Internat. Math. Nachrichten
Nr. 0 (0), 1–16

Geometric Multiscale Analysis: From Wavelets to Parabolic Molecules

Gitta Kutyniok
Technische Universität Berlin

1 Introduction

The 21st century is typically referred to as the century of data. And indeed, today we face a deluge of data even already in daily life arising from, for instance, wireless communications or medical imaging procedures, which need to be acquired, analyzed, transmitted, and stored. These tasks pose very interesting challenges to mathematicians such as developing efficient methodologies for extracting key features from data or to derive optimality results concerning achievable compression rates.

The area of *applied harmonic analysis*, whose origin dates back to the 18th century and the introduction of the Fourier transform, promotes the following general approach. Given a class of data $\mathcal{C}$ in a Hilbert space, the data is decomposed according to

$$\mathcal{C} \ni x \longrightarrow (\langle x, \varphi_i \rangle)_{i \in I},$$

where $(\varphi_i)_{i \in I}$ is a carefully designed representation system. One key idea is that this *decomposition* now allows access to governing features of x. For instance, the location and direction of edges of an image x might be encoded in the set of indices $i \in I$ of those coefficients $\langle x, \varphi_i \rangle$ which are large in absolute value. In general, one might say that the associated coefficients $(\langle x, \varphi_i \rangle)_{i \in I}$ shall present the data in a form convenient for analysis and processing tasks.

A yet different set of applications such as, for instance, PDE solvers require *effi-*

ISSN 0020-7926

cient expansions of some $x \in \mathcal{C}$ in terms of a representation system $(\varphi_i)_{i\in I}$ by

$$x = \sum_{i\in I} c_i \varphi_i.$$

The representation system is ideally chosen such that the coefficient sequence $(c_i)_{i\in I}$ has fast decay in modulus, which is sometimes today coined a *sparse* representation. Certainly, if $(\varphi_i)_{i\in I}$ constitutes an orthonormal basis, the coefficients c_i have to be chosen as $(\langle x, \varphi_i\rangle)_{i\in I}$. In contrast to this, a redundant system allows for optimizing the sparsity of the sequence. One further key issue arising from numerical algorithms – which obviously require an approximation by finite sums – is the question which decay rate of the error of best N-term approximation is achievable. Already on an intuitive level, this shows the relation to the decomposition problem, since if the governing features are contained in the large coefficients, very few terms should already lead to high approximation rates.

Applied harmonic analysis poses certain desiderata to the choice of representation systems for decompositions and expansions. First, typically multiscale systems are chosen to allow different levels of resolution. Second, these representation systems are usually designed according to their partition of Fourier domain. And, third, for both the decomposition and the expansion fast algorithms should be available.

One prominent example are *wavelet systems* which are nowadays used in a variety of both theoretical and practical applications such as, for instance, in optimal schemes for solving elliptic PDEs [4] or in the compression standard JPEG2000 [16]. However, multivariate functions are typically governed by anisotropic – in the sense of directional – features such as singularities on lower dimensional embedded manifolds, which wavelets as isotropic systems cannot efficiently encode. Because of this reason, various novel anisotropic representation systems such as *curvelets* [3] and *shearlets* [13] have been suggested, which even has initiated the new research area of *geometric multiscale analysis*. For many of those systems, optimally sparse approximations have been proven for a particular function class in $L^2(\mathbb{R}^2)$ called cartoon-like functions which serves as a model for functions governed by anisotropic features. Very recently, a general framework called *parabolic molecules* has been proposed in [9], which includes all those systems as special cases and, for the first time, provides a higher level viewpoint on and deep insight into representation systems providing optimally sparse approximations of most types of multivariate functions.

This article shall serve as an introduction to and a survey about geometric multiscale analysis and, in particular, the novel theory of parabolic molecules. For this, we will first give an introduction into wavelet systems (Section 2). After a discussion about the appearance of anisotropic features in multivariate versus univariate functions in Section 3, we will introduce shearlet systems (Section 4) followed by an introduction of curvelet systems (Section 5). Section 6 is then devoted to the theory of parabolic molecules. Finally, an outlook to a framework coined α-*molecules* [8], which covers to some extent even wavelets and ridgelets [2], and a framework called *universal shearlets* [7], which provides a significantly improved flexibility in scaling, is given.

2 Wavelets

We start our endeavour with introducing and discussing wavelets. A wavelet system consists of one or a few generating functions to which scaling and translation operators are applied. To introduce a wavelet system for $L^2(\mathbb{R}^2)$, let us first take a look at the one-dimensional situation.

Definition 2.1. Let $\phi, \psi \in L^2(\mathbb{R})$. Then the associated *wavelet system for* $L^2(\mathbb{R})$ is defined to be

$$\{\phi_m := \phi(\cdot - m) : m \in \mathbb{Z}\} \cup \{\psi_{j,m} := 2^{j/2}\psi(2^j \cdot - m) : j \geq 0, m \in \mathbb{Z}\}.$$

It should be noted that ϕ and ψ can be constructed so that the associated wavelet system forms an orthonormal basis for $L^2(\mathbb{R})$ [16], and one then refers to ϕ as the *scaling function* and ψ as *wavelet*. As it is typical in applied harmonic analysis, this system is designed to partition Fourier domain in a particular way. Figure 1 shows how usually the essential support of the elements in a wavelet system tile the Fourier domain into different frequency bands.

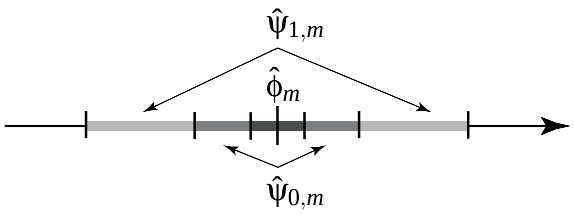


Figure 1: The partition of Fourier domain induced by a wavelet system for $L^2(\mathbb{R})$.

A wavelet system for $L^2(\mathbb{R}^2)$ can now be derived by a tensor product construction, leading to the following system.

Definition 2.2. Let $\phi, \psi \in L^2(\mathbb{R})$. Then the associated *wavelet system for* $L^2(\mathbb{R}^2)$ is defined by

$$\{\phi^{(1)}(x - m) : m \in \mathbb{Z}^2\} \cup \{2^j\psi^{(i)}(2^j x - m) : j \geq 0, m \in \mathbb{Z}^2, i = 1,2,3\},$$

where $\phi^{(1)}(x) = \phi(x_1)\phi(x_2)$, $\psi^{(1)}(x) = \phi(x_1)\psi(x_2)$, $\psi^{(2)}(x) = \psi(x_1)\phi(x_2)$, and $\psi^{(3)}(x) = \psi(x_1)\psi(x_2)$.

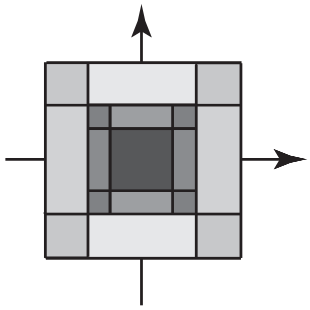


Figure 2: The partition of Fourier domain induced by a wavelet system for $L^2(\mathbb{R}^2)$.

The tiling of Fourier domain now takes the form as displayed in Figure 2, which the reader might want to compare with Figure 1.

One should point out that, on the application side, such 2D wavelet systems are used in the new compression standard JPEG2000. On the mathematical side, wavelet systems are proven to highly efficiently approximate L^2-functions which are smooth except for finitely many point singularities.

3 Anisotropy versus Isotropy

In contrast to univariate functions, multivariate functions cause the additional problem that they do not only exhibit point singularities, but also curvilinear singularities. And in fact, most multivariate functions appearing in applications are governed by such structures, which can be given either explicitly such as edges in images or implicitly such as shock fronts in transport equations.

A suitable model for such functions, called the model of cartoon-like functions, was introduced in [6] in 2001.

Definition 3.1. The set of *cartoon-like functions* $\mathcal{E}^2(\mathbb{R}^2)$ is defined by

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

where $B \subset [0,1]^2$ with ∂B a closed C^2-curve with bounded curvature and $f_0, f_1 \in C_0^2([0,1]^2)$.

An illustration is shown in Figure 3.

Based on this model situation, the following result now provides a benchmark concerning the maximally achievable approximation rate. For the notion of a frame as an extension of orthonormal bases, we refer to Section 4.

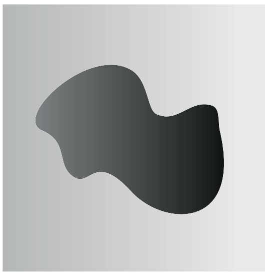


Figure 3: Illustration of a cartoon-like function.

Theorem 3.1 ([6]). *Let* $(\psi_\lambda)_{\lambda\in\Lambda}$ *be a frame for* $L^2(\mathbb{R}^2)$. *Then the optimal asymptotic approximation error of* $f \in \mathcal{E}^2(\mathbb{R}^2)$ *is*

$$\|f - f_N\|_2^2 \asymp N^{-2} \quad \text{as } N \to \infty, \quad \text{where } f_N = \sum_{\lambda\in\Lambda_N} c_\lambda \psi_\lambda$$

is the (non-linear) best N-term approximation.

As might be expected, since wavelets are isotropic objects due to the isotropic scaling matrix, they are only able to deliver a significantly suboptimal approximation rate of

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

The intuitive reason for this failure is illustrated in Figure 4.

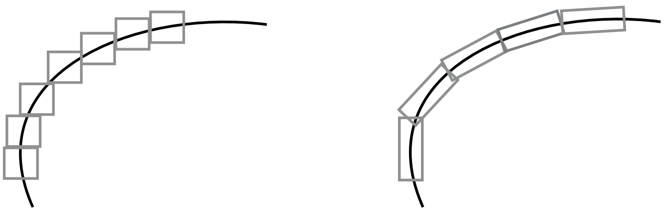


Figure 4: Approximation of a curve by isotropic and anisotropic objects.

This raises not only the question of whether there exist *anisotropic* representation systems $(\psi_\lambda)_{\lambda\in\Lambda}$ which meet this benchmark, but also whether they can be chosen to be 'conveniently' defined. More precisely, the new area of *geometric multiscale analysis* which arose from this question seeks to introduce representation systems which satisfy the following list of desiderata.

(D1) The system should be generated by one or few generating functions.

(D2) The benchmark from Theorem 3.1 should be met.

(D3) The system should allow for compactly supported analyzing elements.
(D4) The continuum and digital realm should be treated uniformly.
(D5) The associated transform should admit a fast implementation.

Item (D3) ensures high spatial localization, whereas item (D4) allows for faithful implementation of the continuum domain theory.

An abundance of approaches have been suggested with the probably most well-known ones being curvelets [3], contourlets [5], and shearlets [11]. We now continue by introducing shearlets, which are by now the only system actually satisfying the previously stated list of desiderata.

4 Shearlets

To accomodate (D1), shearlets are as wavelets based on very few generating functions to which scaling and translation operators are applied. Since these are however anisotropic systems, a third operation is required which changes their orientation.

As scaling, *parabolic scaling* is chosen – the reason being discussed in Section 6 – which is defined by

$$A_{2^j} = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix} \quad \text{and} \quad \tilde{A}_{2^j} = \begin{pmatrix} 2^{j/2} & 0 \\ 0 & 2^j \end{pmatrix}.$$

To change the orientation via rotation would prevent (D4), since rotation does not leave the digital grid $\mathbb{Z}^2$ invariant. Hence *shearing*, given by

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

is selected, and this selection indeed ensures (D4). By using the framework of *affine systems*, which consists of systems of the form

$$\{|\det M|^{1/2}\psi(M \cdot -m) : M \in G \subseteq GL_2,\, m \in \mathbb{Z}^2\}, \quad \psi \in L^2(\mathbb{R}^2),$$

the translation operation is already ‘built-in’.

Since shearing does not provide a resolution of the directions as uniform as rotation, we require two separate systems to handle the more horizontally and the more vertically aligned directions. More precisely, we aim for a partition of Fourier domain as illustrated in Figure 5.

This leads to the following definition, in this form first stated in [14]. It should be noted before that the translation part is made more flexible by the introduction of the matrices M_c and $\tilde{M}_c$ in order to also enable finer sampling. We further remark that those systems are sometimes also referred to as *cone-adapted* shearlet systems.

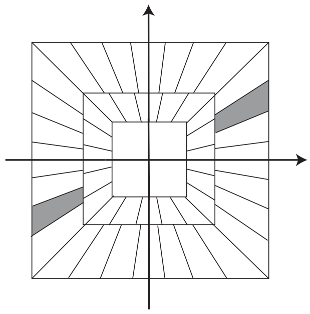


Figure 5: The partition of Fourier domain induced by a shearlet system.

Definition 4.1. For $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ and $c = (c_1, c_2) \in (\mathbb{R}_+)^2$, the *shearlet system* $SH(\phi, \psi, \tilde{\psi}; c)$ is defined by

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

where

$$\Phi(\phi; c_1) := \{\phi_m = \phi(\cdot - c_1 m) : m \in \mathbb{Z}^2\},$$
$$\Psi(\psi; c) := \{\psi_{j,k,m} = 2^{\frac{3}{4}j}\psi(S_k A_{2^j} \cdot - M_c m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\},$$
$$\tilde{\Psi}(\tilde{\psi}; c) := \{\tilde{\psi}_{j,k,m} = 2^{\frac{3}{4}j}\tilde{\psi}(S_k^T \tilde{A}_{2^j} \cdot - \tilde{M}_c m) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\},$$

with

$$M_c = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix} \quad \text{and} \quad \tilde{M}_c = \begin{pmatrix} c_2 & 0 \\ 0 & c_1 \end{pmatrix}.$$

A first large class of functions considered as generators was the following class of band-limited functions, i.e., functions whose Fourier transform is compactly supported.

Example 4.1. *Classical shearlets* are functions $\psi \in L^2(\mathbb{R}^2)$ of the form

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\,\hat{\psi}_2(\tfrac{\xi_2}{\xi_1}),$$

where $\psi_1 \in L^2(\mathbb{R})$ is a discrete wavelet in the sense that it satisfies the discrete Calderón condition, given by

$$\sum_{j \in \mathbb{Z}} |\hat{\psi}_1(2^{-j}\xi)|^2 = 1 \quad \text{for a.e. } \xi \in \mathbb{R},$$

with $\hat{\psi}_1 \in C^\infty(\mathbb{R})$ and $\operatorname{supp}\hat{\psi}_1 \subseteq [-\frac{1}{2}, -\frac{1}{16}] \cup [\frac{1}{16}, \frac{1}{2}]$, and $\psi_2 \in L^2(\mathbb{R})$ is a 'bump function' in the sense that

$$\sum_{k=-1}^{1} |\hat{\psi}_2(\xi + k)|^2 = 1 \quad \text{for a.e. } \xi \in [-1, 1], \tag{4.1}$$

satisfying $\hat{\psi}_2 \in C^\infty(\mathbb{R})$ and $\operatorname{supp}\hat{\psi}_2 \subseteq [-1,1]$.

One might now ask whether there exist generators so that the associated shearlet system constitutes an orthonormal basis for $L^2(\mathbb{R})$. However, no constructions of shearlet orthonormal bases are known to date due to the redundancy coming from the shear component. But a variety of shearlet systems exist which still have superior stability properties in the sense of frames. For those readers not familiar with this functional analytic concept, let us briefly recall the basics from frame theory.

Definition 4.2. A sequence $(g_i)_{i\in I}$ is a *frame for* $L^2(\mathbb{R}^2)$, if there exist constants $0 < A \leq B < \infty$ such that

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

A and B are called the *lower* and *upper frame bound*, respectively. If $A = B$ is possible, $(g_i)_{i\in I}$ is called a *tight* frame. In case $A = B = 1$, it is referred to as a *Parseval frame*.

A frame $(g_i)_{i\in I}$ for $L^2(\mathbb{R}^2)$ allows the analysis of elements in $L^2(\mathbb{R}^2)$ through application of the *analysis operator* given by

$$T : L^2(\mathbb{R}^2) \to \ell_2(I), \quad T(f) = (\langle f, g_i\rangle)_{i\in I}.$$

The associated *frame operator* $Sf = T^*Tf = \sum_{i\in I}\langle f, g_i\rangle g_i$ in turn gives rise to the reconstruction formula

$$f = \sum_{i\in I}\langle f, g_i\rangle S^{-1} g_i \quad \text{for all } f \in L^2(\mathbb{R}^2).$$

Coming back to the situation of shearlets, it was then shown in [11], that $SH(\phi, \psi, \tilde{\psi}; (1,1))$ with suitable ϕ, with $\psi, \tilde{\psi}$ being classical shearlets ($\tilde{\psi}$ with interchanged variables), and with small modifications of the boundary elements forms a Parseval frame for $L^2(\mathbb{R}^2)$.

To however accommodate (D3), we require compactly supported generators. This forces us to give up on optimal stability, i.e., on a Parseval frame. But the following result shows that one still has a certain degree of stability by being able to control the frame bounds.

Theorem 4.1 ([12]). *For* $\alpha > \gamma > 3$, $q > q' > 0$ *and* $q > r > 0$, *let*

$$|\hat{\psi}(\xi_1, \xi_2)| \leq C \cdot \min\{1, |q\xi_1|^\alpha\} \cdot \min\{1, |q'\xi_1|^{-\gamma}\} \cdot \min\{1, |r\xi_2|^{-\gamma}\},$$

and

$$\sum_{j,k} |\hat{\psi}(S_{-k}^T A_{2^{-j}}\xi)|^2 \geq C' > 0,$$

and similar for $\tilde{\psi}$. Then there exists a sampling constant c_0 such that $SH(\phi,\psi,\tilde{\psi};c)$ is a frame for $L^2(\mathbb{R}^2)$ for all $c \leq c_0$ with

$$c^{-2}C_1(\alpha,\gamma,q,q',r,c) \leq A \leq B \leq c^{-2}C_2(\alpha,\gamma,q,q',r,c),$$

where explicit formulas for $C_1(\alpha,\gamma,q,q',r,c)$ and $C_2(\alpha,\gamma,q,q',r,c)$ exist.

Our original goal was though to meet the benchmark from Theorem 3.1. This is the content of the next theorem, which shows that also (D2) is satisfied by shearlets. In fact, it meets it up to a log-factor, wherefore we included the '(almost)'. However, if one regards a log-factor as negligible, this is indeed the optimal achievable rate.

Theorem 4.2 ([14]). *Let $\psi \in L^2(\mathbb{R}^2)$ (similar for $\tilde{\psi}$) be compactly supported such that, for $\alpha > 5$, $\gamma \geq 4$, $h \in L^1(\mathbb{R})$,*

(i) $|\hat{\psi}(\xi)| \leq C \cdot \min\{1,|\xi_1|^\alpha\} \cdot \min\{1,|\xi_1|^{-\gamma}\} \cdot \min\{1,|\xi_2|^{-\gamma}\}$,

(ii) $\left|\frac{\partial}{\partial \xi_2}\hat{\psi}(\xi)\right| \leq |h(\xi_1)| \cdot \left(1+\frac{|\xi_2|}{|\xi_1|}\right)^{-\gamma}$.

Suppose $SH(\phi,\psi,\tilde{\psi};c)$ forms a frame for $L^2(\mathbb{R}^2)$. Then it provides (almost) optimally sparse approximations of cartoon-like functions $f \in \mathcal{E}^2(\mathbb{R}^2)$ in the sense that

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

where f_N is the N-term approximation consisting of the N largest shearlet coefficients.

To provide some intuition, let us give a heuristic argument which shows why the rate N^{-2} can be achieved. Due to the form of f_N, we obtain

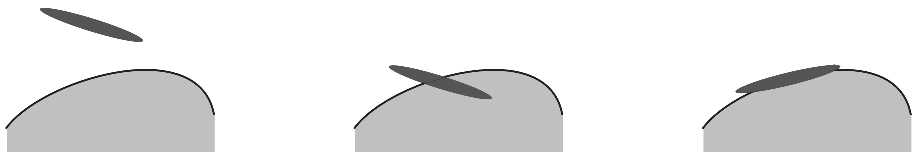
$$\|f - f_N\|_2^2 \leq \frac{1}{A} \cdot \sum_{n>N} (|\langle f,\sigma_\eta\rangle|)_{(n)}^2, \tag{4.2}$$

where $(\sigma_\eta)_\eta$ is the shearlet frame with lower frame bound A and $(|\langle f,\sigma_\eta\rangle|)_{(n)}$ denotes the nth largest shearlet coefficient. To estimate these coefficients, we have to distinguish three cases which are illustrated in Figure 6. In the cases of Figure 6(a)+(b), the coefficients are negligible, mainly because of the assumed (directional) vanishing moment conditions. In the case of Figure 6(c), we can estimate

$$|\langle f,\sigma_\eta\rangle| \leq \|f\|_\infty \|\sigma_\eta\|_1 \leq C \cdot 2^{-\frac{3}{4}j}.$$

Thus, we know that there exist $2^{j/2}$ of such coefficients – recall that each shearlet has a length of $2^{-j/2}$ – and we have an estimate for each of those. This allows us to complete (4.2) by

$$\|f - f_N\|_2^2 \leq \frac{1}{A} \cdot \sum_{n>N} (|\langle f,\sigma_\eta\rangle|)_{(n)}^2 \leq \frac{C}{A} \cdot \sum_{n>N} (n^{-\frac{3}{2}})^2 \leq \frac{C}{A} \cdot N^{-2},$$



(a) (b) (c)

Figure 6: Positions of different shearlets with respect to a discontinuity curve of a cartoon-like function.

thereby finishing the argument.

Finally, also (D5) is satisfied by shearlets. In fact, `www.ShearLab.org` provides an extensive software package for the 2D and 3D shearlet transform based on compactly supported shearlets with accompanying publication [15].

5 Curvelets

The system of curvelets predates shearlets, and was originally introduced based on ridgelets [2], which are systems of certain ridge functions to optimally sparsely approximate ridge-like singularities. The nowadays utilized curvelet system, also called second generation curvelets, was introduced in [3]. As typical for the area of applied harmonic analysis, they are designed to partition Fourier domain in a particular way, which in this case is illustrated in Figure 7. As can be seen, it pro-

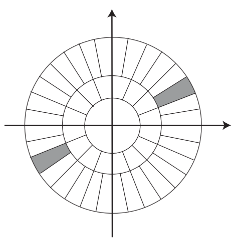


Figure 7: The partition of Fourier domain induced by a curvelet system.

vides a perfect resolution of the different directions in contrast to the approximate one provided by shearlets. This however comes with the disadvantage that there does not exist a faithful implementation of curvelets, which are because of this fact a purely continuum domain theory. It should though be emphasized that curvelets were the first system shown to deliver (almost) optimally sparse approximations of cartoon-like functions [3], which can justifiably be called a breakthrough.

The definition of curvelets is as follows.

Definition 5.1. Let $W \in C^\infty(\mathbb{R})$ be a wavelet with $\mathrm{supp}(W) \subseteq \left(\frac{1}{2}, 2\right)$, and $V \in C^\infty(\mathbb{R})$ be a 'bump function' (cf. (4.1)) with $\mathrm{supp}(V) \subseteq (-1,1)$. Then the *curvelet system* $(\gamma_{(j,\ell,k)})_{(j,\ell,k)\in\Lambda^0}$, where

$$\Lambda^0 := \left\{ (j,\ell,k) \in \mathbb{Z}^4 \ : \ j \geq 0, \ \ell = -2^{\lfloor \frac{j}{2} \rfloor - 1}, \cdots, 2^{\lfloor \frac{j}{2} \rfloor - 1} \right\} \tag{5.1}$$

is defined in polar coordinates by

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

and

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

with $\theta_{(j,\ell)} = \pi\ell/2^{j/2}$ and $x_{(j,\ell,k)} = R_{\theta_{(j,\ell)}} A_{2^{-j}} m$ for $j \geq j_0$, $\ell = 0, \ldots, 2^{\lfloor j/2 \rfloor} - 1$, and $m \in \mathbb{Z}^2$.

It was shown in [3] that this system constitutes a Parseval frame for $L^2(\mathbb{R}^2)$ provided that appropriate functions to address the low frequency part are included.

The following result from 2004 is the first result providing a representation system which (almost) meets the benchmark from Theorem 3.1.

Theorem 5.1 ([3])**.** *The curvelet system provides (almost) optimally sparse approximations of cartoon-like functions $f \in \mathcal{E}^2(\mathbb{R}^2)$ in the sense that*

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

where f_N is the N-term approximation consisting of the N largest curvelet coefficients.

6 Parabolic Molecules

As can be seen from Sections 4 and 5, curvelets differ significantly from shearlets, since they do not form affine systems, they are based on rotation rather than shearing causing problems with faithful implementations, and no compactly supported version is available causing problems with high spatial localization. But there are also striking similarities, since both systems (almost) optimally sparsify cartoon-like functions. Thus the sparsity properties of curvelets and shearlets are similar, and even more, the results for the band-limited version of both systems are proven with resembling proofs [3, 10]. This observation raises the question whether there exists a general framework for such directional systems and what the fundamental concept behind sparse approximation results really is.

The properties we require of a general framework are to cover all systems known to provide optimally sparse approximations of cartoon-like functions, to enable an easy transfer of (sparsity) results between systems, to allow a categorization of systems with respect to sparsity behaviors, and to also be general enough to allow constructions of novel systems.

The common bracket between, in particular, curvelets and shearlets is *parabolic scaling*. This is due to the fact that parabolic scaling is perfectly adapted to the C^2 regularity of the discontinuity curve employed in the cartoon-like model as can be seen heuristically. For this, let $(E(x_2), x_2), x_2 \in I$ be a parametrization of the singularity curve in Figure 8 and, by noting that $E(0) = E'(0) = 0$, consider the

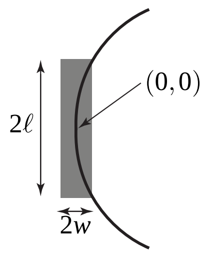


Figure 8: Anisotropic function approximating a curvilinear singularity.

approximation by the Taylor series

$$E(x_2) \approx \frac{1}{2}\kappa x_2^2.$$

For a fixed length 2ℓ, let $2w$ be the width of the smallest rectangle centered at $(0,0)$ containing the entire edge curve E (cf. Figure 8). Then, since $E(\ell) = w$, we obtain

$$w \approx \frac{\kappa}{2}\ell^2,$$

which can be interpreted as '*width* $\approx$ *length*^{2}', which in turn is parabolic scaling.

After agreeing on the type of scaling, a general framework requires a common parameter space, which will be chosen as

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

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

Definition 6.1. A *parametrization* is a pair (Λ, Φ_Λ), where Λ is a discrete index set and Φ_Λ is a mapping

$$\Phi_\Lambda : \begin{cases} \Lambda & \rightarrow & \mathbb{P}, \\ \lambda & \mapsto & (s_\lambda, \theta_\lambda, x_\lambda). \end{cases}$$

We notice that for now, no properties of the map Φ_Λ are required.

Since curvelets are very adapted to the parameter space, a possible associated parametrization can be easily derived as follows.

Example 6.1. The *canonical parametrization* (Λ^0, Φ^0) with Λ^0 being the index set associated with curvelets from (5.1) is defined by

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

The key idea of the definition of parabolic molecules from [9] which now follows is to define maximally flexible systems based on parabolic scaling, rotation, and translation with parameter space $\mathbb{P}$ which allow a different generating function for each index. Decay and smoothness properties of those functions are then governed by the parameters (L, M, N_1, N_2), where L measures spatial localization, M the number of directional (almost) vanishing moments, and N_1, N_2 smoothness.

Definition 6.2. Let (Λ, Φ_Λ) be a parametrization. Then $(m_\lambda)_{\lambda\in\Lambda}$ is a *system of parabolic molecules of order* $(L, M, N_1, N_2) \in (\mathbb{Z}_+ \cup \{\infty\}) \times \mathbb{Z}_+^3$, if, for all $\lambda \in \Lambda$,

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

such that, for all $|\beta| \leq L$,

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

where $\langle x \rangle := (1 + x^2)^{1/2}$.

This framework can be shown to, in particular, include parabolic frames [17], curvelets [3], band-limited shearlets [10], frame decompositions [1], and compactly supported shearlets [12].

Heading towards a general result which poses conditions on the parameters (L, M, N_1, N_2) for a system of parabolic molecules to deliver (almost) optimally sparse approximations of cartoon-like functions, we first state a result which analyzes the decay of the cross-Gramian of two systems of parabolic molecules. This will allow us to transfer the optimal sparse approximation result from curvelets to other systems of parabolic molecules. For this decay result though, we require a distance function between two indices from the common parameter space.

Definition 6.3. Let (Λ, Φ_Λ) and $(\tilde{\Lambda}, \Phi_{\tilde{\Lambda}})$ be parametrizations. For $\lambda \in \Lambda$ and $\mu \in \tilde{\Lambda}$, we define the *index distance* by

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

where

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

We remark that d is nothing else than the Hart Smith's phase space metric on $\mathbb{T}\times\mathbb{R}^2$ (cf. [17]). Using the index distance, we can next state the result on the decay of the cross-Gramian of two systems of parabolic molecules.

Theorem 6.1 ([9]). *Let $N>0$, and let $(m_\lambda)_{\lambda\in\Lambda}$, $(p_\mu)_{\mu\in\tilde{\Lambda}}$ be systems of parabolic molecules of order (L,M,N_1,N_2) with*

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

Then, for all $\lambda\in\Lambda$ and $\mu\in\tilde{\Lambda}$,

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

As already mentioned, this result will be key to transfer the (almost) optimal sparse approximation result from curvelets to various other systems of parabolic molecules. This is however only possible provided that the parametrization of this other system is in some sense 'consistent' with the (canonical) parametrization of curvelets introduced in Example 6.1.

Definition 6.4. A parametrization (Λ,Φ_Λ) is *k-admissible*, if

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

As expected, the curvelet parametrization is, for instance, consistent with itself for all $k>2$, as was shown in [9].

The next main result now reveals a very large class of representation systems parametrized as systems of parabolic molecules which all provide (almost) optimally sparse approximations of cartoon-like functions. It in fact proves that this class consists of all frames of parabolic molecules whose parametrization is consistent with the canonical one and whose associated parameters (L,M,N_1,N_2) are sufficiently large.

Theorem 6.2 ([9]). *Let $(m_\lambda)_{\lambda\in\Lambda}$ be a system of parabolic molecules of order (L,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*

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

Then, for any $\varepsilon > 0$ *and for any* $f \in \mathcal{E}^2(\mathbb{R}^2)$, $(m_\lambda)_{\lambda\in\Lambda}$ *satisfies*

$$\|f - f_N\|_2^2 \leq C \cdot N^{-2+\varepsilon} \quad \textit{as } N \to \infty,$$

It should be emphasized that this result does not only provide a higher level view of the properties which are required of systems to deliver (almost) optimal sparse approximations of cartoon-like functions, but it also allows the construction of a variety of novel systems which automatically exhibit such approximation properties.

7 ...α-Molecules?

Finally, the question arises whether it might be possible to extend the framework of parabolic molecules to include wavelets; maybe even ridgelets. In fact, very recently a much more elaborate framework coined α-*molecules* was introduced in [8], the key idea being to incorporate a parameter $\alpha \in [0,1]$ to measure the amount of anisotropy by considering the scaling matrix

$$A_{\alpha,a} = \begin{pmatrix} a & 0 \\ 0 & a^\alpha \end{pmatrix}, \quad a > 0.$$

In this setting, $\alpha = 1$ corresponds to the scaling associated with wavelets, $\alpha = \frac{1}{2}$ to curvelets or shearlets, and $\alpha = 0$ to ridgelets. In [8] in the spirit of parabolic molecules, an elaborate framework is introduced which then allows to derive sparse approximation results for large classes of systems, thereby on a higher level linking approximation properties to structural properties.

A yet different extension, which should also be mentioned are so-called *universal shearlets*, introduced in [7]. Those systems allow a different scaling matrix for each scale j with $\alpha_j \in [\frac{1}{2}, 1]$, and are specifically designed to parametrize a path from wavelets to shearlets. They were introduced for the purpose of analyzing the ability of wavelets versus shearlets for compressed sensing based inpainting algorithms. Still it can be envisioned that the idea of varying scaling could be incorporated in the framework of α-*molecules* allowing even more flexibility.

References

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

[2] E. J. Candès and D. L. Donoho, *Ridgelets: a key to higher-dimensional intermittency?*. Phil. Trans. R. Soc. Lond. A. **357** (1999), 2495–2509.

[3] 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** (2004), 219–266.
[4] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations: convergence rates*. Math. Comp. **70** (2000), 27–75.
[5] M. N. Do and M. Vetterli, *The contourlet transform: an efficient directional multiresolution image representation*. IEEE Trans. Image Proc. **14** (2006), 2091–2106.
[6] D. L. Donoho, *Sparse components of images and optimal atomic decomposition*, Constr. Approx. **17** (2004), 353–382.
[7] M. Genzel and G. Kutyniok, *Asymptotical Analysis of Inpainting via Universal Shearlet Systems*. Preprint (2014).
[8] P. Grohs, S. Keiper, G. Kutyniok, and M. Schäfer, α-*Molecules*. Preprint (2014).
[9] P. Grohs and G. Kutyniok, *Parabolic Molecules*, Found. Comput. Math., to appear.
[10] K. Guo and D. Labate, *Optimally sparse multidimensional representation using shearlets*. SIAM J. Math Anal. **39** (2007), 298–318.
[11] K. Guo, G. Kutyniok, and D. Labate, *Sparse multidimensional representations using anisotropic dilation and shear operators*, in: Wavelets and Splines (Athens, GA, 2005), Nashboro Press, Nashville, TN, 2006, 189–201.
[12] P. Kittipoom, G. Kutyniok, and W.-Q Lim, *Construction of compactly supported shearlet frames*. Constr. Approx. **35** (2012), 21–72.
[13] G. Kutyniok and D. Labate, *Introduction to Shearlets*, in: Shearlets: Multiscale Analysis for Multivariate Data, Birkhäuser Boston, 2012, 1–38.
[14] G. Kutyniok and W.-Q Lim, *Compactly supported shearlets are optimally sparse*. J. Approx. Theory **163** (2011), 1564–1589.
[15] G. Kutyniok, W.-Q Lim, and R. Reisenhofer, *ShearLab 3D: Faithful Digital Shearlet Transforms based on Compactly Supported Shearlets*. Preprint (2014).
[16] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 2008.
[17] H. Smith, *A parametrix construction for wave equations with $C^{1,1}$-coefficients*, Ann. Inst. Fourier **48** (1998), 797–835.

Author's address:

Gitta Kutyniok
Technische Universität Berlin
Department of Mathematics
Straße des 17. Juni 136 10623 Berlin, Germany