45)

Next, we estimate (31) as in the proof of Theorem 4.3 for \( N \sim J2^{J/2} \). Recall that for estimating (31), the estimates in (27)–(29) and the Lemmata 5.1 and 5.5 are used in the proof of Theorem 4.3. For standard shearlets \( \psi_\lambda \), we can use Lemma 5.1 (i) (see [16]). Lemma 5.1 (ii) is not needed, since there is no additional index \( p \in \mathbb{N}_0 \) for standard shearlets. Further, Lemma 5.5 with \( \rho = 0 \) holds for standard shearlets—see [16].

Therefore, we can apply the same estimates for (I)–(III) in (31) and (A)–(B) in (35) now with \( \rho = 0 \). For the terms (IV) in (31) and (C) in (35), we note that they are all zero, since these terms control \( p \in \mathbb{N}_0 \) and

\[ \Lambda_j^0 = \Lambda_{j,0}^0 \cup \Lambda_{j,1}^0. \]

This yields (23) with \( \rho = 0 \). Finally, (23) with \( \rho = 0 \) and (45) imply our claim.

References

1. Candès, E.J., Donoho, D.L.: New tight frames of curvelets and optimal representations of objects with piecewise \( C^2 \) singularities. Comm. Pure Appl. Math. 56, 216–266 (2004)
2. Candès, E.J., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. 59, 1207–1223 (2006)
3. Casazza, P., Kutyniok, G., (Eds.) Finite Frames: Theory and Applications. Birkhäuser, Boston (2012)
4. Christensen, O.: An Introduction to Frames and Riesz Bases. Birkhäuser, Boston (2003)
5. Dahmen, W., Huang, C., Schwab, C., Welper, G.: Adaptive Petrov–Galerkin methods for first order transport equations. SIAM J. Numer. Anal. 50, 2420–2445 (2012)
6. Dahmen, W., Kutyniok, G., Lim, W.-Q., Schwab, C., Welper, G.: Adaptive anisotropic discretizations for transport equations (under review)
7. Daubechies, I.: Ten Lectures on Wavelets. SIAM, Philadelphia (1992)
8. Davenport, M., Duarte, M., Eldar, Y., Kutyniok, G.: Introduction to compressed sensing. In: Compressed Sensing: Theory and Applications. Cambridge University Press, Cambridge, pp. 1–68 (2012)
9. Donoho, D.L.: Sparse components of images and optimal atomic decomposition. Constr. Approx. 17, 353–382 (2001)
10. Donoho, D.L.: Compressed sensing. IEEE Trans. Inform. Theory 52, 1289–1306 (2006)
11. Grohs, P.: Bandlimited Shearlet Frames with nice Duals. J. Comput. Appl. Math 243, 139–151 (2013)
12. Guo, K., Kutyniok, G., Labate, D.: Sparse multidimensional representations using anisotropic dilation and shear operators. In: Wavelets and Splines (Athens, GA, 2005). Nashboro Press, Nashville, TN, pp. 189–201 (2006)
13. Kutyniok, G., Labate, D.: Introduction to Shearlets. In: Shearlets: Multiscale Analysis for Multivariate Data. Birkhäuser, Boston, pp. 1–38 (2012)
14. Kutyniok, G., Lemvig, J., Lim, W.-Q.: Optimally sparse approximations of 3D functions by compactly supported shearlet frames. SIAM J. Math. Anal. 44, 2962–3017 (2012)
15. Kittipoom, P., Kutyniok, G., Lim, W.-Q.: Construction of compactly supported shearlet frames. Constr. Approx. 35, 21–72 (2012)
16. Kutyniok, G., Lim, W.-Q.: Compactly supported shearlets are optimally sparse. J. Approx. Theory 163, 1564–1589 (2011)
17. Kutyniok, G., Lim, W.-Q., Reisenhofer, R.: ShearLab 3D: faithful digital shearlet transforms based on compactly supported shearlets. ACM Trans. Math. Softw. 42(1), 5:1–5:42 (2016)
18. Nam, S., Davies, M.E., Elad, M., Gribonval, R.: The cosparse analysis model and algorithms. Appl. Comput. Harmon. Anal. 34, 30–56 (2013)