Parabolic Molecules

Philipp Grohs and Gitta Kutyniok

June 9, 2012

Abstract

Anisotropic decompositions using representation systems based on parabolic scaling such as curvelets or shearlets have recently attracted significantly increased attention due to the fact that they were shown to provide optimally sparse approximations of functions exhibiting singularities on lower dimensional embedded manifolds. The literature now contains various direct proofs of this fact and of related sparse approximation results. However, it seems quite cumbersome to prove such a canon of results for each system separately, while many of the systems exhibit certain similarities.

In this paper, with the introduction of the notion of *parabolic molecules*, we aim to provide a comprehensive framework which includes customarily employed representation systems based on parabolic scaling such as curvelets and shearlets. It is shown that pairs of parabolic molecules have the fundamental property to be almost orthogonal in a particular sense. This result is then applied to analyze parabolic molecules with respect to their ability to sparsely approximate data governed by anisotropic features. For this, the concept of *sparsity equivalence* is introduced which is shown to allow the identification of a large class of parabolic molecules providing the same sparse approximation results as curvelets and shearlets. Finally, as another application, smoothness spaces associated with parabolic molecules are introduced providing a general theoretical approach which even leads to novel results for, for instance, compactly supported shearlets.

Keywords: Curvelets, Nonlinear Approximation, Parabolic Scaling, Shearlets, Smoothness Spaces, Sparsity Equivalence

2000 Mathematics Subject Classification. Primary 41AXX, Secondary 41A25, 53B, 22E.

1 Introduction

Recently, a paradigm shift could be observed in applied mathematics, computer science, and electrical engineering. The novel paradigm of sparse approximations now enables not only highly efficient encoding of functions and signals, but also provides intriguing new methodologies, for instance, for recovery of missing data or separation of morphologically distinct components. At about the same time, scientists began to question whether wavelets are indeed perfectly suited for image processing tasks, the main reason being that images are governed by edges while wavelets are isotropic objects. This mismatch becomes also evident when recalling that Besov spaces can be characterized by the decay of wavelet coefficient sequences however Besov models are clearly deficient to adequate capturing of edges.

1.1 Geometric Multiscale Analysis

These two fundamental observations have led to the research area of geometric multiscale analysis whose main goal is to develop representation systems, preferably containing different scales, which are sensitive to anisotropic features in functions/signals and provide sparse approximations of those. Such representation systems are typically based on *parabolic scaling*, and we exemplarily mention (first and second generation) curvelets [8], contourlets [12], and shearlets [25]. Browsing through the literature, it becomes evident that some properties such as sparse approximation of so-called cartoon images are quite similar for some systems such as curvelets and shearlets, whereas other systems such as contourlets show a somehow different behavior. Delving more into the literature we observe that for those systems exhibiting similar sparsity behavior many results were proven with quite resembling proofs. One might ask: Is this cumbersome close repetition of proofs really necessary? We believe that the answer is *no* and that a general framework for representation systems based on parabolic scaling does not only solve this problem but even more provide a fundamental understanding of such systems and allow for a categorization of these.

1.2 Parabolic molecules

The main goal of this paper is to proclaim the framework of parabolic molecules as a general concept encompassing in particular curvelets and both band-limited and compactly supported shearlets. The idea of *molecules* in geometric multiscale analysis dates back to the seminal work by Candès and Demanet [5], in which they studied the curvelet representation of wave propagators by using what they called *curvelet molecules*. Later, Guo and Labate adopted this idea and introduced *shearlet molecules* in [21].

Both such generalization approaches however suffer from the fact that they are solely designed to weaken the conditions of the particular respective systems, namely curvelets and shearlets. In contrast to this, our philosophy is to introduce molecules, which encompass a wide class of directional representation systems by using parabolic scaling as a unifying concept. This is justified by the fact that all known systems providing optimally sparse approximation of cartoon images follow a parabolic scaling law; and it is strongly believed that this is necessary. In fact, our framework is general enough to, for instance, include all known curvelet-type as well as shearlet-type constructions to date.

Our main result (Theorem 2.9) will show that the Gramian matrix between any two systems of parabolic molecules satisfies a strong off-diagonal decay property and is in that sense very close to a diagonal matrix. This will become key to transfer the celebrated properties of curvelet systems to other systems based on parabolic scaling; a fact which we can summarize in the following meta-result:

Meta-Theorem. *All frame systems based on parabolic scaling (specifically curvelets and shearlets) posses the exact same approximation properties, whenever the generating functions are sufficiently smooth, as well as localized in space and frequency.*

This meta-theorem has been verified on a case-by-case basis for a number of different systems. In this paper, for the first time, a rigorous framework is provided which applies to, for instance, all known curvelet or shearlet constructions at once. This will be exemplarily demonstrated by the results on sparse approximation (Theorem 4.6) and anisotropic smoothness spaces (Theorem 4.10) which are indeed universally applicable to all parabolic molecules.

1.3 Sparsity Equivalence

Focussing on the property of sparse approximation of images governed by anisotropic features, it might be even more beneficial to derive a categorization of parabolic molecules according to their approximation behavior. We accommodate this request by introducing the notion of *sparsity equivalence* in Subsection 4.1, which leads to equivalence classes and further to the sought classification. As a byproduct, our framework yields a simple derivation of the results in [20, 27] from [8]. In fact, our results provide a systematic way to analyze the sparse approximation of cartoon images of systems by elements of the class of parabolic molecules categorized by equivalence classes of sparsity equivalence.

1.4 Contribution and Expected Impact

Summarizing, the significance of the notion of parabolic molecules as a higher level viewpoint lies in the fact that it not only provides a general framework which contains various directional representation systems as special cases and enables a quantitative comparison of such, but it moreover allows the transfer of results concerning properties of such systems without repeating quite similar proofs. A few examples, for which this conceptually new approach is fruitful, will be presented in Section 4 including optimally sparse approximations of cartoon-like images.

We therefore anticipate this novel framework to have the following impacts:

- A thorough understanding of the ingredients of representation systems based on parabolic scaling which are crucial for an observed behavior such as sparse approximations of cartoon images, thereby also categorizing different (sparsity) behaviors.

- A framework within which results can be directly transferred from one system to others without repetition of similar proofs. This will allow to establish a desired result for a system based on parabolic scaling by choosing, for instance, a shearlet or curvelet system best suited for the proof, and transfer the result subsequently to any other systems by utilizing the results in this paper.
- An approach to design new representation systems based on parabolic scaling depending on several parameters whose impact on, for instance, sparse approximation behavior is then known in advance.

1.5 Extensions

The framework introduced in this paper and the derived results are amenable to generalizations and extensions. We briefly discuss a few examples.

- *Other Systems.* This general framework supports the introduction of novel systems based on parabolic scaling fulfilling a list of desiderata designed according to a particular application. Such systems can now be objectively compared with other systems according to, for instance, sparse approximation properties.
- *Systems with Continuous Parameters.* One might also ask whether a similar general framework for systems based on parabolic scaling with continuous parameters can be introduced. In light of Subsection 4.1, this however requires a different sparsity model; one conceivable path would be to compare their ability to resolve wavefront sets.
- *Further Properties.* In this paper, we studied the impact of our general framework on the problems of sparse approximation and anisotropic function spaces. This strategy can certainly be also used for other applications such as efficient decomposition of the Radon transform, which has been studied both for shearlets [15] and curvelets [7], as well as the analysis of geometric separation as studied in [13].
- *Weighted Norms.* When aiming at transferring results such as sparse decompositions of curvilinear integrals [6] or sparse decompositions of the Radon transform [7], sometimes weighted ℓ_p norms might need to be analyzed.
- *Higher Dimensions.* We have formulated our results in the bivariate setting. However, an extension to arbitrary dimensions is possible using essentially the same arguments. This is especially relevant since by now several different curvelet and shearlet constructions exist for three-dimensional data [30, 26, 23].

1.6 Outline

This paper is organized as follows. In Section 2, the notion of parabolic molecules is introduced. It is then shown in Section 3 that curvelets and both band-limited and compactly supported shearlets are special cases of this framework. Almost orthogonality of pairs of parabolic molecules is proven in Section 5. Finally, in Section 4, this result is utilized for two applications. First, in Subsection 4.1, using the novel concept of sparsity equivalence a large class of parabolic molecules providing the same sparse approximation results as curvelets and shearlets is identified. Second, in Subsection 4.2, smoothness spaces associated with parabolic molecules are studied.

1.7 Notation

We comment on the notation which we shall use in the present work. Denote by $L_p(\mathbb{R}^d)$ the usual Lebesgue spaces with associated norm $\|\cdot\|_p$. For a discrete set Λ equipped with the counting measure we denote the corresponding Lebesgue space by $\ell_p(\Lambda)$ or ℓ_p if Λ is known from the context. The associated norm will again be denoted $\|\cdot\|_p$. We use the symbol $\langle\cdot,\cdot\rangle$ indiscriminately for the inner product on the Hilbert space $L_2(\mathbb{R}^d)$ as well as for the Euclidean inner product on $\mathbb{R}^d$. The Euclidean norm $\langle x,x\rangle^{1/2}$ of a vector $x \in \mathbb{R}^d$ will be denoted by $|x|$. For a function $f \in L_1(\mathbb{R}^d)$ we can define the Fourier transform $\hat{f}(\omega) := \int_{\mathbb{R}^d} f(x)\exp(-2\pi i\langle x,\omega\rangle)\mathrm{d}x$. By density this definition can be extended to tempered distributions f. We shall also use the notation $\mathbb{T}$ to denote the one-dimensional torus which can be identified with the half-open interval $[0, 2\pi)$. Sometimes we will use the notations $(x)_+ := \max(x, 0)$ $\lfloor x \rfloor := \max\{l \in \mathbb{Z} : l \leq x\}$, and $\langle x \rangle := (1 + x^2)^{1/2}$. Finally, we use the symbol $A \lesssim B$ to indicate that $A \leq CB$ with a uniform constant C.

2 Parabolic Molecules

All anisotropic transforms based on parabolic scaling which have appeared in the literature are indexed by a scale parameter describing the amount of anisotropic scaling, an angular parameter describing the orientation and a spacial parameter describing the location of an element. Nevertheless, these specific constructions are based on different principles: For curvelets the scaling is done by a dilation with respect to polar coordinates and the orientation is enforced by rotations. Shearlets on the other hand are based on affine scaling of a single generator and the directionality is generated by the action of shear matrices.

It is the purpose of this section to introduce the concept of parabolic molecules which distills the essential properties out of these constructions in terms of time-frequency localization properties. As it will turn out, all previous constructions of curvelets and shearlets are instances of this concept, a fact that enables us to operate in a much more general setup than in previous work.

2.1 Definition of Parabolic Molecules

Let us now describe our setup. We start by defining our parameter space

$$\mathbb{P} := \mathbb{R}_+ \times \mathbb{T} \times \mathbb{R}^2,$$

where a point $p = (s, \theta, x) \in \mathbb{P}$ describes a scale 2^s, an orientation θ, and a location x.

Parabolic molecules are defined as systems of functions $(m_\lambda)_{\lambda \in \Lambda}$ with each $m_\lambda \in L_2(\mathbb{R}^2)$ satisfying some additional properties. In particular, each function m_λ will be associated with a unique point in $\mathbb{P}$, as we shall make precise below. Since we are dealing with discrete systems (frames) we would like to operate with discrete sampling sets contained in $\mathbb{P}$. We call such sampling sets *parametrizations* as defined below.

Definition 2.1. *A* parametrization *consists of a pair* (Λ, Φ_Λ) *where* Λ *is a discrete index set and* Φ_Λ *is a mapping*

$$\Phi_\Lambda : \left\{ \begin{array}{ccc} \Lambda & \to & \mathbb{P}, \\ \lambda \in \Lambda & \mapsto & (s_\lambda, \theta_\lambda, x_\lambda)\,. \end{array} \right.$$

which associate with each $\lambda \in \Lambda$ *a* scale s_λ, *a* direction θ_λ *and a* location $x_\lambda \in \mathbb{R}^2$.

There exists a canonical parametrization

$$\Lambda^0 := \left\{ (j, l, k) \in \mathbb{Z}^4 \ : \ j \geq 0,\ l = -2^{\lfloor \frac{j}{2} \rfloor - 1}, \cdots, 2^{\lfloor \frac{j}{2} \rfloor - 1} \right\},$$

where for $\lambda = (j, l, k)$ *we define* $\Phi^0(\lambda) := (s_\lambda, \theta_\lambda, x_\lambda)$ *with* $s_\lambda := j$, $\theta_\lambda := l 2^{-\lfloor j/2 \rfloor} \pi$, *and* $x_\lambda := R_{-\theta_\lambda} D_{2^{-s_\lambda}} k$.

We are now ready to define parabolic molecules. Our definition essentially says that molecules have frequency support in parabolic wedges associated to a certain orientation and spacial support in rectangles with parabolic aspect ratio.

For this, we will use the following notion. We let $R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ denote the rotation matrix of angle θ, and $D_a := \operatorname{diag}(a, \sqrt{a})$ be the anisotropic dilation matrix associated with $a > 0$.

Definition 2.2. *Let* Λ *be a parametrization. A family* $(m_\lambda)_{\lambda \in \Lambda}$ *is called a* family of parabolic molecules *of order* (R, M, N_1, N_2) *if it can be written as*

$$m_\lambda(x) = 2^{3s_\lambda/4} a^{(\lambda)} \left(D_{2^{s_\lambda}} R_{\theta_\lambda} (x - x_\lambda) \right)$$

such that

$$\left| \partial^\beta \hat{a}^{(\lambda)}(\xi) \right| \lesssim \min\left(1, 2^{-s_\lambda} + |\xi_1| + 2^{-s_\lambda/2} |\xi_2| \right)^M \langle |\xi| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2} \tag{1}$$

for all $|\beta| \leq R$. *The implicit constants are uniform over* $\lambda \in \Lambda$.

Remark 2.3. *For convenience our definition only poses conditions on the Fourier transform of m_λ. The number R describes the spatial localization, M the number of directional (almost) vanishing moments and N_1, N_2 describe the smoothness of an element m_λ. We also refer to Figure 1 for an illustration of the approximate frequency support of a parabolic molecule.*

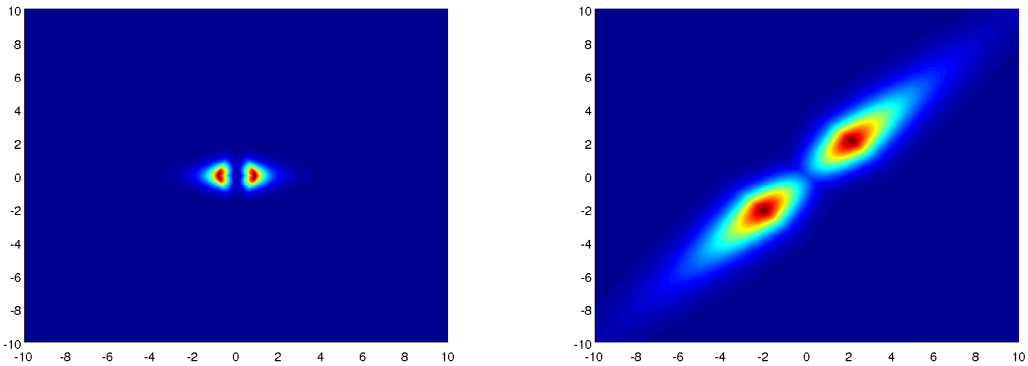

Figure 1: Left: The weight function $\min\left(1, 2^{-s_\lambda} + |\xi_1| + 2^{-s_\lambda/2}|\xi_2|\right)^M \langle|\xi|\rangle^{-N_1} \langle\xi_2\rangle^{-N_2}$ for $s_\lambda = 3$, $M = 3$, $N_1 = N_2 = 2$. Right: Approximate Frequency support of a corresponding molecule $\hat{m}_\lambda$ with $\theta_\lambda = \pi/4$.

We pause to record the following simple estimates: In polar coordinates we have the representation

$$\hat{m}_\lambda(r, \varphi) = 2^{-3s_\lambda/4} \hat{a}^{(\lambda)} \left(2^{-s_\lambda} r \cos(\varphi + \theta_\lambda), 2^{-s_\lambda/2} r \sin(\varphi + \theta_\lambda)\right) \exp\left(2\pi i \langle x_\lambda, \xi\rangle\right).$$

Equation (1) directly implies that in polar coordinates we have the estimate

$$|\hat{m}_\lambda(\xi)| \lesssim 2^{-2s_\lambda/4} \min\left(1, 2^{-s_\lambda}(1+r)\right)^M \left\langle 2^{-s_\lambda} r\right\rangle^{-N_1} \langle 2^{-s_\lambda/2} r \sin(\varphi + \theta_\lambda)\rangle^{-N_2}. \tag{2}$$

2.2 Metric Properties of Parametrizations

In order to proceed we need to introduce some additional (metric) properties of index sets and parametrizations. The parameter space $\mathbb{P}$ can be equipped with a natural notion of (pseudo) distance, see [29], which can be extended to a distance between indices by a pullback via a parametrization.

Definition 2.4. *Following [5, 29], we define for two indices λ, μ the* index distance

$$\omega(\lambda, \mu) := 2^{|s_\lambda - s_\mu|} \left(1 + 2^{s_{\lambda_0}} d(\lambda, \mu)\right),$$

and

$$d(\lambda, \mu) := |\theta_\lambda - \theta_\mu|^2 + |x_\lambda - x_\mu|^2 + |\langle e_\lambda, x_\lambda - x_\mu\rangle|.$$

where $\lambda_0 = argmin(s_\lambda, s_\mu)$ and $e_\lambda = (\cos(\theta_\lambda), \sin(\theta_\lambda))^\top$.

Remark 2.5. *The notation $\omega(\lambda, \mu)$ is actually a slight abuse of notation since ω is acting on $\mathbb{P}$. Therefore it should read*

$$\omega\left(\Phi_\Lambda(\lambda), \Phi_M(\mu)\right)$$

for indices $\lambda \in \Lambda$, $\mu \in M$ with associated parametrizations Φ_Λ, Φ_M. In order not to overload the notation we stick with the shorter but slightly less accurate definition.

Remark 2.6. *We wish to mention that, in fact, real-valued curvelets or shearlets are not associated with an angle but with a ray, i.e., θ and $\theta + \pi$ need to be identified. This is not reflected in the above definition, which is a slight inaccuracy. The 'correct' definition should assume that $|\theta_\lambda| \leq \frac{\pi}{2} \in \mathbb{P}^1$, the projective line. Therefore, it should read*

$$d(\lambda, \mu) := |\{\theta_\lambda - \theta_\mu\}|^2 + |x_\lambda - x_\mu|^2 + |\langle\theta_\lambda, x_\lambda - x_\mu\rangle|$$

with $\{\varphi\}$ the projection of φ onto $\mathbb{P}^1 \cong (-\pi/2, \pi/2]$.

However, for our results it will make no difference which definition is used. Thus we decided to employ Definition 2.4, which avoids additional technicalities.

We need to impose further conditions on an index set Λ in order to arrive at meaningful results. The following definition formalizes a crucial property, which is later on required to be satisfied by an index set in our results.

Definition 2.7. *An index set Λ with associated mapping Φ_Λ is called k-*admissible *if*

$$\sup_{\lambda\in\Lambda} \sum_{\mu\in\Lambda^0} \omega(\lambda,\mu)^{-k} < \infty,$$

and

$$\sup_{\lambda\in\Lambda^0} \sum_{\mu\in\Lambda} \omega(\lambda,\mu)^{-k} < \infty.$$

Lemma 2.8. *The canonical index set Λ^0 is k-admissible for all $k > 2$.*

Proof. We aim to prove that

$$\sup_{\mu\in\Lambda^0} \sum_{\lambda\in\Lambda^0} \omega(\mu,\lambda)^{-k} < \infty. \tag{3}$$

Writing $s_\mu = j'$ in the definition of $\omega(\mu,\lambda)$, we need to consider

$$\sum_{j\in\mathbb{Z}_+} \sum_{\lambda\in\Lambda^0, s_\lambda = j} 2^{-k|j-j'|} \left(1 + 2^{\min(j,j')} d(\mu,\lambda)\right)^{-k}. \tag{4}$$

According to [5, Equation (A.2)], we have

$$\sum_{\lambda\in\Lambda^0, s_\lambda=j} (1 + 2^q d(\mu,\lambda))^{-2} \lesssim 2^{2(j-q)_+} \tag{5}$$

for any q. Hence, for each $k > 2$, (4) can be estimated by

$$\sum_{j\geq 0} 2^{-k|j-j'|} 2^{2|j-j'|} < \infty,$$

which proves (3). □

2.3 Main Result

The main result of this paper essentially states that any two systems of parabolic molecules behave in the same way as far as approximation properties are concerned. Specifically, we show the following theorem, whose proof is quite technical, wherefore we postpone it to Subsection 5.2.

Theorem 2.9. *Let $(m_\lambda)_{\lambda\in\Lambda}$, $(p_\mu)_{\mu\in M}$ be two systems of parabolic molecules of order (R, M, N_1, N_2) with*

$$R \geq 2N, \quad M > 4N - \frac{5}{4}, \quad N_1 \geq 2N + \frac{3}{4}, \quad N_2 \geq 2N. \tag{6}$$

Then

$$|\langle m_\lambda, p_\mu\rangle| \lesssim \omega\left((s_\lambda, \theta_\lambda, x_\lambda), (s_\mu, \theta_\mu, x_\mu)\right)^{-N}.$$

This result shows that the Gramian matrix between any two systems of parabolic molecules satisfies a strong off-diagonal decay property and is in that sense very close to a diagonal matrix. As we shall see in Section 4, this result has a number of immediate applications, most notably for the approximation properties of arbitrary frames which are systems of parabolic molecules (they turn out to be equivalent!).

We find it particularly striking that our framework is general enough to include both curvelet-type, as well as shearlet-type constructions (see Section 3). Therefore, as a consequence of Theorem 2.9, all these systems satisfy the same celebrated properties of the curvelet construction given in [8]. To demonstrate the importance of our result, Section 4 discusses selected applications of Theorem 2.9 such as sparsity equivalence and equivalence of associated smoothness spaces.

3 Examples of Parabolic Molecules

Having defined parabolic molecules in Section 2 above, it is important to examine the versatility of this concept. This is done in the present section. The main findings are that essentially all known constructions in the literature can be cast in our framework and are thus amenable to the techniques and results developed in this paper.

We divide the section into two subsections. In Subsection 3.1 we study so-called curvelet-like constructions. These include curvelets as defined in [4] but also other constructions, such as in [3, 29]. We show that all these function systems are parabolic molecules. In fact, this result should not come to much as a surprise: In [5] a similar concept of curvelet molecules is introduced which includes all the above-mentioned constructions. We also show that curvelet molecules are parabolic molecules.

The real strength of our definition of parabolic molecules is that it includes not only curvelet-type constructions. In fact, we consider it one of the main findings of this paper that also shearlet-type systems can be thought of as instances of parabolic molecules, associated to a specific shearlet parametrization Φ^σ! We show this result, as well as the admissibility of Φ^σ, below in Subsection 3.2. After that, to provide some concrete examples, we study several specific constructions. In particular, we show that compactly supported shearlet constructions (see e.g. [24]) are parabolic molecules.

3.1 Curvelet-like constructions

3.1.1 Second Generation Curvelets

It is easily verified that curvelet molecules as defined in [5] are instances of parabolic molecules associated with the canonical parametrization. In particular, second generation curvelets [4] are parabolic molecules of arbitrary order. We start by describing the construction. Pick two window functions $W(r)$, $V(t)$ which are both real, nonnegative, C^∞ and supported in $\left(\frac{1}{2},2\right)$ and in $(-1,1)$ respectively. We further assume that these windows satisfy

$$\sum_{j\in\mathbb{Z}} W\left(2^j r\right)^2 = 1 \quad \text{for all } r\in\mathbb{R}_+ \quad \text{and} \quad \sum_{l\in\mathbb{Z}} V\left(t-l\right)^2 = 1 \quad \text{for all } t\in\left(-\frac{1}{2},\frac{1}{2}\right).$$

Now define in polar coordinates

$$\hat{\gamma}_{(j,0,0)}(r,\omega) := 2^{-3j/4} W\left(2^{-j}r\right) V\left(2^{\lfloor j/2\rfloor}\omega\right) \text{ and } \gamma_{(j,l,k)}(\cdot) := \gamma_{(j,0,0)}\left(R_{\theta_{(j,l,k)}}\left(\cdot - x_{(j,l,k)}\right)\right),$$

where $(j,l,k)\in\Lambda^0$. With appropriate modifications for the low-frequency case $j=0$ it is possible to show that the system

$$\Gamma^0 := \left\{\gamma_\lambda : \lambda\in\Lambda^0\right\}$$

constitutes a Parseval frame for $L_2(\mathbb{R}^2)$. In order to make the frame elements real-valued, it is possible to identify elements oriented in antipodal directions. This frame is customarily referred to as the tight frame of *second generation curvelets.* We now show that this frame forms a system of parabolic molecules of arbitrary order.

Proposition 3.1. *The second generation curvelet frame constitutes a system of parabolic molecules of arbitrary order associated with the canonical parametrization.*

Proof. Due to rotation invariance we may restrict ourselves to the case $\theta_\lambda = 0$. Therefore, denoting $\gamma_j := \gamma_{(j,0,0)}$, we need to show that the function

$$a^{(\lambda)}(\cdot) := 2^{-3s_\lambda/4}\gamma_j\left(D_{2^{-s_\lambda}}\cdot\right)$$

satisfies (1) for (R,M,N_1,N_2) arbitrary. First note that

$$\hat{a}^{(\lambda)}(\cdot) = 2^{3s_\lambda/4}\hat{\gamma}_j\left(D_{2^{s_\lambda}}\cdot\right).$$

The function $\hat{a}^{(\lambda)}$, together with all its derivatives has compact support in a rectangle away from the ξ_1-axis. Therefore, it only remains to show that, on its support, the function $\hat{a}^{(\lambda)}$ has bounded derivatives, with a bound independent of j. But this follows from elementary arguments, using $r = \sqrt{\xi_1^2 + \xi_2^2}$, $\omega = \arctan(\xi_2/\xi_1)$. Then

$$\hat{a}^{(\lambda)}(\xi) = \hat{\gamma}_{(j,0,0)}(D_{2^j}\xi) = W(\alpha_j(\xi))\, V(\beta_j(\xi)),$$

where

$$\alpha_j(\xi) := 2^{-j}\sqrt{2^{2j}\xi_1^2 + 2^j\xi_2^2}, \text{ and } \beta_j(\xi) := 2^{j/2}\arctan\left(\frac{\xi_2}{2^{j/2}\xi_1}\right).$$

Now it is a simple calculus exercise to show that all derivatives of α_j and β_j are bounded on the support of $\hat{a}^{(\lambda)}$ and uniformly in j. This proves the result. □

3.1.2 Hart Smith's Parabolic Frame

Historically, the first instance of a decomposition into parabolic molecules can be found in Hart Smith's work on Fourier Integral Operators and Wave Equations [29]. This frame, as well as its dual, again forms a system of parabolic molecules of arbitrary order associated with the canonical parametrization. We refer to [29, 1] for the details of the construction which is essentially identical to the curvelet construction, with primal and dual frame being allowed to differ. The same discussion as above for curvelets shows that this system consists of parabolic molecules.

3.1.3 Borup and Nielsen's Construction

Another very similar construction has been given in [3]. In this paper, the focus has been on the study of associated function spaces. Again, it is straightforward to prove that this system constitutes a system of parabolic molecules of arbitrary order associated with the canonical parametrization. As a corollary to our results, it will actually turn out that the spaces defined in [3] coincide with the approximation spaces corresponding to curvelets, shearlets, and Smith's transform.

3.1.4 Curvelet Molecules

In [5] the authors introduced the notion of *curvelet molecules* which are a useful concept in proving sparsity properties of wave propagators. For the sake of completion, we include the exact definition.

Definition 3.2. *Let Λ^0 be the canonical parametrization. A family $(m_\lambda)_{\lambda\in\Lambda^0}$ is called a* family of curvelet molecules *of regularity R if it can be written as*

$$m_\lambda(x) = 2^{3s_\lambda/4} a^{(\lambda)}\left(D_{2^{s_\lambda}} R_{\theta_\lambda}(x - x_\lambda)\right)$$

such that for all $|\beta| \le R$ and each $N = 0, 1, 2, \dots$

$$|\partial^\beta a^{(\lambda)}(x)| \lesssim \langle x\rangle^{-N} \tag{7}$$

and for $M = 0, 1, \dots$

$$|\hat{a}^{(\lambda)}(\xi)| \lesssim \min\left(1, 2^{-s_\lambda} + |\xi_1| + 2^{-s_\lambda/2}|\xi_2|\right)^M. \tag{8}$$

This definition is similar to our definition of parabolic molecules, however with two crucial differences: First, (1) allows for arbitrary rotation angles and is therefore more general. Curvelet molecules on the other hand are only defined for the canonical parametrization Λ^0 (which, in contrast to our definition, is not sufficiently general to also cover shearlet-type systems). Second, the decay conditions analogous to our condition (1) are more restrictive in the sense that it requires infinitely many nearly vanishing moments. In fact, the following result holds:

Proposition 3.3. *A system of curvelet molecules of regularity R constitutes a system of parabolic molecules of order $(\infty, \infty, R/2, R/2)$.*

Proof. The definition of curvelet molecules as above implies that the estimate (8) also holds for all derivatives of $\hat{a}^{(\lambda)}$, see [5]. Furthermore, by (7), all derivatives of $\hat{a}^{(\lambda)}$ can be estimated in modulus by $\langle|\xi|\rangle^{-R}$, which in turn can be estimated by $\langle|\xi|\rangle^{-R/2}\langle\xi_2\rangle^{-R/2}$. This yields the desired estimate. □

3.2 Shearlets

Shearlets were introduced in 2006 as the first directional representation system which not only satisfies the same celebrated properties of curvelets, but also provides a unified treatment of the continuum and digital setting. This key property is achieved through utilization of a shearing matrix as a means to parameterize orientation, which is highly adapted to the digital grid in contrast to rotation. For more information on shearlets, we refer to the book [25].

It is perhaps not surprising that curvelets and their relatives described above fall into the framework of parabolic molecules. Here we show the crucial fact that shearlets can be seen as a special case of parabolic molecules as well. Consider the discrete index set

$$\Lambda^\sigma := \left\{(\varepsilon, j, l, k) \in \mathbb{Z}_2 \times \mathbb{Z}^4 : \varepsilon \in \{0,1\},\ j \geq 0,\ l = -2^{\lfloor \frac{j}{2} \rfloor}, \cdots, 2^{\lfloor \frac{j}{2} \rfloor}\right\}, \tag{9}$$

and the shearlet system

$$\Sigma := \{\sigma_\lambda : \lambda \in \Lambda^\sigma\},$$

with

$$\sigma_{(\varepsilon,0,0,k)}(\cdot) = \varphi(\cdot - k), \quad \sigma_{(\varepsilon,j,l,k)}(\cdot) = 2^{3j/4}\psi^\varepsilon_{j,l,k}\left(D^\varepsilon_{2^j} S^\varepsilon_{l,j} \cdot - k\right), \quad j \geq 1,$$

where $D^0_a = D_a$, $D^1_a := \operatorname{diag}(\sqrt{a}, a)$, $S_{l,j} := \begin{pmatrix} 1 & l2^{-\lfloor j/2 \rfloor} \\ 0 & 1 \end{pmatrix}$ and $S^1_{l,j} = \left(S^0_{l,j}\right)^\top$. Then we define shearlet molecules of order (R, M, N_1, N_2), which is a generalization of shearlets adapted to parabolic molecules, in particular including the classical shearlet molecules introduced in [21], see Subsection 3.2.5.

Definition 3.4. *We call Σ a system of* shearlet molecules *of order (R, M, N_1, N_2) if the functions φ, $\psi^0_{j,l,k}$, $\psi^1_{j,l,k}$ satisfy*

$$|\partial^\beta \hat{\psi}^\varepsilon_{j,l,k}(\xi_1, \xi_2)| \lesssim \min\left(1, |\xi_{1+\varepsilon}|\right)^M \langle |\xi| \rangle^{-N_1} \langle \xi_{2-\varepsilon} \rangle^{-N_2} \tag{10}$$

for every $\beta \in \mathbb{N}^2$ with $|\beta| \leq R$.

Remark 3.5. *In our proofs it is nowhere required that the directional parameter l runs between $-2^{\lfloor \frac{j}{2} \rfloor}$ and $-2^{\lfloor \frac{j}{2} \rfloor}$ Indeed, l running in any discrete interval $-C2^{\lfloor \frac{j}{2} \rfloor}, \ldots, C2^{\lfloor \frac{j}{2} \rfloor}$ would yield the exact same results, as a careful inspection of our arguments shows. Likewise, in certain shearlet constructions, the translational sampling runs not through $k \in \mathbb{Z}^2$ but through $\tau\mathbb{Z}^2$ with $\tau > 0$ a sampling constant. Our results are also valid for this case with the similar proofs. The same remark applies to all curvelet-type constructions.*

Now we can show the main result of this section, namely that shearlet systems with generators satisfying (10) are actually instances of parabolic molecules associated with a specific shearlet-adapted parametrization Φ_σ.

Proposition 3.6. *Assume that the shearlet system Σ constitutes a system of shearlet molecules of order (R, M, N_1, N_2). Then Σ constitutes a system of parabolic molecules of order (R, M, N_1, N_2), associated to the parametrization $(\Lambda^\sigma, \Phi^\sigma)$, where*

$$\Phi^\sigma(\lambda) = (s_\lambda, \theta_\lambda, x_\lambda) := \left(j, \varepsilon\pi/2 + \arctan(-l2^{-\lfloor j/2 \rfloor}), \left(S^\varepsilon_l\right)^{-1} D^\varepsilon_{2^{-j}} k\right).$$

Proof. We confine the discussion to $\varepsilon = 0$, the other case being the same. Further, we will suppress the superscript ε as well as the subscript j, l, k in our notation. We need to show that

$$a^{(\lambda)}(\cdot) := \psi\left(D_{2^{s_\lambda}} S_{l,s_\lambda} R^\top_{\theta_\lambda} D_{2^{-s_\lambda}} \cdot\right)$$

satisfies (1). The Fourier transform of $a^{(\lambda)}$ is given by

$$\hat{a}^{(\lambda)}(\cdot) = \hat{\psi}\left(D_{2^{-s_\lambda}} S^{-\top}_{l,s_\lambda} R^\top_{\theta_\lambda} D_{2^{s_\lambda}} \cdot\right).$$

The matrix $S^{-\top}_{l,s_\lambda} R^\top_{\theta_\lambda}$ has the form

$$S^{-\top}_{l,s_\lambda} R^\top_{\theta_\lambda} = \begin{pmatrix} \cos(\theta_\lambda) & \sin(\theta_\lambda) \\ 0 & -l2^{-\lfloor s_\lambda/2 \rfloor}\sin(\theta_\lambda) + \cos(\theta_\lambda) \end{pmatrix} =: \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}.$$

We claim that the quantities a and c are uniformly bounded from above and below, independent of j, l. To see this, consider the functions

$$\tau(x) := \cos(\arctan(x)) \quad \text{and} \quad \rho(x) := x\sin(\arctan(x)) + \cos(\arctan(x)),$$

which are bounded from above and below on $[-1, 1]$, as an elementary discussion shows (in fact this boundedness holds on any compact interval). Clearly, we have

$$a = \tau\left(-l2^{\lfloor \frac{s_\lambda}{2} \rfloor}\right) \quad \text{and} \quad c = \rho\left(-l2^{\lfloor \frac{s_\lambda}{2} \rfloor}\right)$$

Since we are only considering indices with $\varepsilon = 0$, we have $\left|-l2^{\lfloor \frac{s_\lambda}{2} \rfloor}\right| \leq 1$, which, by the above implies uniform upper and lower boundedness of the quantities a, c, i.e., there exist numbers $0 < \delta_a \leq \Delta_a < \infty$, $0 < \delta_c \leq \Delta_c < \infty$ such that for all j, l we have

$$\delta_a \leq a \leq \Delta_a \quad \text{and} \quad \delta_c \leq c \leq \Delta_c.$$

The matrix $D_{2^{-s_\lambda}} R_{\theta_\lambda}^\top S_{l,s_\lambda}^{-\top} D_{2^{s_\lambda}}$ has the form

$$\begin{pmatrix} a & 2^{-s_\lambda/2} b \\ 0 & c \end{pmatrix}.$$

Using the upper boundedness of a, b, c and the chain rule, we can estimate for any $|\beta| \leq R$:

$$|\partial^\beta \hat{a}^{(\lambda)}(\xi)| \lesssim \sup_{|\gamma| \leq R} \left| \partial^\gamma \hat{\psi}\left(\begin{pmatrix} a & 2^{-s_\lambda/2} b \\ 0 & c \end{pmatrix} \xi\right)\right| \lesssim \left(|\xi_1| + 2^{-s_\lambda/2}|\xi_2|\right)^M.$$

For the last estimate we utilized the moment estimate for $\hat{\psi}$, which is given by (10). This gives us the moment property required in (1).

Now we need to show the decay of $\partial^\beta \hat{a}^{(\lambda)}$ for large frequencies ξ. Again, due to the fact that a, b, c are bounded from above and a, c from below, and utilizing the large frequency decay estimate in (10), we can estimate

$$\begin{aligned} |\partial^\beta \hat{a}^{(\lambda)}(\xi)| &\lesssim \sup_{|\gamma| \leq R} \left| \partial^\gamma \hat{\psi}\left(\begin{pmatrix} a & 2^{-s_\lambda/2} b \\ 0 & c \end{pmatrix} \xi\right)\right| \lesssim \left\langle \left| \begin{pmatrix} a & 2^{-s_\lambda/2} b \\ 0 & c \end{pmatrix} \xi \right| \right\rangle^{-N_1} \langle c\xi_2 \rangle^{-N_2} \\ &\lesssim \langle |\xi| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2}. \end{aligned}$$

The statement is proven. □

The following result shows that, just like the canonical parametrization, the shearlet parametrization Λ^σ is admissible.

Lemma 3.7. *The shearlet parametrization $(\Lambda^\sigma, \Phi^\sigma)$ is k-admissible for $k > 2$.*

Proof. We show the analogue to Equation (5) for the shearlet parametrization, the rest of the proof is analogous to the proof of Lemma 2.8. Hence, we aim to prove that

$$\sum_{\lambda \in \Lambda^\sigma, s_\lambda = j} (1 + 2^q d(\mu, \lambda))^{-2} \lesssim 2^{2(j-q)_+} \tag{11}$$

for any q and $\mu \in \Lambda^0$. Without loss of generality we assume that $\theta_\mu = 0$, $x_\mu = 0$ (the general case follows identical arguments with slightly more notational effort). Further, as before we only restrict ourselves to the case $\varepsilon = 0$, the other case being exactly the same.

First we consider the case $q > j$. In this situation, the expression in (11) can be bounded by a uniform constant.

Now we turn to the other case $j \geq q$. In this case we use the fact that, whenever $|l| \lesssim 2^{-j/2}$, we have

$$\left|\arctan\left(-l2^{-\lfloor \frac{j}{2} \rfloor}\right)\right| \gtrsim \left|l2^{-\lfloor \frac{j}{2} \rfloor}\right| \quad \text{and} \quad |S_l^{-1} D_{2^{-j}} k| \gtrsim |D_{2^{-j}} k|,$$

to estimate (11) by

$$\sum_{l}\sum_{k}\left(1+2^{q}\left(\left|l2^{-\lfloor\frac{j}{2}\rfloor}\right|^{2}+\left|2^{-\lfloor\frac{j}{2}\rfloor}k_2\right|^{2}+\left|2^{-j}k_1-l2^{-\lfloor\frac{j}{2}\rfloor}k_2 2^{-\lfloor\frac{j}{2}\rfloor}\right|\right)\right)^{-2}.$$

This can be interpreted as a Riemann sum and bounded (up to a constant) by the corresponding integral

$$\int_{\mathbb{R}^2}\frac{dx}{2^{-3j/2}}\int_{\mathbb{R}}\frac{dy}{2^{-j/2}}\left(1+2^{q}(y^2+x_2^2+|x_1-x_2y|)\right)^{-2},$$

compare [5, Equation (A.3)]. This integral is bounded by a constant times $2^{2(j-q)}$ as can be seen by making the substitution $x_1\to 2^q x_1$, $x_2\to 2^{q/2}x_2$, $y\to 2^{q/2}y$. This shows (11) and thus completes the proof. □

These results show that the parabolic molecule concept is a unification of previous systems. In the remainder of this section we examine the shearlet constructions which are on the market and show that they indeed fit into our framework.

3.2.1 Bandlimited Shearlets

We start with the classical shearlet construction which yields bandlimited generators. We consider two functions $\psi_1,\ \psi_2$ satisfying

$$\operatorname{supp}\hat{\psi}_1\subset\left[-\frac{1}{2},-\frac{1}{16}\right]\cup\left[\frac{1}{16},\frac{1}{2}\right],\quad \operatorname{supp}\hat{\psi}_2\subset[-1,1],$$

$$\sum_{j\geq 0}\left|\hat{\psi}_1\left(2^{-j}\omega\right)\right|^2=1\quad\text{for } |\omega|\geq\frac{1}{8},$$

and

$$\sum_{l=-2^{\lfloor j/2\rfloor}}^{2^{\lfloor j/2\rfloor}}\left|\hat{\psi}_2\left(2^{\lfloor j/2\rfloor}\omega+l\right)\right|^2=1\quad\text{for } |\omega|\leq 1.$$

Now we define our basic shearlet ψ^0 via

$$\hat{\psi}^0(\xi):=\hat{\psi}_1(\xi_1)\hat{\psi}_2\left(\frac{\xi_2}{\xi_1}\right).$$

It follows from standard arguments that the system

$$\Sigma^0:=\left\{2^{3j/4}\psi^0\left(D^0_{2^j}S^0_{l,j}\cdot-k\right):\ j\geq 0,\ l=-2^{\lfloor\frac{j}{2}\rfloor},\cdots,2^{\lfloor\frac{j}{2}\rfloor}\right\}$$

constitutes a Parseval frame for the Hilbert space $L_2\left(\mathcal{C}\right)^{\vee}$ with

$$\mathcal{C}:=\left\{\xi:\ |\xi_1|\geq\frac{1}{8},\ \frac{|\xi_2|}{|\xi_1|}\leq 1\right\}.$$

In the same way we can construct a Parseval frame Σ^1 for $L_2\left(\mathcal{C}'\right)^{\vee}$,

$$\mathcal{C}':=\left\{\xi:\ |\xi_2|\geq\frac{1}{8},\ \frac{|\xi_1|}{|\xi_2|}\leq 1\right\}.$$

by reversing the coordinate axes. Finally, we can consider a Parseval frame

$$\Phi:=\left\{\varphi(\cdot-k):\ k\in\mathbb{Z}^2\right\}$$

for the Hilbert space $L_2\left(\left[-\frac{1}{8},\frac{1}{8}\right]^2\right)^{\vee}$.

Proposition 3.8. *The system $\Sigma:=\Sigma^0\cup\Sigma^1\cup\Phi$ constitutes a shearlet frame which is a system of parabolic molecules of arbitrary order.*

Proof. To show this, by Proposition 3.6, all we need to show is that the generators $\psi^0,\ \psi^1$ satisfy (10) for arbitrary orders (R,M,N_1,N_2). This, however, follows directly from the fact that the underlying basis functions are bandlimited. □

3.2.2 Bandlimited Shearlets with Nice Duals

The bandlimited shearlet frame Σ as described above suffers from the fact that we do not know much about its dual frames. In particular, we do not know whether there exists a dual frame which is also a system of parabolic molecules. For several results such as those in Subsection 4.2 it is however necessary to have such a construction. In [18] this problem was successfully resolved by carefully glueing together the two bandlimited frames associated with the two frequency cones. In other words, there exist shearlet frames Σ with dual frame Σ' such that both Σ and Σ' form systems of parabolic molecules of arbitrary order.

3.2.3 Smooth Parseval Frames of Shearlets

In [22] another modification of the bandlimited shearlet construction is given by carefully glueing together two boundary elements along the seamlines with angle $\pi/4$. It can be shown that this yields a Parseval frame with smooth and well-localized elements. Again, it is straightforward to check that the system constructed in [22] constitutes a system of parabolic molecules of arbitrary order.

3.2.4 Compactly Supported Shearlets

We next analyze compactly supported shearlets [24], and prove that they also form instances of parabolic molecules. Currently known constructions of compactly supported shearlets involve separable generators, i.e.,

$$\psi^0(x_1, x_2) := \psi_1(x_1)\psi_2(x_2), \quad \psi^1(x_1, x_2) := \psi^0(x_2, x_1). \tag{12}$$

with a wavelet ψ_1 and a scaling function ψ_2. We would like to find conditions on ψ_1, ψ_2 such that (10) is satisfied for given parameters (R, M, N_1, N_2), i.e., that the associated shearlet frame forms a system of shearlet molecules.

First we define the crucial property of vanishing moments for univariate wavelets.

Definition 3.9. *A univariate function g possesses M vanishing moments if*

$$\int_{\mathbb{R}} g(x)x^k dx = 0, \quad \text{for all } k = 0, \dots, M-1.$$

In the frequency domain, vanishing moments are characterized by polynomial decay near zero, as is well known.

Lemma 3.10. *Suppose that $g : \mathbb{R} \to \mathbb{C}$ is continuous, compactly supported and possesses M vanishing moments. Then*

$$|\hat{g}(\xi)| \lesssim \min(1, |\xi|)^M .$$

Proof. First, note that, since g is continuous and compactly supported, it is in $L_1(\mathbb{R})$ and therefore its Fourier transform is bounded. This takes care of frequencies ξ with $|\xi| \geq 1$. For small ξ observe that, up to a constant we have

$$\int_{\mathbb{R}} g(x)x^k dx = \left(\frac{d}{d\xi}\right)^k \hat{g}(0).$$

Hence, if g possesses M vanishing moments, all derivatives of order $< M$ of the Fourier transform $\hat{g}$ vanish at 0. Furthermore, since g is compactly supported, its Fourier transform is analytic. Therefore

$$|\hat{g}(\xi)| \lesssim |\xi|^M,$$

which proves the claim. □

Proposition 3.11. *Assume that $\psi_1 \in C^{N_1}$ is a compactly supported wavelet with $M + R$ vanishing moments, and $\psi_2 \in C^{N_1+N_2}$ is also compactly supported. Then, with ψ^ε defined by (12), the associated shearlet system Σ constitutes a system of parabolic molecules of order (R, M, N_1, N_2).*

Proof. In view of Proposition 3.6 we need to show that the estimate (10) holds. We only consider the case $\varepsilon = 0$ and drop the superscript. The inverse Fourier transform of $\partial^\beta \psi$ is, up to a constant given by $x^\beta \psi(x)$. We first handle the case $|\xi_1| > 1$. By smoothness and compact support of ψ_1, ψ_2, we find that for any $|\beta| \leq R$ the function

$$\partial^{(N_1, N_1+N_2)} x^\beta \psi$$

is in $L_1(\mathbb{R})$, hence it has a bounded Fourier transform which is given, up to a constant by

$$\xi_1^{N_1} \xi_2^{N_1+N_2} \partial^\beta \hat{\psi}(\xi).$$

It follows that the function

$$\langle \xi_1 \rangle^{N_1} \langle \xi_2 \rangle^{N_1+N_2} \partial^\beta \hat{\psi}(\xi)$$

is bounded. Using the simple fact that $\langle x \rangle \langle y \rangle \lesssim \langle \sqrt{x^2+y^2} \rangle$, we get

$$\partial^\beta \psi(\xi) \lesssim \langle \|\xi\| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2}.$$

Now let β be such that $|\beta_1| < R$. Then the function

$$x^\beta \psi(x) = x_1^{\beta_1} x_2^{\beta_2} \psi_1(x_1) \psi_2(x_2),$$

restricted to the variable x_1 possesses at least M vanishing moments, due to the assumption that ψ_1 possesses $M + R$ vanishing moments. Lemma 3.10 then proves the decay of order $\min\left(1, |\xi_1|^M\right)$ for the derivatives of $\hat{\psi}$. □

Remark 3.12. *Several assumptions on the generators ψ^1, ψ^2 could be weakened, for instance the separability of the shearlet generators is not crucial for the arguments to go through. In particular, our arguments nowhere require neither compact support nor bandlimitedness.*

3.2.5 Shearlet Molecules of [21]

In [21] the results of [5] are established for shearlets instead of curvelets. A crucial tool in the proof is the introduction of a certain type of *shearlet molecules* which are similar to curvelet molecules discussed above, but tailored to the shearing operation rather than rotations.

Definition 3.13. *Let Λ^σ be the shearlet index set as in (9). A family $(m_\lambda)_{\lambda \in \Lambda^\sigma}$ is called a* family of shearlet molecules *of regularity R if it can be written as*

$$m_\lambda(x) = 2^{3s_\lambda/4} a^{(\lambda)} \left(D^\varepsilon_{2^{s_\lambda}} S^\varepsilon_{l,j} x - k \right)$$

such that for all $|\beta| \leq R$ and each $N = 0, 1, 2, \ldots$

$$|\partial^\beta a^{(\lambda)}(x)| \lesssim \langle x \rangle^{-N}$$

and for $M = 0, 1, \ldots$

$$|\hat{a}^{(\lambda)}(\xi)| \lesssim \min\left(1, 2^{-s_\lambda} + |\xi_1| + 2^{-s_\lambda/2}|\xi_2|\right)^M.$$

By the results in [21], the shearlet molecules defined therein satisfy the inequality (10) with the choice of parameters $(R, N, N_1, N_2) = (\infty, \infty, R/2, R/2)$. Therefore, in view of Proposition 3.6, shearlet molecules of regularity R as defined in [21] form systems of parabolic molecules of order $(\infty, \infty, R/2, R/2)$. Thus, we derive an analogous result to Lemma 3.3 for shearlet molecules:

Proposition 3.14. *A system of shearlet molecules of regularity R constitutes a system of parabolic molecules of order $(\infty, \infty, R/2, R/2)$.*

Let us finish this section on examples of parabolic molecules by making the following

Remark 3.15. *By now we hope to have convinced the reader that there is a whole zoo of different constructions in the literature which can all be put under one roof using the concept of parabolic molecules.*

4 Applications

In this section we discuss selected applications of the developed theory. A particular focus will be on approximation properties of parabolic molecule systems $(m_\lambda)_{\lambda\in\Lambda}$, especially if they form a frame, e.g.,

$$\|f\|_{L_2(\mathbb{R}^2)} \sim \sum_{\lambda\in\Lambda} |\langle f, m_\lambda\rangle|^2,$$

see e.g., [9]. It is well known, that in this case one can robustly represent any function $f \in L_2(\mathbb{R}^2)$ as a sum

$$f = \sum_{\lambda\in\Lambda} \langle f, m_\lambda\rangle \tilde{m}_\lambda,$$

where $(\tilde{m}_\lambda)_{\lambda\in\Lambda}$ is a *dual frame*. Approximation properties of a frame system $(m_\lambda)_{\lambda\in\Lambda}$ are usually studied in terms of the sparsity of the coefficient sequence $(\langle f, m_\lambda\rangle)_{\lambda\in\Lambda}$. Below, in Subsection 4.1 we show that essentially any frame system which consists of parabolic molecules satisfies the same approximation properties as the curvelet frame constructed in [4]. This, in particular, implies almost optimal approximation results for the class of cartoon images (see below for a definition) for all constructions mentioned in Section 3, for instance compactly supported or bandlimited shearlets. The above-mentioned approximation property may actually be regarded as the main raison d'être of curvelet-like systems and is therefore of central importance.

In Subsection 4.2 we go further and show that practically any reasonable definition of a function space norm based on a coefficient sequence $(\langle f, m_\lambda\rangle)_{\lambda\in\Lambda}$ is equivalent for any two frame systems consisting of parabolic molecules. This shows for instance that finiteness of a function space norm defined via the curvelet frame implies finiteness of the analogous norm defined via compactly supported shearlet frames, whenever the generators possess sufficient smoothness and directional vanishing moments. This result has not been known before.

Remark 4.1. *It is in general not the case that the dual frame $(\tilde{m}_\lambda)_{\lambda\in\Lambda}$ of a frame $(m_\lambda)_{\lambda\in\Lambda}$ of parabolic molecules needs to consist of parabolic molecules, too. However, it can be shown, based on the concept of* intrinsic localization*, that the so-called canonical dual frame of $(m_\lambda)_{\lambda\in\Lambda}$ is of a similar form in a certain sense [19].*

4.1 Sparse Image Approximation

Multivariate problems are typically governed by anisotropic features such as edges in images. A customarily employed model for such data is the class $\mathcal{E}^2(\mathbb{R}^2)$ of so-called *cartoon images* which 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 $B \subset [0,1]^2$ with ∂B a closed C^2-curve and $f_0, f_1 \in C_0^2([0,1]^2)$. Questions of efficient encoding of such a model class can be formulated in terms of optimal approximation properties. Given a frame system $(m_\lambda)_\lambda \subseteq L_2(\mathbb{R}^2)$, an appropriate measure for the approximation behavior is the decay rate of the *error of best N-term approximation*, i.e., of $\|f - f_N\|_2^2$, where f_N denotes the best N-term approximation by $(m_\lambda)_\lambda$ of some $f \in \mathcal{E}^2(\mathbb{R}^2)$, obtained as

$$f_N = \text{argmin}\|f - \sum_{\lambda\in\Lambda_N} c_\lambda m_\lambda\|_2^2 \quad \text{s.t.} \quad \#\Lambda_N \le N.$$

A small technical problem occurs due to the fact that the representation system might not form an orthonormal basis in which case the computation of the best N-term approximation is far from being understood. To circumvent this problem, usually the error of approximation by the N largest coefficients of $(\langle f, m_\lambda\rangle)_{\lambda\in\Lambda}$ is considered, which then certainly also provides a bound for the error of best N-term approximation. Typically, the asymptotics of this error are studied in terms of the ℓ_p-norms of the coefficient sequences of f for small values of p. Indeed, it is easily seen that membership of the coefficient sequence of f in an ℓ_p space for small p implies good N-term approximation rates, whenever the given representation system constitutes a frame, see, e.g., [27, 10].

In [14] it was shown that the optimally achievable decay rate of the error of approximation of some $f \in \mathcal{E}^2(\mathbb{R}^2)$ under the natural assumption of polynomial depth search is

$$\|f - f_N\|_2^2 \asymp N^{-2}, \quad \text{as } N \to \infty.$$

Furthermore, it was proven in [8] and in [20, 27] that both curvelets and shearlets attain this rate up to a log-factor. Apparently, these (parabolic) systems behave similarly concerning sparse approximation of anisotropic features.

The next definition provides a formalization of this concept by introducing the notion of sparsity equivalence. It is based on the connection between best N-term approximation rate and ℓ_p norms.

Definition 4.2. *Let $(m_\lambda)_{\lambda\in\Lambda}$ and $(p_\mu)_{\mu\in M}$ be systems of parabolic molecules of order (R, M, N_1, N_2) and $(\tilde{R}, \tilde{M}, \tilde{N}_1, \tilde{N}_2)$, respectively, and let $0 < p \leq 1$. Then $(m_\lambda)_{\lambda\in\Lambda}$ and $(p_\mu)_{\mu\in M}$ are* sparsity equivalent *in ℓ_p, if*

$$\left\| \left(\langle m_\lambda, p_\mu \rangle \right)_{\lambda\in\Lambda, \mu\in\Lambda^0} \right\|_{\ell_p \to \ell_p} < \infty.$$

Intuitively, systems of parabolic molecules being in the same sparsity equivalence class should have the same sparse approximation behavior with respect to cartoon images. The next result shows that this is indeed the case.

Proposition 4.3. *Let $(m_\lambda)_{\lambda\in\Lambda}$ and $(p_\mu)_{\mu\in M}$ be systems of parabolic molecules of order (R, M, N_1, N_2) and $(\tilde{R}, \tilde{M}, \tilde{N}_1, \tilde{N}_2)$, respectively, which are sparsity equivalent in $\ell_{2/3}$. If $(m_\lambda)_{\lambda\in\Lambda}$ possesses an almost best N-term approximation rate of order $N^{-1+\varepsilon}$ for cartoon images for any $\varepsilon > 0$, then so does $(p_\mu)_{\mu\in M}$.*

Proof. This is a direct consequence of the definition of sparsity equivalence and standard arguments, see for instance [10]. □

This result enables us to provide a very general class of systems of parabolic molecules which optimally sparsely approximate cartoon images by using the known result for curvelets. For this, we first analyze when a system is sparsity equivalent to the tight frame of bandlimited curvelets.

First, we state a simple result concerning operator norms on discrete ℓ_p spaces.

Lemma 4.4. *Let I, J be two discrete index sets, and let $\mathbf{A} : \ell_p(I) \to \ell_p(J)$, $p > 0$ be a linear mapping defined by its matrix representation $\mathbf{A} = (A_{i,j})_{i\in I,\, j\in J}$. Then we have the bound*

$$\|\mathbf{A}\|_{\ell_p(I)\to\ell_p(J)} \leq \max \left(\sup_i \sum_j |A_{i,j}|^r, \sup_j \sum_i |A_{i,j}|^r \right)^{1/r},$$

where $r := \min(1, p)$.

Proof. The proof for $p < 1$ follows easily using the fact that

$$|a + b|^p \leq |a|^p + |b|^p \quad \text{for } a, b \in \mathbb{R}.$$

To show the case $p \geq 1$ one only shows the assertion for $p = 1, \infty$, which is trivial. The claim then follows by interpolation. □

The next theorem proves the central fact that *any* system of parabolic molecules of sufficiently high order is sparsity equivalent to the bandlimited curvelet frame from Subsection 3.1.1.

Theorem 4.5. *Assume that $0 < p \leq 1$, (Λ, Φ_Λ) is a k-admissible parametrization, and $\Gamma^0 = (\gamma_\lambda)_{\lambda\in\Lambda^0}$ the tight frame of bandlimited curvelets. Further, assume that $(m_\lambda)_{\lambda\in\Lambda}$ is a system of molecules associated with Λ of order (R, M, N_1, N_2) such that*

$$R \geq 2\frac{k}{p}, \quad M > 4\frac{k}{p} - \frac{5}{4}, \quad N_1 \geq 2\frac{k}{p} + \frac{3}{4}, \quad N_2 \geq 2\frac{k}{p}.$$

Then $(m_\lambda)_{\lambda\in\Lambda}$ is sparsity equivalent to Γ^0.

Proof. We need to show that

$$\left\|(\langle m_\lambda,\gamma_\mu\rangle)_{\lambda\in\Lambda,\mu\in\Lambda^0}\right\|_{\ell_p\to\ell_p}=\max\left(\sup_{\mu\in\Lambda}\sum_{\lambda\in\Lambda^0}|\langle m_\lambda,\gamma_\mu\rangle|^p,\ \sup_{\lambda\in\Lambda^0}\sum_{\mu\in\Lambda}|\langle m_\lambda,\gamma_\mu\rangle|^p\right)^{1/p}<\infty.$$

By Theorem 2.9, we have

$$|\langle m_\lambda,\gamma_\mu\rangle|\lesssim\omega(\lambda,\mu)^{-\frac{k}{p}}.$$

It follows that

$$\begin{aligned}&\max\left(\sup_{\mu\in\Lambda}\sum_{\lambda\in\Lambda^0}|\langle m_\lambda,\gamma_\mu\rangle|^p,\ \sup_{\lambda\in\Lambda^0}\sum_{\mu\in\Lambda^0}|\langle m_\lambda,\gamma_\mu\rangle|^p\right)\\ \lesssim\ &\max\left(\sup_{\mu\in\Lambda}\sum_{\lambda\in\Lambda^0}\omega(\lambda,\mu)^{-k},\ \sup_{\lambda\in\Lambda^0}\sum_{\mu\in\Lambda^0}\omega(\lambda,\mu)^{-k}\right)<\infty,\end{aligned}$$

due to the k-admissibility of the parametrization of Λ. □

This result in combination with Proposition 4.3 now leads to the main result of this subsection.

Theorem 4.6. *Assume that $(m_\lambda)_{\lambda\in\Lambda}$ is a system of parabolic molecules of order (R,M,N_1,N_2) such that*

(i) $(m_\lambda)_{\lambda\in\Lambda}$ constitutes a frame for $L_2(\mathbb{R}^2)$,

(ii) Λ is k-admissible for all $k>2$,

(iii) it holds that

$$R\geq 6,\quad M>12-\frac{5}{4},\quad N_1\geq 6+\frac{3}{4},\quad N_2\geq 6.$$

Then the frame $(m_\lambda)_{\lambda\in\Lambda}$ possesses an almost best N-term approximation rate of order $N^{-1+\varepsilon}$, $\varepsilon>0$ arbitrary for cartoon images.

Proof. This follows from Proposition 4.3, Theorem 4.5, and the fact, proven in [8], that Γ^0 provides the respective N-term approximation rate. □

We remark that condition (ii) holds in particular for the shearlet parametrization. Hence this result allows a simple derivation of the results in [20, 27] from [8]. In fact, Theorem 4.6 provides a systematic way to, in particular, prove results on sparse approximation of cartoon images.

4.2 Function Spaces

Based on the concept of decomposition spaces introduced in [16], Borup and Nielsen have studied curvelet-like function spaces in [3]. We would like to apply our results to show that these spaces can be characterized by the transform coefficients in any frame which also forms a system of parabolic molecules. Consider the curvelet frame

$$\Gamma^0:=\left\{\gamma_{j,l,k}:(j,l,k)\in\Lambda^0\right\}$$

introduced in Subsection 3.1.1. Following [3] we define for $p,q,\alpha>1$ the function spaces $G^\alpha_{p,q}$ given by the norm

$$\|f\|_{G^\alpha_{p,q}}:=\left(\sum_{j\geq 0,l}\left(2^{\alpha j}\left(\sum_k|\langle f,\gamma_{j,l,k}\rangle|^p\right)^{1/p}\right)^q\right)^{1/q}. \tag{13}$$

This definition might seem somewhat odd, since the summation with respect to the directional parameter l is done with respect to the ℓ_q norm. For this reason and also for some minor technical reasons, we study another, similar family of function spaces, namely the spaces $S^\alpha_{p,q}$ given by the norm

$$\|f\|_{S^\alpha_{p,q}} := \left(\sum_{j\geq 0} \left(2^{\alpha j} \left(\sum_{k,l} |\langle f, \gamma_{j,l,k} \rangle|^p \right)^{1/p} \right)^q \right)^{1/q} . \tag{14}$$

Remark 4.7. *We would like to emphasize that all the results shown in this section also hold for the spaces defined by (13), but with slightly more technical effort arising from the need to handle mixed Lebesgue spaces [2]. We also remark that the function spaces defined via (14) can be interpreted as a decomposition spaces of the form studied in [3] with a mixed Lebesgue space $Y = \ell_q \ell_p$ (see [3] for more information).*

For technical reasons the definition in (14) forces us to work with a slightly stronger notion of admissibility than given in Definition 2.7:

Definition 4.8. *An index set Λ with associated mapping Φ_Λ is called* strongly (k,l)-admissible *if it is k-admissible and if*

$$\sum_{\lambda\in\Lambda_j} (1 + 2^q d(\mu,\lambda))^{-k} \lesssim 2^{l(j-q)_+},$$

where

$$\Lambda_j := \{\lambda \in \Lambda : \ s_\lambda = j\} .$$

Lemma 4.9. *The canonical parametrization (Λ^0, Φ^0) and the shearlet parametrization $(\Lambda^\sigma, \Phi^\sigma)$ are both strongly $(k,2)$-admissible for any $k > 2$.*

Proof. This has already been shown earlier in (5) for the canonical parametrization and in (11) for the shearlet parametrization. □

The aim of this section is to show the following theorem.

Theorem 4.10. *Let $\Sigma = \{\sigma_\lambda : \ \lambda \in \Lambda\}$ be a frame for $L_2(\mathbb{R}^2)$ with dual frame $\tilde{\Sigma} = \{\tilde{\sigma}_\lambda : \ \lambda \in \Lambda\}$. Assume further that Σ, $\tilde{\Sigma}$ are both parabolic molecules of arbitrary order with a strongly (k,l) admissible parametrization for some k,l. Then the following are equivalent norms on $S^\alpha_{p.q}$:*

$$\|f\|_{S^\alpha_{p,q}} \sim \left(\sum_{j\geq 0} \left(2^{\alpha j} \left(\sum_{\lambda\in\Lambda_j} |\langle f, \sigma_\lambda \rangle|^p \right)^{1/p} \right)^q \right)^{1/q} \sim \left(\sum_{j\geq 0} \left(2^{\alpha j} \left(\sum_{\lambda\in\Lambda_j} |\langle f, \tilde{\sigma}_\lambda \rangle|^p \right)^{1/p} \right)^q \right)^{1/q} .$$

Remark 4.11. *Of course it would be possible to show a quantitative version of Theorem 4.10 in the sense that Σ and $\tilde{\Sigma}$ are only required to form a system of parabolic molecules of finite, sufficiently large order, depending on p, q, α.*

Before we start with the proof of Theorem 4.10 we recall the following result which is a very useful inequality, sometimes called the *discrete Hardy inequality*, see [11]. To state this result we define for a sequence $\mathbf{a} = (a_k)_{k\in\mathbb{N}}$ the (quasi) norm

$$\|\mathbf{a}\|_{\ell^\alpha_q} := \left(\sum_{k\in\mathbb{N}} \left(2^{k\alpha}|a_k|\right)^q \right)^{1/q} .$$

The discrete Hardy inequalities are as follows.

Lemma 4.12. *Assume that with $\lambda > \alpha$ and $r \leq q$ we have that either*

$$|b_k| \lesssim 2^{-\lambda k} \left(\sum_{j=0}^{k} \left(2^{\lambda j}|a_j|\right)^r \right)^{1/r} , \tag{15}$$

or

$$|b_k| \lesssim \left(\sum_{j=k}^{\infty} |a_j|^r \right)^{1/r}. \tag{16}$$

Then we have

$$\|\mathbf{b}\|_{\ell_q^\alpha} \lesssim \|\mathbf{a}\|_{\ell_q^\alpha}.$$

Observe that defining $a_k := \left\| (\langle f, \gamma_\lambda \rangle)_{\lambda \in \Lambda_k} \right\|_p$, we have $\|f\|_{S^\alpha_{p,q}} = \|\mathbf{a}\|_{\ell_q^\alpha}$. Armed with these useful facts, we may now proceed with the proof of Theorem 4.10.

Proof of Theorem 4.10. We start by fixing some notation:

$$\mathbf{f}^\Gamma := (\langle f, \gamma_\mu \rangle)_{\mu \in \Lambda^0}, \quad \mathbf{f}^\Gamma_j := (\langle f, \gamma_\mu \rangle)_{\mu \in \Lambda^0_j}, \quad \mathbf{f}^\Sigma := (\langle f, \sigma_\lambda \rangle)_{\lambda \in \Lambda}, \quad \mathbf{f}^\Sigma_j := (\langle f, \sigma_\lambda \rangle)_{\lambda \in \Lambda_j}.$$

Further, we write $\Gamma_j = \left\{ \gamma_\mu : \ \mu \in \Lambda^0_j \right\}$ and similar for the systems Σ, $\tilde{\Sigma}$. Define

$$\mathbf{A} := \langle \Gamma, \Sigma \rangle, \quad \mathbf{A}_{i,j} := \langle \Gamma_i, \Sigma_j \rangle, \quad \tilde{\mathbf{A}} := \langle \tilde{\Sigma}, \Gamma \rangle, \quad \mathbf{A}_{i,j} := \langle \tilde{\Sigma}_i, \Gamma_j \rangle.$$

We have

$$\mathbf{f}^\Sigma = \left(\mathbf{f}^\Gamma\right)^\top \mathbf{A}, \quad \mathbf{f}^\Sigma_i = \sum_{j \geq 0} \left(\mathbf{f}^\Gamma_j\right)^\top \mathbf{A}_{i,j}, \quad \mathbf{f}^\Gamma = \left(\mathbf{f}^\Sigma\right)^\top \mathbf{A}, \quad \mathbf{f}^\Gamma_i = \sum_{j \geq 0} \left(\mathbf{f}^\Sigma_j\right)^\top \mathbf{A}_{i,j}.$$

Let us first assume that $p \geq 1$. Then we would like to show that

$$\left(\sum_{j \geq 0} \left(2^{\alpha j} \left(\sum_{\lambda \in \Lambda_j} |\langle f, \sigma_\lambda \rangle|^p \right)^{1/p} \right)^q \right)^{1/q} < \infty,$$

whenever

$$\left(\sum_{j \geq 0} \left(2^{\alpha j} \left(\sum_{\mu \in \Lambda^0_j} |\langle f, \gamma_\mu \rangle|^p \right)^{1/p} \right)^q \right)^{1/q} < \infty.$$

For this, we obtain

$$b_i := \left\| \mathbf{f}^\Sigma_i \right\|_p = \left\| \sum_{j \geq 0} \left(\mathbf{f}^\Gamma_j\right)^\top \mathbf{A}_{i,j} \right\|_p \leq \sum_{j \geq 0} \left\| \left(\mathbf{f}^\Gamma_j\right)^\top \mathbf{A}_{i,j} \right\|_p = d_i + e_i, \tag{17}$$

where

$$d_i := \sum_{j > i} \left\| \left(\mathbf{f}^\Gamma_j\right)^\top \mathbf{A}_{i,j} \right\|_p \quad \text{and} \quad e_i := \sum_{j \leq i} \left\| \left(\mathbf{f}^\Gamma_j\right)^\top \mathbf{A}_{i,j} \right\|_p.$$

Next, we will prove that the inequalities (15), (16) are satisfied for the sequences d_i, e_i, respectively. By Lemma 4.12, this proves the desired claim.

We start by deriving the following estimate for d_i:

$$d_i \leq \sum_{j > i} \left\| \left(\mathbf{f}^\Gamma_j\right) \|_p \|\mathbf{A}_{i,j} \right\|_{\ell_p \to \ell_p}$$

To further analyze $\|\mathbf{A}_{i,j}\|_{\ell_p \to \ell_p}$, we employ Lemma 4.4 to obtain

$$\|\mathbf{A}_{i,j}\|_{\ell_p \to \ell_p} \leq \max \left(\sup_{\mu \in \Lambda^0_i} \sum_{\lambda \in \Lambda_j} |\langle \gamma_\mu, \sigma_\lambda \rangle|, \sup_{\lambda \in \Lambda_j} \sum_{\mu \in \Lambda^0_i} |\langle \gamma_\mu, \sigma_\lambda \rangle| \right) \tag{18}$$

Using the fact that Γ and Σ are parabolic molecules of arbitrary order, Theorem 2.9 implies that for N arbitrary,

$$|\langle \gamma_\mu, \sigma_\lambda \rangle| \lesssim \omega(\mu, \lambda)^{-N}. \tag{19}$$

By (19) and the fact that the parametrization for Λ is strongly admissible, we can further estimate the first term in (18) by

$$\sup_{\mu\in\Lambda_i^0}\sum_{\lambda\in\Lambda_j}\omega(\mu,\lambda)^{-N} = 2^{-N|i-j|}\sup_{\mu\in\Lambda_i^0}\sum_{\lambda\in\Lambda_j}\left(1+2^{\min(i,j)}d(\mu,\lambda)\right)^{-N} \lesssim 2^{-(N-l)|i-j|}.$$

The second term is treated similarly, and we wind up with

$$\|\mathbf{A}_{i,j}\|_{\ell_p\to\ell_p} \lesssim 2^{-N|i-j|} \tag{20}$$

for N arbitrarily large. In particular, this implies that

$$d_i \lesssim \sum_{j>i}\left\|\left(\mathbf{f}_j^\Gamma\right)\right\|_p \lesssim \left(\sum_{j>i}\left\|\left(\mathbf{f}_j^\Gamma\right)\right\|_p\right)^{1/r}$$

with $r := \min(1,q)$, and this is (16). Similarly we can estimate

$$e_i \lesssim \sum_{j\le i}2^{-N(i-j)}\left\|\left(\mathbf{f}_j^\Gamma\right)\right\|_p = 2^{-Ni}\sum_{j\le i}2^{Nj}\left\|\left(\mathbf{f}_j^\Gamma\right)\right\|_p \lesssim\ 2^{-Ni}\left(\sum_{j\le i}\left(2^{Nj}\left\|\left(\mathbf{f}_j^\Gamma\right)\right\|_p\right)^r\right)^{1/r}$$

which is (15). Applying Lemma 4.12 yields

$$\left(\sum_{j\ge 0}\left(2^{\alpha j}\left(\sum_{\lambda\in\Lambda_j}|\langle f,\sigma_\lambda\rangle|^p\right)^{1/p}\right)^q\right)^{1/q} \lesssim \left(\sum_{j\ge 0}\left(2^{\alpha j}\left(\sum_{\mu\in\Lambda_j^0}|\langle f,\gamma_\mu\rangle|^p\right)^{1/p}\right)^q\right)^{1/q}$$

which proves one half of the desired norm equivalence. The other half (and the case of $\tilde{\Sigma}$) can be shown in exactly the same way. Therefore, for $p \ge 1$, the claim of the theorem is proven.

Let us now turn to the case $p < 1$. For this, we need to replace the estimate (17) with

$$\begin{aligned}
|b_i| &:= \left\|\mathbf{f}_i^\Sigma\right\|_p = \left\|\sum_{j\ge 0}\left(\mathbf{f}_j^\Gamma\right)^\top\mathbf{A}_{i,j}\right\|_p \\
&\lesssim \left\|\sum_{j\le i}\left(\mathbf{f}_j^\Gamma\right)^\top\mathbf{A}_{i,j}\right\|_p + \left\|\sum_{j>i}\left(\mathbf{f}_j^\Gamma\right)^\top\mathbf{A}_{i,j}\right\|_p \\
&\le \left(\sum_{j\le i}\left\|\mathbf{f}_j^\Gamma\right\|_p^p\|\mathbf{A}_{i,j}\|_{\ell_p\to\ell_p}^p\right)^{1/p} + \left(\sum_{j>i}\left\|\mathbf{f}_j^\Gamma\right\|_p^p\|\mathbf{A}_{i,j}\|_{\ell_p\to\ell_p}^p\right)^{1/p} \\
&\lesssim \left(\sum_{j\le i}\left\|\mathbf{f}_j^\Gamma\right\|_p^r\|\mathbf{A}_{i,j}\|_{\ell_p\to\ell_p}^r\right)^{1/r} + \left(\sum_{j>i}\left\|\mathbf{f}_j^\Gamma\right\|_p^r\|\mathbf{A}_{i,j}\|_{\ell_p\to\ell_p}^r\right)^{1/r} \\
&=: d_i + e_i,
\end{aligned}$$

where $r := \min(p,q)$. Now we can use (20) and proceed as above to show that the Hardy inequalities are satisfied for d_i and e_i. Then, the application of Lemma 4.12 finishes the proof. □

As a corollary we can consider the shearlet frame Σ constructed in [18] and briefly described in Subsection 3.2.2 and arrive at the following theorem.

Theorem 4.13. *The curvelet frame Γ^0 and the shearlet frame Σ constructed in [18] span the same approximation spaces.*

We remark that the same conclusion holds for the frames described in Subsections 3.1.2 and 3.1.3. Without proof we also mention that Theorem 4.10 and Theorem 4.13 also hold for the spaces $G^\alpha_{p,q}$. The proof is similar but slightly more technical.

We wish to stress that in fact this result for the first time proves the meta-theorem that curvelet- and shearlet properties are equivalent.

Remark 4.14. *A similar result to Theorem 4.13 has recently been shown in [28]. The proofs in this paper only apply to bandlimited constructions which present considerably less technical difficulty.*

5 Proof of Theorem 2.9

The present sections presents the proof of our main result, namely the almost orthogonality of any two systems of parabolic molecules of sufficient order. Since the argument is quite involved we start by collecting some useful lemmata below in Subsection 5.1 before we go on to the proof of the main result, Theorem 2.9 in Subsection 5.2.

5.1 Some Estimates

Here we collect several estimates which will turn out useful in the proof of Theorem 2.9. The following lemma can be found in [17, Appendix K.1].

Lemma 5.1. *For $N > 1$ and $a, a' \in \mathbb{R}_+$, we have the inequality*

$$\int_{\mathbb{R}} (1 + a|x|)^{-N} (1 + a'|x - y|)^{-N} \, d\varphi \lesssim \max(a, a')^{-1}(1 + \min(a, a')|y|)^{-N}.$$

As a corollary we can show the next result.

Lemma 5.2. *Assume that $|\theta| \le \frac{\pi}{2}$ and $N > 1$. Then we have for $a, a' > 0$ the inequality*

$$\int_{\mathbb{T}} (1 + a|\sin(\varphi)|)^{-N} (1 + a'|\sin(\varphi + \theta)|)^{-N} \, d\varphi \lesssim \max(a, a')^{-1}(1 + \min(a, a')|\theta|)^{-N}. \tag{21}$$

Proof. For $\varphi \in \mathbb{T}$, we have the estimate

$$|\sin(\varphi)| \ge \begin{cases} |\varphi| & \varphi \in I_1 := \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \\ |\varphi - \pi| & \varphi \in I_2 := \left[\frac{\pi}{2}, \pi\right], \\ |\varphi + \pi| & \varphi \in I_3 := \left[-\pi, -\frac{\pi}{2}\right]. \end{cases}$$

In order to use Lemma 5.1 we now split $\mathbb{T}$ into nine intervals depending on $\varphi + \theta, \varphi \in I_1, I_2, I_3$. Then the left-hand side of (21) can be estimated by nine terms of the form

$$\int_{\mathbb{R}} (1 + a|\varphi|)^{-N} (1 + a'|\varphi + \vartheta + \theta|)^{-N} \, d\varphi,$$

where $\vartheta \in \{0, \pm\pi, \pm 2\pi\}$. By Lemma 5.1 this expression can be bounded by a constant times

$$\max(a, a')^{-1}(1 + \min(a, a')|\theta + \vartheta|)^{-N}.$$

Now it remains to note that for $\vartheta \in \{\pm\pi, \pm 2\pi\}$ and $|\theta| \le \frac{\pi}{2}$ we have $|\theta + \vartheta| \ge |\theta|$. This proves the lemma. □

Define the expression

$$S_{\lambda,M,N_1,N_2}(r, \varphi) := \min\left(1, 2^{-s_\lambda}(1 + r)\right)^M \left(1 + 2^{s_\lambda/2}|\sin(\varphi + \theta_\lambda)|\right)^{-N_2} \left(1 + 2^{-s_\lambda} r\right)^{-N_1}.$$

The following lemma will be used in order to decouple the angular and the radial variables.

Lemma 5.3. *We have the estimate*

$$\min\left(1, 2^{-s_\lambda}(1+r)\right)^M \left(1+2^{-s_\lambda}r\right)^{-N_1} \left(1+2^{-s_\lambda/2}r|\sin(\varphi+\theta_\lambda)|\right)^{-N_2} \lesssim S_{\lambda,M-L,N_1,L}(r,\varphi)$$

for every $0 \le L \le N_2$.

Proof. After picking L we can estimate the quantity on the left hand side by

$$\min\left(1, 2^{-s_\lambda}(1+r)\right)^{M-L} \left(1+2^{-s_\lambda}r\right)^{-N_1} \left(\frac{\min\left(1, 2^{-s_\lambda}(1+r)\right)}{1+2^{-s_\lambda/2}r|\sin(\varphi+\theta_\lambda)|}\right)^L.$$

We need to show that

$$\frac{\min\left(1, 2^{-s_\lambda}(1+r)\right)}{1+2^{-s_\lambda/2}r|\sin(\varphi+\theta_\lambda)|} \lesssim \left(1+2^{s_\lambda/2}|\sin(\varphi+\theta_\lambda)|\right)^{-1}. \tag{22}$$

In order to prove (22), we distinguish three cases:

- $\mathbf{r \ge 2^{s_\lambda}}$**:** In this case we derive

$$\begin{aligned}\frac{\min\left(1, 2^{-s_\lambda}(1+r)\right)}{1+2^{-s_\lambda/2}r|\sin(\varphi+\theta_\lambda)|} &\le \frac{1}{1+2^{-s_\lambda/2}r|\sin(\varphi+\theta_\lambda)|} \le \frac{1}{1+2^{-s_\lambda/2}2^{s_\lambda}|\sin(\varphi+\theta_\lambda)|} \\ &\le \left(1+2^{s_\lambda/2}|\sin(\varphi+\theta_\lambda)|\right)^{-1}.\end{aligned}$$

- $\mathbf{r \le 1}$**:** For $r \le 1$ we have

$$\frac{\min\left(1, 2^{-s_\lambda}(1+r)\right)}{1+2^{-s_\lambda/2}r|\sin(\varphi+\theta_\lambda)|} \lesssim 2^{-s_\lambda} \lesssim \left(1+2^{s_\lambda/2}|\sin(\varphi+\theta_\lambda)|\right)^{-1}.$$

- $\mathbf{1 < r < 2^{s_\lambda}}$**:** In this case we have

$$\frac{\min\left(1, 2^{-s_\lambda}(1+r)\right)}{1+2^{-s_\lambda/2}r|\sin(\varphi+\theta_\lambda)|} = \frac{1+r}{r}\frac{1}{\frac{2^{s_\lambda}}{r}+2^{s_\lambda/2}|\sin(\varphi+\theta)|}.$$

  Since $r > 1$ we have that $\frac{1+r}{r} \le 2$ and since $r < 2^{s_\lambda}$, we have that $\frac{2^{s_\lambda}}{r} \ge 1$. This proves the statement.

□

Lemma 5.4. *For* $A, B > 0$ *and*

$$M > A - \frac{5}{4}, \quad N_2 \ge B, \quad N_1 \ge A + 3/4,$$

we have the estimate

$$2^{-\frac{3}{4}(s_\lambda+s_\mu)} \int_{\mathbb{R}_+}\int_{\mathbb{T}} S_{\lambda,M,N_1,N_2}(r,\varphi) S_{\mu,M,N_1,N_2}(r,\varphi)\, r dr d\varphi \lesssim 2^{-A|s_\lambda-s_\mu|}\left(1+2^{\min(s_\lambda,s_\mu)/2}|\theta_\lambda-\theta_\mu|\right)^{-B}.$$

Proof. We assume that $s_\mu \ge s_\lambda$ and start by showing the angular decay: By Lemma 5.2 and $N_2 \ge B$, we have

$$2^{-\frac{3}{4}(s_\lambda+s_\mu)} \int_{\mathbb{R}_+}\int_{\mathbb{T}} S_{\lambda,M,N_1,N_2}(r,\varphi) S_{\mu,M,N_1,N_2}(r,\varphi)\, r dr d\varphi \lesssim \mathcal{S}\cdot 2^{\frac{3}{4}(s_\mu-s_\lambda)}\left(1+2^{s_\lambda/2}|\theta_\lambda-\theta_\mu|\right)^{-B},$$

where

$$\mathcal{S} := 2^{-2s_\mu}\int_{\mathbb{R}_+} \min\left(1, 2^{-s_\lambda}(1+r)\right)^M \min\left(1, 2^{-s_\mu}(1+r)\right)^M \left(1+2^{-s_\lambda}r\right)^{-N_1}\left(1+2^{-s_\mu}r\right)^{-N_1} r dr. \tag{23}$$

The remaining estimate

$$\mathcal{S} \lesssim 2^{-(A+3/4)|s_\lambda - s_\mu|}$$

is established by splitting up this integral into the four cases $r < 1$, $1 \le r < 2^{s_\lambda}$, $2^{s_\lambda} \le r < 2^{s_\mu}$ and $r \ge 2^{s_\mu}$.

Case 1: $0 \le r \le 1$

Here we only use the moment property and estimate

$$\begin{aligned}
(23) \quad &\lesssim \quad 2^{-2s_\mu} \int_0^1 2^{-s_\lambda M} 2^{-s_\mu M} r^{2M+1} dr \\
&\le \quad 2^{-s_\mu(2+M)} 2^{-s_\lambda M} \\
&\le \quad 2^{-(A+3/4)(s_\mu - s_\lambda)}.
\end{aligned}$$

Case 2: $1 \le r \le 2^{s_\lambda}$

For this case, we estimate

$$\begin{aligned}
(23) \quad &\lesssim \quad 2^{-2s_\mu} \int_1^{2^{s_\lambda}} 2^{-s_\mu M} r^M \left(2^{-s_\lambda} r\right)^{-N_1} r dr \\
&= \quad 2^{-(M+2)s_\mu} 2^{N_1 s_\lambda} \int_1^{2^{s_\lambda}} r^{M+1-N_1} dr \\
&\le \quad 2^{-(M+2)s_\mu} 2^{N_1 s_\lambda} 2^{(M+2-N_1)s_\lambda} \\
&= \quad 2^{-(M+2)(s_\mu - s_\lambda)} \\
&\le \quad 2^{-(A+3/4)(s_\mu - s_\lambda)}.
\end{aligned}$$

Case 3: $2^{s_\lambda} \le r \le 2^{s_\mu}$

For this case, we estimate

$$\begin{aligned}
(23) \quad &\lesssim \quad 2^{-2s_\mu} \int_{2^{s_\lambda}}^{2^{s_\mu}} \left(2^{-s_\mu} r\right)^M \left(2^{-s_\lambda} r\right)^{-N_1} r dr \\
&= \quad 2^{-(2+M)s_\mu} 2^{N_1 s_\lambda} \int_{2^{s_\lambda}}^{2^{s_\mu}} r^{M+1-N_1} dr \\
&\lesssim \quad 2^{-(2+M)s_\mu} 2^{N_1 s_\lambda} 2^{(M+2-N_1)s_\mu} \\
&= \quad 2^{-N_1(s_\mu - s_\lambda)} \\
&\le \quad 2^{-(A+3/4)(s_\mu - s_\lambda)}.
\end{aligned}$$

Case 4: $2^{s_\mu} \le r$

For this case, we estimate

$$\begin{aligned}
(23) \quad &\lesssim \quad 2^{-2s_\mu} \int_{2^{s_\mu}}^{\infty} \left(2^{-s_\lambda} r\right)^{-N_1} \left(2^{-s_\mu} r\right)^{-N_1} r dr \\
&= \quad 2^{-2s_\mu} 2^{N_1 s_\mu} 2^{N_1 s_\lambda} \int_{2^{s_\mu}}^{\infty} r^{-2N_1+1} dr \\
&= \quad 2^{-2s_\mu} 2^{N_1 s_\mu} 2^{N_1 s_\lambda} 2^{(-2N_1+2)s_\mu} \\
&= \quad 2^{-N_1(s_\mu - s_\lambda)} \\
&\le \quad 2^{-(A+3/4)(s_\mu - s_\lambda)}.
\end{aligned}$$

The proof is completed. □

5.2 Almost Orthogonality

We now have all ingredients to prove our main result, which is Theorem 2.9.

Proof of Theorem 2.9. To keep the notation simple, we assume that $\theta_\lambda = 0$ and define $s_0 := \min(s_\lambda, s_\mu)$. Further, we set

$$\delta x := x_\lambda - x_\mu, \quad \delta\theta := \theta_\lambda - \theta_\mu.$$

By definition, we can write

$$m_\lambda(\cdot) = 2^{\frac{3}{4}s_\lambda} a^{(\lambda)}\left(D_{2^{s_\lambda}} R_{\theta_\lambda}(\cdot - x_\lambda)\right), \quad p_\mu(\cdot) = 2^{\frac{3}{4}s_\mu} b^{(\mu)}\left(D_{2^{s_\mu}} R_{\theta_\mu}(\cdot - x_\mu)\right),$$

where both $a^{(\lambda)}$ and $b^{(\mu)}$ satisfy (1). We have the equality

$$\begin{aligned}\langle m_\lambda, p_\mu\rangle &= \langle \hat{m}_\lambda, \hat{p}_\mu\rangle = 2^{-\frac{3}{4}(s_\lambda+s_\mu)} \int_{\mathbb{R}^2} \hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right) \overline{\hat{b}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)} \exp\left(-2\pi i \xi\cdot\delta x\right) d\xi \\ &= 2^{-\frac{3}{4}(s_\lambda+s_\mu)} \int_{\mathbb{R}^2} \mathcal{L}^k\left(\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right) \overline{\hat{b}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)}\right) \mathcal{L}^{-k}\left(\exp\left(-2\pi i \xi\cdot\delta x\right)\right) d\xi, \quad (24)\end{aligned}$$

where $\mathcal{L}$ is the symmetric differential operator (acting on the frequency variable) defined by

$$\mathcal{L} := I - 2^{s_0}\Delta_\xi - \frac{2^{2s_0}}{1+2^{s_0}|\delta\theta|^2}\frac{\partial^2}{\partial \xi_1^2}.$$

We have

$$\mathcal{L}^{-k}\left(\exp\left(-2\pi i\xi\cdot\delta x\right)\right) = \left(1 + 2^{s_0}|\delta x|^2 + \frac{2^{2s_0}}{1+2^{s_0}|\delta\theta|}\langle e_\lambda, \delta x\rangle^2\right)^{-k} \exp\left(-2\pi i \xi\cdot\delta x\right), \quad (25)$$

where e_λ denotes the unit vector pointing in the direction described by the angle θ_λ. By Lemma 5.5 and for $k \le \frac{R}{2}$, we have the inequality

$$\mathcal{L}^k\left(\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right) \overline{\hat{b}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)}\right) \lesssim S_{\lambda,M-N_2,N_1,N_2}(\xi) S_{\mu,M-N_2,N_1,N_2}(\xi).$$

Then, by (24) and (25) it follows that

$$|\langle m_\lambda, p_\mu\rangle| \lesssim 2^{-\frac{3}{4}(s_\lambda+s_\mu)} \int_{\mathbb{R}^2} S_{\lambda,M-N_2,N_1,N_2}(\xi) S_{\mu,M-N_2,N_1,N_2}(\xi) d\xi \left(1 + 2^{s_0}|\delta x|^2 + \frac{2^{2s_0}}{1+2^{s_0}|\delta\theta|}\langle e_\lambda, \delta x\rangle^2\right)^{-k}$$

for all $k \le \frac{R}{2}$. Now we can use Lemma 5.4 and the fact that $R \ge 2N$ to establish that

$$\begin{aligned}|\langle m_\lambda, p_\mu\rangle| &\lesssim 2^{-2N|s_\lambda - s_\mu|}\left(1 + 2^{s_0}|\delta\theta|^2\right)^{-N}\left(1 + 2^{s_0}|\delta x|^2 + \frac{2^{2s_0}}{1+2^{s_0}|\delta\theta|}\langle e_\lambda, \delta x\rangle^2\right)^{-N} \\ &\le 2^{-2N|s_\lambda - s_\mu|}\left(1 + 2^{s_0}|\delta\theta|^2 + 2^{s_0}|\delta x|^2 + \frac{2^{2s_0}}{1+2^{s_0}|\delta\theta|}\langle e_\lambda, \delta x\rangle^2\right)^{-N} \\ &\lesssim 2^{-2N|s_\lambda - s_\mu|}\left(1 + 2^{s_0}\left(|\delta\theta|^2 + |\delta x|^2 + |\langle e_\lambda, \delta x\rangle|\right)\right)^{-N} = \omega(\lambda,\mu)^{-N}.\end{aligned}$$

The last inequality follows from the equation in the last line of the proof of [5, Lemma 2.3]. This proves the desired statement. ☐

Lemma 5.5. *Assume that (6) holds for two systems of parabolic molecules of order (R, M, N_1, N_2). Utilizing the notion of the proof of Theorem 2.9, we have*

$$\mathcal{L}^k\left(\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right) \overline{\hat{b}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)}\right) \lesssim S_{\lambda,M-N_2,N_1,N_2}(\xi) S_{\mu,M-N_2,N_1,N_2}(\xi)$$

for all $k \le R/2$.

Proof. We show that

$$\begin{aligned}&\left|\mathcal{L}^k\left(\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right) \overline{\hat{b}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)}\right)\right| \lesssim \\ &\qquad \min\left(1, 2^{-s_\lambda}(1+r)\right)^M \left(1 + 2^{-s_\lambda} r\right)^{-N_1}\left(1 + 2^{-s_\lambda/2} r|\sin(\varphi+\theta_\lambda)|\right)^{-N_2} \\ &\qquad \cdot \min\left(1, 2^{-s_\mu}(1+r)\right)^M \left(1 + 2^{-s_\mu} r\right)^{-N_1}\left(1 + 2^{-s_\mu/2} r|\sin(\varphi+\theta_\mu)|\right)^{-N_2} \quad (26)\end{aligned}$$

which, using Lemma 5.3 with $L = N_2$, implies the desired statement. To show (26) we use induction in k, namely we show that if we have two functions $a^{(\lambda)}, b^{(\mu)}$ satisfying (1) for R, M, N_1, N_2, then the expression

$$\mathcal{L}\left(\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)\overline{\hat{b}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)}\right)$$

can be written as a finite linear combination of terms of the form

$$\hat{c}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)\overline{\hat{d}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)}$$

with c, d satisfying (1) and R replaced by $R-2$, see Lemma 5.6. Iterating this argument we can establish that for $k \leq R/2$

$$\mathcal{L}^k\left(\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)\overline{\hat{b}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)}\right) \tag{27}$$

can be expressed as a finite linear combination of terms of the form

$$\hat{c}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)\overline{\hat{d}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)} \tag{28}$$

with

$$\left|\hat{c}^{(\lambda)}(\xi)\right| \lesssim \min\left(1, 2^{-s_\lambda} + |\xi_1| + 2^{-s_\lambda/2}|\xi_2|\right)^M \langle|\xi|\rangle^{-N_1}\langle\xi_2\rangle^{-N_2}, \tag{29}$$

and an analogous estimate for $d^{(\mu)}$. Combining (28) and (29), we obtain that

$$\begin{aligned}
&|(27)| \lesssim \\
&\min\left(1, 2^{-s_\lambda} + \left|\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)_1\right| + 2^{-s_\lambda/2}\left|\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)_2\right|\right)^M \langle|D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi|\rangle^{-N_1}\left\langle\left|\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)_2\right|\right\rangle^{-N_2} \\
\cdot&\min\left(1, 2^{-s_\mu} + \left|\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)_1\right| + 2^{-s_\mu/2}\left|\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)_2\right|\right)^M \left\langle\left|D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right|\right\rangle^{-N_1}\left\langle\left|\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)_2\right|\right\rangle^{-N_2}.
\end{aligned}$$

Transforming this inequality into polar coordinates as in (2) yields (26). This finishes the proof. □

Lemma 5.6. *Given two functions $a^{(\lambda)}, b^{(\mu)}$ satisfying (1) for R, M, N_1, N_2. Then the expression*

$$\mathcal{L}\left(\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)\overline{\hat{b}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)}\right)$$

can be written as a finite linear combination of terms of the form

$$\hat{c}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)\overline{\hat{d}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)}$$

with c, d satisfying (1) for $R-2, M, N_1, N_2$.

Proof. Recall the definition

$$\mathcal{L} := I - 2^{s_0}\Delta_\xi - \frac{2^{2s_0}}{1 + 2^{s_0}|\delta\theta|^2}\frac{\partial^2}{\partial\xi_1^2}$$

To show this statement we treat the three summands of the operator $\mathcal{L}$ separately. The first part is the identity, and therefore the statement is trivial. To handle the second part, the frequency Laplacian $2^{s_0}\Delta$, we use the product rule

$$\Delta(fg) = 2\left(\partial^{(1,0)}f\partial^{(1,0)}g + \partial^{(0,1)}f\partial^{(0,1)}g\right) + (\Delta f)g + f(\Delta g).$$

Therefore we need to estimate the derivatives of degree 1 and the Laplacians of the two factors in the product

$$\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}} R_{\theta_\lambda}\xi\right)\overline{\hat{b}^{(\mu)}\left(D_{2^{-s_\mu}} R_{\theta_\mu}\xi\right)} =: A(\xi)B(\xi).$$

We start with the first factor,

$$A(\xi) = \hat{a}^{(\lambda)}\left(2^{-s_\lambda}\cos(\theta_\lambda)\xi_1 - 2^{-s_\lambda}\sin(\theta_\lambda)\xi_2, 2^{-s_\lambda/2}\sin(\theta_\lambda)\xi_1 + 2^{-s_\lambda/2}\cos(\theta_\lambda)\xi_2\right).$$

Define

$$A_1(\xi) := \partial^{(1,0)}\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}}R_{\theta_\lambda}\xi\right) \quad \text{and} \quad A_2(\xi) := \partial^{(0,1)}\hat{a}^{(\lambda)}\left(D_{2^{-s_\lambda}}R_{\theta_\lambda}\xi\right).$$

By definition, the functions $A_1,\ A_2$ satisfy (1) with R replaced by $R-1$. An application of the chain rule shows that

$$\partial^{(1,0)}A(\xi) = 2^{-s_\lambda}\cos(\theta_\lambda)A_1(\xi) + 2^{-s_\lambda/2}\sin(\theta_\lambda)A_2(\xi).$$

Analogously, one can compute

$$\partial^{(0,1)}A(\xi) = -2^{-s_\lambda}\sin(\theta_\lambda)A_1(\xi) + 2^{-s_\lambda/2}\cos(\theta_\lambda)A_2(\xi),$$

and the exact same expressions for B using the obvious definitions for $B_1,\ B_2$. We get

$$\begin{aligned}\partial^{(1,0)}A\partial^{(1,0)}B = 2^{-s_\lambda-s_\mu}\cos(\theta_\lambda)\cos(\theta_\mu)A_1B_1 + 2^{-s_\lambda/2-s_\mu}\sin(\theta_\lambda)\cos(\theta_\mu)A_2B_1 \\ + 2^{-s_\mu/2-s_\lambda}\sin(\theta_\mu)\cos(\theta_\lambda)A_1B_2 + 2^{-s_\lambda/2-s_\mu/2}\sin(\theta_\lambda)\sin(\theta_\mu)A_2B_2.\end{aligned}$$

It follows that $2^{s_0}\partial^{(1,0)}A\partial^{(1,0)}B$ can be written as a linear combination as claimed (recall that $s_0 = \min(s_\lambda, s_\mu)$). The same argument applies to the product $2^{s_0}\partial^{(0,1)}A\partial^{(0,1)}B$.

It remains to consider the factor

$$(\Delta A)B + A(\Delta B),$$

where, for symmetry reasons, we only treat the summand

$$(\Delta A)B.$$

In fact, it suffices to only consider

$$(\partial^{(2,0)}A)B = \left(2^{-2s_\lambda}\cos(\theta_\lambda)^2A_{11} + 2^{-3s_\lambda/2+1}\sin(\theta_\lambda)\cos(\theta_\lambda)A_{12} - 2^{-s_\lambda}\sin(\theta_\lambda)^2A_{22}\right)B$$

with A_{ij} defined in an obvious way, satisfying (1) with R replaced by $R-2$. The term $(\partial^{(2,0)}A)B$, and hence $(\Delta A)B$, can be handled in the same way, as can $A(\Delta B)$. This takes care of the term $2^{s_0}\Delta$ in the definition of $\mathcal{L}$.

Finally we need to handle the last term in the definition of $\mathcal{L}$, namely

$$\frac{2^{2s_0}}{1+2^{s_0}|\theta_\mu|^2}\frac{\partial^2}{\partial\xi_1^2}$$

for $\theta_\lambda = 0$ (otherwise the second order derivative would be in the direction of the unit vector with angle θ_λ with obvious modifications in the proof). With our notation and using the product rule we need to consider terms of the form

$$\left(\partial^{(2,0)}A\right)B, \quad \left(\partial^{(1,0)}A\right)\left(\partial^{(1,0)}B\right), \quad A\left(\partial^{(2,0)}B\right),$$

and show that each of them, multiplied by the factor $\frac{2^{2s_0}}{1+2^{s_0}|\theta_\mu|^2}$, satisfies the desired representation. Let us start with

$$\left(\partial^{(2,0)}A\right)B,$$

which, using the fact that $\sin(\theta_\lambda) = 0$, can be written as

$$2^{-2s_\lambda}A_{11}B,$$

and which clearly satisfies the desired assertion. Now consider the expression

$$\left(\partial^{(1,0)}A\right)\left(\partial^{(1,0)}B\right)$$

which can be written as

$$2^{-s_\lambda}2^{-s_\mu}\cos(\theta_\mu)A_1B_1 + 2^{-s_\lambda}2^{-s_\mu/2}\sin(\theta_\mu)A_1B_2.$$

The first summand in this expression clearly causes no problems. To handle the second term we need to show that

$$\frac{2^{2s_0}}{1+2^{s_0}|\theta_\mu|^2}2^{-s_\lambda}2^{-s_\mu/2}\sin(\theta_\mu)\lesssim 1. \tag{30}$$

Here we have to distinguish two cases. First, assume that $|\theta_\mu|\le 2^{-s_0/2}$. Then we can estimate

$$\sin(\theta_\mu)\lesssim 2^{-s_0/2},$$

which readily yields the desired bound for (30). For the case $|\theta_\mu|\ge 2^{-s_0/2}$ we estimate

$$\frac{2^{2s_0}}{1+2^{s_0}|\theta_\mu|^2}2^{-s_\lambda}2^{-s_\mu/2}\sin(\theta_\mu)\lesssim \frac{2^{2s_0}}{1+2^{s_0/2}|\theta_\mu|}2^{-s_0}2^{-s_0/2}|\theta_\mu|\le \frac{2^{2s_0}}{2^{s_0/2}|\theta_\mu|}2^{-s_0}2^{-s_0/2}|\theta_\mu|=1$$

which shows (30) also for this case.

We are left with estimating the term

$$A\left(\partial^{(2,0)}B\right)$$

which can be written as

$$2^{-2s_\mu}\cos(\theta_\mu)^2AB_{11}+2^{-3s_\mu/2+1}\sin(\theta_\mu)\cos(\theta_\mu)AB_{12}+2^{-s_\mu}\sin(\theta_\mu)^2AB_{22}.$$

The first two terms are of a form already treated, and the last term can be handled using the fact that $\sin(\theta_\mu)^2\le\theta_\mu^2$. □

Acknowledgements

G. Kutyniok would like to thank Wolfgang Dahmen, David Donoho, Wang-Q Lim, and Pencho Petrushev for enlightening discussions on this and related topics. She acknowledges support by the Einstein Foundation Berlin, by Deutsche Forschungsgemeinschaft (DFG) Grant SPP-1324 KU 1446/13 and DFG Grant KU 1446/14, and by the DFG Research Center Matheon "Mathematics for key technologies" in Berlin. The research of P. Grohs was in part funded by the European Research Council under Grant ERC Project STAHDPDE No. 247277.

References

[1] F. Andersson, M. de Hoop, H. Smith, and G. Uhlmann. A multi-scale approach to hyperbolic evolution equations with limited smoothness. *Comm. PDE*, 33:988–1017, 2008.

[2] A. Benedek and R. Panzone. The space L^p, with mixed norm. *Duke J.*, 28(3):301–324, 1961.

[3] L. Borup and M. Nielsen. Frame decompositions of decomposition spaces. *J. Fourier Anal. Appl.*, 13:39–70, 2007.

[4] E. Candès and D. Donoho. Continuous curvelet transform: II. Discretization and frames. *Appl. Comput. Harmon. Anal.*, 19(2):198–222, 2005.

[5] E. J. Candès and L. Demanet. The curvelet representation of wave propagators is optimally sparse. *Comm. Pure Appl. Math.*, 58:1472–1528, 2002.

[6] E. J. Candès and D. L. Donoho. Curvelets and curvilinear integrals. *J. Approx. Theory*, 113:59–90, 2000.

[7] E. J. Candès and D. L. Donoho. Recovering edges in ill-posed inverse problems: Optimality of curvelet frames. *Ann. Statist.*, 30:784–842, 2000.

[8] E. J. Candès and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with C^2 singularities. *Comm. Pure Appl. Math.*, 56:219–266, 2004.

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

[10] R. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.

[11] R. DeVore and V. Popov. Interpolation of Besov spaces. *Trans. Amer. Math. Soc.*, 305:397–414, 1988.

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

[13] D. Donoho and G. Kutyniok. Microlocal analysis of the geometric separation problem. *Comm. Pure Appl. Math.* to appear.

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

[15] K. G. F. Colonna, G. R. Easley and D. Labate. Radon transform inversion using the shearlet representation. *Appl. Comput. Harmon. Anal.*, 29:232–250, 2010.

[16] H. G. Feichtinger and P. Gröbner. Banach spaces of distributions defined by decomposition methods. *Math. Nachrichten*, 123:97–120, 1985.

[17] L. Grafakos. *Classical Fourier Analysis.* Springer, 2 edition, 2008.

[18] P. Grohs. Bandlimited shearlet frames with nice duals. 2011. available as SAM report 2011-55 from `http://www.sam.math.ethz.ch/reports`.

[19] P. Grohs. Intrinsic localization of anisotropic frames. 2012. available as SAM report 2012-04 from `http://www.sam.math.ethz.ch/reports`.

[20] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM J. Math Anal.*, 39:298–318, 2007.

[21] K. Guo and D. Labate. Representation of Fourier integral operators using shearlets. *J. Fourier Anal. Appl.*, 14:327–371, 2008.

[22] K. Guo and D. Labate. The construction of smooth parseval frames of shearlets. 2012. preprint.

[23] K. Guo and D. Labate. Optimally sparse representations of $3D$ data with C^2 surface singularities using Parseval frames of shearlets. *SIAM J. Math. Anal.*, 44:851–886, 2012.

[24] P. Kittipoom, G. Kutyniok, and W.-Q. Lim. Construction of compactly supported shearlet frames. *Constr. Approx.*, 35(1):21–72, 2012.

[25] G. Kutyniok and D. Labate, editors. *Shearlets: Multiscale Analysis for Multivariate Data.* Birkhäuser, Boston, 2012.

[26] G. Kutyniok, J. Lemvig, and W.-Q. Lim. Compactly supported shearlet frames and optimally sparse approximations of functions in $L^2(\mathbb{R}^3)$ with piecewise C^α singularities. *SIAM J. Math. Anal.*, to appear.

[27] G. Kutyniok and W.-Q. Lim. Compactly supported shearlets are optimally sparse. *J. Approx. Theory*, 163(11):1564–1589, 2011.

[28] D. Labate, L. Mantovani, and P. Negi. Shearlet smoothness spaces. 2012. preprint.

[29] H. Smith. A parametrix construction for wave equations with $C^{1,1}$-coefficients. *Ann. Inst. Fourier*, 48:797–835, 1998.

[30] L. Ying, L. Demanet, and E. Candès. 3D discrete curvelet transform. In *Proc. SPIE wavelets XI, San Diego*, 2005.