Appl. Comput. Harmon. Anal. 38 (2015) 87–114

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Gabor shearlets

Bernhard G. Bodmann[a,*,1], Gitta Kutyniok[b,2], Xiaosheng Zhuang[c,3]

[a] *Department of Mathematics, University of Houston, USA*
[b] *Institute of Mathematics, Technische Universität Berlin, Germany*
[c] *Department of Mathematics, City University of Hong Kong, China*

ARTICLE INFO

Article history:
Received 27 March 2013
Received in revised form 7 February 2014
Accepted 20 March 2014
Available online 25 March 2014
Communicated by Christopher Heil

Keywords:
Gabor shearlets
Cartoon-like functions
Cone-adapted shearlets
Gabor frames
Orthonormal wavelets
Redundancy
Sparse approximation
Shearlets
Tight frames

ABSTRACT

In this paper, we introduce Gabor shearlets, a variant of shearlet systems, which are based on a different group representation than previous shearlet constructions: they combine elements from Gabor and wavelet frames in their construction. As a consequence, they can be implemented with standard filters from wavelet theory in combination with standard Gabor windows. Unlike the usual shearlets, the new construction can achieve a redundancy as close to one as desired. Our construction follows the general strategy for shearlets. First we define group-based Gabor shearlets and then modify them to a cone-adapted version. In combination with Meyer filters, the cone-adapted Gabor shearlets constitute a tight frame and provide low-redundancy sparse approximations of the common model class of anisotropic features which are cartoon-like functions.

1. Introduction

During the last 10 years, directional representation systems such as curvelets and shearlets were introduced to accommodate the need for sparse approximations of anisotropic features in multivariate data. These anisotropic features, such as singularities on lower dimensional embedded manifolds, called for representation systems to sparsely approximate such data. Prominent examples in the 2-dimensional setting are edge-like structures in images in the regime of explicitly given data and shock fronts in transport equations

* Corresponding author.
E-mail addresses: bgb@math.uh.edu (B.G. Bodmann), kutyniok@math.tu-berlin.de (G. Kutyniok), xzhuang7@cityu.edu.hk (X. Zhuang).

[1] Research was in part supported by National Science Foundation grant NSF DMS-1109545. Part of this work was completed while visiting the Technische Universität Berlin and the Erwin Schrödinger Institut Wien.
[2] Research was supported in part by the Einstein Foundation Berlin, by Deutsche Forschungsgemeinschaft (DFG) Grant SPP-1324 KU 1446/13 and Deutsche Forschungsgemeinschaft Grant KU 1446/14, by the DFG Collaborative Research Center TRR 109 "Discretization in Geometry and Dynamics", and by the DFG Research Center Matheon "Mathematics for key technologies" in Berlin.
[3] Research was supported in part by Research Grants Council of Hong Kong [Project No. CityU 108913].

http://dx.doi.org/10.1016/j.acha.2014.03.006
1063-5203/

in the regime of implicitly given data. Because of their isotropic nature, wavelets are not as well adapted to this task as curvelets [3], contourlets [6], or shearlets [19]. Recently, a general framework for directional representation systems based on parabolic scaling – a scaling adapted to the fact that the regularity of the singularity in the considered model is C^2 – was introduced in [8] seeking to provide a comprehensive viewpoint towards sparse approximations of cartoon-like functions.

Among these representation systems, shearlets distinguished themselves by the fact that they are available as compactly supported systems – which is desirable for applications requiring high spatial localization such as PDE solvers – and also provide a unified treatment of the continuum and digital setting thereby ensuring faithful implementations. Shearlets were introduced in [9] with the early theory focusing on band-limited shearlets, see e.g. [11]. Later, a compactly supported variant was introduced in [18], which again provides optimally sparse approximations of cartoon-like functions [20]. In contrast to those properties, contourlets do not provide optimally sparse approximations and curvelets are neither compactly supported nor do they treat the continuum and digital realm uniformly due to the fact that they are based on rotations in contrast to shearing.

1.1. Key problem

One major problem – which might even be considered a "holy grail" of the area of geometric multiscale analysis – is whether a system can be designed to be

(P1) an orthonormal basis,
(P2) compactly supported,
(P3) possessing a multiresolution structure,
(P4) and providing optimally sparse approximations of cartoon-like functions.

Focusing from now on entirely on shearlets, we observe that bandlimited shearlets satisfy (P4) while replacing (P1) with being a tight frame. Compactly supported shearlets accommodate (P2) and (P4), and form a frame with controllable frame bounds as a substitute for (P1). We are still far from being able to construct a system satisfying all those properties – also by going beyond shearlets – , and it is not even clear whether this is at all possible, cf. also [17]. Several further attempts were already made in the past. In [21], shearlet systems were introduced based on a subdivision scheme, which naturally leads to (P2) and (P3), but not (P1) – not even being tight – and (P4). In [13], a different multiresolution approach was utilized leading to systems which satisfy (P2) and (P3), but not (P4), and (P1) only by forming a tight frame without results on their redundancy.

1.2. What are Gabor shearlets?

The main idea of the present construction is to use a deformation of the group operation with which common shearlet systems are generated, together with a decomposition in the frequency domain to ensure an almost uniform treatment of different directions, while modeling the systems as closely as possible after the one-dimensional multiresolution analysis (MRA) wavelets. To be more precise, the new group operation includes shears and chirp modulations which satisfy the well-studied Weyl–Heisenberg commutation relations. Thus, the shear part naturally leads us to Gabor frame constructions instead of an alternative viewpoint in which shears enter in composite dilations [10]. The filters appearing in this construction can be chosen as the trigonometric polynomials belonging to standard wavelets or to M-band versions of them, or as the smooth filters associated with Meyer's construction. To achieve the optimal approximation rate for cartoon-like functions, we use a cone adaptation procedure. But in contrast to other constructions, we avoid incorporating redundancy in this step.

Due to the different group structure, the authors were not able to show that Gabor shearlets fall into the framework of parabolic molecules (cf. [8]) although they are based on parabolic scaling. Thus, we cannot apply this framework in a straightforward way in our situation for deriving results on sparse approximations by transferring such properties from other systems. We comment on differences with the structure of parabolic molecules in Section 5.

1.3. Our contributions

Gabor shearlets satisfy the following properties, related to Section 1.1:

- (P1*) Gabor shearlets can be chosen to be unit norm and b^{-1}-tight, where b^{-1} – which can be interpreted as the redundancy (cf. Section 2.4) – can be chosen arbitrarily close to one.
- (P2*) Gabor shearlets are not compactly supported, but can be constructed with polynomial decay in the spatial domain.
- (P3) The two-scale relation for the shearlet subband decomposition is implemented with standard filters related to MRA wavelets.
- (P4) In conjunction with a cone-adaptation strategy and Meyer filters, Gabor shearlets provide optimally sparse approximations of cartoon-like functions.

Thus, (P3) and (P4) are satisfied. (P1) is approximately satisfied in the sense that the systems with property (P1*) are nearly orthonormal bases. And (P2) is also approximately satisfied by replacing compact support by polynomial decay in (P2*). It is in this sense that we believe the development of Gabor shearlets contributes to developing a system satisfying the desired properties. Or – if it could be proven that those are not simultaneously satisfiable – providing a close approximation to those.

1.4. Outline of the paper

The remainder of this paper is organized as follows. In Section 2, we set the notation and recall the essential properties of Gabor systems, wavelets, and shearlets which are needed in the sequel. In this section, we also briefly introduce the notion of redundancy first advocated in [1]. In Section 3, after providing some intuition on our approach, we introduce Gabor shearlets based on a group related to chirp modulations and discuss their frame properties and the associated multiresolution structure. The projection of those Gabor shearlets on cones in the frequency domain is then the focus of Section 4, again starting with the construction followed by a discussion of similar properties as before. The last section, Section 5, contains the analysis of sparse approximation properties of cone-adapted Gabor shearlets.

2. Revisited: wavelets, shearlets, and Gabor systems

In this section, we introduce the main notation of this paper, state the basic definitions of Gabor systems, wavelets, and shearlets, and also recall the underlying construction principles, formulated in such a way that Gabor shearlets will become a relatively straightforward generalization. We emphasize that this is not an introduction to Gabor and wavelet theory, and we expect the reader to have some background knowledge, otherwise we refer to [4] or [22]. A good general reference for most of the material presented in this section is the book by Weiss and Hernández [24]. In the last part of this section, we discuss the viewpoint of redundancy from [1], which we adopt in this paper.

In what follows, the Fourier transform of $f \in L^1(\mathbb{R}^n)$ is defined to be $\widehat{f}(\xi) := \int_{\mathbb{R}^n} f(x)e^{-2\pi i x\cdot\xi}\,dx$, where $x \cdot \xi$ is the dot product between x and ξ in $\mathbb{R}^n$. As usual, we extend this integral transform to the unitary map $f \mapsto \widehat{f}$ defined for any function f which is square integrable. The unitarity is captured in the Plancherel identity $\langle f, g\rangle = \langle \widehat{f}, \widehat{g}\rangle$ for any two functions $f, g \in L^2(\mathbb{R}^n)$ with $\langle f, g\rangle := \int_{\mathbb{R}^n} f(x)\overline{g(x)}\, dx$.

2.1. MRA wavelets

Let $\{\phi, \psi\}$ be a pair of a scaling function and a wavelet for $L^2(\mathbb{R})$ associated with a pair of a low-pass filter $H : \mathbb{T} \to \mathbb{C}$ and a high-pass filter $G : \mathbb{T} \to \mathbb{C}$, for convenience defined on the torus $\mathbb{T} = \{z \in \mathbb{C}\colon |z| = 1\}$. We start by recalling the Smith–Barnwell condition for filters.

Definition 2.1. A filter $H : \mathbb{T} \to \mathbb{C}$ satisfies the *Smith–Barnwell condition*, if

$$\left|H(z)\right|^2 + \left|H(-z)\right|^2 = 1$$

for almost every $z \in \mathbb{T}$.

The Smith–Barnwell condition is an essential ingredient in the characterization of localized multiresolution analyses; that is, the scaling functions ϕ are localized in the sense of having faster than polynomial decay: $\int_{\mathbb{R}} (1 + x^2)^n |\phi(x)|^2\, dx < \infty$ for all $n \in \mathbb{N}$.

Theorem 2.1 *(Cohen, as in [24] Theorem 4.23 of Chapter 7). A C^∞ function $H : \mathbb{T} \to \mathbb{C}$ is the low-pass filter of a localized multiresolution analysis with scaling function ϕ given by*

$$\widehat{\phi}(\xi) = \prod_{j=1}^{\infty} H\bigl(e^{-2\pi i\xi/2^j}\bigr)$$

if and only if $H(1) = 1$, H satisfies the Smith–Barnwell condition, and there exists a set $K \subset \mathbb{T}$ which contains 1 and has a finite complement in $\mathbb{T}$ such that $H(z^{2^{-j}}) \neq 0$ for all $j \in \mathbb{Z}$, $j \geqslant 0$, and $z \in K$.

The two-scale relations for ϕ and ψ are conveniently expressed in the frequency domain,

$$\widehat{\phi}(2\xi) = H\bigl(e^{-2\pi i\xi}\bigr)\widehat{\phi}(\xi) \quad \text{and} \quad \widehat{\psi}(2\xi) = G\bigl(e^{-2\pi i\xi}\bigr)\widehat{\phi}(\xi), \quad \text{a.e. } \xi \in \mathbb{R}.$$

The orthonormality of the integer translates of $\{\phi, \psi\}$ is captured in the matrix identity

$$\mathcal{M}(z)\mathcal{M}(z)^* = I_2 \quad \text{with } \mathcal{M}(z) := \begin{bmatrix} H(z) & H(-z) \\ G(z) & G(-z) \end{bmatrix}, \quad \text{for a.e. } z \in \mathbb{T}.$$

Often, only H is specified and the matrix has to be completed to a unitary, with a common choice being $G(z) = -z\overline{H(-z)}$.

The low-pass filter of the Meyer scaling function is of particular use for the construction of Gabor shearlets, which will be shown in Section 5 to yield optimal sparse approximations. The Meyer scaling function ϕ and wavelet function ψ are given by

$$\widehat{\phi}(\xi) = \begin{cases} 1 & \text{if } |\xi| \leqslant \frac{1}{3}, \\ \cos(\frac{\pi}{2}\nu(3|\xi| - 1)) & \text{if } \frac{1}{3} \leqslant |\xi| \leqslant \frac{2}{3}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\widehat{\psi}(\xi) = \begin{cases} -e^{-\pi i\xi} \sin[\frac{\pi}{2}\nu(3|\xi| - 1)] & \text{if } \frac{1}{3} \leqslant |\xi| \leqslant \frac{2}{3}, \\ -e^{-\pi i\xi} \cos[\frac{\pi}{2}\nu(\frac{3}{2}|\xi| - 1)] & \text{if } \frac{2}{3} \leqslant |\xi| \leqslant \frac{4}{3}, \\ 0 & \text{otherwise.} \end{cases}$$

Here, ν is a function satisfying $\nu(x) = 0$ for $x \leqslant 0$, $\nu(x) = 1$ for $x \geqslant 1$, and in addition, $\nu(x) + \nu(1 - x) = 1$ for $0 \leqslant x \leqslant 1$. For example, ν can be defined to be $\nu(x) = x^4(35 - 84x + 70x^2 - 20x^3)$ for $x \in [0, 1]$, which leads to C^3 functions $\widehat{\phi}$ and $\widehat{\psi}$.

The corresponding H is accordingly given by

$$H\big(e^{-2\pi i\xi}\big) = \begin{cases} 1 & |\xi| \leqslant \frac{1}{6}, \\ \cos(\frac{\pi}{2}\nu(6|\xi| - 1)) & \frac{1}{6} \leqslant |\xi| \leqslant \frac{1}{3}, \\ 0 & \frac{1}{3} \leqslant |\xi| \leqslant \frac{1}{2}. \end{cases}$$

We remark that $\xi \mapsto H(e^{-2\pi i\xi})$ is a 1-periodic function and the Meyer wavelet function ψ defined above satisfies $\widehat{\psi}(2\xi) = -e^{-2\pi i\xi}\overline{H(e^{-2\pi i(\xi+\frac{1}{2})})}\widehat{\phi}(\xi)$. Hence the high-pass filter G for ψ is indeed given by $G(z) = -z\overline{H(-z)}$ with $z = e^{-2\pi i\xi}$. For any $k \in \mathbb{N}$, there exists ν such that $\widehat{\phi}$ and $\widehat{\psi}$ are functions in $C^k(\mathbb{R})$. Moreover, ν can be constructed to be C^∞ so that both $\widehat{\phi}$ and $\widehat{\psi}$ are functions in $C^\infty(\mathbb{R})$ and their corresponding filters are functions in $C^\infty(\mathbb{T})$. For more details about Meyer wavelets, we refer to [4] or [22].

2.1.1. Subband decomposition for discrete data

The two-scale relation in combination with downsampling as a simple data reduction strategy is crucial for the efficient decomposition of data.

Let the group of integer translations $\{T_n\}_{n\in\mathbb{Z}}$ acting on $L^2(\mathbb{R})$ be defined by $T_n f(x) = f(x - n)$ for almost every $x \in \mathbb{R}$. The translates of the scaling function ϕ define the core subspace V_0. By definition, each function $f \in V_0$ can be expressed as the series

$$f = \sum_{n\in\mathbb{Z}} c_n T_n \phi$$

with a square summable sequence $\{c_n\}_{n\in\mathbb{Z}}$. This enables us to associate with f the values of the almost everywhere converging series

$$Zf(z) = \sum_{n=-\infty}^{\infty} c_n z^n, \quad z \in \mathbb{T}.$$

We next formalize the decomposition of a function $f \in V_0 = V_{-1} \oplus W_{-1}$ in terms of the Z-transform, where the coarser approximation space V_1 is obtained by the inverse of the usual dyadic dilation D, $V_{-1} = D^{-1}(V_0)$.

Letting now $H : \mathbb{T} \to \mathbb{C}$ be the low-pass filter of a localized multiresolution analysis as specified above, the characterization of the subspace $V_{-1} \subset V_0$ can then be expressed with the help of the filter H. We recall that the orthogonal projection onto V_{-1} in V_0 is given by applying the adjoint filter, downsampling the sequence of coefficients $\{c_n\}_{n\in\mathbb{Z}}$, upsampling and then applying the filter. In terms of the Z-transform, the composition of downsampling and upsampling is a periodization operation. Thus, $f \in V_{-1}$ if and only if it is invariant under the orthogonal projection onto V_{-1},

$$f \in V_{-1} \quad \Longleftrightarrow \quad Zf(z) = H(z)\big(\overline{H(z)}Zf(z) + \overline{H(-z)}Zf(-z)\big) \quad \text{for a.e. } z \in \mathbb{T}.$$

This fact enables us to state a unified characterization of V_{-1} and of $W_{-1} = V_0 \ominus V_{-1}$.

Proposition 2.1. *Let $P_{V_{-1}}$ and $P_{W_{-1}}$ denote the orthogonal projection of V_0 onto V_{-1} and W_{-1}, respectively. Further, letting H be defined as above, define H_+ to be the multiplication operator given by $H_+F(z) = H(z)F(z)$, H_- given by $H_-F(z) = H(-z)F(z)$, and R_2 the reflection operator satisfying $R_2 Zf(z) = Zf(-z)$. Then, we have*

$$ZP_{V_{-1}}f = H_+(I + R_2)\overline{H_+}Zf \quad \textit{and} \quad ZP_{W_{-1}}f = \overline{H_-}(I - R_2)H_-Zf.$$

Proof. By definition, the projection onto V_{-1} satisfies

$$ZP_{V_{-1}}f(z) = H(z)\big(\overline{H(z)}Zf(z) + \overline{H(-z)}Zf(-z)\big) = H_+(I+R_2)\overline{H_+}Zf(z).$$

Similarly, the projection onto W_{-1} is

$$\begin{aligned} ZP_{W_{-1}}f(z) &= G(z)\big(\overline{G(z)}Zf(z) + \overline{G(-z)}Zf(-z)\big) \\ &= -z\overline{H(-z)}\big(-\bar{z}H(-z)Zf(z) + \bar{z}H(z)Zf(-z)\big) \\ &= \overline{H(-z)}\big(H(-z)Zf(z) - H(z)Zf(-z)\big) \\ &= \overline{H_-}(I-R_2)H_-Zf(z). \end{aligned}$$

The proposition is proved. □

The relevance of these identities lies in the fact that $(I+R_2)\overline{H_+}Zf$ is an even function whereas $(I-R_2)H_-Zf$ is odd. Hence knowing every other coefficient in the series expansion is sufficient to determine the projection onto the corresponding subband. Thus, in this case downsampling reduces the data without loss of information.

2.1.2. M-band wavelets

If instead of a dilation factor of 2 in the two-scale relation, a factor of M is used, $M-1$ wavelets are necessary to complement the translates of ϕ to an orthonormal basis of the next higher resolution level. In this situation, it is an $M \times M$ matrix which has to satisfy the orthogonality identity. Generalizing the consideration in the previous subsection, let $\omega = e^{-2\pi i/M}$ and $R_M Zf(z) = Zf(\omega z)$ and the scaling mask H_0 for ϕ satisfy $\sum_{j=0}^{M-1}|H_0(\omega^j z)|^2 = 1$. We then define the orthogonal projection onto V_{-1} in terms of the transform

$$ZP_{V_{-1}}f = H_0\left(\sum_{j=0}^{M-1} R_M^j\right)\overline{H_0}Zf.$$

For a proof that $P_{V_{-1}}$ is indeed a projection, see the more general statement in the next theorem.

We complement the filter H_0 by finding H_n such that $(H_n(\omega^\ell z))_{n,\ell=0}^{M-1}$ is unitary for almost every $z \in \mathbb{T}$. Once the wavelet masks H_ℓ, $\ell = 1,\dots,M-1$ are constructed by matrix extension, the wavelet functions ψ_ℓ, $\ell = 1,\dots,M-1$ are given by $\widehat{\psi_\ell}(M\xi) = H_\ell(e^{-2\pi i\xi})\widehat{\phi}(\xi)$, $\xi \in \mathbb{R}$, $\ell = 1,\dots,M-1$. It is well-known that then $\{\psi_\ell\colon\ \ell = 1,\dots,M-1\}$ generates an orthonormal wavelet basis for $L^2(\mathbb{R})$.

One goal in M-band wavelet design is to choose H_0 and then to complete the matrix so that the filters H_n impart desirable properties on the associated scaling function and wavelets. In fact, one can construct orthonormal scaling functions for any dilation factor $M \geqslant 2$ and the matrix extension technique applies for any dilation factor $M \geqslant 2$. When $M > 2$, the orthonormal bases can be built to be with symmetry [14–16].

In the same terminology as Proposition 2.1, we now have the following result that characterizes the orthogonal projections belonging to M-band wavelets.

Theorem 2.2. *Let $\{H_n\}_{n=0}^{M-1}$ be such that $(H_n(\omega^\ell z))_{n,\ell=0}^{M-1}$ is unitary for almost every $z \in \mathbb{T}$, and let $\{P_{W_{-1,\ell}}\}_{\ell=0}^{M-1}$ be the operators defined by*

$$ZP_{W_{-1,\ell}}f := H_\ell\left(\sum_{j=0}^{M-1} R_M^j\right)\overline{H_\ell}Zf.$$

Then $\{P_{W_{-1,\ell}}\}_{\ell=0}^{M-1}$ are mutually orthogonal projections (note that $P_{W_{-1,0}} = P_{V_{-1}}$).

Proof. We first observe that by the assumed unitarity, every row is normalized, and each pair of rows is mutually orthogonal, i.e.,

$$\sum_{\ell=0}^{M-1} \overline{H_n(\omega^\ell z)} H_m(\omega^\ell z) = \delta_{n,m},$$

where $\delta_{n,m} = 1$ if $n = m$ and $\delta_{n,m} = 0$ otherwise.

Next, we show that each $P_{W_{-1,\ell}}$ is an orthogonal projection. To begin with, we see that $P_{W_{-1,\ell}}$ is Hermitian because the sum $\sum_{j=0}^{M-1} R_M^j$ is and this property is retained when it is conjugated by the multiplication operator H_ℓ. The fact that each $P_{W_{-1,\ell}}$ is idempotent and that the projections are mutually orthogonal is due to the commutation relation

$$R_M \overline{H_\ell} = \overline{H_\ell(\omega \cdot)} R_M$$

and because of the orthogonality of the rows in $(H_m(\omega^\ell z))_{m,\ell=0}^{M-1}$. Let $\ell, k \in \{0, 1, \dots, M-1\}$, we then have for $f \in L^2(\mathbb{R})$ and almost every $z \in \mathbb{T}$,

$$\begin{aligned}
ZP_{W_{-1,\ell}} P_{W_{-1,k}} f(z) &= H_\ell \sum_{m=0}^{M-1} R_M^m \overline{H_\ell} H_k \sum_{n=0}^{M-1} R_M^n \overline{H_k} Zf(z) \\
&= H_\ell \sum_{m=0}^{M-1} \overline{H_\ell(\omega^m z)} H_k(\omega^m z) R_M^m \sum_{n=0}^{M-1} R_M^n \overline{H_k} Zf(z) \\
&= H_\ell \sum_{m=0}^{M-1} \overline{H_\ell(\omega^m z)} H_k(\omega^m z) \sum_{n=0}^{M-1} R_M^n \overline{H_k} Zf(z) \\
&= \delta_{\ell,k} H_\ell \sum_{n=0}^{M-1} R_M^n \overline{H_k} Zf(z) = \delta_{\ell,k} ZP_{W_{-1,\ell}} f(z).
\end{aligned}$$

This shows that each $P_{W_{-1,\ell}}$ is a Hermitian idempotent and that the ranges of any pair $P_{W_{-1,\ell}}, P_{W_{-1,k}}$ with $k \neq l$ are mutually orthogonal, as claimed. □

2.2. From group-based to cone-adapted shearlets

In contrast to wavelets, shearlet systems are based on three operations: scaling, translation, and shearing; the last one to change the orientation of those anisotropic functions. Letting the (parabolic) scaling matrix A_j be defined by

$$A_j = \begin{pmatrix} 4^j & 0 \\ 0 & 2^j \end{pmatrix}, \quad j \in \mathbb{Z},$$

and the shearing matrix S_k be

$$S_k = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix}, \quad k \in \mathbb{Z}.$$

Then, for some generator $\psi \in L^2(\mathbb{R}^2)$, the *group-based shearlet system* is defined by

$$\{2^{\frac{3j}{2}} \psi(S_k A_j \cdot -m) \colon j, k \in \mathbb{Z},\ m \in \mathbb{Z}^2\}.$$

Despite the nice mathematical properties – this system can be regarded as arising from a representation of a locally compact group, the shearlet group – group-based shearlet systems suffer from the fact that they are biased towards one axis which prevents a uniform treatment of directions. Cone-adapted shearlet systems circumvent this problem, by utilizing a particular splitting of the frequency domain into a vertical and horizontal part. For this, we set $A_j^h := A_j$, $S_k^h := S_k$,

$$A_j^v = \begin{pmatrix} 2^j & 0 \\ 0 & 4^j \end{pmatrix}, \quad \text{and} \quad S_k^v = \begin{pmatrix} 1 & 0 \\ -k & 1 \end{pmatrix}, \quad j,k \in \mathbb{Z}.$$

Given a scaling function $\phi \in L^2(\mathbb{R}^2)$ and some $\psi \in L^2(\mathbb{R}^2)$, the *cone-adapted shearlet system* is defined by

$$\begin{aligned} \big\{\phi(\cdot - m)\colon\ m \in \mathbb{Z}^2\big\} &\cup \big\{2^{3j/2}\psi\big(S_k^h A_j^h \cdot - m\big)\colon\ j \geqslant 0,\ |k| \leqslant 2^j,\ m \in \mathbb{Z}^2\big\} \\ &\cup \big\{2^{3j/2}\tilde{\psi}\big(S_k^v A_j^v \cdot - m\big)\colon\ j \geqslant 0,\ |k| \leqslant 2^j,\ m \in \mathbb{Z}^2\big\}, \end{aligned}$$

where $\tilde{\psi}(x_1, x_2) = \psi(x_2, x_1)$. For more details on shearlets, we refer to [19].

Gabor shearlets will also be constructed first as group based systems, and then in a cone-adapted version. However, in contrast to other constructions, we aim at low redundancy in the group-based system and avoid increasing it in the cone adaptation.

2.3. Gabor frames

Like the previous systems, Gabor systems are based on translation and modulation. As usual, we denote the modulations on $L^2(\mathbb{R})$ by $M_m f(\xi) = e^{2\pi i m \xi} f(\xi)$.

By definition of tightness, a square-integrable function $w : \mathbb{R} \to \mathbb{C}$ is the window of a b^{-1}-tight Gabor frame $\{M_{mb} T_n w\colon\ m, n \in \mathbb{Z}\}$, if it is unit norm and for all $f \in L^2(\mathbb{R})$,

$$\|f\|^2 = b \sum_{m,n \in \mathbb{Z}} \big|\langle f, M_{mb} T_n w\rangle\big|^2.$$

For more details on Gabor systems, we refer the reader to [7].

Various ways to construct such a window function w are known. We recall a construction of a b^{-1}-tight Gabor frame with $b^{-1} > 1$ arbitrarily close to 1 [5].

Example 2.1. Let ν be in $C^\infty(\mathbb{R})$ and $\nu(x) = 0$ for $x \leqslant 0$, $\nu(x) = 1$ for $x \geqslant 1$ and $\nu(1-x) + \nu(x) = 1$. Let $w(x) := (\nu((1/2 + \varepsilon - |x|)/2\varepsilon))^{1/2}$, $x \in \mathbb{R}$. Then, it is easy to show that w is a smooth function with support belonging to $[-1/2 - \varepsilon, 1/2 + \varepsilon]$ for any $0 < \varepsilon < 1/2$, $\|w\|^2 = 1$, and $\sum_n |T_n w|^2 = 1$. Consequently, if $b = (1 + 2\epsilon)^{-1}$, then $\{M_{mb} T_n w\colon\ m, n \in \mathbb{Z}\}$ defines a b^{-1}-tight Gabor frame.

2.4. Redundancy

Since we cannot achieve (P1), but would like to approximate this property, besides the classical frame definition, we also require a notion of redundancy. The first more refined definition of redundancy besides the classical "number of elements divided by the dimension" definition was introduced in [1]. The extension of this definition to the infinitely dimensional case can be found in [2]. Since the work [2] is not intended for publication, we make this subsection self-contained.

We start by recalling a redundancy function, which provides a means to measure the concentration of the frame close to one vector. If $\{\varphi_i\}_{i \in I}$ is a frame for a real or complex Hilbert space $\mathcal{H}$ without any zero vectors, and let $\mathbb{S} = \{x \in \mathcal{H}\colon \|x\| = 1\}$, then for each $x \in \mathbb{S}$, the associated *redundancy function* $\mathcal{R} : \mathbb{S} \to \mathbb{R}^+ \cup \{\infty\}$ is defined by

$$\mathcal{R}(x) = \sum_{i \in I} \|\varphi_i\|^{-2} \big|\langle x, \varphi_i\rangle\big|^2.$$

Taking the supremum or the infimum over x in this definition gives rise to the so-called upper and lower redundancy, which is in fact the upper and lower frame bound of the associated normalized frame,

$$\mathcal{R}^+ = \sup_{x \in \mathbb{S}} \mathcal{R}(x) \quad \text{and} \quad \mathcal{R}^- = \inf_{x \in \mathbb{S}} \mathcal{R}(x).$$

For those values, it was proven in [1] that in the finite-dimensional situation, the upper redundancy provides a means to measure the minimal number of linearly independent sets, and the lower redundancy is related to the maximal number of spanning sets, thereby linking analytic to algebraic properties.

It is immediate to see that an orthonormal basis satisfies $\mathcal{R}^- = \mathcal{R}^+ = 1$, and a unit norm A-tight frame $\mathcal{R}^- = \mathcal{R}^+ = A$. This motivates the following definition.

Definition 2.2. A frame $\{\varphi_i\}_{i=\in I}$ for a real or complex Hilbert space has a *uniform redundancy*, if $\mathcal{R}^- = \mathcal{R}^+$, and if it is unit norm and A-tight, then we say that it has *redundancy* A.

In the sequel, we will use the redundancy to determine to which extent (P1) is satisfied.

3. Group-based Gabor shearlets

Let us start with an informal description of the construction of Gabor shearlets in a special case with the goal to first provide some intuition for the reader.

Generally speaking, the shearlet construction in this paper is a Meyer-type modification of a multiresolution analysis based on the Shannon shearlet scaling function $\widehat{\Phi}_{0,0,0} = \chi_K$, where $K = \{\xi \in \mathbb{R}^2\colon |\xi_1| \leqslant 1 \text{ and } |\xi_2/\xi_1| \leqslant 1/2\}$. For an illustration, we refer to Fig. 1.

It is straightforward to verify that chirp modulations

$$\widehat{\Phi}_{0,0,m}(\xi) = \chi_K(\xi) e^{2\pi i m_2 \xi_2/\xi_1} e^{\pi i m_1 \xi_1^3/|\xi_1|}, \quad m = (m_1, m_2) \in \mathbb{Z}^2,$$

define an orthonormal system $\{\Phi_{0,0,m}\colon m \in \mathbb{Z}^2\}$, while the usual modulations

$$\widehat{\Upsilon}_{0,0,m}(\xi) = \chi_K(\xi) e^{2\pi i m_2 \xi_2} e^{\pi i m_1 \xi_1}$$

give a 2-tight frame $\{\Upsilon_{0,0,m}\colon m \in \mathbb{Z}^2\}$ for its span. The same is true when the modulations are augmented with shears, $\widehat{\Phi}_{0,k,m}(\xi) = \widehat{\Phi}_{0,0,m}(\xi_1, \xi_2 - k\xi_1)$ and likewise for $\widehat{\Upsilon}_{0,k,m}$, in order to form the orthonormal or tight systems $\{\Phi_{0,k,m}\colon k \in \mathbb{Z},\ m \in \mathbb{Z}^2\}$ or $\{\Upsilon_{0,k,m}\colon k \in \mathbb{Z},\ m \in \mathbb{Z}^2\}$, respectively. Because both systems are unit-norm, the tightness constant is a good measure for redundancy as detailed in Section 2.4, indicating that chirp modulations are preferable from this point of view. Incorporating parabolic scaling preserves those properties.

In a second step (Section 4) the strategy of shearlets is followed to derive a cone-adapted version, which provides the property of a uniform treatment of directions necessary for optimal sparse approximation results. Apart from directional selectivity, good decay properties form a further necessary ingredient for optimal sparsity. We show that a combination of Gabor frames, Meyer wavelets and a change of coordinates provide smooth alternatives for the characteristic function, yet with still near-orthonormal shearlet systems that are similar to the Shannon shearlets we described.

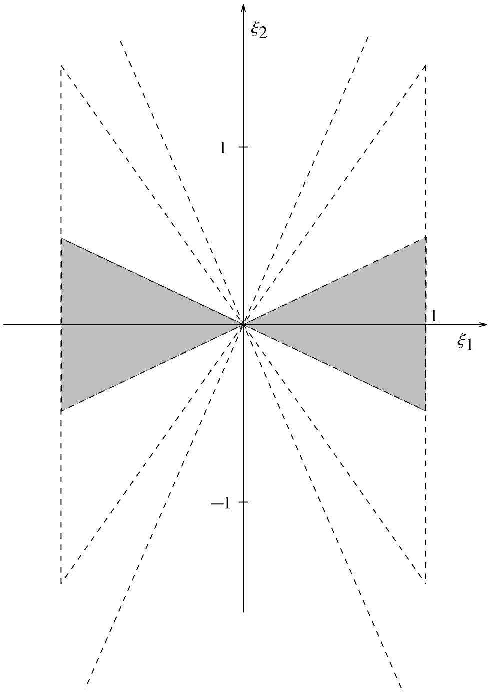

Fig. 1. Support of the scaling function belonging to group-based Shannon shearlets (as well as the group-based Gabor shearlets) in the frequency domain. Additional lines indicate the boundaries of the support for sheared scaling functions.

3.1. Construction using chirp modulations

To begin the shearlet construction, we examine an alternative group of translations acting as chirp modulations in the frequency domain. These modulations do not correspond to the usual Euclidean translations, but for implementations in the frequency domain this is not essential. In the following, we use the notation $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$.

Definition 3.1. Let $\gamma(\xi) := (\gamma_1(\xi), \gamma_2(\xi))$ with $\gamma_1(\xi) := \frac{1}{2}\xi_1^3/|\xi_1| = \frac{1}{2}\operatorname{sgn}(\xi_1)\xi_1^2$ and $\gamma_2(\xi) := \frac{\xi_2}{\xi_1}$ for $\xi = (\xi_1, \xi_2) \in \mathbb{R}^* \times \mathbb{R}$. We define the two-dimensional *chirp-modulations* $\{X_\beta\colon\ \beta \in \mathbb{R}^2\}$ by

$$X_\beta \widehat{f}(\xi) = e^{2\pi i \beta_1 \gamma_1(\xi)} e^{2\pi i \beta_2 \gamma_2(\xi)} \widehat{f}(\xi), \quad \xi \in \mathbb{R}^* \times \mathbb{R}.$$

We emphasize that the set with $\xi_1 = 0$ is excluded from the domain, which does not cause problems since it has measure zero.

Next, notice that the point transformation γ has a Jacobian of magnitude one and is a bijection on $\mathbb{R}^* \times \mathbb{R}$. Therefore, it defines a unitary operator Γ according to

$$\Gamma \widehat{f}(\xi) = \widehat{f}\big(\gamma(\xi)\big), \quad \xi \in \mathbb{R}^* \times \mathbb{R}.$$

As discussed in Section 2.2, the shear operator is a further ingredient of shearlet systems. By abuse of notation, for any $s \in \mathbb{R}$, we will also regard S_s as an operator, that is

$$S_s \widehat{f}(\xi_1, \xi_2) = \widehat{f}(\xi_1, \xi_2 - s\xi_1).$$

The benefit of choosing the chirp-modulations is that shearing and modulation satisfy the well-known Weyl–Heisenberg commutation relations. The proof of the following result is a straightforward calculation, hence we omit it.

Proposition 3.1. *For $s \in \mathbb{R}$ and $\beta \in \mathbb{R}^2$,*

$$S_s X_\beta = e^{-2\pi i \beta_2 s} X_\beta S_s.$$

The last ingredient is a scaling operator which gives parabolic scaling. Again abusing notation, we write the dilation operator with A_j. For $j \in \mathbb{Z}$, we let A_j be the dilation operator acting on $f \in L^2(\mathbb{R}^2)$ by

$$A_j \widehat{f}(\xi_1, \xi_2) = 2^{-3j/2} \widehat{f}\big(2^{-2j}\xi_1, 2^{-j}\xi_2\big)$$

for almost very $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$.

Now we are ready to define group-based Gabor shearlets by the generating functions to which those three operators are then applied. For this, let ϕ be an orthogonal scaling function of a 16-band multiresolution analysis in $L^2(\mathbb{R})$, with associated orthonormal wavelets $\{\psi_\ell\}_{\ell=1}^{15}$, and let w be the unit norm window function of a b^{-1}-tight Gabor frame $\{M_{m_2 b} T_k w\colon\ m_2, k \in \mathbb{Z}\}$ for $L^2(\mathbb{R})$. Then we define the generators

$$\widehat{\Phi_{0,0,0}} := \Gamma \widehat{\phi} \otimes w \quad \text{and} \quad \widehat{\Psi^\ell_{0,0,0}} := \Gamma \widehat{\psi}_\ell \otimes w, \quad \ell = 1, \ldots, 15$$

in $L^2(\mathbb{R}^2) = L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$, based on which we now define group-based Gabor shearlets.

Definition 3.2. Let $\Phi_{0,0,0}$ and $\Psi^\ell_{0,0,0}$, $\ell = 1, \ldots, 15$, and w be defined as above. Let $j_0 \in \mathbb{Z}$. Then the *group-based Gabor shearlet system* is defined by

$$\begin{aligned}
&\mathcal{GGS}_{j_0}\big(\phi, \{\psi_\ell\}_{\ell=1}^{15}; w\big) \\
&\quad := \big\{\Phi_{j_0,k,m}\colon\ k \in \mathbb{Z},\ m \in \mathbb{Z}^2\big\} \cup \big\{\Psi^\ell_{j,k,m}\colon\ j, k \in \mathbb{Z},\ j \geqslant j_0,\ m \in \mathbb{Z}^2,\ \ell = 1, \ldots, 15\big\} \subseteq L^2\big(\mathbb{R}^2\big),
\end{aligned}$$

where

$$\begin{aligned}
\widehat{\Phi_{j,k,m}}(\xi) &= A_j X_{(m_1, m_2 b)} S_k \widehat{\Phi}_{0,0,0} \\
&= 2^{-3j/2} \widehat{\phi}\big(2^{-4j}\gamma_1(\xi)\big) w\big(2^j \gamma_2(\xi) - k\big) e^{2\pi i m_1 2^{-4j}\gamma_1(\xi)} e^{2\pi i m_2 b 2^j \gamma_2(\xi)},
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\Psi^\ell_{j,k,m}}(\xi) &= A_j X_{(m_1, m_2 b)} S_k \widehat{\Psi}^\ell_{0,0,0} \\
&= 2^{-3j/2} \widehat{\psi}_\ell\big(2^{-4j}\gamma_1(\xi)\big) w\big(2^j \gamma_2(\xi) - k\big) e^{2\pi i m_1 2^{-4j}\gamma_1(\xi)} e^{2\pi i m_2 b 2^j \gamma_2(\xi)}.
\end{aligned}$$

The particular choice of dilation factors in the first and second coordinate comes from the need for parabolic scaling and integer dilations. The motivation is that the regularity of the singularity in the cartoon-like model is C^2, and if the generator satisfies *width* $=$ *length*^{2} one can basically linearize the curve inside the support with controllable error by the Taylor expansion. Since we utilize a different group operation, it is not immediately clear which scaling leads to the size constraints *width* $=$ *length*^{2}. An integer value of j requires $4j = j^2$, so $j = 4$. Then one considers the intertwining relationship between the dilation operator A_4 and the standard one-dimensional dyadic dilation D to deduce $A_4 \Gamma = \Gamma D^{-16} \otimes D^4$, which explains the choice of $M = 16$ bands.

3.2. MRA structure

One crucial question is whether the just introduced system is associated with an MRA structure. As a first step, we define associated scaling and wavelet spaces.

Definition 3.3. Let $\varPhi_{j,k,m}$ and $\varPsi^{\ell}_{j,k,m}$, $j,k \in \mathbb{Z}$, $m \in \mathbb{Z}^2$, $\ell = 1,\dots,15$ be defined as in Definition 3.2. For each $j \in \mathbb{Z}$, the *scaling space* V_j is the closed subspace

$$V_j = \overline{\operatorname{span}}\{\varPhi_{j,k,m}\colon\ k \in \mathbb{Z},\ m \in \mathbb{Z}^2\} \subseteq L^2(\mathbb{R}^2),$$

and the associated *wavelet space* W_j is defined by

$$W_j = \overline{\operatorname{span}}\{\varPsi^{\ell}_{j,k,m}\colon\ k \in \mathbb{Z},\ m \in \mathbb{Z}^2,\ \ell = 1,\dots,15\}.$$

Next, we establish that the group-based Gabor shearlet system is indeed associated with an MRA structure, and analyze how close it is to being an orthonormal basis.

Theorem 3.1. *Let $\varPhi_{j,k,m}$ and $\varPsi^{\ell}_{j,k,m}$, $j,k \in \mathbb{Z}$, $m \in \mathbb{Z}^2$, $\ell = 1,\dots,15$ be defined as in Definition 3.2 and let $\{V_j\}_{j\in\mathbb{Z}}$ and $\{W_j\}_{j\in\mathbb{Z}}$ be the associated scaling and wavelet spaces as defined in Definition 3.3. Then, for each $j \in \mathbb{Z}$, the family $\{\varPhi_{j,k,m}\colon\ k \in \mathbb{Z},\ m \in \mathbb{Z}^2\}$ is a unit norm b^{-1}-tight frame for V_j, and $\{\varPsi^{\ell}_{j,k,m}\colon\ k \in \mathbb{Z},\ m \in \mathbb{Z}^2,\ \ell = 1,\dots,15\}$ forms a unit-norm b^{-1}-tight frame for W_j.*

Proof. We first verify that the scaling function generates a b^{-1}-tight frame for a closed subspace of $L^2(\mathbb{R}^2)$. By Proposition 3.1, the operator $\varGamma$ intertwines shears and translations in the second component,

$$S_k \varGamma \widehat{f}(\xi) = \varGamma \widehat{f}(\xi_1, \xi_2 - k\xi_1) = \widehat{f}\big(\gamma_1(\xi), \gamma_2(\xi) - k\big).$$

Moreover, it intertwines chirp modulations with standard modulations. The overall dilation is irrelevant because A_j is unitary, so we can set $j = 0$ for simplicity. Therefore, it is enough to prove that $\{M_{m_1}\widehat{\phi} \otimes M_{m_2 b} T_k w\}$ defines a b^{-1}-tight frame for $\varGamma^{-1}(\widehat{V}_0)$. This follows from the fact that w is the unit norm window function of a b^{-1}-tight Gabor frame and from ϕ being an orthonormal scaling function of an MRA. A similar argument shows that $\{M_{m_1}\widehat{\psi} \otimes M_{m_2 b} T_k w\}$ defines a b^{-1}-tight frame for $\varGamma^{-1}(\widehat{W}_0)$, because ψ is an orthonormal MRA wavelet. $\square$

Theorem 3.2. *The scaling and wavelet subspaces V_0 and W_0 as defined in Definition 3.3 satisfy the two-scale relation*

$$V_0 \oplus W_0 = \check{A}_4(V_0)$$

where $\check{A}_4$ is the inverse Fourier transform of the dilation operator, so for $f \in L^2(\mathbb{R}^2)$, $(\check{A}_4 f)\hat{} = A_4 \hat{f}$.

Proof. We note that the functions

$$\widehat{\phi} \otimes w \quad \text{and} \quad \widehat{\psi}_\ell \otimes w, \quad \ell = 1,\dots,15$$

are orthogonal by assumption, and the orthogonality remains under the usual modulations in the first component. On the other hand, the window function in the second component forms a tight Gabor frame under translations and modulations, so each of the tensor products generates a tight frame for its span.

Since the subspaces $\widehat{\phi} \otimes L^2(\mathbb{R})$, $\widehat{\psi}_\ell \otimes L^2(\mathbb{R})$, $\ell = 1,\dots,15$ are mutually orthogonal, by the unitarity of $\varGamma$ the same is true for their images $\varGamma(\widehat{\phi} \otimes L^2(\mathbb{R}))$ and $\varGamma(\widehat{\psi}_\ell \otimes L^2(\mathbb{R}))$. Finally, the functions $\{\phi, \psi_\ell\}$ satisfy a two-scale relation of an MRA with dilation factor $M = 16$ in $L^2(\mathbb{R})$, so the claim follows from the intertwining relationship $A_4 \varGamma = \varGamma D^{-16} \otimes D^4$. $\square$

Since implementations only concern a finite number of scales, the following result becomes important. It is an easy consequence of Theorem 3.1.

Corollary 3.1. *The group-based Gabor shearlet system $\mathcal{GGS}_{j_0}(\phi, \{\psi_\ell\}_{\ell=1}^{15}; w)$ as defined in Definition 3.2 for any $j_0 \in \mathbb{Z}$, or the system $\{\Psi_{j,k,m}^\ell\colon j,k \in \mathbb{Z},\ m \in \mathbb{Z}^2,\ \ell = 1,\dots,15\}$ forms a unit-norm b^{-1}-tight frame for $L^2(\mathbb{R}^2)$, and consequently it has uniform redundancy $\mathcal{R}^- = \mathcal{R}^+ = b^{-1}$.*

4. Cone-adapted Gabor shearlets

The construction of cone-adapted Gabor shearlets is based on complementing a core subspace V_0 which has the usual MRA properties for $L^2(\mathbb{R}^2)$ under scaling with a dilation factor of 16. The isometric embedding of V_0 in V_1 proceeds in 3 steps:

1. V_1 is split into a direct sum of two coarse-directional subspaces, V_1^h and V_1^v, corresponding to horizontally and vertically aligned details, respectively.
2. Each of these two coarse-directional subspaces is split into a direct sum of high and low pass components. The low-pass subspaces V_0^h and V_0^v combine to $V_0 = V_0^h \oplus V_0^v$.
3. The high pass components are further split into subspaces with a finer directional resolution obtained from shearing.

The first step in the process of constructing the cone-adapted shearlets is a splitting between features that are mostly aligned in the horizontal or in the vertical direction. The shearlets then refine this coarse splitting.

4.1. Cone adaptation

In addition to filters which restrict to cones in the frequency domain, we introduce projections based on quarter rotations for the splitting of horizontal and vertical features. This enables us to define two mutually orthogonal closed subspaces containing functions with support near the usual cones for horizontal and vertical components. As in the case of wavelets, the main goal of this construction is that the smoothness of a function in the frequency domain is not substantially degraded by the projection onto the subspaces.

Again, we use standard filters from wavelets in our construction. For this, we define a version of the Cayley transform $\zeta(\xi) = \frac{1+i\xi}{1-i\xi}$, which maps $\xi \in \mathbb{R}$ to the unit circle $\mathbb{T} = \{z \in \mathbb{C}\colon |z| = 1\}$. The inverse map is defined on $\mathbb{T} \setminus \{-1\}$, $\zeta^{-1}(z) = i\frac{1-z}{1+z}$. We use the map ζ to lift polynomial filters on $\mathbb{T}$ to rational filters on $\mathbb{R}$.

Lemma 4.1. *Let $H : \mathbb{T} \to \mathbb{C}$ satisfy $|H(z)|^2 + |H(-z)|^2 = 1$ for all $z \in \mathbb{T}$, then $\tilde{H}(\xi) := H(\zeta(\xi))$ is a function on $\mathbb{R}$ which satisfies*

$$\left|\tilde{H}(\xi)\right|^2 + \left|\tilde{H}(-1/\xi)\right|^2 = 1.$$

Proof. The Cayley transform intertwines the reflection $\xi \mapsto -1/\xi$ on $\mathbb{R}$ with the reflection about the origin, because

$$\zeta(-1/\xi) = \frac{1 - i/\xi}{1 + i/\xi} = \frac{1 + i\xi}{-1 + i\xi} = -\zeta(\xi).$$

Thus, the property of $\tilde{H}$ is a direct consequence of this coordinate transformation. $\square$

We observe that if $H(z)$ has $N-1$ vanishing derivatives at $z = -1$, $H(-1) = H'(-1) = \cdots = H^{(N-1)}(-1) = 0$, then $\tilde{H}(\xi)$ decays as ξ^{-N} at infinity.

Definition 4.1. Let $H : \mathbb{T} \to \mathbb{C}$ satisfy the Smith–Barnwell condition $|H(z)|^2 + |H(-z)|^2 = 1$ for all $z \in \mathbb{T}$. Its associated filter operators H_+, $\overline{H_+}$, H_- and $\overline{H_-}$, are defined to be the multiplicative operators with the Fourier transform of any $f \in L^2(\mathbb{R}^2)$ in the frequency domain according to $H_+\widehat{f}(\xi) = H(\zeta(\xi_2/\xi_1))\widehat{f}(\xi)$ and $H_-\widehat{f}(\xi) = H(-\zeta(\xi_2/\xi_1))\widehat{f}(\xi)$, the overbar denoting multiplication with the complex conjugate. We denote R to be the rotation operator on $L^2(\mathbb{R}^2)$ given by $R\widehat{f}(\xi_1,\xi_2) = \widehat{f}(\xi_2,-\xi_1)$.

This allows us to introduce a pair of complementary orthogonal projections, which split the group based Gabor shearlets into a vertical and a horizontal part to balance the treatment of directions. The design of these projections is inspired by the description of smooth projections in [24].

We start the construction with isometries associated with the vertical and horizontal cone, which we denote by $\mathcal{C}_v$ and $\mathcal{C}_h$, respectively. By the set inclusion, $L^2(\mathcal{C}_v)$ and $L^2(\mathcal{C}_h)$ naturally embed isometrically in $L^2(\mathbb{R}^2)$. We denote these embeddings by ι_v: $\iota_v f(\xi) = f(\xi)$ if $\xi \in C_v$ and $\iota_v f(\xi) = 0$ otherwise and similarly for ι_h. We wish to find isometries that do not create discontinuities.

Theorem 4.1. *Let $H : \mathbb{T} \to \mathbb{C}$ satisfy $|H(z)|^2 + |H(-z)|^2 = 1$ for all $z \in \mathbb{T}$, and let H_+, $\overline{H_+}$, H_- and $\overline{H_-}$ be defined as in Definition 4.1. Let $\mathcal{C}_v = \{x \in \mathbb{R}^2\colon\ |x_2| \geqslant |x_1|\}$ and $\mathcal{C}_h = \mathbb{R}^2 \setminus \mathcal{C}_v$, then the map $\varXi_v : L^2(\mathcal{C}_v) \to L^2(\mathbb{R}^2)$ given by*

$$\varXi_v f = \overline{H_-}\left(I - \frac{1+i}{2}R - \frac{1-i}{2}R^3\right)\iota_v f$$

is an isometry, and so is the map $\varXi_h : L^2(\mathcal{C}_h) \to L^2(\mathbb{R}^2)$,

$$\varXi_h f = H_+\left(I + \frac{1+i}{2}R + \frac{1-i}{2}R^3\right)\iota_h f.$$

Moreover, the range of $\varXi_v$ is the orthogonal complement of the range of $\varXi_h$ in $L^2(\mathbb{R}^2)$, and $P_v = \varXi_v\varXi_v^*$ and $P_h = \varXi_h\varXi_h^*$ are complementary orthogonal projections on $L^2(\mathbb{R}^2)$.

Proof. We begin by showing that $\varXi_v$ and $\varXi_h$ are isometries. The space $L^2(\mathcal{C}_v)$ splits into even and odd functions. After embedding in $L^2(\mathbb{R}^2)$ these functions then satisfy $R^2\iota_v f = \iota_v f$ or $R^2\iota_v f = -\iota_v f$, respectively.

By the definition of R, the operator $I - \frac{1+i}{2}R - \frac{1-i}{2}R^3$ maps even $\iota_v f$ to

$$\left(I - \frac{1+i}{2}R - \frac{1-i}{2}R^3\right)\iota_v f = \left(\frac{1}{2}I + \frac{1}{2}R^2 - \frac{1}{2}R - \frac{1}{2}R^3\right)\iota_v f,$$

which implies that it is an eigenvector of R, $R(\frac{1}{2}I + \frac{1}{2}R^2 - \frac{1}{2}R - \frac{1}{2}R^3)\iota_v f = -(\frac{1}{2}I + \frac{1}{2}R^2 - \frac{1}{2}R - \frac{1}{2}R^3)\iota_v f$ and for odd f

$$\left(I - \frac{1+i}{2}R - \frac{1-i}{2}R^3\right)\iota_v f = \left(\frac{1}{2}I - \frac{1}{2}R^2 - \frac{i}{2}R + \frac{i}{2}R^3\right)\iota_v f,$$

which gives $R(\frac{1}{2}I - \frac{1}{2}R^2 - \frac{i}{2}R + \frac{i}{2}R^3)\iota_v f = i(\frac{1}{2}I - \frac{1}{2}R^2 - \frac{i}{2}R + \frac{i}{2}R^3)\iota_v f$.

Similarly, the operator $(I + \frac{1+i}{2}R + \frac{1-i}{2}R^3)$ maps the even functions into functions that are invariant under R, whereas the odd functions give eigenvectors of R corresponding to eigenvalue $-i$. We verify that for even $\iota_h f$,

$$\left(I + \frac{1+i}{2}R + \frac{1-i}{2}R^3\right)\iota_h f = \left(\frac{1}{2}I + \frac{1}{2}R^2 + \frac{1}{2}R + \frac{1}{2}R^3\right)\iota_h f.$$

Hence we get the eigenvalue equation $R(\frac{1}{2}I+\frac{1}{2}R^2+\frac{1}{2}R+\frac{1}{2}R^3)\iota_h f = (\frac{1}{2}I+\frac{1}{2}R^2+\frac{1}{2}R+\frac{1}{2}R^3)\iota_h f$. Analogously, for odd f,

$$\left(I+\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)\iota_h f = \left(\frac{1}{2}I-\frac{1}{2}R^2+\frac{i}{2}R-\frac{i}{2}R^3\right)\iota_h f$$

which yields $R(\frac{1}{2}I-\frac{1}{2}R^2+\frac{i}{2}R-\frac{i}{2}R^3)\iota_h f = (-i)(\frac{1}{2}I-\frac{1}{2}R^2+\frac{i}{2}R-\frac{i}{2}R^3)\iota_h f$.

Since R is unitary, the eigenvector equations imply that the orthogonality between even and odd functions is preserved by the embedding followed by the symmetrization with $(I+\frac{1+i}{2}R+\frac{1-i}{2}R^3)$ or $(I-\frac{1+i}{2}R-\frac{1-i}{2}R^3)$. Thus, the identity

$$\left\|\left(I-\frac{1+i}{2}R-\frac{1-i}{2}R^3\right)\iota_v f\right\|^2_{L^2(\mathbb{R}^2)} = 2\|f\|^2_{L^2(\mathcal{C}_v)} \quad \text{for all } f \in L^2(\mathcal{C}_v)$$

can be verified by checking it separately for even and odd functions. Next, multiplying by $\overline{H_-}$ and using that $|H_-|^2+R^{-1}|H_-|^2R = |H_-|^2+|H_+|^2 = 1$ gives by the orthogonality of $R\iota_v f$ and $\iota_v f$ the isometry

$$\begin{aligned}
&\left\|\overline{H_-}\left(I-\frac{1+i}{2}R-\frac{1-i}{2}R^3\right)\iota_v f\right\|^2_{L^2(\mathbb{R}^2)} \\
&\quad = \|\overline{H_-}\iota_v f\|^2_{L^2(\mathbb{R}^2)} + \left\|H_-\left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)\iota_v f\right\|^2_{L^2(\mathbb{R}^2)} \\
&\quad = \|H_-\iota_v f\|^2_{L^2(\mathbb{R}^2)} + \left\|H_+\left(\frac{1+i}{2}+\frac{1-i}{2}R^2\right)\iota_v f\right\|^2_{L^2(\mathbb{R}^2)} \\
&\quad = \|H_-\iota_v f\|^2_{L^2(\mathbb{R}^2)} + \|H_+\iota_v f\|^2_{L^2(\mathbb{R}^2)} = \|f\|^2_{L^2(\mathcal{C}_v)}.
\end{aligned}$$

The same proof applies to Ξ_h.

To show that the ranges are orthogonal complements of each other, we first establish that the projections $P_v = \Xi_v\Xi_v^*$ and $P_h = \Xi_h\Xi_h^*$ have the more convenient expressions

$$P_h = H_+\left(I+\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)\overline{H_+} \quad \text{and} \quad P_v = \overline{H_-}\left(I-\frac{1+i}{2}R-\frac{1-i}{2}R^3\right)H_-.$$

To this end, we note that if M_v is the multiplication operator $M_v f(\xi) = \chi_{\mathcal{C}_v}(\xi)f(\xi)$ with $\chi_{\mathcal{C}_v}$ the characteristic function of $\mathcal{C}_v$, and similarly for M_h, $M_h f(\xi) = \chi_{\mathcal{C}_h}(\xi)f(\xi)$, then by definition

$$P_h = \Xi_h\Xi_h^* = H_+\left(I+\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)M_h\left(I+\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)\overline{H_+}.$$

We simplify this expression using that $M_hR = RM_v$, $M_v+M_h = I$ and $R^2M_h = M_hR^2$, which gives the identities

$$\begin{aligned}
&\left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)M_h + M_h\left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right) \\
&\quad = \left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)M_h + \left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)M_v = \frac{1+i}{2}R+\frac{1-i}{2}R^3
\end{aligned}$$

and

$$\left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)M_h\left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right) = \left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)^2M_v = M_v.$$

Inserting this in the expression for P_h results in

$$H_+\left(I+\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)M_h\left(I+\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)\overline{H_+}$$
$$=H_+\left(M_h+\left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)+M_v\right)\overline{H_+}=H_+\left(I+\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)\overline{H_+}.$$

The identities for P_v are completely analogous.

Finally, we show that the two orthogonal projections are complementary. To this end, we use

$$P_h=H_+\overline{H_+}+H_+\overline{H_-}\left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)$$

and

$$P_v=\overline{H_-}H_- - \overline{H_-}H_+\left(\frac{1+i}{2}R+\frac{1-i}{2}R^3\right)$$

which gives $P_h+P_v=I$ after elementary cancellations and $H_+\overline{H_+}+\overline{H_-}H_-=I$. Since P_h is by definition an orthogonal projection, $P_v=I-P_h$ is the complementary one. Thus, the ranges of Ξ_h and Ξ_v, or equivalently, the ranges of P_h and P_v, are orthogonal complements in $L^2(\mathbb{R}^2)$. □

For later use, we denote the range spaces of Ξ_h and Ξ_v by

$$L_h^2(\mathbb{R}^2)=\Xi_h\big(L^2(\mathcal{C}_h)\big)\quad\text{and}\quad L_v^2(\mathbb{R}^2)=\Xi_v\big(L^2(\mathcal{C}_v)\big),$$

which are orthogonal complements in $L^2(\mathbb{R}^2)$.

Under the isometries Ξ_v or Ξ_h, a unit norm tight frame for $L^2(\mathcal{C}_v)$ or $L^2(\mathcal{C}_h)$ is mapped to a unit norm tight frame for $L_h^2(\mathbb{R}^2)$ or $L_v^2(\mathbb{R}^2)$. This means, in principle we would only need to construct shearlets for the horizontal and vertical cones, not for all of $\mathbb{R}^2$. However, in order to achieve smoothness, we need to enlarge the cones slightly.

To prepare this construction, we first refer to a result of Søndergaard from [23].

Theorem 4.2. *(See [23].) Let $N_0,\tau\in\mathbb{N}$, $\tau<N_0$ and let $\alpha_0=2/N_0$, $\beta_0=\tau/2$. If w is a function in the Feichtinger algebra, and if $\{M_{m\beta_0}T_{k\alpha_0}w\}_{m,k\in\mathbb{Z}}$ is an $\frac{N_0}{\tau}$-tight Gabor frame for $L^2(\mathbb{R})$, then the periodization w°,*

$$w^\circ(\xi)=\sum_{n\in\mathbb{Z}}w(\xi-2n)\quad\textit{for a.e. }\xi\in\mathbb{R},$$

defines an $\frac{N_0}{\tau}$-tight Gabor frame $\{M_{m\beta_0}T_{k\alpha_0}w^\circ\colon\ 0\leqslant k\leqslant N_0-1,\ m\in\mathbb{Z}\}$ for $L^2([-1,1])$.

In order to achieve smoothness of the cone-adapted shearlets in the frequency domain, we need a slightly dilated Gabor frame. For $\epsilon>0$, we denote the dilated window $w_\epsilon^\circ(x)=\frac{1}{\sqrt{1+\epsilon}}w^\circ(x/(1+\epsilon))$.

Corollary 4.1. *Let $N_0,\tau\in\mathbb{N}$, $\tau<N_0$ and let $\alpha=2(1+\epsilon)/N_0$, $\beta=(1+\epsilon)^{-1}\tau/2$. If w satisfies the assumptions in Theorem 4.2 then the dilated window gives rise to the N_0/τ-tight Gabor frame $\{M_{m\beta}T_{k\alpha}w_\epsilon^\circ\}$ for $L^2([-1-\epsilon,1+\epsilon])$.*

Of particular interest to us is the next corollary, which we can draw from this result. We remark that tightness is preserved when periodizing the window, and if its support is sufficiently small then so is the norm.

Corollary 4.2. *The uniform redundancy $\mathcal{R}^- = \mathcal{R}^+ = \frac{N_0}{\tau}$ of the Gabor frame $\{M_{m\beta}T_{k\alpha}w_\epsilon^\circ\}$ defined in Corollary 4.1 can be chosen as close to one as desired by choosing $N_0, \tau \in \mathbb{N}$ sufficiently large.*

Before stating the definition of cone-adapted Gabor shearlets, we require the following additional ingredients. We consider the change of variables $(\xi_1, \xi_2) \mapsto \gamma^\iota(\xi) = (\gamma_1^\iota(\xi), \gamma_2^\iota(\xi))$, $\iota \in \{h, v\}$, defined by

$$\gamma_1^h(\xi) = \frac{1}{2}\operatorname{sgn}(\xi_1)\xi_1^2, \qquad \gamma_2^h(\xi) = \frac{\xi_2}{\xi_1} \quad \text{and} \quad \gamma_1^v(\xi) = \frac{1}{2}\operatorname{sgn}(\xi_2)\xi_2^2, \qquad \gamma_2^v(\xi) = \frac{\xi_1}{\xi_2}.$$

We let Γ_h and Γ_v denote the associated unitary operators, $\Gamma_h f(\xi) = f(\gamma^h(\xi))$ and $\Gamma_v f(\xi) = f(\gamma^v(\xi))$. For each orientation v or h, we define the appropriate dilation, shear, and modulation operators by

$$A_j^h \equiv A_j, \qquad X_m^h \equiv X_m, \quad \text{and} \quad S_k^h \equiv S_k,$$

and if $\widehat{f}(\xi_1, \xi_2) = \widehat{g}(\xi_2, \xi_1)$, then

$$A_j^v \widehat{f}(\xi_1, \xi_2) = A_j^h \widehat{g}(\xi_2, \xi_1), \qquad X_m^v \widehat{f}(\xi_1, \xi_2) = X_m^h \widehat{g}(\xi_2, \xi_1), \quad \text{and} \quad S_k^v \widehat{f}(\xi_1, \xi_2) = S_k^h \widehat{g}(\xi_2, \xi_1).$$

Definition 4.2. Let ϕ be an orthogonal scaling function of a 16-band multiresolution analysis in $L^2(\mathbb{R})$, with associated orthonormal wavelets $\{\psi_\ell\colon \ \ell = 1, \dots, 15\}$, and let $\epsilon > 0$, $N_0, \tau \in \mathbb{N}$ such that w is the unit norm window function of an $\frac{N_0}{\tau}$-tight Gabor frame $\{M_{m_2\beta}T_{k\alpha}w\colon \ m_2, k \in \mathbb{Z}\}$ for $L^2(\mathbb{R})$, with the periodization w_ϵ° as described in Corollary 4.1. For any $j_0 \in \mathbb{Z}$, the associated *cone-adapted Gabor shearlet system* is defined by

$$\begin{aligned}
&\mathcal{CGS}_{j_0}\big(\phi, \{\psi_\ell\}_{\ell=1}^{15}; w\big) \\
&\quad := \big\{\Phi_{j_0,k,m}^h, \Phi_{j_0,k,m}^v\colon \ k \in \mathbb{Z}, \ |k/N_0| \leqslant 2^{j-1}, \ m \in \mathbb{Z}^2\big\} \\
&\qquad \cup \big\{\Psi_{j,k,m}^{h,\ell}, \Psi_{j,k,m}^{v,\ell}\colon \ j, k \in \mathbb{Z}, \ j \geqslant j_0, \ |k/N_0| \leqslant 2^{j-1}, \ m \in \mathbb{Z}^2, \ \ell = 1, \dots, 15\big\} \subseteq L^2\big(\mathbb{R}^2\big),
\end{aligned}$$

where

$$\widehat{\Phi_{j,k,m}^h} = P_h A_j^h X_{(m_1, m_2\beta)}^h S_{k\alpha}^h \Gamma_h \widehat{\phi} \otimes w_\epsilon^\circ \quad \text{and} \quad \widehat{\Phi_{j,k,m}^v} = P_v A_j^v X_{(m_1, m_2\beta)}^v S_{k\alpha}^v \Gamma_v w_\epsilon^\circ \otimes \widehat{\phi}$$

and accordingly

$$\widehat{\Psi_{j,k,m}^{h,\ell}} = P_h A_j^h X_{(m_1, m_2\beta)}^h S_{k\alpha}^h \Gamma_h \widehat{\psi}_\ell \otimes w_\epsilon^\circ \quad \text{and} \quad \widehat{\Psi_{j,k,m}^{v,\ell}} = P_v A_j^v X_{(m_1, m_2\beta)}^v S_{k\alpha}^v \Gamma_v w_\epsilon^\circ \otimes \widehat{\psi}_\ell.$$

For an illustration of the support of the special case of cone-adapted Shannon shearlets and the more general cone-adapted Gabor shearlets, we refer to Fig. 2.

4.2. MRA structure

By classical results from frame theory, the system consisting of the functions $\Phi_{j,k,m}^h$, $\Phi_{j,k,m}^v$ forms a tight frame. To show this, we denote the sets $C_{v,\epsilon} = \{\xi \in \mathbb{R}^2\colon \ |\xi_1/\xi_2| < 1+\epsilon\}$ and $C_{h,\epsilon} = \{\xi \in \mathbb{R}^2\colon \ |\xi_2/\xi_1| < 1+\epsilon\}$.

Theorem 4.3. *Let $H : \mathbb{T} \to \mathbb{C}$ satisfy $|H(z)|^2 + |H(-z)|^2 = 1$ for all $z \in \mathbb{T}$, and let $\epsilon > 0$ be such that H_+, $\overline{H_+}$, H_- and $\overline{H_-}$ defined as in Definition 4.1 have the support of H_+ contained in $C_{h,\epsilon}$ and that of H_- contained in $C_{v,\epsilon}$. Let $N_0 \in 2\mathbb{N}$, $\tau \in \mathbb{N}$, and let w be the unit norm window function of an $\frac{N_0}{\tau}$-tight Gabor frame $\{M_{m_2\tau/2}T_{2k/N_0}w\colon \ m_2, k \in \mathbb{Z}\}$ for $L^2(\mathbb{R})$. Let $\alpha = 2(1+\epsilon)/N_0$ and $\beta = \tau(1+\epsilon)^{-1}/2$, such that w_ϵ° is the window of a unit-norm Gabor frame $\{M_{m_2\beta}T_{k\alpha}w_\epsilon^\circ\}$ for $L^2([-1-\epsilon, 1+\epsilon])$, then the system $\mathcal{CGS}_{j_0}(\phi, \{\psi_\ell\}_{\ell=1}^{15}; w)$ for any $j_0 \in \mathbb{Z}$, or the system*

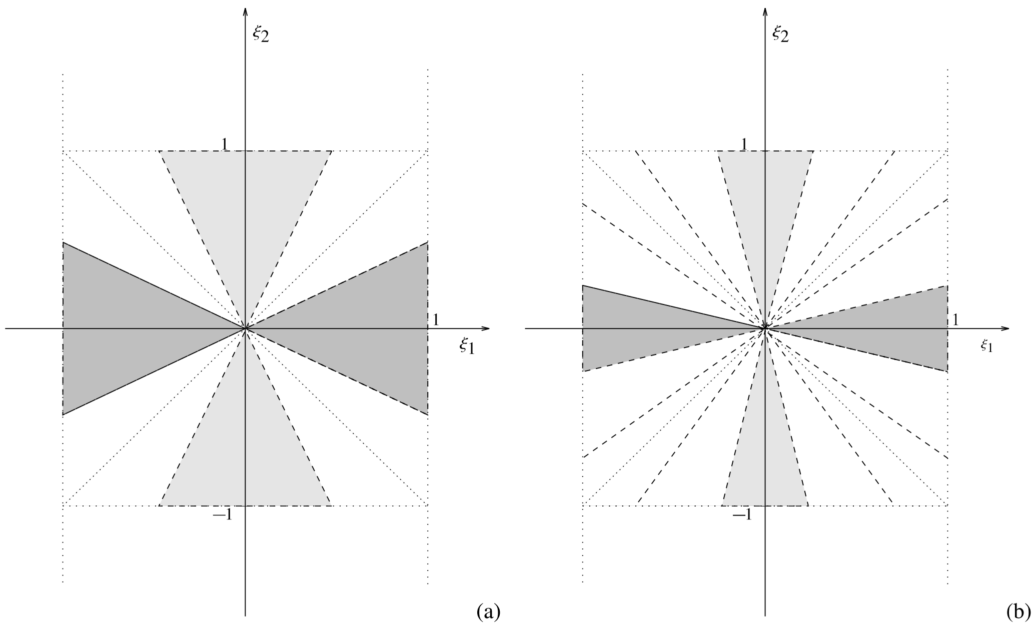

Fig. 2. (a) Support of the cone-adapted Shannon shearlet scaling functions in the frequency domain, in horizontal and vertical orientations; (b) Support of a cone-adapted Gabor shearlet scaling function in the frequency domain, corresponding to a Gabor frame with $N_0 = 4$. The smallest achievable redundancy with $N_0 = 4$ is obtained by setting $\tau = 3$, resulting in $N_0/\tau = 4/3$. With sufficiently large values of N_0, and the implicit finer directional resolution, N_0/τ can be chosen as close to one as desired.

$$\left\{\Psi^{h,\ell}_{j,k,m}, \Psi^{v,\ell}_{j,k,m}\colon\ j,k \in \mathbb{Z},\ |k/N_0| \leqslant 2^{j-1},\ m \in \mathbb{Z}^2,\ \ell = 1,\dots,15\right\}$$

is a unit-norm $\frac{N_0}{\tau}$-tight frame for $L^2(\mathbb{R}^2)$. Moreover, if the wavelets of the 16-band multiresolution analysis $\{\psi_\ell\colon\ \ell = 1,\dots,15\}$, H and w are in addition C^∞, then the shearlets $\{\Psi^{h,\ell}_{j,k,m}, \Psi^{v,\ell}_{j,k,m}\colon\ j,k \in \mathbb{Z},\ |k/N_0| \leqslant 2^{j-1},\ m \in \mathbb{Z}^2,\ \ell = 1,\dots,15\}$ have polynomial decay in $\mathbb{R}^2$.

Proof. The family

$$\left\{A^h_j X^h_{(m_1,m_2\beta)} S^h_{k\alpha} \Gamma_h \widehat{\psi}_\ell \otimes w^\circ_\epsilon\colon\ j,k \in \mathbb{Z},\ |k/N_0| \leqslant 2^{j-1},\ m \in \mathbb{Z}^2;\ \ell = 1,\dots,15\right\}$$

is an $\frac{N_0}{\tau}$-tight frame for $L^2(\mathcal{C}_{h,\epsilon})$. We use the orthogonal projection operator P_h for mapping this family to $L^2(\mathbb{R}^2)$.

Since the functions in range of the projection P_h are supported in $\mathcal{C}_{h,\epsilon}$, the tight frame is mapped to a tight frame for the range of P_h. Consequently, under the projection,

$$\left\{\widehat{\Psi^{h,\ell}_{j,k,m}} = P_h A^h_j X^h_{(m_1,m_2\beta)} S^h_{k\alpha} \Gamma_h \widehat{\psi}_\ell \otimes w^\circ_\epsilon\colon\ j,k \in \mathbb{Z},\ |k/N_0| \leqslant 2^{j-1},\ m \in \mathbb{Z}^2,\ \ell = 1,\dots,15\right\}$$

is an $\frac{N_0}{\tau}$-tight frame for $P_h(L^2(\mathbb{R}^2))$. Similarly,

$$\left\{\widehat{\Psi^{v,\ell}_{j,k,m}} = P_v A^v_j X^v_{(m_1,m_2\beta)} S^v_{k\alpha} \Gamma_v \widehat{\psi}_\ell \otimes w^\circ_\epsilon\colon\ j,k \in \mathbb{Z},\ |k/N_0| \leqslant 2^{j-1},\ m \in \mathbb{Z}^2,\ \ell = 1,\dots,15\right\}$$

is a $\frac{N_0}{\tau}$-tight frame for $P_v(L^2(\mathbb{R}^2))$. By the orthogonality of the ranges for P_h and P_v, the union

$$\left\{\Psi^{h,\ell}_{j,k,m}, \Psi^{v,\ell}_{j,k,m}\colon\ j,k \in \mathbb{Z},\ |k/N_0| \leqslant 2^{j-1},\ m \in \mathbb{Z}^2,\ \ell = 1,\dots,15\right\}$$

is an $\frac{N_0}{\tau}$-tight frame for $L_2(\mathbb{R}^2)$. The proof for the case of $\mathcal{CGS}_{j_0}(\phi, \{\psi_\ell\}_{\ell=1}^{15}; w)$ is similar.

The polynomial decay is a consequence of the smoothness of $\widehat{\Psi_{j,k,m}^{v,\ell}}$ and of $\widehat{\Psi_{j,k,m}^{h,\ell}}$ which follows from the construction. $\square$

5. Optimal sparse approximations

In this section, we show that the cone-adapted Gabor shearlet system $\mathcal{CGS}_{j_0}(\phi, \{\psi_\ell\}_{\ell=1}^{15}; w)$ provides optimally sparse approximation of cartoon-like functions, similar to 'classical' shearlets (see [11,20]). Due to the asymptotic nature of the optimally approximation results, which involve only shearlets with a large scale j, without loss of generality, we consider $\mathcal{CGS}_{j_0}(\phi, \{\psi_\ell\}_{\ell=1}^{15}; w)$ with $j_0 = 0$ and denote the system as $\mathcal{CGS}(\Psi^h, \Psi^v)$. We will first state the main result and the core proof in the following subsection, and postpone the technical parts of the proof to later subsections.

Although we can use a similar strategy is in the proof for classical shearlets [11,20], certain adaptations are necessary. One might have hoped to eliminate repeating some of the tedious estimates by a more elegant, universal result such as showing that Gabor shearlets fall into the class of parabolic molecules [8] or perhaps at least that they have the same asymptotic properties. However, it is not possible to simply linearize the chirp modulations used in the construction of Gabor shearlets while incurring a negligible error on the support of the shearlets because under parabolic scaling, even for shearlets in the high frequency regime, the support size stays commensurate with the scale given by the modulations. In the absence of a universal proof, we have chosen to examine the details of previous proofs and modify them when necessary for our purposes.

5.1. Main result

We first require the definition of cartoon-like functions. For this, we recall that in [3] $\mathcal{E}^2(A)$ denotes the set of *cartoon-like functions* f, which are C^2 functions away from a C^2 edge singularity: $f = f_0 + f_1\chi_B$, where $f_0, f_1 \in C^2([0,1]^2)$ and $\|f\|_{C^2} := \sum_{|v|\leqslant 2} \|\partial^v f\|_\infty \leqslant 1$ with $\partial^v = \partial_1^{v_1}\partial_2^{v_2}$ being the 2D differential operator with order $\partial_1 = \frac{\partial}{\partial x_1}$, $\partial_2 = \frac{\partial}{\partial x_2}$, and $v = (v_1, v_2)$. More precisely, in polar coordinates, let $\rho(\theta) : [0, 2\pi) \mapsto [0,1]^2$ be a radius function satisfying $\sup_\theta |\rho''(\theta)| \leqslant A$ and $\rho \leqslant \rho_0 \leqslant 1$. The set $B \subset \mathbb{R}^2$ is given by $B = \{x \in [0,1]^2\colon \ ||x||_2 \leqslant \rho(\theta)\}$. In particular, the boundary ∂B is given by the curve in $\mathbb{R}^2$: $\beta(\theta) = (\rho(\theta)\cos\theta, \rho(\theta)\sin\theta)$.

Utilizing this notion, we can now formulate out main result concerning optimal sparse approximation of such cartoon-like functions by our cone-adapted Gabor shearlet system as follows. Here, the N-term approximation f_N of f is given by $f_N = \frac{\tau}{N_0}\sum_{j=1}^N \langle f, \Psi_{\mu_j}\rangle \Psi_{\mu_j}$ with $\langle f, \Psi_{\mu_j}\rangle$, $j = 1, \dots, N$ being the N largest coefficients in magnitude.

Theorem 5.1. *Let $\mathcal{CGS}(\Psi^h, \Psi^v)$ satisfy all the assumptions in Theorem 4.3, let $f \in \mathcal{E}^2(A)$ and f_N be the N-term approximation of f from the N largest cone-adapted Gabor shearlet coefficients $\{\langle f, \Psi_\mu\rangle\colon \Psi_\mu \in \mathcal{CGS}(\Psi^h, \Psi^v)\}$ in magnitude, then*

$$\|f - f_N\|_2^2 \leqslant c \cdot N^{-2} \cdot (\log N)^3.$$

To prove this theorem, we follow the main idea as in [3,11]. In a nutshell, we first use a smooth partition of unity that decomposes a cartoon-like function f into small dyadic cubes of size about $2^{-j} \times 2^{-j}$. If j is large enough, then there are only two types of dyadic cubes: one intersects with the singularity of the function, namely, the edge fragments, and the other only contains the smooth region of the function. We then analyze the decay property of the shearlet coefficients. Eventually, by combining the decay estimation of each dyadic cube, we can prove Theorem 5.1.

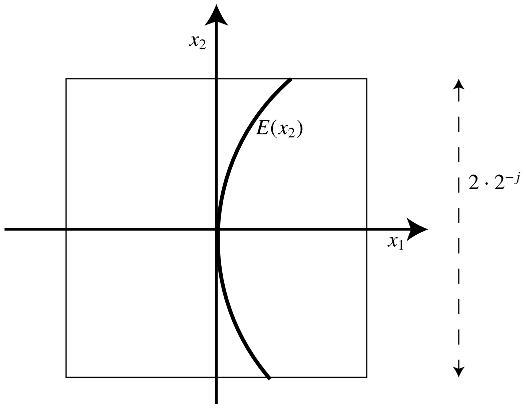

Fig. 3. An edge fragment.

Although the main steps are similar to [3,11], we would like to point out that some of the key steps require slightly technical extensions of results in [3,11]. For the results available in [3,11], we simply state them here without proof for the purpose of readability.

Let us next state some necessary auxiliary results, including Theorem 5.2 for the decay estimate with respect to those edge fragments and Theorem 5.3 for the decay estimate with respect to those smooth regions.

An edge fragment (see Fig. 3) is of the form

$$f(x_1, x_2) = w_0\big(2^j x_1, 2^j x_2\big) g(x_1, x_2) \mathbf{1}_{\{x_1 \geqslant E(x_2)\}}, \tag{1}$$

where w_0, g are smooth functions supported on $[-1,1]^2$ and $|E''(x)| \leqslant A$.

Let $\mathcal{Q}_j$ be the collection of dyadic cubes of the form $Q = [m_1/2^j, (m_1+1)/2^j] \times [m_2/2^j, (m_2+1)/2^j]$. For w_0 a nonnegative C^∞ function with support in $[-1,1]^2$, we can define a smooth partition of unity

$$\sum_{Q \in \mathcal{Q}_j} w_Q(x) = 1, \quad x \in \mathbb{R}^2$$

with $w_Q = w_0(2^j x_1 - m_1, 2^j x_2 - m_2)$. If $Q \in \mathcal{Q}_j$ intersects with the curve singularity, then $f_Q := f w_Q$ is an edge fragment.

Let $\mathcal{Q}_j^0$ be the collection of those dyadic cubes $Q \in \mathcal{Q}_j$ such that the edge singularity intersects with the support of w_Q. Then the cardinality

$$\big|\mathcal{Q}_j^0\big| \leqslant c \cdot 2^j. \tag{2}$$

Similarly, $\mathcal{Q}_j^1 := \mathcal{Q}_j \backslash \mathcal{Q}_j^0$ are those cubes that do not intersect with the edge singularity. We have

$$\big|\mathcal{Q}_j^1\big| \leqslant c \cdot 2^{2j} + 4 \cdot 2^j. \tag{3}$$

Let $\{s_\mu\}$ be a sequence. We define $|s_\mu|_{(N)}$ to be the Nth largest entry of the $\{|s_\mu|\}$. The *weak-ℓ^p quasi-norm* $\|\cdot\|_{w\ell^p}$ of $\{s_\mu\}$ is defined to be

$$\|s_\mu\|_{w\ell^p} := \sup_{N>0} \big(N^{1/p} \cdot |s_\mu|_{(N)}\big),$$

which is equivalent to

$$\|s_\mu\|_{w\ell^p} = \Big(\sup_{\epsilon>0} \big(\big|\{\mu\colon\ |s_\mu| > \epsilon\}\big| \cdot \epsilon^p\big)\Big)^{1/p}.$$

We abbreviate indices for elements in $\mathcal{CGS}(\Psi^h, \Psi^v)$ and write Ψ_μ with $\mu = (j, k, m; \iota, \ell)$. The index set at scale j is $\Lambda_j := \{\mu = (j, k, m; \iota, \ell)\colon\ k \in \mathbb{Z},\ |k/N_0| \leqslant 2^{j-1},\ m \in \mathbb{Z}^2;\ \ell = 1, \ldots, 15,\ \iota = h, v\}$.

Now similar to [11, Theorem 1.3], we have the following result which provides a decay estimate of the coefficients with respect to those $Q \in \mathcal{Q}_j^0$.

Theorem 5.2. *Let $f \in \mathcal{E}^2(A)$ and $f_Q := f w_Q$. For $Q \in \mathcal{Q}_j^0$ with $j \geqslant 0$ fixed, the sequence of coefficients $\{\langle f_Q, \Psi_\mu\rangle\colon\ \mu \in \Lambda_j\}$ obeys*

$$\left\|\langle f_Q, \Psi_\mu\rangle\right\|_{w\ell^{2/3}} \leqslant c \cdot 2^{-3j/2}$$

for some constant c independent of Q and j.

Similarly, for the smooth part, we can show that the sequence of coefficients $\{\langle f_Q, \Psi_\mu\rangle\colon\ \mu \in \Lambda_j\}$ with $Q \in \mathcal{Q}_j^1$ obeys the following estimate (c.f. [11, Theorem 1.4]).

Theorem 5.3. *Let $f \in \mathcal{E}^2(A)$. For $Q \in \mathcal{Q}_j^1$ with $j \geqslant 0$ fixed, the sequence of coefficients $\{\langle f_Q, \Psi_\mu\rangle\colon\ \mu \in \Lambda_j\}$ obeys*

$$\left\|\langle f_Q, \Psi_\mu\rangle\right\|_{w\ell^{2/3}} \leqslant c \cdot 2^{-3j}$$

for some constant independent of Q and j.

The proofs of Theorems 5.2 and 5.3 are very technical and require extension of results in [3,11]. We therefore postpone their detailed proofs to the next two subsections. As a consequence of Theorem 5.2 and Theorem 5.3, it is easy to show the following result.

Corollary 5.1. *Let $f \in \mathcal{E}^2(A)$ and for $j \geqslant 0$, let $s_j(f)$ be the sequence of $s_j(f) = \{\langle f, \Psi_\mu\rangle\colon\ \mu \in \Lambda_j\}$. Then*

$$\left\|s_j(f)\right\|_{w\ell^{2/3}} \leqslant c$$

Proof. By the triangle inequality,

$$\begin{aligned}
\left\|s_j(f)\right\|_{w\ell^{2/3}}^{2/3} &\leqslant \sum_{Q \in \mathcal{Q}_j} \left\|\langle f_Q, \Psi_\mu\rangle\right\|_{w\ell^{2/3}}^{2/3} \\
&\leqslant \sum_{Q \in \mathcal{Q}_j^0} \left\|\langle f_Q, \Psi_\mu\rangle\right\|_{w\ell^{2/3}}^{2/3} + \sum_{Q \in \mathcal{Q}_j^1} \left\|\langle f_Q, \Psi_\mu\rangle\right\|_{w\ell^{2/3}}^{2/3} \\
&\leqslant c \cdot \left|\mathcal{Q}_j^0\right| \cdot 2^{-j} + c \cdot \left|\mathcal{Q}_j^1\right| \cdot 2^{-2j} \\
&\leqslant c. \qquad \square
\end{aligned}$$

Now, we can give the decay rate of our cone-adapted Gabor shearlet coefficients as follows.

Theorem 5.4. *Let $f \in \mathcal{E}^2(A)$ and $s(f) := \{\langle f, \Psi_\mu\rangle\colon\ \Psi_\mu \in \mathcal{CGS}(\Psi^h, \Psi^v)\}$ be the cone-adapted Gabor shearlet coefficients associated with f. Let $\{|s(f)|_{(N)}\colon\ N = 1, 2, \ldots\}$ be the sorted sequence of the absolute values of $s(f)$ in descending order. Then*

$$\sup_{f \in \mathcal{E}^2(A)} \left|s(f)\right|_{(N)} \leqslant c \cdot N^{-3/2} \cdot (\log N)^{3/2}.$$

Proof. From Definition 4.2, we have $\widehat{\Psi^{h,\ell}_{j,k,m}} = P_h g^h_{j,k,m}$ with $g^{h,\ell}_{j,k,m} = A^h_j X^h_{(m_1,m_2\beta)} S^h_{k\alpha} \Gamma_h \widehat{\psi_\ell} \otimes w^\circ_\epsilon$. Then,

$$\begin{aligned}
\widehat{\Psi^{h,\ell}_{j,k,m}}(\xi_1,\xi_2) &= \overline{H_+}\left(I + \frac{1+i}{2}R + \frac{1-i}{2}R^3\right) H_+ g^{h,\ell}_{j,k,m}(\xi_1,\xi_2) \\
&= \left|H\big(\zeta(\xi_2/\xi_1)\big)\right|^2 g^{h,\ell}_{j,k,m}(\xi_1,\xi_2) \\
&\quad + \overline{H\big(\zeta(\xi_2/\xi_1)\big)} H\big(-\zeta(\xi_2/\xi_1)\big)\left(\frac{1+i}{2} g^{h,\ell}_{j,k,m}(\xi_2,-\xi_1) + \frac{1-i}{2} g^{h,\ell}_{j,k,m}(-\xi_2,\xi_1)\right) \\
&=: g_1 + g_2.
\end{aligned}$$

For analyzing the optimal sparsity, we first consider $\Theta^{h,\ell}_{j,k,m} = g_1$, which can be rewritten as follows:

$$\Theta^{h,\ell}_{j,k,m}(\xi_1,\xi_2) \equiv \sigma^\ell_{j,k}\big(\gamma^h(\xi)\big) \cdot e_{j,m}\big(\gamma^h(\xi)\big)$$

with

$$\sigma^\ell_{j,k}\big(\gamma^h(\xi)\big) := \left|H\big(\zeta\big(\gamma^h(\xi)\big)\big)\right|^2 \widehat{\psi}_\ell\big(2^{-4j}\gamma^h_1(\xi)\big) w^\circ_\epsilon\big(2^j\gamma^h_2(\xi) - k\alpha\big) \tag{4}$$

and

$$e_{j,m}\big(\gamma^h(\xi)\big) := 2^{-3j/2} e^{2\pi i m_1 2^{-4j}\gamma^h_1(\xi)} e^{2\pi i\beta m_2 2^j \gamma^h_2(\xi)}. \tag{5}$$

For simplicity, we use again the compact notation $\Theta_\mu(\xi) := \sigma^\ell_{j,k}(\gamma^\iota(\xi)) e_{j,m}(\gamma^\iota(\xi))$ with $\mu = (j,k,m;\iota,\ell) \in \Lambda_j$. The index set Λ_j at scale j is as before.

By Corollary 5.1, we have

$$R(j,\epsilon) := \left|\left\{\mu \in \Lambda_j\colon\ \left|\langle f, \Theta_\mu\rangle\right| > \epsilon\right\}\right| \leqslant c \cdot \epsilon^{-2/3}.$$

Also,

$$\left|\langle f, \Theta_\mu\rangle\right| \leqslant c \cdot 2^{-3j/2}.$$

Therefore, $R(j,\epsilon) = 0$ for $j > \frac{2}{3}\log_2(\epsilon^{-1})$. Thus

$$\left|\left\{\mu\colon\ \left|\langle f, \Psi_\mu\rangle\right| > \epsilon\right\}\right| \leqslant \sum_{j\geqslant 0} R(j,\epsilon) \leqslant c \cdot \epsilon^{-2/3} \cdot \log_2\big(\epsilon^{-1}\big).$$

Repeating the steps for the second term in the definition of $\Psi^{h,\ell}_{j,k,m}$ shows that we can replace Θ_μ by Ψ_μ at the cost of a change of the constant c. This can be seen from the fact that the term $\frac{1+i}{2} g^{h,\ell}_{j,k,m}(\xi_2,-\xi_1) + \frac{1-i}{2} g^{h,\ell}_{j,k,m}(-\xi_2,\xi_1)$ is supported in the vertical cone and thus g_2 can be viewed as composed of two quarter-rotated elements of the form of g_1. The same strategy applies to the vertical cone elements. The theorem is proved. □

Now we can prove Theorem 5.1 using the above results.

Proof of Theorem 5.1. $f_N = \sum_{\mu\in I_N}\langle f, \Psi_\mu\rangle \Psi_\mu$ where I_N is the set of indices corresponding to the N largest entry of $\{|\langle f, \Psi_\mu\rangle|\colon\ \mu\}$. By the tight frame property and Theorem 5.4, we have

$$\|f - f_N\|^2 \leqslant \sum_{n>N} \left|s(f)\right|^2_{(N)} \leqslant c \cdot \sum_{n>N} N^{-3}\log(N)^3 \leqslant c \cdot N^{-2} \cdot \log(N)^3.$$

This finishes the proof of the theorem. □

5.2. Analysis of the edge fragments

We shall focus on proving Theorem 5.2 next. To that end, we need some auxiliary results first. From [11, Theorem 2.2] or [3, Theorem 6.1], we have the following result, which gives the estimate of the decay of the edge fragment in the Fourier domain along a fixed direction. We assume without loss of generality that the edge fragment in this subsection satisfying $E(0) = E'(0) = 0$ and $\sup_{|x_2|\leqslant 2^{-j}} |E(x_2)| \leqslant \frac{1}{2}\sup_{|x_2|\leqslant 2^{-j}} |E''(x_2)|$.

Theorem 5.5. *Let f be an edge fragment as defined in (1) and $I_j := [2^{2j-\sigma}, 2^{2j+\mu}]$ with $\sigma \in \{0,1,2,3,4\}$ and $\mu \in \{0,1,2\}$, then*

$$\int\limits_{|\lambda|\in I_j} \left|\widehat{f}(\lambda\cos\theta, \lambda\sin\theta)\right|^2 d\lambda \leqslant c\cdot 2^{-4j}\cdot\left(1+2^j|\sin\theta|\right)^{-5}.$$

Theorem 5.5 can be extended for a general edge fragment. Interested readers are referred to [3, Section 7.2] and [12, Section 4.5].

Using Theorem 5.5, one can prove the following result (cf. [11, Proposition 2.1]).

Corollary 5.2. *Let f be an edge fragment as defined in (1), then*

$$\int\limits_{\mathbb{R}^2} \left|\widehat{f}(\xi)\right|^2 \left|\sigma^{\ell}_{j,k}\left(\gamma^h(\xi)\right)\right|^2 d\xi \leqslant c\cdot 2^{-3j}\left(1+|k|\right)^{-5}.$$

Note that although $\sigma^{\ell}_{j,k}(\gamma^h(\xi))$ might not be compactly supported compared to [11, Proposition 2.1], it does not affect the result here, since proofs related to the support of $\sigma^{\ell}_{j,k}(\gamma^h(\xi))$ can be passed through its essential support and the estimate outside the essential supported is absorbed in the constant c. For elements in the vertical cone $\sigma^{\ell}_{j,k}(\gamma^v(\xi))$, similarly to the above result, one can show that the decay estimate is of order less than $2^{-3j}(1+|k|)^{-5}$.

From [11, Corollary 2.4] or [3, Corollary 6.6], we have the following result about the decay of the derivative of the edge fragment in the Fourier domain along a fixed direction.

Corollary 5.3. *Let f be an edge fragment as defined in (1) and $v = (v_1, v_2)$, then*

$$\int\limits_{|\lambda|\in I_j} \left|\partial^v \widehat{f}(\lambda\cos\theta, \lambda\sin\theta)\right|^2 d\lambda \leqslant c_v\cdot 2^{-2j|v|}\cdot 2^{-2jv_1}\cdot 2^{-4j}\cdot\left(1+2^j|\sin\theta|\right)^{-5} + c_v\cdot 2^{-2j|v|}\cdot 2^{-10j}.$$

We also need the following lemma (see [11, Lemma 2.5]), which follows from a direct computation.

Lemma 5.1. *Let $\sigma^{\ell}_{j,k}(\gamma^h(\xi))$ be given as above, then, for each $v = (v_1, v_2) \in \mathbb{N}^2$, $v_1, v_2 \in \{0,1,2\}$,*

$$\left|\partial^v \sigma^{\ell}_{j,k}\left(\gamma^h(\xi)\right)\right| \leqslant c_v\cdot 2^{-(2v_1+v_2)j}\cdot\left(1+|k|\right)^{v_1},$$

where $|v| = v_1 + v_2$ and c_v is independent of j and k.

Use the above results, we can prove the following result, which is an extension of [11, Proposition 2.3] and can be proved with a similar approach.

Corollary 5.4. *Let f be an edge fragment defined as in (1), $\sigma^{\ell}_{j,k}(\gamma^h(\xi))$ be defined as above, and L_t be the differential operator defined by*

$$L_t = \left(t \cdot I - \left(\frac{2^{2j}}{2\pi(1+|k|)}\right)^2 \partial_1^2\right)\left(I - \left(\frac{2^j}{2\pi}\right)^2 \partial_2^2\right),$$

where $t > 0$ is a fixed constant. Then

$$\int_{\mathbb{R}^2} \left| L_t\left(\widehat{f}(\xi)\sigma_{j,k}^{\ell}\left(\gamma^h(\xi)\right)\right)\right|^2 d\lambda \leqslant c_t \cdot 2^{-3j}\left(1+|k|\right)^{-5}$$

for some positive constant c_t independent of j and k.

Now we are ready to prove Theorem 5.2.

Proof of Theorem 5.2. Fix $j \geqslant 0$, for simplicity, let $f = f_Q$ be the edge fragment as in (1). We have

$$\langle f, \Psi_\mu \rangle = \int_{\mathbb{R}^2} \widehat{f}(\xi)\overline{\sigma_{j,k}^{\ell}\left(\gamma^h(\xi)\right)} \cdot \overline{e_{j,m}\left(\gamma^h(\xi)\right)}\, d\xi.$$

We have

$$\partial_1 e_{j,m}\left(\gamma^h(\xi)\right) = \left(2\pi i m_1 2^{-4j} \operatorname{sgn}(\xi_1)\xi_1 - 2\pi i m_2 \beta 2^j \frac{\xi_2}{\xi_1^2}\right) e_{j,m}\left(\gamma^h(\xi)\right).$$

$$\partial_1^2 e_{j,m}\left(\gamma^h(\xi)\right) = \left(2\pi i m_1 2^{-4j} \operatorname{sgn}(\xi_1) + 2\pi i m_2 \beta 2^{j+1} \frac{\xi_2}{\xi_1^3}\right) e_{j,m}\left(\gamma^h(\xi)\right)$$
$$+ \left(2\pi i m_1 2^{-4j} \operatorname{sgn}(\xi_1)\xi_1 - 2\pi i m_2 \beta 2^j \frac{\xi_2}{\xi_1^2}\right)^2 e_{j,m}\left(\gamma^h(\xi)\right).$$

Also,

$$\partial_2^2 e_{j,m}\left(\gamma^h(\xi)\right) = \left(2\pi i m_2 \beta \frac{2^j}{\xi_1}\right)^2 e_{j,m}\left(\gamma^h(\xi)\right).$$

Let L_t be the differential operator defined in Corollary 5.4. Then,

$$L_t\left(e_{j,m}\left(\gamma^h(\xi)\right)\right) = g_{j,k}^m(\xi) e_{j,m}\left(\gamma^h(\xi)\right)$$

with

$$g_{j,k}^m(\xi) = \left[t + \frac{\left(\frac{m_1}{2\pi i}\operatorname{sgn}(\xi_1) + \frac{1}{2\pi i} m_2 \beta \frac{2^{5j+1}\xi_2}{\xi_1^3}\right) + \left(m_1 \frac{\operatorname{sgn}(\xi_1)\xi_1}{2^{2j}} - m_2 \beta \frac{2^{3j}\xi_2}{\xi_1^2}\right)^2}{(1+|k|)^2}\right] \cdot \left[1 + \left(m_2 \beta \frac{2^{2j}}{\xi_1}\right)^2\right].$$

Let $W_{j,k}$ be the essential support of $\sigma_{j,k}^{\ell}(\gamma^h(\xi))$ defined as

$$W_{j,k} := \left\{(\lambda, \theta)\colon\ 2^{2j} a' \leqslant |\lambda| \leqslant 2^{2j} b',\ \arctan\left(2^{-j}(k\alpha - 1)\right) \leqslant \theta \leqslant \arctan\left(2^{-j}(k\alpha + 1)\right)\right\}. \tag{6}$$

For $\xi \in W_{j,k}$, we have $|\xi_1| \approx 2^{2j}$ and one can show that $\frac{2^j \xi_2}{\xi_1} \approx 2^j \tan\theta \approx k$. Consequently, we can choose a large $t > 0$ independent of j, k, m such that

$$\sup_{\xi \in W_{j,k}} \left|g_{j,k}^m(\xi)\right| \geqslant c \cdot \left[1 + \frac{(m_1 - m_2 k)^2}{(1+|k|)^2}\right] \cdot \left[1 + m_2^2\right] =: c \cdot G_k(m) \tag{7}$$

for some positive constant c independent of j, k, and m. For $G_k(m)$, we have

$$G_k(m) = \begin{cases} (1+m_1^2)(1+m_2^2) & \text{for } k=0 \\ [1+\frac{(\frac{m_1}{k}-m_2)^2}{(1+|k|)^2/|k|^2}]\cdot[1+m_2^2] & \text{for } k\neq 0. \end{cases}$$

Consequently,

$$\begin{aligned} \langle f, \Psi_\mu\rangle &= \int_{\mathbb{R}^2} \widehat{f}(\xi)\overline{\sigma_{j,k}^{\ell}\big(\gamma^h(\xi)\big)}\cdot \overline{e_{j,m}\big(\gamma^h(\xi)\big)}\,d\xi \\ &= \int_{\mathbb{R}^2} L_t\big(\widehat{f}(\xi)\overline{\sigma_{j,k}^{\ell}\big(\gamma^h(\xi)\big)}\big)\cdot L_t^{-1}\big(\overline{e_{j,m}\big(\gamma^h(\xi)\big)}\big)\,d\xi \\ &= \int_{\mathbb{R}^2} \frac{L_t(\widehat{f}(\xi)\overline{\sigma_{j,k}^{\ell}(\gamma^h(\xi))})}{\overline{g_{j,k}^{m}(\xi)}}\cdot\overline{e_{j,m}\big(\gamma^h(\xi)\big)}\,d\xi. \end{aligned}$$

For $k\neq 0$ and $\tilde{m} := (\tilde{m}_1, \tilde{m}_2) \in \mathbb{Z}^2$, define $R_{\tilde{m}} := \{m = (m_1, m_2) \in \mathbb{Z}^2\colon \frac{m_1}{k} \in [\tilde{m}_1, \tilde{m}_1 + 1),\ m_2 = \tilde{m}_2\}$. Since for j, k fixed, $\{e_{j,m}(\gamma^h(\xi))\colon k \in \mathbb{Z}^2\}$ is an orthonormal basis for L^2 functions supported on $W_{j,k}$, we obtain

$$\begin{aligned} \sum_{m\in R_{\tilde{m}}} \big|\langle f,\Psi_\mu\rangle\big|^2 &\leqslant \int_{\mathbb{R}^2}\left|\frac{L_t(\widehat{f}(\xi)\overline{\sigma_{j,k}^{\ell}(\gamma^h(\xi))})}{\overline{g_{j,k}^{m}(\xi)}}\right|^2 d\xi \\ &\leqslant \sup_{\xi\in W_{j,k}}\frac{1}{|g_{j,k}^{m}(\xi)|^2}\int_{\mathbb{R}^2}\big|L_t\big(\widehat{f}(\xi)\overline{\sigma_{j,k}^{\ell}\big(\gamma^h(\xi)\big)}\big)\big|^2\,d\xi \\ &\leqslant c\cdot\frac{1}{[(1+(m_1/k-m_2)^2)(1+m_2^2)]^2}\int_{\mathbb{R}^2}\big|L_t\big(\widehat{f}(\xi)\overline{\sigma_{j,k}^{\ell}\big(\gamma^h(\xi)\big)}\big)\big|^2\,d\xi. \end{aligned}$$

By Corollary 5.4, we have

$$\sum_{m\in R_{\tilde{m}}}\big|\langle f,\Psi_\mu\rangle\big|^2 \leqslant c\cdot G_{\tilde{m}}^{-2}\cdot 2^{-3j}\big(1+|k|\big)^{-5}$$

with $G_{\tilde{m}} := (1+(\tilde{m}_1-\tilde{m}_2)^2)(1+\tilde{m}_2^2)$. For $k=0$, similarly, we have $G_{\tilde{m}} = (1+\tilde{m}_1^2)(1+\tilde{m}_2^2)$.

Let $N_{j,k,\tilde{m}}(\epsilon) := |\{m \in R_{\tilde{m}}\colon |\langle f,\Psi_\mu\rangle| > \epsilon\}|$. Then $N_{j,k,\tilde{m}}(\epsilon) \leqslant c\cdot(1+|k|)$ and the above inequality implies

$$N_{j,k,\tilde{m}}(\epsilon) \leqslant c\cdot G_{\tilde{m}}^{-2}\cdot 2^{-3j}\cdot\epsilon^{-2}\cdot\big(1+|k|\big)^{-5}.$$

Thus,

$$N_{j,k,\tilde{m}}(\epsilon) \leqslant c\cdot\min\big(1+|k|, G_{\tilde{m}}^{-2}\cdot 2^{-3j}\cdot\epsilon^{-2}\cdot\big(1+|k|\big)^{-5}\big),$$

which implies

$$\sum_{k=-2^j}^{2^j} N_{j,k,\tilde{m}}(\epsilon) \leqslant c\cdot G_{\tilde{m}}^{-2/3}\cdot 2^{-j}\cdot\epsilon^{-2/3}.$$

Since $\sum_{\tilde{m}\in\mathbb{Z}^2} G_{\tilde{m}}^{-2/3} < \infty$, by above inequality, we obtain

$$\left|\left\{\mu \in M_j\colon \left|\langle f, \Psi_\mu\rangle\right| > \epsilon\right\}\right| \leqslant \sum_{m\in\mathbb{Z}^2}\sum_{k=-2^j}^{2^j} N_{j,k,\tilde{m}}(\epsilon) \leqslant c\cdot 2^{-j}\epsilon^{-2/3},$$

which is equivalent to the conclusion that

$$\left\|\langle f_Q, \Psi_\mu\rangle\right\|_{w\ell^{2/3}} \leqslant c\cdot 2^{-3j/2}. \qquad \square$$

5.3. Analysis of the smooth region

Now, we shall focus on proving Theorem 5.3. Let us provide some lemmas first. From [3, Lemma 8.1] or [11, Lemma 2.6], we have

Lemma 5.2. *Let $f = gw_Q$, where $g \in \mathcal{E}^2(A)$ and $Q \in \mathcal{Q}_j^1$. Then*

$$\int\limits_{W_{j,k}} \left|\widehat{f}(\xi)\right|^2 d\xi \leqslant c\cdot 2^{-10j},$$

where $W_{j,k}$ is the essential support of $\sigma_{j,k}^{\ell}(\gamma^h(\xi))$ as in (6).

From [11, Lemma 2.7] we have

Lemma 5.3. *for $v = (v_1, v_2) \in \mathbb{N}^2$,*

$$\sum_{k=-2^j}^{2^j} \left|\partial^v \sigma_{j,k}^{\ell}\left(\gamma^h(\xi)\right)\right|^2 \leqslant c\cdot 2^{-2|v|j}.$$

Using the above two lemmas, one can easily prove the following result, which is an extension of [11, Lemma 2.8] and can be proved by a similar approach.

Lemma 5.4. *Let $f = gw_Q$, where $g \in \mathcal{E}^2(A)$ and $Q \in \mathcal{Q}_j^1$. Define the differential operator $L_t := (tI - \frac{2^{2j}}{(2\pi)^2}\Delta)$ with $t > 0$ and $\Delta = \partial_1^2 + \partial_2^2$. Then,*

$$\int\limits_{\mathbb{R}^2}\sum_{k=-2^j}^{2^j} \left|L_t^2\left(\widehat{f}(\xi)\sigma_{j,k}^{\ell}\left(\gamma^h(\xi)\right)\right)\right|^2 d\xi \leqslant c_t\cdot 2^{-10j}$$

for some positive constant c_t independent of j.

Now we are ready to prove Theorem 5.3.

Proof of Theorem 5.3. Let $f = f_Q = gw_Q$ and L_t as defined in Lemma 5.4. We have

$$L_t\left(e_{j,m}\left(\gamma^h(\xi)\right) = g_{j,k}^m(\xi)e_{j,m}\left(\gamma^h(\xi)\right),\right.$$

where

$$g_{j,k}^m(\xi) = \left[t + \frac{m_1 2^{-2j}\operatorname{sgn}(\xi_1) + m_2\beta 2^{3j+1}\frac{\xi_2}{\xi_1^3}}{2\pi i} + 2^{2j}\left(m_1 2^{-4j}\operatorname{sgn}(\xi_1)\xi_1 - m_2\beta 2^j\frac{\xi_2}{\xi_1^2}\right)^2 \right.$$
$$\left. + \left(m_2\beta\frac{2^{2j}}{\xi_1}\right)^2\right] e_{j,m}\big(\gamma^h(\xi)\big).$$

Similar argument to the proof of Theorem 5.2, we can choose $t > 0$ large enough so that

$$\sup_{\xi\in W_{j,k}} \left|g_{j,k}^m(\xi)\right| \geqslant c\cdot\left[1 + 2^{-2j}(m_1 - m_2k)^2 + m_2^2\right].$$

For $\tilde{m} := (\tilde{m}_1, \tilde{m}_2) \in \mathbb{Z}^2$, define $R_{\tilde{m}} := \{m = (m_1, m_2) \in \mathbb{Z}^2\colon\ 2^{-2j}(m_1 - m_2k) \in [\tilde{m}_1, \tilde{m}_1 + 1),\ m_2 = \tilde{m}_2\}$. Observe that for each $\tilde{m}$, there are only $1 + 2^{2j}$ choices for m_1 in $R_{\tilde{m}}$. Hence $|R_{\tilde{m}}| \leqslant 1 + 2^{2j}$. Again, similar argument to the proof of Theorem 5.2, we have

$$\sum_{m\in R_{\tilde{m}}} \left|\langle f, \Psi_\mu\rangle\right|^2 \leqslant c\cdot \sup_{\xi\in W_{j,k}} \frac{1}{|g_{j,k}^m(\xi)|^4} \int_{\mathbb{R}^2} \left|L_t^2\big(\hat{f}(\xi)\overline{\sigma_{j,k}^\ell\big(\gamma^h(\xi)\big)}\big)\right|^2 d\xi$$
$$\leqslant c\cdot\frac{1}{[1 + 2^{-2j}(m_1 - m_2k)^2 + m_2^2]^4} \int_{\mathbb{R}^2} \left|L_t^2\big(\hat{f}(\xi)\overline{\sigma_{j,k}^\ell\big(\gamma^h(\xi)\big)}\big)\right|^2 d\xi.$$

Then by Lemma 5.4,

$$\sum_{k=-2^j}^{2^j} \sum_{m\in R_{\tilde{m}}} \left|\langle f, \Psi_\mu\rangle\right|^2 \leqslant c\cdot G_{\tilde{m}}^{-4} \cdot \int_{\mathbb{R}^2} \sum_{k=-2^j}^{2^j} \left|L_t^2\big(\hat{f}(\xi)\overline{\sigma_{j,k}^\ell\big(\gamma^h(\xi)\big)}\big)\right|^2 d\xi$$
$$\leqslant c\cdot G_{\tilde{m}}^{-4}\cdot 2^{-10j}$$

where $G_{\tilde{m}} := 1 + \tilde{m}_1^2 + \tilde{m}_2^2$.

Using the Hölder inequality

$$\sum_{m=1}^N |a_m|^p \leqslant \left(\sum_{m=1}^N |a_m|^2\right)^{p/2} N^{1-p/2}, \quad 1/2 < p < 2.$$

Since the cardinality of $R_{\tilde{m}}$ is bounded by $1 + 2^{2j}$, we have

$$\sum_{k=-2^j}^{2^j} \sum_{m\in R_{\tilde{m}}} \left|\langle f, \Psi_\mu\rangle\right|^p \leqslant c\cdot\left(2^{2j}\right)^{1-p/2}\cdot G_{\tilde{m}}^{-2p}\cdot 2^{-5pj}.$$

Moreover, since $p > 1/2$, $\sum_{\tilde{m}\in\mathbb{Z}^2} G_{\tilde{m}}^{-2p} < \infty$. Consequently,

$$\sum_{\mu\in M_j} \left|\langle f, \Psi_\mu\rangle\right|^p \leqslant c\cdot 2^{2j(1-p/2)-5pj} = c\cdot 2^{2j(1-3p)}.$$

In particular

$$\left\|\langle f, \Psi_\mu\rangle\right\|_{\ell^{2/3}} \leqslant c\cdot 2^{-3j}. \qquad \square$$

References

[1] B.G. Bodmann, P.G. Casazza, G. Kutyniok, A quantitative notion of redundancy for finite frames, Appl. Comput. Harmon. Anal. 30 (2011) 348–362.
[2] J. Cahill, P.G. Casazza, A. Heinecke, A quantitative notion of redundancy for infinite frames, preprint.
[3] E.J. Candès, D.L. Donoho, New tight frames of curvelets and optimal representations of objects with C^2 singularities, Comm. Pure Appl. Math. 56 (2004) 219–266.
[4] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conf. Ser. in Appl. Math., vol. 61, SIAM, Philadelphia, PA, 1992.
[5] I. Daubechies, A. Grossmann, Y. Meyer, Painless nonorthogonal expansions, J. Math. Phys. 27 (1986) 1271–1283.
[6] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (2005) 2091–2106.
[7] K. Gröchenig, Foundations of Time–Frequency Analysis, Birkhäuser, Boston, 2001.
[8] P. Grohs, G. Kutyniok, Parabolic molecules, Found. Comput. Math. 14 (2) (2014) 299–337.
[9] K. Guo, G. Kutyniok, D. Labate, Sparse multidimensional representations using anisotropic dilation and shear operators, in: Wavelets and Splines, Athens, GA, 2005, Nashboro Press, Nashville, TN, 2006, pp. 189–201.
[10] K. Guo, D. Labate, W. Lim, G. Weiss, E. Wilson, Wavelets with composite dilations and their MRA properties, Appl. Comput. Harmon. Anal. 20 (2006) 231–249.
[11] K. Guo, D. Labate, Optimally sparse multidimensional representation using shearlets, SIAM J. Math. Anal. 39 (2007) 298–318.
[12] K. Guo, D. Labate, Optimally sparse representations of 3D data with C^2 surface singularities using Parseval frames of shearlets, SIAM J. Math. Anal. 44 (2012) 851–886.
[13] B. Han, G. Kutyniok, Z. Shen, A unitary extension principle for Shearlet Systems, SIAM J. Numer. Anal. 49 (2011) 1921–1946.
[14] B. Han, S. Kwon, X. Zhuang, Generalized interpolating refinable function vectors, J. Comput. Appl. Math. 227 (2009) 254–270.
[15] B. Han, X. Zhuang, Matrix extension with symmetry and its application to symmetric orthonormal multiwavelets, SIAM J. Math. Anal. 42 (2010) 2297–2317.
[16] B. Han, X. Zhuang, Algorithms for matrix extension and orthogonal wavelet filter banks over algebraic number fields, Math. Comp. 82 (2013) 459–490.
[17] R. Houska, The nonexistence of shearlet scaling functions, Appl. Comput. Harmon. Anal. 32 (2012) 28–44.
[18] P. Kittipoom, G. Kutyniok, W.-Q. Lim, Construction of compactly supported shearlet frames, Constr. Approx. 35 (2012) 21–72.
[19] G. Kutyniok, D. Labate, Shearlets: Multiscale Analysis for Multivariate Data, Birkhäuser, Boston, 2012.
[20] G. Kutyniok, W.-Q. Lim, Compactly supported shearlets are optimally sparse, J. Approx. Theory 163 (2011) 1564–1589.
[21] G. Kutyniok, T. Sauer, Adaptive directional subdivision schemes and shearlet multiresolution analysis, SIAM J. Math. Anal. 41 (2009) 1436–1471.
[22] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, 1998.
[23] P.L. Søndergaard, Gabor frames by sampling and periodization, Adv. Comput. Math. 27 (2007) 355–373.
[24] G. Weiss, E. Hernández, A First Course on Wavelets, CRC Press, Boca Raton, 1996.