TREE APPROXIMATION AND OPTIMAL IMAGE CODING WITH SHEARLETS

Philipp Grohs

Seminar for Applied Mathematics,
ETH Zürich, Switzerland
Email: pgrohs@math.ethz.ch

ABSTRACT

It is by now classical that various anisotropic frame decompositions such as curvelets or shearlets guarantee (almost) optimal N-term approximation rates for functions which are C^2 apart from a C^2 discontinuity curve. However, if no structure is present in the set of retained indices, the cost of transmitting the location of the indices might dominate the cost of transmitting the actual coefficients. Therefore, as far as bit rate coding is concerned, simply storing the N largest (curvelet- or shearlet-) coefficients possibly leads to non-optimal codes. In the wavelet case this issue can be resolved by requiring that the set of indices which are kept possesses a tree structure which can be encoded more efficiently. In the present work we show how an analogous procedure can be carried out for curvelets or shearlets. The main result is that the N-term approximation rate can be essentially retained while imposing the additional constraint that the set of indices is a tree.

Keywords— Shearlets, Curvelets, Tree Approximation, Optimal Encoding, Kolmogorov Entropy

1. INTRODUCTION

A simple and popular method to transform a function $f \in L_2(\mathbb{R}^d)$ into a discrete sequence of numbers of length N is to expand f in a frame $\Phi := (\psi_\lambda)_{\lambda\in\Lambda}$ (see [2] for information about frames) satisfying

$$\|f\|_2^2 \sim \sum_{\lambda\in\Lambda} |\langle f, \psi_\lambda\rangle|^2 \qquad (1)$$

and keeping only the N largest (in modulus) coefficients $(\langle f, \psi_\lambda\rangle)_{\lambda\in\Lambda^{(N)}(f)}$ of the sequence $(\langle f, \psi_\lambda\rangle)_{\lambda\in\Lambda}$. It is well known that every frame Φ possesses a dual frame $\tilde{\Phi} := \left(\tilde{\psi}_\lambda\right)_{\lambda\in\Lambda}$ such that

$$f = \sum_{\lambda\in\Lambda} \langle f, \psi_\lambda\rangle \tilde{\psi}_\lambda$$

in $L_2(\mathbb{R}^d)$. If (1) holds with '=' instead of '$\sim$' we call Φ a *Parseval frame*. In this case $\tilde{\Phi} = \Phi$. Given the sequence $(\langle f, \psi_\lambda\rangle)_{\lambda\in\Lambda^{(N)}(f)}$ one can compute an approximation of f by $f_N := \sum_{\lambda\in\Lambda^{(N)}(f)} \langle f, \psi_\lambda\rangle \tilde{\psi}_\lambda$. A central question in nonlinear approximation is to determine the rate at which f_N converges to f in $L_2(\mathbb{R}^d)$ (or in more general function space norms). For univariate functions f which are smooth except for pointwise discontinuities, it can be shown that the N-term approximation f_N in a wavelet basis converges to f at an optimal rate [5, 4]. Unfortunately, this optimality property does not hold anymore for bivariate functions which are smooth apart from curvilinear singularities. Such functions serve as popular models for images. To be more specific, we define the class

$$\begin{aligned}\mathcal{F} := \{f_0 + f_1\chi_B : \text{supp } f_0, \text{supp } f_1 \subset [0,1]^2,\\ \|f_0\|_{C^2}, \|f_1\|_{C^2} \le 1 \text{ and } \chi_B\\ \text{is the indicator function of a domain } B \text{ with non-}\\ \text{selfintersecting boundary curve with curvature} \le 1\}.\end{aligned}$$

We regard $\mathcal{F}$ as a reasonably realistic model for images and would like to encode its elements as efficiently as possible. In a wavelet basis, the N-term approximation error decays of order $N^{-1/2}$ which is far from the optimal rate of N^{-1} [7]. In a breakthrough work, Candes and Donoho constructed so-called curvelet frames which reach this optimal approximation rate if one disregards logarithmic terms [1]. The amazing thing about this result is the simplicity of the approximation procedure, namely hard thresholding of the frame coefficients in a fixed, nonadaptive frame. Following this work, in [13, 9] shearlet systems were constructed which satisfy the same approximation properties but also a number of other desirable properties, such as compact support, see [11].

Despite these strong results, in view of constructing a coding scheme for functions in $f \in \mathcal{F}$, this solves only part of the problem: If we want to store the N-term approximation f_N e.g. in a shearlet frame, we first need to transform the coefficients $\langle f, \psi_\lambda\rangle$, $\lambda \in \Lambda^{(N)}(f)$ into bit sequences of finite length (quantization). Further, we also need to store the locations of the coefficients $\Lambda^{(N)}(f)$. This latter task may actually be quite costly, and even dominate the cost of storing the coefficients. A way out of this problem is to impose the additional constraint, that the index set $\Lambda^{(N)}(f)$ possesses a tree structure, in which case it can be encoded much more efficiently. The central question to answer in this regard is whether this additional constraint deteriorates the approximation rate of the N-term approxima-

This work has been funded by the European Research Council under Project STAHDPDE No. 247277.

tion. For the case of wavelets (and functions f in certain Besov spaces) this has been carried out in [3], where it is shown that the approximation rate *can* be retained. In the present paper we describe analogous results obtained in [8] for encoding functions in $\mathcal{F}$ using curvelets or shearlets. The main outcome is a close-to-optimal encoding/decoding pair in the sense of rate-distortion coding.

2. SHEARLETS

Shearlets have had a big impact in the field of applied harmonic analysis in the past few years due to their ability to represent anisotropic features efficiently. Compared to curvelets, they have the additional property of being defined over a uniform grid which will also turn out to be beneficial when we introduce the tree structure on the index set below. Shearlets are built using the operations of translation, anisotropic dilation and shearing. We follow [9] 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 [9] it is shown that there exist functions $\varphi, \psi^{(0)}, \psi^{(1)}$ such that with

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

the system

$$\{\varphi_k : k \in \mathbb{Z}\} \cup \left\{\psi_{(j,k,l,d)} : j \geq 0, -2^j \leq l \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\right\}$$

constitutes a Parseval frame for $L_2(\mathbb{R}^2)$ which means that (1) holds with '$\sim$' replaced by '$=$'. With $\Lambda_{-1} := \mathbb{Z}^2$ and $\Lambda_j := \left\{(j,l,k,d) : -2^j \leq l \leq 2^j - 1, k \in \mathbb{Z}^2, d = 0, 1\right\}$, the shearlet index set $\Lambda = \dot{\bigcup}_{j \geq -1} \Lambda_j$ 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 1. *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 $k \in B_d^{\lceil l/2 \rceil - l'}(A_d k' + \mathcal{E}_d)$ (see Figure 1). 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 $\mathcal{T}$.

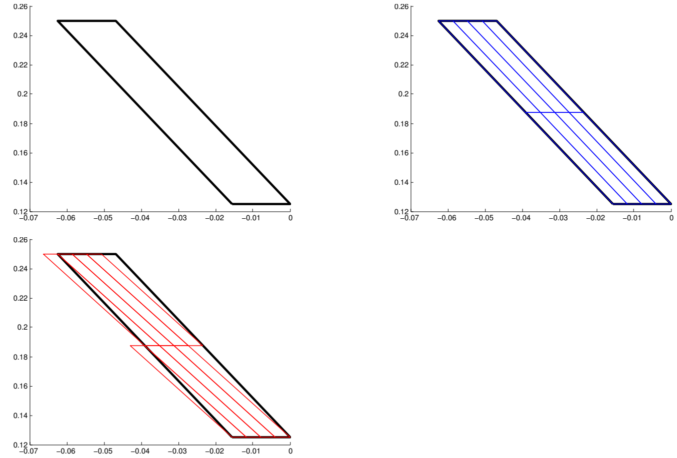

Fig. 1. Top left: Essential support of ψ_λ with $j = 3, l = 3, d = 0, k = (2,1)$. Top right: Essential support of its children with $l = 6$. Bottom left: Essential support of its children with $l = 7$.

3. NONLINEAR APPROXIMATION

Recall that we want to approximate the class $\mathcal{F}$ as defined in the introduction with as few as possible bits. One central concept in this direction is the N-term approximation rate which is defined by

$$\sigma_N(\mathcal{F}) := \sup_{f \in \mathcal{F}} \inf_{g \in \Sigma^N} \|f - g\|_2,$$

where $\Sigma^N := \left\{\sum_{\lambda \in \Lambda^{(N)}} c_\lambda \psi_\lambda : \text{card } \Lambda^{(N)} \leq N\right\}$ and ψ_λ denotes the shearlet corresponding to the index λ. A central result (see [10]) is that

$$\sigma_N(\mathcal{F}) \lesssim N^{-1+\varepsilon} \text{ for any } \varepsilon > 0$$

which gives a nearly optimal rate, see [6].

Essentially this means that one can encode any $f \in \mathcal{F}$ with N coefficients up to an error which is approximately bounded by N^{-1}. Moreover, the best approximation of f in Σ_N can be constructed by only using the N terms $\langle f, \psi_\lambda \rangle \psi_\lambda$ in the representation $f = \sum_{\lambda \in \Lambda} \langle f, \psi_\lambda \rangle \psi_\lambda$ with the largest coefficients $|\langle f, \psi_\lambda \rangle|$. This method has the drawback that the overall cost of storing the indices of the N largest coefficients may actually dominate the total cost of storing the coefficients themselves. For wavelet methods it can be shown that this problem can be circumvented if one imposes the additional requirement that the set of stored indices possesses a tree structure. Indeed, it can be easily seen that encoding a tree with N elements is in general much cheaper than encoding an arbitrary set without any structure. The main question is whether one has to sacrifice approximation power if one approximates with trees. We show that this is not the case. First some definitions. For $N \geq 1$ let

$$\Sigma_N^t := \left\{\sum_{\lambda \in \mathcal{T}} c_\lambda \psi_\lambda : \mathcal{T} \subset \Lambda \text{ is a tree with} \leq N \text{ elements}\right\}$$

and

$$t_N(\mathcal{F}) := \sup_{f \in \mathcal{F}} \inf_{g \in \Sigma_N^t} \|f - g\|_2.$$

Our main result is the following:

Theorem 2 ([8]). *For any $\varepsilon > 0$ we have the approximation rate*

$$t_N(\mathcal{F}) \lesssim N^{-1+\varepsilon}.$$

Proof (sketch). We give a rough sketch of the proof. For $f \in \mathcal{F}$ define $\Lambda(f,\eta) := \{\lambda \in \Lambda : |\langle f, \psi_\lambda\rangle| \geq \eta\}$ and $\mathcal{T}(f,\eta)$ the smallest tree containing $\Lambda(f,\eta)$. The main task is to bound the growth of the function card $\mathcal{T}(f,\eta)$. Indeed, we can show that

$$\text{card}\, \mathcal{T}(f,\eta) \lesssim \eta^{-2/3-\varepsilon} \text{ for any } \varepsilon > 0. \quad (2)$$

Using this fact, one can then show that the approximant $\mathcal{S}(f,\eta) := \sum_{\lambda\in\mathcal{T}(f,\eta)} \langle f, \psi_\lambda\rangle \psi_\lambda$ satisfies the approximation rate

$$\|f - \mathcal{S}(f,\eta)\|_2 \lesssim \eta^{2/3-\varepsilon} \text{ for all } \varepsilon > 0. \quad (3)$$

Together, (2) and (3) establish the desired claim. □

We would like to add that this result remains valid also for curvelets and for other shearlet systems [8]. In particular one gets the same approximation results for compactly supported shearlet frames as constructed in [11], provided that some additional smoothness- and moment conditions on the generators are satisfied, see [12]. Actually, the compact support will turn out to be crucial in the next section where we construct (almost) optimal encoding schemes for functions $f \in \mathcal{F}$.

4. OPTIMAL IMAGE CODING

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* D 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

$$d(E,D) := \sup_{f\in\mathcal{F}} \|f - D(E(f))\|_2. \quad (4)$$

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 [3]. 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.* In order to construct good shearlet coding procedures for $\mathcal{F}$ it is therefore essential to establish the fact that the set

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

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

Lemma 3. *If $\varphi, \psi^{(0)}, \psi^{(1)}$ are compactly supported, then 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 supp $\psi_\lambda \subset m + D$. 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 supp $\psi_\lambda \subset A_d^{-j} B_d^{-l} k + 2^{-j} B$, 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 supp $\psi_{\lambda_n} \subset \bigcup_{e\in\mathcal{E}} m + \sum_{i=2}^{n+1} A_{\mu_i}^{-1} e + 2^{-n} B \subset m + \sum_{i\in\mathbb{N}} 2^{-i}[0,4]^2 + B$. It follows that for all children λ of m we have supp $\psi_\lambda \subset m + D$ with a bounded set D. This proves the assertion. □

Moreover, as we have mentioned in the previous section, the conclusion of Theorem 2 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 [3, 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 [6].

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 4. *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. $\square$

The result about the Kolmogorov entropy is essentially known, see e.g. [7]. 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 [14].

5. CONCLUSION

In this short note we have demonstrated the ability of shearlets to perform nearly optimally for bit-rate image coding. The same results also hold for curvelet systems but due to the fact that they are not defined over a uniform grid things become considerably more cumbersome. A more detailed exposition of these results together with the full proofs is given in [8].

Acknowledgements

We would like to thank Christoph Schwab for useful discussions.

6. REFERENCES

[1] E. Candes and D. 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.

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

[3] A. Cohen, W. Dahmen, I. Daubechies, and R. Devore. Tree approximation and optimal encoding. *Applied and Computational Harmonic Analysis*, 11:192–226, 2001.

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

[5] R. DeVore. Nonlinear approximation. *Acta Numerica*, pages 51–150, 1998.

[6] D. Donoho. Wedgelets: Nearly minimax estimation of edges. *Annals of Statistics*, 27:353–382, 1998.

[7] D. Donoho. Sparse components of images and optimal-donoho wedgelets atomic decompositions. *Constructive Approximation*, 17:353–382, 1999.

[8] P. Grohs. Tree approximation with anisotropic decompositions. Technical report, 2010.

[9] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM journal on mathematical analysis*, 39(1):298–318, 2008.

[10] K. Guo and D. Labate. Representation of fourier integral operators using shearlets. *Journal of Fourier Analysis and Applications*, 14(3):327–371, 2008.

[11] P. Kittipoom, G. Kutyniok, and W.-Q. Lim. Construction of compactly supported shearlet frames. Technical report, 2010.

[12] G. Kutyniok and W.-Q. Lim. Compactly supported shearlets are optimally sparse. Technical report, 2010.

[13] D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets. In *Proceedings of the SPIE*, pages 254–262, 2005.

[14] E. LePennec and S. Mallat. Sparse geometric image representation with bandelets. *IEEE Transactions on Image Processing*, 14:423–438, 2005.