

Tree approximation with anisotropic decompositions

P. Grohs

Research Report No. 2010-44
December 2010

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

Tree approximation with anisotropic decompositions

P. Grohs

December 6, 2010

Abstract

In recent years anisotropic transforms like the shearlet or curvelet transform have received a considerable amount of interest due to their ability to efficiently capture anisotropic features in terms of nonlinear N -term approximation. In this paper we study tree-approximation properties of such transforms where the N -term approximant has to satisfy the additional constraint that the set of kept indices possesses a tree structure. The main result of this paper is that for shearlet- and related systems, this additional constraint does not deteriorate the approximation rate. As an application of our results we construct (almost) optimal encoding schemes for cartoon images.

Contents

1	Introduction	1
1.1	Outline	2
1.2	Notation	2
2	Preliminaries	2
2.1	Cartoon Images	2
2.2	Shearlets	3
3	The Optimality Result	4
3.1	Optimal Tree Approximation for other Systems	8
4	Applications in Image Coding	11

1 Introduction

In many applications of mathematics one has to deal with piecewise smooth functions where the discontinuity arises along a smooth submanifold of the domain of definition. A particular case is given by bivariate functions which are smooth except for a smooth discontinuity curve. To give some examples of interest we mention that such functions have become widely recognized as a suitable model for image data and also arise as solutions to transport problems. It is therefore of eminent interest to come up with simple and accurate schemes to encode such data. Until a few years ago, only adaptive schemes have been available for this task, where adaptive means that one essentially has to track down the discontinuity curve and then adapt the approximation procedure to the curve [10, 16, 11, 9] (in [3] this is called the *Lagrangian Viewpoint*). In a remarkable work, in 2002 Candes and Donoho for the first time came up with a nonadaptive approximation procedure for bivariate functions which is very simple – it is defined by hard thresholding of the transform coefficients in a curvelet frame – and which possesses (almost) optimal convergence properties [4]. These results have been followed by other, similar constructions, most notably shearlets [25] and contourlets [14]. While hard thresholding of the transform coefficients gives optimal approximation rates in terms of the number of kept coefficients, there remains the question of an optimal scheme for transforming the list of kept coefficients into a sequence of bits, as would certainly be necessary for practical purposes. The key problem that may arise is that the storage cost of the indices of the kept coefficients might actually dominate the whole cost. For wavelet compression (or more general compression with orthobases), there exist several ways to remedy this problem, see for example [15, 7].

We would like to focus on [7] which is based on the idea of requiring that the set of kept indices possesses a tree structure. In this way, using the fact that trees can be encoded very efficiently, optimal bit-rate codes for unit balls of Besov spaces can be constructed with wavelets.

The main purpose of this paper is to show analogous results for the approximation of bivariate functions with smooth discontinuity curves with anisotropic transformations based on parabolic scaling, e.g. shearlets or curvelets.

Our main result, Theorem 4 is that the additional requirement of possessing a tree structure does not deteriorate the N -term approximation rate.

This result is of much broader interest than simply constructing optimal encoding schemes, since for implementational purposes it is often beneficial to store the index set as a tree.

1.1 Outline

We give an outline of this paper: Below, in Section 2 we collect various definitions and results that will be needed later on. For convenience we have decided to put a focus on the shearlet transform and therefore we explain the classical construction of a shearlet Parseval frame. We also introduce the tree structure that is inherently present in the shearlet index set. Section 3 contains our main optimality result for tree approximation. We first prove the result for the shearlet Parseval frame introduced in Section 2. Then we introduce a localization concept that allows us to transfer this result to other systems such as curvelets or different shearlet systems, compactly supported shearlets being one important example. Finally, in Section 4 we apply the results obtained in Section 3 and show how to construct a simple coding procedure which performs (almost) optimally in the sense of rate-distortion coding [1]. As an application we give a bound on the Kolmogorov entropy of the class \mathcal{F} of cartoon images defined in Section 2.

1.2 Notation

We will use the asymptotic notation $A \lesssim B$ to indicate that A is bounded by a uniform constant times B in magnitude. If $A \lesssim B$ and $B \lesssim A$ we write $A \sim B$. For a tempered distribution f we denote by \hat{f} its Fourier transform (the specific choice of normalization will not be relevant for us). The symbol $\lceil x \rceil$ denotes the smallest integer which is greater than x . We will use the symbol $|\cdot|$ in three instances: to denote the absolute value of a complex number, to denote the cardinality of a set and to denote the scale of a shearlet index (see below).

2 Preliminaries

2.1 Cartoon Images

For several years it has been popular to model image data as elements of (the unit ball of) the space of functions of bounded variation or Besov spaces. For these models wavelet methods can be shown to perform optimally in the task of encoding an image [12, 8]. However, this model does not fully pay tribute to the fact that an image is mostly defined by its edges, i.e. discontinuities along curves. Recently another model for so-called *cartoon images* has found a growing interest in the community. Following [4, 16] we introduce the class of functions we wish to approximate. Let $\text{STAR}^2(\nu)$ be the class of indicator functions χ_B of sets B with $B \subset [0, 1]^2$ and ∂B a C^2 -curve with curvature $\leq \nu$. More precisely $\text{STAR}^2(\nu)$ consists of indicator functions of sets B which are (modulo translation) of the form

$$B = \{x \in \mathbb{R}^2 : |x| \leq \rho(\varphi), x = (|x|, \varphi) \text{ in polar coordinates} \}$$

with

$$\sup_{\varphi} |\rho''(\varphi)| \leq \nu, \quad \sup_{\varphi} |\rho(\varphi)| < 1.$$

Then we define the set of *cartoon images* as

$$\mathcal{F}(\nu) := \{f = f_0 + f_1 \chi_B : \text{supp } f_i \subset [0, 1]^2, \chi_B \in \text{STAR}^2(\nu)$$

$$\text{and } \|f_0\|_{C^2}, \|f_1\|_{C^2} \leq 1\},$$

where we write

$$\|f\|_{C^2} := \sum_{|\alpha| \leq 2} \|D^\alpha f\|_\infty,$$

D^α denoting the partial derivative w.r.t. $\alpha \in \mathbb{R}^2$. Essentially this definition means that a function is in $\mathcal{F}(\nu)$ if it is smooth except for a C^2 discontinuity curve. Since the dependence on the parameter ν will not appear in our results, we will from now on simply write \mathcal{F} instead of $\mathcal{F}(\nu)$. This set of functions has served as a popular model for cartoon images for a while and therefore it is a crucial question how well one can approximate functions in \mathcal{F} . In the seminal paper [4], it has been shown that one can actually get (almost) optimal approximation performances for \mathcal{F} if one expands a function in terms of a curvelet frame and keeps only the largest coefficients. This stands in contrast to wavelet methods which can be shown to converge only at half the rate of curvelets. If one is willing to agree on the fact that \mathcal{F} is a more realistic model for images than for instance unit balls in Besov spaces, then this shows that curvelets are superior to wavelets for the encoding of images. Despite these theoretical results, there remain several issues regarding a simple and fast implementation of a curvelet transform. Indeed, since curvelets are defined by applying rotations to various basis functions, and since it is not clear how to translate this operation to a digital grid, the actual implementations of curvelet transforms are usually not fully faithful to the continuous theory. As a remedy to this problem shearlets have been introduced in [25]. There, the operation of rotation is replaced by a shearing operation which *can* be defined on a digital grid. Moreover, the desirable approximation properties of curvelets still remain valid for shearlets.

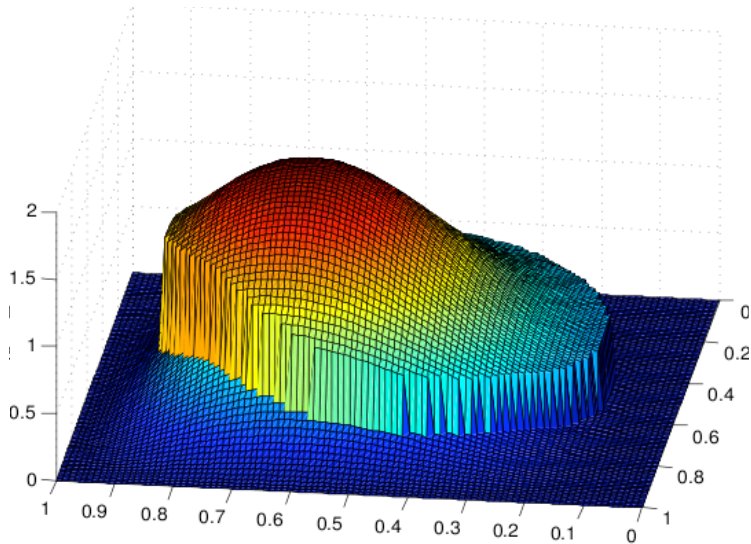


Figure 1: A typical element of \mathcal{F} .

2.2 Shearlets

The main goal of this paper is to show that \mathcal{F} can still be (almost) optimally approximated if one imposes the additional constraint on the kept indices to form a tree. This is highly desirable for deriving efficient coding procedures as well as certain implementational issues. We chose to present our main result for shearlet frames instead of curvelet frames for the following reasons:

- Shearlets are defined over a uniform grid which makes it much easier to define a suitable parent-child relation on the index set,
- there exist constructions of compactly supported shearlet frames, [23], a property that will turn out essential for constructing (almost) optimal coding schemes.

We now describe the main definitions and notation related to shearlets. First we need the concept of a *frame*, [6].

Definition 1. A system $\Psi = (\psi_\lambda)_{\lambda \in \Lambda}$ of elements ψ_λ in a Hilbert space \mathcal{H} , indexed by a countable index set Λ is called a frame if

$$\sum_{\lambda \in \Lambda} |\langle f, \psi_\lambda \rangle|^2 \sim \|f\|_2^2 \quad \text{for all } f \in \mathcal{H}. \quad (1)$$

If (1) holds with '=' instead of ' \sim ', then Ψ is called a Parseval frame. In this case we have the representation

$$f = \sum_{\lambda \in \Lambda} \langle f, \psi_\lambda \rangle \psi_\lambda \quad (2)$$

in \mathcal{H} .

In the following we will only be interested in frames of the Hilbert space $L_2(\mathbb{R}^2)$. Shearlets are built from a finite set of basis functions using the operations of translation, anisotropic dilation and shearing.

We follow [21] in defining a shearlet Parseval frame for $L_2(\mathbb{R}^2)$. Let $A_0 := \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$, $A_1 := \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, $B_0 := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $B_1 := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. In [21] it is shown that there exist functions $\varphi, \psi^{(0)}, \psi^{(1)}$ such that with

$$\sigma_{(j,l,k,d)} := 2^{3j/2} \psi^{(d)} \left(B_d^l A_d^j \cdot -k \right), \quad \sigma_k := \varphi(\cdot - k),$$

the system

$$\Sigma := \{\sigma_k : k \in \mathbb{Z}\} \cup \{\sigma_{(j,k,l,d)} : j \geq 0, -2^j \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\}$$

constitutes a Parseval frame for $L_2(\mathbb{R}^2)$ (see also [19, Theorem 2.1]). With

$$\Lambda_{-1} := \mathbb{Z}^2 \text{ and } \Lambda_j := \{(j, l, k, d) : -2^j \leq l \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\},$$

we define the shearlet index set $\Lambda = \dot{\bigcup}_{j \geq -1} \Lambda_j$ and get the representation

$$f = \sum_{\lambda \in \Lambda} \langle f, \sigma_\lambda \rangle \sigma_\lambda$$

in $L_2(\mathbb{R}^2)$. The shearlet index set Λ carries a natural tree structure which we will now describe. For an index $\lambda \in \Lambda$ we write $|\lambda|$ to denote the unique integer j with $\lambda \in \Lambda_j$. Further we write

$$\mathcal{E}_0 := \{(0, 0), (1, 0), (2, 0), (3, 0), (0, 1), (1, 1), (2, 1), (3, 1)\}$$

and

$$\mathcal{E}_1 := \{(0, 0), (0, 1), (0, 2), (0, 3), (1, 0), (1, 1), (1, 2), (1, 3)\}.$$

Definition 2. An index $(0, l, k, d) \in \Lambda_0$ is called a child of $m \in \Lambda_{-1}$ if $k = B_d^l m$. An index $(j, l, k, d) \in \Lambda_j$ is called a child of (j', l', k', d') if $d = d'$, $j = j' + 1$, $l \in \{2l', 2l' + 1\}$ and $k \in B_d^{[l/2] - l'} (A_d k' + \mathcal{E}_d)$ (see Figure 2). We can transitively extend this relation and write $\lambda' \preceq \lambda$ if either $\lambda = \lambda'$ or λ' is a child of λ .

Every $\lambda \in \Lambda_j$ possesses a unique parent in Λ_{j-1} , $j \geq 0$ and 16 children in Λ_{j+1} for $j \geq 0$ and 4 children for $j = -1$. We call a subset $\mathcal{T} \subset \Lambda$ a tree if for every $\lambda \in \mathcal{T}$ also its parent is contained in \mathcal{T} .

3 The Optimality Result

In this section we will prove our main result, namely that we can still get close-to-optimal N -term approximation performance if we only keep index sets forming a tree. The proofs utilize various concepts from nonlinear approximation [12] and wavelet tree approximation [7].

We define the approximation spaces

$$\Sigma_n^t := \left\{ \sum_{\lambda \in \mathcal{T}} c_\lambda \sigma_\lambda : \mathcal{T} \text{ is tree, and } |\mathcal{T}| \leq n \right\}.$$

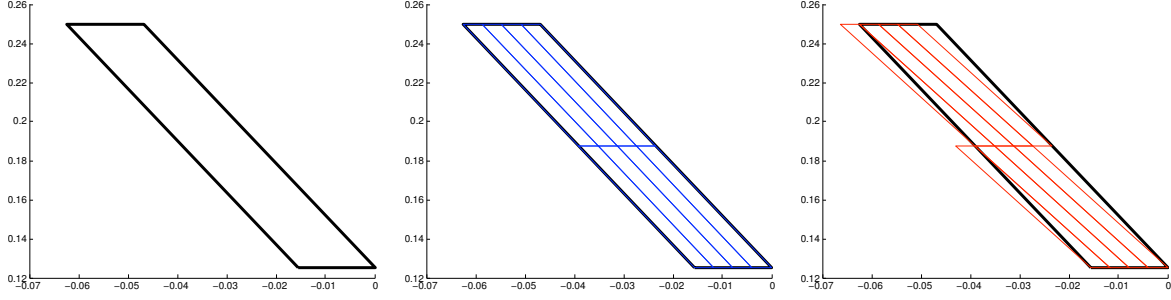


Figure 2: Left: Essential support of σ_λ with $j = 3, l = 3, d = 0, k = (2, 1)$. Middle: Essential support of its children with $l = 6$. Right: Essential support of its children with $l = 7$.

The main concern of this paper is to answer the following question:

What is the asymptotic rate of the error

$$t_n(f) := \inf_{g \in \Sigma_n^t} \|f - g\|_2,$$

where f ranges in \mathcal{F} ?

This question is of basic importance in nonlinear approximation. If we skip the requirement that the index set must have a tree structure, then there exists the following result:

Theorem 3 ([21]). *With*

$$\Sigma_n := \left\{ \sum_{\lambda \in \mathcal{I}} c_\lambda \sigma_\lambda : |\mathcal{I}| \leq n \right\}$$

we have for all $\varepsilon > 0$ the approximation

$$\sigma_n(f) := \inf_{g \in \Sigma_n} \|f - g\|_2 \lesssim n^{-1+\varepsilon},$$

where f ranges in \mathcal{F} . Moreover this convergence rate is optimal except for the (arbitrarily small) ε .

Theorem 3 has been inspired by the analogous result for curvelets in [4]. The goal of this section is to show the analogous statement for tree approximation with shearlets:

Theorem 4. *We have for all $\varepsilon > 0$ the approximation*

$$\sup_{f \in \mathcal{F}} t_n(f) := \sup_{f \in \mathcal{F}} \inf_{g \in \Sigma_n^t} \|f - g\|_2 \lesssim n^{-1+\varepsilon}.$$

In order to prove this theorem we must introduce some more notation. We let

$$\Lambda(f, \eta) := \{\lambda \in \Lambda : |\langle f, \sigma_\lambda \rangle| \geq \eta\}, \quad \Lambda_j(f, \eta) := \Lambda(f, \eta) \cap \Lambda_j.$$

Furthermore, we define $\mathcal{T}(f, \eta)$ to be the smallest tree containing $\Lambda(f, \eta)$ and $\mathcal{T}_j(f, \eta) := \mathcal{T}(f, \eta) \cap \Lambda_j$. Clearly, we have $\mathcal{T}(f, \eta) \subset \mathcal{T}(f, \eta')$ for $\eta \geq \eta'$.

Lemma 5. *Let $f \in \mathcal{F}$. Then for any $\varepsilon > 0$ there exists $\delta > 0$ such that we have the estimate*

$$\left\| (\langle f, \sigma_\lambda \rangle)_{\lambda \in \Lambda_j} \right\|_{2/3+\varepsilon} \lesssim 2^{-\delta j} \quad (3)$$

and hence

$$|\Lambda_j(f, \eta)| \lesssim 2^{-\delta j} \eta^{-(2/3+\varepsilon)} \quad \text{for all } j \geq 0. \quad (4)$$

Proof. Our starting point is a result of Guo and Labate in [21] where the unit square is partitioned into dyadic squares of sidelength $\sim 2^{-j}$ and f is localized onto each such square using a smooth partition of

unity. We denote by \mathcal{Q} the collection of all these squares tiling the unit square. Further, we denote the localization of f onto a dyadic square Q by f_Q and consider the coefficient sequence

$$\Sigma_Q := (\langle f_Q, \sigma_\lambda \rangle)_{\lambda \in \Lambda_j}.$$

There are two different types of elements in \mathcal{Q} : Those which intersect the singularity curve and those which do not. We call the collection of squares of the first type \mathcal{Q}^0 and the collection of squares of the latter type \mathcal{Q}^1 . We shall now use two key results that have been proven in [21]. We use the notion of weak l_p spaces (a.k.a. Lorentz spaces) which are defined on countable complex-valued sequences $\Theta = (\theta_n)_{n \in \mathbb{Z}}$ using the quasinorm

$$\|\Theta\|_{wl_p} := \sup_{N>0} N^{1/p} |\theta_N^*|,$$

where $\Theta^* = (\theta_N^*)_{N \in \mathbb{N}}$ denotes *decreasing rearrangement* of Θ , [13].

Lemma 6 ([21], Theorems 1.3, 1.4). *For $Q \in \mathcal{Q}^0$ we have*

$$\|\Sigma_Q\|_{wl_{2/3}} \lesssim 2^{-3j/2}.$$

For $Q \in \mathcal{Q}^1$ we have

$$\|\Sigma_Q\|_{2/3} \lesssim 2^{-3j}.$$

The implicit constants are independent of scale j .

Using this result we can prove

Corollary 7. *For all $\varepsilon > 0$ and $Q \in \mathcal{Q}^0$ we have*

$$\|\Sigma_Q\|_{2/3+\varepsilon} \lesssim 2^{-3j/2}.$$

For $Q \in \mathcal{Q}^1$ we have

$$\|\Sigma_Q\|_{2/3+\varepsilon} \lesssim 2^{-3j}.$$

The implicit constants are independent of scale j .

Proof. Let $Q \in \mathcal{Q}_0$ and denote by $\Sigma_Q^* = ((\gamma_Q^*)_N)_{N \in \mathbb{N}}$ the decreasing rearrangement of Σ_Q . By Theorem 6, we have

$$|(\gamma_Q^*)_N| \lesssim 2^{-3j/2} N^{-3/2}.$$

Therefore

$$\|\Sigma_Q\|_{2/3+\varepsilon}^{2/3+\varepsilon} = \sum_{n \in \mathbb{N}} |(\gamma_Q^*)_N|^{2/3+\varepsilon} \lesssim 2^{-(3j/2)(2/3+\varepsilon)} \sum_{n \in \mathbb{N}} N^{-1-3\varepsilon/2} \lesssim 2^{-(3j/2)(2/3+\varepsilon)}.$$

The case $Q \in \mathcal{Q}_1$ is the same. □

Now we use the p -triangle inequality for $p = 2/3 + \varepsilon$ with Corollary 7 and compute

$$\begin{aligned} \left\| (\langle f, \sigma_\lambda \rangle)_{\Lambda_j} \right\|_{2/3+\varepsilon}^{2/3+\varepsilon} &\leq \sum_{Q \in \mathcal{Q}^0} \|\Sigma_Q\|_{2/3+\varepsilon}^{2/3+\varepsilon} + \sum_{Q \in \mathcal{Q}^1} \|\Sigma_Q\|_{2/3+\varepsilon}^{2/3+\varepsilon} \\ &\lesssim 2^j 2^{-j-3j\varepsilon/2} + 4^j 2^{-2j-3\varepsilon} \lesssim 2^{-3j\varepsilon/2}. \end{aligned}$$

We have used the fact that $|\mathcal{Q}^0| \lesssim 2^j$ and $|\mathcal{Q}^1| \lesssim 4^j$. It is a well-known fact that for general sequences (θ_λ) we have the inequality

$$\sup_{\eta>0} |\{\lambda : |\theta_\lambda| \geq \eta\}| \eta^p \leq \|(\theta_\lambda)\|_p.$$

With $p = 2/3 + \varepsilon$ this implies that

$$\sup_{\eta>0} |\{\lambda \in \Lambda_j : |\langle f, \sigma_\lambda \rangle| \geq \eta\}| \eta^p \lesssim 2^{-\delta j},$$

where $\delta := \frac{3\varepsilon/2}{2/3+\varepsilon} > 0$. This proves the desired statement. □

Lemma 8. We have for any $\varepsilon > 0$

$$|\mathcal{T}(f, \eta)| \lesssim \eta^{-(2/3+\varepsilon)}. \quad (5)$$

Proof. We show the estimate

$$|\mathcal{T}_j(f, \eta)| \lesssim 2^{-\delta j} \eta^{-(2/3+\varepsilon)},$$

which implies the desired result. Every element in $\mathcal{T}_j(f, \eta)$ is either in $\Lambda_j(f, \eta)$ or it is the unique parent of some $\lambda' \in \Lambda_{j'}(f, \eta)$, $j' > j$. Therefore we have

$$|\mathcal{T}_j(f, \eta)| \leq \sum_{j' \geq j} |\Lambda_{j'}(f, \eta)|.$$

From (5) we know that for $f \in \mathcal{F}$ with some $\delta > 0$ depending only on ε we have the estimate

$$|\Lambda_{j'}(f, \eta)| \lesssim 2^{-\delta j'} \eta^{-(2/3+\varepsilon)}.$$

This implies the desired result. \square

Having a bound for the cardinality of $\mathcal{T}(f, \eta)$ we now approximate f by only keeping the indices in $\mathcal{T}(f, \eta)$.

Definition 9. Define the tree approximant

$$\mathcal{S}(f, \eta) := \sum_{\lambda \in \mathcal{T}(f, \eta)} \langle f, \sigma_\lambda \rangle \sigma_\lambda.$$

Lemma 10. For any $\varepsilon > 0$ we have the approximation rate

$$\|f - \mathcal{S}(f, \eta)\|_2 \lesssim \eta^{2/3-\varepsilon}, \quad (6)$$

uniformly over $f \in \mathcal{F}$.

Proof. Define

$$s_l := \sum_{\lambda \in \mathcal{T}(f, 2^{-(l+1)}\eta) \setminus \mathcal{T}(f, 2^{-l}\eta)} \langle f, \sigma_\lambda \rangle \sigma_\lambda.$$

Due to the frame property of Σ we can estimate for any $\varepsilon' > 0$

$$\begin{aligned} \|s_l\|_2 &\leq \left(\sum_{\lambda \in \mathcal{T}(f, 2^{-(l+1)}\eta) \setminus \mathcal{T}(f, 2^{-l}\eta)} |\langle f, \sigma_\lambda \rangle|^2 \right)^{1/2} \\ &\leq 2^{-l} \eta (|\mathcal{T}(f, 2^{-l-1}\eta)|)^{1/2} \lesssim 2^{-l(2/3-\varepsilon'/2)} \eta^{2/3-\varepsilon'/2}. \end{aligned}$$

We have used (5) in the last estimate. Now, since

$$\|f - \mathcal{S}(f, \eta)\|_2 \leq \sum_l \|s_l\| \lesssim \eta^{2/3-\varepsilon'/2}$$

we arrive at the desired result by setting $\varepsilon' := 2\varepsilon$. \square

We are ready to conclude the proof of our main result.

Proof of Theorem 4. Let $\eta^{2/3+\varepsilon/2} := n^{-1}$ with $\varepsilon > 0$ fixed. Then by (5) we have

$$|\mathcal{T}(f, \eta)| \lesssim \eta^{-(2/3+\varepsilon/2)} = n$$

and by (6) we have

$$\begin{aligned} \|f - \mathcal{S}(f, \eta)\|_2 &\lesssim \eta^{2/3-\varepsilon/6} = n^{-\frac{2/3-\varepsilon/6}{2/3+\varepsilon/2}} \\ &= n^{-1} n^{\frac{\varepsilon/2+\varepsilon/6}{2/3+\varepsilon/2}} \leq n^{-1} n^{\varepsilon \frac{2/3}{2/3}} = n^{-1+\varepsilon}. \end{aligned}$$

\square

3.1 Optimal Tree Approximation for other Systems

In the proof of our main theorem we have assumed that we are given a tight frame of bandlimited shearlets in order to make use of the results in [22]. Naturally, the question arises whether these assumptions are crucial. Actually, they are not. We can get the same approximation rate for tree approximation with any shearlet or curvelet system provided that the underlying basis functions are sufficiently smooth, sufficiently localized in space and possess sufficiently many anisotropic vanishing moments. The main idea is to build on Theorem 4 and to examine the cross Gramian matrix $(\langle \sigma_\lambda, \sigma'_{\lambda'} \rangle)$ between two shearlet frames Σ, Σ' , or even a curvelet frame Σ and a shearlet frame Σ' . Before we can state our main result we need some definitions.

Consider two hierachical index sets Λ', Λ'' , meaning that we have a disjoint union $\Lambda' = \dot{\bigcup}_{j \geq 0} \Lambda'_j$ and $\Lambda'' = \dot{\bigcup}_{j \geq 0} \Lambda''_j$. We can associate to each $\lambda' \in \Lambda'$ its scale $|\lambda'|$ which is the unique index j such that $\lambda' \in \Lambda'_j$. The same can be done with Λ'' . We now define a localization concept that will turn out to be useful for us.

Definition 11. *Two systems $\Sigma' := (\sigma'_{\lambda'})_{\lambda' \in \Lambda'}$, $\Sigma'' := (\sigma''_{\lambda''})_{\lambda'' \in \Lambda''}$ are called L -localized if*

$$|\langle \sigma'_{\lambda'}, \sigma''_{\lambda''} \rangle| \lesssim 4^{-L} |\lambda'|^{-|\lambda''|} \omega(\lambda', \lambda'')^{-L},$$

where ω is a distance function satisfying

$$\sup_{\lambda' \in \Lambda'_{j'}} \sum_{\lambda'' \in \Lambda''_{j''}} \omega(\lambda', \lambda'')^{-2} \lesssim 4^{2|j' - j''|} \quad (7)$$

with the implicit constant independent of j', j'' .

Lemma 12. *Assume that Σ' is L -localized with the shearlet frame Σ and $L > 21/6$. Then for any $f \in \mathcal{F}$ and for any $\varepsilon > 0$ there exists $\delta > 0$ (depending only on ε) such that*

$$\left\| \left(\langle f, \sigma'_{\lambda'} \rangle \right)_{\lambda' \in \Lambda'} \right\|_{2/3+\varepsilon} \lesssim 2^{-\delta j}. \quad (8)$$

Proof. Let $f \in \mathcal{F}$. Then by Lemma 5 we have with $c_\lambda := \langle f, 2^{\delta|\lambda|} \sigma_\lambda \rangle$ and $\delta > 0$ small

$$f = \sum_{\lambda \in \Lambda} c_\lambda 2^{-\delta|\lambda|} \sigma_\lambda, \quad (9)$$

where $\|(c_\lambda)_{\lambda \in \Lambda}\|_p < \infty$ and $p = 2/3 + \varepsilon$. We want to show that

$$\left\| \left(2^{\delta|\lambda'|} \langle f, \sigma_{\lambda'} \rangle \right)_{\lambda' \in \Lambda'} \right\|_p < \infty \quad (10)$$

which implies the desired claim. Clearly (10) follows if we can establish that the mapping $(c_\lambda)_{\lambda \in \Lambda} \mapsto \left(2^{\delta|\lambda'|} \langle f, \sigma_{\lambda'} \rangle \right)_{\lambda' \in \Lambda'}$ is bounded in l_p . The matrix of this mapping is given by

$$\left(\left\langle 2^{\delta|\lambda'|} \sigma'_{\lambda'}, 2^{-\delta|\lambda|} \sigma_\lambda \right\rangle \right)_{\lambda \in \Lambda, \lambda' \in \Lambda'}$$

and therefore in view of Schur's lemma we need to show that

$$\sup_{\lambda \in \Lambda} \sum_{\lambda' \in \Lambda'} \left| \left\langle 2^{-\delta|\lambda'|} \sigma'_{\lambda'}, 2^{\delta|\lambda|} \sigma_\lambda \right\rangle \right|^p < \infty. \quad (11)$$

Using the localization property of Σ' we estimate

$$\begin{aligned}
\sup_{\lambda \in \lambda} \sum_{\lambda' \in \Lambda'} \left| \left\langle 2^{-\delta|\lambda'|} \sigma'_{\lambda'}, 2^{\delta|\lambda|} \sigma_{\lambda} \right\rangle \right|^p &= \sup_{\lambda \in \lambda} \sum_{j \geq 0} \sum_{\lambda' \in \Lambda'_j} \left| \left\langle 2^{-\delta|\lambda'|} \sigma'_{\lambda'}, 2^{\delta|\lambda|} \sigma_{\lambda} \right\rangle \right|^p \\
&\leq \sup_{\lambda \in \lambda} \sum_{j \geq 0} 2^{p\delta|\lambda|-j} \sum_{\lambda' \in \Lambda'_j} |\langle \sigma'_{\lambda'}, \sigma_{\lambda} \rangle|^p \\
&\leq \sup_{\lambda \in \lambda} \sum_{j \geq 0} 2^{p\delta|\lambda|-j} \sum_{\lambda' \in \Lambda'_j} 4^{-Lp|\lambda|-j} \omega(\lambda, \lambda')^{-Lp} \\
&\leq \sup_{\lambda \in \lambda} \sum_{j \geq 0} 2^{p\delta|\lambda|-j} 4^{(2-L)p|\lambda|-j} \\
&= \sup_{\lambda \in \lambda} \sum_{j \geq 0} 4^{p(2+\delta/2-L)|\lambda|-j} < \infty,
\end{aligned}$$

whenever $L > 21/6$ and δ sufficiently small. \square

Remark 13. *It is certainly possible to require instead of (7) that*

$$\sup_{\lambda' \in \Lambda'_{j'}} \sum_{\lambda'' \in \Lambda''_{j''}} \omega(\lambda', \lambda'')^{-\alpha} \lesssim 4^{\beta|j'-j''|} \quad (12)$$

for some $\alpha, \beta > 0$. The conclusion of Lemma 12 would still hold, possibly with another constant than $21/6$. The reason why we chose $\alpha = \beta = 2$ is simply that for this choice, for many anisotropic frame decompositions, condition (7) can be verified.

We now assume that the system Σ' constitutes a frame for $L_2(\mathbb{R}^2)$, i.e.

$$\|f\|_2^2 \sim \sum_{\lambda' \in \Lambda'} |\langle f, \sigma'_{\lambda'} \rangle|^2$$

It is well known that in this case a dual frame $\tilde{\Sigma}' = (\tilde{\sigma}'_{\lambda'})_{\lambda' \in \Lambda'}$ exists with

$$f = \sum_{\lambda' \in \Lambda'} \langle f, \sigma'_{\lambda'} \rangle \tilde{\sigma}'_{\lambda'}.$$

Having a tree structure on Λ' we can define the set

$$\Sigma_n^{t'} := \left\{ \sum_{\lambda' \in \mathcal{T}} c_{\lambda'} \tilde{\sigma}'_{\lambda'} : \mathcal{T} \text{ is tree, and } |\mathcal{T}| \leq n \right\}$$

and consider the quantity

$$t'_n(f) := \inf_{g \in \Sigma_n^{t'}} \|f - g\|_2.$$

Theorem 14. *Assume that Σ' constitutes a frame for $L_2(\mathbb{R}^2)$ and that the index set Λ' possesses a tree structure. Assume moreover that Σ, Σ' are L -localized with $L > 21/6$. Then the conclusion of Theorem 4 holds with Σ replaced by Σ' , meaning that*

$$\sup_{f \in \mathcal{F}} t'_n(f) \lesssim n^{-1+\varepsilon} \quad (13)$$

for all $\varepsilon > 0$.

Proof. The proof goes by repeating the arguments leading to Theorem 4 and using the frame property of Σ' and (8). \square

The reason why Theorem 14 is interesting, is that a number of anisotropic systems are localized with Σ and therefore possess the same approximation rates.

Example 15. We give some examples of systems Σ' which are L -localized with Σ (without proof): Arbitrary systems of curvelet molecules of sufficient regularity are L -localized with Σ and $L > 21/6$, see [2] for the definition and [18] for other results in this direction. Another example is given by the tight frame Φ_J constructed in [5, Section 5.2].

In this paper we would like to focus on systems Σ' of *shearlet molecules* as defined in [22]:

Definition 16. A system $\Sigma' = ((m_\lambda)_{\lambda \in \Lambda})$ of functions is called a system of shearlet molecules of regularity R if we can write

$$m_\lambda(\cdot) = 2^{3j/2} a^{(\lambda)} \left(B_d^j A_d^l \cdot - \delta k \right), \quad \lambda = (j, l, k, d) \in \Lambda$$

with a sampling constant $\delta > 0 \in \mathbb{R}$ and functions $a^{(\lambda)}$ satisfying

$$\left| D^\mu a^{(\lambda)}(\cdot) \right| \lesssim (1 + |\cdot|)^{-N} \quad \text{for all } \mu \in \mathbb{N}^2, |\mu| \leq R, N \in \mathbb{N} \quad (14)$$

and

$$\left| \hat{a}^{(\lambda)}(\xi) \right| \lesssim (4^{-j} + |\xi_{1+d}|)^{-R} (1 + |\xi|)^{-R}, \quad \xi = (\xi_1, \xi_2) \in \mathbb{R}^2. \quad (15)$$

For a system Σ' with sampling constant δ we write

$$x_\lambda := A_d^{-j} B_d^{-l} \delta k. \quad (16)$$

We also write e_λ for the unit vector $(\cos(\theta_\lambda), \sin(\theta_\lambda))$, where $\theta_\lambda := \arctan(2^{-j}l)$. By examining the construction in [21] it is easy to see that the Parseval frame Σ is a system of shearlet molecules of arbitrary regularity. We now define a notion of distance between two indices of two (possibly different) systems of shearlet molecules. This definition follows [22] which in turn is based on [4, 27].

Definition 17. We define a distance between shearlet indices via

$$\omega(\lambda', \lambda'') := \left(1 + 4^{\min(\lambda', \lambda'')} d(\lambda', \lambda'') \right),$$

where

$$d(\lambda', \lambda'') := |2^{j'} l' - 2^{-j''} l''|^2 + |x_{\lambda'} - x_{\lambda''}|^2 + |\langle e_{\lambda'}, x_{\lambda'} - x_{\lambda''} \rangle|.$$

It is not difficult to see that this distance satisfies (7), as shown in [22] (see also [4]). With respect to this distance, any two systems of sufficient regularity are almost orthogonal as shown in [22, Theorem 1.3 and 1.4]:

Theorem 18. For any $L > 0$, there exists $R > 0$ such that any two systems Σ', Σ'' of shearlet molecules with regularity R satisfy

$$|\langle m'_{\lambda'}, m''_{\lambda''} \rangle| \lesssim 4^{-L} \|\lambda' - \lambda''\| \omega(\lambda', \lambda'')^{-L}. \quad (17)$$

In particular, Σ', Σ'' are L -localized.

In particular this implies by Theorem 14 that any system of shearlet molecules of sufficient regularity forming a frame satisfies the same tree approximation rate as the Parseval frame Σ .

Corollary 19. There exists $R_0 > 0$ such that for all systems Σ' of shearlet molecules of regularity $R > R_0$, which also form a frame for $L_2(\mathbb{R}^2)$, the conclusion of Theorem 4 is valid.

Of course it would be possible to make the dependence of R on L in Theorem 18 explicit and to compute R_0 , but that would be beyond the scope of this paper. Rather we would like to single out a particular system of shearlet molecules that will turn out useful in constructing encoding schemes, namely the construction given in [23]. The main idea is to do the same construction as for Σ but with compactly supported functions $\varphi', \psi'^{(0)}, \psi'^{(1)}$. By choosing these functions appropriately (sufficiently smooth, sufficiently many directional vanishing moments), with

$$\sigma'_{(j,l,k,d)} := 2^{3j/2} \psi'^{(d)} \left(B_d^l A_d^j \cdot - \delta k \right), \quad \sigma'_k := \varphi'(\cdot - \delta k),$$

the system

$$\Sigma' := \{\sigma'_k : k \in \mathbb{Z}\} \cup \left\{ \sigma'_{(j,k,l,d)} : j \geq 0, -2^j \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1 \right\} \quad (18)$$

constitutes a frame for $L_2(\mathbb{R}^2)$ for δ small enough, see [23]. Moreover (by appropriate choice of $\varphi', \psi^{(0)}, \psi^{(1)}$), the system Σ' is a system of shearlet molecules of regularity $> R_0$ and therefore, by Theorem 14, the conclusion of Theorem 4 is valid for Σ' :

Theorem 20. *There exist compactly supported functions $\varphi', \psi^{(0)}, \psi^{(1)}$ such that with Σ' as in (18) we have for any $\varepsilon > 0$*

$$\sup_{f \in \mathcal{F}} t'_n(f) \lesssim n^{-1+\varepsilon}.$$

We would like to remark that Theorem 20 could also be proven more directly by using the results in [24] instead of our results on localization.

4 Applications in Image Coding

The near-optimality of tree approximation leads to a near-optimal encoding strategy in the same way as in [7] for wavelets. An encoding scheme for \mathcal{F} consists of an *encoder* E which maps an $f \in \mathcal{F}$ to a bitstream $E(f)$, i.e. a sequence of zeros and ones. A *decoder* maps a bitstream onto a function $f \in L_2([0, 1]^2)$.

The *distortion* of the encoding/decoding pair (E, D) is defined as

$$d(E, D) := \sup_{f \in \mathcal{F}} \|f - D(E(f))\|_2. \quad (19)$$

For an encoder E we define its *runlength* as

$$M(E) := \sup_{f \in \mathcal{F}} |E(f)|,$$

where $|E(f)|$ denotes the length of the bitstream $E(f)$. A general encoding/decoding scheme for wavelets is constructed in [7]. The main property that is used is the fact that a general tree can be encoded much less expensively than an unstructured set of indices, *provided that the number of roots in the tree is uniformly bounded*, this is shown in [7, Lemma 6.1]. Therefore, in order directly apply the results and constructions of [7, Section 6] for constructing good shearlet coding procedures for \mathcal{F} , it is essential to establish the fact that the set

$$\mathcal{D}_0 := \{\lambda \in \Lambda_{-1} : \exists f \in \mathcal{F}, \lambda' \preceq \lambda : \langle f, \sigma'_{\lambda'} \rangle \neq 0\}$$

of possible roots is finite. Fortunately, this is the case if the shearlet frame consists of compactly supported functions:

Lemma 21. *If $\varphi', \psi^{(0)}, \psi^{(1)}$ are compactly supported and Σ' is constructed as in (18), then $\text{card } \mathcal{D}_0 < \infty$.*

Proof. We show that for all $m \in \mathbb{Z}^2$, there exists a bounded set D in \mathbb{Z}^2 such that for all $\lambda \preceq m$ we have $\text{supp } \sigma'_\lambda \subset m + D$. Since all $f \in \mathcal{F}$ are supported in $[0, 1]^2$, this implies that only a finite number of indices $m \in \Lambda_{-1}$ can occur as possible root. For any $\lambda = (j, l, k, d) \in \Lambda$ it is not hard to see that the compact support of the basis functions implies that $\text{supp } \sigma'_\lambda \subset A_d^{-j} B_d^{-l} k + 2^{-j} B$, where B is some bounded set in \mathbb{R}^2 . We will now write A_λ for the dilation matrix $B_d^l A_d^j$ associated with an index $\lambda = (j, l, k, d)$. The children of m in Λ_0 are given by all indices $\lambda_0 = (0, l_0, k_0, d_0)$ with $k_0 \in B_{d_0}^{l_0} A_{d_0}^0 m$. We shall now drop the subscript d for the matrices A, B and \mathcal{E} . The children of m in Λ_1 are given by all indices $\lambda_1 = (1, l_1, k_1, d_1)$ with $k_1 \in B^\nu A k_0 + B^\nu \mathcal{E}$, where $\nu \in \{0, 1\}$ and $k_0 \in B^{l_0} A^0 m$ for some l_0 and therefore $k_1 \in A_{\lambda_1} m + A_{\lambda_1} A_{\lambda_0}^{-1} A^{-1} \mathcal{E}$. Iterating this argument shows that $\lambda_n \in \Lambda_n$ is a child of m only if $k_n \in A_{\lambda_n} \left(m + \sum_{i=2}^{n+1} A_{\mu_i}^{-1} \mathcal{E} \right)$ with some indices $\mu_i \in \Lambda_i$. An elementary computation shows that $\|A_{\mu_i}^{-1}\| \lesssim 2^{-i}$ uniformly for all $\mu_i \in \Lambda_i$. It follows that for $\lambda_n \in \Lambda_n$ we have $\text{supp } \sigma'_{\lambda_n} \subset \bigcup_{e \in \mathcal{E}} m + \sum_{i=2}^{n+1} A_{\mu_i}^{-1} e + 2^{-n} B \subset m + \sum_{i \in \mathbb{N}} 2^{-i} [0, 4]^2 + B$. It follows that for all children λ of m we have $\text{supp } \sigma'_\lambda \subset m + D$ with a bounded set D . This proves the assertion. \square

Moreover, by Theorem 20, the conclusion of Theorem 4 remains valid for compactly supported shearlet frames.

Using the fact that the set \mathcal{D}_0 of roots is finite, we can perform the exact same encoding construction as in [7, Section 6] and construct an encoder E_N which has length $M(E_N) \lesssim 2^{(2/3+\varepsilon)N}$ for all $\varepsilon > 0$ and $N \in \mathbb{N}$ and a decoder D_N with

$$d(E_N, D_N) \lesssim 2^{-(2/3-\varepsilon)N}.$$

It follows that

$$d(E_N, D_N) \lesssim M(E_N)^{-1+\varepsilon}$$

for all $\varepsilon > 0$, a result that is optimal if we disregard the arbitrarily small ε , compare [16].

Having a close-to-optimal bit rate coding procedure allows us to draw some conclusions regarding the Kolmogorov entropy of \mathcal{F} . We equip \mathcal{F} with the metric inherited from $L_2(\mathbb{R}^2)$. It is not difficult to see that \mathcal{F} is contained in a compact subset of $L_2(\mathbb{R}^2)$. For any $\nu > 0$ there exists a minimal number N_ν such that \mathcal{F} can be covered by N_ν balls with diameter ν . The Kolmogorov ν -entropy H_ν is defined by

$$H_\nu := \log N_\nu.$$

Corollary 22. *For any $\varepsilon > 0$ the Kolmogorov ν -entropy satisfies*

$$H_\nu \lesssim \nu^{-1+\varepsilon}$$

Proof. Using the encoding/decoding pair described above, we can consider the image of \mathcal{F} under the mapping E_N which has cardinality $\lesssim 2^{M(E_N)}$. Now consider the system of balls with midpoints $\{D_N(E_N(f)) : f \in \mathcal{F}\}$ and radius $\sim M(E_N)^{-1+\varepsilon}$. By the fact that $d(E_N, D_N) \lesssim M(E_N)^{-1+\varepsilon}$, it follows that this system is a covering of \mathcal{F} . On the other hand, the number of elements in this covering is $2^{M(E_N)}$ and therefore $H_{M(E_N)^{-1+\varepsilon}} \lesssim M(E_N)$. This proves the statement. \square

Of course there exist several other methods to bound the Kolmogorov entropy of \mathcal{F} , see e.g. [17, 26]. However, the method outlined in this section provides a particularly simple proof. Also the coding procedure which we presented is very simple: It is based on simple hard thresholding of the frame coefficients of f with respect to a *nonadaptive* frame. This stands in contrast to other adaptive methods like for instance bandelets [26].

Acknowledgements

The author would like Christoph Schwab for useful discussions. This work has been funded by the European Research Council under Grant ERC Project STAHPDE No. 247277. A part of the results of the present work has been announced in [20].

References

- [1] Toby Berger. *Rate Distortion Theory*. Prentice Hall, 1970.
- [2] Emmanuel Candes and Laurent Demanet. The curvelet representation of wave propagators is optimally sparse. *Communications in Pure and Applied Mathematics*, 58:1472–1528, 2004.
- [3] Emmanuel Candes and David Donoho. Curvelets - a surprisingly effective nonadaptive representation for objects with edges. In *Curves and Surfaces*. Vanderbilt University Press, 1999.
- [4] Emmanuel Candes and David Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Communications in Pure and Applied Mathematics*, 57:219–266, 2002.
- [5] Emmanuel Candes and David Donoho. Continuous curvelet transform: II. discretization and frames. *Applied and Computational Harmonic Analysis*, 19(2):198–222, 2005.
- [6] Ole Christensen. *An Introduction to Frames and Riesz Bases*. Birkhuser, 2003.

- [7] Albert Cohen, Wolfgang Dahmen, Ingrid Daubechies, and Ronald DeVore. Tree approximation and optimal encoding. *Applied and Computational Harmonic Analysis*, 11:192–226, 2001.
- [8] Albert Cohen, Ronald DeVore, Pencho Petrushev, and Hong Xu. Nonlinear approximation and the space $BV(\mathbb{R}^2)$. *American Journal of Mathematics*, 121:587–628, 1999.
- [9] Albert Cohen, Nira Dyn, Frederic Hecht, and Jean-Marie Mirebeau. Adaptive multiresolution analysis based on anisotropic triangulations. Technical report, 2009.
- [10] Ronald Coifman and Mladen Wickerhauser. Entropy based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 32:712–718, March 1992.
- [11] Laurent Demaret, Nira Dyn, and Armin Iske. Image compression by linear splines over adaptive triangulations. *Signal Processing*, 86:1604–1616, 2006.
- [12] Ronald DeVore. Nonlinear approximation. *Acta Numerica*, pages 51–150, 1998.
- [13] Ronald DeVore and George Lorentz. *Constructive Approximation*. Springer, 1993.
- [14] Minh Do and Martin Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14:2091–2106, 2005.
- [15] David Donoho. Unconditional bases and bit-level compression. *Applied and Computational Harmonic Analysis*, 3, 1996.
- [16] David Donoho. Wedgelets: Nearly minimax estimation of edges. *Annals of Statistics*, 27:353–382, 1998.
- [17] David Donoho. Sparse components of images and optimal atomic decompositions. *Constructive Approximation*, 17:353–382, 1999.
- [18] David Donoho and Gitta Kutyniok. Sparsity equivalence of anisotropic decompositions. Technical report, 2010.
- [19] Glenn Easley, Wang-Q Lim, and Demetrio Labate. Sparse directional image representations using the discrete shearlet transform. *Applied and Computational Harmonic Analysis*, 25:25–46, 2008.
- [20] Philipp Grohs. Tree approximation and optimal image coding with shearlets. In *Proceedings of the SAMPTA 2011 conference*, 2011. submitted.
- [21] Kanghui Guo and Demetrio Labate. Optimally sparse multidimensional representation using shearlets. *SIAM journal on mathematical analysis*, 39(1):298–318, 2008.
- [22] Kanghui Guo and Demetrio Labate. Representation of Fourier Integral Operators using shearlets. *Journal of Fourier Analysis and Applications*, 14(3):327–371, 2008.
- [23] Pisamai Kittipoom, Gitta Kutyniok, and Wang-Q Lim. Construction of compactly supported shearlet frames. Technical report, 2010.
- [24] Gitta Kutyniok and Wang-Q Lim. Compactly supported shearlets are optimally sparse. Technical report, 2010.
- [25] Demetrio Labate, Wang-Q Lim, Gitta Kutyniok, and Guido Weiss. Sparse multidimensional representation using shearlets. In *Proceedings of the SPIE*, pages 254–262, 2005.
- [26] Emmanuel LePennec and Stephane Mallat. Sparse geometric image representation with bandelets. *IEEE Transactions on Image Processing*, 14:423–438, 2005.
- [27] Hart Smith. A Hardy space for Fourier Integral Operators. *Journal of Geometric Analysis*, 8:629–653, 1998.

Research Reports

No.	Authors/Title
10-44	<i>P. Grohs</i> Tree approximation with anisotropic decompositions
10-43	<i>J. Li, H. Liu, H. Sun and J. Zou</i> Reconstructing acoustic obstacles by planar and cylindrical waves
10-42	<i>E. Kokiopoulou, D. Kressner and Y. Saad</i> Linear dimension reduction for evolutionary data
10-41	<i>U.S. Fjordholm</i> Energy conservative and -stable schemes for the two-layer shallow water equations
10-40	<i>R. Andreev and Ch. Schwab</i> Sparse tensor approximation of parametric eigenvalue problems
10-39	<i>R. Hiptmair, A. Moiola and I. Perugia</i> Stability results for the time-harmonic Maxwell equations with impedance boundary conditions
10-38	<i>I. Hnětynková, M. Plešinger, D.M. Sima, Z. Strakoš and S. Van Huffel</i> The total least squares problem in $AX \approx B$. A new classification with the relationship to the classical works
10-37	<i>S. Mishra</i> Robust finite volume schemes for simulating waves in the solar atmosphere
10-36	<i>C. Effenberger, D. Kressner and C. Engström</i> Linearization techniques for band structure calculations in absorbing photonic crystals
10-35	<i>R. Hiptmair and C. Jerez-Hanckes</i> Multiple traces boundary integral formulation for Helmholtz transmission problems
10-34	<i>H. Harbrecht and Ch. Schwab</i> Sparse tensor finite elements for elliptic multiple scale problems
10-33	<i>K. Grella and C. Schwab</i> Sparse tensor spherical harmonics approximation in radiative transfer
10-32	<i>P. Kauf, M. Torrilhon and M. Junk</i> Scale-induced closure for approximations of kinetic equations
10-31	<i>M. Hansen</i> On tensor products of quasi-Banach spaces