Multivariate α-Molecules

Axel Flinth   Martin Schäfer [*]

October 15, 2018

Abstract

The suboptimal performance of wavelets with regard to the approximation of multivariate data gave rise to new representation systems, specifically designed for data with anisotropic features. Some prominent examples of these are given by ridgelets, curvelets, and shearlets, to name a few.

The great variety of such so-called directional systems motivated the search for a common framework, which unites many under one roof and enables a simultaneous analysis, for example with respect to approximation properties. Building on the concept of parabolic molecules, the recently introduced framework of α-molecules does in fact include the previous mentioned systems. Until now however it is confined to the bivariate setting, whereas nowadays one often deals with higher dimensional data. This motivates the extension of this unifying theory to dimensions larger than 2, put forward in this work. In particular, we generalize the central result that the cross-Gramian of any two systems of α-molecules will to some extent be localized.

As an exemplary application, we investigate the sparse approximation of video signals, which are instances of 3D data. The multivariate theory allows us to derive almost optimal approximation rates for a large class of representation systems.

Keywords: Wavelets, Shearlets, Anisotropic Scaling, α-Molecules, Multiscale Analysis, Nonlinear Approximation.

2010 MSC: 41A30, 41A63, 42C40

1 Introduction

[1]One of the most influential modern developments in applied harmonic analysis has undeniably been the introduction of *wavelets* [13]. Their construction is based on dilations and translations of a (finite) set of generating functions $(g_\lambda)_\lambda \subseteq L^2(\mathbb{R}^d)$. By carefully choosing the generators, the resulting systems can become frames or even orthonormal bases of the space $L^2(\mathbb{R}^d)$. Furthermore, additional properties can be obtained, such as e.g. smoothness or compact support. Some real-world applications of wavelets today are e.g. data compression (e.g. JPEG2000) or restoration tasks in imaging sciences [1]. In the field of PDE's wavelets nowadays play a central role in solving elliptic equations [7].

The great success of wavelet systems – besides their elegant construction principle and available fast numerical implementations – rests upon the fact that they provide efficient multiscale representations for various types of data. In particular, they optimally sparsely approximate functions $f : \mathbb{R}^d \to \mathbb{C}$, which are smooth apart from (a finite number of) point singularities, in the sense of fast decay of the N-term approximation error. Since such singularities are the only ones that occur in 'reasonable' 1D data, we can safely say that wavelets are optimal for approximating one-dimensional functions.

Moving up a dimension however, the situation changes completely. Two-dimensional data may well have singularities along curves and is often governed by such anisotropic features – think of edges in images for instance. A widely used model for such data is the class of cartoon-like functions [15], i.e. functions which are smooth except for a curve-like singularity (see Section 4 for details). For this class, wavelets do not perform optimally any more [15], and hence other approaches have to be considered.

[*]Both authors are affiliated to the Institut für Mathematik, Technische Universität Berlin.

E-mail addresses: flinth@math.tu-berlin.de, schafer@math-tu.berlin.de

1.1 Directional Representation Systems

The reason for the non-optimal performance of wavelets in a multivariate setting is due to their isotropic scaling law, which is not optimally suited for resolving anisotropic, i.e. directional, features. Therefore, many systems employing some form of anisotropic scaling and thus better suited for this task have been considered in recent years. Subsequently, we briefly recall a few of these constructions for motivation purposes, but by no means this shall be a complete overview.

1.1.1 Ridgelets

Aiming to approximate functions with line singularities, Candès defined so-called *ridgelets* [2] as translated, rotated and dilated versions of a ridge function, which is constant orthogonal to some specified direction $\eta \in \mathbb{S}^{d-1}$. Since such functions are not square-integrable, the concept was adjusted by Donoho [3] allowing ridgelets a slow decay orthogonal to the η-direction, leading to a modified notion adopted for instance in [23, 26, 25]. In the new sense, a system of ridgelets is constructed performing rotations, translations and directional scaling on a generator $g \in L^2(\mathbb{R}^d)$ with corresponding scaling matrix

$$A_{0,s} = \begin{pmatrix} s & 0 \\ 0 & 1 \end{pmatrix}, \qquad s > 0.$$

Tight ridgelet frames of this type were constructed e.g. in [23, 25].

1.1.2 Curvelets and Shearlets

A true breakthrough was achieved by Candès and Donoho in 2002 with the introduction of *curvelets* [5], the first system to provide a provably (almost) optimal approximation rate for a certain class of cartoonlike functions. Again, the idea is to apply certain rotation, translation and scaling operations to a generating function. The major novelty was the use of *parabolic scaling*, a compromise between directional scaling used for ridgelets and isotropic scaling used for wavelets, described by a matrix of the form

$$A_{\frac{1}{2},s} = \begin{pmatrix} s & 0 \\ 0 & s^{1/2} \end{pmatrix}, \qquad s > 0. \tag{1}$$

This type of scaling is specifically adapted to data with C^2-discontinuity curves, since it leaves the parabola invariant and produces functions with essential support in a rectangle of size *'width $\approx$ length^2'*. We mention that in the actual construction of the classical tight frame of curvelets [5], the translations and rotations are applied to a set of generators, related to each other by a parabolic scaling law realised not by (1) but by dilations with respect to polar coordinates.

A few years after curvelets in 2005, *shearlets* were developed mainly by Kutyniok, Labate, Lim, and Weiss [40]. They also scale parabolically and feature the same celebrated approximation properties as curvelets for cartoon-like functions [30, 38]. The main difference is that shearings and not rotations are used for the change of direction. The choice of shears makes shearlets more adapted to a digital grid, since shearings given by the matrices

$$S_h = \begin{pmatrix} 1 & 0 \\ h & 1 \end{pmatrix} \quad \text{and} \quad S_h^T = \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix}, \quad h \in \mathbb{R}, \tag{2}$$

leave the digital grid invariant. This is favorable in a discrete setting and bears the advantage of a unified treatment of the continuum and digital realm. It should be noted that some actual constructions of shearlet systems are not entirely faithful to the original idea of applying shears, translations, and parabolic scalings to a single generator: The probably most notable and widely used adjustment is the idea of *cone-adaption*, where several generators with different orientations are used in order to avoid large shear parameters. We will discuss this strategy in greater detail later in the article.

Nowadays, shearlets are a widely used directional representation system with applications ranging from imaging science [16], simulations of inverse scattering problems [43] to solvers for transport equations [12]. For more information we refer to the book [39].

1.1.3 α-Curvelets (and -Shearlets)

The systems we have presented so far all utilize different versions of the scaling matrix

$$A_{\alpha,s} = \begin{pmatrix} s & 0 \\ 0 & s^\alpha \end{pmatrix}, \qquad s > 0, \tag{3}$$

where the parameter $\alpha \in [0,1]$ specifies the degree of anisotropy in the scaling: $\alpha = 1$ corresponds to wavelets, $\alpha = \frac{1}{2}$ to curvelets and shearlets, and $\alpha = 0$ to ridgelets. This observation was used in [25] to define *α-curvelets*, and associated bandlimited tight frames were constructed for every $\alpha \in [0,1]$. Similar to α-curvelets, the notion of a shearlet can be generalized to comprise α-scaling. The resulting *α-shearlets* have been defined and examined in [35, 41] (for the range $\alpha \in [\frac{1}{2}, 1)$).

1.2 A common Framework

The directional systems described above are all constructed using the same idea: start with a set of generators, and then perform scalings (with some degree of anisotropicity), changes of direction (e.g. rotations or shears) and translations. Further, in order to obtain systems with desirable properties, some regularity condition on the generators has to be posed. Having this in mind, it seems possible to regard all such systems as certain instances of a more general concept.

1.2.1 Parabolic Molecules

In 2011, Grohs and Kutyniok introduced the concept of *parabolic molecules* [27], which allows to derive classical curvelets and shearlets as special instances of the same general construction process. Starting from a set of generators, a system of parabolic molecules is obtained via parabolic dilations, rotations, and translations. The essential novelty is that the generators can, apart from a certain time-frequency localization, be chosen freely and each function may have its own generator. Together with the utilization of so-called parametrizations to allow generic indexing, the 'variability' of the generators provides the flexibility to cast rotation- and shear-based systems as instances of one unifying construction principle. Moreover, it becomes possible to relax the vanishing moment conditions – important for high approximation rates – imposed on the generators. Rather to demand rigid conditions as in most classical constructions, it suffices to require the moments of the variable generators to vanish asymptotically at high scales, without changing the asymptotic approximation behavior.

1.2.2 α-Molecules

The scope of parabolic molecules is limited to parabolically scaled systems, wherefore a major generalization was pursued in [28], namely the extension to *α-molecules*. These incorporate more general α-scaling (3) and can thus bridge the gap between wavelets and ridgelets, as well as curvelets and shearlets in between.

   However, like the framework of parabolic molecules, they are confined to a 2-dimensional setting. Since nowadays higher dimensional data plays an ever increasing role, an extension of the theory to higher dimensions is appreciable. A first step in this direction was taken by one of the authors [20] with an extension of the parabolic molecules framework to 3D. In this paper, we aim to generalize the framework to arbitrary dimensions $d \in \mathbb{N}$, $d \geq 2$ and general scaling parameters $\alpha \in [0,1]$.

1.2.3 Why α-Molecules?

The concept of (multivariate) α-molecules covers a great variety of directional multiscale systems and unifies their treatment and analysis, e.g. with respect to approximation properties. The foundational result behind this is Theorem 2.5, i.e. the fact that the localization of the cross-Gramian of two systems of α-molecules – in the sense of a strong off-diagonal decay – merely depends on their respective parametrizations and orders. Hence, the parametrization and the order of a system of α-molecules alone is sufficient information to determine the corresponding approximation behavior. This is illustrated by Theorem 4.4, where a large class of directional representation systems is specified with almost optimal

approximation performance with respect to cartoon video data. Finally, we remark that apart from the analysis aspects the framework also promises new design approaches for novel constructions.

1.3 Our Contributions

As mentioned before, the goal of this paper is to generalize the concept of α-molecules to arbitrary dimensions. The multivariate formulation extends the earlier results from [27, 28, 20] and gives valuable insights, e.g. on how they scale with the dimension. One should emphasize that the extension beyond dimension 2 comes with several delicacies such as to determine a suitable definition of the so-called α-scale index distance. The technical effort to prove the mentioned results is considerably higher when dealing with more than 2 dimensions, mainly because many arguments take place on the unit sphere instead of the unit circle. It also gets significantly harder to prove that shearlets, i.e. the main example of multidimensional directional systems in dimensions higher than 2, can be included in the framework. Hence already the core results of the theoretical framework themselves do not generalize straightforwardly.

1.4 Outline

The paper is organized as follows. The core part of the theory, in particular Theorem 2.5, is presented in Section 2. Since the corresponding proof is quite involved it is outsourced to Section 6. The abstract theory is further developed in Section 3 with a focus on approximation theory. Here Theorem 3.7 derives sufficient conditions for two systems of α-molecules to be sparsity equivalent. In Section 4 we then exemplarily apply the theory to video data and identify a large class of representation systems providing almost optimal sparse approximation in Theorem 4.4. Finally, Section 5 is devoted to a large class of concrete systems of α-molecules, namely multivariate α-shearlet molecules. As specific examples we present pyramid-adapted shearlet systems, in particular those generated by compactly supported functions and the smooth Parseval frame of band-limited shearlets by Guo and Labate. This frame is presented in greater detail in Subsection 5.3.1, thereby fixing some inaccuracies of the original definition.

1.5 Notation

The (strictly) positive real numbers are denoted by $\mathbb{R}_+$. The vector space $\mathbb{R}^d$, where $d \in \mathbb{N}$, is equipped with the usual Euclidean scalar product denoted by $\langle \cdot, \cdot \rangle$. $\mathbb{N}$ hereby denotes the set of positive integers, while $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For the unit sphere in $\mathbb{R}^d$ the symbol $\mathbb{S}^{d-1}$ is used. The standard unit vectors are given by $e_1, \ldots, e_d$ and for a vector $x \in \mathbb{R}^d$ we use the notation $[x]_i := \langle x, e_i \rangle$, $i \in \{1, \ldots, d\}$, for the i:th component. Its p-(quasi-)norm in the range $0 < p \leq \infty$ is denoted by $|x|_p$. In case of the Euclidean norm $|x|_2 = \sqrt{\langle x, x \rangle}$, we will usually omit the subindex. We further define $|x|_{[d-1]} := |([x]_1, \ldots, [x]_{d-1}, 0)^T|_2$. For a matrix $M \in \mathbb{R}^{m,n}$, we denote its operator norm as a mapping from the Euclidean $\mathbb{R}^n$ to the Euclidean $\mathbb{R}^m$ by $\|M\|_{2 \to 2}$, and its entries by $[M]_i^j$, $i = 1, \ldots, m$, $j = 1, \ldots n$.

The usual Lebesgue spaces on $\mathbb{R}^d$ are denoted by $L^p(\mathbb{R}^d)$, where $0 < p \leq \infty$. The corresponding sequence spaces are given by $\ell^p(\Lambda)$, where Λ is a countable index set. In both cases we use the symbol $\| \cdot \|_p$ for the associated (quasi-)norms. Further, the symbol $\langle \cdot, \cdot \rangle$ will also be used for the inner products on the Hilbert spaces $L^2(\mathbb{R}^d)$ and $\ell^2(\Lambda)$. For the weak versions of the sequence spaces we use the notation $\omega \ell^p(\Lambda)$ with associated (quasi-)norms $\| \cdot \|_{\omega \ell^p}$. For their definition we refer to Subsection 3.1.

In addition, we need the following function spaces on $\mathbb{R}^d$: the space of continuous functions $C(\mathbb{R}^d)$, the space of n-times continuously differentiable functions $C^n(\mathbb{R}^d)$ for $n \in \mathbb{N} \cup \{\infty\}$, as well as their respective restrictions $C_c(\mathbb{R}^d)$ and $C_c^n(\mathbb{R}^d)$ to functions with compact support.

The Fourier transform $\hat{f}$ of a function f in the space of Schwarz functions $\mathscr{S}(\mathbb{R}^d)$ is given by

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) \exp(-2\pi i \langle \xi, x \rangle) dx.$$

As usual, it extends to the space of tempered distributions $\mathscr{S}'(\mathbb{R}^d)$.

For two entities x, y, usually dependent on a certain set of parameters, the notation '$x \lesssim y$' shall mean that $x \leq Cy$ for some fixed constant $C > 0$, which is independent of the involved parameters. If both $x \lesssim y$ and $y \lesssim x$ we write '$x \asymp y$'. We further need the ceiling function on $\mathbb{R}$ given by

$\lceil x \rceil := \min\{\ell \in \mathbb{Z} : \ell \geq x\}$. A useful abbreviation is also the ubiquitous 'analyst's bracket' defined by $\langle x \rangle := \sqrt{1 + x^2}$ for $x \in \mathbb{R}$.

2 α-Molecules in d Dimensions

Recalling the definition in [28], a system of bivariate α-molecules consists of functions in $L^2(\mathbb{R}^2)$ obtained by applying α-scaling, rotations, and translations to a set of generating functions, which need to be sufficiently localized in time and frequency. Due to this construction, every α-molecule is naturally associated with a certain scale, orientation and spatial position, which – in the 2-dimensional case – is conveniently represented by a point in the corresponding parameter space $\mathbb{P}_2 = \mathbb{R}_+ \times \mathbb{S}^1 \times \mathbb{R}^2$.

Aiming for a multivariate generalization, we thus first need a d-dimensional version of this parameter space. We let $\mathbb{S}^{d-1}$ denote the unit sphere in $\mathbb{R}^d$ and put

$$\mathbb{P}_d = \mathbb{R}_+ \times \mathbb{S}^{d-1} \times \mathbb{R}^d.$$

Each function $m_\lambda \in L^2(\mathbb{R}^d)$ of a system of d-dimensional α-molecules $(m_\lambda)_{\lambda \in \Lambda}$ shall by definition then be associated with a unique point $(s_\lambda, e_\lambda, x_\lambda) \in \mathbb{P}_d$, where the variable $s_\lambda \in \mathbb{R}_+$ shall represent its scale, the vector $e_\lambda \in \mathbb{S}^{d-1}$ its orientation in $\mathbb{R}^d$, and $x_\lambda \in \mathbb{R}^d$ the spatial location. The relation between the index λ of a molecule m_λ and its position $(s_\lambda, e_\lambda, x_\lambda)$ in $\mathbb{P}_d$ is described by a so-called parametrization, analogue to [28].

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

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

which associates with each $\lambda \in \Lambda$ *a scale* $s_\lambda \in \mathbb{R}_+$, *a direction* $e_\lambda \in \mathbb{S}^{d-1}$, *and a location* $x_\lambda \in \mathbb{R}^d$.

For practical purposes it is more convenient to represent an orientation $\eta \in \mathbb{S}^{d-1}$ by a set of angles. Therefore we define the rotation matrix R_θ for $\theta = (\theta_1, \ldots, \theta_{d-2}) \in \mathbb{R}^{d-2}$ by

$$R_\theta = \begin{pmatrix} \cos(\theta_1) & & -\sin(\theta_1) \\ & I_{d-2} & \\ \sin(\theta_1) & & \cos(\theta_1) \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} \cos(\theta_{d-2}) & & -\sin(\theta_{d-2}) & \\ & 1 & & \\ \sin(\theta_{d-2}) & & \cos(\theta_{d-2}) & \\ & & & I_{d-3} \end{pmatrix},$$

where I_d for $d \in \mathbb{N}$ denotes the d-dimensional identity matrix. Furthermore, we introduce for $\varphi \in \mathbb{R}$ the matrix

$$R_\varphi = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) & \\ -\sin(\varphi) & \cos(\varphi) & \\ & & I_{d-2} \end{pmatrix}.$$

Note that these definitions pose an inconsistency in the notation, since they depend on the particular naming of the index. However, since we always use these particular indices, this will not lead to any problems while improving the readability significantly.

Each orientation $\eta \in \mathbb{S}^{d-1}$ can now be uniquely represented by a set of angles $(\theta_1, \ldots, \theta_{d-2}, \varphi) \in [0, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]^{d-3} \times [0, 2\pi]$ via the relation

$$\eta = R_\varphi^T R_\theta^T e_d, \tag{4}$$

where e_d is the dth unit vector of $\mathbb{R}^d$. Explicitly, it is given by

$$\eta(\theta, \varphi) = \begin{pmatrix} \eta_1(\theta, \varphi) \\ \vdots \\ \eta_d(\theta, \varphi) \end{pmatrix} = \begin{pmatrix} \cos(\varphi)\cos(\theta_{d-2}) \cdot \cdots \cdot \cos(\theta_2)\sin(\theta_1) \\ \sin(\varphi)\cos(\theta_{d-2}) \cdot \cdots \cdot \cos(\theta_2)\sin(\theta_1) \\ -\sin(\theta_{d-2})\cos(\theta_{d-3}) \cdots \cos(\theta_2)\sin(\theta_1) \\ \vdots \\ -\sin(\theta_3)\cos(\theta_2)\sin(\theta_1) \\ -\sin(\theta_2)\sin(\theta_1) \\ \cos(\theta_1) \end{pmatrix}.$$

Next, we adapt the α-scaling matrix (3) to the multivariate setting. For $\alpha \in [0,1]$, we set

$$A_{\alpha,s} = \begin{pmatrix} s^\alpha I_{d-1} & \\ & s \end{pmatrix}. \tag{5}$$

In case $\alpha = 1$ this matrix scales $\mathbb{R}^d$ isotropically, in the range $\alpha \in [0,1)$ it scales uniformly in all directions except for the e_d-direction. Note that here we choose e_d as the distinguished direction in which the scaling is stronger – in contrast to the 2-dimensional case [28], where e_1 was chosen.

After this preparation we are ready to give the definition of d-variate α-molecules, $d \in \mathbb{N}, d \geq 2$, which essentially reduces to the original definition from [28] for $d = 2$, except for the interchanged roles of the directions e_1 and e_d.

Definition 2.2. *Let $\alpha \in [0,1]$, $d \in \mathbb{N}$, $d \geq 2$ and $L, M, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$. Further let (Λ, Φ_Λ) be a parametrization with $\Phi_\Lambda(\lambda) = (s_\lambda, e_\lambda, x_\lambda) \in \mathbb{P}_d$ for $\lambda \in \Lambda$. The corresponding angles (4) for e_λ shall be denoted by $(\theta_\lambda, \varphi_\lambda)$. A family of functions $(m_\lambda)_{\lambda \in \Lambda} \subseteq L^2(\mathbb{R}^d)$ is called a* system of d-dimensional α-molecules *of order (L, M, N_1, N_2) with respect to the parametrization (Λ, Φ_Λ), if each m_λ is of the form*

$$m_\lambda = s_\lambda^{\frac{1+\alpha(d-1)}{2}} g_\lambda\left(A_{\alpha,s_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}(x - x_\lambda)\right) \tag{6}$$

with generators $g_\lambda \in L^2(\mathbb{R}^d)$ satisfying for every multi-index $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L$ the condition

$$|\partial^\rho \hat{g}_\lambda(\xi)| \lesssim \min\left(1, s_\lambda^{-1} + |[\xi]_d| + s_\lambda^{-(1-\alpha)} |\xi|_{[d-1]}\right)^M \langle|\xi|\rangle^{-N_1} \langle|\xi|_{[d-1]}\rangle^{-N_2}. \tag{7}$$

The implicit constant in (7) is required to be uniform in Λ. In case that a control parameter takes the value ∞, this shall mean that the condition (7) is fulfilled with the respective quantity arbitrarily large.

A system of α-molecules is thus obtained by applying rotations, translations, and α-scaling to a set of generating functions $(g_\lambda)_\lambda$, which are required to obey a prescribed time-frequency localization. Every molecule m_λ is thereby allowed to have its own individual generator g_λ.

The definition only poses conditions on the Fourier transform of the generators g_λ. The number L describes the spatial localization, M the number of directional (almost) vanishing moments, and N_1, N_2 the smoothness of an element m_λ. Also note that the weighting function on the right hand side of (7) is symmetric with respect to rotations around the e_d-axis, as well as reflections along this axis.

Applying $A_{\alpha,s}$ with $\alpha < 1$ and $s > 1$ to the unit ball $B = \{x \in \mathbb{R}^d : |x| \leq 1\}$ stretches B in the e_d-direction. This results in plate-like support of the characteristic function $\chi_B(A_{\alpha,s}\cdot)$ for large $s \in \mathbb{R}_+$, with the 'plate' lying in the plane spanned by the vectors $\{e_1, \ldots, e_{d-1}\}$. Thus, at high scales α-molecules can be thought of as plate-like objects in the spatial domain. The approximate frequency support on the other hand is concentrated in a pair of opposite cones in the direction of the respective orientation.

Remark 2.3. *It may seem more natural to choose a rotation R_η from e_d to $\eta \in \mathbb{S}^{d-1}$ in the (e_d, η)-plane to adjust the orientation in (6). Due to the symmetries of the weighting function of the generators, this choice is however not necessary. Since it is easier to use fixed rotation planes, we stick to this more pragmatic choice of rotation parameters.*

Let us conclude this paragraph with some comments on the use of the term 'molecule'. In the theory of function spaces, 'atoms' originally refer to the basic building blocks of a function space. In the widest sense, an atomic decomposition of a function space is a countable subset containing functions called 'atoms', which allow to represent every function of the space as a countable linear combination.

In the theory of atomic decomposition of so called *Hardy spaces* H^p, an atom is defined as a function a supported in some cube Q possessing vanishing moments and satisfying some norm bound, e.g. $\|a\|_2 \leq |Q|^{\frac{1}{2} - \frac{1}{p}}$ [48]. This definition has also been adapted in a slightly less rigid form in other parts of mathematics. Here, the term "atom" often simply refers to a function having compact support and possessing many vanishing moments. Typically, these atoms also possess additional properties, e.g. they may be bounded or fulfill smoothness conditions.

Furthermore, in many instances, the atoms are also related to each other. Classical coorbit spaces e.g. possess atomic decompositions with atoms obtained from the action of a group on a single generator [18]. A primary example are the (homogeneous) Besov-Triebel-Lizorkin spaces with (compactly supported) wavelets as atoms.

A system of α-molecules somewhat resembles this structure, whereas in a relaxed form. The compact support condition is replaced with a less restrictive decay condition, and the moments are only asymptotically vanishing. In addition, the molecules are related to each other via certain group transformations, namely translations, rotations, and dilations. However, this relation shall not be understood in a strict sense, since the generators are allowed to vary to some extent. They just need to fulfill a uniform localization condition. For the functions of a system featuring these kind of soft conditions the term 'molecules' was coined and used e.g. in [4, 32, 27, 28].

2.1   Index Distance in d Dimensions

A central ingredient of the theory of bivariate α-molecules [28] is the fact that the parameter space $\mathbb{P}_2$ can be equipped with a natural (pseudo-)metric with the property, that the distance between two points in $\mathbb{P}_2$ 'anti-correlates' with the size of the scalar products of α-molecules associated with those points: The greater the distance between two indices $\lambda, \lambda' \in \mathbb{P}_2$, the smaller the scalar product of the corresponding α-molecules $m_\lambda, m_{\lambda'} \in L^2(\mathbb{R}^d)$.

Our next aim is to find a suitable analogon of this (pseudo-)metric for the parameter space $\mathbb{P}_d$. As for $\mathbb{P}_2$, the distance between two points $(s_\lambda, e_\lambda, x_\lambda)$, $(s_\mu, e_\mu, x_\mu) \in \mathbb{P}_d$ must certainly take into account their spatial, scale, and orientational relation. The spatial distance is measured by a rightly balanced combination of an isotropic term $|x_\lambda - x_\mu|^2$ and a non-isotropic component $|\langle e_\lambda, x_\lambda - x_\mu \rangle|$, which depends on the orientation of the molecules. For the distance between the orientations e_λ and e_μ, it seems natural to consider the angle $d_\mathbb{S}(e_\lambda, e_\mu) = \arccos(\langle e_\lambda, e_\mu \rangle)$ with $d_\mathbb{S}(e_\lambda, e_\mu) \in [0, \pi]$. Due to the symmetries of the weighting function in (7) however, the angle $d_\mathbb{S}(e_\lambda, e_\mu)$ is projected onto the interval $[-\pi/2, \pi/2)$, with the projected angle $\{d_\mathbb{S}(e_\lambda, e_\mu)\}$ being the unique element of the set $\{d_\mathbb{S}(e_\lambda, e_\mu) + n\pi \mid n \in \mathbb{Z}\}$ in the interval $[-\pi/2, \pi/2)$. A suitable measure for the orientational distance is then $|\{d_\mathbb{S}(e_\lambda, e_\mu)\}|^2$. This definition is in fact consistent with the one in [28], since in two dimensions we have $|\{d_\mathbb{S}(e_\lambda, e_\mu)\}| = |\{\varphi_\lambda - \varphi_\mu\}|$. Finally, the distance between different scales s_λ, $s_\mu > 0$ is measured by the ratio $\max\{s_\lambda/s_\mu, s_\mu/s_\lambda\}$.

Altogether, this leads to the following definition. It directly generalizes the metric introduced in [29], which is a simplified version of the original metric from [28].

Definition 2.4. *Let* $\alpha \in [0, 1]$. *For given parametrizations* (Λ, Φ_Λ) *and* (Δ, Φ_Δ) *with* $(s_\lambda, e_\lambda, x_\lambda) = \Phi_\Lambda(\lambda)$ *and* $(s_\mu, e_\mu, x_\mu) = \Phi_\Delta(\mu)$, *the* α-*scaled index distance* $\omega_\alpha : \Lambda \times \Delta \to [1, \infty)$ *is defined by*

$$\omega_\alpha(\lambda, \mu) = \max\left\{\frac{s_\lambda}{s_\mu}, \frac{s_\mu}{s_\lambda}\right\}(1 + d_\alpha(\lambda, \mu)), \qquad \lambda \in \Lambda, \ \mu \in \Delta,$$

where with $s_0 = \min\{s_\lambda, s_\mu\}$

$$d_\alpha(\lambda, \mu) = s_0^{2\alpha} |x_\lambda - x_\mu|^2 + s_0^{2(1-\alpha)} |\{d_\mathbb{S}(e_\lambda, e_\mu)\}|^2 + s_0 |\langle e_\lambda, x_\lambda - x_\mu \rangle|.$$

Let (Λ, Φ_Λ) be a parametrization. Then the induced index distance on Λ is pseudosymmetric and satisfies a pseudo triangle inequality. More precisely, it has the following properties:

(i) $\omega_\alpha(\lambda, \lambda) = 1$ for all $\lambda \in \Lambda$,

(ii) $\omega_\alpha(\lambda, \lambda') \asymp \omega_\alpha(\lambda', \lambda)$ for all $\lambda, \lambda' \in \Lambda$,

(iii) $\omega_\alpha(\lambda, \lambda') \lesssim \omega_\alpha(\lambda, \lambda'')\omega_\alpha(\lambda'', \lambda')$ for all $\lambda, \lambda', \lambda'' \in \Lambda$.

Hence the function ω_α can be viewed as a kind of multiplicative pseudo-metric. A proof of these properties for the 2-dimensional case can be found in [29], which translates very well to higher dimensions.

Now we are in a position to formulate the main theorem of this paper. It states that the index distance – in a certain sense – measures the size of the scalar products of α-molecules.

Theorem 2.5. *Let $\alpha \in [0,1]$, $d \in \mathbb{N}$, $d \geq 2$, and let $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ be two systems of d-dimensional α-molecules of order (L, M, N_1, N_2) with respect to parametrizations (Λ, Φ_Λ) and (Δ, Φ_Δ), respectively. Further assume that there exists some constant $c > 0$ such that*

$$s_\lambda \geq c \quad and \quad s_\mu \geq c \quad for\ all\ \lambda \in \Lambda, \mu \in \Delta\ with\ (s_\lambda, e_\lambda, x_\lambda) = \Phi_\Lambda(\lambda), (s_\mu, e_\mu, x_\mu) = \Phi_\Delta(\mu).$$

If $N_1 > \frac{d}{2}$ and if there exists some positive integer $N \in \mathbb{N}$ such that

$$L \geq 2N, \quad M > 3N - d + \frac{1 + \alpha(d-1)}{2}, \quad N_1 \geq N + \frac{1 + \alpha(d-1)}{2}, \quad N_2 \geq 2N + d - 2,$$

then we have

$$|\langle m_\lambda, p_\mu \rangle| \lesssim \omega_\alpha(\lambda, \mu)^{-N}.$$

The proof of Theorem 2.5 is very long and technical and for this reason not presented here but in Section 6. Let us instead discuss the significance of this result. It states that – with appropriate assumptions on the parametrizations – the cross-Gramian of two systems of α-molecules is well-localized, in the sense of a fast off-diagonal decay with respect to the index distance ω_α. Put differently, the matrix is close to a diagonal matrix and the corresponding systems are almost orthogonal to each other. This property has many implications, see for instance [29, 24]. Its significance with respect to sparsity equivalence is elaborated in the next section.

3 Sparse Approximation with α-Molecules

Based on Theorem 2.5 it is possible to develop a general methodology to categorize frames of α-molecules according to their sparse approximation behavior. A central concept in this context is the notion of sparsity equivalence.

Another question that naturally arises in this context is if it is possible develop a theory of smoothness spaces associated with frames of α-molecules. For such an investigation, coorbit space theory provides an appropriate abstract framework. Usually however, due to the lack of group structure, this question can not be handled within the classical theory developed in [18, 17, 19]. In subsequent contributions, among others [10, 11, 8], the classical theory has seen significant extensions beyond the group setting. In fact, it is possible to base the theory solely on the notion of an abstract continuous frame, see [21, 47, 46, 36]. In this general setup coorbits associated to frames of α-molecules can be defined and investigated.

Up to now, this has only been carried out to some extent and not yet systematically, i.e., for certain special instances of α-molecules. A particular example are cone-adapted bivariate shearlet systems [44]. The authors believe that the development of a theory of α-molecule smoothness spaces is beyond the scope of this paper, but certainly a very interesting possibility for future research. In this paper, we concentrate on sparse approximations.

3.1 Sparse Approximation and Sparsity Equivalence

Let us briefly recall some aspects of approximation theory in a general (separable) Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$. Utilizing a system $(m_\lambda)_{\lambda \in \Lambda} \subseteq \mathcal{H}$, a signal $f \in \mathcal{H}$ can be represented by the coefficients $c_\lambda \in \mathbb{C}$ of the expansion

$$f = \sum_{\lambda \in \Lambda} c_\lambda m_\lambda. \tag{8}$$

Suitable representation systems are provided e.g. by so-called frame systems [6], which ensure stable measurement of the coefficients and also stable reconstruction. A system $(m_\lambda)_{\lambda \in \Lambda}$ in $\mathcal{H}$ forms a *frame* if there exist constants $A, B > 0$, called the *frame bounds*, such that

$$A\|f\|^2 \leq \sum_{\lambda \in \Lambda} |\langle f, m_\lambda \rangle|^2 \leq B\|f\|^2 \text{ for all } f \in \mathcal{H}.$$

If A and B can be chosen equal, the frame is called *tight*. In case $A = B = 1$, one speaks of a *Parseval frame*. The associated frame operator $S : \mathcal{H} \to \mathcal{H}$ is given by $Sf = \sum_{\lambda \in \Lambda} \langle f, m_\lambda \rangle m_\lambda$.

Since S is always invertible, the system $(S^{-1} m_\lambda)_\lambda$ is also a frame, referred to as the *canonical dual frame*. It can be used to compute a particular sequence of coefficients in the expansion (8) via

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

This sequence however is usually not the only one possible. Unlike the expansion in a basis, a representation with respect to a frame need not be unique. The canonical dual frame can also be used to express f in terms of the *frame coefficients* $(\langle f, m_\lambda \rangle)_\lambda$ by

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

In general, any system $(\tilde{m}_\lambda)_{\lambda \in \Lambda}$ satisfying this reconstruction formula is called an associated *dual frame*.

Let us turn to the question of efficient encoding. In practice we have to restrict to finite expansions (8), which usually leads to an approximation error. Given a positive integer N, the best N-term approximation f_N of some element $f \in \mathcal{H}$ with respect to the system $(m_\lambda)_\lambda$ is defined by

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

For efficient approximation, it is desirable to find representation systems which provide good sparse approximation for the considered data, in the sense that the approximation error $\|f - f_N\|$ decays quickly for $N \to \infty$. Typically one wants to approximate signals in some subclass $\mathcal{C} \subseteq \mathcal{H}$. The approximation performance of a system with respect to such a class is then usually judged by the worst-case scenario, i.e. the worst possible decay rate of the error $\|f - f_N\|$ for $f \in \mathcal{C}$. In this sense a system $(m_\lambda)_\lambda$ provides *optimally sparse approximations* with respect to $\mathcal{C}$, if its worst-case approximation rates are the best among all systems.

It is common to consider not the best N-term approximation but the N-term approximation, obtained by keeping the N largest coefficients. This approximation is better understood and provides a bound for the best N-term approximation error. We will also denote it by f_N, since the context will always make the meaning clear.

The N-term approximation rate achieved by a frame is closely related to the decay of the corresponding frame coefficients, often measured by a strong or weak ℓ^p-(quasi-)norm with $p > 0$. The weak ℓ^p-(quasi-)norm is defined by

$$\|(c_\lambda)_\lambda\|_{\omega \ell^p} := \left(\sup_{\varepsilon > 0} \varepsilon^p \cdot \#\{\lambda : |c_\lambda| > \varepsilon\} \right)^{1/p},$$

and by definition a sequence $(c_\lambda)_\lambda \in \omega \ell^p$ if $\|(c_\lambda)_\lambda\|_{\omega \ell^p} < \infty$. Since $\|(c_\lambda)_\lambda\|_{\omega \ell^p} \leq \|(c_\lambda)_\lambda\|_p$ for every sequence $(c_\lambda)_\lambda$, we have the embedding $\ell^p \hookrightarrow \omega \ell^p$. Note also that every non-increasing rearrangement $(c_n^*)_{n \in \mathbb{N}}$ of a sequence $(c_\lambda)_\lambda \in \omega \ell^p$ satisfies

$$\sup_{n > 0} n^{1/p} |c_n^*| = \|(c_\lambda)_\lambda\|_{\omega \ell^p}.$$

The well-known result below (see [14]), whose proof can be found e.g. in [28], shows that membership of the expansion coefficients in an ℓ^p space for small $p > 0$ implies good N-term approximation rates.

Lemma 3.1 ([28, Lemma 5.1]). *Let $(m_\lambda)_{\lambda \in \Lambda}$ be a frame in $\mathcal{H}$ and $f = \sum c_\lambda m_\lambda$ an expansion of $f \in \mathcal{H}$ with respect to this frame. If $(c_\lambda)_\lambda \in \omega \ell^{2/(p+1)}(\Lambda)$ for some $p > 0$, then the N-term approximation rate for f achieved by keeping the N largest coefficients is at least of order $N^{-p/2}$, i.e.*

$$\|f - f_N\|_2^2 \lesssim N^{-p}.$$

In particular, the error of best N-term approximation decays at least with order $N^{-p/2}$.

As illustrated by Lemma 3.1 the decay rate of the frame coefficients determines the N-term approximation rate. In particular, if the sequence $(\langle f, m_\lambda \rangle)_{\lambda \in \Lambda}$ of frame coefficients lies in ℓ^p for $p < 2$, the best approximation rate of the dual frame $(\tilde{m}_\lambda)_{\lambda \in \Lambda}$ is at least of order $N^{-(1/p-1/2)}$.

Let us now assume that we have two frames $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ in the Hilbert space $\mathcal{H}$ and expansion coefficients for $f \in \mathcal{H}$ with respect to these two systems. Then these frames provide the same N-term approximation rate for f, if the corresponding expansion coefficients have similar decay, e.g. if they belong to the same ℓ^p-space. We recall [28, Proposition 5.2] and formulate this result in an abstract Hilbert space setting.

Proposition 3.2 ([28, Proposition 5.2]). *Let $p \in (0,2)$, and let $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ be frames in a Hilbert space $\mathcal{H}$ and $(\tilde{m}_\lambda)_{\lambda \in \Lambda}$ a dual frame for $(m_\lambda)_{\lambda \in \Lambda}$ such that*

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

Then $(\langle f, \tilde{m}_\lambda \rangle)_\lambda \in \ell^p(\Lambda)$ implies $(\langle f, p_\mu \rangle)_\mu \in \ell^p(\Delta)$. In particular, $f \in \mathcal{H}$ can be encoded by the N largest frame coefficients from $(\langle f, p_\mu \rangle)_\mu$ up to accuracy $\lesssim N^{-(1/p-1/2)}$.

This proposition motivates the following notion of sparsity equivalence initially introduced in [27] for parabolic molecules.

Definition 3.3 ([27, Definition 4.2]). *Let $p \in (0, \infty]$, and let $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ be frames in a Hilbert space $\mathcal{H}$. Then $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ are sparsity equivalent in ℓ^p, if*

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

Sparsity equivalence, as pointed out in [28], is not an equivalence relation. Nevertheless, it allows to transfer approximation properties from one anchor system to other systems.

3.2 Consistency of Parametrizations

Our aim in this section is to categorize frames of α-molecules in $L^2(\mathbb{R}^d)$ with respect to their approximation behavior, building upon the notion of sparsity equivalence. We emphasize that a system of α-molecules does not per se constitute a frame. In fact, the question if a system of functions in $L^2(\mathbb{R}^d)$ is a frame is decoupled from the question if it forms a system of α-molecules.

Theorem 3.7 will provide sufficient conditions for two frames of α-molecules to be sparsity equivalent, based upon the notion of (α, k)-consistency originally introduced in [28]. To motivate this concept, we recall a simple estimate from [27] for the operator norm of a matrix on discrete ℓ^p spaces.

Lemma 3.4 ([27, Lemma 4.4]). *Let Λ, Δ be two discrete index sets, and let $\mathbf{A} : \ell^p(\Lambda) \to \ell^p(\Delta)$, $p > 0$ be a linear mapping defined by its matrix representation $\mathbf{A} = (A_{\lambda,\mu})_{\lambda \in \Lambda, \, \mu \in \Delta}$. Then we have the bound*

$$\|\mathbf{A}\|_{\ell^p(\Lambda) \to \ell^p(\Delta)} \leq \max \left\{ \sup_\lambda \sum_\mu |A_{\lambda,\mu}|^{\min\{1,p\}}, \sup_\mu \sum_\lambda |A_{\lambda,\mu}|^{\min\{1,p\}} \right\}^{1/\min\{1,p\}}.$$

We apply this lemma to the cross-Gramian $\mathbf{A}$ of two systems of α-molecules, and aim for sufficient conditions for the right hand side to be finite. The following notion was introduced in [28].

Definition 3.5 ([28, Definition 5.5]). *Let $\alpha \in [0,1]$ and $k > 0$. Two parametrizations (Λ, Φ_Λ) and (Δ, Φ_Δ) are called (α, k)-consistent, if*

$$\sup_{\lambda \in \Lambda} \sum_{\mu \in \Delta} \omega_\alpha (\lambda, \mu)^{-k} < \infty \quad and \quad \sup_{\mu \in \Delta} \sum_{\lambda \in \Lambda} \omega_\alpha (\lambda, \mu)^{-k} < \infty. \tag{9}$$

Remark 3.6. *Note that due to symmetry of the distance function ω_α it suffices to check only one of the two conditions in equation (9) to prove (α, k)-consistency.*

In view of Theorem 2.5, the consistency of the parametrizations of two systems of α-molecules provides a convenient sufficient condition for their sparsity equivalence.

Theorem 3.7. *Let $\alpha \in [0,1]$, $d \in \mathbb{N}$, $d \geq 2$, $k > 0$, and $p \in (0, \infty]$. Let $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ be two frames of d-dimensional α-molecules of order (L, M, N_1, N_2) with (α, k)-consistent parametrizations (Λ, Φ_Λ) and (Δ, Φ_Δ) satisfying*

$$s_\lambda \geq c, s_\mu \geq c \text{ for all } \lambda \in \Lambda, \mu \in \Delta$$

and with $q := \min\{1, p\}$

$$L \geq 2\frac{k}{q}, \quad M > 3\frac{k}{q} - d + \frac{1 + \alpha(d-1)}{2}, \quad N_1 > \frac{d}{2}, \; N_1 \geq \frac{k}{q} + \frac{1 + \alpha(d-1)}{2}, \quad N_2 \geq 2\frac{k}{q} + d - 2.$$

Then $(m_\lambda)_{\lambda \in \Lambda}$ and $(p_\mu)_{\mu \in \Delta}$ are sparsity equivalent in ℓ^p.

Proof. Let $q := \min\{1, p\}$. By Lemma 3.4, it suffices to prove that

$$\max\left\{\sup_{\lambda \in \Lambda} \sum_{\mu \in \Delta} |\langle m_\lambda, p_\mu\rangle|^q, \sup_{\mu \in \Delta} \sum_{\lambda \in \Lambda} |\langle m_\lambda, p_\mu\rangle|^q\right\}^{1/q} < \infty.$$

Since, by Theorem 2.5, we have $|\langle m_\lambda, p_\mu\rangle| \lesssim \omega_\alpha(\lambda, \mu)^{-k/q}$, we can conclude that

$$\max\left\{\sup_{\lambda \in \Lambda} \sum_{\mu \in \Delta} |\langle m_\lambda, p_\mu\rangle|^q, \sup_{\mu \in \Delta} \sum_{\lambda \in \Lambda} |\langle m_\lambda, p_\mu\rangle|^q\right\} \lesssim \max\left\{\sup_{\lambda \in \Lambda} \sum_{\mu \in \Delta} \omega_\alpha(\lambda, \mu)^{-k}, \sup_{\mu \in \Delta} \sum_{\lambda \in \Lambda} \omega_\alpha(\lambda, \mu)^{-k}\right\}$$

with the expression on the right hand side being finite due to the (α, k)-consistency of the parametrizations (Λ, Φ_Λ) and (Δ, Φ_Δ). The proof is completed. $\qquad \square$

This theorem allows to categorize frames of α-molecules according to their sparse approximation behavior. The general strategy is as follows. If an approximation result for a specific system of α-molecules is known and a class of α-molecules satisfies the hypotheses of Theorem 3.7, i.e. they are all sparsity equivalent to this specific system, they automatically inherit its known approximation behavior. In this way, one in the end obtains a stand-alone result for frames of α-molecules to exhibit sparse approximation, depending solely on the parametrization and the order.

4 Sparse Approximation of Video Data

In this section we demonstrate with a specific example how the machinery of α-molecules can be applied in practice. In our exemplary application, we are interested in the sparse representation of video signals, modelled by the class of cartoon-like functions $\mathcal{E}^2(\mathbb{R}^3)$ introduced below.

Following the general methodology, we first need a suitable anchor system, for which a sparse approximation result with respect to $\mathcal{E}^2(\mathbb{R}^3)$ is known. Utilizing Theorem 3.7 the framework can then transfer the approximation rate from this reference system to other systems. In this way we will be able to identify a large class of representation systems, which provide almost optimal sparse approximation for $\mathcal{E}^2(\mathbb{R}^3)$.

4.1 Cartoon-like Functions

A suitable model for image and video data is provided by the class of cartoon-like functions, first introduced by Donoho [15] and later extended e.g. in [41]. We shall use the following simplified model in dimensions $d = 2$ and $d = 3$.

Definition 4.1 ([15],[41, Definition 2.1])**.** *For fixed $\nu > 0$ and $d \in \{2, 3\}$ the class $\mathcal{E}^2(\mathbb{R}^d)$ of cartoon-like functions consists of functions $f : \mathbb{R}^d \to \mathbb{C}$ of the form*

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

where $B \subset [0,1]^d$ and $f_i \in C^2(\mathbb{R}^d)$ with supp $f_i \subset [0,1]^d$ and $\|f_i\|_{C^2} \leq 1$ for each $i = 0, 1$. For dimension $d = 2$, we assume that the boundary ∂B is a closed C^2-curve with curvature bounded by ν, and, for $d = 3$, the discontinuity ∂B shall be a closed C^2-surface with principal curvatures bounded by ν.

This model is justified by the observation that real-life images and video data typically consist of smooth regions, separated by piecewise smooth boundaries. Without loss of generality we restrict in Definition 4.1 to cartoon-like functions with smooth boundaries.

4.2 Optimal Approximation for Cartoon-like Functions

In [15, 41] the optimal approximation rates for $\mathcal{E}^2(\mathbb{R}^2)$ and $\mathcal{E}^2(\mathbb{R}^3)$, achievable with algorithms satisfying a *polynomial depth search* constraint, were derived. An algorithm for sparse approximation is thereby restricted to polynomial depth search in a given (countable) dictionary, if there exists a polynomial π such that the algorithm only chooses from the first $\pi(N)$ vectors of the dictionary when forming the N:th sparse approximation [15]. Such a constraint is very natural from a practical standpoint and has the nice side-effect that it ensures the existence of the best N-term approximation. In general, this may not be the case, as the instance of a countable dense subset of $L^2(\mathbb{R}^d)$ as a dictionary shows. Here the best 1-term approximation does not always exist.

We now cite the result of [15, 41]. In [41] it is conjectured that the result also generalizes to higher dimensions.

Theorem 4.2 ([15, Theorem 7.2],[41, Theorem 3.2]). *Let $d \in \{2, 3\}$. The best N-term approximation rate for $\mathcal{E}^2(\mathbb{R}^d)$, achieved by an arbitrary dictionary under the restriction of polynomial depth search, cannot exceed*

$$\|f - f_N\|_2^2 \lesssim N^{-\frac{2}{d-1}},$$

where f_N is the best N-term approximation of $f \in \mathcal{E}^2(\mathbb{R}^d)$.

For $d = 2$ it has been shown that this rate can indeed be achieved [15] using so-called wedgelets, which are adaptive to the data. Moreover, there are several examples of non-adaptive frames in two and three dimensions which almost provide these optimal rates [5, 31, 38, 33, 41], typically up to log-terms. In particular, it was proven by Guo and Labate in [33] that the smooth Parseval frame of band-limited 3D-shearlets SH constructed by them in [34] sparsely approximates the class $\mathcal{E}^2(\mathbb{R}^3)$ with an almost optimal approximation rate. We recall the definition of this particular system in Subsection 5.3.1 and state the approximation result below.

Theorem 4.3 ([33, Theorem 3.1]). *Let $SH = \{\psi_\lambda\}_{\lambda \in \Lambda_{SH}}$ be the smooth Parseval frame of 3D-shearlets defined in Subsection 5.3.1. Then the sequence of shearlet coefficients $\theta_\lambda(f) := \langle f, \psi_\lambda \rangle$, $\lambda \in \Lambda_{SH}$, associated with $f \in \mathcal{E}^2(\mathbb{R}^3)$ satisfies*

$$\sup_{f \in \mathcal{E}^2(\mathbb{R}^3)} |\theta_\lambda(f)|_N \lesssim N^{-1} \cdot \log(N),$$

where $|\theta_\lambda(f)|_N$ denotes the N:th largest shearlet coefficient.

Theorem 4.3 shows that the shearlet coefficients belong to $\omega \ell^p(\Lambda_{SH})$ for every $p > 1$. In view of Lemma 3.1, for every $f \in \mathcal{E}^2(\mathbb{R}^3)$, the frame SH therefore provides at least the approximation rate

$$\|f - f_N\|_2^2 \lesssim N^{-1+\varepsilon} \quad , \varepsilon > 0 \text{ arbitrary,} \tag{10}$$

where f_N denotes the N-term approximation obtained from the N largest coefficients. According to Theorem 4.2, this is almost the optimal approximation rate achievable for cartoon-like functions $\mathcal{E}^2(\mathbb{R}^3)$. For small $\varepsilon > 0$, we get arbitrarily close to the optimal rate.

The idea to use 3D-shearlets for processing video data has been tested in practice in [45]. The authors of said article develop a clever discretisation procedure, allowing a fast computation of the shearlet coefficients. They then test how their discrete shearlet transform can be used to denoise and enhance video sequences, with promising results.

4.3 Transfer of the Approximation Rate

Our final goal is to find a large class of representation systems which achieve the almost optimal rate (10). For this, we put the machinery of α-molecules to work. Via Theorem 3.7 it is possible to transfer the approximation rate (10) established for the smooth Parseval frame of 3D-shearlets SH to other systems of 3-dimensional $\frac{1}{2}$-molecules. In fact, the frame SH is a suitable choice for the reference system, since by Proposition 5.9 it constitutes a system of 3-dimensional $\frac{1}{2}$-molecules of order $(\infty, \infty, \infty, \infty)$ with respect to the parametrization $(\Lambda_{SH}, \Phi_{SH})$ The following result is then a direct application of the general theory.

Theorem 4.4. *Assume that a frame* $(m_\lambda)_{\lambda \in \Lambda}$ *of 3-dimensional* $\frac{1}{2}$*-molecules satisfies, for some* $k > 0$*, the following two conditions:*

(i) *its parametrization* (Λ, Φ_Λ) *is* $(\frac{1}{2}, k)$*-consistent with* $(\Lambda_{SH}, \Phi_{SH})$*,*

(ii) *its order* (L, M, N_1, N_2) *satisfies*

$$L \geq 2k, \quad M \geq 3k - 2, \quad N_1 \geq k + 1, \quad N_1 > 3/2, \quad N_2 \geq 2k + 1.$$

Then each dual frame $(\tilde{m}_\lambda)_{\lambda \in \Lambda}$ *possesses an almost optimal N-term approximation rate for the class of cartoon-like functions* $\mathcal{E}^2(\mathbb{R}^3)$*, i.e. for all* $f \in \mathcal{E}^2(\mathbb{R}^3)$

$$\|f - f_N\|_2^2 \lesssim N^{-1+\varepsilon}, \quad \varepsilon > 0 \text{ arbitrary,}$$

where f_N *denotes the N-term approximation obtained from the N largest frame coefficients.*

Proof. Let $SH = \{\psi_\lambda\}_{\lambda \in \Lambda_{SH}}$ be the Parseval frame of 3D-shearlets from Subsection 5.3.1, and take $f \in \mathcal{E}^2(\mathbb{R}^3)$. By Theorem 4.3 the sequence of shearlet coefficients $(\theta_\lambda)_\lambda$ given by $\theta_\lambda = \langle f, \psi_\lambda \rangle$ belongs to $\omega \ell^p(\Lambda_{SH})$ for every $p > 1$. Since $\omega \ell^p \hookrightarrow \ell^{p+\epsilon}$ for arbitrary $\varepsilon > 0$, this further implies $(\theta_\lambda)_\lambda \in \ell^p(\Lambda_{SH})$ for every $p > 1$. Let now

$$f = \sum_{\mu \in \Lambda} c_\mu \tilde{m}_\mu$$

be the canonical expansion of f with respect to the dual frame $(\tilde{m}_\mu)_\mu$, with frame coefficients $(c_\mu)_\mu$. Note that since SH is a Parseval frame, the canonical dual frame of SH is equal to SH itself. Therefore the coefficents are given by

$$c_\mu = \langle f, m_\mu \rangle = \left\langle \sum_\lambda \theta_\lambda \psi_\lambda, m_\mu \right\rangle = \sum_\lambda \langle \psi_\lambda, m_\mu \rangle \theta_\lambda.$$

Thus, they are related to the shearlet coefficients $(\theta_\lambda)_\lambda$ by the cross-Gramian $(\langle \psi_\lambda, m_\mu \rangle)_{\lambda,\mu}$. By Theorem 3.7, conditions (i) and (ii) guarantee that the frame $(m_\mu)_{\mu \in \Lambda}$ is sparsity equivalent to $(\psi_\lambda)_{\lambda \in \Lambda_{SH}}$ in ℓ^p for every $p > 1$. This implies that the cross-Gramian is a bounded operator $\ell^p(\Lambda_{SH}) \to \ell^p(\Lambda)$, which maps $(\theta_\lambda)_\lambda$ to $(c_\mu)_\mu$. Hence, $(c_\mu)_\mu \in \ell^p(\Lambda)$ for every $p > 1$. The embedding $\ell^p \hookrightarrow \omega \ell^p$ then proves $(c_\mu)_\mu \in \omega \ell^p(\Lambda)$ for every $p > 1$. Finally, for arbitrary $\varepsilon > 0$, the application of Lemma 3.1 yields

$$\|f - f_N\|_2^2 \lesssim N^{-1+\varepsilon},$$

where f_N denotes the N-term approximation with respect to the system $(\tilde{m}_\mu)_\mu$ obtained by choosing the N largest coefficients. $\square$

Theorem 4.4 specifies a large class of multiscale systems with almost optimal approximation performance for video data $\mathcal{E}^2(\mathbb{R}^3)$. According to Remark 5.11 condition (i) is in particular fulfilled by every $\frac{1}{2}$-shearlet parametrization (see Theorem 5.3) for $k > 3$. Hence, due to condition (ii) all systems of 3-dimensional $\frac{1}{2}$-shearlet molecules of order

$$L \geq 7, \quad M \geq 8, \quad N_1 \geq 5, \quad N_2 \geq 8,$$

provide almost optimal approximation for $\mathcal{E}^2(\mathbb{R}^3)$.

Taking into account Proposition 5.12, the statement of Theorem 4.4 in particular includes the following result for compactly supported shearlet frames.

Corollary 4.5. *Any dual frame of a shearlet frame of the form* (27) *generated by compactly supported functions* $\phi, \psi^1, \psi^2, \psi^3 \in L^2(\mathbb{R}^3)$, *so that* $\phi \in C^{13}(\mathbb{R}^3)$ *and so that for each* $\varepsilon \in \{1, 2, 3\}$

(i) $\partial^\gamma \psi^\varepsilon$ *exists and is continuous for every* $\gamma = (\gamma_1, \gamma_2, \gamma_3)^T \in \mathbb{N}_0^3$ *with* $|\gamma|_\infty \leq 13$ *and* $\gamma_\varepsilon \leq 5$,

(ii) ψ^ε *has at least 15 vanishing directional moments in direction* e_ε,

provides the almost optimal approximation rate (10) *for the cartoon video class* $\mathcal{E}^2(\mathbb{R}^3)$.

This corollary is a new result for compactly supported shearlets on its own. A similar result was proved in [41]. In comparison the most intriguing fact is the simplicity of its deduction: The framework of α-molecules enables a simple transport of the decay rates.

Remark 4.6. *The 2-dimensional counterpart of Theorem 4.4 is contained in [28, Theorem 5.12] for the choice* $\alpha = \frac{1}{2}$. *In [28] it is further shown that not only any system of bivariate* $\frac{1}{2}$-*shearlet molecules satisfies the conditions of this theorem, provided that the order is sufficiently high, but also every system of sufficiently high order* $\frac{1}{2}$-*curvelet molecules. This implies that also curvelet-like constructions yield almost optimal approximation for the class* $\mathcal{E}^2(\mathbb{R}^2)$, *including the classical curvelet frame.*

In contrast, a 'true' curvelet construction in 3D is not known to the authors. Note that, despite the misleading name, the construction of the '3D discrete curvelet transform' in [49] is actually shear-based. Still, it can be expected that any curvelet-like system would fall into our framework, and thus Theorem 4.4 would immediately establish the almost optimal approximation rate.

5 Shearlet Systems in d Dimensions

We introduce a very general class of shear-based systems, namely *systems of α-shearlet molecules*. The definition in d dimensions is analogue to the 2-dimensional case [28]. Roughly speaking, they are shear-based systems obtained from variable generators, where similar to α-molecules the conditions on the generators have been relaxed to a mere time-frequency localization requirement. The notion of α-shearlet molecules comprises many specific shear-based constructions and simplifies the treatment of such systems within the general framework of α-molecules.

5.1 Multidimensional α-Shearlet Molecules

As explained in Section 1, shearlet-like constructions are based on anisotropic scaling, shearings, and translations. For the change of scale, we utilize α-scaling as defined by (5). The change of orientation is provided by shearings, in d dimensions given by the shearing matrices

$$S_h = \begin{pmatrix} I_{d-1} & 0 \\ h^T & 1 \end{pmatrix} \quad \text{and} \quad S_h^T = \begin{pmatrix} I_{d-1} & h \\ 0 & 1 \end{pmatrix}, \quad h \in \mathbb{R}^{d-1},$$

which are the natural generalizations of (2). The matrix S_h^T shears parallel to the $(e_1, \ldots, e_{d-1})$-plane and the shear vector $h \in \mathbb{R}^{d-1}$ determines the direction of the shearing in this plane. The transformations associated with shearings and α-scalings naturally form a group [9].

To avoid directional bias, the frequency domain is divided into cone-like regions along the coordinate axes and a coarse-scale box for the low frequencies. Note that this comes at the cost of losing the group properties mentioned above. This division procedure is however crucial for applications, and also, as the subsequent arguments will show, for including α-shearlets in the concept of α-molecules. The cone-like regions along the e_ε-axes shall be called *pyramids* and are explicitly given by

$$\mathcal{P}_\varepsilon = \left\{ (\xi_1, \ldots, \xi_d)^T \in \mathbb{R}^d \mid \forall i \in \{1, \ldots, d\} : |\xi_i| \leq |\xi_\varepsilon| \right\},$$

where $\varepsilon \in \{1, \ldots, d\}$. $\varepsilon = 0$ shall refer to a coarse-scale box of the form $\mathcal{R} = \{\xi \in \mathbb{R}^d : |\xi|_\infty \leq C\}$, where $C > 0$ is a suitably chosen constant. In the sequel we will always stay in this so-called cone-adapted setting. For an illustration of this specific setting in 3D, we refer to Subsection 5.3.

In each cone we require different versions of the scaling and shearing operators. The cyclic permutation matrix

$$Z = \begin{pmatrix} 0 & 1 \\ I_{d-1} & 0 \end{pmatrix} \tag{11}$$

allows to elegantly define these operators associated with the respective cones by $Z^\varepsilon S_h Z^{-\varepsilon}$ and $Z^\varepsilon A_{\alpha,s} Z^{-\varepsilon}$.

Before we come to the definition of α-shearlet molecules, we need to introduce a set of characteristic parameters, associated with these systems. The resolution of the underlying sampling grid is determined by the parameters $\sigma > 1$, $\tau_1, \ldots, \tau_d > 0$, and a sequence $\Theta = (\eta_j)_{j \in \mathbb{N}_0} \subset \mathbb{R}_+$. The parameter σ specifies the fineness of the scale sampling. The parameters τ_ε, $\varepsilon \in \{1, \ldots, d\}$, determine the spatial resolution in the e_ε-direction. For convenience they are summarized in the diagonal matrix $\mathcal{T} := \mathrm{diag}(\tau_1, \ldots, \tau_d) \in \mathbb{R}^{d \times d}$. The angular resolution at each scale $j \in \mathbb{N}_0$ is given by the value η_j of the sequence Θ. Last but not least, in each cone $\varepsilon \in \{1, \ldots, d\}$ and at each scale $j \in \mathbb{N}_0$ the shearing parameter ℓ is restricted to a set $\mathscr{L}_{\varepsilon,j} \subseteq \mathbb{Z}^{d-1}$. These sets are collected in $\mathscr{L} := \{\mathscr{L}_{\varepsilon,j} : \varepsilon \in \{1, \ldots, d\}, j \in \mathbb{N}_0\}$.

After the introduction of this sampling data $\mathbb{D} := \{\sigma, \Theta, \mathscr{L}, \mathcal{T}\}$ we can now give the definition of a system of α-shearlet molecules in d dimensions, depending on $\mathbb{D}$. The scale-dependent step size η_j of the directional sampling is assumed to satisfy $\eta_j \asymp \sigma^{-j(1-\alpha)}$ for $j \in \mathbb{N}_0$. Further, we require the upper bounds $\mathbb{L}_j := \max\{|\ell|_\infty : \ell \in \mathscr{L}_{\varepsilon,j}, \varepsilon \in \{1, \ldots, d\}\}$, $j \in \mathbb{N}_0$, to fulfill the complementary condition $\mathbb{L}_j \lesssim \sigma^{j(1-\alpha)}$. We remark that the translation parameters τ_ε may also vary with the indices (ε, j, ℓ), as long as their values are restricted to some fixed interval $[\tau_{min}, \tau_{max}]$ with $0 < \tau_{min} \leq \tau_{max} < \infty$. However, this is not indicated in the notation.

Definition 5.1. *Let $\alpha \in [0,1]$, $d \in \mathbb{N}$, $d \geq 2$, and $L, M, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$. Further the sampling data $\mathbb{D}$ shall be given as above. For $\varepsilon \in \{1, \ldots, d\}$, a system of functions*

$$\Sigma_\varepsilon := \Big\{ m^\varepsilon_{j,\ell,k} \in L^2(\mathbb{R}^d) \ : \ (j,\ell,k) \in \Lambda^s_\varepsilon \Big\},$$

indexed by the set $\Lambda^s_\varepsilon := \big\{ (j,\ell,k) \ : \ j \in \mathbb{N}_0, \ \ell \in \mathscr{L}_{\varepsilon,j} \subseteq \mathbb{Z}^{d-1}, \ k \in \mathbb{Z}^d \big\}$, is called a system of d-dimensional α-shearlet molecules of order (L, M, N_1, N_2) associated with the orientation ε, if it is of the form

$$m^\varepsilon_{j,\ell,k}(x) = \sigma^{\frac{(1+\alpha(d-1))j}{2}} \gamma^\varepsilon_{j,\ell,k}\big(Z^\varepsilon A^j_{\alpha,\sigma} S_{\ell\eta_j} Z^{-\varepsilon} x - \mathcal{T}k\big) \tag{12}$$

with generating functions $\gamma^\varepsilon_{j,\ell,k} \in L^2(\mathbb{R}^d)$ satisfying for every $\rho \in \mathbb{N}^d_0$ with $|\rho|_1 \leq L$

$$\big|\partial^\rho \hat{\gamma}^\varepsilon_{j,\ell,k}(\xi)\big| \lesssim \frac{\min\{1, \sigma^{-j} + \sigma^{-(1-\alpha)j}|Z^{-\varepsilon}\xi|_{[d-1]} + |[Z^{-\varepsilon}\xi]_d|\}^M}{\langle|\xi|\rangle^{N_1} \langle|Z^{-\varepsilon}\xi|_{[d-1]}\rangle^{N_2}}. \tag{13}$$

The implicit constant is required to be uniform over Λ^s_ε. If one of the parameters L, M, N_1, N_2 takes the value ∞, this shall mean that condition (13) is fulfilled with the respective quantity arbitrarily large.

Combining systems of α-shearlet molecules of order (L, M, N_1, N_2) for each orientation $\varepsilon \in \{1, \ldots, d\}$ with a system of coarse-scale elements

$$\Sigma_0 := \Big\{ m^0_{0,\mathbf{0},k} := \gamma^0_{0,\mathbf{0},k}(\cdot - \mathcal{T}k) \ : \ k \in \mathbb{Z}^d \Big\}, \tag{14}$$

where the generators $\gamma^0_{0,\mathbf{0},k} \in L^2(\mathbb{R}^d)$ fulfill $|\partial^\rho \hat{\gamma}^0_{0,\mathbf{0},k}(\xi)| \lesssim \langle|\xi|\rangle^{-N_1} \langle|\xi|_{[d-1]}\rangle^{-N_2}$ for every $\rho \in \mathbb{N}^d_0$ with $|\rho|_1 \leq L$, yields a *system of α-shearlet molecules* of order (L, M, N_1, N_2). The associated index set is

$$\Lambda^s_0 := \Big\{ (0, \mathbf{0}, k) : k \in \mathbb{Z}^d \Big\} \subseteq \mathbb{N}_0 \times \mathbb{Z}^{d-1} \times \mathbb{Z}^d.$$

Definition 5.2. *For each $\varepsilon \in \{1, \ldots, d\}$, let Σ_ε be a system of α-shearlet molecules of order (L, M, N_1, N_2) associated with the respective orientation. Further, let Σ_0 be a system of coarse-scale scale elements defined as in (14). Then the union*

$$\Sigma := \bigcup_{\varepsilon=0}^d \Sigma_\varepsilon$$

is called a system of α-shearlet molecules of order (L, M, N_1, N_2). *The associated* shearlet index set *is given by*

$$\Lambda^s = \{(\varepsilon, j, \ell, k) : \varepsilon \in \{0, \ldots, d\}, (j, \ell, k) \in \Lambda^s_\varepsilon\}.$$

Next, we prove that the system Σ is a system of α-molecules.

Theorem 5.3. *Let $\alpha \in [0, 1]$ and Σ be a system of α-shearlet molecules of order (L, M, N_1, N_2). Then Σ constitutes a system of α-molecules of the same order. The associated α-shearlet parametrization (Λ^s, Φ^s) is given by the map $\Phi^s(\lambda) = (s_\lambda, e_\lambda, x_\lambda) \in \mathbb{P}_d$ for $\lambda = (\varepsilon, j, \ell, k) \in \Lambda^s$, where*

$$s_\lambda = \sigma^j, \quad e_\lambda = n_\lambda \cdot Z^\varepsilon \begin{pmatrix} \eta_j \ell \\ 1 \end{pmatrix}, \quad x_\lambda = Z^\varepsilon S^{-1}_{\ell\eta_j} A^{-j}_{\alpha,\sigma} Z^{-\varepsilon} \mathcal{T} k, \tag{15}$$

and $n_\lambda = (1 + \eta_j^2 |\ell|_2^2)^{-1/2}$ is a normalization constant.

In particular for $\varepsilon = 0$ we have $s_\lambda = 1$, $e_\lambda = e_d$ and $x_\lambda = \mathcal{T} k$ for every $\lambda = (0, 0, \mathbf{0}, k) \in \Lambda^s$.

Proof. Since a finite union of systems of α-molecules is itself a system of α-molecules, we can prove this theorem separately for each system Σ_ε, $\varepsilon \in \{0, \ldots, d\}$. For Σ_0 the statement is obvious. For the other systems it suffices to give the proof for $\varepsilon = d$, since they are all related by a mere permutation of indices. We subsequently drop the index ε to simplify the notation and note $Z^\varepsilon = I$ for $\varepsilon = d$.

For the proof we introduce the index set $\Lambda^{s,d} := \{(d, j, \ell, k) : (j, \ell, k) \in \Lambda^s_d\}$. Let $\lambda = (d, j, \ell, k) \in \Lambda^{s,d}$ and $m_\lambda := m^d_{j,\ell,k}$ the associated α-shearlet molecule with corresponding generating function $\gamma_\lambda := \gamma^d_{j,\ell,k}$. As usual we denote the angles representing the orientation e_λ by $(\theta_\lambda, \varphi_\lambda)$, i.e. $e_\lambda = R^T_{\varphi_\lambda} R^T_{\theta_\lambda} e_d$. The molecule m_λ can clearly be written in the form (6) with respect to the generator

$$g_\lambda(x) := \gamma_\lambda(A^j_{\alpha,\sigma} S_{\ell\eta_j} R^T_{\varphi_\lambda} R^T_{\theta_\lambda} A^{-j}_{\alpha,\sigma} x), \quad x \in \mathbb{R}^d.$$

It remains to check condition (7) for these functions. On the Fourier side we have

$$\hat{g}_\lambda(\xi) = \hat{\gamma}_\lambda(A^{-j}_{\alpha,\sigma} S^{-T}_{\ell\eta_j} R^T_{\varphi_\lambda} R^T_{\theta_\lambda} A^j_{\alpha,\sigma} \xi), \quad \xi \in \mathbb{R}^d.$$

For $\lambda = (d, j, \ell, k) \in \Lambda^{s,d}$ let us first examine the matrix

$$M_\lambda := S^{-T}_{\ell\eta_j} R^T_{\varphi_\lambda} R^T_{\theta_\lambda}. \tag{16}$$

A simple calculation shows $M_\lambda e_d = S^{-T}_{\ell\eta_j} R^T_{\varphi_\lambda} R^T_{\theta_\lambda} e_d = S^{-T}_{\ell\eta_j} e_\lambda = S^{-T}_{\ell\eta_j} n_\lambda(\eta_j\ell, 1)^T = n_\lambda e_d$. Hence, the entries of the last column of M_λ vanish except for the last one. Next, we prove the uniform boundedness of the set of operators $\{M_\lambda\}_{\lambda \in \Lambda^{s,d}}$. It holds uniformly for $\lambda \in \Lambda^{s,d}$

$$\|M_\lambda\|_{2 \to 2} = \|S^{-T}_{\ell\eta_j}\|_{2 \to 2} = \sqrt{1 + \eta_j^2 |\ell|_2^2} \lesssim \sqrt{1 + \eta_j^2 \mathbb{L}_j^2} \lesssim 1.$$

Note that this implies that each entry in M_λ is bounded in modulus. Since similar considerations hold for the inverse $M^{-1}_\lambda = R_{\theta_\lambda} R_{\varphi_\lambda} S^T_{\ell\eta_j}$, we can conclude that both $\widetilde{M}_\lambda := A^{-j}_{\alpha,\sigma} M_\lambda A^j_{\alpha,\sigma}$ and its inverse $\widetilde{M}^{-1}_\lambda$ have the form

$$\begin{pmatrix} * & \ldots & * & 0 \\ \vdots & \ddots & \vdots & \vdots \\ * & \ldots & * & 0 \\ \square & \ldots & \square & * \end{pmatrix},$$

where the entries $*$ are the same as in M_λ (or M^{-1}_λ) and the entries $\square$ are of the form $\sigma^{-j(1-\alpha)}[M_\lambda e_i]_d$ (or $\sigma^{-j(1-\alpha)}[M^{-1}_\lambda e_i]_d$) for $i \in \{1, \ldots, d-1\}$. In particular, the entries of $\widetilde{M}_\lambda$ and $\widetilde{M}^{-1}_\lambda$ are uniformly bounded in modulus. This implies $\|\widetilde{M}_\lambda\|_{2 \to 2} \lesssim 1$ and $\|\widetilde{M}^{-1}_\lambda\|_{2 \to 2} \lesssim 1$. Altogether, we obtain

$$|\widetilde{M}_\lambda \xi| \asymp |\xi| \text{ uniformly for } \xi \in \mathbb{R}^d \text{ and } \lambda \in \Lambda^{s,d}. \tag{17}$$

16

Due to the structure of the last column of $\widetilde{M}_\lambda$ we further have for $\xi = (\xi_1, \ldots, \xi_d)^T \in \mathbb{R}^d$

$$|\widetilde{M}_\lambda \xi|_{[d-1]} = |\widetilde{M}_\lambda(\xi_1, \ldots, \xi_{d-1}, 0)^T|_{[d-1]} \leq \|\widetilde{M}_\lambda\|_{2\to 2} |(\xi_1, \ldots, \xi_{d-1}, 0)^T| = \|\widetilde{M}_\lambda\|_{2\to 2} |\xi|_{[d-1]}.$$

For the inverse $\widetilde{M}^{-1}$ it holds analogously $|\widetilde{M}_\lambda^{-1} \xi|_{[d-1]} \leq \|\widetilde{M}_\lambda^{-1}\|_{2\to 2} |\xi|_{[d-1]}$. We conclude

$$|\widetilde{M}_\lambda \xi|_{[d-1]} \asymp |\xi|_{[d-1]} \text{ uniformly for } \xi \in \mathbb{R}^d \text{ and } \lambda \in \Lambda^{s,d}. \tag{18}$$

Finally, the following estimate holds uniformly for $\xi = (\xi_1, \ldots, \xi_d)^T \in \mathbb{R}^d$ and $\lambda \in \Lambda^{s,d}$,

$$|[\widetilde{M}_\lambda \xi]_d| \leq |[M_\lambda]_d^d||\xi_d| + \sum_{i=1}^{d-1} \sigma^{-j(1-\alpha)} |[M_\lambda]_d^i||\xi_i| \lesssim \sigma^{-(1-\alpha)j} |\xi|_{[d-1]} + |\xi_d|. \tag{19}$$

Finally, we can prove (7) for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L$,

$$
\begin{aligned}
|\partial^\rho \hat{g}_\lambda(\xi)| &\lesssim \sup_{|\beta|_1 \leq L} \left| (\partial^\beta \hat{\gamma}_\lambda)(\widetilde{M}_\lambda \xi) \right| \lesssim \frac{\min\{1, \sigma^{-j} + \sigma^{-(1-\alpha)j} |\widetilde{M}_\lambda \xi|_{[d-1]} + |[\widetilde{M}_\lambda \xi]_d|\}^M}{\langle |\widetilde{M}_\lambda \xi| \rangle^{N_1} \langle |\widetilde{M}_\lambda \xi|_{[d-1]} \rangle^{N_2}} \\
&\lesssim \frac{\min\{1, \sigma^{-j} + \sigma^{-(1-\alpha)j} |\xi|_{[d-1]} + |[\xi]_d|\}^M}{\langle |\xi| \rangle^{N_1} \langle |\xi|_{[d-1]} \rangle^{N_2}}.
\end{aligned}
$$

The first estimate holds true, since $\hat{g}_\lambda(\xi) = \hat{\gamma}_\lambda(\widetilde{M}_\lambda \xi)$ and the entries of $\widetilde{M}_\lambda$ are uniformly bounded in λ. The second estimate is due to (13). For the last estimate we used (17), (18), and (19). The observation $s_\lambda = \sigma^j$ finishes the proof. $\qquad\square$

5.2   Consistency of the Shearlet Parametrizations

The properties of α-molecules depend essentially on their parametrizations. In view of Theorem 3.7 the consistency is of particular interest when investigating approximation properties. In this paragraph we shall prove, in Proposition 5.6, that shearlet parametrizations are consistent with each other. This allows to establish approximation rates for various shearlet-like constructions simultaneously, as long as they fall under the umbrella of the shearlet-molecule concept.

Lemma 5.4. *Consider the* gnomonic projection $\phi : \mathbb{R}^d \backslash \{x \in \mathbb{R}^d \mid [x]_d = 0\} \to \mathbb{R}^d, x \mapsto \frac{1}{[x]_d} x$, *and let* $1 \geq c > 0$ *be fixed. For* $v, w \in \mathbb{S}^{d-1} \cap \{x \in \mathbb{R}^d : [x]_d \geq c\}$ *we then have* $|\phi(v) - \phi(w)| \asymp |v - w|$ *and* $|v - w|_{[d-1]} \asymp |v - w|$.

Proof. First note that $|v - w|_{[d-1]} = |\pi(v) - \pi(w)|$, where π is the orthogonal projection of $\mathbb{R}^d$ onto the $(e_1, \ldots, e_{d-1})$-plane. On the set $\mathbb{S}^{d-1} \cap \{x \in \mathbb{R}^d : [x]_d \geq c\}$, the mappings ϕ and π are diffeomorphisms with bounded derivatives in both directions. This implies the statement. $\qquad\square$

In order to apply the previous lemma it is useful to record the following observation.

Lemma 5.5. *Let* $1 \geq c > 0$ *be fixed and let* $w = (w_1, \ldots, w_d)^T \in \mathbb{S}^{d-1}$ *be a vector such that* $w_d < c$. *Then there exists a point* $\tilde{w} = (\tilde{w}_1, \ldots, \tilde{w}_d)^T \in \mathbb{S}^{d-1}$ *with* $\tilde{w}_d \geq c$ *such that*

$$|\tilde{w} - v| \leq |w - v| \qquad \text{for every} \quad v \in \mathbb{S}^{d-1} \cap \{x \in \mathbb{R}^d : [x]_d \geq c\}.$$

Proof. If $w_d \leq -c$ simply take $\tilde{w}$ to be the reflection of w at the $(e_1, \ldots, e_{d-1})$-plane. Then we have $\tilde{w}_d \geq c > 0$ and we can conclude for every $v = (v_1, \ldots, v_d)^T \in \mathbb{S}^{d-1}$ with $v_d \geq c > 0$

$$|v - w|^2 = |v - w|_{[d-1]}^2 + |v_d - w_d|^2 = |v - \tilde{w}|_{[d-1]}^2 + |v_d + \tilde{w}_d|^2 \geq |v - \tilde{w}|_{[d-1]}^2 + |v_d - \tilde{w}_d|^2 = |v - \tilde{w}|^2.$$

In the other case $|w_d| < c$ we argue as follows. Applying a rotation about the e_d-axis, we may assume that w is of the form $[\sqrt{1 - w_d^2}, 0, \ldots 0, w_d]^T$. The vector $\tilde{w} = [\sqrt{1 - c^2}, 0, \ldots, 0, c]^T$ then has the desired properties. To verify this it suffices to show $\langle \tilde{w}, v \rangle \geq \langle w, v \rangle$, because then

$$|\tilde{w} - v|^2 = |\tilde{w}|^2 + |v|^2 - 2\langle \tilde{w}, v \rangle = 2 - 2\langle \tilde{w}, v \rangle \leq |w|^2 + |v|^2 - 2\langle w, v \rangle = |w - v|^2.$$

In order to show $0 \leq \langle \tilde{w} - w, v \rangle = v_1(\sqrt{1-c^2} - \sqrt{1-w_d^2}) + v_d(c - w_d)$, we first observe that $\sqrt{1-c^2} < \sqrt{1-w_d^2}$ since $|w_d| < c$. Moreover, every $v \in \mathbb{S}^{d-1} \cap \{x \in \mathbb{R}^d : [x]_d \geq c\}$ satisfies $|v_1| \leq \sqrt{1-v_d^2}$. It follows

$$v_1\left(\sqrt{1-c^2} - \sqrt{1-w_d^2}\right) + v_d(c - w_d) \geq \sqrt{1-v_d^2}\left(\sqrt{1-c^2} - \sqrt{1-w_d^2}\right) + v_d(c - w_d).$$

Hence, it just remains to prove the inequality $\sqrt{1-v_d^2}\left(\sqrt{1-c^2} - \sqrt{1-w_d^2}\right) + v_d(c - w_d) \geq 0$ under the condition $0 \leq |w_d| < c \leq v_d \leq 1$. The associated angles $\theta_w, \theta_c, \theta_v \in [0, \pi]$ defined by $\cos\theta_w = w_d$, $\cos\theta_c = c$, and $\cos\theta_v = v_d$, satisfy $\pi \geq \theta_w \geq \theta_c \geq \theta_v \geq 0$, and the inequality reads as

$$\cos(\theta_c - \theta_v) = \langle(\sin\theta_v, \cos\theta_v), (\sin\theta_c, \cos\theta_c)\rangle \geq \langle(\sin\theta_v, \cos\theta_v), (\sin\theta_w, \cos\theta_w)\rangle = \cos(\theta_w - \theta_v).$$

This however is obviously true since $0 \leq \theta_c - \theta_v \leq \theta_w - \theta_v \leq \pi$. $\qquad\square$

After this preparation we are in the position to prove the consistency.

Proposition 5.6. *Let $\alpha \in [0,1]$ and assume that (Λ, Φ_Λ) and (Δ, Φ_Δ) are α-shearlet parametrizations (possibly with different parameters). Then (Λ, Φ_Λ) and (Δ, Φ_Δ) are (α, k)-consistent for every $k > d$.*

Proof. As already was noted in Remark 3.6, it suffices to prove that for $N > d$ it holds

$$\sup_{\mu \in \Delta} \sum_{\lambda \in \Lambda} \omega_\alpha(\lambda, \mu)^{-N} < \infty.$$

For this task it is convenient to decompose the shearlet index set $\Lambda = \Lambda_0 \cup \cdots \cup \Lambda_d$ into the sets Λ_ε associated with the respective pyramidal regions $\mathcal{P}_\varepsilon$ for $\varepsilon \in \{1, \ldots, d\}$ and the low-frequency box $\mathcal{R}$ for $\varepsilon = 0$. The sum then splits accordingly into $d + 1$ parts, which we handle separately below.

$\underline{\Lambda_0:}$ Let $\mu \in \Delta$ and $\lambda = (0, 0, \mathbf{0}, k) \in \Lambda_0$ with $k \in \mathbb{Z}^d$. The shearlet parametrization (15) yields $s_\lambda = 1$, $e_\lambda = e_d$, and $x_\lambda = \mathcal{T}k$. Furthermore, $s_\mu \geq 1$ for all $\mu \in \Lambda$. Hence we have

$$\omega_\alpha(\lambda, \mu) = s_\mu(1 + |\mathcal{T}k - x_\mu|^2 + |\{d_\mathbb{S}(e_d, e_\mu)\}|^2 + |\langle e_d, \mathcal{T}k - x_\mu\rangle|) \geq s_\mu(1 + |\mathcal{T}k - x_\mu|^2).$$

We conclude

$$\sum_{\lambda \in \Lambda_0} \omega_\alpha(\lambda, \mu)^{-N} \leq \sum_{k \in \mathbb{Z}^d} s_\mu^{-N}(1 + |\mathcal{T}k - x_\mu|^2)^{-N} \lesssim \sum_{k \in \mathbb{Z}^d} (1 + |k|^2)^{-N},$$

where for $N > d/2$ the sum on the right converges.

$\underline{\Lambda_\varepsilon, \; \varepsilon \in \{1, \ldots, d\}:}$ We only deal with the special case $\varepsilon = d$, since the other cases can be transformed to this case via rotations. Let $\mu \in \Delta$ and write $s_\mu = \sigma^{j'}$ with $j' \in \mathbb{R}$. In view of $s_\lambda = \sigma^j$ for $\lambda = (d, j, \ell, k) \in \Lambda_d$ we then have

$$\sum_{\lambda \in \Lambda_d} \omega_\alpha(\lambda, \mu)^{-N} = \sum_{j \in \mathbb{N}_0} \sigma^{-N|j-j'|} \sum_{\substack{\lambda \in \Lambda_d \\ s_\lambda = \sigma^j}} (1 + d_\alpha(\lambda, \mu))^{-N}.$$

If we can prove that

$$\mathcal{S} := \sum_{\substack{\lambda \in \Lambda_d \\ s_\lambda = \sigma^j}} (1 + d_\alpha(\lambda, \mu))^{-N} \lesssim \sigma^{d|j-j'|}, \tag{20}$$

independently of $j \in \mathbb{N}_0$ and $\mu \in \Delta$, we are done, since $\sigma > 1$, $s_\mu = \sigma^{j'}$, $\max\{s_\lambda/s_\mu, s_\mu/s_\lambda\} = \sigma^{|j-j'|}$ and thus if $N > d$

$$\sum_{\lambda \in \Lambda_d} \omega_\alpha(\lambda, \mu)^{-N} \lesssim \sum_{j \in \mathbb{N}_0} \sigma^{(d-N)|j-j'|} \leq 2 \sum_{j \in \mathbb{N}_0} \sigma^{(d-N)j} = \frac{2}{1 - \sigma^{d-N}} < \infty.$$

18

Putting in the definition of $d_\alpha(\lambda, \mu)$ and abbreviating $j_0 := \min\{j, j'\}$, the sum $\mathcal{S}$ becomes

$$\mathcal{S} = \sum_{\substack{\lambda \in \Lambda_d \\ s_\lambda = \sigma^j}} \left(1 + \sigma^{2\alpha j_0} |x_\lambda - x_\mu|^2 + \sigma^{2(1-\alpha)j_0} |\{d_\mathbb{S}(e_\lambda, e_\mu)\}|^2 + \sigma^{j_0} |\langle e_\lambda, x_\lambda - x_\mu\rangle|\right)^{-N}. \tag{21}$$

In order to prove the estimate (20) for $\mathcal{S}$, we first study the different terms of the summand independently. Let $\lambda = (d, j, \ell, k) \in \Lambda_d$ and recall the matrix M_λ from (16). It holds

$$M_\lambda^T = R_{\theta_\lambda} R_{\varphi_\lambda} S_{\ell \eta_j}^{-1},$$

and – according to the discussion of M_λ in the proof of Theorem 5.3 – its last row is given by $(0, \ldots, 0, n_\lambda)$ with $n_\lambda = (1 + \eta_j^2 |\ell|_2^2)^{-\frac{1}{2}}$. Since $\eta_j \asymp \sigma^{-j(1-\alpha)}$ and $|\ell|_2 \lesssim \mathbb{L}_j \lesssim \sigma^{j(1-\alpha)}$ this implies $n_\lambda \asymp 1$ uniformly for all $\lambda \in \Lambda_d$.

As a direct consequence $[M_\lambda^T x]_d = n_\lambda [x]_d \asymp [x]_d$ uniformly for $\lambda \in \Lambda_d$ and $x \in \mathbb{R}^d$. In addition, we have $|M_\lambda^T x| \asymp |x|$ uniformly for $\lambda \in \Lambda_d$ and $x \in \mathbb{R}^d$ since $\|M_\lambda^T\|_{2\to 2} = \|M_\lambda\|_{2\to 2} \lesssim 1$ and also $\|M_\lambda^{-T}\|_{2\to 2} = \|M_\lambda^{-1}\|_{2\to 2} \lesssim 1$.

These observations allow the following estimate,

$$\begin{aligned}
|x_\lambda - x_\mu| &= \left|S_{\ell\eta_j}^{-1} A_{\alpha,\sigma}^{-j} \mathcal{T} k - x_\mu\right| = \left|R_{\theta_\lambda} R_{\varphi_\lambda} S_{\ell\eta_j}^{-1} A_{\alpha,\sigma}^{-j} \mathcal{T} k - R_{\theta_\lambda} R_{\varphi_\lambda} x_\mu\right| \\
&= \left|M_\lambda^T \left(\mathcal{T} A_{\alpha,\sigma}^{-j} k - M_\lambda^{-T} R_{\theta_\lambda} R_{\varphi_\lambda} x_\mu\right)\right| \asymp \left|A_{\alpha,\sigma}^{-j} k - \mathcal{T}^{-1} M_\lambda^{-T} R_{\theta_\lambda} R_{\varphi_\lambda} x_\mu\right| \\
&\gtrsim \left|A_{\alpha,\sigma}^{-j} k - \mathcal{T}^{-1} S_{\ell\eta_j} x_\mu\right|_{[d-1]} = \left|\sigma^{-j\alpha} k - \mathcal{T}^{-1} S_{\ell\eta_j} x_\mu\right|_{[d-1]}.
\end{aligned} \tag{22}$$

In view of $e_\lambda = R_{\varphi_\lambda}^T R_{\theta_\lambda}^T e_d$ we also have the estimate

$$\begin{aligned}
|\langle e_\lambda, x_\lambda - x_\mu\rangle| &= \left|\langle e_\lambda, S_{\ell\eta_j}^{-1} A_{\alpha,\sigma}^{-j} \mathcal{T} k - x_\mu\rangle\right| = \left|\langle e_d, R_{\theta_\lambda} R_{\varphi_\lambda} S_{\ell\eta_j}^{-1} A_{\alpha,\sigma}^{-j} \mathcal{T} k - R_{\theta_\lambda} R_{\varphi_\lambda} x_\mu\rangle\right| \\
&= \left|\langle e_d, M_\lambda^T (\mathcal{T} A_{\alpha,\sigma}^{-j} k - M_\lambda^{-T} R_{\theta_\lambda} R_{\varphi_\lambda} x_\mu)\rangle\right| \asymp \left|\langle e_d, A_{\alpha,\sigma}^{-j} k - \mathcal{T}^{-1} S_{\ell\eta_j} x_\mu\rangle\right| \\
&= \left|\langle e_d, \sigma^{-j} k - \mathcal{T}^{-1} S_{\ell\eta_j} x_\mu\rangle\right|.
\end{aligned} \tag{23}$$

According to the shearlet parametrization (15) we have $e_\lambda = n_\lambda (\ell\eta_j, 1)^T$, where $n_\lambda \asymp 1$ as shown above. Hence, there is a constant $c > 0$ such that $n_\lambda \geq c$ for all $\lambda \in \Lambda_d$. It follows $[e_\lambda]_d \geq c > 0$ for all $\lambda \in \Lambda_d$. Without loss of generality we can further assume that $[e_\mu]_d \geq 0$ since $|\{d_\mathbb{S}(e_\lambda, -e_\mu)\}| = |\{d_\mathbb{S}(e_\lambda, e_\mu)\}|$.

In this situation Lemma 6.3 applies and tells us that $|\{d_\mathbb{S}(e_\lambda, e_\mu)\}| \asymp |e_\lambda - e_\mu|$. Moreover, possibly after changing e_μ to $\tilde{e}_\mu$ as in Lemma 5.5 to enforce $[\tilde{e}_\mu]_d \geq c$, we have $|e_\lambda - e_\mu| \geq |e_\lambda - \tilde{e}_\mu|$ for all $\lambda \in \Lambda_d$.

Let ϕ be the gnomonic projection from Lemma 5.4. Then $\phi(e_\lambda) = (\ell\eta_j, 1)^T \in \mathbb{R}^d$ and Lemma 5.4 together with the observation $|\phi(v) - \phi(w)|_{[d-1]} = |\phi(v) - \phi(w)|$ implies the estimate,

$$|e_\lambda - \tilde{e}_\mu| \asymp |(\ell\eta_j, 1)^T - \phi(\tilde{e}_\mu)| = |(\ell\eta_j, 1)^T - \phi(\tilde{e}_\mu)|_{[d-1]}.$$

Subsequently, let $\nu_\mu \in \mathbb{R}^{d-1}$ be defined by $\phi(\tilde{e}_\mu) = (\nu_\mu, 1)^T$. Altogether we arrive at

$$|\{d_\mathbb{S}(e_\lambda, e_\mu)\}| \gtrsim |(\ell\eta_j, 1)^T - \phi(\tilde{e}_\mu)|_{[d-1]} = |\ell\eta_j - \nu_\mu|. \tag{24}$$

We now use (22), (23) and (24) to estimate the sum $\mathcal{S}$ in (21). Introducing the quantities $q_1(\ell) := \sigma^{j\alpha} \mathcal{T}^{-1} S_{\ell\eta_j} x_\mu$, $q_2(\ell) := \sigma^j \mathcal{T}^{-1} S_{\ell\eta_j} x_\mu$, and $q_3 := \eta_j^{-1} \nu_\mu$, and taking into account $\eta_j \asymp \sigma^{-(1-\alpha)j}$ we obtain

$$\mathcal{S} \lesssim \sum_{k \in \mathbb{Z}^d} \sum_{\ell \in \mathscr{L}_{d,j}} \left(1 + \sigma^{2\alpha(j_0-j)} |k - q_1(\ell)|_{[d-1]}^2 + \sigma^{(j_0-j)} |\langle e_d, k - q_2(\ell)\rangle| + \sigma^{2(1-\alpha)(j_0-j)} |\ell - q_3|^2\right)^{-N}.$$

Next, we distinguish the cases $j > j'$ and $j \leq j'$. For $j \leq j'$ we have $j_0 = j$ and we obtain

$$\begin{aligned}
\mathcal{S} &\lesssim \sum_{k \in \mathbb{Z}^d} \sum_{\ell \in \mathbb{Z}^{d-1}} \left(1 + |k - q_1(\ell)|_{[d-1]}^2 + |\langle e_d, k - q_2(\ell)\rangle| + |\ell - q_3|^2\right)^{-N} \\
&\lesssim \sum_{k \in \mathbb{Z}^d} \sum_{\ell \in \mathbb{Z}^{d-1}} \left(1 + |k|_{[d-1]}^2 + |\langle e_d, k\rangle| + |\ell|_{[d-1]}^2\right)^{-N} < \infty.
\end{aligned}$$

In case $j > j'$ it holds $j_0 = j'$ and $p := j' - j \leq 0$. The term $\sigma^{pd}\mathcal{S}$ is – up to a multiplicative constant – bounded by

$$\sum_{\ell \in \mathbb{Z}^{d-1}} \sigma^{p(1-\alpha)(d-1)} \sum_{k \in \mathbb{Z}^d} \sigma^{p\alpha(d-1)} \sigma^p \big(1 + \sigma^{2\alpha p} |k - q_1(\ell)|_{[d-1]}^2 + \sigma^p |\langle e_d, k - q_2(\ell)\rangle| + \sigma^{2(1-\alpha)p} |\ell - q_3|^2\big)^{-N}.$$

The last sum can be interpreted as a Riemann sum, which is bounded – up to a multiplicative constant independent of j and j' as long as $j > j'$ – by the corresponding integral

$$\int_{y \in \mathbb{R}^{d-1}} \int_{x \in \mathbb{R}^d} \big(1 + |x - \sigma^{\alpha p} q_1(y)|_{[d-1]}^2 + |\langle e_d, x - \sigma^p q_2(y)\rangle| + |y - \sigma^{(1-\alpha)p} q_3|^2\big)^{-N} dx\, dy.$$

All in all we end up with $\mathcal{S} \lesssim \sigma^{d(j-j')} \int_{y \in \mathbb{R}^{d-1}} \int_{x \in \mathbb{R}^d} \big(1 + |x|_{[d-1]}^2 + |\langle e_d, x\rangle| + |y|^2\big)^{-N} dx\, dy$. To see that the integral converges for $N > d$, we carry out the integration over x_d, which yields up to a fixed constant

$$\int_{y \in \mathbb{R}^{d-1}} \int_{\tilde{x} \in \mathbb{R}^{d-1}} (1 + |\tilde{x}|_{[d-1]}^2 + |y|^2)^{-(N-1)} d\tilde{x}\, dy = \int_{z \in \mathbb{R}^{2(d-1)}} (1 + |z|^2)^{-(N-1)}.$$

The integral on the right converges precisely for $N > d$. This observation concludes the proof. $\qquad\square$

5.3 Pyramid-adapted Shearlet Systems in 3D

In the sequel we focus on some concrete shearlet systems in 3 dimensions, which are already on the market and for which we verify that they are instances of α-shearlet molecules for $\alpha = \frac{1}{2}$ and thus by Theorem 5.3 also systems of 3-dimensional parabolic molecules.

Let us first recall the classic definition [39] of a shearlet system in $L^2(\mathbb{R}^3)$. A system of 3D-shearlets is defined as a collection of functions in $L^2(\mathbb{R}^3)$ of the form

$$\left\{ \psi_{j,\ell,k} = 2^j \psi(S_\ell A_{\frac{1}{2},2}^j \cdot - k) \ : \ j \in \mathbb{Z}, \ell \in \mathbb{Z}^2, k \in \mathbb{Z}^3 \right\}, \tag{25}$$

where $\psi \in L^2(\mathbb{R}^3)$ is a suitable generating function. The classical choice for the generator is a function ψ defined on the frequency domain by

$$\hat{\psi}(\xi) = w(\xi_3) v(\tfrac{\xi_1}{\xi_3}) v(\tfrac{\xi_2}{\xi_3}), \quad \xi = (\xi_1, \xi_2, \xi_3)^T \in \mathbb{R}^3, \tag{26}$$

where $v \in C_c^\infty(\mathbb{R})$ is a bump function and $w \in C_c^\infty(\mathbb{R})$ the Fourier transform of a suitable univariate discrete wavelet. It is possible to choose v and w so that (25) becomes a Parseval frame for $L^2(\mathbb{R}^3)$.

Unfortunately, the shearlet system (25) has a directional bias due to the fact, that for large shearings the frequency support of the elements becomes more and more elongated along the (e_1, e_2)-plane. This directional bias affects negatively the approximation properties of (25) and makes this system impractical in most applications.

To avoid this problem, the Fourier domain is usually partitioned into three pyramidal regions

$$\mathcal{P}_1 = \left\{ (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 \ : \ |\tfrac{\xi_2}{\xi_1}| \leq 1, |\tfrac{\xi_3}{\xi_1}| \leq 1 \right\},$$

$$\mathcal{P}_2 = \left\{ (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 \ : \ |\tfrac{\xi_1}{\xi_2}| \leq 1, |\tfrac{\xi_3}{\xi_2}| \leq 1 \right\},$$

$$\mathcal{P}_3 = \left\{ (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 \ : \ |\tfrac{\xi_1}{\xi_3}| \leq 1, |\tfrac{\xi_2}{\xi_3}| \leq 1 \right\},$$

and for each pyramid a separate shearlet system is used. Then, since each system only has to cover one pyramid, the shear parameters can be restricted avoiding large shears. To take care of low frequencies, it is common to use distinguished coarse-scale elements with frequencies in a centered box. Here this will be the cube

$$\mathcal{R} = \left\{ \xi \in \mathbb{R}^3 \ : \ |\xi|_\infty \leq \tfrac{1}{8} \right\}.$$

This cube together with the truncated pyramids $\tilde{\mathcal{P}}_1 := \mathcal{P}_1 \backslash \mathcal{R}$, $\tilde{\mathcal{P}}_2 := \mathcal{P}_2 \backslash \mathcal{R}$, and $\tilde{\mathcal{P}}_3 := \mathcal{P}_3 \backslash \mathcal{R}$ partitions the Fourier domain into 4 regions.

With each of these regions, different operators are associated. The coarse-scale functions are only translated, in the other regions we also scale and shear. The scaling and shearing operators associated with the respective regions $\varepsilon \in \{1, 2, 3\}$ are given by $A_{\frac{1}{2}, s}^{(\varepsilon)} = Z^\varepsilon A_{\frac{1}{2}, s} Z^{-\varepsilon}$ and $S_h^{(\varepsilon)} = Z^\varepsilon S_h Z^{-\varepsilon}$, and take the concrete form

$$
A_{\frac{1}{2}, s}^{(1)} = \begin{pmatrix} s & 0 & 0 \\ 0 & s^{\frac{1}{2}} & 0 \\ 0 & 0 & s^{\frac{1}{2}} \end{pmatrix}, \quad A_{\frac{1}{2}, s}^{(2)} = \begin{pmatrix} s^{\frac{1}{2}} & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s^{\frac{1}{2}} \end{pmatrix}, \quad A_{\frac{1}{2}, s}^{(3)} = A_{\frac{1}{2}, s} = \begin{pmatrix} s^{\frac{1}{2}} & 0 & 0 \\ 0 & s^{\frac{1}{2}} & 0 \\ 0 & 0 & s \end{pmatrix},
$$

for $s > 0$, and for $h \in \mathbb{R}^2$

$$
S_h^{(1)} = \begin{pmatrix} 1 & h_1 & h_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad S_h^{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ h_2 & 1 & h_1 \\ 0 & 0 & 1 \end{pmatrix}, \quad S_h^{(3)} = S_h = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h_1 & h_2 & 1 \end{pmatrix}.
$$

Now we are ready to define a modified shearlet system, which is adapted to our partition of the Fourier domain and therefore called *cone-adapted*. These systems do not exhibit the directional bias as (25). In the 3D cone-adapted setting they are called *pyramid-adapted 3D shearlet systems*.

Definition 5.7 ([41, 42]). *For fixed $\tau_1, \tau_2 > 0$ let $\mathcal{T} = \mathrm{diag}(\tau_1, \tau_2, \tau_2) \in \mathbb{R}^{3 \times 3}$. The (affine) pyramid-adapted 3D shearlet system generated by the functions $\phi \in L^2(\mathbb{R}^3)$ and $\psi^\varepsilon \in L^2(\mathbb{R}^3)$, $\varepsilon \in \{1, 2, 3\}$, is defined as the union*

$$
SH(\phi, \psi^1, \psi^2, \psi^3; \tau_1, \tau_2) := \Phi(\phi; \tau_1) \cup \Psi_1(\psi^1; \tau_1, \tau_2) \cup \Psi_2(\psi^2; \tau_1, \tau_2) \cup \Psi_3(\psi^3; \tau_1, \tau_2) \tag{27}
$$

of the coarse-scale functions $\Phi(\phi; \tau_1) := \{\phi_k = \phi(\cdot - \tau_1 k) : k \in \mathbb{Z}^3\}$ and the functions

$$
\Psi_\varepsilon(\psi^\varepsilon; \tau_1, \tau_2) := \left\{ \psi_{j, \ell, k}^\varepsilon = 2^j \psi^\varepsilon(Z^\varepsilon S_\ell A_{\frac{1}{2}, 2}^j Z^{-\varepsilon} \cdot - Z^\varepsilon \mathcal{T} Z^{-\varepsilon} k) : \ j \in \mathbb{N}_0, \ell \in \mathbb{Z}^2, |\ell|_\infty \le \lceil 2^{j/2} \rceil, k \in \mathbb{Z}^3 \right\}
$$

associated with the pyramids $\tilde{\mathcal{P}}_\varepsilon$ for $\varepsilon \in \{1, 2, 3\}$.

These pyramid-adapted affine systems are the prime examples of $\frac{1}{2}$-shearlet-molecules. In practice, one wants them to be frames, especially tight frames. However, the construction of tight frames of pyramid-adapted shearlets is not trivial.

The simplest way to obtain a Parseval frame of pyramid-adapted shearlets starts with a Parseval shearlet frame of the type (25), which is easier to construct. It yields a shearlet system associated with the pyramid $\tilde{\mathcal{P}}_3$ by removing all elements, whose frequency support does not intersect $\tilde{\mathcal{P}}_3$. Truncating the remaining functions in the frequency domain outside of $\tilde{\mathcal{P}}_3$, one obtains a Parseval frame for the space

$$
L^2(\tilde{\mathcal{P}}_3)^\vee := \{f \in L^2(\mathbb{R}^3) \ : \ \mathrm{supp}\, \hat{f} \subset \tilde{\mathcal{P}}_3\}.
$$

A similar procedure yields Parseval frames associated with the the other parts of the Fourier domain, namely for $L^2(\tilde{\mathcal{P}}_\varepsilon)^\vee$, $\varepsilon \in \{1, 2\}$, and $L^2(\mathcal{R})^\vee$. The union of these frames then is a Parseval frame for the whole space $L^2(\mathbb{R}^3)$.

This approach has the drawback that it leads to bad spatial localization of the shearlets due to their lack of smoothness in the frequency domain, which is a consequence of the truncation. A different approach was taken by Candès, Demanet, and Ying in [49]. They gave up on the affine structure of the system and found a shearlet-like construction of a Parseval frame. Guo and Labate modified this approach [33, 34] and found another shearlet-type construction, which is even close to affine.

5.3.1 Bandlimited Tight Frames of Pyramid-adapted Shearlets

The construction we present here is the one due to Guo and Labate [33, 34]. It starts with a Meyer scaling function $\phi \in \mathscr{S}(\mathbb{R})$ satisfying $0 \le \hat{\phi} \le 1$, supp $\hat{\phi} \subseteq [-\frac{1}{8}, \frac{1}{8}]$, and $\hat{\phi} = 1$ on $[-\frac{1}{16}, \frac{1}{16}]$, which is used to define $\Phi \in \mathscr{S}(\mathbb{R}^3)$ on the Fourier side by

$$
\hat{\Phi}(\xi) := \hat{\phi}(\xi_1) \hat{\phi}(\xi_2) \hat{\phi}(\xi_3), \quad \xi = (\xi_1, \xi_2, \xi_3)^T \in \mathbb{R}^3. \tag{28}
$$

Then the function

$$W(\xi) := \sqrt{\hat{\Phi}^2(2^{-2}\xi) - \hat{\Phi}^2(\xi)}, \quad \xi \in \mathbb{R}^3,$$

is defined, with $\operatorname{supp} W \subseteq [-\frac{1}{2}, \frac{1}{2}]^3 \backslash (-\frac{1}{16}, \frac{1}{16})^3$ and $W = 1$ on $[-\frac{1}{4}, \frac{1}{4}]^3 \backslash (-\frac{1}{8}, \frac{1}{8})^3$, such that $\hat{\Phi}^2(\xi) + \sum_{j \geq 0} W^2(2^{-2j}\xi) = 1$.for every $\xi \in \mathbb{R}^3$. These functions thus produce a smooth tiling of the frequency domain into cartesian coronae. To take care of the directional scaling, a bump-like function $v \in C_c^\infty(\mathbb{R})$ is used, which satisfies $\operatorname{supp} v \subseteq [-1, 1]$ and

$$|v(t-1)|^2 + |v(t)|^2 + |v(t+1)|^2 = 1 \quad \text{for } t \in [-1, 1],$$
$$v(0) = 1 \text{ and } v^{(n)}(0) = 0 \text{ for } n \geq 1.$$

For an explicit construction of such a function we refer to [31]. Then $V \in C^\infty(\mathbb{R}^3)$ is defined by

$$V(\xi) = v(\tfrac{\xi_1}{\xi_3})v(\tfrac{\xi_2}{\xi_3}), \quad \xi = (\xi_1, \xi_2, \xi_3)^T \in \mathbb{R}^3,$$

with support fully contained in the pyramid $\mathcal{P}_3$. After this preparation the smooth Parseval frame of band-limited 3D-shearlets introduced by Guo and Labate in [34] can be defined.

The *coarse-scale functions*, which take care of the low frequencies in $\mathcal{R}$, are translates of the function Φ from (28),

$$SH_0 := \left\{ \psi_{0,\mathbf{0},k}^0(x) = \Phi(\cdot - k) \ : \ k \in \mathbb{Z}^3 \right\}. \tag{29}$$

The shearlets, whose frequency support is fully contained in the respective pyramids, are called *interior shearlets*. They are defined as the collection of functions

$$SH_{int} := \left\{ \psi_{j,\ell,k}^\varepsilon \ : \ (\varepsilon, j, \ell, k) \in \Lambda_{int} \right\}$$

indexed by $\Lambda_{int} := \{(\varepsilon, j, \ell, k) \in \{1, 2, 3\} \times \mathbb{N}_0 \times \mathbb{Z}^2 \times \mathbb{Z}^3 : |\ell|_\infty < 2^j\}$ and with Fourier transforms

$$\hat{\psi}_{j,\ell,k}^\varepsilon(\xi) = 2^{-2j}W(2^{-2j}\xi)V(S_\ell^{-T}A_{\frac{1}{2},4}^{-j}Z^{-\varepsilon}\xi)\exp(-2\pi i\langle Z^\varepsilon S_\ell^{-T}A_{\frac{1}{2},4}^{-j}Z^{-\varepsilon}\xi, k\rangle), \quad \xi \in \mathbb{R}^3, \tag{30}$$

where W and V are the auxiliary functions from above. Observe that

$$V(S_\ell^{-T}A_{\frac{1}{2},4}^{-j}\xi) = v(2^j\tfrac{\xi_1}{\xi_3} - \ell_1)v(2^j\tfrac{\xi_2}{\xi_3} - \ell_2),$$

and compare this to (26). Finally, the so-called *boundary shearlets*, obtained by carefully glueing together shearlets from adjacent pyramidal regions, are added to obtain a smooth well-localized frame. The glueing process is rather delicate, since it is important to select the right shearlets matching together. In fact, even in the original construction [34] there are some inaccuracies. We correct them here, which leads to a slight modification of the original definition.

We make a distinction between boundary shearlets defined between two pyramidal regions and those at the corners, where three pyramidal regions meet. Following [34] we also distinguish between the scales $j \geq 1$ and $j = 0$.

Let us begin with the scales $j \geq 1$ and the functions at the boundary where only two pyramidal regions meet. We define, for $j \geq 1$, $\varepsilon_1 \in \{-1, 1\}$, $\ell_1 = \varepsilon_1 2^j$, $|\ell_2| < 2^j$, and $k \in \mathbb{Z}^3$,

$$\hat{\psi}_{j,\ell,k}^1(\xi) = \begin{cases} 2^{-2j-3}W(2^{-2j}\xi)v(2^j\tfrac{\xi_2}{\xi_1} - \ell_1)v(2^j\tfrac{\xi_3}{\xi_1} - \ell_2)\exp(-2\pi i\langle 2^{-2}ZS_\ell^{-T}A_{\frac{1}{2},2^{-2j}}Z^{-1}\xi, k\rangle) & ,\xi \in \mathcal{P}_1; \\ 2^{-2j-3}W(2^{-2j}\xi)v(2^j\tfrac{\xi_1}{\xi_2} - \ell_1)v(2^j\tfrac{\xi_3}{\xi_2} - \varepsilon_1\ell_2)\exp(-2\pi i\langle 2^{-2}ZS_\ell^{-T}A_{\frac{1}{2},4}^{-j}Z^{-1}\xi, k\rangle) & ,\xi \in \mathcal{P}_2; \end{cases}$$

$$\hat{\psi}_{j,\ell,k}^2(\xi) = \begin{cases} 2^{-2j-3}W(2^{-2j}\xi)v(2^j\tfrac{\xi_3}{\xi_2} - \ell_1)v(2^j\tfrac{\xi_1}{\xi_2} - \ell_2)\exp(-2\pi i\langle 2^{-2}Z^2S_\ell^{-T}A_{\frac{1}{2},4}^{-j}Z^{-2}\xi, k\rangle) & ,\xi \in \mathcal{P}_2; \\ 2^{-2j-3}W(2^{-2j}\xi)v(2^j\tfrac{\xi_2}{\xi_3} - \ell_1)v(2^j\tfrac{\xi_1}{\xi_3} - \varepsilon_1\ell_2)\exp(-2\pi i\langle 2^{-2}Z^2S_\ell^{-T}A_{\frac{1}{2},4}^{-j}Z^{-2}\xi, k\rangle) & ,\xi \in \mathcal{P}_3; \end{cases}$$

$$\hat{\psi}_{j,\ell,k}^3(\xi) = \begin{cases} 2^{-2j-3}W(2^{-2j}\xi)v(2^j\tfrac{\xi_1}{\xi_3} - \ell_1)v(2^j\tfrac{\xi_2}{\xi_3} - \ell_2)\exp(-2\pi i\langle 2^{-2}S_\ell^{-T}A_{\frac{1}{2},4}^{-j}\xi, k\rangle) & ,\xi \in \mathcal{P}_3; \\ 2^{-2j-3}W(2^{-2j}\xi)v(2^j\tfrac{\xi_3}{\xi_1} - \ell_1)v(2^j\tfrac{\xi_2}{\xi_1} - \varepsilon_1\ell_2)\exp(-2\pi i\langle 2^{-2}S_\ell^{-T}A_{\frac{1}{2},4}^{-j}\xi, k\rangle) & ,\xi \in \mathcal{P}_1. \end{cases} \tag{31}$$

Note that we only give the definition in the regions where the functions have non-trivial support. Outside they are supposed to be zero.

Next, we come to the corners where three pyramidal regions meet. Here we have to glue together three shearlet parts, each coming from a different pyramid. For convenience the corner elements will be associated to the first pyramid $\mathcal{P}_1$. For $j \geq 1$, $\ell_1 = \varepsilon_1 2^j$, $\ell_2 = \varepsilon_2 2^j$ with $\varepsilon_1, \varepsilon_2 \in \{-1, 1\}$, and $k \in \mathbb{Z}^3$, they are defined as follows:

$$\hat{\psi}^1_{j,\ell,k}(\xi) = \begin{cases} 2^{-2j-3}W(2^{-2j}\xi)v(2^j\frac{\xi_2}{\xi_1} - \ell_1)v(2^j\frac{\xi_3}{\xi_1} - \ell_2)\exp(-2\pi i\langle 2^{-2}ZS_\ell^{-T}A_{\frac{1}{2},4}^{-j}Z^{-1}\xi, k\rangle) & ,\xi \in \mathcal{P}_1; \\ 2^{-2j-3}W(2^{-2j}\xi)v(2^j\frac{\xi_1}{\xi_2} - \ell_1)v(2^j\frac{\xi_3}{\xi_2} - \varepsilon_1\ell_2)\exp(-2\pi i\langle 2^{-2}ZS_\ell^{-T}A_{\frac{1}{2},4}^{-j}Z^{-1}\xi, k\rangle) & ,\xi \in \mathcal{P}_2; \\ 2^{-2j-3}W(2^{-2j}\xi)v(2^j\frac{\xi_2}{\xi_3} - \varepsilon_2\ell_1)v(2^j\frac{\xi_1}{\xi_3} - \ell_2)\exp(-2\pi i\langle 2^{-2}ZS_\ell^{-T}A_{\frac{1}{2},4}^{-j}Z^{-1}\xi, k\rangle) & ,\xi \in \mathcal{P}_3. \end{cases} \tag{32}$$

As in the original construction the definition of the boundary shearlets is slightly different at the lowest scale $j = 0$. For $j = 0$, $\ell_1 = \pm 1$, $\ell_2 = 0$, and $k \in \mathbb{Z}^3$, we set

$$\hat{\psi}^1_{j,\ell,k}(\xi) = \begin{cases} W(\xi)v(\frac{\xi_2}{\xi_1} - \ell_1)v(\frac{\xi_3}{\xi_1})\exp(-2\pi i\langle\xi, k\rangle) & ,\xi \in \mathcal{P}_1; \\ W(\xi)v(\frac{\xi_1}{\xi_2} - \ell_1)v(\frac{\xi_3}{\xi_2})\exp(-2\pi i\langle\xi, k\rangle) & ,\xi \in \mathcal{P}_2; \end{cases}$$

$$\hat{\psi}^2_{j,\ell,k}(\xi) = \begin{cases} W(\xi)v(\frac{\xi_3}{\xi_2} - \ell_1)v(\frac{\xi_1}{\xi_2})\exp(-2\pi i\langle\xi, k\rangle) & ,\xi \in \mathcal{P}_2; \\ W(\xi)v(\frac{\xi_2}{\xi_3} - \ell_1)v(\frac{\xi_1}{\xi_3})\exp(-2\pi i\langle\xi, k\rangle) & ,\xi \in \mathcal{P}_3; \end{cases} \tag{33}$$

$$\hat{\psi}^3_{j,\ell,k}(\xi) = \begin{cases} W(\xi)v(\frac{\xi_1}{\xi_3} - \ell_1)v(\frac{\xi_2}{\xi_3})\exp(-2\pi i\langle\xi, k\rangle) & ,\xi \in \mathcal{P}_3; \\ W(\xi)v(\frac{\xi_3}{\xi_1} - \ell_1)v(\frac{\xi_2}{\xi_1})\exp(-2\pi i\langle\xi, k\rangle) & ,\xi \in \mathcal{P}_1. \end{cases}$$

Finally, we come to the corner elements at the scale $j = 0$. Again they are associated with the first pyramid. Let $\varepsilon_1, \varepsilon_2 \in \{-1, 1\}$. Then we define them for $j = 0$, $\ell = (\varepsilon_1, \varepsilon_2)$, and $k \in \mathbb{Z}^3$ by

$$\hat{\psi}^1_{j,\ell,k}(\xi) = \begin{cases} W(\xi)v(\frac{\xi_2}{\xi_1} - \ell_1)v(\frac{\xi_3}{\xi_1} - \ell_2)\exp(-2\pi i\langle\xi, k\rangle) & ,\xi \in \mathcal{P}_1; \\ W(\xi)v(\frac{\xi_1}{\xi_2} - \ell_1)v(\frac{\xi_3}{\xi_2} - \varepsilon_1\ell_2)\exp(-2\pi i\langle\xi, k\rangle) & ,\xi \in \mathcal{P}_2; \\ W(\xi)v(\frac{\xi_2}{\xi_3} - \varepsilon_2\ell_1)v(\frac{\xi_1}{\xi_3} - \ell_2)\exp(-2\pi i\langle\xi, k\rangle) & ,\xi \in \mathcal{P}_3. \end{cases} \tag{34}$$

All boundary shearlets are collected in the family SH_{bound}. Together with the coarse-scale functions SH_0 and the interior shearlets SH_{int} they provide a Parseval frame for $L^2(\mathbb{R}^3)$.

Theorem 5.8 ([34]). *The system of $3D$-shearlets*

$$SH := SH_0 \cup SH_{int} \cup SH_{bound}$$

is a smooth (well-localized) Parseval frame for $L^2(\mathbb{R}^3)$, consisting of band-limited Schwarz functions.

The corresponding index set $\Lambda_{SH} \subset \{0, 1, 2, 3\} \times \mathbb{N}_0 \times \mathbb{Z}^2 \times \mathbb{Z}^3$ is given by

$$\Lambda_{SH} := \Big\{(0, 0, \mathbf{0}, k) : k \in \mathbb{Z}^3\Big\} \cup \Big\{(\varepsilon, j, \ell, k) \; : \; \varepsilon \in \{1, 2, 3\}, j \in \mathbb{N}_0, \ell \in \mathscr{L}_{\varepsilon,j} \subseteq \mathbb{Z}^2, k \in \mathbb{Z}^3\Big\} \tag{35}$$

with the shear parameters $\mathscr{L}_{\varepsilon,j} = \{(\ell_1, \ell_2) \in \mathbb{Z}^2 : |\ell_1| \leq 2^j, |\ell_2| < 2^j\}$ for $\varepsilon \in \{2, 3\}$ and $\mathscr{L}_{1,j} = \{(\ell_1, \ell_2) \in \mathbb{Z}^2 : |\ell_1| \leq 2^j, |\ell_2| < 2^j\} \cup \{(\pm 2^j, \pm 2^j)\}$ at each scale $j \in \mathbb{N}_0$.

The system SH is another instance of a system of $\frac{1}{2}$-shearlet molecules (at least after an appropriate re-indexing). In particular, it falls within the framework of α-molecules for $\alpha = \frac{1}{2}$.

Proposition 5.9. *Appropriately re-indexed, the smooth Parseval frame of band-limited shearlets SH constitutes a system of 3-dimensional $\frac{1}{2}$-shearlet molecules of order $(\infty, \infty, \infty, \infty)$.*

Proof. We first re-index the coarse-scale functions (29) as well as the boundary elements (33) and (34) at scale $j = 0$. For this we utilize the index set $\Gamma \subset \{0, 1, 2, 3\} \times \mathbb{N}_0 \times \mathbb{Z}^2$ given by

$$\Gamma := \Big\{(0, 0, \mathbf{0})\Big\} \cup \Big\{(\varepsilon, 0, \ell) : \varepsilon \in \{1, 2, 3\}, \ell = (\pm 1, 0)\Big\} \cup \Big\{(1, 0, \ell) : \ell = (\pm 1, \pm 1)\Big\}.$$

The functions we want to re-index are then precisely those functions $\psi_{j,\ell,k}^\varepsilon \in SH$ with $(\varepsilon, j, \ell, k) \in \Delta :=$ $\Gamma \times \mathbb{Z}^3$. It holds $\#\Gamma = 11$, and we enumerate the set Γ in lexicographic order from 0 to 10, described by a bijective function $\mathcal{N} : \Gamma \to \{0, \ldots, 10\}$. For $(\varepsilon, j, \ell, k) \in \Delta$ we then re-index as follows. Writing $\tilde{k}(\varepsilon, j, \ell, k) := (k_1, k_2, 11 \cdot k_3 + \mathcal{N}(\varepsilon, j, \ell)) \in \mathbb{Z}^3$ the re-indexed elements are given by

$$\widetilde{\psi}_{0,\mathbf{0},\tilde{k}(\varepsilon,j,\ell,k)}^0 := \psi_{j,\ell,k}^\varepsilon \quad \text{for } (\varepsilon, j, \ell, k) \in \Delta.$$

For $(\varepsilon, j, \ell, k) \in \Lambda_{SH} \setminus \Delta$ the functions $\tilde{\psi}_{j,\ell,k}^\varepsilon := \psi_{j,\ell,k}^\varepsilon$ remain the same. The relabelling is thus given by

$$F_{label} : \Lambda_{SH} \to \widetilde{\Lambda}_{SH}, \ (\varepsilon, j, \ell, k) \mapsto \begin{cases} (0, 0, \mathbf{0}, \tilde{k}(\varepsilon, j, \ell, k)) & , (\varepsilon, j, \ell, k) \in \Delta; \\ (\varepsilon, j, \ell, k) & , (\varepsilon, j, \ell, k) \notin \Delta. \end{cases} \tag{36}$$

The newly obtained system

$$\widetilde{SH} := \{\widetilde{\psi}_{j,\ell,k}^\varepsilon\}_{(\varepsilon,j,\ell,k)\in\widetilde{\Lambda}_{SH}} \tag{37}$$

is equipped with a shearlet index set $\tilde{\Lambda}_{SH} \subset \{0, 1, 2, 3\} \times \mathbb{N}_0 \times \mathbb{Z}^2 \times \mathbb{Z}^3$ similar to (35), however with the modified shear parameters $\mathscr{L}_{1,0} = \mathscr{L}_{2,0} = \mathscr{L}_{3,0} = \{(0,0)\}$ at scale $j = 0$.

In the remainder we show that (37) is a system of $\frac{1}{2}$-shearlet molecules of order $(\infty, \infty, \infty, \infty)$. As parameters we fix $\sigma = 4$ and $\eta_j = 2^{-j}$ for $j \in \mathbb{N}_0$. The translation parameters $\mathcal{T} = \text{diag}(\tau_1, \tau_2, \tau_3)$ vary with the indices and are chosen suitably later. Clearly, we have $S_\ell A_{\frac{1}{2},4}^j = A_{\frac{1}{2},4}^j S_{\ell\eta_j}$. Thus, defining

$$\gamma_{j,\ell,k}^\varepsilon(x) := 2^{-2j}\tilde{\psi}_{j,\ell,k}^\varepsilon\big(Z^\varepsilon A_{\frac{1}{2},4}^{-j} S_\ell^{-1} Z^{-\varepsilon}(x + \mathcal{T}k)\big), \quad x \in \mathbb{R}^3, \tag{38}$$

for every $(\varepsilon, j, \ell, k) \in \widetilde{\Lambda}_{SH}$ we get the desired representation (12), i.e.

$$\tilde{\psi}_{j,\ell,k}^\varepsilon(x) = 2^{2j}\gamma_{j,\ell,k}^\varepsilon(Z^\varepsilon S_\ell A_{\frac{1}{2},4}^j Z^{-\varepsilon}x - \mathcal{T}k) = 2^{2j}\gamma_{j,\ell,k}^\varepsilon(Z^\varepsilon A_{\frac{1}{2},4}^j S_{\ell\eta_j} Z^{-\varepsilon}x - \mathcal{T}k), \quad x \in \mathbb{R}^3.$$

On the Fourier side the generators (38) take the form

$$\hat{\gamma}_{j,\ell,k}^\varepsilon(\xi) = 2^{2j}\hat{\tilde{\psi}}_{j,\ell,k}^\varepsilon\big(Z^\varepsilon A_{\frac{1}{2},4}^j S_\ell^T Z^{-\varepsilon}\xi\big)\exp(2\pi i\langle\xi, \mathcal{T}k\rangle), \quad \xi \in \mathbb{R}^3, \tag{39}$$

and it remains to show that these functions satisfy (13). For this task we distinguish between the coarse-scale elements, the interior shearlets and the boundary shearlets of (37).

Coarse-Scale: The coarse-scale elements $\{\tilde{\psi}_{0,\mathbf{0},k}^0\}_{k\in\mathbb{Z}^3}$ of (37) are precisely those functions $\psi_{j,\ell,k}^\varepsilon \in SH$ where $(\varepsilon, j, \ell, k) \in \Delta$. For them (38) simplifies to $\gamma_{0,\mathbf{0},k}^0 = \tilde{\psi}_{0,\mathbf{0},k}^0(\cdot + \mathcal{T}k)$, and we just need to show (13) for $\hat{\gamma}_{0,\mathbf{0},k}^0$ with $k \in \mathbb{Z}^3$. For this $\mathcal{T} = \text{diag}(1, 1, \frac{1}{11})$ is the right choice. A calculation yields for every $(\varepsilon, j, \ell, k) \in \Delta$ and $\tilde{k} = \tilde{k}(\varepsilon, j, \ell, k) \in \mathbb{Z}^3$

$$\hat{\gamma}_{0,\mathbf{0},\tilde{k}(\varepsilon,j,\ell,k)}^0(\xi) = \hat{\psi}_{j,\ell,k}^\varepsilon(\xi)\exp(\tfrac{2\pi i}{11}\xi_3 \cdot \mathcal{N}(\varepsilon, j, \ell)), \quad \xi = (\xi_1, \xi_2, \xi_3)^T \in \mathbb{R}^3.$$

Hence, looking at the definitions (29), (33), (34), it follows $\hat{\gamma}_{0,\mathbf{0},k}^0 \in C_c^\infty(\mathbb{R}^3)$ with support in $[-\frac{1}{2}, \frac{1}{2}]^3$.

Interior Shearlets: Choosing $\mathcal{T} = \text{diag}(1, 1, 1)$, Equation (39) together with (30) yields for $\xi \in \mathbb{R}^3$

$$\hat{\gamma}_{j,\ell,k}^\varepsilon(\xi) = W(Z^\varepsilon M_{j,\ell} Z^{-\varepsilon}\xi)V(Z^{-\varepsilon}\xi),$$

where the matrix $M_{j,\ell} := 2^{-2j}A_{\frac{1}{2},4}^j S_\ell^T$ has the form

$$M_{j,\ell} = \begin{pmatrix} 2^{-j} & 0 & \ell_1 2^{-j} \\ 0 & 2^{-j} & \ell_2 2^{-j} \\ 0 & 0 & 1 \end{pmatrix}. \tag{40}$$

24

We check (13) exemplarily for the case $\varepsilon = 3$ where $Z^3 = I$. A calculation yields

$$M_{j,\ell}\xi = \begin{pmatrix} 2^{-j}(\xi_1 + \ell_1\xi_3) \\ 2^{-j}(\xi_2 + \ell_2\xi_3) \\ \xi_3 \end{pmatrix}, \quad \xi = (\xi_1, \xi_2, \xi_3)^T \in \mathbb{R}^3.$$

Subsequently we prove that for $j \in \mathbb{N}_0$ and $|\ell|_\infty \leq 2^j$

$$\mathrm{supp}\,(\hat{\gamma}^3_{j,\ell,k}) \subseteq \mathcal{P}_3 \cap \left\{\xi \in \mathbb{R}^3 \;:\; \tfrac{1}{32} \leq |\xi_3| \leq \tfrac{1}{2}\right\}. \tag{41}$$

First observe that clearly $\mathrm{supp}\,(\hat{\gamma}^3_{j,\ell,k}) \subseteq \mathcal{P}_3$. Now take $\xi \in \mathcal{P}_3$ and $\xi \notin \left\{\xi \in \mathbb{R}^3 : \tfrac{1}{32} \leq |\xi_3| \leq \tfrac{1}{2}\right\}$. With $|\xi_3| > \tfrac{1}{2}$ also $|[M_{j,\ell}\xi]_3| > \tfrac{1}{2}$ and thus $M_{j,\ell}\xi \notin \mathrm{supp}\,W$. If on the other hand $|\xi_3| < \tfrac{1}{32}$ then $|[M_{j,\ell}\xi]_3| < \tfrac{1}{32}$ and, since $j \geq 0$, $|\xi_1| \leq |\xi_3|$ and $|\ell|_\infty \leq 2^j$,

$$|[M_{j,\ell}\xi]_1| = |2^{-j}(\xi_1 + \ell_1\xi_3)| \leq 2^{-j}(1 + |\ell_1|)|\xi_3| \leq 2^{-j}(1 + 2^j)|\xi_3| \leq (1 + 2^{-j})|\xi_3| \leq 2|\xi_3| < \tfrac{1}{16}.$$

Analogously, one obtains $|[M_{j,\ell}\xi]_2| < \tfrac{1}{16}$. Altogether, this yields $|M_{j,\ell}\xi|_\infty < \tfrac{1}{16}$ if $|\xi_3| < \tfrac{1}{32}$, which implies $M_{j,\ell}\xi \notin \mathrm{supp}\,W$. Consequently, $\hat{\gamma}^3_{j,\ell,k}(\xi) = 0$ outside $\mathcal{P}_3 \cap \{\xi \in \mathbb{R}^3 : \tfrac{1}{32} \leq |\xi_3| \leq \tfrac{1}{2}\}$, which proves (41).

Using analogous estimates for $\varepsilon \in \{2,3\}$ it follows that $\mathrm{supp}\,\hat{\gamma}^\varepsilon_{j,\ell,k} \subseteq [-\tfrac{1}{2}, \tfrac{1}{2}]^3 \setminus (-\tfrac{1}{32}, \tfrac{1}{32})^3$ for every $(\varepsilon, j, \ell, k) \in \Lambda_{int} = \{(\varepsilon, j, \ell, k) \in \{1,2,3\} \times \mathbb{N}_0 \times \mathbb{Z}^2 \times \mathbb{Z}^3 : |\ell|_\infty < 2^j\}$.

The derivatives of $\hat{\gamma}^\varepsilon_{j,\ell,k}$ are linear combinations of functions $\partial^\beta W(Z^\varepsilon M_{j,\ell}Z^{-\varepsilon}\xi)\partial^\delta V(Z^{-\varepsilon}\xi)$ with coefficients uniformly bounded, since the entries of $M_{j,\ell}$ are. This fact together with the support condition implies that (13) is fulfilled for arbitrary order.

Boundary Shearlets: The boundary shearlets of (37) satisfy $j \geq 1$ and are precisely the functions in (31) and (32). We exemplarily handle the functions (31), the argumentation for the functions (32) is similar. Let us first look at the boundary between $\mathcal{P}_1$ and $\mathcal{P}_3$, i.e. the functions $\tilde{\psi}^\varepsilon_{j,\ell,k}$, where $\varepsilon = 3$, $j \geq 1$, $\ell_1 = \pm 2^j$, $|\ell_2| < 2^j$, and $k \in \mathbb{Z}^3$. Plugging (31) into (39) and choosing $\mathcal{T} = \tfrac{1}{4}\mathrm{diag}(1,1,1)$ we obtain

$$\hat{\gamma}^3_{j,\ell,k}(\xi) = 2^{-3}\begin{cases} W(M_{j,\ell}\xi)V(\xi) & ,\xi \in M_{j,\ell}^{-1}\mathcal{P}_3; \\ W(M_{j,\ell}\xi)V(N_{j,\ell}\xi) & ,\xi \in M_{j,\ell}^{-1}\mathcal{P}_1; \end{cases}$$

for the generators, where $M_{j,\ell}$ is the matrix (40) and where

$$N_{j,\ell} := \begin{pmatrix} -\ell_1 2^{-j} & 0 & 2^j - \ell_1^2 2^{-j} \\ -\,\mathrm{sgn}(\ell_1)\ell_2 2^{-j} & 1 & \ell_2 - 2^{-j}|\ell_1|\ell_2 \\ 2^{-j} & 0 & 2^{-j}\ell_1 \end{pmatrix}.$$

Similar calculations as for the interior shearlets then prove

$$\mathrm{supp}\,(\hat{\gamma}^3_{j,\ell,k}) \subseteq \left\{\xi \in \mathbb{R}^3 \;:\; |\tfrac{\xi_1}{\xi_3}| \leq 2, |\tfrac{\xi_2}{\xi_3}| \leq 4\right\} \cap \left\{\xi \in \mathbb{R}^3 \;:\; \tfrac{1}{64} \leq |\xi_3| \leq \tfrac{1}{2}\right\}.$$

For $\varepsilon \in \{1,2\}$ complementary results hold true. Altogether, it follows $\mathrm{supp}\,\hat{\gamma}^\varepsilon_{j,\ell,k} \subseteq [-2,2]^3 \setminus (-\tfrac{1}{64}, \tfrac{1}{64})^3$ for the generators of the functions (31). Hence, condition (13) is fulfilled for arbitrary order. $\qquad\square$

Proposition 5.9 in particular shows that the system SH is a system of $\tfrac{1}{2}$-molecules, however not with respect to a shearlet parametrization because of the necessary relabelling F_{label} in (36). The actual parametrization is given by

$$\Phi_{SH} := \widetilde{\Phi}_{SH} \circ F_{label},$$

where $\widetilde{\Phi}_{SH}$ denotes the $\tfrac{1}{2}$-shearlet parametrization of the relabeled system.

Corollary 5.10. *The smooth shearlet frame SH is a system of 3-dimensional $\tfrac{1}{2}$-molecules of order $(\infty, \infty, \infty, \infty)$ with respect to the parametrization $(\Lambda_{SH}, \Phi_{SH})$.*

Remark 5.11. *Although $(\Lambda_{SH}, \Phi_{SH})$ is not a shearlet parametrization, it clearly is $(\tfrac{1}{2}, k)$-consistent with every $\tfrac{1}{2}$-shearlet parametrization for $k > 3$. This follows from Proposition 5.6 and the observation that relabelling of elements does not make any difference here.*

5.3.2 Compactly Supported Pyramid-adapted Shearlets

There also exist shearlet frames for $L^2(\mathbb{R}^2)$ and $L^2(\mathbb{R}^3)$ consisting of compactly supported functions. Compactly supported frames of the form (27) have been constructed in [41, 37]. They are also instances of $\frac{1}{2}$-shearlet molecules and their order can be controlled by the regularity of the generators.

Proposition 5.12. *Let $\phi, \psi^1, \psi^2, \psi^3 \in L^2(\mathbb{R}^3)$ be compactly supported and $L, M, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$. If $\phi \in C^{N_1+N_2}(\mathbb{R}^3)$ and if, for every $\varepsilon \in \{1, 2, 3\}$,*

(i) the derivatives $\partial^\gamma \psi^\varepsilon$ exist and are continuous for every $\gamma \in \mathbb{N}_0^3$ with $[Z^\varepsilon \gamma]_1, [Z^\varepsilon \gamma]_2 \leq N_1 + N_2$ and $[Z^\varepsilon \gamma]_3 \leq N_1$, where Z is the cyclic permutation matrix (11),

(ii) the generator ψ^ε has $M + L$ directional vanishing moments in e_ε-direction, i.e.

$$\forall (x_1, x_2) \in \mathbb{R}^2 : \int_\mathbb{R} \psi^\varepsilon(Z^\varepsilon x) x_3^N \, dx_3 = 0 \quad \text{for every } N \in \{0, \ldots, M + L - 1\},$$

then the system (27) obtained from these generators is a system of $\frac{1}{2}$-shearlet molecules of order (L, M, N_1, N_2).

Proof. It is obvious that – rightly indexed – a system of the form (27) constitutes a system of $\frac{1}{2}$-shearlet molecules. Hence we just need to verify the order of the system. For this, little more is needed than utilizing the facts that spatial decay implies smoothness in Fourier domain (and vice versa), and that vanishing moments in spatial domain implies estimates of the form $|\hat{g}(\xi)| \lesssim \min(1, |\xi|)^M$ in Fourier domain. We refer to [27, Proposition 3.11] for details, where a similar two-dimensional version of the theorem is proven. $\square$

6 Proof of Theorem 2.5

This final section is devoted to the technical proof of Theorem 2.5. It is split up into several pieces and has the same general structure as the proof of the corresponding 2-dimensional result [28, Theorem 4.2]. In d dimensions however it takes more effort and the arguments are more involved.

6.1 Auxiliary Lemmas

Let us first collect some simple elementary facts, which turn out to be useful. Subsequently $\mathcal{O}(d, \mathbb{R})$ shall denote the orthogonal group of $\mathbb{R}^d$. Recall also the notation $\{\theta\}$ for the 'projection' of $\theta \in \mathbb{R}$ onto the interval $[-\frac{\pi}{2}, \frac{\pi}{2})$ in the sense of Subsection 2.1.

Lemma 6.1. *For $\theta \in \mathbb{R}$ let $\{\theta\}$ denote its 'projection' onto the interval $[-\frac{\pi}{2}, \frac{\pi}{2})$ as introduced in Subsection 2.1. It then holds $|\{\theta\}| \asymp |\sin(\theta)|$.*

Proof. Due to π-periodicity it suffices to verify the relation for $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2})$. In this range we have $\frac{2}{\pi}|\theta| \leq |\sin(\theta)| \leq |\theta|$. $\square$

An immediate corollary is the following result. Recall the notation $d_\mathbb{S}(v, w)$ for the angle $\arccos(\langle v, w \rangle) \in [0, \pi]$ between two vectors $v, w \in \mathbb{S}^{d-1}$.

Lemma 6.2. *Let $e_d \in \mathbb{R}^d$ be the d:th unit vector. For $\eta \in \mathbb{S}^{d-1}$ we have $|\{d_\mathbb{S}(\eta, e_d)\}| \asymp |\eta|_{[d-1]}$.*

Proof. Using a suitable rotation $R \in \mathcal{O}(d, \mathbb{R})$ of the form

$$R = \begin{pmatrix} R_{d-1} & 0 \\ 0 & 1 \end{pmatrix},$$

where $R_{d-1} \in \mathcal{O}(d-1, \mathbb{R})$, we can achieve $R\eta = (\sin(\theta), 0, \ldots, 0, \cos(\theta))^T$ with $\theta = d_\mathbb{S}(\eta, e_d)$. Since $|\eta|_{[d-1]} = |\eta - e_d|_{[d-1]} = |R(\eta - e_d)|_{[d-1]} = |R\eta - e_d|_{[d-1]} = |\sin(\theta)|$, it just remains to prove $|\sin(\theta)| \asymp |\{\theta\}|$, which is true by Lemma 6.1. $\square$

Lemma 6.3. *Let $c > 0$ be a constant. Then we have for all $v, w \in \mathbb{S}^{d-1}$ with $[v]_d \geq c$ and $[w]_d \geq 0$*

$$|\{d_{\mathbb{S}}(v, w)\}| \asymp |v - w|.$$

Proof. Under the assumptions there exists $\varepsilon > 0$ dependent on c, such that $0 \leq d_{\mathbb{S}}(v, w) \leq \pi - \varepsilon$. It follows $\frac{\varepsilon}{\pi - \varepsilon} |d_{\mathbb{S}}(v, w)| \leq |\{d_{\mathbb{S}}(v, w)\}| \leq |d_{\mathbb{S}}(v, w)|$. The observation $d_{\mathbb{S}}(v, w) \asymp |v - w|$ finishes the proof. $\square$

Lemma 6.4. *Let $R \in \mathcal{O}(d, \mathbb{R})$ be a rotation and $\theta_0 = d_{\mathbb{S}}(e_d, Re_d) \in [0, \pi]$ the angle between the d:th unit vector $e_d \in \mathbb{R}^d$ and its image Re_d under R. Then it holds for all $\eta \in \mathbb{S}^{d-1}$*

$$|R\eta|_{[d-1]} = \sin(d_{\mathbb{S}}(R\eta, e_d)) \geq \min\{|\sin(d_{\mathbb{S}}(\eta, e_d) + \theta_0)|, |\sin(d_{\mathbb{S}}(\eta, e_d) - \theta_0)|\}.$$

Note $d_{\mathbb{S}}(\eta, e_d) = d_{\mathbb{S}}(R\eta, Re_d)$.

Proof. Let $\eta = (\eta_1, \ldots, \eta_d)^T \in \mathbb{S}^{d-1}$ and put $\theta_1 := d_{\mathbb{S}}(\eta, e_d) = \arccos(\langle \eta, e_d \rangle) \in [0, \pi]$. The rotation $R \in \mathcal{O}(d, \mathbb{R})$ can be decomposed in the form $R = \tilde{R} R_{\theta_0}$ with $\tilde{R}, R_{\theta_0} \in \mathcal{O}(d, \mathbb{R})$ such that

$$\tilde{R} = \begin{pmatrix} R_{d-1} & \\ & 1 \end{pmatrix} \quad \text{and} \quad R_{\theta_0} = \begin{pmatrix} \cos(\theta_0) & & -\sin(\theta_0) \\ & I_{d-2} & \\ \sin(\theta_0) & & \cos(\theta_0) \end{pmatrix},$$

where $R_{d-1} \in \mathcal{O}(d-1, \mathbb{R})$ is some $(d-1)$-dimensional rotation matrix and I_{d-2} is the $(d-2)$-dimensional identity matrix. The rotation $\tilde{R}$ leaves $|\cdot|_{[d-1]}$ invariant, whence

$$|R\eta|_{[d-1]} = |\tilde{R} R_{\theta_0} \eta|_{[d-1]} = |R_{\theta_0} \eta|_{[d-1]}.$$

Using $\eta_d = \cos(\theta_1)$ and $|\eta|_{[d-1]}^2 = \eta_1^2 + \eta_2^2 + \ldots + \eta_{d-1}^2 = 1 - \eta_d^2$, it further follows

$$\begin{aligned} |R_{\theta_0} \eta|_{[d-1]}^2 &= (\cos(\theta_0)\eta_1 - \sin(\theta_0)\eta_d)^2 + \eta_2^2 + \cdots + \eta_{d-1}^2 \\ &= \cos^2(\theta_0)\eta_1^2 + \sin^2(\theta_0)\cos^2(\theta_1) - 2\cos(\theta_0)\sin(\theta_0)\eta_1 \cos(\theta_1) + (1 - \eta_1^2 - \cos^2(\theta_1)) \\ &= 1 - (\eta_1 \sin(\theta_0) + \cos(\theta_1)\cos(\theta_0))^2. \end{aligned}$$

The last expression is a second-degree polynomial in the variable η_1 with a negative leading coefficient. Since $\eta_1^2 \leq 1 - \eta_d^2 = 1 - \cos^2(\theta_1) = \sin^2(\theta_1)$, the variable η_1 can take values only in $[-\sin(\theta_1), \sin(\theta_1)]$. The polynomial attains its minimum on this interval at the endpoints. Hence, we can conclude

$$\begin{aligned} |R_{\theta_0} \eta|_{[d-1]}^2 &\geq \min_{\epsilon \in \{-1,1\}} \left\{ 1 - (\epsilon \sin(\theta_1)\sin(\theta_0) + \cos(\theta_1)\cos(\theta_0))^2 \right\} \\ &= \min_{\epsilon \in \{-1,1\}} \left\{ 1 - \cos^2(\theta_1 - \epsilon\theta_0) \right\} = \min_{\epsilon \in \{-1,1\}} \left\{ \sin^2(\theta_1 - \epsilon\theta_0) \right\}, \end{aligned}$$

which proves the claim. $\square$

6.2 Integral Estimates

We start with an estimate for the generators in (6), which will allow us to work in polar coordinates.

Lemma 6.5. *Let the family of functions $\{g_\lambda\}_{\lambda \in \Lambda}$ satisfy (7) uniformly for a multi-index $\rho \in \mathbb{N}_0^d$, and assume that there is a constant $c > 0$ such that $s_\lambda \geq c$ for all $\lambda \in \Lambda$. Then the following estimate holds true uniformly for $\lambda \in \Lambda$ and $\xi \in \mathbb{R}^d$*

$$\left| (\partial^\rho \hat{g}_\lambda)(A_{\alpha, s_\lambda}^{-1} R_{\theta_\lambda} R_{\varphi_\lambda} \xi) \right| \lesssim \frac{\min\left\{ 1, s_\lambda^{-1}(1 + |\xi|) \right\}^M}{\left(1 + s_\lambda^{-1}|\xi| \right)^{N_1} \left(1 + s_\lambda^{-\alpha} |R_{\theta_\lambda} R_{\varphi_\lambda} \xi|_{[d-1]} \right)^{N_2}}. \tag{42}$$

Proof. We have $|A_{\alpha, s_\lambda}^{-1} \xi| \geq \min\{s_\lambda^{-1}, s_\lambda^{-\alpha}\} |\xi| \gtrsim s_\lambda^{-1} |\xi|$ uniformly for $\xi \in \mathbb{R}^d$ and $\lambda \in \Lambda$, since $s_\lambda \geq c > 0$ for every $\lambda \in \Lambda$. It follows $|A_{\alpha, s_\lambda}^{-1} R_{\theta_\lambda} R_{\varphi_\lambda} \xi| \gtrsim s_\lambda^{-1} |R_{\theta_\lambda} R_{\varphi_\lambda} \xi| = s_\lambda^{-1} |\xi|$. Further, we observe $|A_{\alpha, s_\lambda}^{-1} \xi|_{[d-1]} = s_\lambda^{-\alpha} |\xi|_{[d-1]}$ and $|[A_{\alpha, s_\lambda}^{-1} \xi]_d| = s_\lambda^{-1} |[\xi]_d|$. Finally, it holds $\langle |\xi| \rangle \asymp 1 + |\xi|$ and $|[\xi]_d| + |\xi|_{[d-1]} \asymp |\xi|$. Collecting all of these estimates, one obtains

$$\left|(\partial^\rho \hat{g}_\lambda)(A^{-1}_{\alpha,s_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}\xi)\right| \lesssim \frac{\min\left\{1, s_\lambda^{-1} + |[A^{-1}_{\alpha,s_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}\xi]_d| + s_\lambda^{-(1-\alpha)}|A^{-1}_{\alpha,s_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}\xi|_{[d-1]}\right\}^M}{\langle|A^{-1}_{\alpha,s_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}\xi|\rangle^{N_1}\langle|A^{-1}_{\alpha,s_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}\xi|_{[d-1]}\rangle^{N_2}}$$

$$\lesssim \frac{\min\left\{1, s_\lambda^{-1}(1+|\xi|)\right\}^M}{\left(1+s_\lambda^{-1}|\xi|\right)^{N_1}\left(1+s_\lambda^{-\alpha}|R_{\theta_\lambda} R_{\varphi_\lambda}\xi|_{[d-1]}\right)^{N_2}}.$$

$\square$

The expression on the right hand side of (42) can further be estimated by the function

$$S_{\lambda,M,N_1,N_2}(\xi) := \frac{\min\left\{1, s_\lambda^{-1}(1+|\xi|)\right\}^M}{(1+s_\lambda^{-1}|\xi|)^{N_1}(1+s_\lambda^{1-\alpha}|R_{\theta_\lambda} R_{\varphi_\lambda}(\xi/|\xi|)|_{[d-1]})^{N_2}}, \quad \xi \in \mathbb{R}^d. \tag{43}$$

As already discussed in [28], this function can be separated into angular and radial components, allowing us to treat these parts independently in the integration later. Since the notation in this article differs slightly from the one in [28], we choose to state the lemma once more here, but refer to said article for a proof.

Lemma 6.6. *([28, Lemma 6.4]) Assume that $s_\lambda \geq c > 0$ for all $\lambda \in \Lambda$. For every $M, N_1, N_2, K \in \mathbb{N}_0$ such that $K \leq N_2$ we have with respect to $\lambda \in \Lambda$ and $\xi \in \mathbb{R}^d$ the uniform estimate*

$$\frac{\min\left\{1, s_\lambda^{-1}(1+|\xi|)\right\}^M}{\left(1+s_\lambda^{-1}|\xi|\right)^{N_1}\left(1+s_\lambda^{-\alpha}|R_{\theta_\lambda} R_{\varphi_\lambda}\xi|_{[d-1]}\right)^{N_2}} \lesssim S_{\lambda,M-K,N_1,K}(\xi).$$

Next, we want to estimate the scalar product of two functions of the form (43). Before the actual result, Lemma 6.10, we need some preparation. This is the part of the proof of Theorem 2.5 which differs the most from the situation in two dimensions.

Lemma 6.7. *Let $a \geq a' > 0$, $d \in \mathbb{N}$, $d \geq 2$, and $N > 1$. Then we have uniformly for $y \in \mathbb{R}$*

$$\int_\mathbb{R} \frac{|x|^{d-2}\,dx}{(1+a\,|x|)^{N+d-2}(1+a'\,|x-y|)^{N+d-2}} \lesssim a^{-(d-1)}(1+a'\,|y|)^{-N}.$$

Proof. Utilizing the following result from Grafakos [22][Appendix K.1]

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

we can estimate

$$\int_\mathbb{R} \frac{|x|^{d-2}\,dx}{(1+a\,|x|)^{N+d-2}(1+a'\,|x-y|)^{N+d-2}} = a^{-(d-2)}\int_\mathbb{R} \frac{|ax|^{d-2}\,dx}{(1+a\,|x|)^{N+d-2}(1+a'\,|x-y|)^{N+d-2}}$$

$$\leq a^{-(d-2)}\int_\mathbb{R} \frac{(1+|ax|)^{d-2}\,dx}{(1+a\,|x|)^{N+d-2}(1+a'\,|x-y|)^{N+d-2}} = a^{-(d-2)}\int_\mathbb{R} \frac{dx}{(1+a\,|x|)^N(1+a'\,|x-y|)^{N+d-2}}$$

$$\lesssim a^{-(d-2)}\max\{a,a'\}^{-1}(1+\min\{a,a'\}\,|y|)^{-N} = a^{-(d-1)}(1+a'\,|y|)^{-N}.$$

$\square$

We can immediately deduce the following corollary.

Corollary 6.8. *Let $a \geq a' > 0$, $d \in \mathbb{N}\backslash\{1\}$, and $N > 1$. Then we have uniformly for $\theta_0 \in \mathbb{R}$*

$$\int_0^\pi \frac{|\sin^{d-2}(\theta)|\,d\theta}{(1+a\,|\sin(\theta)|)^{N+d-2}(1+a'\,|\sin(\theta-\theta_0)|)^{N+d-2}} \lesssim a^{-(d-1)}\left(1+a'\,|\{\theta_0\}|\right)^{-N}.$$

Proof. Let us call the integral to be estimated $\mathcal{S}$. Since the integrand on the left hand side is π-periodic, we may change the domain of integration to $[-\pi/2, \pi/2]$. Applying Lemma 6.1, we can further conclude

$$\mathcal{S} \asymp \int\limits_{-\pi/2}^{\pi/2} \frac{|\theta|^{d-2}\, d\theta}{(1+a\,|\theta|)^{N+d-2}(1+a'\,|\{\theta-\theta_0\}|)^{N+d-2}},$$

Since $|\{\theta_0\}| \leq \frac{\pi}{2}$ we can estimate

$$\mathcal{S} \lesssim \sum_{\vartheta \in \{-\pi,0,\pi\}} \int_{\mathbb{R}} \frac{|\theta|^{d-2}\, d\theta}{(1+a\,|\theta|)^{N+d-2}(1+a'\,|\theta - (\{\theta_0\}+\vartheta)|)^{N+d-2}},$$

We now use Lemma 6.7 to estimate this by

$$\mathcal{S} \lesssim \sum_{\vartheta \in \{-\pi,0,\pi\}} a^{-(d-1)}(1+a'\,|\{\theta_0\}+\vartheta|)^{-N} \lesssim a^{-(d-1)}(1+a'\,|\{\theta_0\}|)^{-N}.$$

$\square$

This result is used to estimate the integral of the angular parts of (43) over the sphere $\mathbb{S}^{d-1}$.

Lemma 6.9. *Let $a, a' > 0$, $d \in \mathbb{N}$, $d \geq 2$, $\theta_\lambda, \theta_\mu \in [0,\pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]^{d-3}$, $\varphi_\lambda, \varphi_\mu \in [0, 2\pi]$ and $N > 1$. Further, let $d\sigma$ denote the standard surface measure on the sphere $\mathbb{S}^{d-1}$. We then have the estimate*

$$\int_{\mathbb{S}^{d-1}} \frac{d\sigma(\eta)}{(1+a|R_{\theta_\mu}R_{\varphi_\mu}\eta|_{[d-1]})^{N+d-2}(1+a'\,|R_{\theta_\lambda}R_{\varphi_\lambda}\eta|_{[d-1]})^{N+d-2}}$$

$$\lesssim \max\{a, a'\}^{-(d-1)}\big(1+\min\{a,a'\}|\{d_\mathbb{S}(e_\lambda, e_\mu)\}|\big)^{-N},$$

where $e_\lambda = R_{\varphi_\lambda}^T R_{\theta_\lambda}^T e_d$ and $e_\mu = R_{\varphi_\mu}^T R_{\theta_\mu}^T e_d$.

Proof. Note the symmetry of the statement with respect to interchanging the entities a, a' and λ, μ. Without loss of generality we can therefore restrict to the case $a \geq a' > 0$.

Since the mapping $R_{\theta_\mu} R_{\varphi_\mu}$ is an isometry, the integral is equal to

$$\mathcal{S} := \int_{\mathbb{S}^{d-1}} \frac{d\sigma(\eta)}{(1+a\,|\eta|_{[d-1]})^{N+d-2}(1+a'|R_{\theta_\lambda}R_{\varphi_\lambda}R_{\varphi_\mu}^T R_{\theta_\mu}^T \eta|_{[d-1]})^{N+d-2}}.$$

For the integration we parameterize the sphere $\mathbb{S}^{d-1}$ by standard spherical coordinates, i.e. coordinates $(\theta_1, \ldots \theta_{d-2}, \varphi) \in [0,\pi]^{d-2} \times [0, 2\pi)$ such that for $\eta \in \mathbb{S}^{d-1}$

$$\eta(\theta, \varphi) = \begin{pmatrix} \sin(\theta_1) \cdots \cdots \sin(\theta_{d-2})\cos(\varphi) \\ \sin(\theta_1) \cdots \cdots \sin(\theta_{d-2})\sin(\varphi) \\ \sin(\theta_1) \cdots \sin(\theta_{d-3})\cos(\theta_{d-2}) \\ \vdots \\ \cos(\theta_1) \end{pmatrix}.$$

Observe that $\langle \eta, e_d \rangle = \cos(\theta_1)$ and thus $\theta_1 = d_\mathbb{S}(\eta, e_d)$. Also note $|\eta|_{[d-1]} = |\sin(\theta_1)|$. Letting $\theta_0 := d_\mathbb{S}(e_\lambda, e_\mu) \in [0, \pi]$ denote the angle between e_λ and e_μ we have $\theta_0 = d_\mathbb{S}(e_d, R_{\theta_\lambda}R_{\varphi_\lambda}R_{\varphi_\mu}^T R_{\theta_\mu}^T e_d)$. Since $R_{\theta_\lambda}R_{\varphi_\lambda}R_{\varphi_\mu}^T R_{\theta_\mu}^T \in \mathcal{O}(d, \mathbb{R})$ we can apply Lemma 6.4 to estimate $|R_{\theta_\lambda}R_{\varphi_\lambda}R_{\varphi_\mu}^T R_{\theta_\mu}^T \eta|_{[d-1]}$. We obtain

$$\mathcal{S} \leq \int_0^{2\pi} \int_0^\pi \cdots \int_0^\pi \frac{\sin^{d-2}(\theta_1)\sin^{d-3}(\theta_2)\ldots \sin(\theta_{d-2})d\theta_1 d\theta_2 \ldots d\theta_{d-2}d\varphi}{(1+a|\sin(\theta_1)|)^{N+d-2}(1+a'\min\{|\sin(\theta_1+\theta_0)|, |\sin(\theta_1-\theta_0)|\})^{N+d-2}}$$

$$\lesssim \int_0^\pi \frac{\sin^{d-2}(\theta_1)\, d\theta_1}{(1+a|\sin(\theta_1)|)^{N+d-2}(1+a'\min\{|\sin(\theta_1+\theta_0)|, |\sin(\theta_1-\theta_0)|\})^{N+d-2}}$$

$$\leq \sum_{\epsilon \in \{-1,1\}} \int_0^\pi \frac{|\sin(\theta_1)|^{d-2}\, d\theta_1}{(1+a|\sin(\theta_1)|)^{N+d-2}(1+a'\,|\sin(\theta_1-\epsilon\theta_0)|)^{N+d-2}}.$$

Using Corollary 6.8 we finally arrive at $\mathcal{S} \lesssim \max\{a, a'\}^{-(d-1)}\big(1+\min\{a,a'\}\,|\{\theta_0\}|\big)^{-N}$. $\square$

29

With this estimate for the angular components in our toolbox, we proceed to prove the main result concerning the correlation of functions of the form (43).

Lemma 6.10. *Let $\alpha \in [0,1]$, $d \in \mathbb{N}$, $d \geq 2$, and $M, N_1, N_2 \in \mathbb{N}_0$. Further, let (Λ, Φ_Λ) and (Δ, Φ_Δ) be parametrizations with $(s_\lambda, e_\lambda, x_\lambda) = \Phi_\Lambda(\lambda)$ and $(s_\mu, e_\mu, x_\mu) = \Phi_\Delta(\mu)$ for $\lambda \in \Lambda$, $\mu \in \Delta$, such that $c \leq s_\lambda$ and $c \leq s_\mu$ for a fixed constant $c > 0$. Then for $A > 0$ and $B > 1$ satisfying*

$$N_1 > \frac{d}{2}, \quad M + d > N_1 \geq A + \frac{1 + \alpha(d-1)}{2}, \quad \text{and} \quad N_2 \geq B + d - 2$$

the following estimate holds true with an implicit constant independent of $\lambda \in \Lambda$ and $\mu \in \Delta$,

$$(s_\lambda s_\mu)^{-\frac{1+\alpha(d-1)}{2}} \int_{\mathbb{R}^d} S_{\lambda, M, N_1, N_2}(x) S_{\mu, M, N_1, N_2}(x)\, dx \lesssim \max\left\{\frac{s_\lambda}{s_\mu}, \frac{s_\mu}{s_\lambda}\right\}^{-A} \left(1 + \min\{s_\lambda, s_\mu\}^{1-\alpha} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|\right)^{-B}.$$

Proof. Without loss of generality we subsequently assume $s_\lambda \leq s_\mu$. The strategy is to separate the integration into an angular and a radial part and estimate these independently. For the estimate of the angular part we can use Lemma 6.9, which yields

$$(s_\lambda s_\mu)^{-\frac{(1+\alpha(d-1))}{2}} \int_0^\infty \int_{\mathbb{S}^{d-1}} S_{\lambda, M, N_1, N_2}(\eta, r) S_{\mu, M, N_1, N_2}(\eta, r) r^{d-1}\, d\sigma(\eta) dr$$

$$\lesssim (s_\lambda s_\mu)^{-\frac{1+\alpha(d-1)}{2}} s_\mu^{-(1-\alpha)(d-1)} s_\mu^d \left(1 + s_\lambda^{1-\alpha} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|\right)^{-B} \cdot \mathcal{S}$$

with a remaining radial integral

$$\mathcal{S} := s_\mu^{-d} \int_0^\infty \frac{\min\left\{1, s_\lambda^{-1}(1+r)\right\}^M}{(1 + s_\lambda^{-1} r)^{N_1}} \frac{\min\left\{1, s_\mu^{-1}(1+r)\right\}^M}{(1 + s_\mu^{-1} r)^{N_1}} r^{d-1}\, dr.$$

Note that for the estimate we used the assumptions $s_\lambda \leq s_\mu$, $B > 1$ and $N_2 \geq B + d - 2$. It remains to verify the relation $(s_\mu s_\lambda)^{-(1+\alpha(d-1))/2} s_\mu^{-(1-\alpha)(d-1)} s_\mu^d \cdot \mathcal{S} \lesssim (s_\mu/s_\lambda)^{-A}$, or equivalently

$$\mathcal{S} \lesssim \left(\frac{s_\mu}{s_\lambda}\right)^{-A - \frac{1+\alpha(d-1)}{2}}.$$

To prove this, we split the integration of $\mathcal{S}$ into three parts $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ corresponding to the integration ranges $0 \leq r \leq 1$, $1 \leq r \leq s_\mu$, and $s_\mu \leq r$ respectively.

$\underline{0 \leq r \leq 1}$: Here we estimate $\min\left\{1, s_\lambda^{-1}(1+r)\right\}^M \leq s_\lambda^{-M}(1+r)^M \leq 2^M s_\lambda^{-M}$ and $(1 + s_\lambda^{-1} r)^{N_1} \geq 1$, and similarly for the index μ. Hence, the integral over this part can be estimated by

$$\mathcal{S}_1 \lesssim s_\mu^{-d} s_\lambda^{-M} s_\mu^{-M} \int_0^1 r^{d-1}\, dr \asymp s_\mu^{-(M+d)} s_\lambda^{-M} \lesssim \left(\frac{s_\mu}{s_\lambda}\right)^{-(M+d)},$$

where the last inequality holds because of the uniform lower bound $0 < c \leq s_\lambda$ for $\lambda \in \Lambda$. Finally observe that the assumed inequalities imply $M + d > A + \frac{1+\alpha(d-1)}{2}$.

$\underline{1 \leq r \leq s_\mu}$ We estimate the terms involving μ as follows: $(1 + s_\mu^{-1} r)^{N_1} \geq 1$ and $(r + 1) \leq 2r$. Hence

$$\min\left\{1, s_\mu^{-1}(1+r)\right\}^M \leq s_\mu^{-M}(1+r)^M \leq s_\mu^{-M}(r+r)^M \leq 2^M s_\mu^{-M} r^M.$$

For the terms with λ's, we have $(1 + s_\lambda^{-1} r)^{N_1} \geq s_\lambda^{-N_1} r^{N_1}$ and $\min\left\{1, s_\lambda^{-1}(1+r)\right\}^M \leq 1$. The integral $\mathcal{S}_2$ hence satisfies

$$\mathcal{S}_2 \lesssim s_\mu^{-d} s_\lambda^{N_1} s_\mu^{-M} \int_1^{s_\mu} r^{M - N_1 + d - 1}\, dr \lesssim s_\mu^{-(M+d)} s_\lambda^{N_1} s_\mu^{M+d-N_1} = \left(\frac{s_\mu}{s_\lambda}\right)^{-N_1},$$

where it was used that $M + d > N_1$, which implies $M + d - N_1 - 1 > -1$, for the integration. By assumption $N_1 \geq A + \frac{1+\alpha(d-1)}{2}$, giving the desired result.

30

$\underline{s_\mu \leq r}$ We estimate both terms like the λ-terms above to obtain

$$\mathcal{S}_3 \lesssim s_\mu^{-d} s_\lambda^{N_1} s_\mu^{N_1} \int_{s_\mu}^{\infty} r^{d-1-2N_1}\, dr \lesssim s_\mu^{N_1-d} s_\lambda^{N_1} s_\mu^{d-2N_1} \lesssim \left(\frac{s_\mu}{s_\lambda}\right)^{-N_1}.$$

The integral converges since $N_1 > \frac{d}{2}$. Since $N_1 \geq A + \frac{1+\alpha(d-1)}{2}$ the proof is finished. $\qquad\square$

6.3 Cancellation Estimate

Theorem 2.5 provides estimates for the scalar products of α-molecules. To derive them we evaluate these scalar products on the Fourier side, where we can take advantage of cancellation phenomena. Technically, the method is based on a clever integration by parts involving the following differential operator, depending on $\lambda \in \Lambda$, $\mu \in \Delta$,

$$\mathscr{L}_{\lambda,\mu} := \mathcal{I} - s_0^{2\alpha}\Delta - \frac{s_0^2}{1 + s_0^{2(1-\alpha)}\,|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2}\langle e_\lambda, \nabla\rangle^2, \tag{44}$$

where $s_0 = \min\{s_\lambda, s_\mu\}$, $\mathcal{I}$ is the identity operator, ∇ the gradient and Δ the standard Laplacian.

Lemma 6.11 shows how $\mathscr{L}_{\lambda,\mu}$ acts on products of functions a_λ, b_μ which satisfy (7).

Lemma 6.11. *Let a_λ and b_μ satisfy (7) for every multi-index $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L$ and assume $s_\lambda, s_\mu \geq c > 0$. Then we can write the expression*

$$\mathscr{L}_{\lambda,\mu}\left(a_\lambda(A_{\alpha,s_\lambda}^{-1} R_{\theta_\lambda} R_{\varphi_\lambda}\xi) b_\mu(A_{\alpha,s_\mu}^{-1} R_{\theta_\mu} R_{\varphi_\mu}\xi)\right)$$

as a finite linear combination of terms of the form

$$p_\lambda(A_{\alpha,s_\lambda}^{-1} R_{\theta_\lambda} R_{\varphi_\lambda}\xi) q_\mu(A_{\alpha,s_\mu}^{-1} R_{\theta_\mu} R_{\varphi_\mu}\xi)$$

with functions p_λ, q_μ, which satisfy (7) for all multi indices $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L - 2$.

Proof. For convenience we introduce the operators $O_\lambda := A_{\alpha,s_\lambda}^{-1} R_{\theta_\lambda} R_{\varphi_\lambda}$ and $O_\mu := A_{\alpha,s_\mu}^{-1} R_{\theta_\mu} R_{\varphi_\mu}$. Further, we define the functions $\widetilde{a}_\lambda(\xi) := a_\lambda(O_\lambda\xi)$ and $\widetilde{b}_\mu(\xi) := b_\mu(O_\mu\xi)$. We also abbreviate $\xi_\lambda := O_\lambda\xi$ and $\xi_\mu := O_\mu\xi$. Taking into account $s_\lambda \gtrsim 1$, we observe $\|O_\lambda\|_{2\to2} = \|A_{\alpha,s_\lambda}^{-1}\|_{2\to2} = \max\{s_\lambda^{-\alpha}, s_\lambda^{-1}\} \lesssim s_\lambda^{-\alpha}$. Analogously, it holds $\|O_\mu\|_{2\to2} = \|A_{\alpha,s_\mu}^{-1}\|_{2\to2} \lesssim s_\mu^{-\alpha}$. Finally, we introduce the 'transfer' matrix

$$T_{\lambda,\mu} := R_{\theta_\mu} R_{\varphi_\mu} R_{\varphi_\lambda}^T R_{\theta_\lambda}^T \in \mathcal{O}(d,\mathbb{R}). \tag{45}$$

After these remarks we turn to the proof, where we treat the components of $\mathscr{L}_{\lambda,\mu}$ separately.

$\underline{\mathcal{I}}$ This term causes no pain.

$\underline{s_0^{2\alpha}\Delta}$ By the product rule we have

$$\Delta(\widetilde{a}_\lambda \widetilde{b}_\mu) = \underbrace{2\langle\nabla\widetilde{a}_\lambda, \nabla\widetilde{b}_\mu\rangle}_{\mathbf{A}} + \underbrace{\widetilde{a}_\lambda\Delta\widetilde{b}_\mu + \widetilde{b}_\mu\Delta\widetilde{a}_\lambda}_{\mathbf{B}}.$$

In the following we first treat part **A** and then part **B**.

A The chain rule yields $\nabla\widetilde{a}_\lambda(\xi) = O_\lambda^T \nabla a_\lambda(\xi_\lambda)$ for every $\xi \in \mathbb{R}^d$ and an analogous formula for $\widetilde{b}_\mu$. Thus we obtain

$$\langle\nabla\widetilde{a}_\lambda(\xi), \nabla\widetilde{b}_\mu(\xi)\rangle = \langle O_\lambda^T \nabla a_\lambda(\xi_\lambda), O_\mu^T \nabla b_\mu(\xi_\mu)\rangle = \langle O_\mu O_\lambda^T \nabla a_\lambda(\xi_\lambda), \nabla b_\mu(\xi_\mu)\rangle.$$

The expression $\langle O_\mu O_\lambda^T \nabla a_\lambda, \nabla b_\mu\rangle$ is a linear combination of the products $\partial_i a_\lambda \partial_j b_\mu$, where $i,j \in \{1,\ldots,d\}$, with the entries of the matrix $O_\mu O_\lambda^T$ as coefficients. The functions $\partial_i a_\lambda$ and $\partial_j b_\mu$ clearly satisfy (7) for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L - 1$. Moreover, the entries of the matrix $O_\mu O_\lambda^T$ are bounded in modulus by $\|O_\mu O_\lambda^T\|_{2\to2}$, which in turn obeys the estimate

$$\|O_\mu O_\lambda^T\|_{2\to2} = \|A_{\alpha,s_\mu}^{-1} T_{\lambda,\mu} A_{\alpha,s_\lambda}^{-1}\|_{2\to2} \leq \|A_{\alpha,s_\mu}^{-1}\|_{2\to2}\|A_{\alpha,s_\lambda}^{-1}\|_{2\to2} \lesssim (s_\mu s_\lambda)^{-\alpha} \leq s_0^{-2\alpha},$$

where $s_0 = \min\{s_\lambda, s_\mu\}$. This shows that the function $s_0^{2\alpha}\mathbf{A}$ can be written as claimed.

B   Due to symmetry it suffices to treat the term $\widetilde{b}_\mu \Delta \widetilde{a}_\lambda$. Since $\widetilde{b}_\mu(\xi) = b_\mu(\xi_\mu)$ for $\xi \in \mathbb{R}^d$ and since b_μ fulfills condition (7) for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L$, the function b_μ is a suitable first factor with the required properties. Let us investigate the second factor $\Delta \widetilde{a}_\lambda$.

The second derivative of $\widetilde{a}_\lambda$ is at each $\xi \in \mathbb{R}^d$ a bilinear mapping $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, which by the chain rule satisfies for $v, w \in \mathbb{R}^d$

$$\widetilde{a}_\lambda''(\xi)[v, w] = a_\lambda''(\xi_\lambda)[O_\lambda v, O_\lambda w].$$

Thus, we have the expansion

$$\Delta \widetilde{a}_\lambda(\xi) = \sum_{i=1}^d \widetilde{a}_\lambda''(\xi)[e_i, e_i] = \sum_{i=1}^d a_\lambda''(\xi_\lambda)[O_\lambda e_i, O_\lambda e_i].$$

Let $\rho \in \mathbb{N}_0^d$ be a multi-index with $|\rho|_1 \leq L - 2$. Then the partial derivative with respect to ρ of the function $\xi \mapsto \sum_{i=1}^d s_0^{2\alpha} a_\lambda''(\xi)[O_\lambda e_i, O_\lambda e_i]$ clearly exists. It remains to prove the frequency localization (7). In view of $\partial^\rho(a_\lambda'') = (\partial^\rho a_\lambda)''$ we can estimate for every $i \in \{1, \ldots, d\}$ and every $\xi \in \mathbb{R}^d$

$$s_0^{2\alpha} |\partial^\rho a_\lambda''(\xi)[O_\lambda e_i, O_\lambda e_i]| \leq s_0^{2\alpha} \|\!|\partial^\rho a_\lambda''(\xi)|\!\| \|O_\lambda\|_{2 \to 2}^2 \lesssim \|\!|\partial^\rho a_\lambda''(\xi)|\!\|.$$

The norm of the bilinear mapping is given by $\|\!|\partial^\rho a_\lambda''(\xi)|\!\| = \sup_{|v|, |w|=1} |\partial^\rho a_\lambda''(\xi)[v, w]|$. This is equal to the spectral norm of the corresponding Hesse matrix. Therefore we can deduce $\|\!|\partial^\rho a_\lambda''(\xi)|\!\| \lesssim \sup_{|\beta|_1=2} |\partial^\beta \partial^\rho a_\lambda(\xi)|$. The functions $\partial^\beta \partial^\rho a_\lambda$ satisfy (7) for every $\beta \in \mathbb{N}_0^d$ with $|\beta|_1 = 2$ due to the assumption on a_λ. The required frequency localization follows.

$\dfrac{s_0^2(1 + s_0^{2(1-\alpha)} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2)^{-1} \langle e_\lambda, \nabla \rangle^2}{s_0^{1+\alpha} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^{-1}}$ First we put $w_1 := s_0^2$, $w_2 := s_0^{2\alpha} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^{-2}$, and $w_3 := $ and notice that the pre-factor satisfies

$$s_0^2(1 + s_0^{2(1-\alpha)} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2)^{-1} \leq \min\{w_1, w_2, w_3\}. \tag{46}$$

The first two estimates are obvious. For the third, recall that $1 + t^2 \geq 2t$ for all $t \in \mathbb{R}$. Hence,

$$s_0^2(1 + s_0^{2(1-\alpha)} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2)^{-1} \leq \tfrac{1}{2} s_0^2(s_0^{1-\alpha} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|)^{-1} \leq s_0^{1+\alpha} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^{-1}.$$

We begin with the product rule, which yields

$$\langle e_\lambda, \nabla \rangle^2 \big(\widetilde{a}_\lambda \widetilde{b}_\mu\big) = \widetilde{b}_\mu \langle e_\lambda, \nabla \rangle^2 \widetilde{a}_\lambda + 2(\langle e_\lambda, \nabla \rangle \widetilde{a}_\lambda)(\langle e_\lambda, \nabla \rangle \widetilde{b}_\mu) + \widetilde{a}_\lambda \langle e_\lambda, \nabla \rangle^2 \widetilde{b}_\mu. \tag{47}$$

Recall that $e_\lambda = R_{\varphi_\lambda}^T R_{\theta_\lambda}^T e_d$. We calculate with the chain rule for $\xi \in \mathbb{R}^d$

$$\langle e_\lambda, \nabla \widetilde{a}_\lambda(\xi) \rangle = \langle O_\lambda e_\lambda, \nabla a_\lambda(\xi_\lambda) \rangle = \big\langle A_{\alpha, s_\lambda}^{-1} e_d, \nabla a_\lambda(\xi_\lambda) \big\rangle = s_\lambda^{-1} \partial_d a_\lambda(\xi_\lambda),$$

where we used $O_\lambda e_\lambda = A_{\alpha, s_\lambda}^{-1} e_d$. Using the 'transfer' matrix $T_{\lambda, \mu}$ from (45), we similarly obtain

$$\langle e_\lambda, \nabla \widetilde{b}_\mu(\xi) \rangle = \langle O_\mu e_\lambda, \nabla b_\mu(\xi_\mu) \rangle = \big\langle A_{\alpha, s_\mu}^{-1} T_{\lambda, \mu} e_d, \nabla b_\mu(\xi_\mu) \big\rangle.$$

Next, we note that $\langle e_\lambda, \nabla \rangle^2 \widetilde{a}_\lambda(\xi) = \widetilde{a}_\lambda''(\xi)[e_\lambda, e_\lambda]$. Together with the chain rule, this implies

$$\langle e_\lambda, \nabla \rangle^2 \widetilde{a}_\lambda(\xi) = a_\lambda''(\xi_\lambda)[O_\lambda e_\lambda, O_\lambda e_\lambda] = a_\lambda''(\xi_\lambda)[A_{\alpha, s_\lambda}^{-1} e_d, A_{\alpha, s_\lambda}^{-1} e_d] = s_\lambda^{-2} \partial_d^2 a_\lambda(\xi_\lambda).$$

We also obtain

$$\langle e_\lambda, \nabla \rangle^2 \widetilde{b}_\mu(\xi) = b_\mu''(\xi_\mu)[O_\mu e_\lambda, O_\mu e_\lambda] = \big(\langle A_{\alpha, s_\mu}^{-1} T_{\lambda, \mu} e_d, \nabla \rangle^2 b_\mu\big)(\xi_\mu).$$

Let us henceforth use the abbreviation $\eta := T_{\lambda, \mu} e_d \in \mathbb{S}^{d-1}$. Plugging the above calculations into (47) leads to the following expression for $\langle e_\lambda, \nabla \rangle^2 \big(\widetilde{a}_\lambda \widetilde{b}_\mu\big)(\xi)$ at $\xi \in \mathbb{R}^d$

$$s_\lambda^{-2} b_\mu(\xi_\mu) \cdot \partial_d^2 a_\lambda(\xi_\lambda) + 2 s_\lambda^{-1} \partial_d a_\lambda(\xi_\lambda) \cdot \langle A_{\alpha, s_\mu}^{-1} \eta, \nabla b_\mu(\xi_\mu) \rangle + a_\lambda(\xi_\lambda) \cdot \big(\langle A_{\alpha, s_\mu}^{-1} \eta, \nabla \rangle^2 b_\mu\big)(\xi_\mu). \tag{48}$$

For the first summand of (48) we consider the product of the functions $s_\lambda^{-2}\partial_d^2 a_\lambda$ and b_μ. Since $s_\lambda^{-2} \leq s_0^{-2}$ and in view of (46) the pre-factor w_1 is compensated. Due to the assumptions on a_λ and b_μ the product is thus of the desired form.

Let us put $\eta_{[d-1]} := (\eta_1, \ldots, \eta_{d-1}, 0)^T \in \mathbb{R}^d$ and $\eta_{[d]} := (0, \ldots, 0, \eta_d)^T \in \mathbb{R}^d$ and observe that

$$A_{\alpha,s_\mu}^{-1}\eta = A_{\alpha,s_\mu}^{-1}(\eta_{[d-1]} + \eta_{[d]}) = s_\mu^{-\alpha}\eta_{[d-1]} + s_\mu^{-1}\eta_{[d]}.$$

The second summand of (48) then becomes – up to the factor 2 –

$$\partial_d a_\lambda(\xi_\lambda) \cdot \left(s_\lambda^{-1}s_\mu^{-\alpha}\langle\eta_{[d-1]}, \nabla b_\mu(\xi_\mu)\rangle + s_\lambda^{-1}s_\mu^{-1}\eta_d\partial_d b_\mu(\xi_\mu)\right).$$

We choose the function $\partial_d a_\lambda$ as the first factor, which clearly has the required properties, and the function

$$\xi \mapsto s_\lambda^{-1}s_\mu^{-\alpha}\langle\eta_{[d-1]}, \nabla b_\mu(\xi)\rangle + s_\lambda^{-1}s_\mu^{-1}\eta_d\partial_d b_\mu(\xi).$$

as the second factor. The second component of this function causes no problems because $|\eta_d| \leq 1$ and the pre-factor w_1 is compensated due to $(s_\lambda s_\mu)^{-1} \leq s_0^{-2}$. To deal with the other term, notice that by Lemma 6.2 $|\eta_{[d-1]}| = |\eta|_{[d-1]} \asymp |\{d_\mathbb{S}(e_\lambda, e_\mu)\}|$. Thus

$$s_\lambda^{-1}s_\mu^{-\alpha}|\langle\eta_{[d-1]}, \nabla b_\mu\rangle| \lesssim s_\lambda^{-1}s_\mu^{-\alpha}|\{d_\mathbb{S}(e_\lambda, e_\mu)\}|\,|\nabla b_\mu|.$$

The fact that $\partial_i b_\mu$, $i \in \{1, \ldots, d\}$, satisfy (7) by assumption, and that $s_\lambda^{-1}s_\mu^{-\alpha}|\{d_\mathbb{S}(e_\lambda, e_\mu)\}|$ compensates w_3, implies that also the first component satisfies the required properties.

Let us turn to the last summand of (48). The first factor a_λ is of the desired form. For the second factor we expand the function $\langle A_{\alpha,s_\mu}^{-1}\eta, \nabla\rangle^2 b_\mu$ in the form

$$s_\mu^{-2\alpha}\langle\eta_{[d-1]}, \nabla\rangle^2 b_\mu + 2s_\mu^{-1-\alpha}\eta_d\langle\eta_{[d-1]}, \nabla\rangle\partial_d b_\mu + s_\mu^{-2}\eta_d^2\partial_d^2 b_\mu.$$

Its partial derivatives of order $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L - 2$ clearly exist, and we get the estimate

$$|\langle A_{\alpha,s_\mu}^{-1}\eta, \nabla\rangle^2\partial^\rho b_\mu| \lesssim s_0^{-2\alpha}|d_\mathbb{S}(e_\lambda, e_\mu)|^2\sum_{i,j=1}^{d-1}|\partial_i\partial_j\partial^\rho b_\mu| + 2s_0^{-1-\alpha}|d_\mathbb{S}(e_\lambda, e_\mu)||\nabla\partial_d\partial^\rho b_\mu| + s_0^{-2}|\partial_d^2\partial^\rho b_\mu|.$$

Here we again used that $|\eta_{[d-1]}| \asymp |\{d_\mathbb{S}(e_\lambda, e_\mu)\}|$ according to Lemma 6.2. This estimate completes the proof, taking into account the estimate (46) of the pre-factor and the fact that the partial derivatives of b_μ up to order L satisfy (7). $\qquad\square$

6.4   Proof of Theorem 2.5

At last we have all the tools available to prove Theorem 2.5. Write $\Delta x = x_\mu - x_\lambda$. An application of the Plancherel identity yields

$$\langle m_\lambda, p_\mu\rangle = \langle\hat{m}_\lambda, \hat{p}_\mu\rangle$$
$$= (s_\lambda s_\mu)^{-\frac{\alpha(d-1)+1}{2}}\int_{\mathbb{R}^d}\hat{a}_\lambda(A_{\alpha,s_\lambda}^{-1}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,s_\mu}^{-1}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\exp(2\pi i\langle\xi, \Delta x\rangle)\,d\xi$$

for two α-molecules m_λ and p_μ with respective generators a_λ and b_μ. According to Lemma 6.5 the Fourier transforms of the generators therefore satisfy (42) for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L$

Next, we want to exploit cancellation. For this we utilize the differential operator $\mathscr{L}_{\lambda,\mu}$ from (44). First, we observe that partial integration yields

$$\left\langle\mathscr{L}_{\lambda,\mu}^N\exp(2\pi i\langle\xi, \Delta x\rangle), \hat{a}_\lambda(A_{\alpha,s_\lambda}^{-1}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,s_\mu}^{-1}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\right\rangle$$
$$= \left\langle\exp(2\pi i\langle\xi, \Delta x\rangle), \mathscr{L}_{\lambda,\mu}^N\left(\hat{a}_\lambda(A_{\alpha,s_\lambda}^{-1}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,s_\mu}^{-1}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\right)\right\rangle,$$

since the boundary terms vanish due to the decay properties (7) of the generators and its derivatives. Note that we assume $N_1 > d/2$ and $L \geq 2N$. Second, we calculate for $\xi \in \mathbb{R}^d$

$$\mathscr{L}_{\lambda,\mu}^N\big(\exp(2\pi i\langle\xi,\Delta x\rangle)\big) = \left(1 + 4\pi^2 s_0^{2\alpha}|\Delta x|^2 + \frac{4\pi^2 s_0^2\langle e_\lambda,\Delta x\rangle^2}{1 + s_0^{2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2}\right)^N \exp(2\pi i\langle\xi,\Delta x\rangle).$$

Consequently, we have

$$\langle m_\lambda, p_\mu\rangle = \left(1 + 4\pi^2 s_0^{2\alpha}|\Delta x|^2 + \frac{4\pi^2 s_0^2\langle e_\lambda,\Delta x\rangle^2}{1 + s_0^{2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2}\right)^{-N} \cdot \mathcal{S}_{\lambda,\mu},$$

with

$$\mathcal{S}_{\lambda,\mu} := (s_\lambda s_\mu)^{-\frac{\alpha(d-1)+1}{2}} \int_{\mathbb{R}^d} \mathscr{L}_{\lambda,\mu}^N\big(\hat{a}_\lambda(A_{\alpha,s_\lambda}^{-1}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,s_\mu}^{-1}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\big)\exp(2\pi i\langle\xi,\Delta x\rangle)\,d\xi.$$

Since $L \geq 2N$ by assumption, Lemma 6.11 can iteratively be applied N times, and we conclude that

$$\mathscr{L}_{\lambda,\mu}^N\big(\hat{a}_\lambda(A_{\alpha,s_\lambda}^{-1}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,s_\mu}^{-1}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\big)$$

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

$$p_\lambda(A_{\alpha,s_\lambda}^{-1}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)q_\mu(A_{\alpha,s_\mu}^{-1}R_{\theta_\mu}R_{\varphi_\mu}\xi)),$$

where p_λ and q_μ satisfy (7) (for the multi-index just containing zeros).

Using Lemma 6.5 and putting $K = 2N + d - 2 \leq N_2$ in Lemma 6.6 then yields

$$\big|\mathscr{L}_{\lambda,\mu}^N\big(\hat{a}_\lambda(A_{\alpha,s_\lambda}^{-1}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,s_\mu}^{-1}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\big)\big|$$
$$\lesssim S_{\lambda,M-(2N+d-2),N_1,2N+d-2}(\xi)S_{\mu,M-(2N+d-2),N_1,2N+d-2}(\xi).$$

Due to the assumptions, we can further choose a number $\widetilde{N} \leq N_1$ which satisfies

$$(M - (2N + d - 2)) + d > \widetilde{N} \geq N + \frac{1 + \alpha(d-1)}{2}. \tag{49}$$

Since $\widetilde{N} \leq N_1$ we have the estimate $S_{\eta,M-(2N+d-2),N_1,2N+d-2} \leq S_{\eta,M-(2N+d-2),\widetilde{N},2N+d-2}$ for $\eta = \lambda, \mu$. Hence, we obtain

$$|\mathcal{S}_{\lambda,\mu}| \lesssim (s_\lambda s_\mu)^{-\frac{\alpha(d-1)+1}{2}} \int_{\mathbb{R}^d} S_{\lambda,M-(2N+d-2),N_1,2N+d-2}(\xi)S_{\mu,M-(2N+d-2),N_1,2N+d-2}(\xi)\,d\xi$$
$$\lesssim (s_\lambda s_\mu)^{-\frac{\alpha(d-1)+1}{2}} \int_{\mathbb{R}^d} S_{\lambda,M-(2N+d-2),\widetilde{N},2N+d-2}(\xi)S_{\mu,M-(2N+d-2),\widetilde{N},2N+d-2}(\xi)\,d\xi$$
$$\lesssim \max\left\{\frac{s_\lambda}{s_\mu},\frac{s_\mu}{s_\lambda}\right\}^{-N}(1 + s_0^{(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|)^{-2N}.$$

Here we used (49) and Lemma 6.10 in the last line (using this S and setting $\tilde{M} = M - (2N + d - 2)$, $A = N$ and $B = 2N$ ($B > 1$, $A > 0$ since $N > 1$)).

Altogether, we arrive at the desired estimate

$$|\langle m_\lambda, p_\mu\rangle| \lesssim \max\left\{\frac{s_\lambda}{s_\mu},\frac{s_\mu}{s_\lambda}\right\}^{-N}\left(1 + s_0^{2\alpha}|\Delta x|^2 + \frac{s_0^2\langle e_\lambda,\Delta x\rangle^2}{1 + s_0^{2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2}\right)^{-N}\left(1 + s_0^{(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|\right)^{-2N}$$

$$\lesssim \max\left\{\frac{s_\lambda}{s_\mu},\frac{s_\mu}{s_\lambda}\right\}^{-N}\left(1 + s_0^{2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2 + s_0^{2\alpha}|\Delta x|^2 + \frac{s_0^2\langle e_\lambda,\Delta x\rangle^2}{1 + s_0^{2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2}\right)^{-N}$$

$$\lesssim \omega_\alpha(\lambda,\mu)^{-N}.$$

For the last estimate observe that the inequality between the arithmetic and the geometric mean

$$\left(1 + s_0^{2(1-\alpha)}\left|\{d_\mathbb{S}(e_\lambda, e_\mu)\}\right|^2\right) + \frac{s_0^2\langle e_\lambda, \Delta x\rangle^2}{1 + s_0^{2(1-\alpha)}\left|\{d_\mathbb{S}(e_\lambda, e_\mu)\}\right|^2} \geq 2s_0|\langle e_\lambda, \Delta x\rangle|$$

implies

$$1 + s_0^{2(1-\alpha)}\left|\{d_\mathbb{S}(e_\lambda, e_\mu)\}\right|^2 + s_0^{2\alpha}\left|\Delta x\right|^2 + \frac{s_0^2\langle e_\lambda, \Delta x\rangle^2}{1 + s_0^{2(1-\alpha)}\left|\{d_\mathbb{S}(e_\lambda, e_\mu)\}\right|^2} \geq$$

$$\frac{1}{2}\left(1 + s_0^{2(1-\alpha)}\left|\{d_\mathbb{S}(e_\lambda, e_\mu)\}\right|^2 + s_0^{2\alpha}\left|\Delta x\right|^2\right) + \frac{1}{2}\left(1 + s_0^{2(1-\alpha)}\left|\{d_\mathbb{S}(e_\lambda, e_\mu)\}\right|^2 + \frac{s_0^2\langle e_\lambda, \Delta x\rangle^2}{1 + s_0^{2(1-\alpha)}\left|\{d_\mathbb{S}(e_\lambda, e_\mu)\}\right|^2}\right)$$

$$\gtrsim 1 + s_0^{2(1-\alpha)}\left|\{d_\mathbb{S}(e_\lambda, e_\mu)\}\right|^2 + s_0^{2\alpha}\left|\Delta x\right|^2 + s_0\left|\langle e_\lambda, \Delta x\rangle\right| = 1 + d_\alpha(\lambda, \mu).$$

This concludes the proof. $\qquad\qquad\square$

Acknowledgements

References

[1] J. Cai, B. Dong, S. Osher, and Z. Shen. Image restoration: total variation, wavelet frames, and beyond. *J. Amer. Math. Soc.*, 25(4):1033–1089, 2012.

[2] E. Candès. Ridgelets: Theory and applications. PhD thesis, Stanford University, 1998.

[3] E. Candès and D. L. Donoho. Ridgelets: a key to higher-dimensional intermittency? *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 357:2495–2509, 1999.

[4] E. J. Candès and L. Demanet. The curvelet representation of wave propagators is optimally sparse. *Comm. Pure Appl. Math.*, 58(11):1472–1528, 2002.

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

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

[7] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comp.*, 70:27–75, 2001.

[8] S. Dahlke, M. Fornasier, H. Rauhut, G. Steidl, and G. Teschke. Generalized coorbit theory, Banach frames, and the relation to alpha-modulation spaces. *Proc. London Math. Soc. (3)*, 96:464–506, 2008.

[9] S. Dahlke, G. Kutyniok, P. Maaß, C. Sagiv, and H.-G. Stark. The uncertainty principle associated with the continuous shearlet transform. *Int. J. Wavelets Multiresolut. Inf. Process.*, 6(2):157–181, 2008.

[10] S. Dahlke, G. Steidl, and G. Teschke. Coorbit spaces and Banach frames on homogeneous spaces with applications to the sphere. *Adv. Comput. Math.*, 21(1-2):147–180, 2004.

[11] S. Dahlke, G. Steidl, and G. Teschke. Weighted coorbit spaces and Banach frames on homogeneous spaces. *J. Fourier Anal. Appl.*, 10(5):507–539, 2004.

[12] W. Dahmen, C. Huang, G. Kutyniok, C. Schwab, and G. Welper. Efficient resolution of anisotropic structures. In *Extraction of Quantifiable Information from Complex Systems*, volume 102 of *Lecture Notes in Computational Science and Engineering*, pages 25–51. Springer, 2014.

[13] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

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

[15] D. L. Donoho. Sparse components of images and optimal atomic decomposition. *Constr. Approx.*, 17(3):353–382, 2001.

[16] G. Easley and D. Labate. Image processing using shearlets. In G. Kutyniok and D. Labate, editors, *Shearlets: Multiscale Analysis for Multivariate Data*, pages 283–320. Birkhäuser Boston, 2012.

[17] H. G. Feichtinger and K. Gröchenig. A unified approach to atomic decompositions via integrable group representations. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Mathematics*, pages 52–73. Springer, Berlin, 1988.

[18] H. G. Feichtinger and K. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions, I. *J. Funct. Anal.*, 86(2):307–340, 1989.

[19] H. G. Feichtinger and K. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions, II. *Monatsh. Math.*, 108(2-3):129–148, 1989.

[20] A. Flinth. 3D parabolic molecules. Bachelor's thesis, TU Berlin, 2013.

[21] M. Fornasier and H. Rauhut. Continuous frames, function spaces, and the discretization problem. *J. Fourier Anal. Appl.*, 11(3):245–287, 2005.

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

[23] P. Grohs. Ridgelet-type frame decompositions for Sobolev spaces related to linear transport. *J. Fourier Anal. Appl.*, 18(2):309–325, 2012.

[24] P. Grohs. Intrinsic localization of anisotropic frames. *Appl. Comput. Harmon. Anal.*, 35(2):264–283, 2013.

[25] P. Grohs, S. Keiper, G. Kutyniok, and M. Schäfer. Cartoon approximation with α-curvelets. *J. Fourier Anal. Appl.* To appear.

[26] P. Grohs, S. Keiper, G. Kutyniok, and M. Schäfer. α-Molecules: Curvelets, shearlets, ridgelets, and beyond. In *Wavelets and Sparsity XV(San Diego, CA), SPIE Proc.*, volume 8858. SPIE, 2013.

[27] P. Grohs and G. Kutyniok. Parabolic molecules. *Found. Comput. Math.*, 14(2):299–337, 2014.

[28] P. Grohs, S.Keiper, G. Kutyniok, and M. Schäfer. α-Molecules. 2014. Submitted. arXiv:1407.4424.

[29] P. Grohs and S. Vigogna. Intrinsic localization of anisotropic frames II: α-molecules. *J. Fourier Anal. Appl.*, 21(1):182–205, 2014.

[30] K. Guo, G. Kutyniok, and D. Labate. Sparse multidimensional representations using anisotropic dilation and shear operators. In *Wavelets and Splines (Athens, GA, 2005)*, pages 189–201. Nashboro Press, Nashville, TN, 2006.

[31] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM J. Math. Anal.*, 39(1):298–318, 2007.

[32] K. Guo and D. Labate. Representation of Fourier integral operators using shearlets. *J. Fourier Anal. Appl.*, 14(3):327–371, 2008.

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

[34] K. Guo and D. Labate. The construction of smooth Parseval frames of shearlets. *Math. Model. Nat. Phenom.*, 8(1):82–105, 2013.

[35] S. Keiper. A flexible shearlet transform - sparse approximation and dictionary learning. Bachelor's thesis, TU Berlin, 2013.

[36] H. Kempka, M. Schäfer, and T. Ullrich. General coorbit space theory for quasi-Banach spaces and inhomogeneous function spaces with variable smoothness and integrability. 2015. Preprint.

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

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

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

[40] G. Kutyniok, D. Labate, W. Q. Lim, and G. Weiss. Sparse multidimensional representation using shearlets. In *Wavelets XI(San Diego, CA), SPIE Proc.*, volume 5914, pages 254–262. SPIE, 2005.

[41] G. Kutyniok, J. Lemvig, and W. Q. Lim. Compactly supported shearlet frames and optimally sparse approximations of functions in $L^2(\mathbb{R}^3)$ with piecewise C^α singularities. *SIAM J. Math. Anal.*, 44:2962–3017, 2012.

[42] G. Kutyniok, J. Lemvig, and W. Q. Lim. Compactly supported shearlets. In M. Neamtu and L. Schumaker, editors, *Approximation Theory XIII: San Antonio 2010*, volume 13 of *Springer Proceedings in Mathematics*, pages 163–186. Springer New York, 2012.

[43] G. Kutyniok, V. Mehrmann, and P. Petersen. Regularization and numerical solution of the inverse scattering problem using shearlet frames. 2014. Submitted. arXiv:1407.7349.

[44] D. Labate, L. Mantovani, and P. Negi. Shearlet smoothness spaces. *J. Fourier Anal. Appl.*, 19(3):577–611, 2013.

[45] P. S. Negi and D. Labate. 3D discrete shearlet transform and video processing. *IEEE Trans. Image Process.*, 21(6):2944–2954, 2012.

[46] M. Schäfer. Generalized coorbit space theory for quasi-Banach spaces. Diplomarbeit, Rheinische Friedrich-Wilhelms-Universität Bonn, 2012.

[47] T. Ullrich and H. Rauhut. Generalized coorbit space theory and inhomogeneous function spaces of Besov-Lizorkin-Triebel type. *J. of Funct. Anal.*, 260(11):3299–3362, 2011.

[48] J. M. Wilson. On the atomic decomposition for Hardy spaces. *Pac. J. Math.*, 116(1):201–207, 1985.

[49] L. Ying, L. Demanet, and E. Candès. 3D discrete curvelet transform. In *Wavelets XI(San Diego, CA), SPIE Proc.*, volume 5914. SPIE, 2005.