ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Parabolic Molecules: Curvelets, Shearlets, and Beyond

P. Grohs and S. Keiper and G. Kutyniok and M. Schaefer

Research Report No. 2013-47
November 2013

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Parabolic Molecules: Curvelets, Shearlets, and Beyond

Philipp Grohs, Sandra Keiper, Gitta Kutyniok, and Martin Schäfer

Abstract Anisotropic representation systems such as curvelets and shearlets have had a significant impact on applied mathematics in the last decade. The main reason for their success is their superior ability to optimally resolve anisotropic structures such as singularities concentrated on lower dimensional embedded manifolds, for instance, edges in images or shock fronts in solutions of transport dominated equations. By now, a large variety of such anisotropic systems has been introduced, for instance, second generation curvelets, bandlimited shearlets, and compactly supported shearlets, all based on a parabolic dilation operation. These systems share similar approximation properties, which is usually proven on a case-by-case basis for each different construction. The novel concept of parabolic molecules, which was recently introduced by two of the authors, allows for a unified framework encompassing all known anisotropic frame constructions based on parabolic scaling. The main result essentially states that all such systems share similar approximation properties. One main consequence is that at once all the desirable approximation properties of one system within this framework can be deduced for virtually any other system based on parabolic scaling. The present paper motivates and surveys recent results in this direction.

1 Introduction

Wavelets have had a tremendous impact on applications requiring an efficient representation system such as image compression or PDE solvers. However, multivariate

Philipp Grohs
ETH Zürich, Seminar for Applied Mathematics, 8092 Zürich, Switzerland, e-mail: philipp.grohs@sam.math.ethz.ch

Sandra Keiper • Gitta Kutyniok • Martin Schäfer
Technische Universität Berlin, Department of Mathematics, 10623 Berlin, Germany, e-mail: keiper@math.tu-berlin.de; kutyniok@math.tu-berlin.de; schaefer@math.tu-berlin.de

data does typically exhibit the distinct property of being governed by anisotropic features, which wavelets – as an isotropic system – are not capable of resolving optimally in the sense of optimal approximation rates. In imaging sciences, this fact is even backed up by neurophysiology, since it is today generally accepted that neurons are highly directional based, thereby reacting most strongly to curvelike structures.

This observation has led to the introduction of various novel representation systems, which are designed to accommodate the anisotropic nature of most multivariate data. The considered model situation are functions with singularities along lower dimensional embedded manifolds such as edges or rays in imaging applications, with the goal to provide optimally sparse approximations of these objects. Some of the most well-known nowadays termed *directional representation systems* are ridgelets [4], curvelets [5], and shearlets [19, 28]. With the introduction of such a variety of systems the appeal has grown to extract the underlying principles of these new constructions and build an abstract common framework, which can unite many of these systems 'under one roof'. The framework should be general enough to include as many constructions as possible, while on the other hand also be specific enough to still capture their main features and properties. Such a framework would help to gain deeper insights into the properties of such systems. Moreover, it bears an obvious economical advantage. Up to now the properties of each new system, e.g. their approximation rates of anisotropic features, have been proven more or less from scratch, although the proofs often resemble one another in many ways. From the higher level viewpoint provided by such a framework, it becomes possible to provide proofs, which build upon abstract properties and are therefore independent of the specific constructions. Thus, results can be established for many systems simultaneously.

The introduction of *parabolic molecules* in 2011 by two of the authors [17] was a first step in this direction. A system of parabolic molecules can be regarded as being generated from a set of functions via parabolic dilations, rotations and translations. Each element in a system of parabolic molecules is therefore naturally associated with a certain scale, orientation and spatial location. The central conceptual idea is now to allow the generators to vary, as long as they obey a prescribed time-frequency localization, which also explains the terminology 'molecules'.

At the heart of this is the fundamental observation that it is foremost the time-frequency localizations of the functions in a system, which determine its properties and performance. This concept of *variable generators*, where in the extreme case every element is allowed to have its own individual generator, is a key feature of the framework and gives it a great amount of flexibility. Additional flexibility is achieved by *parametrizations* to allow generic indexing of the elements. Another fruitful idea is the relaxation of the rigid vanishing moment conditions imposed on the generators of most classical constructions by requiring the moments to only *vanish asymptotically* at high scales without changing the asymptotic behavior of the approximation.

It was shown in [17] that the concept of parabolic molecules can unify shear-based and rotation-based constructions under one roof. In particular, it enables to treat the classical shearlets and curvelets simultaneously, although these specific

constructions are based on different construction 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. As an example application, in [17] parabolic molecules were used to show that these systems feature a similar approximation behavior, thereby not only unifying the approximation results for curvelets [5] and shearlets [20, 27], but proving optimal sparse approximations for a much larger class of systems belonging to the class of parabolic molecules.

Our exposition is organized as follows. We begin with a general introduction to the problem of sparsely representing multivariate data in Section 2. The main issue with such data is the possible occurrence of anisotropic phenomena, which impair the otherwise good performance of classical wavelet systems. This motivates the need for so-called directional representation systems, some classical constructions of which we present in the next Section 3, namely classical curvelets and shearlets. Here we emphasize their similar approximation performance, which is almost optimal for cartoon-like images.

After this exposition we turn to parabolic molecules as a unifying framework. We first establish the basic concepts in Section 4 and state one main result, namely that the cross-Gramian of two systems of parabolic molecules exhibits a strong off-diagonal decay. This property will become essential in Section 6, where we discuss the approximation behavior of parabolic molecules. Before moving there, however, we pause for a while in Section 5 to illustrate the versatility of the framework by giving some examples. After we have convinced the reader of their applicability, we then turn to the section on approximation, where we essentially prove that any two systems of parabolic molecules, which are consistent and have sufficiently high order, exhibit the same approximation behavior.

2 Representation of Multivariate Data

Most applications require efficient encoding of multivariate data in the sense of optimal (sparse) approximation rates by a suitable representation system. This is typically phrased as a problem of best N-term approximation (see Subsection 2.1). The performance of an approximation scheme is then usually analyzed with respect to certain subclasses of the Hilbert space $L^2(\mathbb{R}^d)$, which is the standard continuum domain model for d-dimensional data, in particular, in imaging science. As elaborated upon before, the key feature of most multivariate data is the appearance of anisotropic phenomena. Hence such a subclass of $L^2(\mathbb{R}^d)$ is required to provide a suitable model for this fact, which, for $d = 2$, is fulfilled by the subclass of so-called cartoon-like images as introduced in Subsection 2.2. It can then be easily seen that wavelets do not deliver optimal approximation rates (Subsection 2.3), which then naturally leads to the theory of directional representation systems.

In the sequel, we will use the 'analyst's brackets' $\langle x \rangle := \sqrt{1+x^2}$ for $x \in \mathbb{R}$. Also, for two quantities $A, B \in \mathbb{R}$, which may depend on several parameters we shall write $A \lesssim B$, if there exists a constant $C > 0$ such that $A \leq CB$, uniformly in the parameters. If the converse inequality holds true, we write $A \gtrsim B$ and if both inequalities hold we shall write $A \asymp B$.

2.1 Sparse Approximation

Let us start by briefly discussing some aspects of approximation theory. From a practical standpoint, a function $f \in L^2(\mathbb{R}^2)$ is a rather intractable object. In order to analyze f, the most common approach is to represent it with respect to some representation system $(m_\lambda)_{\lambda \in \Lambda} \subseteq L^2(\mathbb{R}^2)$, i.e., to expand f as

$$f = \sum_{\lambda \in \Lambda} c_\lambda m_\lambda, \tag{1}$$

and then consider the coefficients $c_\lambda \in \mathbb{R}$. In practice we have to account for noise, hence it is necessary to ensure the robustness of such a representation. This leads to the notion of a frame (cf. [9, 8]).

A frame is a generalization of the notion of an orthonormal basis to include redundant systems, while still ensuring stability. More precisely, a system $(m_\lambda)_{\lambda \in \Lambda} \subseteq L^2(\mathbb{R}^2)$ forms a *frame* for $L^2(\mathbb{R}^2)$, if there exist constants $0 < A \leq B < \infty$ such that

$$A\|f\|_2^2 \leq \sum_{\lambda \in \Lambda} |\langle f, m_\lambda \rangle|^2 \leq B\|f\|_2^2 \quad \text{for all } f \in L^2(\mathbb{R}^2).$$

A frame is called *tight*, if $A = B$ is possible, and *Parseval*, if $A = B = 1$. Since the *frame operator* $S : L^2(\mathbb{R}^2) \to L^2(\mathbb{R}^2)$ defined by $Sf = \sum_{\lambda \in \Lambda} \langle f, m_\lambda \rangle m_\lambda$ is invertible, it follows that one sequence of coefficients in (1) – note that for a redundant system this sequence is not unique anymore – can be computed as

$$c_\lambda = \langle f, S^{-1} m_\lambda \rangle, \quad \lambda \in \Lambda,$$

where $(S^{-1} m_\lambda)_\lambda$ is usually referred to as the *canonical dual frame*. This particular coefficient sequence has the distinct property that it minimizes the ℓ_2-norm.

When representing f with respect to a frame $(m_\lambda)_\lambda \subseteq L^2(\mathbb{R}^2)$, we are confronted with yet another problem. Since in real world applications infinitely many coefficients are infeasible, the function f has to be approximated by a finite subset of this system. Letting N be the number of elements allowed in this approximation, we obtain what is called an *N-term approximation* for f with respect to $(m_\lambda)_\lambda$. The *best N-term approximation*, typically denoted by f_N, is optimal among those in terms of a minimal approximation error and is defined by

$$f_N = \operatorname*{argmin}_{(c_\lambda)_{\lambda \in \Lambda_N}} \|f - \sum_{\lambda \in \Lambda_N} c_\lambda m_\lambda\|_2^2 \quad \text{subject to} \quad \#\Lambda_N \leq N.$$

An appropriate measure for the approximation behavior of a system $(m_\lambda)_\lambda$ for a subclass $\mathscr{C}$, say, of $L^2(\mathbb{R}^2)$ is the decay of the L^2-error of the best N-term approximation $\|f - f_N\|_2$ as $N \to \infty$, thus the *asymptotic approximation rate*. As discussed before, 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. The delicacy of this problem can for instance be seen in [13]. A typical approach to circumvent this problem is to consider instead the N-term approximation by the N largest coefficients $(c_\lambda)_{\lambda\in\Lambda}$. It is evident that this error also provides a bound for the error of best N-term approximation.

There indeed exists a close relation between the N-term approximation rate achieved by a frame, and the decay rate of the corresponding frame coefficients. By measuring this decay rate in terms of the ℓ_p-(quasi)-norms for $p > 0$, the following lemma shows that membership of the coefficient sequence to an ℓ_p-space for small p implies 'good' N-term approximation rates. For the proof, we refer to [10, 27].

Lemma 1. *Let $f = \sum c_\lambda m_\lambda$ be an expansion of $f \in L^2(\mathbb{R}^2)$ with respect to a frame $(m_\lambda)_{\lambda\in\Lambda}$. Further, assume that the coefficients satisfy $(c_\lambda)_\lambda \in \ell^{2/(2k+1)}$ for some $k > 0$. Then the best N-term approximation rate is at least of order N^{-k}, i.e.*

$$\|f - f_N\|_2 \lesssim N^{-k}.$$

2.2 Image Data and Anisotropic Phenomena

To model the fact that multivariate data appearing in applications is typically governed by anisotropic features – in the 2-dimensional case curvilinear structures –, the so-called *cartoon-like functions* were introduced in [11]. This class is by now widely used as a standard model in particular for natural images. It mimics the fact that natural images often consist of nearly smooth parts separated by discontinuities, which is illustrated in Figure 2.2.

The first rigorous mathematical definition was given in [11] and extensively employed starting from the work in [5]. It postulates that images consist of $C^2(\mathbb{R}^2)$-regions separated by smooth $C^2(\mathbb{R})$-curves. This leads to the next definition (see also Figure 2.2).

Definition 1. The *class $\mathscr{E}^2(\mathbb{R}^2)$ of cartoon-like functions* is the set of functions $f : \mathbb{R}^2 \to \mathbb{C}$ of the form

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

where $B \subset [0,1]^2$ is a set with ∂B being a continuous and piecewise C^2-curve with bounded curvature and $f_i \in C^2(\mathbb{R}^2)$ are functions with supp $f_0 \subset [0,1]^2$ and $\|f_i\|_{C^2} \leq 1$ for each $i = 0, 1$.

We remark that by now several extensions of this model have been introduced and studied, starting with the extended model in [26].

(1) (2a) (2b)

Fig. 1 (1): Illustration of the appearance of 'cartoon-like parts' in natural images. (2): Illustration of the fact that the human brain is able to deduce the image (2a) just from its 'cartoon-like' ingredients (2b).

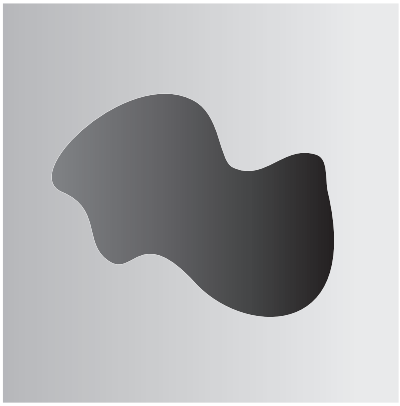


Fig. 2 Example of a cartoon-like function.

Having agreed on a suitable subclass of functions, one might now ask whether there exists a maximal asymptotic approximation rate leading to a notion of optimality. Indeed, such a benchmark result was derived by Donoho in [11].

Theorem 1 ([11]). *Let* $(m_\lambda)_{\lambda\in\Lambda} \subseteq L^2(\mathbb{R}^2)$. *Under the assumption of polynomial depth search for the representation coefficients used in the N-term approximation, the associated asymptotic approximation rate of some* $f \in \mathscr{E}^2(\mathbb{R}^2)$ *satisfies*

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

It is in this sense that a system satisfying this approximation rate is coined to deliver *optimally sparse approximations*.

2.3 2D Wavelet Systems

Nowadays, wavelet systems are widely utilized representation systems both for theoretical purposes as well as for engineering applications, for instance for the decomposition of elliptic operators or for the detection of anomalies in signals. Their success stems from the fact that wavelets deliver optimal sparse approximations for data being governed by isotropic features – which is in particular the case for elliptic operator equations whose solutions may exhibit point singularities (for instance if re-entrant corners are present in the computational domain) as well as in the 1-dimensional setting – and from the fast numerical realization of the wavelet transform.

Let us first recall a certain type of wavelet system in $L^2(\mathbb{R}^2)$, obtained by the following tensor product construction, see e.g. [30] for details. Starting with a given multiresolution analysis of $L^2(\mathbb{R})$ with scaling function $\phi^0 \in L^2(\mathbb{R})$ and wavelet $\phi^1 \in L^2(\mathbb{R})$, for every index $e = (e_1, e_2) \in E$, $E = \{0,1\}^2$, the generators $\psi^e \in L^2(\mathbb{R}^2)$ are defined as the tensor products

$$\psi^e = \phi^{e_1} \otimes \phi^{e_2}.$$

Definition 2. Let $\phi^0, \phi^1 \in L^2(\mathbb{R})$ and $\psi^e \in L^2(\mathbb{R}^2)$, $e \in E$, be defined as above. For fixed sampling parameters $q > 1$, $\tau > 0$ we define the *discrete wavelet system*

$$\begin{aligned} W(\phi^0, \phi^1; q, \tau) = &\left\{ \psi^{(0,0)}(\cdot - \tau k) \,:\, k \in \mathbb{Z}^2 \right\} \\ &\cup \left\{ q^j \psi^e(q^j \cdot - \tau k) \,:\, e \in E \backslash \{(0,0)\},\, j \in \mathbb{N}_0,\, k \in \mathbb{Z}^2 \right\}. \end{aligned}$$

The associated index set is given by

$$\Lambda^w = \left\{ ((0,0),0,k) \,:\, k \in \mathbb{Z}^2 \right\} \cup \left\{ (e,j,k) \,:\, e \in E \backslash \{(0,0)\},\, j \in \mathbb{N}_0,\, k \in \mathbb{Z}^2 \right\}.$$

Next, we recall the definition of vanishing moments for univariate wavelets, which says that the associated wavelet system annihilates polynomials up to some degree.

Definition 3. A function $g \in L^2(\mathbb{R})$ is said to possess M vanishing moments, if

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

It is well known that this property can be characterized by polynomial decay near zero of the associated Fourier transform. For the convenience of the reader, we provide the short proof.

Lemma 2. *Suppose that $g \in L^2(\mathbb{R}) \cap C(\mathbb{R})$ is 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, $g \in L^1(\mathbb{R})$ and hence $\hat{g}$ is bounded. This shows that the claimed inequality holds for $|\xi| \leq 1$.

Let now $\xi \in R$ satisfy $|\xi| > 1$. For this, observe that, up to a constant,

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

Since g possesses M vanishing moments, it follows that all derivatives of order $k < M$ of $\hat{g}$ vanish at 0. Furthermore, since g is compactly supported, its Fourier transform is analytic. Thus

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

which proves the claim. □

We now assume that $\phi^0, \phi^1 \in L^2(\mathbb{R})$ satisfy $\widehat{\phi^0}, \widehat{\phi^1} \in C^\infty(\mathbb{R})$ and that there are $0 < a$ and $0 < b < c$ such that

$$\text{supp } \widehat{\phi^0} \subset [-a,a] \quad \text{and} \quad \text{supp } \widehat{\phi^1} \subset [-c,c]\backslash[-b,b].$$

These conditions are fulfilled, for instance, if $\phi^0, \phi^1 \in L^2(\mathbb{R})$ are the generators of a Lemarié-Meyer wavelet system. In this case, it is well-known that the associated tensor product wavelets are indeed suboptimal for approximation of anisotropic features modeled by cartoon-like functions.

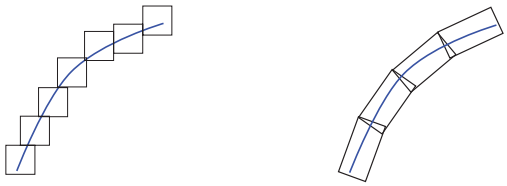
Theorem 2. *For $f \in \mathscr{E}^2(\mathbb{R}^2)$, the wavelet system $W\left(\phi^0, \phi^1; a, b\right)$ provides an asymptotic L^2-error of best N-term approximation given by*

$$\|f - f_N\|_2^2 \asymp N^{-1}, N \to \infty.$$

3 Directional Representation Systems

The reason for the failure of wavelets to provide optimally sparse approximations of cartoon-like functions is the fact that wavelets are inherently *isotropic* objects and thus not optimally suited for approximating *anisotropic* objects. To overcome this problem, in recent years various directional representation systems were introduced, among which are ridgelets, curvelets, and shearlets, to name just a few. Their main advantage lies in their anisotropic support, which is much better suited to align with curvilinear structures (see Figure 3), thereby already intuitively promoting a fast error decay of the best N-term approximation.

In this section, we now first introduce the second generation curvelet system, which was in fact also the first system to provide (almost) optimally sparse approximations of cartoon-like functions (cf. Subsection 3.1). This is followed by a discussion of different versions of shearlets in Subsection 3.2.



(1) (2)

Fig. 3 Approximation of a curve by isotropic-shaped (1) and anisotropic-shaped (2) elements.

3.1 Second Generation Curvelets

Second generation curvelets were introduced in 2004 by Candès and Donoho in the seminal work [5]. It is this curvelet system which is referred to today when curvelets are mentioned. The anisotropy of these systems is induced into this system by enforcing a parabolic scaling so that the shape of the support essentially follows the parabolic scaling law '$length^2 \approx width$'. Intuitively, this seems a compromise between the isotropic scaling, as utilized for wavelets, and scaling in only one coordinate direction, as utilized for ridgelets. However, the reason is much deeper, since this law is particularly suited for approximating C^2-singularity curves, which is the type of curves our model is based upon.

We now describe the original construction. For this, let W and V be two window functions, which are both real, nonnegative, C^∞, and supported in $\left(\frac{1}{2},2\right)$ and in $(-1,1)$, respectively. We further require that these windows satisfy

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

For every scale $j \geq 0$, we now define the functions $\gamma_{(j,0,0)}$ in polar coordinates by

$$\hat{\gamma}_{(j,0,0)}(r,\omega) := 2^{-3j/4} W\left(2^{-j} r\right) V\left(2^{\lfloor j/2 \rfloor}\omega\right).$$

For $j \in \mathbb{Z}$ and $\theta \in \mathbb{T}$, the parabolic scaling matrix A_j and the rotation matrix R_θ are defined by

$$A_j := \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix} \quad \text{and} \quad R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

The definition of curvelets then reads as follows:

$$\gamma_{(j,\ell,k)}(\cdot) := \gamma_{(j,0,0)}\left(R_{\theta_{j,\ell}} \cdot - x_{j,k}\right),$$

where $\theta_{j,\ell} = \ell 2^{-\lfloor j/2 \rfloor}\pi$, $x_{j,k} = A_j^{-1} k$, and $(j,\ell,k) \in \Lambda^0$ with set of curvelet indices given by

$$\Lambda^0 := \left\{ (j,\ell,k) \in \mathbb{Z}^4 : j \geq 0, \quad \ell = -2^{\lfloor j/2 \rfloor - 1}, \ldots, 2^{\lfloor j/2 \rfloor - 1} \right\}. \tag{2}$$

With appropriate modifications for the low-frequency case $j = 0$ – for details we refer to [7] –, the system

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

constitutes a Parseval frame for $L^2(\mathbb{R}^2)$, which is customarily referred to as the frame of *second generation curvelets*. When identifying frame elements oriented in antipodal directions, this system becomes a frame with real-valued elements.

Let us next discuss the approximation properties of Γ^0 proved in [5]. Ignoring log-like factors, this frame indeed attains the optimal achievable approximation rate for the class of cartoon-like functions $\mathscr{E}^2(\mathbb{R}^2)$. Moreover, this rate is achieved by simple thresholding, which is even more surprising, since this approximation scheme is intrinsically non-adaptive.

Theorem 3 ([5]). *The second generation curvelet frame Γ^0 provides (almost) optimal sparse approximations of cartoon-like functions $f \in \mathscr{E}^2(\mathbb{R}^2)$, i.e.,*

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

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

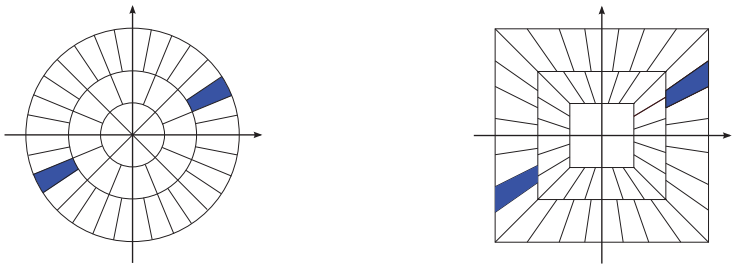
The implicit constant in (3) only depends on the maximal curvature of the singularity curve of f, the number of corner points, and the minimal opening angle in the corners. In particular, the approximation rate is uniform over all functions whose singularity curve has maximal curvature bounded by a fixed constant.

Finally we remark that due to the construction the frame elements of Γ^0 are bandlimited functions. Up to now no constructions of compactly supported curvelets are known.

3.2 Shearlet Systems

Shearlets were introduced in 2006 [19] as the first directional representation system which not only satisfies the same celebrated properties of curvelets, but is also more adapted to the digital realm. In fact, shearlets enable a unified treatment of the continuum and digital setting, which allows implementations faithful to the continuum domain theory. This key property is achieved through utilization of a shearing matrix instead of rotations as a means to parameterize orientation, thereby preserving the structure of the integer grid. The resulting different tilings of frequency domain are illustrated in Figure 4.

We next introduce a selection of the variety of available shearlet systems, namely bandlimited shearlets (Subsection 3.2.1), the so-called smooth Parseval frames of shearlets (Subsection 3.2.3), and compactly supported shearlets (Subsection 3.2.2).



(1) (2)

Fig. 4 Frequency tiling induced by a curvelet system (1) and a shearlet system (2).

For a more detailed exposition of shearlets than given below, we refer to the book [28].

3.2.1 Bandlimited Shearlets

We first present the classical cone-adapted shearlet construction of band-limited shearlets presented in [19]. It is worth emphasizing that due to the shearing operator, the frequency domain needs to be split into four cones to ensure an almost uniform treatment of the different directions, which comes naturally for rotation as a means to change the orientation (compare Figure 4).

First, let $\psi_1,\ \psi_2 \in L^2(\mathbb{R})$ be chosen such that

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

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

and

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

Then the classical mother shearlet ψ is defined by

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

For $j,\ell \in \mathbb{Z}$ let now the parabolic scaling matrix A_j and the shearing matrix S_ℓ be defined by

$$A_j := \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix} \quad \text{and} \quad S_\ell := \begin{pmatrix} 1 & \ell \\ 0 & 1 \end{pmatrix}.$$

Further, for a domain $\Omega \subset \mathbb{R}^2$ let us define the space

$$L^2(\Omega)^\vee := \left\{ f \in L^2(\mathbb{R}^2) : \text{ supp } \widehat{f} \subset \Omega \right\}.$$

It was then shown in [19] that the system

$$\Sigma^0 := \left\{ 2^{3j/4} \psi(S_\ell A_j \cdot - k) : \ j \geq 0, \ \ell = -2^{\lfloor j/2 \rfloor}, \cdots, 2^{\lfloor j/2 \rfloor}, k \in \mathbb{Z}^2 \right\}$$

constitutes a Parseval frame for the Hilbert space $L^2(\mathscr{C})^\vee$ on the frequency cone

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

By reversing the coordinate axes, also a Parseval frame Σ^1 for $L^2(\mathscr{C}')^\vee$, where

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

can be constructed. Finally, we can consider a Parseval frame

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

for the Hilbert space $L^2\left(\left[-\frac{1}{8}, \frac{1}{8} \right]^2 \right)^\vee$. Combining those systems, we obtain the *bandlimited shearlet frame*

$$\Sigma := \Sigma^0 \cup \Sigma^1 \cup \Phi.$$

In [20], it was shown that bandlimited shearlet frames achieve (almost) optimal sparse approximations for elements of $\mathscr{E}^2(\mathbb{R}^2)$, similar to curvelets and in fact even with the *same* log-like factor.

Theorem 4 ([20]). *The bandlimited shearlet frame Σ provides (almost) optimal sparse approximations of cartoon-like functions $f \in \mathscr{E}^2(\mathbb{R}^2)$, i.e.,*

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

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

3.2.2 Smooth Parseval Frames of Shearlets

Following [22], a slight modification of the bandlimited shearlet construction, namely by carefully glueing together boundary elements along the seamlines with angle $\pi/4$, yields a Parseval frame with smooth and well-localized elements.

3.2.3 Compactly Supported Shearlets

In 2011, compactly supported shearlets were introduced by one of the authors and her collaborators in [27]. Currently known constructions of compactly supported shearlets involve separable generators, i.e.,

$$\psi(x_1,x_2) := \psi_1(x_1)\psi_2(x_2), \quad \widetilde{\psi}(x_1,x_2) := \psi(x_2,x_1). \tag{4}$$

with a wavelet ψ_1 and a scaling function ψ_2. Following [27], the cone-adapted discrete shearlet system is then defined as follows, where $A_j := \operatorname{diag}(2^j, 2^{j/2})$ as before and $\widetilde{A}_j := \operatorname{diag}(2^{j/2}, 2^j)$.

Definition 4. For some fixed sampling parameter $c > 0$, the *cone-adapted discrete shearlet system* $SH\big(\phi,\psi,\widetilde{\psi};c\big)$ generated by $\phi,\psi,\widetilde{\psi} \in L^2(\mathbb{R}^2)$ is defined by

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

where

$$\begin{aligned}
\Phi(\phi;c) &= \{\sigma_k = \phi(\cdot - k) : k \in c\mathbb{Z}^2\},\\
\Psi(\psi;c) &= \big\{\sigma_{j,\ell,k} = 2^{3j/4}\psi(S_\ell A_j \cdot - k) : j \geq 0, |\ell| \leq \lceil 2^{j/2}\rceil, k \in c\mathbb{Z}^2\big\},\\
\widetilde{\Psi}(\widetilde{\psi};c) &= \big\{\widetilde{\sigma}_{j,\ell,k} = 2^{3j/4}\widetilde{\psi}(S_\ell^T \widetilde{A}_j \cdot - k) : j \geq 0, |\ell| \leq \lceil 2^{j/2}\rceil, k \in c\mathbb{Z}^2\big\}.
\end{aligned}$$

Under certain assumptions on $c,\psi,\widetilde{\psi}$ this shearlet system forms a frame with controllable frame bounds [24].

In [27], it was shown that compactly supported shearlet frames, under assumptions on the separable behavior and the directional vanishing moments of the generators, also achieve (almost) optimal sparse approximations for elements of $\mathscr{E}^2(\mathbb{R}^2)$.

Theorem 5 ([27]). *Let $c > 0$ and let $\phi,\psi,\widetilde{\psi} \in L^2(\mathbb{R}^2)$ be compactly supported. Suppose that, in addition, for all $\xi = (\xi_1,\xi_2) \in \mathbb{R}^2$, the shearlet ψ satisfies*

(i) $|\hat{\psi}(\xi)| \leq C_1 \min(1,|\xi_1|^\alpha)\min(1,|\xi_1|^{-\gamma})\min(1,|\xi_2|^{-\gamma})$ *and*
(ii) $\left|\frac{\partial}{\partial \xi_2}\hat{\psi}(\xi)\right| \leq |h(\xi_1)|\left(1+\frac{\xi_2}{\xi_1}\right)^{-\gamma}$,

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

Then the shearlet frame $SH(\phi,\psi,\widetilde{\psi};c)$ provides (almost) optimal sparse approximations of cartoon-like functions $f \in \mathscr{E}^2(\mathbb{R}^2)$, i.e.,

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

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

With this theorem we end our presentation of directional representation systems, although there do exist more constructions. It is a striking fact, that the three presented examples all exhibit the same approximation behavior, although they are construction-wise quite different. The framework of parabolic molecules, which we will present in the subsequent sections, will reveal the fundamental common ingredients in these systems which ensure (almost) optimal sparse approximations of cartoon-like functions.

4 Parabolic Molecules

The concept of parabolic molecules took shape by distilling the essential principles, which underly many of the newly constructed directional representation systems, in particular curvelets and shearlets. It provides a framework, which comprises many of these classic systems, and allows the design of new constructions with pre-defined approximation properties.

Moreover, the approximation properties of some new system are usually proven more or less from scratch. By adopting the higher level viewpoint of time-frequency localization, the parabolic molecule framework is very general and independent of specific constructions. This has the advantage, that it enables a unified treatment of many systems. In particular, it can be used to establish approximation results for many systems simultaneously.

A system of parabolic molecules consists of functions, obtained from a set of generators via parabolic dilations, rotations and translations. Similar to curvelets, each function in a system of parabolic molecules is therefore naturally associated with a certain scale, orientation and spatial location.

A central feature of the framework, which explains the terminology ‘molecules’, is the concept of variable generators: In order to gain flexibility the generators are allowed to vary, as long as they obey a prescribed time-frequency localization. At the heart of this is the fundamental observation that it is foremost the time-frequency localization, which determines the approximation properties and performance of a system.

A nice side-effect of this less rigid construction principle is the fact that the strict vanishing moment conditions, usually imposed on the generators of classical constructions, can be relaxed without changing the asymptotic approximation behavior of the system. It suffices to require the moments to vanish asymptotically at high scales.

4.1 Definition of Parabolic Molecules

Let us now delve into the details of the framework of parabolic molecules. A system of parabolic molecules is a family of functions $(m_\lambda)_{\lambda\in\Lambda}$ obtained from a set of

generators via parabolic dilations, rotations and translations. Each function m_λ is therefore associated with a unique point in the parameter space $\mathbb{P}$, sometimes also referred to as phase space, given by

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

where a point $p = (s, \theta, x) \in \mathbb{P}$ specifies a scale $2^s \in \mathbb{R}_+$, an orientation $\theta \in \mathbb{T}$, and a location $x \in \mathbb{R}^2$.

The relation between the index λ of a molecule m_λ and its location $(s_\lambda, \theta_\lambda, x_\lambda)$ in the parameter space $\mathbb{P}$ is described via so-called parametrizations.

Definition 5. A *parametrization* consists of a pair (Λ, Φ_Λ), where Λ is a discrete index set and Φ_Λ is a mapping

$$\Phi_\Lambda : \Lambda \to \mathbb{P}, \quad \lambda \mapsto (s_\lambda, \theta_\lambda, x_\lambda),$$

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

By using parametrizations, the actual indices of the molecules can be de-coupled from their associated locations in $\mathbb{P}$. This gives the freedom to assign generic indices to the molecules, a feature, which is essential to include systems into the framework, whose constructions are based on different principles, like e.g. shearlet-like and curvelet-like systems. Another benefit of this approach is that a parameterization does not have to sample phase space in a regular fashion. The only property it needs to satisfy for our results to be applicable is consistency as defined below in Subsection 6.2.

Before defining parabolic molecules we fix the following notation. As defined in Section 3, let R_θ denote the rotation matrix by an angle θ, and A_j the parabolic scaling matrix associated with $j \geq 0$.

Definition 6. Let Λ be a parametrization. A family $(m_\lambda)_{\lambda \in \Lambda}$ of functions $m_\lambda \in L^2(\mathbb{R}^2)$ is called a *family of parabolic molecules* of order (R, M, N_1, N_2) if it can be written as

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

such that

$$\left|\partial^\beta \hat{a}^{(\lambda)}(\xi)\right| \lesssim \min\left(1, 2^{-s_\lambda} + |\xi_1| + 2^{-s_\lambda/2}|\xi_2|\right)^M \langle|\xi|\rangle^{-N_1} \langle\xi_2\rangle^{-N_2} \tag{5}$$

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

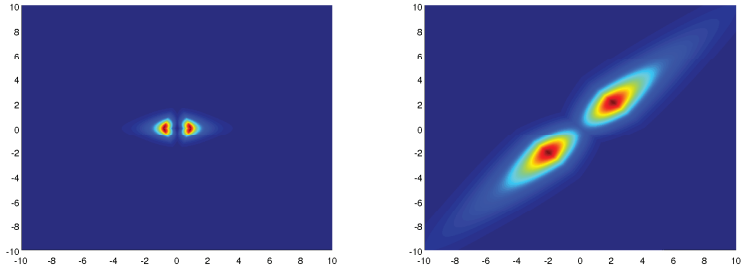
Remark 1. To simplify notation we did not explicitly refer to the utilized parametrization Φ_Λ.

Notice that a system of parabolic molecules $(m_\lambda)_{\lambda \in \Lambda}$ is generated by parabolically scaling, rotating and translating a set of generators $(a^{(\lambda)})_{\lambda \in \Lambda}$. In contrast to many classical constructions, where the set of generators is usually small, each

molecule is allowed to have its own individual generator. We only require these generators to uniformly obey a prescribed time-frequency localization.

Let us remark, that for convenience the time-frequency conditions in the definition are formulated on the Fourier side. Thus, the number R actually 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.

According to the definition the frequency support of a parabolic molecule is concentrated in a parabolic wedge associated to a certain orientation, and in the spatial domain its essential support lies in a rectangle with parabolic aspect ratio. For illustration purposes, the approximate frequency support of two parabolic molecules at different scales and orientations is depicted in Figure 5.



(1) (2)

Fig. 5 (1): 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 $s_\lambda = 3$, $M = 3$, $N_1 = N_2 = 2$. (2): Approximate Frequency support of a corresponding molecule $\hat{m}_\lambda$ with $\theta_\lambda = \pi/4$.

Changing into polar coordinates we obtain the representation

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

which directly implies the estimate

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

4.2 Index Distance

An essential ingredient for the theory is the fact that the parameter space $\mathbb{P}$ can be equipped with a natural (pseudo-)metric. It was first introduced by Hart-Smith [29], albeit in a different context, and is therefore sometimes termed the *Hart-Smith pseudo metric*. Later it was also used in [3].

Definition 7. Following [3, 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 $\lambda_0 = \operatorname{argmin}(s_\lambda, s_\mu)$ and $e_\lambda = (\cos(\theta_\lambda), \sin(\theta_\lambda))^\top$.

Remark 2. The notation $\omega(\lambda,\mu)$ is a slight abuse of notation, since ω is acting on $\mathbb{P}$. Therefore it should read

$$\omega(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))$$

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

Remark 3. We also mention that there is a slight inaccuracy in the above definition. 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. 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 $\{\varphi\}$ being 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 7, which avoids additional technicalities.

We point out that the Hart-Smith pseudo metric is not a distance in the strict sense, e.g. we have $\omega(\lambda,\lambda) = 1 \neq 0$. As we shall see later, it somehow measures the correlation of a pair of parabolic molecules associated to the corresponding points in $\mathbb{P}$. The following proposition, whose proof can be found in [3], collects some of its properties.

Proposition 1 ([3]). *For indices λ, μ, ν we have*

(i) Symmetry: $\omega(\lambda,\mu) \asymp \omega(\mu,\lambda)$.
(ii) Triangle Inequality: $d(\lambda,\mu) \leq C(\omega(\lambda,\nu) + \omega(\nu,\mu))$ for some constant $C > 0$.
(iii) Composition: For every integer $N > 0$ and some positive constant C_N it holds

$$\sum_\nu \omega(\lambda,\nu)^{-N} \omega(\nu,\mu)^{-N} \leq C_N \omega(\lambda,\mu)^{-N-1}.$$

4.3 Decay of the Cross-Gramian

Given two systems $(m_\lambda)_{\lambda\in\Lambda}$ and $(p_\mu)_{\mu\in\Delta}$ of parabolic molecules we are interested in the magnitudes of the cross-correlations $|\langle m_\lambda, p_\mu\rangle|$. A fast decay will be key to, for instance, transferring sparse approximation properties from one system of parabolic molecules to another.

The following theorem establishes a relation to the index distance on $\mathbb{P}$. It states, that a high distance of two indices can be interpreted as a low cross-correlation of the associated molecules. The proof is quite technical and we refer to [17] for the details.

Theorem 6 ([17]). *Let $(m_\lambda)_{\lambda\in\Lambda}$, $(p_\mu)_{\mu\in\Delta}$ be two systems of parabolic molecules of order (R,M,N_1,N_2) with*

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

Then

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

This result shows that the Gramian matrix between two systems of parabolic molecules satisfies a strong off-diagonal decay property and is in that sense very close to a diagonal matrix. In Section 6 we will present several immediate applications of this result, most notably for the approximation properties of parabolic molecules.

5 Examples of Parabolic Molecules

Before going deeper into the theory of parabolic molecules and further exploring their properties, we pause for a while and give some examples for illustration. This will give evidence about the versatility of the concept. In particular, we show that both rotation-based and shear-based constructions fit well into the framework. It will also be proven that earlier constructions, which also employ the 'molecule' concept, can be viewed as subclasses of the more general parabolic molecules.

5.1 Curvelet-Like Systems

We begin with the review of curvelet-like systems, i.e. constructions based on rotation. Due to their similar construction principles, it may not come as a surprise that second generation curvelets are instances of parabolic molecules. It is also easily verified that curvelet molecules as defined in [3] fall into this framework.

5.1.1 Second Generation Curvelets

We start by specifying the parametrization, which we utilize for fitting second generation curvelets into the framework of parabolic molecules.

Definition 8. Let

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

be the curvelet index from (2) and define $\Phi^0 : \Lambda^0 \to \mathbb{P}$ by

$$\Phi^0(j,\ell,k) := (j, \ell 2^{-\lfloor j/2 \rfloor}\pi, R_{-\theta_\lambda} A_{-s_\lambda} k).$$

Then (Λ^0, Φ^0) is called the *canonical parametrization*.

We next prove that the frame Γ^0 of second generation curvelets as defined in Subsection 3.1 forms a system of parabolic molecules of arbitrary order.

Proposition 2 ([17]). *The second generation curvelet frame Γ^0 constitutes a system of parabolic molecules of arbitrary order associated with the canonical parametrization.*

Proof. Let $\lambda \in \Lambda^0$. Due to rotation invariance, we may restrict ourselves to the case $\theta_\lambda = 0$. Therefore, denoting $\gamma_j := \gamma_{(j,0,0)}$, it is sufficient to prove that the function

$$a^{(\lambda)}(\cdot) := 2^{-3s_\lambda/4} \gamma_j \left(A_{s_\lambda}^{-1} \cdot \right)$$

satisfies (5) for (R, M, N_1, N_2) arbitrary. For this, first note that

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

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 $r = \sqrt{\xi_1^2 + \xi_2^2}$, $\omega = \arctan(\xi_2/\xi_1)$, which yields

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

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

By a straightforward calculation, all derivatives of α_j and β_j are bounded on the support of $\hat{a}^{(\lambda)}$ and uniformly in j. The proposition is proved. □

5.1.2 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 also this system is a special instance of the framework of parabolic molecules.

5.1.3 Borup and Nielsen's Construction

Another very similar construction has been given in [2]. 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.

5.1.4 Curvelet Molecules

The final concept of parabolic molecules had many predecessors. In [3] the authors also employed the idea of molecules and introduced the notion of *curvelet molecules*. It proved to be a useful concept for showing sparsity properties of wave propagators. Let us first give their exact definition.

Definition 9. 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

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

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 $M = 0, 1, \ldots$

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

This definition is similar to our definition of parabolic molecules, however with two crucial differences: First, (5) 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 (5) are more restrictive in the sense that it requires infinitely many nearly vanishing moments.

In fact, the following result can be proven using similar arguments as for Proposition 2.

Proposition 3 ([17]). *A system of curvelet molecules of regularity R constitutes a system of parabolic molecules of order $(\infty, \infty, R/2, R/2)$.*

5.2 Shearlet-Like Systems

It is perhaps not surprising that curvelets and their relatives described above fall into the framework of parabolic molecules. However, we will next show that even shearlets as a very different directional representation system are examples of parabolic molecules. In this regard we would like to draw the reader's attention to the parametrization chosen for fitting shearlets into this framework.

5.2.1 Shearlet Molecules

Shearlet molecules as introduced in [17] provide a framework for shearlet-like systems in the spirit of curvelet molecules. For their definition we require the index set

$$\Lambda^\sigma := \left\{(\varepsilon, j, l, k) \in \mathbb{Z}_2 \times \mathbb{Z}^4 : \varepsilon \in \{0,1\},\ j \geq 0,\ \ell = -2^{\lfloor j/2 \rfloor}, \cdots, 2^{\lfloor j/2 \rfloor}\right\} \tag{6}$$

and generating functions $\phi, \psi_{j,\ell,k}, \widetilde{\psi}_{j,\ell,k} \in L^2(\mathbb{R}^2)$ for $(j,\ell,k) \in \Lambda^\sigma$. The associated shearlet system

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

is then defined by setting $\sigma_{(\varepsilon,0,0,k)}(\cdot) = \phi(\cdot - k)$ and for $j \geq 1$:

$$\begin{aligned}
\sigma_{(0,j,\ell,k)}(\cdot) &= 2^{3j/4}\psi_{j,\ell,k}\left(A_j S_{\ell,j} \cdot - k\right), \\
\sigma_{(1,j,\ell,k)}(\cdot) &= 2^{3j/4}\widetilde{\psi}_{j,\ell,k}\left(\widetilde{A}_j S_{\ell,j}^T \cdot - k\right).
\end{aligned}$$

Here $S_{\ell,j}$ denotes the shearing matrix

$$S_{\ell,j} := \begin{pmatrix} 1 & \ell 2^{-\lfloor j/2 \rfloor} \\ 0 & 1 \end{pmatrix}.$$

We proceed to 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 5.2.5.

Definition 10. We call Σ a system of *shearlet molecules* of order (R, M, N_1, N_2), if the functions $\psi_{j,\ell,k}$ satisfy

$$|\partial^\beta \hat{\psi}_{j,\ell,k}(\xi_1, \xi_2)| \lesssim \min\left(1, 2^{-j} + |\xi_1| + 2^{-j/2}|\xi_2|\right)^M \langle|\xi|\rangle^{-N_1} \langle\xi_2\rangle^{-N_2} \tag{7}$$

and

$$|\partial^\beta \hat{\phi}(\xi_1, \xi_2)| \lesssim \langle|\xi|\rangle^{-N_1} \langle\xi_2\rangle^{-N_2} \tag{8}$$

for every $\beta \in \mathbb{N}^2$ with $|\beta| \leq R$, and if the functions $\widetilde{\psi}_{j,\ell,k}$ satisfy (7) with the roles of ξ_1 and ξ_2 reversed.

Remark 4. In our proofs it is nowhere required that the directional parameter ℓ runs between $-2^{\lfloor j/2 \rfloor}$ and $-2^{\lfloor j/2 \rfloor}$. Indeed, ℓ running in any discrete interval $-C2^{\lfloor j/2 \rfloor}$, $\ldots$, $C2^{\lfloor j/2 \rfloor}$ 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 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 (7) and (8) are actually instances of parabolic molecules associated with a specific shearlet-adapted parametrization $(\Lambda^\sigma, \Phi^\sigma)$. This result shows that the concept of parabolic molecules is indeed a unification of in particular curvelet and shearlet systems.

Proposition 4 ([17]). *Assume that the shearlet system Σ constitutes a system of shearlet molecules of order (R, M, N_1, N_2). Then Σ forms a system of parabolic molecules of order (R, M, N_1, N_2), associated to the parametrization $(\Lambda^\sigma, \Phi^\sigma)$, where with $A_j^0 = A_j$, $A_j^1 = \widetilde{A}_j$, $S_{\ell,j}^0 = S_{\ell,j}$, $S_{\ell,j}^1 = S_{\ell,j}^T$ the map Φ^σ is given by*

$$\Phi^\sigma(\lambda) = (s_\lambda, \theta_\lambda, x_\lambda) := \left(j, \varepsilon\pi/2 + \arctan(-\ell 2^{-\lfloor j/2 \rfloor}), \left(S_{\ell,j}^\varepsilon\right)^{-1} \left(A_j^\varepsilon\right)^{-1} k \right).$$

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

$$a^{(\lambda)}(\cdot) := \psi\left(A_{s_\lambda} S_{\ell,s_\lambda} R_{\theta_\lambda}^T A_{-s_\lambda} \cdot \right)$$

satisfies (5).

We first observe that the Fourier transform of $a^{(\lambda)}$ is given by

$$\hat{a}^{(\lambda)}(\cdot) = \hat{\psi}\left(A_{-s_\lambda} S_{\ell,s_\lambda}^{-T} R_{\theta_\lambda}^T A_{s_\lambda} \cdot \right),$$

and the matrix $S_{\ell,s_\lambda}^{-T} R_{\theta_\lambda}^T$ has the form

$$S_{\ell,s_\lambda}^{-T} R_{\theta_\lambda}^T = \begin{pmatrix} \cos(\theta_\lambda) & \sin(\theta_\lambda) \\ 0 & -\ell 2^{-\lfloor s_\lambda/2 \rfloor} \sin(\theta_\lambda) + \cos(\theta_\lambda) \end{pmatrix} =: \begin{pmatrix} u & v \\ 0 & w \end{pmatrix}.$$

We next claim that the quantities u and w are uniformly bounded from above and below, independent of j, ℓ. To prove this claim, 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 elementary arguments show. In fact, this boundedness holds on any compact interval. We have

$$u = \tau\left(-\ell 2^{\lfloor s_\lambda/2 \rfloor}\right) \quad \text{and} \quad w = \rho\left(-\ell 2^{\lfloor s_\lambda/2 \rfloor}\right).$$

Since we are only considering indices with $\varepsilon = 0$, we have $\left|-\ell 2^{\lfloor s_\lambda/2 \rfloor}\right| \leq 1$, which now implies uniform upper and lower boundedness of the quantities u, w. Hence, there exist constants $0 < \delta_u \leq \Delta_u < \infty$ and $0 < \delta_w \leq \Delta_w < \infty$ such that for all j, ℓ it holds

$$\delta_u \leq u \leq \Delta_u \quad \text{and} \quad \delta_w \leq w \leq \Delta_w.$$

Observing that the matrix $A_{-s_\lambda} R_{\theta_\lambda}^T S_{\ell, s_\lambda}^{-T} A_{s_\lambda}$ has the form

$$\begin{pmatrix} u & 2^{-s_\lambda/2} v \\ 0 & w \end{pmatrix},$$

and by using the upper boundedness of u, v, w and the chain rule, for any $|\beta| \leq R$, we obtain

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

For the last estimate we utilized the moment estimate for $\hat{\psi}$, which is given by (7). This proves the moment property required in (5).

Finally, we need to show the decay of $\partial^\beta \hat{a}^{(\lambda)}$ for large frequencies ξ. Again, due to the fact that u, v, w are bounded from above and u, w from below, and utilizing the large frequency decay estimate in (7), we can estimate

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

The statement is proven. □

In the remainder of this section we examine the main shearlet constructions which are known today and show that they indeed fit into the framework of parabolic molecules.

5.2.2 Classical Shearlets

For the band-limited shearlet system Σ defined in Subsection 3.2.1, the following results can be shown using Proposition 4.

Proposition 5 ([17]). *The system $\Sigma := \Sigma^0 \cup \Sigma^1 \cup \Phi$ constitutes a shearlet frame which is a system of parabolic molecules of arbitrary order.*

It is also straightforward to check that the related Parseval frame constructed in [22] constitutes a system of parabolic molecules of arbitrary order.

5.2.3 Bandlimited Shearlets with Nice Duals

The bandlimited shearlet frame Σ as described above suffers from the fact that its dual frames are unknown. In particular, it is not known whether, in general, there exists a dual frame which also forms a system of parabolic molecules. In particular for applications, such a construction is however required. For general frames Σ of parabolic molecules it can be shown that the canonical dual frame Σ' constitutes a system of parabolic molecules of lower order [16]. However, the result of that paper is mostly of a qualitative nature and in particular it is difficult to compute the order of the dual frame for a given construction. In [15] this problem was successfully resolved by carefully glueing together the two bandlimited frames associated with the two frequency cones. The result in this paper in fact provides a construction of shearlet frames Σ with a dual frame Σ' such that both Σ and Σ' form systems of parabolic molecules of arbitrary order.

5.2.4 Compactly Supported Shearlets

Again by using the general result Proposition 4, it can be shown that the compactly supported shearlets as introduced in Subsection 3.2.3 also constitute a system of parabolic molecules, this time with the order being dependent in a more delicate way on the chosen generators.

Proposition 6 ([17]). *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 and $\widetilde{\psi}$ defined by (4), the associated shearlet system Σ constitutes a system of parabolic molecules of order (R,M,N_1,N_2).*

We remark that several assumptions on the generators ψ, $\widetilde{\psi}$ could be weakened, for instance the separability of the shearlet generators is not crucial for the arguments of the associated proof. More precisely, neither compact support nor bandlimitedness is necessary.

5.2.5 Shearlet Molecules of [21]

In [21] the results of [3] 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 11. Let Λ^σ be the shearlet index set as in (6) and A_j^ε, $S_{\ell,j}^\varepsilon$ be defined as in Proposition 4. A family $(m_\lambda)_{\lambda\in\Lambda^\sigma}$ is called a *family of shearlet molecules* of regularity R, if it can be written as

$$m_\lambda(x) = 2^{3s_\lambda/4} a^{(\lambda)} \left(A_j^\varepsilon S_{\ell,j}^\varepsilon x - k \right),$$

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 (7) with the choice of parameters $(R, N, N_1, N_2) = (\infty, \infty, R/2, R/2)$. Therefore, in view of Proposition 4, shearlet molecules of regularity R as defined in [21] form systems of parabolic molecules of order $(\infty, \infty, R/2, R/2)$.

Proposition 7 ([17]). *A system of shearlet molecules of regularity R constitutes a system of parabolic molecules of order $(\infty, \infty, R/2, R/2)$.*

6 Sparse Approximation with Parabolic Molecules

This section is devoted to one prominent application of the framework of parabolic molecules, and, in particular, the result of the decay of the cross-Gramian (Theorem 6), namely to sparse approximation behavior. This result will also show that the viewpoint of time-frequency localization as adopted by the framework of parabolic molecules provides the right angle to view questions of approximation behavior.

After introducing a measure for determining similar sparsity behavior, two main results will be presented: First, it will be shown that any two systems of parabolic molecules, which are consistent in a certain sense made precise later, of sufficiently high order exhibit the same approximation behavior. Second, by linking an arbitrary system to the curvelet frame, we obtain a 'stand-alone result' in the sense of sufficient conditions on the order of a system of parabolic molecules for providing (almost) optimally sparse approximations of cartoon-like functions.

6.1 Sparsity Equivalence

In light of Lemma 1, two frames should possess similar sparse approximation behavior, provided that the corresponding coefficient sequences have the same sparsity. This gave rise to the notion of sparsity equivalence from [17], which is a useful tool to compare such behavior. It is based on the close connection between the best N-term approximation rate of a frame and the ℓ_p-(quasi-)norm of the associated coefficient sequence.

Definition 12. Let $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ be systems of parabolic molecules and let $0 < p \leq 1$. Then $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ are *sparsity equivalent in ℓ_p*, if

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

Intuitively, systems of parabolic molecules being in the same sparsity equivalence class have similar approximation properties. This will subsequently be elaborated more deeply.

6.2 Consistency of Parametrizations

The next goal will be to find conditions which ensure that two systems of parabolic molecules are sparsity equivalent. It seems clear from an intuitive viewpoint that this requires some 'consistency' of the associated parametrizations. The next definition provides the correct notion for making this mathematically precise.

Definition 13. Two parametrizations (Λ, Φ_Λ) and (Δ, Φ_Δ) are called *k-consistent* for $k > 0$, if

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

In combination with Theorem 6, consistency is the essential tool to decide whether two frames of parabolic molecules are sparsity equivalent. We emphasize that although the original definition of systems of parabolic molecules does not require those system to form a frame, in the context of approximation theory, however, the frame property becomes important.

The following result states a sufficient condition for sparsity equivalence.

Theorem 7 ([17]). *Two frames $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ of parabolic molecules of order (R, M, N_1, N_2) with k-consistent parametrizations for some $k > 0$, are sparsity equivalent in ℓ_p, $0 < p \leq 1$, if*

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

Proof. By Schur's test, a well-known result from operator theory, we have

$$\left\| \left(\langle m_\lambda, p_\mu \rangle \right)_{\lambda \in \Lambda, \mu \in \Delta} \right\|_{\ell_p \to \ell_p} \leq \max \left(\sup_{\mu \in \Delta} \sum_{\lambda \in \Lambda} |\langle m_\lambda, p_\mu \rangle|^p, \sup_{\lambda \in \Lambda} \sum_{\mu \in \Delta} |\langle m_\lambda, p_\mu \rangle|^p \right)^{1/p}.$$

By Theorem 6, this implies that

$$\left\| \left(\langle m_\lambda, p_\mu \rangle \right)_{\lambda \in \Lambda, \mu \in \Delta} \right\|_{\ell_p \to \ell_p} \lesssim \max \left(\sup_{\mu \in \Delta} \sum_{\lambda \in \Lambda} \omega(\lambda, \mu)^{-k}, \sup_{\lambda \in \Lambda} \sum_{\mu \in \Delta} \omega(\lambda, \mu)^{-k} \right)^{1/p}.$$

But the term on the right hand side is finite, due to the k-consistency of the parametrizations (Λ, Φ_Λ) and (Δ, Φ_Δ). This proves that $(m_\lambda)_{\lambda\in\Lambda}$ and $(p_\mu)_{\mu\in\Delta}$ are sparsity equivalent in ℓ_p. □

Thus, as long as the parametrizations are consistent, the sparsity equivalence can be controlled by the order of the molecules.

In the remainder, we fix the frame of second generation curvelets Γ^0 from Section 3.1 as a reference frame. Recall that with respect to the canonical parametrization $(\Lambda^0, \Phi_{\Lambda^0})$, this frame constitutes a system of parabolic molecules, justifying the following definition.

Definition 14. A parametrization (Λ, Φ_Λ) is called *k-admissible* for $k > 0$, if it is k-consistent with the canonical parametrization $(\Lambda^0, \Phi_{\Lambda^0})$.

Before stating our main results, it seems natural to ask whether the curvelet and shearlet parametrization are k-admissible. This is the content of the next two lemmata.

Lemma 3 ([17]). *The canonical parametrization $(\Lambda^0, \Phi_{\Lambda^0})$ is k-admissible for all $k > 2$.*

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

$$\sum_{j\in\mathbb{Z}_+} \sum_{\lambda\in\Lambda^0, s_\lambda = j} 2^{-k|j-j'|} \left(1 + 2^{\min(j,j')} d(\mu,\lambda)\right)^{-k} < \infty. \tag{9}$$

By [3, Equation (A.2)], for any q, we have

$$\sum_{\lambda\in\Lambda^0, s_\lambda = j} (1 + 2^q d(\mu,\lambda))^{-2} \lesssim 2^{2(j-q)_+}. \tag{10}$$

Hence, for each $k > 2$, (9) can be estimated by

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

which finishes the proof. □

Lemma 4 ([17]). *The shearlet parametrization $(\Lambda^\sigma, \Phi^\sigma)$ is k-admissible for $k > 2$.*

Proof. The proof follows the same arguments as the proof of Lemma 3, except deriving the analogue to (10), i.e.,

$$\sum_{\lambda\in\Lambda^\sigma, s_\lambda = j} (1 + 2^q d(\mu,\lambda))^{-2} \lesssim 2^{2(j-q)_+}, \quad \text{for any } q \text{ and } \mu \in \Lambda^0, \tag{11}$$

requires a bit more work.

Without loss of generality we assume that $\theta_\mu = 0$ and $x_\mu = 0$. Also, we only restrict ourselves to the case $\varepsilon = 0$, the other case being exactly the same. In the

case $q > j$, the term on the left hand side of (11) can be bounded by a uniform constant.

Thus, it remains to deal with the $j \geq q$. Now we use the fact that, whenever $|\ell| \lesssim 2^{-j/2}$, we have

$$\left|\arctan\left(-\ell 2^{-\lfloor j/2 \rfloor}\right)\right| \gtrsim \left|\ell 2^{-\lfloor j/2 \rfloor}\right| \quad \text{and} \quad |S_\ell^{-1} A_{-j} k| \gtrsim |A_{-j} k|,$$

to estimate (11) by

$$\sum_\ell \sum_k \left(1 + 2^q \left(\left|\ell 2^{-\lfloor j/2 \rfloor}\right|^2 + \left|2^{-\lfloor j/2 \rfloor} k_2\right|^2 + \left|2^{-j} k_1 - \ell 2^{-\lfloor j/2 \rfloor} k_2 2^{-\lfloor j/2 \rfloor}\right|\right)\right)^{-2}.$$

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

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

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

6.3 Sparse Approximations

The next theorem now states the central fact that any system of parabolic molecules of sufficiently high order, whose parametrization is *k*-admissible, is sparsity equivalent to the second generation curvelet frame from Subsection 3.1. This theorem can interpreted as a means to transfer sparse approximation results from one system of parabolic molecules to another one, which is also key to Theorem 9.

Theorem 8 ([17]). *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 parabolic molecules associated with Λ of order (R, M, N_1, N_2) such that*

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

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

Recall that it was shown by Donoho in [11] (cf. Theorem 1) that (under natural conditions) the optimally achievable decay rate of the approximation error for the class $\mathscr{E}^2(\mathbb{R}^2)$ is given by

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

As discussed before, in [5, 20, 24] rotation-based as well as shear-based systems were constructed, which attain this rate up to a log-factor. Since these systems are instances of parabolic molecules with consistent parametrizations, their similar approximation behavior is no coincidence, as we will see in the next result.

Theorem 9 ([17]). *Assume that $(m_\lambda)_{\lambda\in\Lambda}$ is a system of parabolic molecules of order (R,M,N_1,N_2) with respect to the parametrization (Λ,Φ_Λ) such that*

(i) $(m_\lambda)_{\lambda\in\Lambda}$ constitutes a frame for $L^2(\mathbb{R}^2)$,
(ii) (Λ,Φ_Λ) is k-admissible for every $k>2$,
(iii) it holds that

$$R\geq 6,\quad M>9-\frac{5}{4},\quad N_1\geq 3+\frac{3}{4},\quad N_2\geq 6.$$

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 the cartoon image class $\mathscr{E}^2(\mathbb{R}^2)$.

We remark that condition (ii) holds in particular for the shearlet parametrization. Hence this result allows a simple derivation of the results in [20, 24] from [5]. In fact, Theorem 9 provides a systematic way to, in particular, prove results on sparse approximation of cartoon-like functions. It moreover enables us to provide a very general class of systems of parabolic molecules which optimally sparsely approximate cartoon-like functions by using the known result for curvelets.

7 Outlook and Further Generalizations

Finally, we discuss some possible extensions and directions for future research.

- *Higher Dimensional Setting.* A general framework such as parabolic molecules would also be of benefit for higher dimensional functions, in particular for the 3-dimensional setting which then includes videos with time as third dimension. The model of cartoon-like functions was already extended to this situation in [26]. Then, in [12], a general framework of parabolic molecules for functions in $L^2(\mathbb{R}^3)$ was introduced allowing, in particular, a similar result on the cross-Gramian of two systems of 3D parabolic molecules. We expect that the 3D framework now indicates a natural extension to higher dimensional settings.
- *General Scaling Matrix.* Another key question concerns the inclusion of other types of scaling laws: Can the framework of parabolic molecules be extended to also include, in particular, wavelets and ridgelets as well as newer hybrid constructions such as [26] or [23]? In the parabolic molecule framework the degree of anisotropic scaling is confined to parabolic scaling, but one approach to cover more scaling laws consists in the introduction of a parameter $\alpha\in[0,1]$, which measures the degree of anisotropy. More precisely, one then considers scaling matrices of the type $\mathrm{diag}(a,a^\alpha)$ for $\alpha\in[0,1]$, $\alpha=0$ corresponding to ridgelets,

$\alpha = \frac{1}{2}$ to curvelets and shearlets, and $\alpha = 1$ to wavelets. First results using this approach to introduce an extension of parabolic molecules coined α-*molecules* have been derived in [18].

- *Continuum Setting.* It would be highly desirable to also introduce such a framework for the continuum setting, i.e., with continuous parameter sets, adapted to the continuous shearlet and curvelet transform [6, 25, 14]. This would, for instance, allow the transfer of characterization results of microlocal smoothness spaces between different representation systems.

References

1. Andersson, F., de Hoop, M., Smith, H., Uhlmann, G.: A multi-scale approach to hyperbolic evolution equations with limited smoothness. Comm. PDE **33**, 988–1017 (2008)
2. Borup, L., Nielsen, M.: Frame decompositions of decomposition spaces. J. Fourier Anal. Appl. **13**, 39–70 (2007)
3. Candès, E. J., Demanet, L.,: The curvelet representation of wave propagators is optimally sparse. Comm. Pure Appl. Math. **58**, 1472–1528 (2002)
4. Candès, E. J., Donoho, D. L.: Ridgelets: a key to higher-dimensional intermittency? Phil. Trans. R. Soc. Lond. A. **357**, 2495-2509 (1999)
5. Candès, E. J., Donoho, D. L.: New tight frames of curvelets and optimal representations of objects with C^2 singularities. Comm. Pure Appl. Math. **56**, 219–266 (2004)
6. Candès, E. J., Donoho, D. L.: Continuous curvelet transform: I. Resolution of the wavefront set. Appl. Comput. Harmon. Anal. **19**, 162–197 (2005)
7. Candès, E. J., Donoho, D. L.: Continuous curvelet transform: II. Discretization and frames. Appl. Comput. Harmon. Anal. **19**, 198–222 (2005)
8. Casazza, P. G., Kutyniok, G., eds.: Finite frames: Theory and applications. Birkhäuser Boston (2012)
9. Christensen, O.: An introduction to frames and Riesz bases. Birkhäuser (2003)
10. DeVore, R. A.: Nonlinear approximation. Acta Numerica **7**, 51–150 (1998)
11. Donoho, D. L.: Sparse components of images and optimal atomic decomposition. Constr. Approx. **17**, 353–382 (2001)
12. Flinth, A.: 3D parabolic molecules. Bachelor's thesis, Technische Universität Berlin (2013)
13. Gribonval, R., Nielsen, M.: Nonlinear approximation with dictionaries. I. Direct estimates. J. Fourier Anal. Appl. **10**, 51–71 (2004)
14. Grohs, P.: Continuous Shearlet frames and Resolution of the Wavefront Set. Monatsh. Math. **164** 393–426, (2011)
15. Grohs, P.: Bandlimited shearlet frames with nice duals. J. Comput. Appl. Math. **244**, 139–151 (2013)
16. Grohs, P.: Intrinsic localization of anisotropic frames. Appl. Comput. Harmon. Anal. **35**, 264-283 (2013)
17. Grohs P., Kutyniok, G.: Parabolic molecules. Found. Comput. Math. (to appear).
18. Grohs, P., Keiper, S., Kutyniok, G., Schäfer, M.: α-Molecules. *preprint* (2013)
19. Guo, K., Kutyniok, G., Labate, D.: Sparse multidimensional representations using anisotropic dilation and shear operators. Wavelets and Splines (Athens, GA, 2005), Nashboro Press., 189–201 (2006)
20. Guo, K., Labate, D.: Optimally sparse multidimensional representation using shearlets. SIAM J. Math Anal. **39**, 298–318 (2007)
21. Guo, K., Labate, D.: Representation of Fourier integral operators using shearlets. J. Fourier Anal. Appl. **14**, 327–371 (2008)

22. Guo, K., Labate, D.: The construction of smooth parseval frames of shearlets. Math. Model Nat. Phenom. **8**, 82–105 (2013)
23. Keiper, S.: A flexible shearlet transform – sparse approximation and dictionary learning. Bachelor's thesis, Technische Universität Berlin (2013)
24. Kittipoom, P., Kutyniok, G., Lim, W.-Q: Construction of compactly supported shearlet frames. Constr. Approx. **35**, 21–72 (2012)
25. Kutyniok, G., Labate, D.: Resolution of the wavefront set using continuous shearlets. Trans. Amer. Math. Soc. **361**, 2719–2754 (2009)
26. Kutyniok, G., Lemvig, J., Lim, W.-Q: Optimally sparse approximations of 3D functions by compactly supported shearlet frames. SIAM J. Math. Anal. **44**, 2962–3017 (2012)
27. Kutyniok, G., Lim, W.-Q: Compactly supported shearlets are optimally sparse. J. Approx. Theory **163**, 1564–1589 (2011)
28. Kutyniok, G., Labate, D., eds.: Shearlets: Multiscale analysis for multivariate data. Birkhäuser Boston (2012)
29. Smith, H.: A parametrix construction for wave equations with $C^{1,1}$-coefficients. Ann. Inst. Fourier **48**, 797–835 (1998)
30. Wojtaszczyk, P.: A mathematical introduction to wavelets. Cambridge University Press, Cambridge (1997)

Recent Research Reports

Nr.	Authors/Title

2013-37 M. Hairer and M. Hutzenthaler and A. Jentzen
Loss of regularity for Kolmogorov equations

2013-38 W. E and A. Jentzen and H. Shen
Renormalized powers of Ornstein-Uhlenbeck processes and well-posedness of stochastic Ginzburg-Landau equations

2013-39 D. Schoetzau and Ch. Schwab and T.P. Wihler
hp-dGFEM for Second-Order Mixed Elliptic Problems in Polyhedra

2013-40 S. Mishra and F. Fuchs and A. McMurry and N.H. Risebro
Explicit and implicit finite volume schemes for radiation mhd and the effects of radiation on wave propagation in stratified atmospheres.

2013-41 J. Ernest and P. LeFloch and S. Mishra
Schemes with Well controlled Dissipation (WCD) I: Non-classical shock waves

2013-42 S. Mishra and N. Risebro and F. Weber
Convergence rates of finite difference schemes for the wave equation with rough coeffiicients

2013-43 U. Koley and S. Mishra and N. Risebro and F. Weber
Robust finite difference schemes for a nonlinear variational wave equation modeling liquid crystals

2013-44 G. Coclite and S. Mishra and N. Risebro and F. Weber
Analysis and Numerical approximation of Brinkman regularization of two-phase flows in porous media

2013-45 M. Hutzenthaler and A. Jentzen
Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients

2013-46 G. Da Prato and A. Jentzen and M. Röckner
A mild Itô formula for SPDEs