DUALIZABLE SHEARLET FRAMES AND SPARSE APPROXIMATION

GITTA KUTYNIOK AND WANG-Q LIM

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 proven 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 efficiently computed. We also show that dualizable shearlet frames still provide optimally sparse approximations of anisotropic features.

1. INTRODUCTION

During the last years, methodologies utilizing sparse approximations have had a tremendous impact on data science. This is foremost due to the method of compressed sensing (see [3, 11] or [9]), which played a major role in the initiation of today's paradigm that any type of data admits a sparse representation within a suitably chosen orthonormal basis, or, more generally, a frame [5]. 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 are cartoon-like functions, which are – coarsely speaking – compactly supported

piecewise smooth functions. Shearlet systems [14], 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 (non-tight) 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 in providing decompositions and expansions for data processing, and the reader might consult [4] 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 *tight frame*; in 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

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

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 to have 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:

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

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

$$f = \sum_{i \in I} c_i \varphi_i, \tag{3}$$

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, related to compressed sensing. Instead we expect that the sequence of frame coefficients $(\langle f, \varphi_i \rangle)_{i \in I}$ is sparse. In [19], 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 ℓ_1-norm on such coefficient sequences and mimimizing 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, which models features images are assumed to be governed by.

As such a class typically so-called *cartoon-like functions* introduced in [10] are taken, which 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 [6, 7].

Donoho then proved in [10] 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

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

This results 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 [13] as a directional representation system which meets this benchmark result, but which – in contrast to the previously advocated system of curvelets [2] – 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 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* $\mathcal{SH}(\phi, \psi, \tilde{\psi}; c)$ *is defined by*

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

where

$$\Phi(\phi; c_1) = \{\phi_m := \phi(\cdot - c_1 m) : m \in \mathbb{Z}^2\},$$
$$\Psi(\psi; c) = \{\psi_{j,k,m} := |\det(A_j)|^{1/2} \psi(S_k A_j \cdot - \mathrm{diag}(c_1, c_2)m) : j \geq 0, |k| \leq 2^{\lceil j/2 \rceil}, m \in \mathbb{Z}^2\},$$
$$\tilde{\Psi}(\tilde{\psi}; c) = \{\tilde{\psi}_{j,k,m} := |\det(\tilde{A}_j)|^{1/2} \tilde{\psi}(S_k^T \tilde{A}_j \cdot - \mathrm{diag}(c_2, c_1)m) : j \geq 0, |k| \leq 2^{\lceil j/2 \rceil}, m \in \mathbb{Z}^2\},$$

with $\tilde{A}_j = \mathrm{diag}(2^{\lfloor j/2 \rfloor}, 2^j)$.

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`) [18]. As shown in [16], 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 proven in [17] 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 [12]. 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 which can be regarded as being of the form $\mathcal{SH}(\phi, \psi, \tilde{\psi}; c)$ and satisfies the following properties:

- The system is compactly supported and forms a frame for $L^2(\mathbb{R}^2)$.
- The system delivers optimal sparse approximations of cartoon-like functions.
- An associated dual frame can be stated in closed form and 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 [17] for the class of compactly supported shearlet frames introduced in [16] with respect to the exponent of the additional log-term (see Corollary 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 [2].

1.7. **Outline.** The paper is organized as follows. The construction of dualizable shearlet systems is presented and discussed in Section 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 optimally sparse approximations of cartoon-like functions is presented in Section 4 as Theorem 4.3. Since the proof is very involved, the key steps and the core part are presented in Section 5 with the proofs of several preliminary lemmata being outsourced to Section 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 Subsection 2.1. Since those elements lack directionality in the sense of wedgelike shape elements, they are subsequently filtered, yielding the desired dualizable shearlet systems (see Subsection 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 Subsection 1.4 –; the vertically aligned system will be derived by switching the variables, or rotation by 90 degrees denoted by $R_{\frac{\pi}{2}}$.

Since the construction is rather technical in nature, it is not initially clear that the term "shearlets" is justified; and we argue in Subsection 2.3 in favor of this expression by comparing the novel systems to customarily defined shearlets (cf. Subsection 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, which satisfy the support condition

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

as well as, for some $\rho \in (0, \frac{2}{13})$, $\alpha \geq \frac{6}{\rho} + 1$, and $\beta > \alpha + 1$, the decay conditions

$$\left| \left(\frac{d}{d\xi}\right)^\ell \hat{\psi}_1(\xi) \right| \lesssim \frac{\min\{1, |\xi|^\alpha\}}{(1 + |\xi|)^\beta} \quad \text{and} \quad \left| \left(\frac{d}{d\xi}\right)^\ell \hat{\varphi}_1(\xi) \right| \lesssim \frac{1}{(1 + |\xi|)^\beta} \quad \text{for } \ell = 0, 1. \tag{5}$$

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 [8].

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 Figure 1(a), and then apply shearing operators.

To generate the anticipated tiling, for $x = (x_1, x_2) \in \mathbb{R}^2$, we set

$$\psi^0(x) := \psi_1(x_1)\varphi_1(x_2) \quad \text{and} \quad \psi^p(x) := 2^{(p-1)/2}\psi_1(x_1)\psi_1(2^{p-1}x_2) \text{ for } p > 0$$

as well as

$$\varphi^0(x) := \varphi_1(x_1)\varphi_1(x_2) \quad \text{and} \quad \varphi^p(x) := 2^{(p-1)/2}\varphi_1(x_1)\psi_1(2^{p-1}x_2) \text{ for } p > 0.$$

Notice that the parameter p will be utilized to derive the dyadic substructure in vertical direction.

For a fixed integer $j_0 \geq 0$, we next consider the system given by

$$\{|\det(A_{j_0})|^{1/2}\varphi^p(A_{j_0} \cdot - D_p m), \ |\det(A_j)|^{1/2}\psi^p(A_j \cdot - D_p m) : j \geq j_0, p \geq 0\},$$

where $D_p = \text{diag}(1, 2^{-\max\{p-1,0\}})$, which achieves the tiling of Fourier domain as depicted in Figure 1(a). 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 "normal" shearlets, the shear parameter should equal $\frac{k}{2^{\lceil j/2 \rceil}}$ with $|k| \leq 2^{\lceil j/2 \rceil}$. Since we later need to parameterize by those quotients, we require a unique representation without ambiguity. For this, we define the map

$$s : \{(j,q) : j = 0, q = 0\} \cup \{j : j \geq 0\} \times \{q : |q| \leq 2^j, q \in 2\mathbb{Z} + 1\} \to [-1,1], \ \ s(j,q) := \frac{q}{2^j},$$

which is obviously injective. Thus from now on, we consider the set of shear parameters given by

$$\mathbb{S} = \{s(j,q) = 0 : j = 0, q = 0\} \cup \{s(j,q) : j \geq 0, |q| \leq 2^j, q \in 2\mathbb{Z} + 1\}.$$

Armed with 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 $D_p := \mathrm{diag}(1, 2^{-d_p})$ 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 $s = \frac{q_0}{2^{\lceil j_0/2 \rceil}}$, we define the* shearlet-type wavelet system $\Psi_s(\varphi_1, \psi_1)$ *by*

$$\Psi_s(\varphi_1, \psi_1) := \{\varphi_{j_0,s,m,p}, \psi_{j,s,m,p} : j \geq j_0, m \in \mathbb{Z}^2, p \geq 0\},$$

where

$$\varphi_{j,s,m,p} := |\det(A_j)|^{1/2} \varphi^p(A_j S_s \cdot - D_p m) \quad and \quad \psi_{j,s,m,p} := |\det(A_j)|^{1/2} \psi^p(A_j S_s \cdot - D_p m).$$

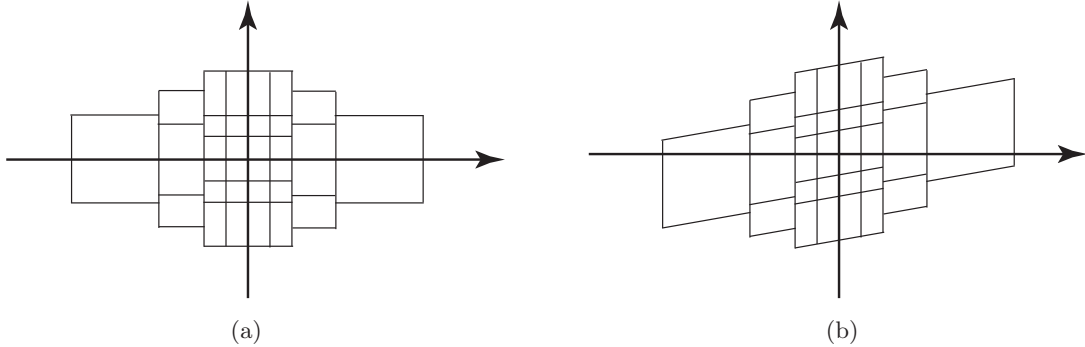The tiling of Fourier domain by shearlet-type wavelet systems is depicted in Figure 1.

(a)   (b)

FIGURE 1. Tiling of Fourier domain by Ψ_s with (a) $s = 0$ and (b) $s = \frac{1}{4}$.

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,

$$\varphi_{0,0,m,0}(x) = \varphi_1(x_1 - m_1)\varphi_1(x_2 - m_2) \ \text{ and } \ \varphi_{0,0,m,p}(x) = \varphi_1(x_1 - m_1)2^{\frac{d_p}{2}}\psi_1(2^{d_p}x_2 - m_2)$$

as well as, in addition for $j \geq 0$ and $p > 0$,

$$\psi_{j,0,m,0}(x) = 2^{\frac{j}{2}}\psi_1(2^j x_1 - m_1)2^{\frac{1}{2}\lfloor\frac{j}{2}\rfloor}\varphi_1(2^{\lfloor\frac{j}{2}\rfloor}x_2 - m_2)$$

and

$$\psi_{j,0,m,p}(x) = 2^{\frac{j}{2}}\psi_1(2^j x_1 - m_1)2^{\frac{d_p}{2}+\frac{1}{2}\lfloor\frac{j}{2}\rfloor}\psi_1(2^{d_p}(2^{\lfloor\frac{j}{2}\rfloor}x_2) - m_2).$$

Next, for each $j \geq 0$, let V_j and W_j be the subspaces of $L^2(\mathbb{R}^2)$ defined by

$$V_j = \overline{\text{span}\{2^{\frac{j}{2}}\varphi_1(2^j \cdot -m) : m \in \mathbb{Z}\}} \quad \text{and} \quad W_j = \overline{\text{span}\{2^{\frac{j}{2}}\psi_1(2^j \cdot -m) : m \in \mathbb{Z}\}}.$$

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 $V_0 \otimes W_{d_p}$, 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 $W_j \otimes V_0$ and $W_j \otimes W_{d_p}$, respectively.

Since $V_0 \perp W_j$ and $W_j \perp W_{j'}$ for $j, j' \geq 0$, $j \neq j'$, those subspaces are mutually orthogonal and, for each $j \geq 0$,

$$\left(V_0 \otimes V_0\right) \oplus \left(\bigoplus_{p=1}^{\infty} V_0 \otimes W_{d_p}\right) \oplus \left(W_j \otimes V_0\right) \oplus \left(\bigoplus_{p=1}^{\infty} W_j \otimes W_{d_p}\right) = (V_0 \otimes L^2(\mathbb{R})) \oplus (W_j \otimes L^2(\mathbb{R})).$$

From this, we finally obtain

$$\left(V_0 \otimes L^2(\mathbb{R})\right) \oplus \left(\bigoplus_{j=0}^{\infty} W_j \otimes L^2(\mathbb{R})\right) = L^2(\mathbb{R}^2).$$

This proves our claim. $\qquad\square$

We wish to mention that the definition of a dualizable shearlet system in Definition 2.5 will also require the systems derived by switching the variable in $\Psi_s(\varphi_1, \psi_1)$, i.e., by rotation by $R_{\frac{\pi}{2}}$.

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

$$\delta_g = \inf_{\xi \in \Omega_g} |\hat{g}(\xi)| > 0, \quad \text{where } \Omega_g = \left\{\xi \in \mathbb{R}^2 : |\tfrac{\xi_2}{\xi_1}| < 1, \tfrac{1}{2} < |\xi_1| < 1\right\}, \tag{6}$$

as well as the decay condition

$$\left|\left(\frac{\partial}{\partial \xi_2}\right)^{\ell}\hat{g}(\xi)\right| \lesssim \frac{\min\{1, |\xi_1|^{\alpha}\}}{(1 + |\xi_1|)^{\beta}(1 + |\xi_2|)^{\beta}} \quad \text{for } \ell = 0, 1, \tag{7}$$

with α and β chosen as before (i.e., $\rho \in (0, \frac{2}{13})$, $\alpha \geq \frac{6}{\rho} + 1$, 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 [16].

Lemma 2.3 ([16]). *Letting $\varphi_1, \psi_1 \in L^2(\mathbb{R})$ and $g \in L^2(\mathbb{R}^2)$ be defined as before, we have*

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

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

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

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

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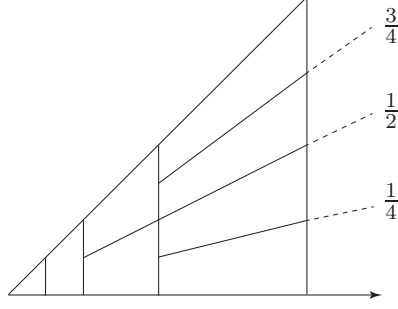

FIGURE 2. Frequency tiling with $\hat{G}_s$ for $s = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ (for $\xi_1 > 0$).

The following result provides an identity, which 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*

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

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

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*

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

with index set

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

where

$$\psi_\lambda^0 = \begin{cases} G_s * \varphi_{j_0, s, m, p} & : \quad \lambda = (-1, s, m, p) \in \Lambda_s, \\ G_s * \psi_{j, s, m, p} & : \quad \lambda = (j, s, m, p) \in \Lambda_s, \end{cases}$$

and $\psi_\lambda^1 = \psi_\lambda^0 \circ R_{\frac{\pi}{2}}$.

We immediately observe that the constructed system is compactly supported.

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

Proof. Let $\mathcal{SH}(\varphi_1, \psi_1; g)$ 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

$$\operatorname{supp}(G_s) \subset S_s^{-1} A_{j_0}^{-1} [-C_1, C_1]^2$$

and the elements of the shearlet-type wavelet systems satisfy

$$\operatorname{supp}(\varphi_{j_0,s,m,p}) \subset S_s^{-1} A_{j_0}^{-1} [-C_2, C_2]^2, \ \operatorname{supp}(\psi_{j,s,m,p}) \subset S_s^{-1} A_j^{-1} [-C_2, C_2]^2.$$

Hence, there exists some $C > 0$ such that

$$\operatorname{supp}(G_s * \varphi_{j_0,s,m,p}), \ \operatorname{supp}(G_s * \psi_{j,s,m,p}) \subset S_s^{-1} A_{j_0}^{-1} [-C, C]^2,$$

which proves the claim. $\qquad\square$

2.3. Comparison with Customarily Defined Shearlet Systems.

We now aim to justify the term "shearlets" by rewriting the elements of $\mathcal{SH}(\varphi_1, \psi_1; g)$ such that the resemblance with cone-adapted shearlet systems (cf. Subsection 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, which 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 $\mathcal{SH}(\varphi_1, \psi_1; g)$ be a dualizable shearlet system. Define*

$$\hat{\Theta}(\xi) = |\hat{\varphi}^0(\xi)|^2 + \sum_{j=0}^{\infty} |\hat{g}(A_j^{-1}\xi)|^2 \quad and \quad \hat{\Theta}_\ell(\xi) = \sum_{j=-\ell}^{\infty} |\hat{g}(A_j^{-1}\xi)|^2, \quad \ell \geq 0.$$

Then, for the elements of $\mathcal{SH}(\varphi_1, \psi_1; g)$, the following hold.

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

$$\psi_\lambda^0 = (\Theta * \varphi^p)(\cdot - D_p m) \quad with \ \lambda = (-1, 0, m, p) \in \Lambda_0$$

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

$$\psi_\lambda^0 = |\det(A_j)|^{1/2} (\Theta * \psi^p)(A_j \cdot - D_p m) \quad with \ \lambda = (j, 0, m, p) \in \Lambda_0$$

(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 $m \in \mathbb{Z}^2$, $p \geq 0$,*

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

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

$$\psi_\lambda^0 = |\det(A_j)|^{1/2} (\Theta_{j-j_0} * \psi^p)(S_k A_j \cdot - D_p m) \quad with \ \lambda = (j, s, m, p) \in \Lambda_s$$

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

$$A_j^{-1} S_s^{-T} = S_k^{-T} A_j^{-1}.$$

Using this relation, we obtain

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

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

3. The 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 optimally sparse approximation of cartoon-like functions. The first item will now be formally stated and proved.

Theorem 3.1. *Let* $\mathcal{SH}(\varphi_1, \psi_1; g) = \{\psi_\lambda^\ell : \lambda \in \Lambda_s, s \in \mathbb{S}, \ell = 0, 1\}$ *be a dualizable shearlet system, which constitutes a frame for* $L^2(\mathbb{R}^2)$. *Then*

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

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

$$\hat{\tilde{\psi}}_\lambda^0 = \frac{\hat{G}_s \cdot \hat{\psi}_\lambda^0}{\sum_{s' \in \mathbb{S}} |\hat{G}_{s'}|^2 + |\hat{G}_{s'} \circ R_{\frac{\pi}{2}}|^2} \quad and \quad \tilde{\psi}_\lambda^1 = \tilde{\psi}_\lambda^0 \circ R_{\frac{\pi}{2}}.$$

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

$$
\begin{aligned}
\sum_{s \in \mathbb{S}} \sum_{\lambda \in \Lambda_s} |\langle f, \psi_\lambda^0 \rangle|^2 = &= \sum_{s=s(\lceil j_0/2 \rceil, q_0) \in \mathbb{S}} \Big(\sum_{m \in \mathbb{Z}^2} \sum_{p \in \mathbb{N}_0} |\langle \overline{G}_s * f, \varphi_{j_0, s(\lceil j_0/2 \rceil, q_0), m, p} \rangle|^2 \\
&+ \sum_{j=j_0}^{\infty} \sum_{m \in \mathbb{Z}^2} \sum_{p \in \mathbb{N}_0} |\langle \overline{G}_s * f, \psi_{j, s(\lceil j_0/2 \rceil, q_0), m, p} \rangle|^2 \Big).
\end{aligned}
$$

Using the orthonormal basis property proven in Lemma 2.2, we can conclude that

$$\sum_{s \in \mathbb{S}} \sum_{\lambda \in \Lambda_s} |\langle f, \psi_\lambda^0 \rangle|^2 = \int_{\mathbb{R}^2} \Big(\sum_{s \in \mathbb{S}} |\hat{G}_s(\xi)|^2 \Big) |\hat{f}(\xi)|^2 d\xi. \tag{9}$$

Similarly, we can show that

$$\sum_{s \in \mathbb{S}} \sum_{\lambda \in \Lambda_s} |\langle f, \psi_\lambda^1 \rangle|^2 = \int_{\mathbb{R}^2} \Big(\sum_{s \in \mathbb{S}} |\hat{G}_s(R_{\frac{\pi}{2}} \xi)|^2 \Big) |\hat{f}(\xi)|^2 d\xi. \tag{10}$$

By Lemmata 2.3 and 2.4, it follows that

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

Combining this inequality with (9) and (10) 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 φ_1 and g, namely (4) and (6), which imply

$$\sum_{s\in\mathbb{S}} |\hat{G}_s(\xi)|^2 + |\hat{G}_s(R_{\frac{\pi}{2}}\xi)|^2$$

$$\geq \left(\chi_{[-1/2,1/2]^2} + \sum_{j\geq 0} \sum_{|k|\leq 2^{\lceil j/2\rceil}} \chi_{S^T_{k/2^{j/2}} A_j \Omega_g} + \chi_{S^T_{k/2^{j/2}} A_j \Omega_g} \circ R_{\frac{\pi}{2}} \right) \cdot (\min\{\delta_{\varphi_1}^2, \delta_g\})^2$$

$$\geq (\min\{\delta_{\varphi_1}^2, \delta_g\})^2 > 0 \quad \text{for a.e. } \xi.$$

The frame property of $\widetilde{\mathcal{SH}}(\varphi_1, \psi_1; g)$ can be shown by similar arguments.

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

$$\sum_{s\in\mathbb{S}} \sum_{\lambda\in\Lambda_s} \langle \hat{f}, \hat{\psi}_\lambda^0 \rangle \hat{\tilde{\psi}}_\lambda^0 = \sum_{s=s(\lceil j_0/2\rceil, q_0)\in\mathbb{S}} \frac{\hat{G}_s}{\sum_{s'\in\mathbb{S}} |\hat{G}_{s'}|^2 + |\hat{G}_{s'} \circ R_{\frac{\pi}{2}}|^2} \left(\sum_{m\in\mathbb{Z}^2} \sum_{p\in\mathbb{N}_0} \langle \overline{\hat{G}}_s \hat{f}, \hat{\varphi}_{j_0,s,m,p} \rangle \hat{\varphi}_{j_0,s,m,p} \right.$$

$$\left. + \sum_{j=j_0}^{\infty} \sum_{m\in\mathbb{Z}^2} \sum_{p\in\mathbb{N}_0} \langle \overline{\hat{G}}_s \hat{f}, \hat{\psi}_{j,s,m,p} \rangle \hat{\psi}_{j,s,m,p} \right).$$

Lemma 2.2 again implies

$$\sum_{s\in\mathbb{S}} \sum_{\lambda\in\Lambda_s} \langle \hat{f}, \hat{\psi}_\lambda^0 \rangle \hat{\tilde{\psi}}_\lambda^0 = \sum_{s\in\mathbb{S}} \frac{|\hat{G}_s|^2 \cdot \hat{f}}{\sum_{s'\in\mathbb{S}} |\hat{G}_{s'}|^2 + |\hat{G}_{s'} \circ R_{\frac{\pi}{2}}|^2}. \tag{11}$$

Similarly,

$$\sum_{s\in\mathbb{S}} \sum_{\lambda\in\Lambda_s} \langle \hat{f}, \hat{\psi}_\lambda^1 \rangle \hat{\tilde{\psi}}_\lambda^1 = \sum_{s\in\mathbb{S}} \frac{|\hat{G}_s \circ R_{\frac{\pi}{2}}|^2 \cdot \hat{f}}{\sum_{s'\in\mathbb{S}} |\hat{G}_{s'}|^2 + |\hat{G}_{s'} \circ R_{\frac{\pi}{2}}|^2}. \tag{12}$$

Using the filter properties as well as (11) and (12) finally yields

$$\sum_{\ell=0}^{1} \sum_{s\in\mathbb{S}} \sum_{\lambda\in\Lambda_s} \langle \hat{f}, \hat{\psi}_\lambda^\ell \rangle \hat{\tilde{\psi}}_\lambda^\ell = \hat{f}.$$

The theorem is proved. $\qquad\square$

We remark that the dual frame does not form a (dualizable) shearlet system. However, this was also not to be expected, since already for wavelet frames, only very few dual frames do have the form of a wavelet system [1].

4. Sparse Approximation Properties

We now turn to analyzing sparse approximation properties of dualizable shearlet systems with respect to anisotropic features which are modeled by the class of cartoon-like functions. We start by formally introducing this class, which was first defined in [10]. We remark that the superscript 2 in the notion $\mathcal{E}^2(\mathbb{R}^2)$ 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 [15].

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, ∂B has bounded curvature, and $f_i \in C^2(\mathbb{R}^2)$ satisfies $\mathrm{supp} f_i \subseteq [0,1]^2$ and $\|f_i\|_{C^2} \le 1$ for $i = 0, 1$.

We now let

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

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

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

with associated dual frame

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

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 $(\langle f, \psi_\lambda^\ell \rangle)_{\ell, \lambda}$ (cf. Section 1).

Without loss of generality, we will only consider shearlet elements $\psi_\lambda^0 \in \mathcal{SH}(\varphi_1, \psi_1; g)$ associated with one frequency cone. Since the elements ψ_λ^1 just arise as rotation by 90 degrees, they can be dealt with similarly. Hence, for the sake of brevity, we from now on omit the superscript "0", i.e., we write

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

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 [10].

Theorem 4.2 ([10]). *Let $(h_i)_{i \in I} \subseteq L^2(\mathbb{R}^2)$ be a frame for $L^2(\mathbb{R}^2)$. Then, for any $f \in \mathcal{E}^2(\mathbb{R}^2)$, the L^2-error of best N-term approximation by f_N with respect to $(h_i)_{i \in I}$ satisfies*

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

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 $\mathcal{SH}(\varphi_1, \psi_1; g) = \{\psi_\lambda^\ell : \lambda \in \Lambda_s, s \in \mathbb{S}, \ell = 0,1\}$ be a dualizable shearlet system, and let ρ be the smoothness parameter defined in (5) and (7) associated with φ_1, ψ_1, g. Further, let $f \in \mathcal{E}^2(\mathbb{R}^2)$. Then*

$$\|f - f_N\|_2 \lesssim N^{-1+\frac{15}{2}\rho} \cdot \log(N) \qquad as \ N \to \infty,$$

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

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 [17] from $\log(N)^{3/2}$ to $\log(N)$.

Corollary 4.4. *For each $f \in \mathcal{E}^2(\mathbb{R}^2)$, the compactly supported shearlet system defined in [17] provides an approximation rate of*

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

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

The proof of this corollary follows the proof of Theorem 4.3 quite closely except for slight modifications which we describe in Subsection 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 $\lambda = (j, s, m, p) \in \Lambda_s$, $s = s(\lceil j_0/2 \rceil, q_0) \in \mathbb{S}$, $j \geq j_0$ and $k \in \mathbb{Z}$ with $s = \frac{k}{2^{\lceil j/2 \rceil}}$,

$$\psi_\lambda = |\det(A_j)|^{1/2}(\Theta_{j-j_0} * \psi^p)(S_k A_j \cdot -D_p m). \tag{13}$$

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 ψ_λ 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 "normal" shearlets, the steps of the proof of the (optimal sparse) approximation result from [17] 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 [17] in the sense of Corollary 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^∞ and for an L^2 function. This is made precise in the following lemma, whose technically natured proof can be found in Subsection 6.2.

Lemma 5.1.   (i) *For $f \in L^\infty(\mathbb{R}^2)$, we have*

$$|\langle f, \psi_\lambda \rangle| \lesssim 2^{-\frac{3}{4}j} \cdot \|f\|_\infty \quad for \ all \ \lambda = (j, s, m, p) \in \Lambda.$$

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

$$|\langle f, \psi_\lambda \rangle| \lesssim 2^{-\frac{\alpha}{2}p} \cdot \|f\|_2 \quad for \ all \ \lambda = (j, s, p, m) \in \Lambda.$$

One main difficulty in proving this result is the analysis of the function $\Theta_{j-j_0} * \psi^p$, which is now the generator of the shearlet element in (13). In fact, we require a universal upper bound for this function, which is given by the following result. For its proof, we refer to Subsection 6.1.

Lemma 5.2. *There exists a universal constant C such that*

$$\|\Theta_{j-j_0} * \psi^p\|_1 \leq C \quad \text{for all } j \geq j_0, p \geq 0.$$

Aiming to drive an at least similar overall strategy as in the proof of [17], let us recall the hypotheses from this paper. One key condition is that customarily defined shearlets generated by A_j and S_k are supported in $S_s^{-1} A_j^{-1}[-C, C]^2$ for some $C > 0$ with $s = \frac{k}{2^{\lceil j/2 \rceil}}$. In this area, the C^2 curvilinear singularity of a cartoon-like function is well approximated by its tangent. However, Lemma 2.6 shows that for dualizable shearlets, we can only conclude that $\text{supp}(\psi_\lambda) \subset S_s^{-1} A_{j_0}^{-1}[-C, C]^2$. It can be computed that $\text{supp}(\psi_\lambda)$ is in fact essentially the same as the support of customarily defined shearlets 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 ψ_λ, we will approximate ψ_λ by more suitable functions of smaller supports comparable to the size of the supports of "normal" shearlets with controllable error bound. This is the essence of the following result, whose proof is outsourced to Subsection 6.3. For the following lemmata, the parameter ρ 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*

$$\hat{\psi}_\lambda^\sharp := \sum_{j'=\max\{\lfloor j(1-\rho) \rfloor, j_0\}}^{\infty} |\hat{g}(A_{j'}^{-1} S_s^{-T} \xi)|^2 \hat{\psi}_{j,s,m,p}.$$

Then the following hold.

   (i) *There exists some $C > 0$ such that*

$$\text{supp}(\psi_\lambda^\sharp) \subset S_s^{-1} A_j^{-1}\Big([-2^{j\rho}C, 2^{j\rho}C] \times [-2^{j\rho/2}C, 2^{j\rho/2}C] + m\Big).$$

   (ii) *We have*

$$|\langle f, (\psi_\lambda - \psi_\lambda^\sharp) \rangle| \lesssim 2^{-\frac{j}{2}\rho\alpha} \|f\|_2 \quad \text{for all } f \in L^2(\mathbb{R}^2).$$

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

$$\psi_\lambda^\sharp = |\det(A_j)|^{1/2} (\Theta_{j-\max\{\lfloor j(1-\rho) \rfloor, j_0\}} * \psi^p)(S_k A_j \cdot -D_p m)$$

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 $s = \frac{k}{2^{\lceil j/2 \rceil}}$. The proof of the following result is provided in Subsection 6.4.

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

$$\left| \left(\frac{\partial}{\partial \xi_2}\right)^\ell (\widehat{\Theta_{j-\max\{\lfloor j(1-\rho) \rfloor, j_0\}} * \psi^p})(\xi) \right| \lesssim \frac{\min\{1, |\xi_1|\}^{\alpha-1}}{(1+|\xi_1|)^{\beta-\gamma}(1+|\xi_2|)^\gamma}.$$

One main last ingredient, which we state as a lemma before providing the complete proof of Theorem 4.3, are decay rate of the shearlet coefficients $\langle f, \psi_\lambda \rangle$ for cartoon-like functions, where we now carefully insert conditions related to the functions $\psi_\lambda^\sharp$. Again, the proof can be found in Subsection 6.5.

Lemma 5.5. *Assume that $f \in \mathcal{E}^2(\mathbb{R}^2)$ with C^2 discontinuity curve given by $x_1 = E(x_2)$. For $\psi_\lambda \in \mathcal{SH}(\varphi_1, \psi_1; g)$ 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 $\psi_\lambda^\sharp \in L^2(\mathbb{R}^2)$ be defined as in Lemma 5.4. Let $\hat{x}_2 \in \mathbb{R}$ so that $(E(\hat{x}_2), \hat{x}_2) \in \mathrm{supp}(\psi_\lambda^\sharp)$ and $\hat{s} = E'(\hat{x}_2)$. Also let $k_s \in \mathbb{Z}$ so that $s = \frac{k_s}{2^{\lceil j/2 \rceil}}$. Then the following hold.*

(i) *If $|\hat{s}| \leq 3$, then*

$$|\langle f, \psi_\lambda \rangle| \lesssim \min\left\{ 2^{-\frac{3}{4}j}, \frac{2^{-\frac{3}{4}j} 2^{3\rho j}}{|k_s + 2^{\lceil j/2 \rceil} \hat{s}|^3} \right\}.$$

(ii) *If $|\hat{s}| > \frac{3}{2}$, then*

$$|\langle f, \psi_\lambda \rangle| \lesssim 2^{3\rho j} 2^{-\frac{9}{4}j}.$$

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 Γ of $f \in \mathcal{E}^2(\mathbb{R}^2)$ is then given by

$$\mathcal{Q}_j = \{Q_{j,\ell} : \mathrm{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset\},$$

where $\mathrm{int}(Q_{j,\ell})$ is the interior set of $Q_{j,\ell}$.

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

$$\Gamma \cap \mathrm{int}(Q_{j,\ell}) = \{(x_1, x_2) \in \mathrm{int}(Q_{j,\ell}) : x_1 = E_{j,\ell}(x_2)\}.$$

This allows us to defined

$$\mathcal{Q}_j^0 := \{Q_{j,\ell} \in \mathcal{Q}_j : \|E_{j,\ell}'\|_\infty \leq 3\}$$

and

$$\mathcal{Q}_j^1 := \mathcal{Q}_j \cap (\mathcal{Q}_j^0)^c.$$

Notice that, for all $E_{j,\ell}$ associated with $\mathcal{Q}_{j,\ell} \in Q_j^1$, we may assume

$$\inf_{(x_1,x_2)\in\mathrm{int}(Q_{j,\ell})} |E_{j,\ell}'(x_2)| > 3/2$$

for sufficiently large j.

We further define the orientation of the discontinuity curve Γ in each dyadic cube $Q_{j,\ell}$ by

$$\hat{s}_{j,\ell} = E_{j,\ell}'(\hat{x}_2) \quad \text{for some } (E_{j,\ell}(\hat{x}_2), \hat{x}_2) \in \mathrm{int}(Q_{j,\ell}) \cap \Gamma.$$

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

$$\mathbb{S}_{J/2} = \{s(j, q) = 0 : j = 0, q = 0\} \cup \{s(\lceil j/2 \rceil, q) : 0 \leq j \leq J, |q| \leq 2^{\lceil j/2 \rceil}, q \in 2\mathbb{Z} + 1\}.$$

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

$$s = \frac{k_{j,s}}{2^{\lceil j/2 \rceil}} \quad \text{for } s = s(\lceil j_0/2 \rceil, q_0) \in S_{J/2} \text{ and } j \geq j_0.$$

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

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

The smooth part, i.e., shearlet coefficients not intersecting the discontinuity curve Γ, can now be handled quickly. Following the proof of Proposition 2.1 in [17], for the approximated shearlet elements $\psi_\lambda^\sharp$ defined in Lemma 5.3, one can show that

$$\sum_{\lambda \in (\Lambda_N)^c \cap \Lambda^1} |\langle f, \psi_\lambda^\sharp \rangle|^2 \lesssim 2^{-J(1 - \frac{13}{2}\rho)}$$

with $N \sim J 2^{\frac{J(1+\rho)}{2}}$ as $J \to \infty$. By Lemma 5.3(ii), this implies

$$\sum_{\lambda \in (\Lambda_N)^c \cap \Lambda^1} |\langle f, \psi_\lambda \rangle|^2 \lesssim 2^{-J(1 - \frac{13}{2}\rho)} \quad \text{as } J \to \infty. \tag{14}$$

We now turn to analyze shearlets corresponding to Λ^0, aiming to prove that, again for $N \sim J 2^{\frac{J(1+\rho)}{2}}$,

$$\sum_{\lambda \in (\Lambda_N)^c \cap \Lambda^0} |\langle f, \psi_\lambda \rangle|^2 \lesssim 2^{-J(1 - \frac{13}{2}\rho)} \quad \text{as } J \to \infty. \tag{15}$$

For this, we fix some $J \geq 0$. Then we define subsets Λ_j^0, $j \geq 0$, of Λ^0 by

$$\Lambda_j^0 := \{\lambda = (j', s, m, p) \in \Lambda^0 : j' = j\}. \tag{16}$$

Notice that $\Lambda^0 = \bigcup_{j=0}^\infty \Lambda_j^0$. Further, for $r = 0, 1$, define again subsets of those sets by

$$\Lambda_{j,r}^0 := \Big\{\lambda = (j, s, m, p) \in \Lambda_j^0 : \text{int}(\text{supp}(\psi_\lambda^\sharp)) \cap \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset,$$

$$\text{for } p \leq \max\Big(\frac{j\rho}{2}, \frac{J\rho}{2}\Big), Q_{j,\ell} \in \mathcal{Q}_j^r\Big\} \tag{17}$$

corresponding to areas in which the discontinuity curve has a certain slope.

We further aim to collect all indices from the sets Λ_j^0 which 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 $\mathcal{Q}_j^0$, by choosing

$$\tilde{\Lambda}_j^0 := \Big\{\lambda = (j, s, m, p) \in \Lambda_j^0 : p \leq \frac{J\rho}{2}\Big\} \quad \text{for } j = 0, 1, \ldots, \lceil J/4 \rceil - 1$$

and

$$\tilde{\Lambda}_j^0 := \left\{ \lambda = (j, s, m, p) \in \Lambda_j^0 : \text{int}(\text{supp}(\psi_\lambda^\sharp)) \cap \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset \right.$$

$$\left. \text{and } |k_{j,s} + 2^{\lceil j/2 \rceil} \hat{s}_{j,\ell}| \leq 2^{\frac{J-j}{4}} \text{ for } Q_{j,\ell} \in \mathcal{Q}_j^0 \text{ and } p \leq \tfrac{J_\rho}{2} \right\} \quad \text{for } j \geq \lceil J/4 \rceil.$$

Similar considerations lead to the following selection related to the set $\mathcal{Q}_j^1$:

$$\tilde{\Lambda}_j^1 := \left\{ \lambda = (j, s, m, p) \in \Lambda_j^0 : \text{int}(\text{supp}(\psi_\lambda^\sharp)) \cap \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset \text{ for } Q_{j,\ell} \in \mathcal{Q}_j^1 \text{ and } p \leq \tfrac{J_\rho}{2} \right\}.$$

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

$$\tilde{\Lambda}^0 := \Big(\bigcup_{j=0}^{J} \tilde{\Lambda}_j^0 \Big) \bigcup \Big(\bigcup_{j=\lceil J/4 \rceil}^{\lceil J/3 \rceil} \tilde{\Lambda}_j^1 \Big). \tag{18}$$

We now turn to estimate $\#(\tilde{\Lambda}^0)$. Using the same argument as in [17](page 19), for $s \in \mathbb{S}$, $j \geq 0$ fixed and each $Q_{j,\ell} \in \mathcal{Q}_j^0$, we obtain

$$\#\Big(\Big\{ \lambda = (j, s, m, p) \in \Lambda_j^0 : \text{int}(\text{supp}(\psi_\lambda^\sharp)) \cap \text{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset \text{ for } p \leq \tfrac{J_\rho}{2} \Big\} \Big)$$

$$\lesssim 2^{\frac{J_\rho}{2}} (1 + |\hat{k}_{j,\ell}(s)|), \tag{19}$$

where $\hat{k}_{j,\ell}(s) = k_{j,s} + 2^{\lceil j/2 \rceil} \hat{s}_{j,\ell}$ and the additional factor $2^{\frac{J_\rho}{2}}$ comes from oversampling parameter p associated with the sampling matrix D_p in (13) for $\lambda = (j, s, m, p)$. Also, for $p \in \mathbb{N}_0$ and $j \geq 0$ fixed, it is immediate that

$$\#(\{\lambda = (j', s, m, p') \in \Lambda_j^0 : j' = j \text{ and } p' = p\}) \lesssim 2^{2j} 2^p \quad \text{for } j = 0, \ldots, \lceil J/4 \rceil - 1. \tag{20}$$

Finally, we obtain

$$\#(\{\lambda = (j', s, m, p') \in \Lambda_{j,r}^0 : j' = j \text{ and } p' = p\}) \lesssim 2^{\frac{3}{2}j} 2^p \quad \text{for } j \geq 0, \; r = 0, 1, \tag{21}$$

from arguing as follows: There exist at most about 2^{j+p} shearlets ψ_λ whose approximated part $\psi_\lambda^\sharp$ intersects Γ for $\lambda = (j, s, m, p)$ with fixed j, s, and p. Also, if $\lambda = (j, s, m, p) \in \Lambda_j^0$, then $s \in \mathbb{S}_{j/2}$ and $\#(\mathbb{S}_{j/2}) \lesssim 2^{j/2}$. 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 (21).

By (19)–(21), we now derive an estimate for $\#(\tilde{\Lambda}^0)$ as follows:

$$\#(\tilde{\Lambda}^0) \lesssim \sum_{j=0}^{\lceil J/4 \rceil - 1} \#(\tilde{\Lambda}_j^0) + \sum_{j=\lceil J/4 \rceil}^{\lceil J/3 \rceil} \#(\tilde{\Lambda}_j^1) + \sum_{j=\lceil J/4 \rceil}^{J} \#(\tilde{\Lambda}_j^0)$$

$$\lesssim \sum_{j=0}^{\lceil J/4 \rceil - 1} (2^{2j} 2^{\frac{J_\rho}{2}}) + \sum_{j=\lceil J/4 \rceil}^{\lceil J/3 \rceil} (2^{\frac{3}{2}j}) 2^{\frac{J_\rho}{2}}$$

$$+ \sum_{j=\lceil J/4 \rceil}^{J} \sum_{\{\ell : Q_{j,\ell} \in \mathcal{Q}_j^0\}} \sum_{\{s \in \mathbb{S}_{j/2} : |\hat{k}_{j,\ell}(s)| \leq 2^{\frac{J-j}{4}}\}} 2^{\frac{J_\rho}{2}} (1 + (1 + |\hat{k}_{j,\ell}(s)|))$$

$$\lesssim \quad 2^{\frac{J}{2}(1+\rho)} + 2^{\frac{J\rho}{2}} \sum_{j=J/4}^{J} \#(\mathcal{Q}_j^0)(2^{\frac{J-j}{2}})$$

$$\lesssim \quad J2^{\frac{J}{2}(1+\rho)}. \tag{22}$$

Let now $N > 0$ be given. Then we choose $J > 0$ such that $N \sim J2^{\frac{J}{2}(1+\rho)}$. Without loss of generality, we may assume that $N \geq \#(\tilde{\Lambda}^0)$. Then we have

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

$$\lesssim \sum_{j=\lceil J/4 \rceil}^{J} \sum_{\lambda \in \Lambda_{j,0}^0 \cap (\tilde{\Lambda}_j^0)^c} |\langle f, \psi_\lambda \rangle|^2 + \sum_{j=\lceil J/4 \rceil}^{\lceil J/3 \rceil} \sum_{\lambda \in \Lambda_{j,1}^0 \cap (\tilde{\Lambda}_j^1)^c} |\langle f, \psi_\lambda \rangle|^2$$

$$+ \sum_{j=\lceil J/3 \rceil+1}^{J} \sum_{\lambda \in \Lambda_{j,1}^0} |\langle f, \psi_\lambda \rangle|^2 + \sum_{j=0}^{J} \sum_{\lambda \in \Lambda_j^0 \cap (\Lambda_{j,0}^0 \cup \Lambda_{j,1}^0)^c} |\langle f, \psi_\lambda \rangle|^2$$

$$+ \sum_{j \geq J} \sum_{\lambda \in \Lambda_j^0} |\langle f, \psi_\lambda \rangle|^2$$

$$= \quad (I) + (II) + (III) + (IV) + (V).$$

For the second inequality above, we used the fact that $\tilde{\Lambda}_j^0 = \Lambda_{j,0}^0 \cup \Lambda_{j,1}^0$ for $j < \lceil J/4 \rceil$.

We now estimate (I) – (V). For this, for each $s \in \mathbb{S}_{j/2}$ and $Q_{j,\ell} \in \mathcal{Q}_j^0$, 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 (19), we obtain

$$(I) \lesssim \sum_{j=\lceil J/4 \rceil}^{J} \sum_{\{\ell : Q_{j,\ell} \in \mathcal{Q}_j^0\}} \sum_{\{s \in \mathbb{S}_{j/2} : |\hat{k}_{j\ell}(s)| > 2^{\frac{J-j}{4}}\}} 2^{\frac{J\rho}{2}}(1 + |\hat{k}_{j,\ell}(s)|) \left(\frac{2^{3\rho j} 2^{-\frac{3}{4}j}}{|\hat{k}_{j,\ell}(s)|^3} \right)^2$$

$$\lesssim \sum_{j=\lceil J/4 \rceil}^{J} \sum_{\{\ell : Q_{j,\ell} \in \mathcal{Q}_j^0\}} 2^{\frac{J\rho}{2}} 2^{6\rho j} 2^{-\frac{3}{2}j} (2^{-\frac{J-j}{4}})^4$$

$$\lesssim \sum_{j=\lceil J/4 \rceil}^{J} (2^{j/2}) 2^{\frac{J\rho}{2}} 2^{6\rho j} 2^{-\frac{3}{2}j} (2^{-\frac{J-j}{4}})^4 \lesssim 2^{-J(1-\frac{13}{2}\rho)}. \tag{23}$$

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 (21) to obtain

$$(III) \lesssim \sum_{j=\lceil J/3 \rceil}^{J} 2^{\frac{J\rho}{2}}(2^{\frac{3}{2}j})(2^{3\rho j} 2^{-\frac{9}{4}j})^2 \lesssim \sum_{j=\lceil J/3 \rceil}^{J} 2^{\frac{J\rho}{2}} 2^{-3j(1-2\rho)} \lesssim 2^{\frac{J\rho}{2}} 2^{-J(1-2\rho)} = 2^{-J(1-5/2\rho)}. \tag{24}$$

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

Term (IV) is estimated by using Lemma 5.1(ii) and (20). Using also $\alpha \geq \frac{6}{\rho} + 1$, we have

$$
\begin{aligned}
\text{(IV)} &\lesssim \sum_{j=0}^{J}\sum_{p=0}^{\infty}(2^{\frac{J\rho}{2}+p})(2^{2j})(2^{-\alpha(\frac{J}{4}\rho+\frac{p}{2})})^2 \\
&\lesssim \sum_{j=0}^{J}\Big(\sum_{p=0}^{\infty}2^{-p(\alpha-1)}\Big)2^{\frac{J\rho}{2}}2^{2j}2^{-\alpha\frac{J\rho}{2}} \\
&\lesssim \sum_{j=0}^{J}2^{2J-\alpha\frac{J\rho}{2}+\frac{J\rho}{2}} \lesssim 2^{-J}.
\end{aligned}
\tag{25}
$$

Finally, the last term can be estimated by

$$
\begin{aligned}
\text{(V)} &\leq \sum_{j\geq J}\sum_{\lambda\in\Lambda_{j,0}^0}|\langle f,\psi_\lambda\rangle|^2 + \sum_{j\geq J}\sum_{\lambda\in\Lambda_{j,1}^0}|\langle f,\psi_\lambda\rangle|^2 + \sum_{j\geq J}\sum_{\lambda\in\Lambda_j\cap(\Lambda_{j,0}^0\cup\Lambda_{j,1}^0)^c}|\langle f,\psi_\lambda\rangle|^2 \\
&= \text{(A)} + \text{(B)} + \text{(C)}.
\end{aligned}
\tag{26}
$$

It remains to analyze the terms (A) – (C). We start with (A). By Lemmata 5.1(i) and 5.5(i) we well as (19), we obtain

$$
\begin{aligned}
\text{(A)} &\lesssim \sum_{j\geq J}\sum_{\{\ell:Q_{j,\ell}\in\mathcal{Q}_j^0\}}2^{\frac{j\rho}{2}}\sum_{s\in\mathbb{S}_{j/2}}(1+|\hat{k}_{j,\ell}(s)|)\min\Big\{2^{-\frac{3}{4}j},\frac{2^{3\rho j}2^{-\frac{3}{4}j}}{|\hat{k}_{j,\ell}(s)|^3}\Big\}^2 \\
&\lesssim \sum_{j\geq J}2^{\frac{j\rho}{2}}(\#(\mathcal{Q}_j^0)2^{-\frac{3}{2}j}2^{6\rho j}) \\
&\lesssim \sum_{j\geq J}2^{-j}2^{\frac{13j\rho}{2}} \lesssim 2^{-J(1-\frac{13\rho}{2})}.
\end{aligned}
$$

The terms (B) and (C) can be estimated by using Lemmata 5.5(ii) and 5.1(ii) as well as equations (20) and (21) to obtain

$$
\text{(B)} \lesssim \sum_{j\geq J}(2^{\frac{3}{2}j})(2^{\frac{j\rho}{2}})(2^{-\frac{9}{4}j}2^{3\rho j})^2 = \sum_{j\geq J}2^{-3j}2^{6\rho j}2^{\frac{j\rho}{2}} \lesssim 2^{-3J+\frac{13}{2}\rho J}.
$$

and

$$
\text{(C)} \lesssim \sum_{j\geq J}\sum_{p=0}^{\infty}(2^{\frac{j\rho}{2}+p})(2^{2j})(2^{-\alpha(j\frac{\rho}{4}+\frac{p}{2})})^2 = \sum_{j\geq J}\Big(\sum_{p=0}^{\infty}2^{-p(\alpha-1)}\Big)2^{2j+\frac{j\rho}{2}-j\frac{\alpha\rho}{2}} \lesssim 2^{-J}.
$$

Thus, continuing (26),

$$
\text{(V)} \lesssim 2^{-J(1-\frac{13\rho}{2})}.
\tag{27}
$$

Summarizing, (23)—(27) imply (15).

Finally, using (14), (15), and the frame property of the shearlet system $\mathcal{SH}(\varphi_1,\psi_1;g)$, we can conclude that

$$
\|f-f_N\|_2^2 \lesssim \sum_{\lambda\notin\Lambda_N}|\langle f,\psi_\lambda\rangle|^2 \lesssim 2^{-J(1-\frac{13\rho}{2})}
$$

with $N \sim J2^{\frac{J(1+\rho)}{2}}$, which implies our claim. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

6. Proofs of Preliminary Lemmata and Corollary 4.4

6.1. **Proof of Lemma 5.2.** We first observe that

$$(\widehat{Q_{j-j_0} * \psi^p})(\xi) = \sum_{j'=j_0-j}^{\infty} \hat{h}_{j',p}(\xi), \quad \text{where } \hat{h}_{j',p}(\xi) := |\hat{g}(A_{j'}^{-1}\xi)|^2 \hat{\psi}^p(\xi),$$

with

$$\operatorname{supp}(h_{j',p}) \subset A_{\max\{-j',0\}}[-C,C]^2$$

for some $C > 0$. We can then estimate $\|Q_{j-j_0} * \psi^p\|_1$ by

$$
\begin{aligned}
\|Q_{j-j_0} * \psi^p\|_1 &= \int_{\mathbb{R}^2} \Big| \int_{\mathbb{R}^2} (\widehat{Q_{j-j_0} * \psi^p})(\xi) e^{2\pi i \langle \xi, x \rangle} d\xi \Big| dx \\
&\lesssim \sum_{j'=0}^{\infty} \int_{[-C,C]^2} \int_{\mathbb{R}^2} |\hat{h}_{j',p}(\xi)| d\xi dx + \sum_{j'=0}^{j-j_0} \int_{A_{j'}[-C,C]^2} \int_{\mathbb{R}^2} |\hat{h}_{-j',p}(\xi)| d\xi dx \\
&= (\mathrm{I}) + (\mathrm{II}).
\end{aligned}
$$

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

$$\int_{\mathbb{R}} \frac{2^{-j_1} \min\{1, |2^{-j_2}x|\}^{\alpha}}{(1 + |2^{-j_1}x|)^{\beta}} dx \lesssim 2^{-\alpha(j_2 - j_1)}. \qquad (28)$$

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

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

Second,

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

$$\lesssim \sum_{j=0}^{\infty} \int_{A_j[-C,C]^2} 2^{-j} 2^{-\alpha j} \int_{\mathbb{R}} \frac{\min\{1, |2^{-p}\xi_2|\}^2}{(1 + |2^{j/2}\xi_2|)^2 (1 + |2^{-p}\xi_2|)^2} d\xi_2 dx$$

$$\lesssim \sum_{j=0}^{\infty} \int_{A_j[-C,C]^2} 2^{-j} 2^{-\alpha j} \int_{\mathbb{R}} \frac{2^{-j/2}}{(1 + |\xi_2|)^2} d\xi_2 dx$$

$$\lesssim \sum_{j=0}^{\infty} 2^{\frac{3}{2} \cdot j} 2^{-j} 2^{-\alpha j} 2^{-j/2} \leq 2.$$

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

$$\psi_\lambda = |\det(A_j)|^{1/2} (\Theta_{j-j_0} * \psi^p)(S_k A_j \cdot - D_p m),$$

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

$$\begin{aligned}
|\langle f, \psi_\lambda \rangle| &\leq |\det(A_j)|^{1/2} \int_{\mathbb{R}^2} |(\Theta_{j-j_0} * \psi^p)(S_k A_j x - D_p m) f(x)| dx \\
&\leq 2^{-\frac{3}{4}j} \int_{\mathbb{R}^2} |f(A_j^{-1} S_k^{-1}(y + D_p m))| |(\Theta_{j-j_0} * \psi^p)(y)| dy \\
&\leq 2^{-\frac{3}{4}j} \|f\|_\infty \|\Theta_{j-j_0} * \psi^p\|_1.
\end{aligned}$$

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

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

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

Note that (5) and (7) imply that

$$|\hat{\psi}_{j,0,m,p}(\xi)|^2 \lesssim 2^{-\frac{3}{2}j-p} \cdot \frac{\min\{1, |2^{-j}\xi_1|^{2\alpha}\}}{(1 + |2^{-j}\xi_1|)^{2\beta}} \cdot \frac{\min\{1, |2^{-j/2-p}\xi_2|^{2\alpha}\}}{(1 + |2^{-j/2-p}\xi_2|)^{2\beta}}. \tag{29}$$

Hence, using (7), (28) and (29), can estimate (I) by

$$
\begin{aligned}
\text{(I)} \quad &\lesssim \quad \sum_{j'=0}^{j+p} 2^{-j/2-p} \int_{\mathbb{R}} \frac{\min\{1, |2^{-j/2-p}\xi_2|^{2\alpha}\}}{(1+|2^{-j'/2}\xi_2|)^{2\beta}} d\xi_2 \int_{\mathbb{R}} \frac{2^{-j}}{(1+|2^{-j}\xi_1|)^{2\beta}} d\xi_1 \\
&\lesssim \quad \sum_{j'=0}^{j+p} 2^{-j/2-p} \int_{\mathbb{R}} \frac{\min\{1, |2^{-j/2-p}\xi_2|^{2\alpha}\}}{(1+|2^{-j'/2}\xi_2|)^{2\beta}} d\xi_2 \\
&\lesssim \quad 2^{-\alpha p}.
\end{aligned}
$$

Similarly, we can estimate (II) as follows:

$$
\begin{aligned}
\text{(II)} \quad &\lesssim \quad \sum_{j'=j+p}^{\infty} 2^{-j} \int_{\mathbb{R}} \frac{\min\{1, |2^{-j'}\xi_1|^{2\alpha}\}}{(1+|2^{-j}\xi_1|)^{2\beta}} d\xi_1 \int_{\mathbb{R}} \frac{2^{-j/2-p}}{(1+|2^{-j/2-p}\xi_2|)^{2\beta}} d\xi_2 \\
&\lesssim \quad \sum_{j'=j+p}^{\infty} 2^{-j} \int_{\mathbb{R}} \frac{\min\{1, |2^{-j'}\xi_1|^{2\alpha}\}}{(1+|2^{-j}\xi_1|)^{2\beta}} d\xi_1 \\
&\lesssim \quad 2^{-2\alpha p}.
\end{aligned}
$$

This proves (ii).

6.3. Proof of Lemma 5.3.
First, note that (i) is obvious from the definition of $\psi_\lambda^\sharp$ and

$$
|\langle f, (\psi_\lambda - \psi_\lambda^\sharp) \rangle|^2 \le \|f\|_2^2 \|\psi_\lambda - \psi_\lambda^\sharp\|_2^2.
$$

Next, we estimate $\|\psi_\lambda - \psi_\lambda^\sharp\|_2^2$ to show (ii). By using (5), (7), and (28), we obtain

$$
\begin{aligned}
\|\psi_\lambda - \psi_\lambda^\sharp\|_2^2 \quad &\lesssim \quad \int_{\mathbb{R}^2} \Big| \sum_{j'=0}^{\lfloor j(1-\rho) \rfloor} |\hat{g}(A_{j'}^{-1}\xi)|^2 |\hat{\psi}_{j,0,0,p}(\xi)| \Big|^2 d\xi \\
&\lesssim \quad \int_{\mathbb{R}^2} \sum_{j'=0}^{\lfloor j(1-\rho) \rfloor} |\hat{g}(A_{j'}^{-1}\xi)|^2 |\hat{\psi}_{j,0,0,p}(\xi)|^2 d\xi \\
&\lesssim \quad \sum_{j'=0}^{\lfloor j(1-\rho) \rfloor} \int_{\mathbb{R}^2} \frac{2^{-j'} \min\{1, |2^{-j}\xi_1|\}^\alpha}{(1+|2^{-j'}\xi_1|)^\beta} \cdot \frac{2^{-p-j/2}}{(1+|2^{-j/2-p}\xi_2|)^{2\beta}} d\xi \\
&\lesssim \quad \sum_{j'=0}^{\lfloor j(1-\rho) \rfloor} 2^{-(j-j')\alpha} \lesssim 2^{-j\rho\alpha}.
\end{aligned}
$$

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

$$
\left| \left(\frac{\partial}{\partial \xi_2} \right) \widehat{\Theta_{j-\max\{\lfloor j(1-\rho) \rfloor, j_0\}} * \psi^p}(\xi) \right|
$$

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

$$+ \sum_{j=1}^{\infty}\left|\left(\frac{\partial}{\partial\xi_2}\right)|\hat{g}(A_j^{-1}\xi)|^2\hat{\psi}^p(\xi)\right| + \sum_{j=1}^{\infty}\left||\hat{g}(A_j^{-1}\xi)|^2\left(\frac{\partial}{\partial\xi_2}\right)\hat{\psi}^p(\xi)\right|$$

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

$$+ \sum_{j=1}^{\infty}\frac{\min\{1,|2^{-j}\xi_1|^\alpha\}\min\{1,|\xi_1|^\alpha\}\min\{1,|2^{-p}\xi_2|^\alpha\}}{(1+|2^{-j}\xi_1|)^\beta(1+|2^{-j/2}\xi_2|)^\beta(1+|\xi_1|)^\beta(1+|2^{-p}\xi_2|)^\beta}$$

$$= \text{(I)} + \text{(II)}$$

Then we have

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

$$\lesssim \quad \frac{\min\{1,|\xi_1|^{\alpha-1}\}}{(1+|\xi_2|)^{\beta-1}(1+|\xi_1|)^\beta}$$

and

$$\text{(II)} \quad \lesssim \quad \sum_{j=1}^{\infty}\frac{\min\{1,|\xi_1|^\alpha\}2^{-p\gamma}|\xi_2|^\gamma 2^{\frac{j}{2}\gamma}2^{-j\gamma}|\xi_1|^\gamma}{(1+|\xi_1|)^\beta(1+|2^{-p}\xi_2|)^\gamma|\xi_2|^\gamma}$$

$$\lesssim \quad \frac{\min\{1,|\xi_1|^\alpha\}}{(1+|\xi_1|)^{\beta-\gamma}(1+|\xi_2|)^\gamma}.$$

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

$$q_{j,k_s,m} := |\det(A_j)|^{1/2}q(S_{k_s}A_j \cdot -m)$$

for $j \geq 0$, $k_s \in \mathbb{Z}$ and $m \in \mathbb{Z}^2$. Further, assume that

$$\text{supp}(q_{j,k_s,m}) \subset A_j^{-1}S_{k_s}^{-1}\Big([-2^{\rho j}L, 2^{\rho j}L] \times [-2^{\rho j/2}L, 2^{\rho j/2}L] + m\Big)$$

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 [17], we can show that

$$|\langle f, q_{j,k_s,m}\rangle| \lesssim \frac{2^{-\frac{3}{4}j}2^{3\rho j}}{|k_s + 2^{\lceil j/2\rceil}\hat{s}|^3} \quad \text{if } |\hat{s}| \leq 3 \tag{30}$$

and

$$|\langle f, q_{j,k_s,m}\rangle| \lesssim 2^{3\rho j}2^{-\frac{9}{4}} \quad \text{if } |\hat{s}| > 3/2. \tag{31}$$

We next choose $q_{j,k_s,m} := \psi_\lambda^\sharp$ for $\lambda = (j,s,m,p) \in \Lambda_s$ with $s = s(\lceil j_0/2 \rceil, q_0) \in \mathbb{S}$, $s = \frac{k_s}{2^{\lceil j/2 \rceil}}$, and $j \geq j_0$. Hence, in particular, $q = \Theta_{j-\max\{\lfloor j(1-\rho)\rfloor, j_0\}} * \psi^p$. By Lemma 5.3(i) and Lemma 5.4, we derive (30) and (31) for this choice. Thus, by Lemma 5.3(ii),

$$|\langle f, \psi_\lambda \rangle| \lesssim \frac{2^{-\frac{3}{4}j}2^{3\rho j}}{|k_s + 2^{\lceil j/2 \rceil}\hat{s}|^3} + 2^{-\frac{j}{2}\rho\alpha} \quad \text{if } |\hat{s}| \leq 3 \tag{32}$$

and

$$|\langle f, \psi_\lambda \rangle| \lesssim 2^{3\rho j}2^{-\frac{9}{4}j} + 2^{-\frac{j}{2}\rho\alpha} \quad \text{if } |\hat{s}| > 3/2. \tag{33}$$

Finally, Lemma 5.1(i), (32), and (33) imply (i) and (ii) for $\alpha \geq \frac{6}{\rho}$.

6.6. Proof of Corollary 4.4.

We will retain all notations used in the proof of Theorem 4.3. In the considered case, a compactly supported function $\psi \in L^2(\mathbb{R}^2)$ can be chosen so that shearlets are defined by

$$\psi_\lambda = |\det(A_j)|^{1/2}\psi(S_k A_j \cdot - \mathrm{diag}(c_1, c_2)m)$$

for $\lambda \in \Lambda_s$ with $s = s(\lceil j_0/2 \rceil, q_0) \in \mathbb{S}$ and $k \in \mathbb{Z}$ with $s = \frac{k}{2^{\lceil j/2 \rceil}}$. We emphasize that the additional oversampling matrix D_p is not needed and the index set Λ_s originally defined in Definition 2.5 is given as

$$\Lambda_s = \{\lambda = (j,s,m) : j \geq j_0, m \in \mathbb{Z}^2\}.$$

Further, the shearlet generator ψ can be chosen so that it satisfies a directional vanishing moment condition (compare Lemma 5.4) in frequency, and that the shearlets ψ_λ satisfy a support condition (compare Lemma 5.3(i)) with $\rho = 0$, which yields

$$\mathrm{supp}(\psi_\lambda) \subset S_s^{-1}A_j^{-1}([-C, C]^2 + m) \quad \text{for some } C > 0. \tag{34}$$

These two conditions imply Lemma 5.5 (i)–(ii) with $\rho = 0$, which can be derived by using similar arguments as in the proof of Proposition 2.2 in [17].

Now consider index sets Λ_j^0, $\Lambda_{j,r}^0$, $\tilde{\Lambda}_j^r$ and $\tilde{\Lambda}^0$ as defined in (16) – (18) for $r = 0, 1$. Notice that for those index sets, the additional index $p \in \mathbb{N}_0$ for the sampling matrix D_p is not needed, and we have

$$\Lambda_j^0 = \{\lambda = (j', s, m) \in \Lambda^0 : j' = j\}$$

as well as

$$\Lambda_{j,r}^0 = \{\lambda = (j,s,m) \in \Lambda_j^0 : \mathrm{int}(\mathrm{supp}(\psi_\lambda)) \cap \mathrm{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset, \text{ for } Q_{j,\ell} \in \mathcal{Q}_j^r\}$$

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

$$\tilde{\Lambda}_j^0 = \{\lambda = (j,s,m) \in \Lambda_j^0 : \mathrm{int}(\mathrm{supp}(\psi_\lambda)) \cap \mathrm{int}(Q_{j,\ell}) \cap \Gamma \neq \emptyset, |k_{j,s} + 2^{\lceil j/2 \rceil}\hat{s}_{j,\ell}| \leq 2^{\frac{J-j}{4}}, Q_{j,\ell} \in \mathcal{Q}_j^0\}.$$

Also,

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

Applying the same estimate as (22) with those index sets $\tilde{\Lambda}_j^r$, we obtain

$$\#(\tilde{\Lambda}^0) \lesssim J2^{\frac{J}{2}}$$

Note that $\rho = 0$ in (22), since the additional index $p \in \mathbb{N}_0$ is not required for $\tilde{\Lambda}_j^r$ in this case. Applying (23)—(27) with (34) and Lemma 5.5 (i)–(ii) with $\rho = 0$, we obtain

$$\|f - f_N\|_2^2 \lesssim \sum_{\lambda \notin \Lambda_N} |\langle f, \psi_\lambda \rangle|^2 \lesssim 2^{-J}$$

with $N \sim J2^{\frac{J}{2}}$. This proves our claim.

References

[1] M. Bownik and J. Lemvig, *The canonical and alternate duals of a wavelet frame*, Appl. Comput. Harmon. Anal. **23** (2007), 263–272.

[2] E. J. Candès and D. L. Donoho, *New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities*, Comm. Pure Appl. Math. **56** (2004), 216–266.

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

[4] P. Casazza and G. Kutyniok, eds., *Finite Frames: Theory and Applications*, Birkhäuser Boston, 2012.

[5] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, Boston, 2003.

[6] W. Dahmen, C. Huang, C. Schwab, and G. Welper, *Adaptive Petrov-Galerkin methods for first order transport equations*, SIAM J. Numer. Anal. **50** (2012), 2420-2445.

[7] W. Dahmen, G. Kutyniok, W.-Q Lim, C. Schwab, and G. Welper, *Adaptive anisotropic discretizations for transport equations*, preprint.

[8] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[9] M. Davenport, M. Duarte, Y. Eldar, and G. Kutyniok, *Introduction to Compressed Sensing*, in: Compressed Sensing: Theory and Applications, 1-68, Cambridge University Press, 2012.

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

[11] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory **52** (2006), 1289–1306.

[12] P. Grohs, *Bandlimited Shearlet Frames with nice Duals*, J. Comput. Appl. Math **243** (2013), 139–151.

[13] K. Guo, G. Kutyniok, and D. Labate, *Sparse multidimensional representations using anisotropic dilation and shear operators*, in: Wavelets and Splines (Athens, GA, 2005), Nashboro Press, Nashville, TN, 2006, 189–201.

[14] G. Kutyniok and D. Labate, *Introduction to Shearlets*, in: Shearlets: Multiscale Analysis for Multivariate Data, 1–38, Birkhäuser Boston, 2012.

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

[16] P. Kittipoom, G. Kutyniok and W.-Q Lim, *Construction of compactly supported shearlet frames*, Constr. Approx. **35** (2012), 21–72.

[17] G. Kutyniok and W.-Q Lim, *Compactly supported shearlets are optimally sparse*, J. Approx. Theory, **163** (2011), 1564–1589.

[18] G. Kutyniok, W.-Q Lim, and R. Reisenhofer, *ShearLab 3D: Faithful Digital Shearlet Transforms based on Compactly Supported Shearlets*, preprint.

[19] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, *The Cosparse Analysis Model and Algorithms*, Appl. Comput. Harmon. Anal., **34** (2013), 30–56.

Department of Mathematics, Technische Universität Berlin, 10623 Berlin, Germany
E-mail address: kutyniok@math.tu-berlin.de

Department of Mathematics, Technische Universität Berlin, 10623 Berlin, Germany
E-mail address: lim@math.tu-berlin.de