

# $\alpha$ -Molecules: Curvelets, Shearlets, Ridgelets, and Beyond

Philipp Grohs<sup>a</sup>, Sandra Keiper<sup>b</sup>, Gitta Kutyniok<sup>b</sup>, and Martin Schäfer<sup>b</sup>

<sup>a</sup>ETH Zürich, Seminar for Applied Mathematics, 8092 Zürich, Switzerland;

<sup>b</sup>Technische Universität Berlin, Department of Mathematics, 10623 Berlin, Germany

## ABSTRACT

The novel framework of parabolic molecules provides for the first time a unifying framework for (sparse) approximation properties of directional representation systems by, in particular, including curvelets and shearlets. However, the considered common bracket is parabolic scaling, which excludes systems such as ridgelets and wavelets. In this paper, we therefore provide a generalization of this framework, which we coin  $\alpha$ -molecules, by introducing an additional parameter  $\alpha$ , which specifies the extent of anisotropy in the scaling. We show that, for instance, both ridgelets and wavelets are in fact  $\alpha$ -molecules. As an application of the concept, we then analyze the sparse approximation behavior of  $\alpha$ -molecules. Utilizing the idea of sparsity equivalence, it is possible to identify large classes of  $\alpha$ -molecules providing the same sparse approximation behavior.

**Keywords:** Ridgelets, Wavelets, Curvelets, Nonlinear Approximation, Anisotropic Scaling, Shearlets, Sparsity Equivalence, Parabolic Molecules, Cartoon Images

## 1. INTRODUCTION

In recent years, starting from the classical system of wavelets, various novel representation systems have been introduced to efficiently represent 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, curvelets, and shearlets. 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 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. approximation rates, 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 give 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 [1] 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. 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

---

Further author information: (Send correspondence to G.K.)

P.G.: E-mail: philipp.grohs@sam.math.ethz.ch, Telephone: +41 44 632 3200

S.K.: E-mail: keiper@math.tu-berlin.de, Telephone: +49 30 314 27826

G.K.: E-mail: kutyniok@math.tu-berlin.de, Telephone: +49 30 314 25758

M.S.: E-mail: schaefer@math.tu-berlin.de, Telephone: +49 30 314 27379

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 [1] 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 application, it was proven that these systems feature a similar approximation behavior.

However, in the parabolic molecule framework the degree of anisotropic scaling is confined to parabolic scaling, excluding, for instance, ridgelets, wavelets as well as newer hybrid constructions. The aim of this paper is therefore to further extend the framework by allowing more general ‘ $\alpha$ -scaling’, where the parameter  $\alpha \in [0, 1]$  specifies the degree of anisotropy in the scaling. This conceptual idea ensures that the aforementioned systems can be included.

After establishing the basic concept in Section 2, accompanied by some examples, we prove as one main result of this paper that the cross-Gramian of two systems of  $\alpha$ -molecules exhibits a strong off-diagonal decay. This property will become essential in Section 3, where the approximation behavior of  $\alpha$ -molecules is studied. In fact, we can essentially prove that any two systems of  $\alpha$ -molecules, which are consistent and have sufficiently high order, exhibit the same approximation behavior. This, in particular, provides a systematic way to prove results on sparse approximation of cartoon images. For further details such as for the proofs of the presented results, we refer the reader to [2].

## 2. $\alpha$ -MOLECULES

In this section we introduce the concept of  $\alpha$ -molecules, give some examples and finally show the crucial property that under certain conditions the cross-Gramian of two systems of  $\alpha$ -molecules is almost orthogonal. Not surprisingly, the exposition follows the same lines as [1] for parabolic molecules.

Throughout the paper the notation  $a \lesssim b$  shall indicate that the entities  $a, b$ , possibly depending on some context dependent parameters, satisfy  $a \leq C \cdot b$  for a positive constant  $C > 0$ , which is independent of the parameters. If both  $a \lesssim b$  and  $b \lesssim a$  we denote this by  $a \asymp b$ . For  $x \in \mathbb{R}$  we use  $\langle x \rangle$  to abbreviate  $(1 + x^2)^{\frac{1}{2}}$ .

### 2.1 Definition of $\alpha$ -Molecules

We start by defining the *parameter space*

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

where  $\mathbb{R}_+ = (0, \infty)$  and  $\mathbb{T} = [-\frac{\pi}{2}, \frac{\pi}{2}]$  denotes the torus, where the endpoints are identified. A point  $p = (s, \theta, x) \in \mathbb{P}$  describes a scale  $s \in \mathbb{R}_+$ , an orientation  $\theta \in \mathbb{T}$ , and a location  $x \in \mathbb{R}^2$ .

$\alpha$ -molecules are defined as systems of functions  $(m_\lambda)_{\lambda \in \Lambda}$ , where each  $m_\lambda \in L_2(\mathbb{R}^2)$  has to satisfy some additional properties. In particular, each function  $m_\lambda$  will be associated with a unique point in  $\mathbb{P}$ , which is done via a *parametrization* as defined below.

DEFINITION 2.1. A parametrization consists of a pair  $(\Lambda, \Phi_\Lambda)$  where  $\Lambda$  is an index set and  $\Phi_\Lambda$  is a mapping

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

which associates with each  $\lambda \in \Lambda$  a scale  $s_\lambda$ , a direction  $\theta_\lambda$  and a location  $x_\lambda$ .

Let  $R_\theta := \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$  denote the rotation matrix by an angle  $\theta \in \mathbb{R}$  and  $D_a := \text{diag}(a, a^\alpha)$  the anisotropic dilation matrix associated with  $a > 0$  and  $\alpha \in [0, 1]$ . Now we have collected all necessary ingredients for defining  $\alpha$ -molecules.

DEFINITION 2.2. Let  $(\Lambda, \Phi_\Lambda)$  be a parametrization. A family  $(m_\lambda)_{\lambda \in \Lambda}$  is called a family of  $\alpha$ -molecules with respect to the parametrization  $(\Lambda, \Phi_\Lambda)$  of order  $(R, M, N_1, N_2)$ , if it can be written as

$$m_\lambda(x) = s_\lambda^{(1+\alpha)/2} a^{(\lambda)}(D_{s_\lambda} R_{\theta_\lambda}(x - x_\lambda))$$

such that

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

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

REMARK 2.3. We comment on a minor notational difference between the present work and [1]. In [1] the parameter  $s_\lambda$  corresponds to the logarithm of the scale, i.e. the scale  $s_\lambda$  in the present paper should be understood as  $2^{s_\lambda}$  from [1].

According to the definition above an  $\alpha$ -molecule essentially has frequency support in a pair of opposite wedges associated to a certain orientation, and essential spatial support in a rectangle, whose side lengths depend on the scale.

As in [1] our definition poses conditions on the Fourier transform of  $m_\lambda$  which can however be interpreted also in terms of spatial localization properties: The number  $R$  describes the spatial localization,  $M$  the number of directional (almost) vanishing moments and  $N_1, N_2$  describe the smoothness of an element  $m_\lambda$ . We refer to Figure 1 for an illustration of the approximate frequency support of an  $\alpha$ -molecule in the case  $\alpha = \frac{1}{2}$ .

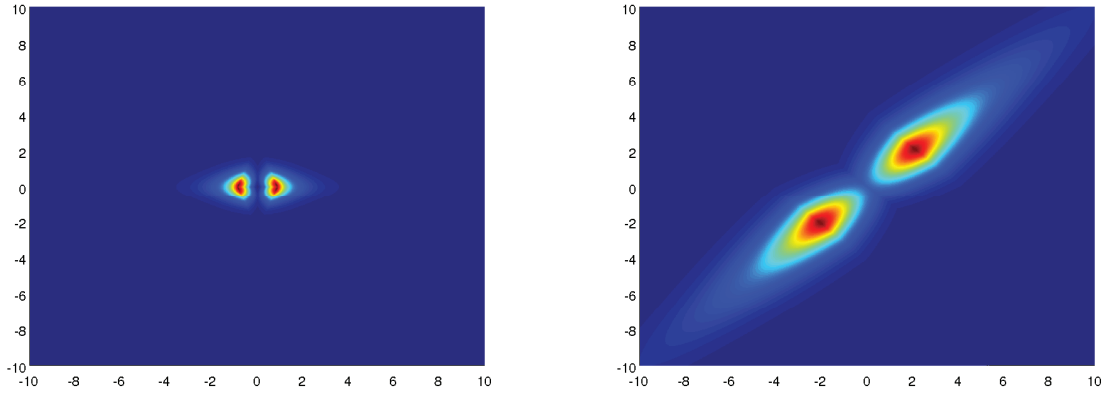


Figure 1. Left: The weight function  $\min\left(1, s_\lambda^{-1} + |\xi_1| + s_\lambda^{-(1-\alpha)} |\xi_2|\right)^M \langle |\xi| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2}$  for  $s_\lambda = 8$ ,  $M = 3$ ,  $N_1 = N_2 = 2$ . Right: Approximate frequency support of a corresponding  $\frac{1}{2}$ -molecule  $\hat{m}_\lambda$  with  $\theta_\lambda = \pi/4$ .

We proceed to specify two important subclasses of  $\alpha$ -molecules. The classification is performed via the choice of certain parametrizations.

### 2.1.1 $\alpha$ -Curvelet Molecules

In [3] the authors introduced the notion of *curvelet molecules*, a notion closely related to curvelets (see Section 2.2.3). We generalize their definition and specify the class of curvelet parametrizations.

DEFINITION 2.4. Let  $\alpha \in [0, 1]$  and  $\tau > 0$ ,  $g > 1$  be some fixed parameters. Further, let  $(\omega_j)_{j \in \mathbb{N}_0}$  be a sequence of positive real numbers with  $\omega_j \asymp g^{-j(1-\alpha)}$ , i.e. there are constants  $C, c > 0$  independent of  $j \in \mathbb{N}_0$  such that  $cg^{-j(1-\alpha)} \leq \omega_j \leq Cg^{-j(1-\alpha)}$ . An  $\alpha$ -curvelet parametrization is given by an index set of the form

$$\Lambda^c := \{(j, \ell, k) : j \in \mathbb{N}_0, \ell \in \mathbb{Z} \text{ with } |\ell| \leq L_j \text{ for some } L_j \in \mathbb{N}_0 \cup \{\infty\}, k \in \mathbb{Z}^2\},$$

and a mapping  $\Phi^c$ , which assigns to each index  $\lambda = (j, \ell, k)$  a point  $(s_\lambda, \theta_\lambda, x_\lambda)$  in the parameter space  $\mathbb{P}$  via  $s_\lambda := g^j$ ,  $\theta_\lambda := \ell \cdot \omega_j$  and  $x_\lambda := \tau \cdot R_{-\theta_\lambda} D_{s_\lambda}^{-1} k$ .

The parameters  $g > 1$  and  $\tau > 0$  are sampling constants, which determine the fineness of the sampling grid,  $g$  for the scale parameters and  $\tau$  for the space parameters. The numbers  $(\omega_j)_{j \in \mathbb{N}_0}$  prescribe the step size of the angular sampling at each scale  $j \in \mathbb{N}_0$ .

Using curvelet parametrizations we can give a concise definition of  $\alpha$ -curvelet molecules.

**DEFINITION 2.5.** *Let  $\alpha \in [0, 1]$ . A family of  $\alpha$ -curvelet molecules is a family of  $\alpha$ -molecules with respect to an  $\alpha$ -curvelet parametrization.*

### 2.1.2 $\alpha$ -Shearlet Molecules

The concept of *shearlet molecules* first appeared in [4] and was later generalized in [1]. We extend the concept even further and introduce systems of  $\alpha$ -shearlet molecules, which resemble cone-adapted shearlet systems (see Section 2.2.4).

Let  $\alpha \in [0, 1]$  and fix constants  $g > 1$  and  $\tau > 0$ . Further, let  $(\eta_j)_{j \in \mathbb{N}_0}$  be a sequence of positive real numbers which obey  $\eta_j \asymp g^{-j(1-\alpha)}$  for  $j \in \mathbb{N}_0$ . Consider the discrete index set

$$\Lambda^s := \Lambda_0^s \cup \left\{ (\varepsilon, j, \ell, k) : \varepsilon \in \{0, 1\}, j \in \mathbb{N}, \ell \in \mathbb{Z} \text{ with } |\ell| \leq L_j \text{ and } L_j \lesssim g^{j(1-\alpha)}, k \in \mathbb{Z}^2 \right\}, \quad (1)$$

where  $\Lambda_0^s := \{(0, 0, 0, k) : k \in \mathbb{Z}^2\}$ . We define the shearlet system

$$\Sigma := \left\{ \sigma_\lambda : \lambda \in \Lambda^s \right\}, \quad (2)$$

with

$$\sigma_{(\varepsilon, j, \ell, k)}(\cdot) := g^{(1+\alpha)j/2} \psi_{j, \ell, k}^\varepsilon \left( D_{g^j}^\varepsilon S_{\ell, j}^\varepsilon \cdot -\tau k \right),$$

where  $D_a^0 := D_a$ ,  $D_a^1 := \text{diag}(a^\alpha, a)$ ,  $S_{\ell, j}^0 := \begin{pmatrix} 1 & \ell \eta_j \\ 0 & 1 \end{pmatrix}$ ,  $S_{\ell, j}^1 := \left( S_{\ell, j}^0 \right)^T$ , and  $\psi_{j, \ell, k}^\varepsilon \in L^2(\mathbb{R}^2)$ .

Under certain assumptions on the generators  $\psi_{j, \ell, k}^\varepsilon$  the system  $\Sigma$  constitutes a system of  $\alpha$ -shearlet molecules.

**DEFINITION 2.6.** *We call  $\Sigma$  a system of  $\alpha$ -shearlet molecules of order  $(R, M, N_1, N_2)$  if the generating functions  $\psi_{j, \ell, k}^0$  and  $\psi_{j, \ell, k}^1$  satisfy for every  $\beta \in \mathbb{N}_0^2$  with  $|\beta| \leq R$  the estimate*

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

with an implicit constant independent of the indices  $(\varepsilon, j, \ell, k) \in \Lambda^s$ .

Next we show the crucial fact that  $\alpha$ -shearlet molecules are instances of  $\alpha$ -molecules associated with a specific shearlet parametrization  $\Phi^s$ .

**DEFINITION 2.7.** *Let  $\Lambda^s$  be the  $\alpha$ -shearlet index set defined in (1). For each shearlet index  $(\varepsilon, j, \ell, k) = \lambda \in \Lambda^s$ , the associated  $\alpha$ -shearlet parametrization  $(\Lambda^s, \Phi^s)$  is given by*

$$\Phi^s(\lambda) = (s_\lambda, \theta_\lambda, x_\lambda) := \left( g^j, \varepsilon \pi/2 + \arctan(-\ell \eta_j), (S_{\ell, j}^\varepsilon)^{-1} D_{g^{-j}}^\varepsilon k \right).$$

Although their construction is based on shears instead of rotations, we can give a characterization of  $\alpha$ -shearlet molecules as  $\alpha$ -molecules in the same fashion as for  $\alpha$ -curvelet molecules, compare Definition 2.5.

**PROPOSITION 2.8** ([2]). *Let  $\Lambda^s$  be the  $\alpha$ -shearlet index set defined in (1). A family  $\Sigma = (\sigma_\lambda)_{\lambda \in \Lambda^s}$  defined as in (2) constitutes a system of  $\alpha$ -shearlet molecules of order  $(R, M, N_1, N_2)$  if and only if  $\Sigma$  constitutes a system of  $\alpha$ -molecules of the same order with respect to the associated  $\alpha$ -shearlet parametrization.*

## 2.2 Examples of $\alpha$ -Molecules

Many representation systems which exist in the literature fall into the framework of  $\alpha$ -molecules.

### 2.2.1 Wavelets

As a first example, we prove that the classical wavelet systems in  $L^2(\mathbb{R}^2)$ , obtained by the following tensor product construction, see e.g. [5] for details, are instances of  $\alpha$ -molecules.

Starting with a given multi-resolution analysis of  $L^2(\mathbb{R})$  with scaling function  $\phi^0 \in L^2(\mathbb{R})$  and wavelet  $\phi^1 \in L^2(\mathbb{R})$ , the functions  $\psi^e \in L^2(\mathbb{R}^2)$  are defined for every index  $e = (e_1, e_2) \in E$ , where  $E = \{0, 1\}^2$ , as the tensor products

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

These functions serve as the generators for the wavelet system defined below.

**DEFINITION 2.9.** *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. Further, let  $g > 1$ ,  $\tau > 0$  be fixed sampling parameters. We define the wavelet system*

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

The associated index set is given by

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

For simplicity, we restrict our further investigations to bandlimited systems with infinitely many vanishing moments. For this, we assume that  $\hat{\phi}^0, \hat{\phi}^1 \in C^R(\mathbb{R})$  for some  $R \in \mathbb{N}_0 \cup \{\infty\}$ , and that there are  $0 < a$  and  $0 < b < c$  such that

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

These conditions are fulfilled e.g. if  $\phi^0, \phi^1 \in L^2(\mathbb{R})$  are the generators of a Lemarie-Meyer wavelet system. We have the following result.

**PROPOSITION 2.10** ([2]). *Let  $g > 1$ ,  $\tau > 0$  be fixed, and assume that the functions  $\phi^0, \phi^1$  satisfy the assumptions above. Then the wavelet system  $W(\phi^0, \phi^1; g, \tau)$  constitutes a system of 1-molecules of order  $(R, \infty, \infty, \infty)$  with respect to the parametrization*

$$\Lambda^w \rightarrow \mathbb{P}, \quad (e, j, k) \mapsto (g^j, 0, \tau g^{-j} k).$$

Finally, we remark that also more general wavelet systems, among them systems of compactly supported wavelets, fall into the framework of  $\alpha$ -molecules.

### 2.2.2 Ridgelets

The observation that wavelets have a suboptimal approximation rate for images with anisotropic features led to the quest for new better suited representation systems. One approach, even predating curvelets, is the idea of *ridgelets*, which are designed to optimally approximate data with straight line singularities. Since up to now there does not exist a uniform definition in the literature, we shortly review their historical evolution and try to extract the main conceptual ideas. This will then motivate the definition of *ridgelet molecules* as suggested in [6].

The earliest version of the ridgelet transform was introduced by Candès [7] in 1998. It uses a univariate wavelet  $\psi$  to map a function  $f \in L^2(\mathbb{R}^d)$  to its transform coefficients

$$\langle f, \sqrt{a} \psi(as \cdot x - t) \rangle, \quad s \in \mathbb{S}^{d-1}, t \in \mathbb{R}, a \in \mathbb{R}_+,$$

where  $\cdot$  denotes the Euclidean inner product in  $\mathbb{R}^d$ . The function  $x \mapsto \sqrt{a} \psi(as \cdot x - t)$  is a *ridge function* (hence the name ridgelet) which only varies in the direction  $s$ . Unfortunately, since this function is not in  $L^2(\mathbb{R}^d)$ , the definition, as it stands, does not make sense for every  $f \in L^2(\mathbb{R}^d)$ . Similar to the continuous Fourier transform, however, the continuous version of this transform can be well-defined.

In order to avoid the problems associated with the lack of integrability of ridge functions, Donoho [8] relaxed the definition of a ridgelet a little, allowing them a slow decay in the other directions. In the spirit of this more

general approach, as pointed out by Grohs [6], one might define a ridgelet system as a system of functions of the form

$$\sqrt{a}\rho(D_a R_s x - t) \quad (3)$$

utilizing some generator  $\rho \in L^2(\mathbb{R}^d)$ , which needs to be oscillatory in one coordinate direction. Note that here  $D_a = \text{diag}(a, 1, \dots, 1) \in \mathbb{R}^{d \times d}$  and  $R_s$  denotes rotation by  $s \in \mathbb{S}^{d-1}$ .

A yet more general viewpoint, also adopted in [6], is to characterize ridgelets by their localization properties in space and frequency, without enforcing the rigid condition of being exactly of the form (3). This leads to the notion of *ridgelet molecules*, which in the language of  $\alpha$ -molecules can easily be defined for  $d = 2$ .

DEFINITION 2.11. *A system of 0-curvelet molecules is called a system of ridgelet molecules.*

We see that ridgelet molecules readily fall into the more general concept of  $\alpha$ -curvelet molecules, which we already have at hand. They are just the special case for  $\alpha = 0$ .

In [6] bandlimited tight ridgelet frames were constructed. In the following section we will adapt this construction process to obtain bandlimited tight  $\alpha$ -curvelet frames for every  $\alpha \in [0, 1]$ . In particular, we will obtain a tight frame of ridgelet molecules. The following statement is included in Proposition 2.13.

PROPOSITION 2.12 ([2]). *The tight ridgelet frame  $C_0(W^{(0)}, W^{(1)}, V)$  constructed in (4) is a system of ridgelet molecules of order  $(\infty, \infty, \infty, \infty)$  with parameters  $g = 2$  and  $\tau = 1$ .*

### 2.2.3 Curvelets and Hybrid Curvelets

In 2002 Candès and Donoho [9] introduced the second generation of curvelets, which we nowadays simply refer to as *curvelets*. Their construction involves a parabolic scaling law, which as pointed out in the original paper has many unique properties. Parabolic scaling can be viewed as a natural compromise between isotropic scaling, as utilized for wavelets, and 0-scaling, i.e. scaling in only one coordinate direction, as utilized for ridgelets.

Allowing more general  $\alpha$ -scaling for  $\alpha \in [0, 1]$  yields, what we will call,  $\alpha$ -curvelets or hybrid curvelets. Thus, we obtain a whole scale of representation systems, which interpolates between wavelets for  $\alpha = 1$  on the one end and ridgelets for  $\alpha = 0$  on the other end. In this sense, curvelets can be viewed as lying in between ridgelets and wavelets.

Following the construction principle of the tight ridgelet frames in [6] we now construct bandlimited tight frames of  $\alpha$ -curvelets for every  $\alpha \in [0, 1]$ . This construction can also be seen as a variation of the classical second generation curvelet frame from [9]. As in [6, 9] we separate the radial and angular components to simplify the construction.

Let us begin with the construction of the radial functions  $W^{(j)}$  for  $j \in \mathbb{N}_0$ . We first define the  $C^\infty$ -functions  $\widetilde{W}^{(0)} : \mathbb{R}_+ \rightarrow [0, 1]$  and  $\widetilde{W} : \mathbb{R}_+ \rightarrow [0, 1]$  with the following properties:

$$\begin{aligned} \text{supp } \widetilde{W}^{(0)} &\subset [0, 2), & \widetilde{W}^{(0)}(r) &= 1 \quad \text{for all } r \in [0, \frac{3}{2}], \\ \text{supp } \widetilde{W} &\subset (\frac{1}{2}, 2), & \widetilde{W}(r) &= 1 \quad \text{for all } r \in [\frac{3}{4}, \frac{3}{2}]. \end{aligned}$$

Then we put for  $j \in \mathbb{N}$  and  $r \in \mathbb{R}_+$

$$\widetilde{W}^{(j)}(r) := \widetilde{W}(2^{-j}r).$$

In a final step we rescale for every  $j \in \mathbb{N}_0$  (to obtain an integer grid later)

$$W^{(j)}(r) := \widetilde{W}^{(j)}(8\pi r) \quad , r \in \mathbb{R}_+.$$

Next, we define the angular functions  $V^{(j, \ell)} : \mathbb{S}^1 \rightarrow [0, 1]$ , where  $\mathbb{S}^1 \subset \mathbb{R}^2$  denotes the unit circle,  $j \in \mathbb{N}$  and the index  $\ell$  runs through  $0, \dots, L_j - 1$  with

$$L_j = 2^{\lfloor j(1-\alpha) \rfloor}, \quad j \in \mathbb{N}.$$

We start with a  $C^\infty$ -function  $V : \mathbb{R} \rightarrow [0, 1]$ , living on the whole of  $\mathbb{R}$ , satisfying

$$\text{supp } V \subset [-\frac{3}{4}\pi, \frac{3}{4}\pi] \quad \text{and} \quad V(t) = 1 \text{ for all } t \in [-\frac{\pi}{2}, \frac{\pi}{2}].$$

For every  $j \in \mathbb{N}$  we let  $\tilde{V}^{(j,0)} : \mathbb{S}^1 \rightarrow [0, 1]$  be the restriction of the scaled version  $V(2^{\lfloor j(1-\alpha) \rfloor} \cdot)$  of the function  $V$  to the interval  $[-\pi, \pi]$ . Since  $[-\pi, \pi]$  can be identified via  $\varphi : t \mapsto e^{it}$  with  $\mathbb{S}^1$ , this yields a function  $\tilde{V}^{(j,0)}$  on  $\mathbb{S}^1$ . Notice, that in our case this is a  $C^\infty$ -function.

In order to obtain real-valued curvelets, we put in a symmetrization step and define

$$V^{(j,0)}(\xi) := \tilde{V}^{(j,0)}(\xi) + \tilde{V}^{(j,0)}(-\xi) \quad \text{for } \xi \in \mathbb{S}^1.$$

We define the angles  $\omega_j = \pi 2^{-\lfloor j(1-\alpha) \rfloor}$  for  $j \in \mathbb{N}$  and let for  $\ell = 0, 1, \dots, L_j - 1$

$$R_{j,\ell} = \begin{pmatrix} \cos(\ell\omega_j) & -\sin(\ell\omega_j) \\ \sin(\ell\omega_j) & \cos(\ell\omega_j) \end{pmatrix}$$

be the rotation matrix by the angle  $\ell\omega_j$ . By rotating  $V^{(j,0)}$  we finally get  $V^{(j,\ell)} : \mathbb{S}^1 \rightarrow [0, 1]$ ,

$$V^{(j,\ell)}(\xi) := V^{(j,0)}(R_{j,\ell}\xi) \quad \text{for } \xi \in \mathbb{S}^1.$$

In order to secure the tightness of the frame we need the function

$$\Phi(\xi) := W^{(0)}(|\xi|)^2 + \sum_{j,\ell} W^{(j)}(|\xi|)^2 V^{(j,\ell)}\left(\frac{\xi}{|\xi|}\right)^2.$$

Now we can define the functions  $\psi_0$  and  $\psi_{j,\ell}$  on the Fourier domain by

$$\hat{\psi}_0(\xi) := \frac{W^{(0)}(|\xi|)}{\sqrt{\Phi(\xi)}} \quad \text{and} \quad \hat{\psi}_{j,\ell}(\xi) = \frac{W^{(j)}(|\xi|) V^{(j,\ell)}\left(\frac{\xi}{|\xi|}\right)}{\sqrt{\Phi(\xi)}}.$$

It is obvious that  $\hat{\psi}_0, \hat{\psi}_{j,\ell} \in C^\infty(\mathbb{R}^2)$ . Moreover, these functions are compactly supported.

The final frame is obtained by taking translates in the time domain of the above functions. For  $y \in \mathbb{R}^2$  let  $T_y$  denote the translation operator, given by  $T_y f(x) = f(x - y)$  for  $x \in \mathbb{R}^2$ . Further, let  $D_s = \text{diag}(s, s^\alpha)$  for  $s \in \mathbb{R}_+$  be the  $\alpha$ -scaling operator. We define the collection of functions

$$\psi_{0,k} := T_k \psi_0 \quad \text{and} \quad \psi_{j,\ell,k} := 2^{-j(1+\alpha)/2} \cdot T_{x_{j,\ell,k}} \psi_{j,\ell},$$

where  $j \in \mathbb{N}$ ,  $\ell = 0, 1, \dots, L_j - 1$ ,  $k \in \mathbb{Z}^2$  and  $x_{j,\ell,k} = R_{j,\ell}^{-1} D_{2^j}^{-1} k$ . We will use the following notation for this system,

$$C_\alpha(W^{(0)}, W^{(1)}, V) = \left\{ \psi_{0,k} : k \in \mathbb{Z}^2 \right\} \cup \left\{ \psi_{j,\ell,k} : j \in \mathbb{N}, k \in \mathbb{Z}^2, \ell \in \{0, 1, \dots, L_j - 1\} \right\}. \quad (4)$$

**PROPOSITION 2.13** ([2]). *Let  $\alpha \in [0, 1]$ . The  $\alpha$ -curvelet system  $C_\alpha(W^{(0)}, W^{(1)}, V)$  constructed above constitutes a tight frame for  $L^2(\mathbb{R}^2)$ . Moreover, it is a system of  $\alpha$ -curvelet molecules of order  $(\infty, \infty, \infty, \infty)$  with parameters  $g = 2$  and  $\tau = 1$ .*

#### 2.2.4 Shearlets and Hybrid Shearlets

*Shearlets* were introduced in 2005 by Kutyniok, Labate, Lim and Weiss [10]. A shearlet system is generated by a function by changing its scale, spatial positions and orientations. Like curvelets it uses parabolic scaling for the scale change. The main difference to curvelets is that in order to change their orientation shears are employed instead of rotations. This makes shearlets more adapted to a digital grid, which is favorable in a discrete setting.

To avoid large shears and still cover all orientations, usually two generators with orthogonal orientations are used. Moreover, a distinct generator is utilized for the coarse-scale elements. Such shearlet systems are called *cone-adapted*, since one can picture the frequency plane as divided into a horizontal and a vertical cone, as well as a coarse-scale box, associated with the respective generators.

Similar to hybrid curvelets, the notion of a shearlet can be generalized to comprise  $\alpha$ -scaling. These hybrid shearlets or  $\alpha$ -shearlets have been defined and examined in [11], at least for the range  $\alpha \in [\frac{1}{2}, 1)$ .

Before we recall their definition, we remark that the anisotropy of the scaling is parametrized by a different parameter in [11], for which we will use the notation  $\beta$  to distinguish it from the  $\alpha$  utilized here. The parameter  $\beta$  ranges in  $(1, \infty)$ , which corresponds to  $\alpha \in (0, 1)$  in our setup, since they are related by  $\alpha = \beta^{-1}$ . For  $j \in \mathbb{N}_0$  the scaling matrices  $A_{2^j} = \text{diag}(2^{j\beta/2}, 2^{j/2})$  and  $\tilde{A}_{2^j} = \text{diag}(2^{j/2}, 2^{j\beta/2})$  are defined, and for  $s \in \mathbb{R}$  the shear matrix

$$S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.$$

Following [11] the cone-adapted discrete shearlet system is then defined as follows.

DEFINITION 2.14. *For parameters  $c \in \mathbb{R}_+$  and  $\beta \in (1, \infty)$  the cone-adapted discrete shearlet system  $SH(\phi, \psi, \tilde{\psi}; c, \beta)$  generated by  $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$  is defined by*

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

where

$$\begin{aligned} \Phi(\phi; c, \beta) &= \{\phi_k = \phi(\cdot - k) : k \in c\mathbb{Z}^2\}, \\ \Psi(\psi; c, \beta) &= \{\psi_{j,\ell,k} = 2^{j(\beta+1)/4} \psi(S_\ell A_{2^j} \cdot -k) : j \geq 0, |\ell| \leq \lceil 2^{j(\beta-1)/2} \rceil, k \in c\mathbb{Z}^2\}, \\ \tilde{\Psi}(\tilde{\psi}; c, \beta) &= \{\tilde{\psi}_{j,\ell,k} = 2^{j(\beta+1)/4} \tilde{\psi}(S_\ell^T \tilde{A}_{2^j} \cdot -k) : j \geq 0, |\ell| \leq \lceil 2^{j(\beta-1)/2} \rceil, k \in c\mathbb{Z}^2\}. \end{aligned}$$

In the following we want to analyze compactly supported systems. For simplicity we assume the separability of the generator  $\psi \in L^2(\mathbb{R}^2)$ , i.e.  $\psi(x_1, x_2) = \psi_1(x_1)\psi_2(x_2)$ , and let  $\tilde{\psi}$  be the rotation of  $\psi$  by  $\pi/2$ . We make the following assumptions,

$$\psi_1 \in C_0^{N_1}(\mathbb{R}) \quad \text{and} \quad \psi_2 \in C_0^{N_1+N_2}(\mathbb{R}).$$

Moreover, for  $\psi_1$  we assume  $M \in \mathbb{N}_0$  vanishing moments.

PROPOSITION 2.15 ([2]). *If the generators  $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$  satisfy the assumptions above, the cone-adapted shearlet system  $SH(\phi, \psi, \tilde{\psi}; c, \beta)$  is a system of  $\alpha$ -shearlet molecules for  $\alpha = \beta^{-1}$  of order  $(R, M - R, N_1, N_2)$ , where  $R \in \{0, \dots, M\}$ . The parameters of the  $\alpha$ -shearlet parametrization are given by  $\tau = c$ ,  $g = 2^{\beta/2}$ ,  $\eta_j = g^{-j(1-\alpha)}$  and  $L_j = \lceil g^{j(1-\alpha)} \rceil$ .*

These results show that the  $\alpha$ -molecule concept indeed has the ability to unite many representation systems under one roof. In the next section we will see what we gain by this.

### 2.3 Index Distance and Decay of the Cross-Gramian

An essential ingredient for the theory is the fact that the parameter space  $\mathbb{P}$  can be equipped with a natural (pseudo-)metric. In [12] Hart-Smith introduced a distance on  $\mathbb{P}$ , which was used in [1] for parabolic molecules, and which can be seen as a predecessor to this metric.

We should mention that this metric is not a distance in the strict sense. As we will see later, it measures the correlation of a pair of  $\alpha$ -molecules, associated to the corresponding points in  $\mathbb{P}$ .

DEFINITION 2.16. *For two indices  $\lambda \in \Lambda$  and  $\mu \in M$  we define the index distance*

$$\omega(\lambda, \mu) := \max\left(\frac{s_\lambda}{s_\mu}, \frac{s_\mu}{s_\lambda}\right) (1 + d(\lambda, \mu)),$$



and

$$d(\lambda, \mu) := s_0^{2(1-\alpha)} |\theta_\lambda - \theta_\mu|^2 + s_0^{2\alpha} |x_\lambda - x_\mu|^2 + \frac{s_0^2}{1 + s_0^{2(1-\alpha)} |\theta_\lambda - \theta_\mu|^2} |\langle e_\lambda, x_\lambda - x_\mu \rangle|^2,$$

where  $s_0 = \min(s_\lambda, s_\mu)$  and  $e_\lambda = (\cos(\theta_\lambda), -\sin(\theta_\lambda))^\top = R_{-\theta_\lambda} e_1$  is the co-direction.

Note, that in essence this definition provides a notion of distance on the parameter space  $\mathbb{P}$ . In order not to overload the notation, we did not explicitly specify the parametrizations used to transfer this distance to the generic index sets  $\Lambda$  and  $M$ .

The following theorem expresses the relation between the index distance on  $\mathbb{P}$  and the correlation of  $\alpha$ -molecules. It states, that a high distance of two indices can be interpreted as a low cross correlation of the associated  $\alpha$ -molecules. The proof is quite technical and we refer to [2] for the details.

**THEOREM 2.17 ([2]).** *Let  $\alpha \in [0, 1]$  and  $(m_\lambda)_{\lambda \in \Lambda}, (p_\mu)_{\mu \in M}$  be systems of  $\alpha$ -molecules of order  $(R, M, N_1, N_2)$ , such that  $s_\lambda, s_\mu \geq c > 0$  for all  $\lambda \in \Lambda$  and  $\mu \in M$ . Further assume, that for some  $N \in \mathbb{N}$  it holds*

$$R \geq 2N, \quad M > 3N - \frac{3-\alpha}{2}, \quad N_1 \geq N + \frac{1+\alpha}{2}, \quad N_2 \geq 2N.$$

Then

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

Put in yet another way, this result shows that the Gramian matrix between two systems of  $\alpha$ -molecules of high order satisfies a strong off-diagonal decay property and is in that sense very close to a diagonal matrix. As we shall see in the next section this has a number of immediate consequences for the approximation properties of  $\alpha$ -molecules.

### 3. SPARSE APPROXIMATION WITH $\alpha$ -MOLECULES

In this section we analyze the approximation properties of  $\alpha$ -molecules. The main result will be that any two systems of alpha molecules, which are consistent in a yet to be defined sense and which have sufficiently high order, exhibit the same approximation behavior.

Prior to this investigation, we briefly discuss 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 handle  $f$ , one usually represents it with respect to some representation system  $(m_\lambda)_{\lambda \in \Lambda} \subset L^2(\mathbb{R}^2)$ ,

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

since the coefficients  $c_\lambda \in \mathbb{R}$  of this expansion are much handier. In practice, since we have to account for noise, it is also necessary to ensure the robustness of such a representation. Here the notion of a frame comes into play, see e.g. [13].

Frames ensure stable measurement and stable reconstruction. Moreover, the coefficients can be calculated by  $\langle f, \tilde{m}_\lambda \rangle$  using a dual frame  $(\tilde{m}_\lambda)_\lambda$ . It is important to note, that a representation with respect to a frame need not be unique. However, using the canonical dual frame for the computation of the coefficients yields a special sequence, which is customarily referred to as *the* sequence of frame coefficients.

When trying to represent  $f$  with respect to a frame system  $(m_\lambda)_\lambda \subseteq L_2(\mathbb{R}^2)$ , we are confronted with yet another problem. Since in the real world it is not possible to store infinitely many coefficients, we have to approximate  $f$  utilizing only a finite subset of this system. If the subset is restricted to contain at most  $N$  elements, we obtain what is called an  *$N$ -term approximation* for  $f$  with respect to  $(m_\lambda)_\lambda$ . The best  $N$ -term approximation, which is the best we can do if we only have storage capacity for  $N$  coefficients, is denoted by  $f_N$  and defined by

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

It is clear, that for efficient encoding it is necessary that the error  $\|f - f_N\|_2$  of best  $N$ -term approximation decays quickly with growing  $N$ . Since the computation of the best  $N$ -term approximation is far from being understood, usually instead the error of the  $N$ -term approximation, obtained by keeping the  $N$  largest coefficients  $(\langle f, \tilde{m}_\lambda \rangle)_{\lambda \in \Lambda}$ , is considered. This error certainly also provides a bound for the error of best  $N$ -term approximation.

There is a close relationship between the  $N$ -term approximation rate, achieved by a frame, and the decay rate of the corresponding frame coefficients. The decay of the frame coefficients can be measured in terms of the  $\ell_p$ -(quasi)-norms for  $p > 0$ . The following lemma shows that membership of the coefficient sequence in an  $\ell_p$ -space for small  $p$  implies good  $N$ -term approximation rates, see also [14, 15].

**LEMMA 3.1.** *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.*

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

### 3.1 Sparsity Equivalence and Consistency of Parametrizations

As discussed above, two frames have similar approximation properties, if the corresponding coefficient sequences exhibit the same sparsity. The notion of sparsity equivalence from [1] is a useful tool to compare such sparsity.

**DEFINITION 3.2.** *Let  $(m_\lambda)_{\lambda \in \Lambda}$  and  $(p_\mu)_{\mu \in M}$  be systems of  $\alpha$ -molecules and let  $0 < p \leq 1$ . Then  $(m_\lambda)_{\lambda \in \Lambda}$  and  $(p_\mu)_{\mu \in M}$  are called sparsity equivalent in  $\ell_p$ , if*

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

Our next goal is to find conditions which ensure that two systems of  $\alpha$ -molecules are sparsity equivalent. Here the consistency of parametrizations comes into play.

**DEFINITION 3.3.** *Two parametrizations  $(\Lambda, \Phi_\Lambda)$  and  $(\Gamma, \Phi_\Gamma)$  are called  $k$ -consistent for  $k > 0$  if*

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

In combination with Theorem 2.17 the consistency is the essential tool to decide whether two frames of  $\alpha$ -molecules are sparsity equivalent.

**THEOREM 3.4** ([2]). *Two frames  $(m_\lambda)_{\lambda \in \Lambda}$  and  $(p_\mu)_{\mu \in \Gamma}$  of  $\alpha$ -molecules of order  $(R, M, N_1, N_2)$  with  $k$ -consistent parametrizations for some  $k > 0$ , are sparsity equivalent in  $\ell_p$ ,  $0 < p \leq 1$ , if*

$$R \geq 2\frac{k}{p}, \quad M > 3\frac{k}{p} - \frac{3-\alpha}{2}, \quad N_1 \geq \frac{k}{p} + \frac{1+\alpha}{2}, \quad \text{and} \quad N_2 \geq 2\frac{k}{p}.$$

We see that, as long as the parametrizations are consistent, the sparsity equivalence can be controlled by the order of the molecules. The proof relies on a well-known result concerning operator norms.

**LEMMA 3.5.** *Let  $I, J$  be two discrete index sets,  $p > 0$  and  $r = \min(1, p)$ . Further, let  $\mathbf{A} : \ell_p(I) \rightarrow \ell_p(J)$  be a linear mapping defined by its matrix representation  $\mathbf{A} = (A_{i,j})_{i \in I, j \in J}$ . Then we have the bound*

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

### 3.1.1 Consistency of Curvelet and Shearlet Parametrizations

It is a central result that curvelet and shearlet parametrizations are consistent.

PROPOSITION 3.6 ([2]). *Let  $\alpha \in [0, 1]$  and  $(\Lambda, \Phi_\Lambda)$  and  $(\Gamma, \Phi_\Gamma)$  be either  $\alpha$ -curvelet or  $\alpha$ -shearlet parametrizations. Then  $(\Lambda, \Phi_\Lambda)$  and  $(\Gamma, \Phi_\Gamma)$  are  $k$ -consistent for  $k > 2$ .*

The proof relies on the following lemma.

LEMMA 3.7 ([2]). *Let  $(\Lambda, \Phi_\Lambda)$  be an  $\alpha$ -curvelet or  $\alpha$ -shearlet parametrization and let  $\mu = (s_\mu, \theta_\mu, x_\mu)$  be an arbitrary fixed point of the parameter space  $\mathbb{P}$ . For  $N > 2$  it holds*

$$\sum_{\substack{\lambda \in \Lambda \\ s_\lambda \text{ fixed}}} (1 + d(\lambda, \mu))^{-N} \leq C \cdot \max\left(\frac{s_\lambda}{s_\mu}, 1\right)^2,$$

where the constant  $C > 0$  is independent of  $\mu$  and  $s_\lambda$ .

### 3.2 Sparse Image Approximation

A customarily employed model for image data is the class  $\mathcal{E}^\beta(\mathbb{R}^2)$  of *cartoon images* defined by

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

where  $\beta \in \mathbb{R}$ ,  $B \subset [0, 1]^2$  with  $\partial B$  being a closed  $C^\beta$ -curve, and  $f_0, f_1 \in C_0^\beta([0, 1]^2)$ .

It was shown [11, 16] that for  $\beta \in (1, 2]$  the optimally achievable decay rate of the approximation error for the class  $\mathcal{E}^\beta(\mathbb{R}^2)$ , under the natural assumption of polynomial depth search, is

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

Furthermore, in [9, 11, 14, 17] rotation-based as well as shear-based systems were constructed, which attain this rate up to a log-factor. The similar approximation behavior of these systems should not come as a surprise, since these systems are instances of  $\alpha$ -molecules with consistent parametrizations.

For convenience, we fix the construction (4) as a reference system and make the following definition.

DEFINITION 3.8. *We call a parametrization  $(\Lambda, \Phi_\Lambda)$   $k$ -admissible for  $k > 0$ , if it is  $k$ -consistent with the  $\alpha$ -curvelet parametrization of the frame of  $\alpha$ -molecules  $C_\alpha(W^{(0)}, W^{(1)}, V)$  constructed in Section 2.2.3.*

Now we can formulate our final result.

THEOREM 3.9 ([2]). *Let  $\alpha \in [\frac{1}{2}, 1)$  and  $\beta = \alpha^{-1}$ . Assume that  $(m_\lambda)_{\lambda \in \Lambda}$  is a frame of  $\alpha$ -molecules of order  $(R, M, N_1, N_2)$  with respect to the parametrization  $(\Lambda, \Phi_\Lambda)$  such that*

(i)  $(\Lambda, \Phi_\Lambda)$  is  $k$ -admissible for every  $k > 2$ ,

(ii) it holds that  $R \geq 2 + \frac{2}{\alpha}$ ,  $M > \frac{3}{2} + \frac{3}{\alpha} + \frac{\alpha}{2}$ ,  $N_1 \geq \frac{3}{2} + \frac{1}{\alpha} + \frac{\alpha}{2}$ , and  $N_2 \geq 2 + \frac{2}{\alpha}$ .

Then the frame  $(m_\lambda)_{\lambda \in \Lambda}$  possesses an almost best  $N$ -term approximation rate of order  $N^{-\frac{\beta}{2} + \varepsilon}$ ,  $\varepsilon > 0$  arbitrary, for the cartoon image class  $\mathcal{E}^\beta(\mathbb{R}^2)$ .

We remark that condition (i) holds in particular for the curvelet and shearlet parametrizations. Hence this result allows a simple derivation of the results in [9, 11, 14, 17]. In fact, Theorem 3.9 provides a systematic way to prove results on sparse approximation of cartoon images.

### ACKNOWLEDGMENTS

S.K. is supported by the DFG Collaborative Research Center TRR 109 ‘‘Discretization in Geometry and Dynamics’’. G.K. acknowledges support by the Einstein Foundation Berlin, by Deutsche Forschungsgemeinschaft (DFG) Grant SPP-1324 KU 1446/13 and DFG Grant KU 1446/14, by the DFG Collaborative Research Center TRR 109 ‘‘Discretization in Geometry and Dynamics’’, and by the DFG Research Center MATHEON ‘‘Mathematics for key technologies’’ in Berlin, Germany..

## REFERENCES

- [1] Grohs, P. and Kutyniok, G., Parabolic molecules, *Found. Comput. Math.* (to appear).
- [2] Grohs, P., Keiper, S., Kutyniok, G., and Schäfer, M.,  $\alpha$ -Molecules, *preprint* (2013).
- [3] Candès, E. J. and Demanet, L., The curvelet representation of wave propagators is optimally sparse, *Comm. Pure Appl. Math.* **58**, 1472–1528 (2002).
- [4] Guo, K. and Labate, D., Representation of Fourier integral operators using shearlets, *J. Fourier Anal. Appl.* **14**, 327–371 (2008).
- [5] Wojtaszczyk, P., *A Mathematical Introduction to Wavelets*, Cambridge University Press, Cambridge (1997).
- [6] Grohs, P., Ridgelet-type frame decompositions for Sobolev spaces related to linear transport, *J. Fourier Anal. Appl.* **18**(2), 309–325 (2012).
- [7] Candès, E., *Ridgelets: Theory and Applications*, PhD thesis, Stanford University (1998).
- [8] Donoho, D. L., Orthonormal ridgelets and linear singularities, *SIAM J. Math. Anal.* **31**(5), 1062–1099 (2000).
- [9] Candès, E. J. and 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).
- [10] Kutyniok, G., Labate, D., Lim, W.-Q, and Weiss, G., Sparse multidimensional representation using shearlets, *Wavelets XI(San Diego, CA), SPIE Proc.* **5914**, 254–262 (2005).
- [11] Keiper, S., *A Flexible Shearlet Transform - Sparse Approximation and Dictionary Learning*, Bachelor's thesis, TU Berlin (2013).
- [12] Smith, H., A parametrix construction for wave equations with  $C^{1,1}$ -coefficients, *Ann. Inst. Fourier* **48**, 797–835 (1998).
- [13] Christensen, O., *An Introduction to Frames and Riesz Bases*, Birkhäuser (2003).
- [14] Kutyniok, G. and Lim, W.-Q, Compactly supported shearlets are optimally sparse, *J. Approx. Theory* **163**(11), 1564–1589 (2011).
- [15] DeVore, R. A., Nonlinear approximation, *Acta Numerica* **7**, 51–150 (1998).
- [16] Donoho, D. L., Sparse components of images and optimal atomic decomposition, *Constr. Approx.* **17**, 353–382 (2001).
- [17] Guo, K. and Labate, D., Optimally sparse multidimensional representation using shearlets, *SIAM J. Math. Anal.* **39**, 298–318 (2007).