Compactly supported shearlets are optimally sparse

Presentation by Lukas Weißhaupt and Viktor Glombik

Paper by Prof. Dr. Gitta Kutyniok and Wang-Q. Lim.

Seminar Applied Functional Analysis
10.01.2020

Motivation I: Going to the doctor

Fig. 1:

Motivation I: Going to the doctor

Fig. 1: We strive to eliminate noise while keeping important information. [6]

Motivation II: Why does this work

- In data relevant information is sparse
  - $\rightsquigarrow$ data storage, transmission, denoising, feature extraction, classification ...
- Our visual system takes in 10^7 bits/sec only processes 50 bits/sec.

"Neuroscientists have identified edge processing neurons in the earliest and most fundamental stages of the processing pipeline upon which visual processing is built."

— Candès and Donoho in [2] summarising the work of nobel prize winning neurophysiologist David Hubel.

Table of Contents

Sparse Approximation

or: the ABC of harmonic analysis

- represent complicated objects well by few objects
- construct analysing elements $\{\varphi_i\}$ which can best capture most relevant information of data

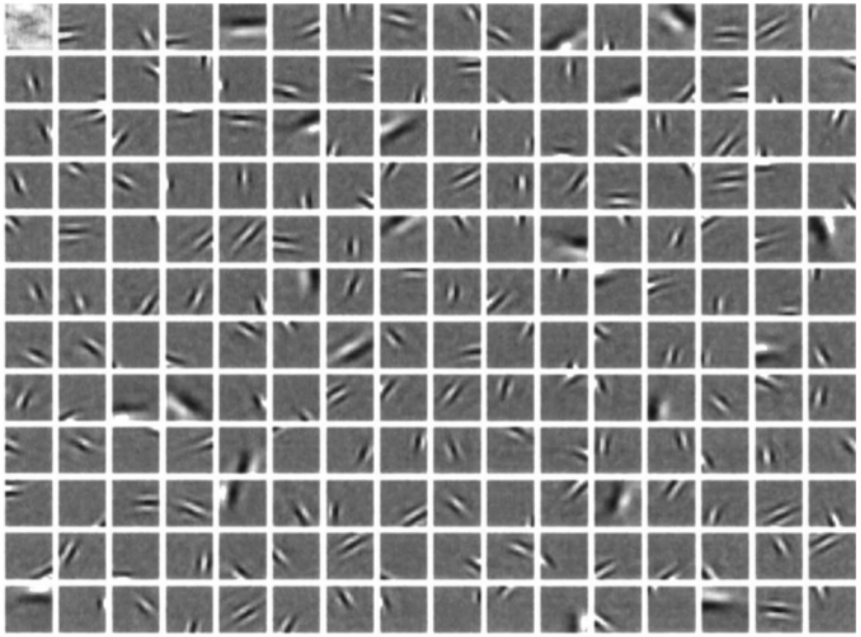
$$f = \sum_{i \in I} c_i \varphi_i$$

Sparse Approximation

or: the ABC of harmonic analysis

- represent complicated objects well by few objects
- construct analysing elements $\{\varphi_i\}$ which can best capture most relevant information of data

$$f = \sum_{i \in I} c_i \varphi_i$$



Sparse Components of Naturally-Occurring Image Data learned by unsupervised algorithm. [5]

Directional Representation Systems

From Bases to Frames

Given vectors $\{\varphi_k\}_k$ in a Hilbert space V we want to express all $v \in V$ as $v = \sum_k v_k \varphi_k$ efficiently.

Directional Representation Systems

From Bases to Frames

Given vectors $\{\varphi_k\}_k$ in a Hilbert space V we want to express all $v \in V$ as $v = \sum_k v_k \varphi_k$ efficiently.

Problem

Possibly $(\varphi_k)_k$ not linearly independent
$\implies$ $(v_k)_k$ not unique.

Directional Representation Systems

From Bases to Frames

Given vectors $\{\varphi_k\}_k$ in a Hilbert space V we want to express all $v \in V$ as $v = \sum_k v_k \varphi_k$ efficiently.

Problem

Possibly $(\varphi_k)_k$ not linearly independent
$\implies$ $(v_k)_k$ not unique.

Solution

It suffices if there exists $B \geq A > 0$ such that for all $v \in V$

$$A\|v\|^2 \leq \sum_{k\in\mathbb{N}} |\langle v, \varphi_k\rangle|^2 \leq B\|v\|^2. \qquad \text{(Frame condition)}$$

Directional Representation Systems

From Bases to Frames

Given vectors $\{\varphi_k\}_k$ in a Hilbert space V we want to express all $v \in V$ as $v = \sum_k v_k \varphi_k$ efficiently.

Problem

Possibly $(\varphi_k)_k$ not linearly independent
$\implies$ $(v_k)_k$ not unique.

Solution

It suffices if there exists $B \geq A > 0$ such that for all $v \in V$

$$A\|v\|^2 \leq \sum_{k\in\mathbb{N}} |\langle v, \varphi_k\rangle|^2 \leq B\|v\|^2. \qquad \text{(Frame condition)}$$

– The set $(\varphi_k)_k$ might be "overcomplete"

Frame theory

- Analysis operator: $T : V \to \ell_2,\ v \mapsto \{v_k = \langle v, \varphi_k \rangle\}_k$.

  $\rightsquigarrow$ Frame condition: $A\|v\|^2 \leq \|Tv\|_{\ell_2}^2 \leq B\|v\|^2$

- Synthesis operator: $T^* : \ell^2 \to V,\ \{v_k\}_k \mapsto v = \sum_k v_k \varphi_k$

- Define $S := TT^* : V \to V,\ v \mapsto \sum_k \langle v, \varphi_k \rangle \varphi_k$.

- With $\{\tilde{\varphi}_k = S^{-1}\varphi_k\}_k$ we have

$$v = \sum_k \langle v, \tilde{\varphi}_k \rangle \varphi_k = \sum_k \langle v, \varphi_k \rangle \tilde{\varphi}_k$$

Optimal sparsity of a directional representation systems

"Quality" of a directional representation system $(\varphi)_{i\in I}$ is measured by asymptotic $(N \to \infty)$ behaviour of its

Optimal sparsity of a directional representation systems

"Quality" of a directional representation system $(\varphi)_{i\in I}$ is measured by asymptotic ($N \to \infty$) behaviour of its

Definition (best N-term (non-linear) approximation)

The N-term approximation of $f \in \mathcal{E}^2(v)$ is given by

$$\|f - f_N\|_2^2, \qquad \text{where } f_N := \sum_{i\in I_n} \langle f, \varphi_i, \rangle \tilde{\varphi}_i$$

and $(\langle f, \varphi_i \rangle)_{i\in I_N}$ are the N largest coefficients in magnitude.

The Cartoon-like model I

Images are governed by anisotropic features.

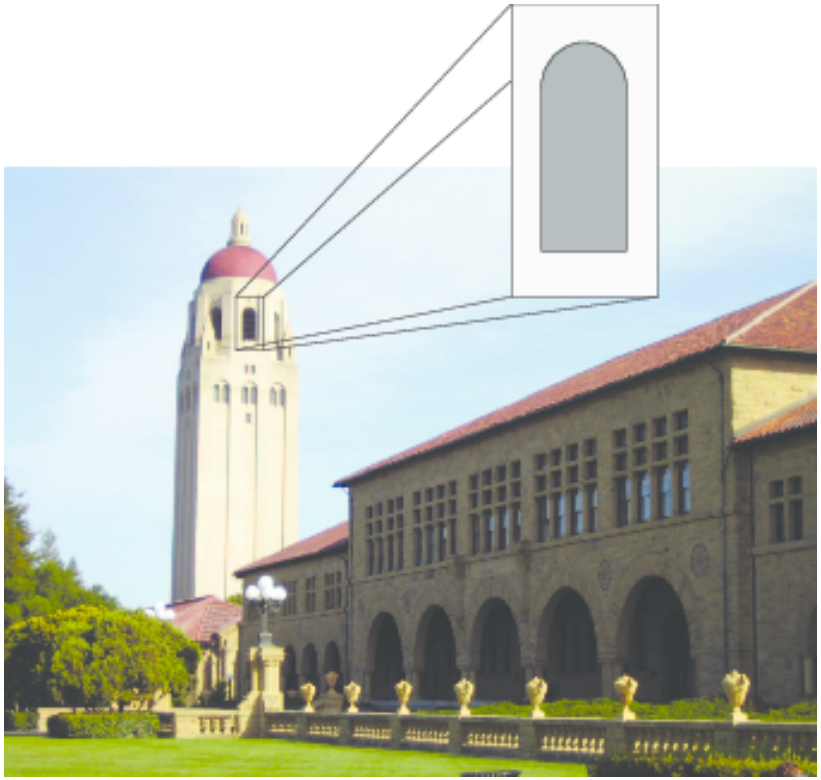


Fig. 2: [3] Images are made up of cartoon-like images

The Cartoon-like model II

Definition ($S^2(v)$ and cartoon-like images)

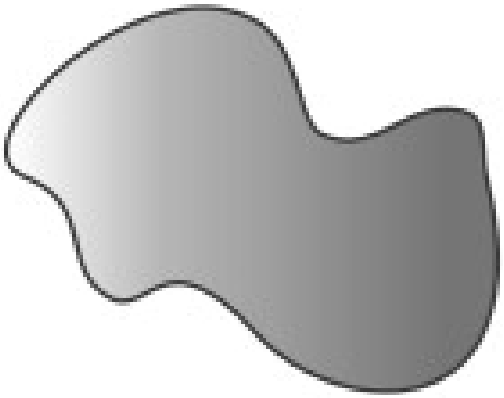
For $v > 0$, $S^2(v)$ is the set of $B \subset [0,1]^2$ being translates of

$$A := \{x \in \mathbb{R}^2 : |x| \leq \rho(\theta),\ x = (r,\theta)\},$$

with a radius function $\rho : [0, 2\pi) \to [0,1]$ with $\sup |\rho''(\theta)| \leq v$.

$$\mathcal{E}^2(v) := \{f_0 + f_1 \mathbb{1}_B : f_{0,1} \in \mathcal{C}^2(\mathbb{R}^2),\ \operatorname{supp}(f_{0,1}) \subset [0,1]^2, B \in S^2(v)\}$$

with $\|f_0 + f_1 \mathbb{1}_B\|_{\mathcal{C}^2} \leq 1$ is the class of cartoon-like functions.



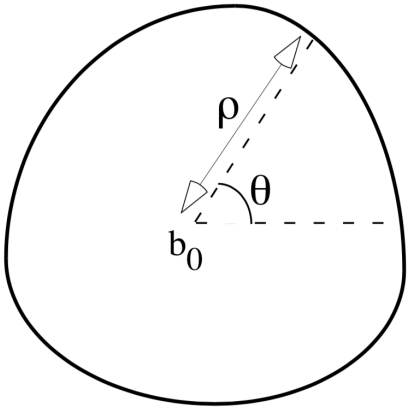


Table of Contents

Shearing

- scaling via parabolic scaling law $\mathsf{length}^2 \approx \mathsf{width}$
- directionality via parameterising slope

$$A_{2^j} := \begin{pmatrix} 2^j & 0 \\ 0 & 2^{\frac{j}{2}} \end{pmatrix}, \qquad \tilde{A}_{2^j} := \begin{pmatrix} 2^{\frac{j}{2}} & 0 \\ 0 & 2^j \end{pmatrix}, \qquad S_k := \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$$

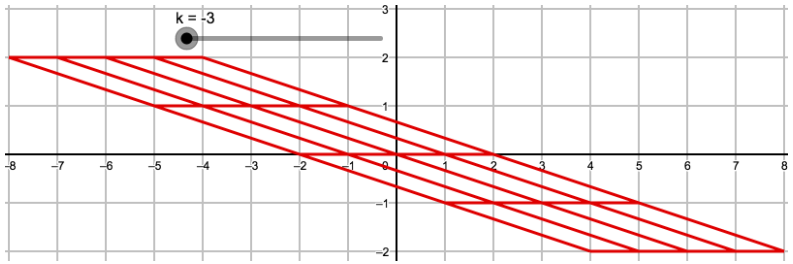


Fig. 4: The shearing matrix S_k leaves the integer grid $\mathbb{Z}^2$ invariant.

The cone-adapted shearlet system

Definition For some sampling constant $c > 0$ the cone-adapted shearlet system generated by a scaling function $g \in L^2(\mathbb{R}^2)$ and generators $h_1, h_2 \in L^2(\mathbb{R}^2)$ is

$$\mathcal{S}(g, h_1, h_2; c) := G(g; c) \cup H_1(h_1; c) \cup H_2(h_2; c),$$

where

$$G(g; c) := \{g_m := g(\cdot - c \cdot m) : m \in \mathbb{Z}^2\},$$
$$H_i(g; c) := \{h^{(i)}_{j,k,m} : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\},\ i \in \{1, 2\}$$

and

$$h^{(1)}_{j,k,m} := 2^{3j/4} h_1(S_k A_{2^j} \cdot - c \cdot m),\ h^{(2)}_{j,k,m} := 2^{3j/4} h_2(S_k^T \tilde{A}_{2^j} \cdot - c \cdot m).$$

Cones

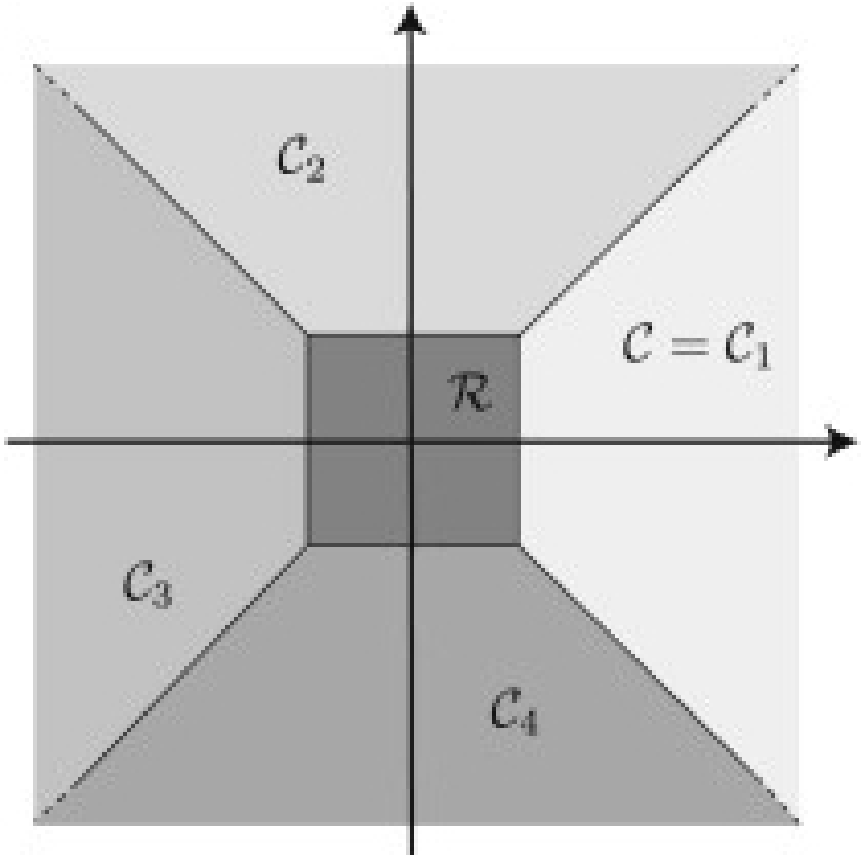


Fig. 5: Partitioning the frequency plane in four cones $\mathcal{C}_i$ and a rectangle $\mathcal{R}$. [1]

What are g, h_1 and h_2?

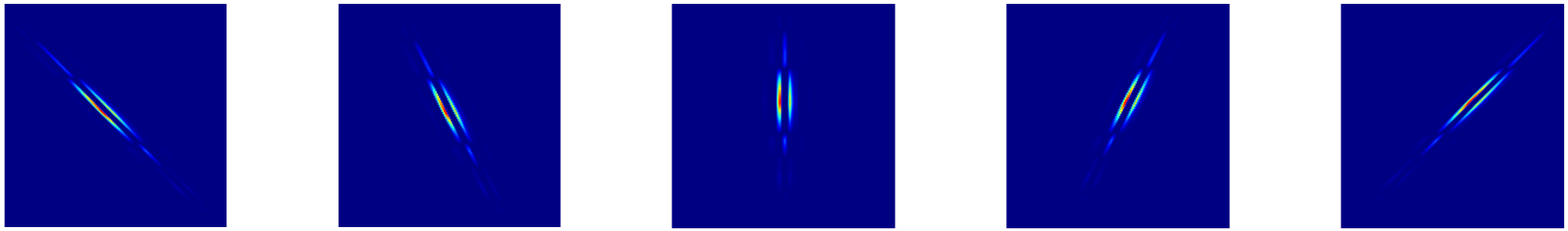


Fig. 6: Display of compactly supported shearlets $h^{(1)}_{2,k,0}$ in the spatial domain for $|k| \leq 2$

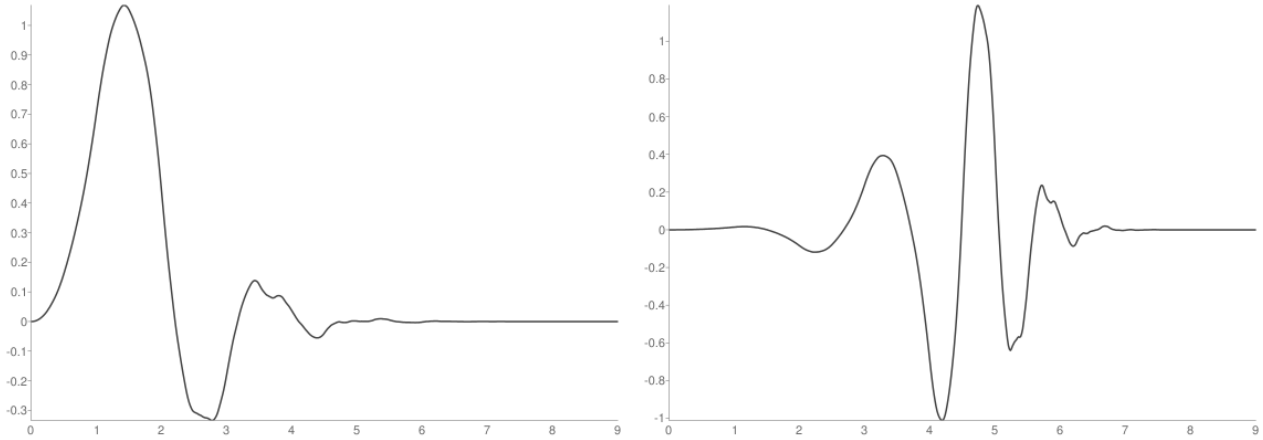


Fig. 7: Scaling function of a Daubechies Wavelet and the wavelet itself

Advantages of shearlets

+	−
sparse approximations	non-tight frame
unified treatment of continuum and digital	$\rightsquigarrow$ inverse transform difficult
fast algorithm	

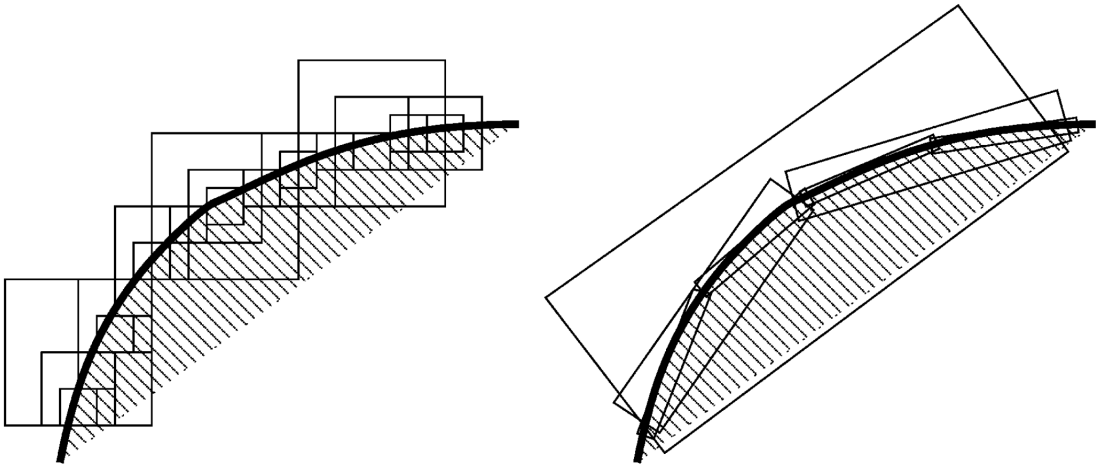


Fig. 8: [4]

Table of Contents

The main result

Theorem (Compactly supported shearlets are optimally sparse)

Let $c > 0$ and $g, h_1, h_2 \in L^2(\mathbb{R}^2)$ be compactly supported and h_1 and h_2 satisfy the weak directional vanishing moment conditions. If $\mathcal{S}(g, h_1, h_2; c)$ forms a frame for $L^2(\mathbb{R}^2)$, for any $v > 0$ the shearlet frame $\mathcal{S}(g, h_1, h_2; c)$ provides optimally space approximations of functions $f \in \mathcal{E}^2(v)$; there exists some $C > 0$ such that

$$\|f - f_N\|_2^2 \leq CN^{-2}(\log(N))^3 \qquad \text{as } N \to \infty,$$

where f_N is the nonlinear N-term approximation obtained by choosing the N largest shearlet coefficients of f.

Directional vanishing moment conditions (DC)

For all $x \in \mathbb{R}^2$ the shearlet h_1 satisfies

(i) $\left|\hat{h}_1(x)\right| \leq C_1 \cdot \mathsf{m}(|x_1|^{\alpha})\, \mathsf{m}(|x_1|^{-\gamma})\, \mathsf{m}(|x_2|^{-\gamma})$ $\quad (\star)$

(ii) $\left|\partial_2 \hat{h}_1(x)\right| \leq |t(x_1)| \cdot \left(1 + \frac{|x_2|}{|x_1|}\right)^{-\gamma}$,

where $\mathsf{m}(x) := \min(1, x)$, $\alpha > 5$, $\gamma \geq 4$, $t \in L^1(\mathbb{R})$ and C_1 is a constant. Furthermore, the shearlet h_2 satisfies (i) and (ii) with the roles of x_i reversed.

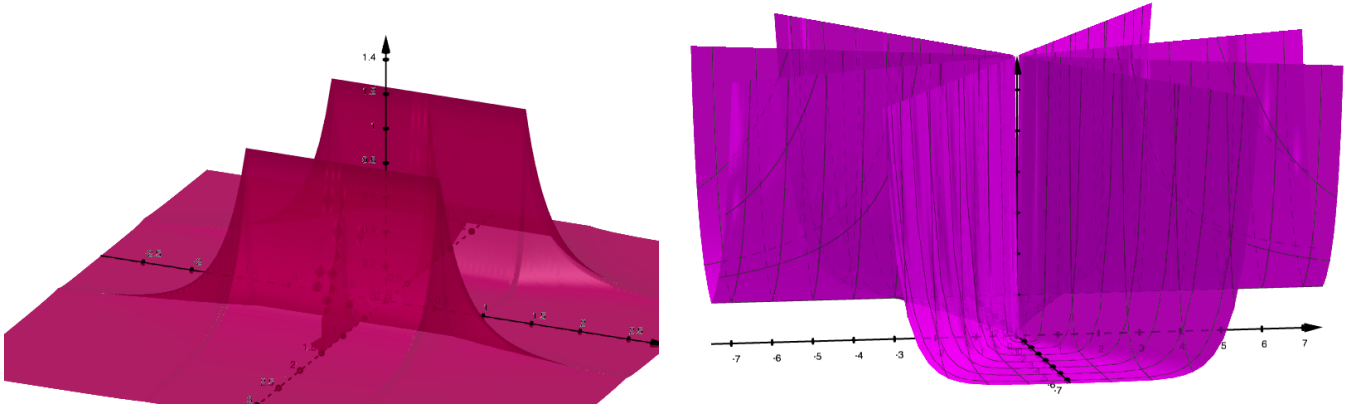


Fig. 9: The LHS of (i) for $\alpha = 6$ and $\gamma = 4$ and the LHS of (ii) for $t = \mathrm{id}$.

Proof strategy

Splitting $\mathcal{SH}(g, h_1, h_2; c)$ in two cases:

(i) $\operatorname{int}(\operatorname{supp}(h_\lambda^{(1)})) \cap \partial B = \emptyset$

(ii) $\operatorname{int}(\operatorname{supp}(h_\lambda^{(1)})) \cap \partial B \neq \emptyset$

For shearlets interacting with the discontinuity curve (ii):

- partition $\mathbb{R}^2$ into cubes
- for any cube Q examine shearlets intersecting ∂B in Q
- analyse slope of tangent to ∂B in Q and use it to bound N^{o} of intersecting shearlets

Proof of the main theorem
Preparation

- WLOG only consider $H_1(g;c)$ and $\|h_1\| < 1$.
  $\rightsquigarrow j \leq \frac{4}{3}\log_2(\varepsilon^{-1})$
- suffices to show that

$$\sum_{n>N} |\theta(f)|_n^2 \leq C \cdot N^{-2} \cdot (\log N)^3 \quad \text{as } N \to \infty,$$

as

$$||f - f_N||_2^2 \leq \frac{1}{A} \sum_{n>N} |\theta(f)|_n^2$$

by the frame condition.

Table of Contents

Analysis of coefficients associated with the smooth regions

Lemma (Optimal Sparsity away from discontinuities)

Let $g \in \mathcal{C}^2(\mathbb{R}^2)$ *with* $\operatorname{supp}(g) \subset [0,1]^2$ *and* $h \in L^2(\mathbb{R}^2)$ *compactly supported with*

$$|\hat{h}(x,y)| \leq C_1\, \mathsf{m}(|x|^a)\, \mathsf{m}(|x|^{-\gamma})\, \mathsf{m}(|y|^{-\gamma}) \quad \forall (x,y) \in \mathbb{R}^2, \qquad (\star)$$

where $\gamma > 3$, $a > \gamma + 2$ *and* C_1 *is a constant.*
Then there exists a $C > 0$ *such that*

$$\sum_{n>N} |\theta(g)|_n^2 \leq C N^{-2} \quad N \to \infty,$$

where $|\theta(f)|_n$ *denotes the* n*-th largest shearlet coefficient* $\langle f, h_\lambda \rangle$ *in absolute value.*

Auxiliary corollary

Corollary (Kittipoom, Kutyniok, Lim, 2010)

For a shearlet generator $h \in L^2(\mathbb{R}^2)$ satisfying $(\star)$ there exists some $C > 0$, such that for all $g \in \mathcal{C}^2(\mathbb{R}^2)$ with $\operatorname{supp}(g) \subset [0,1]^2$

$$\sum_{j=0}^{\infty} \sum_{\substack{k\in\mathbb{Z}:\\ |k|\leq\lceil 2^{j/2}\rceil}} \sum_{m\in\mathbb{Z}^2} 2^{4j} \left|\langle g, h_{j,k,m}\rangle\right|^2 \leq C \left\| \frac{\partial^2}{\partial x_1^2} g \right\|_2^2 .$$

holds.

Proof of lemma 1

Let

$$\zeta_j := \{\lambda \in \Lambda_j : \operatorname{supp}(h_\lambda) \cap \operatorname{supp}(g) \neq \emptyset\}.$$

We have

$$N_J := \left| \bigcup_{j=0}^{J-1} \zeta_j \right| = \sum_{j=0}^{J-1} |\zeta_j| \sim \sum_{j=0}^{J-1} \underbrace{2^{3j/2}}_{\text{position}} \underbrace{2^{j/2}}_{\text{shear}} \sim 2^{2J}.$$

as $\operatorname{supp}(h_\lambda)$ is contained in a rectangle of dimensions $2^j \times 2^{j/2}$ and $|k| \leq \lceil 2^{j/2} \rceil$.

Here $a_j \sim b_j \iff \exists c_1, c_2$ such that $c_1 a_j \leq b_j \leq c_2 a_j$.

Proof of lemma 1

There exists some $C > 0$ such that

$$\begin{aligned}
\sum_{j\in\mathbb{N}} 2^{4j} \sum_{n>N_j} |\theta(g)|_n^2 &\leq C \sum_{j=0}^{\infty} \sum_{\ell=j}^{\infty} \sum_{k,m} 2^{4j} \left|\langle g, h_{\ell,k,m}\rangle\right|^2 \\
&= C \sum_{\ell=0}^{\infty} \sum_{k,m} 2^{4j} \left|\langle g, h_{\ell,k,m}\rangle\right|^2 \left(\sum_{j=1}^{\ell} 2^{4j}\right) \\
&\leq C_1 \sum_{\ell=0}^{\infty} \sum_{k,m} 2^{4\ell} \left|\langle g, h_{\ell,k,m}\rangle\right|^2 \\
&\leq C_2 \left\| \frac{\partial^2}{\partial x_1^2} g \right\|_2^2 < \infty. \quad \text{(Auxillary Corollary)}
\end{aligned}$$

Proof of lemma 1

Hence for all $j \geq 0$,

$$\sum_{n>N_j} |\theta(g)|_n^2 \leq C_1 2^{-4j} = C_1 (2^{2j})^{-2} \leq C_2 N_j^{-2}.$$

For $N > 0$ there exists a $j_0 \in \mathbb{N}_{>0}$ such that

$$N \sim N_{j_0} \sim 2^{2j_0}.$$

$\square$

Table of Contents

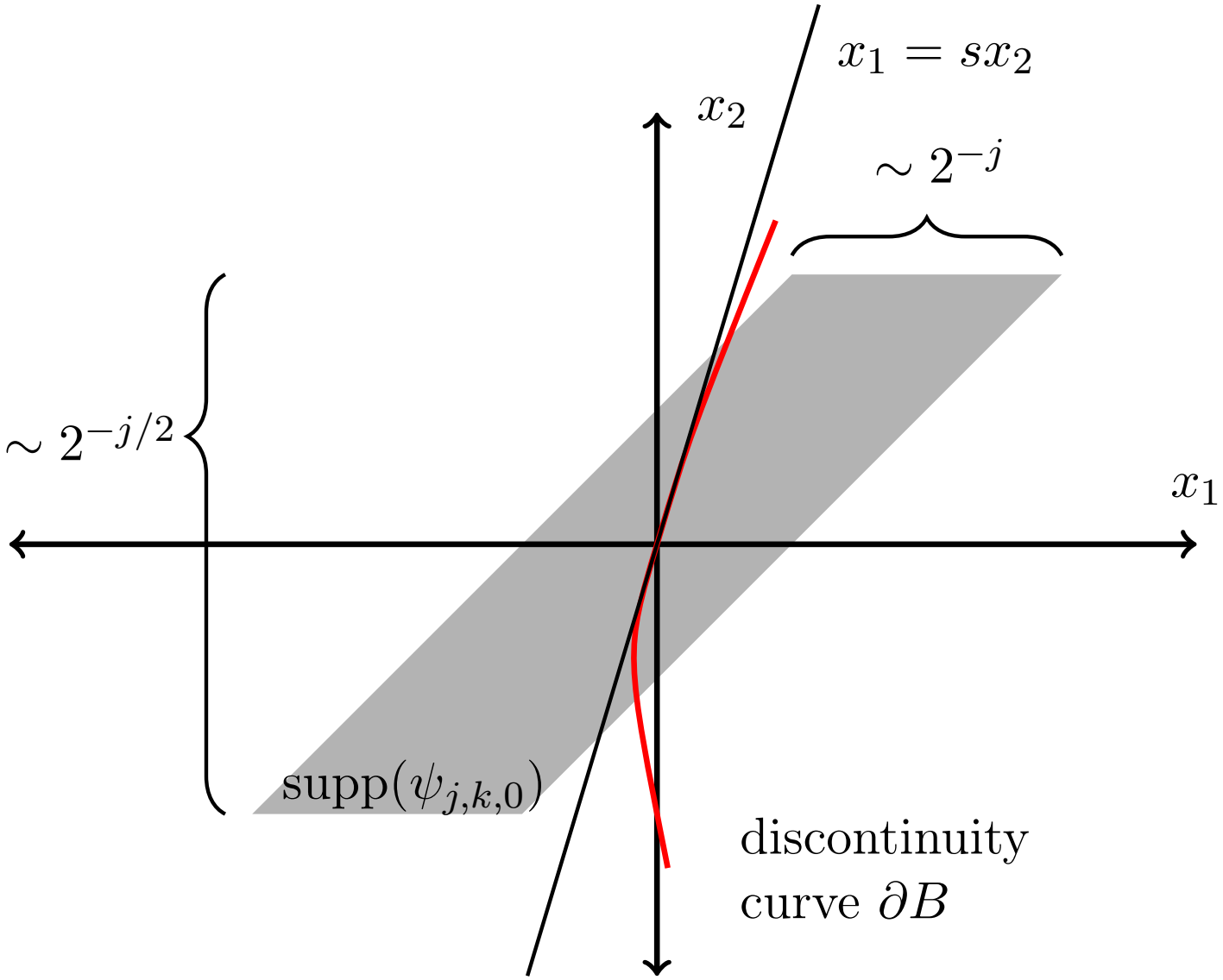
Proof of the main theorem
Notation

- For $j \geq 0$ let

$$\begin{aligned} Q_{j,0} &= [-2^{-j/2}, 2^{-j/2}]^2, \\ \Lambda_j &= \{\lambda \in \Lambda_j : \operatorname{int}(\operatorname{supp}(h_\lambda)) \cap \operatorname{int}(Q_j) \cap \partial B \neq \emptyset\}, \\ \Lambda_j(\epsilon) &= \{\lambda \in \Lambda_j : |\langle f, h_\lambda \rangle| > \epsilon\}) \end{aligned}$$

- parameterize ∂B by either $(x_1, E(x_1))$ or $(E(x_2), x_2)$.

Visualisation



Proof of the main theorem

- number of indices bounded by number of shearlets intersecting tangent curve (up to constants)
- If $m_2 \neq 0$, there's no intersection $\rightsquigarrow$ suffices to estimate number of possible m_1.
- $|m_1| \leq |\hat{k}|$ or $m_1 = 1$ if $\hat{k} = 0$.
- estimate independent of choice of $\hat{x}$.

Proof of the main theorem

- in case $|s| \leq 2$ by lemma 2 we have

$$|\langle f, h_\lambda \rangle| > \varepsilon \implies |\hat{k}| \leq C \cdot \epsilon^{-1/3} \cdot 2^{-j/4}$$

  and in turn

$$|\Lambda_j(\varepsilon)| \leq C \cdot \sum_k (|\hat{k}| + 1) \leq C \cdot (\varepsilon^{-1/3} \cdot 2^{-j/4} + 1)^2$$

- additionally, in case $|s| > 2$ by lemma 2 it suffices to consider

$$j \leq \frac{9}{4} \log_2(\varepsilon^{-1}) + C$$

  and we have

$$|\Lambda_j(\varepsilon)| \leq C \cdot 2^j$$

Proof of the main theorem

Combining everything

$$
\begin{aligned}
\left|\bigcup_j \Lambda_j(\epsilon)\right| \cdot \#(Q_j) &\leq \left(\sum_{|s|\leq 2} |\Lambda_j(\varepsilon)| + \sum_{|s|>2} |\Lambda_j(\varepsilon)|\right) \cdot \#(Q_j) \\
&\leq C \underbrace{\sum_{j=0}^{\lfloor \frac{4}{3}\log_2(\varepsilon^{-1})\rfloor} 2^{j/2}(\varepsilon^{-1/3} \cdot 2^{-j/4} + 1)^2}_{\sim \varepsilon^{-2/3} \cdot \log_2(\varepsilon^{-1})} \\
&\quad + C_1 \underbrace{\sum_{j=0}^{\lfloor \frac{4}{9}\log_2(\varepsilon^{-1})+C\rfloor} 2^{j/2} \cdot 2^j}_{\sim \varepsilon^{-1}} \\
&\leq C_2 \varepsilon^{-2/3} \log_2(\varepsilon^{-1}).
\end{aligned}
$$

Proof of the main theorem

The end

Setting $N := \left|\bigcup_j \Lambda_j(\varepsilon)\right|$, this yields

$$\varepsilon \leq CN^{-3/2} \cdot (\log(N))^{3/2}$$

implying

$$|\theta(f)|_N^2 \leq CN^{-3} \cdot (\log(N))^3.$$

Hence, as $N \to \infty$

$$\sum_{n>N} |\theta(f)|_n^2 \leq CN^{-2}(\log(N))^3.$$

Outlook

Further related topics

- 3D-Shearlets,
- Discontinuous shearlets [Kutyniok, Lim, 2015]
- algorithmic implementation.

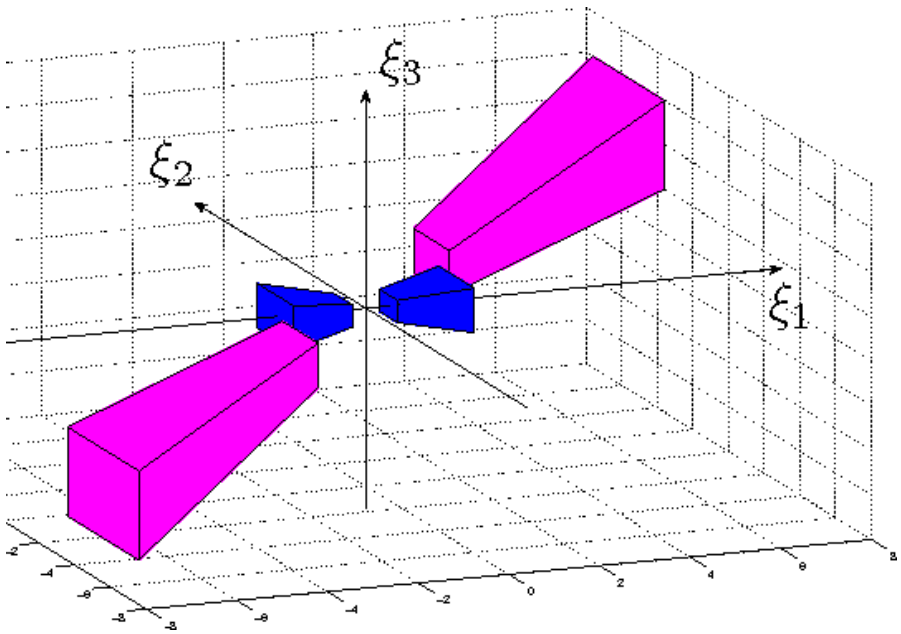


Fig. 10: 3D-shearlets. [7]

References I

[1] Gitta Kutyniok and Wang-Q Lim.
Compactly supported shearlets are optimally sparse.
2011.

[2] Emmanuel J. Candès, David L. Donoho.
New Tight Frames of Curvelets and Optimal Representations of Objects with $\mathcal{C}^2$ Singularities.
2002.

[3] Gitta Kutyniok, Demetrio Labate.
Shearlets. Multiscale analysis for multivariate data
Birkhäuser 2011

References II

[4] Minh N. Do and Martin Vetterli.
The Contourlet Transform: An Efficient Directional Multiresolution Image Representation,
IEEE Transanctions on image processing, Vol. 14, No. 12, December 2005

[5] Bruno A. Olshausen, David J. Field.
Emergence of simple-cell receptive field properties by learning a sparse code for natural images
Nature Volume 381, pages 607–609, 1996.

[6] Wei-Chun Lin, Jing-Wein Wang.
Edge detection in medical images with quasi high-pass filter based on local statistics
Biomedical Signal Processing and Control, Volume 39, Jan 2018

References III

[7] Kanghui Guo, Demetrio Labate.
Analysis and detection of surface discontinuities using the 3D continuous shearlet transform.
Applied and Computational Harmonic Analysis. Volume 30, Issue 2, March 2011.

[8] Selfmade in Geogebra.