OPTIMALLY SPARSE APPROXIMATIONS OF MULTIVARIATE FUNCTIONS USING COMPACTLY SUPPORTED SHEARLET FRAMES

Gitta Kutyniok[1*], *Jakob Lemvig*[1], *Wang-Q Lim*[1†]

[1]Institute of Mathematics, University of Osnabrück,
49069 Osnabrück, Germany
Emails: {kutyniok, jlemvig, wlim}@math.uni-osnabrueck.de

ABSTRACT

In this paper, we introduce pyramid-adapted shearlet systems for the three-dimensional setting, and show how one can construct frames for $L^2(\mathbb{R}^3)$ with this particular shearlet structure. We then introduce a generalized three-dimensional cartoon-like image model class of piecewise C^2 smooth functions with discontinuities on a C^α smooth surface with $1 < \alpha \leq 2$ and show that pyramid-adapted shearlet systems provide a nearly optimally sparse approximation error rate within this model class measured by means of non-linear, best n-term approximations.

Keywords— Anisotropic Features, Shearlets, Cartoon-like Images, Non-linear Approximations, Sparse Approximations

1. INTRODUCTION

Many important problem classes such as neuro, satellite, and seismic imaging, or partial differential equations with shock waves or boundary layers, are governed by anisotropic characteristics. Such anisotropic features could, e.g., be singularities concentrated on lower dimensional embedded manifolds or edges between image 'objects' and image 'background'. Advances in modern technology have pushed forward the need to efficiently handle enormous, multidimensional data with types of anisotropic characteristics. Over the last decade there has therefore been an intense study in developing efficient multivariate, directional representation systems. To analyze the ability of representation systems to reliably capture and sparsely represent anisotropic structures, Candés and Donoho [1] introduced the model situation of so-called cartoon-like images, i.e., two-dimensional functions which are piecewise C^2 smooth apart from a C^2 discontinuity curve. In recent years, it has been shown that curvelets, contourlets, and shearlets all have the ability to essentially optimal sparsely approximate cartoon-like images measured by the L^2-error of the (best) n-term approximation. Traditionally, this type of results has only been available for band-limited generators [1,3,5,6], but recently Kutyniok and Lim [8] showed that optimal sparsity also holds for compactly supported shearlet generators under weak moment conditions in dimension two.

*DFG Grant SPP-1324, KU 1446/13, DFG Grant SPP-1324, KU 1446/14
†DFG Grant SPP-1324, KU 1446/13

In the present paper we will consider sparse approximations of cartoon-like images using shearlets in dimension *three*. When passing from the two-dimensional setting to the three-dimensional setting, the complexity of anisotropic structures changes significantly. In 2D one 'only' has to handle one type of anisotropic features, namely curves, whereas in 3D one has to handle *two* geometrically very different anisotropic structures: Curves as one-dimensional features and surfaces as two-dimensional anisotropic features. Moreover, the analysis of sparse approximations in dimension two depends heavily on reducing the analysis to affine subspaces of $\mathbb{R}^2$. Clearly, these subspaces always have dimension and co-dimension one in 2D. In dimension three, however, we have subspaces of co-dimension one and two, and one therefore needs to perform the analysis on subspaces of the 'correct' co-dimension.

The generalized class of cartoon-like images in 3D considered in this paper consists of three-dimensional piecewise C^2 smooth functions with discontinuities on a C^α surface for $\alpha \in (1, 2]$. We will give the precise definition as well as the optimal rate of approximation with this model in Section 2. In Section 3 we construct so-called pyramid-adapted shearlet frames with compactly supported generators. Finally, in Section 4, we prove that such shearlet systems indeed deliver nearly optimal sparse approximations of three-dimensional cartoon-like images.

We mention that our model image class can be extended further to also contain three-dimensional cartoon-like images with a *piecewise* C^α smooth discontinuity surface. Allowing piecewise C^α smoothness instead of C^α smoothness everywhere is an essential way to model singularities along surfaces *as well as* along curves. The sparsity results presented in this paper can be extended to this generalized model class, but for brevity we will not do this here and simply refer to [7].

We remark that even though the present paper only deals with construction of shearlet frames for $L^2(\mathbb{R}^3)$ and sparse approximations of such, it also illustrates how many of the problems that arises when passing to higher dimensions can be handled. The reason for this observation is the fact that the 3D setting is the first dimension which exhibits (non-trivial, proper) subspaces and anisotropic features of different dimensions. Therefore the extension of the presented result in $L^2(\mathbb{R}^3)$ to higher dimensions $L^2(\mathbb{R}^n)$ should be, if not straightforward, then at least be achievable by the methodologies developed.

2. CARTOON-LIKE IMAGE CLASS

We start by defining the 3D cartoon-like image class. Fix $\mu, \nu > 0$. By $\mathcal{E}_2^2(\mathbb{R}^3)$ we denote the set of functions $f : \mathbb{R}^3 \to \mathbb{C}$ of the form

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

where $B \subset [0,1]^3$ with ∂B a closed C^2-surface for which the principal curvatures are bounded by ν and $f_i \in C^2(\mathbb{R}^3)$ with $\operatorname{supp} f_0 \subset [0,1]^3$ and $\|f_i\|_{C^2} \le \mu$ for each $i = 0, 1$. A very simple cartoon-like function $f = \chi_B$ with $\mu = 1$ and $\nu = 1/r$ is depicted in Fig. 1, where r is the radius of a ball B in $\mathbb{R}^3$. The requirement that the 'edge' ∂B is C^2 might be too restrictive in some applications, and we therefore enlarge the cartoon-like image model class to allow less regular images, where ∂B is C^α smooth for $1 < \alpha \le 2$, and not necessarily a C^2-surface. We speak of $\mathcal{E}_\alpha^2(\mathbb{R}^3)$ as consisting of cartoon-like 3D images having C^2 smoothness apart from a C^α discontinuity surface.

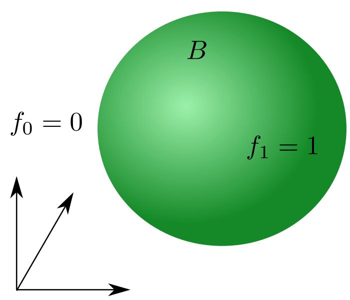

Fig. 1. A cartoon-like image $f = \chi_B$, where ∂B is a sphere.

In [7], it was shown using information theoretic arguments (cf. [4]) that the optimal approximation rate for such 3D cartoon-like image models $f \in \mathcal{E}_\alpha^2(\mathbb{R}^3)$ which can be achieved for almost any representation system is

$$\|f - f_n\|_2^2 = O(n^{-\alpha/2}), \quad n \to \infty,$$

where f_n is the best n-term approximation of f. This optimal rate can be used as a benchmark for measuring the performance of different representation systems as illustrated in the following example.

Example 1. For a simple cartoon-like image of the form $f = \chi_B$, where B is a ball contained in $[0,1]^3$, see Fig. 1, we clearly have $f \in \mathcal{E}_2^2(\mathbb{R}^3)$. By the later presented Thm. 4, the error rate of the n-term shearlet approximation f_n decays as $\|f - f_n\|_{L^2}^2 = O(n^{-1}(\log n)^2)$, which is optimal up a polylog factor. On the other hand, the corresponding error of the best n-term Fourier series approximation of f decays asymptotically as $n^{-1/3}$. This can be seen as follows. Let $I_n = \{k \in \mathbb{Z}^3 : \|k\|_2 \le n\}$ and let f_{I_n} be the partial Fourier sum with terms from I_n. Since the Fourier transform of f decays like $|\hat{f}(\xi)| \asymp \|\xi\|_2^{-2}$ as $\|\xi\|_2 \to \infty$, we have

$$\begin{aligned}\|f - f_{I_n}\|_{L^2}^2 &= \sum_{k \notin I_n} |c_k|^2 \asymp \int_{\|\xi\|_2 > n} \|\xi\|_2^{-4} \, d\xi \\ &= \int_n^\infty r^{-4} r^2 dr = n^{-1},\end{aligned}$$

where $h(n) \asymp g(n)$ means that h is bounded both above and below by g asymptotically. The conclusion now follows from the cardinality of $|I_n| \asymp n^3$ as $n \to \infty$. A best n-term approximation of f using a wavelet basis can be shown to perform slightly better with asymptotic behavior as $n^{-1/2}$. However, this is still far from the optimally achievable rate obtained by shearlet frames.

3. SHEARLETS

The *Pyramid-Adapted Shearlet Systems* are defined as follows. We partition the frequency space into three pairs of pyramids:

$$\begin{aligned}\mathcal{P} &= \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 : |\xi_1| \ge 1, \, |\xi_2/\xi_1| \le 1, \, |\xi_3/\xi_1| \le 1\}, \\ \tilde{\mathcal{P}} &= \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 : |\xi_2| \ge 1, \, |\xi_1/\xi_2| \le 1, \, |\xi_3/\xi_2| \le 1\}, \\ \breve{\mathcal{P}} &= \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 : |\xi_3| \ge 1, \, |\xi_1/\xi_3| \le 1, \, |\xi_2/\xi_3| \le 1\},\end{aligned}$$

and a centered cube:

$$\mathcal{C} = \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 : \|(\xi_1, \xi_2, \xi_3)\|_\infty < 1\}.$$

We first consider the shearlet system associated with the pyramid pair $\mathcal{P}$.

Let $\alpha \in (1, 2]$. We scale according to the *scaling matrix* A_{2^j}, $j \in \mathbb{Z}$, and represent directionality by the *shear matrix* S_k, $k = (k_1, k_2) \in \mathbb{Z}^2$, defined by

$$A_{2^j} = \begin{pmatrix} 2^{j\alpha/2} & 0 & 0 \\ 0 & 2^{j/2} & 0 \\ 0 & 0 & 2^{j/2} \end{pmatrix}, \qquad S_k = \begin{pmatrix} 1 & k_1 & k_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

respectively. The case $\alpha = 2$ corresponds to paraboloidal scaling. As α decreases, the scaling becomes less anisotropic, and allowing $\alpha = 1$ would yield isotropic scaling. The translation lattices will be generated by the matrix $M_c = \operatorname{diag}(c_1, c_2, c_2)$, where $c_1 > 0$ and $c_2 > 0$. The shearlet system associated with the pyramid $\mathcal{P}$ generated by $\psi \in L^2(\mathbb{R}^3)$ is then defined as

$$\begin{aligned}\Psi(\psi; c) = \{&2^{j(\alpha+2)/4} \psi(S_k A_{2^j} \cdot - m) : \\ &j \ge 0, |k_1|, |k_2| \le \lceil 2^{j(\alpha-1)/2} \rceil, m \in M_c \mathbb{Z}^3\}.\end{aligned}$$

The shearlet systems $\tilde{\Psi}(\tilde{\psi}; c)$ and $\breve{\Psi}(\breve{\psi}; c)$ associated with $\tilde{\mathcal{P}}$ and $\breve{\mathcal{P}}$, respectively, are defined in a similar manner (simply switch the role of the variables in the definitions of the scaling, shear, and translation matrix).

The partition of the frequency space into pyramids allows us to restrict the range of the shear parameters. Without such a partitioning as, e.g., in 'shearlet group' systems, one must

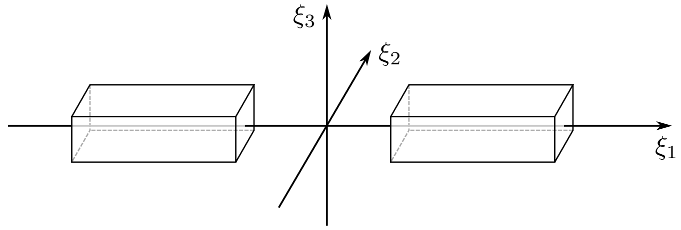

Fig. 2. The essential frequency concentration of the shearlet $\hat{\psi}$.

allow arbitrarily large shear parameters. For the 'pyramid-adapted' systems, we can, however, restrict the shear parameters to $[-\lceil 2^{(\alpha-1)j/2} \rceil, \lceil 2^{(\alpha-1)j/2} \rceil]$. This fact is pivotal for providing the shearlet system with a more uniform treatment of the directional features.

We are now ready to introduce our 3D shearlet system. For fixed $\alpha \in (1,2]$ and $c = (c_1, c_2) \in (\mathbb{R}_+)^2$, the *pyramid-adapted 3D shearlet system* $SH(\phi, \psi, \tilde{\psi}, \breve{\psi}; c, \alpha)$ generated by $\phi, \psi, \tilde{\psi}, \breve{\psi} \in L^2(\mathbb{R}^3)$ is defined by

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

where

$$\Phi(\phi; c_1) = \{\phi(\cdot - m) \,:\, m \in c_1\mathbb{Z}^3\}.$$

The functions $\psi, \tilde{\psi}, \breve{\psi} \in L^2(\mathbb{R}^3)$ are called shearlets, and the function ϕ is a scaling function associated the centered cube $\mathcal{C}$.

We are now ready to state the general sufficient conditions for the construction of shearlet frames.

Theorem 2 ([7]). *Let $\phi, \psi \in L^2(\mathbb{R}^3)$ be functions such that*

$$|\hat{\phi}(\xi)| \leq C_1 \min\{1, |\xi_1|^{-\gamma}\} \cdot \min\{1, |\xi_2|^{-\gamma}\} \cdot \min\{1, |\xi_3|^{-\gamma}\},$$

and

$$|\hat{\psi}(\xi)| \leq C_2 \cdot \min\{1, |\xi_1|^{\delta}\} \cdot \min\{1, |\xi_1|^{-\gamma}\} \cdot \min\{1, |\xi_2|^{-\gamma}\} \cdot \min\{1, |\xi_3|^{-\gamma}\},$$

for some constants $C_1, C_2 > 0$ and $\delta > 2\gamma > 6$. Define $\tilde{\psi}(x) = \psi(x_2, x_1, x_3)$ and $\breve{\psi}(x) = \psi(x_3, x_2, x_1)$ for $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. Then there exists the sampling constant $c_0 > 0$ such that the shearlet system $SH(\phi, \psi, \tilde{\psi}, \breve{\psi}; c)$ forms a frame for $L^2(\mathbb{R}^3)$ for all $c = (c_1, c_2)$ with $c_2 \leq c_1 \leq c_0$ provided that there exists a positive constant $M > 0$ such that

$$|\hat{\phi}(\xi)|^2 + \sum_{j\geq 0} \sum_{k_1, k_2 \in K_j} |\hat{\psi}(S_k^T A_{2^j}\xi)|^2 + |\hat{\tilde{\psi}}(\tilde{S}_k^T \tilde{A}_{2^j}\xi)|^2 + |\hat{\breve{\psi}}(\breve{S}_k^T \breve{A}_{2^j}\xi)|^2 > M$$

for a.e $\xi \in \mathbb{R}^3$, where $K_j := [-\lceil 2^{(\alpha-1)j/2} \rceil, \lceil 2^{(\alpha-1)j/2} \rceil]$.

Thm. 2 allows us to construct compactly supported shearlet frames generated by separable functions.The following example gives us explicitly a family of shearlets satisfying the assumptions of Thm. 2.

Example 3. Let $K, L \in \mathbb{N}$ be such that $L \geq 10$ and $\frac{3L}{2} \leq K \leq 3L - 2$, and define a shearlet $\psi \in L^2(\mathbb{R}^3)$ by

$$\hat{\psi}(\xi) = m_1(4\xi_1)\hat{\phi}(\xi_1)\hat{\phi}(2\xi_2)\hat{\phi}(2\xi_3), \quad \xi = (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3,$$

where the function m_0 is the low pass filter satisfying

$$|m_0(\xi_1)|^2 = (\cos(\pi\xi_1))^{2K} \sum_{n=0}^{L-1} \binom{K-1+n}{n} (\sin(\pi\xi_1))^{2n},$$

for $\xi_1 \in \mathbb{R}$, the function m_1 is the associated bandpass filter defined by

$$|m_1(\xi_1)|^2 = |m_0(\xi_1 + 1/2)|^2, \quad \xi_1 \in \mathbb{R},$$

and ϕ the scaling function is given by

$$\hat{\phi}(\xi_1) = \prod_{j=0}^{\infty} m_0(2^{-j}\xi_1), \quad \xi_1 \in \mathbb{R}.$$

In [7], it is shown that there exists a sampling constant $c_0 > 0$ such that the shearlet system $\Psi(\psi; c)$ forms a frame for $\check{L}^2(\mathcal{P}) := \{f \in L^2(\mathbb{R}^3) \,:\, \operatorname{supp} \hat{f} \subset P\}$ for any sampling matrix M_c with $c = (c_1, c_2) \in (\mathbb{R}_+)^2$ and $c_2 \leq c_1 \leq c_0$.

To obtain a frame for all of $L^2(\mathbb{R}^3)$ we simply set $\tilde{\psi}(x) = \psi(x_2, x_1, x_3)$ and $\breve{\psi}(x) = \psi(x_3, x_2, x_1)$ as in Thm. 2, and choose $\phi(x) = \phi(x_1)\phi(x_2)\phi(x_3)$ as scaling function for $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. Then the corresponding shearlet system $SH(\phi, \psi, \tilde{\psi}, \breve{\psi}; c, \alpha)$ forms a frame for $L^2(\mathbb{R}^3)$. We refer again to [7] for a proof of this result.

4. SPARSE APPROXIMATIONS

We now consider approximations of three-dimensional cartoon-like images using shearlet frames introduced in the previous section. For an orthonormal basis, the best n-term approximation is obtained by keeping the n largest coefficients in modulus. Since our shearlet system forms a non-tight frame, and, in particular, not a basis, we need to comment on the approximation procedure used to construct our n-term shearlet approximation. Suppose $SH(\phi, \psi, \tilde{\psi}, \breve{\psi}; c, \alpha)$ forms a frame for $L^2(\mathbb{R}^3)$ with frame bounds A and B. Since the shearlet system is a countable set of functions, we can write it in the form $(\sigma_i)_{i\in I}$ for some countable index set I. By basic frame theory [2], there exists a canonical dual frame $(\tilde{\sigma}_i)_{i\in I}$ of $(\sigma_i)_{i\in I}$ with frame bounds B^{-1} and A^{-1}. As our n-term approximation f_n of a cartoon-like image $f \in \mathcal{E}^2_\alpha(\mathbb{R}^3)$ by the frame $SH(\phi, \psi, \tilde{\psi}; c)$, we then take

$$f_n = \sum_{i\in I_n} \langle f, \sigma_i \rangle \tilde{\sigma}_i,$$

where $(\langle f, \sigma_i \rangle)_{i\in I_n}$ are the n largest coefficients $\langle f, \sigma_i \rangle$ in magnitude. Using the frame property this yields, for any positive integer n,

$$\|f - f_n\|_2^2 \leq \frac{1}{A} \sum_{i>n} |\theta(f)|_i^2,$$

where $|\theta(f)|_i$ denote the ith largest shearlet coefficient in absolute value. The approximation procedure does not always yield the best n-term approximation, but, surprisingly, even with this rather crude selection procedure, we can prove an nearly optimally sparse approximation rate for the class of generalized 3D cartoon-like images as the following result shows.

Theorem 4 ([7]). *Let $\alpha \in (1, 2]$, $c \in (\mathbb{R}_+)^2$, and let $\phi, \psi, \tilde{\psi}, \breve{\psi} \in L^2(\mathbb{R}^3)$ be compactly supported. Suppose that, for all $\xi = (\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3$, the function ψ satisfies:*

(i) $|\hat{\psi}(\xi)| \le C \cdot \min\{1, |\xi_1|^\delta\} \cdot \min\{1, |\xi_1|^{-\gamma}\} \cdot \min\{1, |\xi_2|^{-\gamma}\} \cdot \min\{1, |\xi_3|^{-\gamma}\}$,

(ii) $\left|\frac{\partial}{\partial \xi_i}\hat{\psi}(\xi)\right| \le |h(\xi_1)| \left(1 + \frac{|\xi_2|}{|\xi_1|}\right)^{-\gamma} \left(1 + \frac{|\xi_3|}{|\xi_1|}\right)^{-\gamma}$,

for $i = 1, 2$, where $\delta > 8$, $\gamma \ge 4$, $h \in L^1(\mathbb{R})$, and C a constant, and suppose that $\tilde{\psi}$ and $\breve{\psi}$ satisfy analogous conditions with the obvious change of coordinates. Further, suppose that the shearlet system $SH(\phi, \psi, \tilde{\psi}, \breve{\psi}; c, \alpha)$ forms a frame for $L^2(\mathbb{R}^3)$.

Then, for any $\nu > 0$ and $\mu > 0$, the shearlet frame $SH(\phi, \psi, \tilde{\psi}, \breve{\psi}; c, \alpha)$ provides nearly optimally sparse approximations of functions $f \in \mathcal{E}^2_\alpha(\mathbb{R}^3)$ in the sense that

$$\|f - f_n\|^2_{L^2} = \begin{cases} O(n^{-\alpha/2+\tau}), & \text{if } \alpha < 2, \\ O(n^{-1}(\log n)^2), & \text{if } \alpha = 2, \end{cases} \qquad (4.1)$$

as $n \to \infty$, where

$$\tau = \tau(\alpha) = \frac{3(2-\alpha)(\alpha-1)(\alpha+2)}{2(9\alpha^2 + 17\alpha - 10)},$$

and f_n is the n-term approximation obtained by choosing the n largest shearlet coefficients of f.

We remark that a large class of generators $\psi, \tilde{\psi}$, and $\breve{\psi}$ satisfy conditions (i) and (ii) in Thm. 4, e.g., the shearlet system from Example 3, when L and K is chosen sufficiently large.

Condition (i) in Thm. 4 can be interpreted as both a condition ensuring almost separable behavior and a control of the essential support support of the shearlets in frequency domain (see Fig. 2). Conditions (i) and (ii) together guarantee weak directional vanishing moments of the shearlets (see [3] for a precise definition), which is crucial for having fast decay of the shearlet coefficients when the corresponding shearlet intersects the discontinuity surface ∂B in a non-tangential way.

For $\alpha = 2$, we say that shearlets provide almost optimally sparse approximations since the error rate in (4.1) is only a polylog factor $(\log n)^2$ away from the optimal rate $O(n^{-1})$. However, for $\alpha \in (1, 2)$ we are a power of N with exponent τ away from the optimal rate. The exponent τ is small, and satisfies $0 < \tau(\alpha) < 0.04$ for $\alpha \in (1, 2)$ and $\tau(\alpha) \to 0$ for $\alpha \to 1+$ or $\alpha \to 2-$, see also Figure 3.

Let us mention that a slightly better estimate $\tau(\alpha)$ can be obtained satisfying $\tau(\alpha) < 0.037$ for $\alpha \in (1, 2)$; we can, however, with the current proof of Thm. 4 not make $\tau(\alpha)$ arbitrarily small.

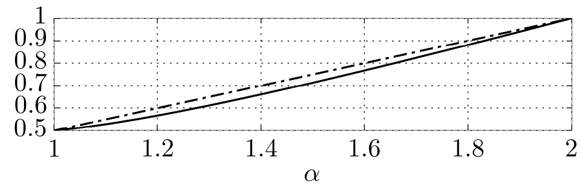

Fig. 3. Graph of $\frac{\alpha}{2} - \tau(\alpha)$ and the optimal rate $\frac{\alpha}{2}$ (dashed) as a function of α. The difference $\tau(\alpha)$ between the two graphs shows the optimality gap.

5. REFERENCES

[1] E. J. Candés and D. L. Donoho, *New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities*, Comm. Pure and Appl. Math. **56** (2004), 216–266.

[2] O. Christensen, *An introduction to frames and Riesz bases*, Birkhäuser Boston Inc., Boston, MA, 2003.

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

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

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

[6] K. Guo and D. Labate, *Optimally sparse representations of 3D data with C^2 surface singularities using Parseval frames of shearlets*, preprint.

[7] 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*, preprint.

[8] G. Kutyniok and W.-Q Lim, *Compactly supported shearlets are optimally sparse*, preprint.