Wavelets, Shearlets and Geometric Frames: Part I

Philipp Grohs[1] and Axel Obermeier[2]

October 20, 2014

[1]ETH Zürich
[2]ETH Zürich, supported by SNF grant 146356

Course Syllabus

MON	20th century methods (PG)
TUE	21st century methods (PG)
WED	Computer class (Axel Obermeier)

Outline for Today

0. Motivation
1. Nonlinear Approximation
2. Wavelets and Point Singularities

0. Motivation

'BIG DATA'

Source: The Economist (2010)

The Big Challenge

Find efficient representations for complicated data

What are Good Data Representations

Data typically modeled as function $f : \Omega \to M$, given either explicitly or (more often) implicitly as $F(f) = 0$ (operator equation or inverse problem).

What are Good Data Representations

Data typically modeled as function $f : \Omega \to M$, given either explicitly or (more often) implicitly as $F(f) = 0$ (operator equation or inverse problem).

Aim to find 'dictionary' of atoms $\{\varphi_i\}_{i \in I}$ such that each signal f can be represented as a 'sparse' linear combination of atoms:

$$f \sim \sum_{i \in J} c_i \varphi_i, \quad J \subset I, \ \#J = N, \ \text{where } N \text{ is small.}$$

What are Good Data Representations

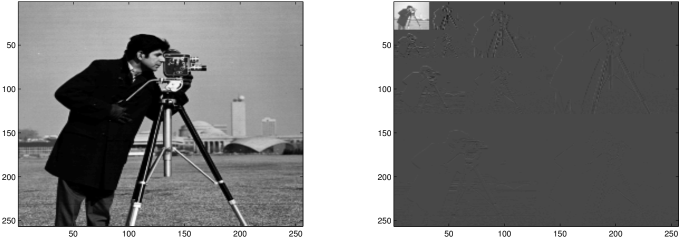
Data typically modeled as function $f : \Omega \to M$, given either explicitly or (more often) implicitly as $F(f) = 0$ (operator equation or inverse problem).

Aim to find 'dictionary' of atoms $\{\varphi_i\}_{i \in I}$ such that each signal f can be represented as a 'sparse' linear combination of atoms:

$$f \sim \sum_{i \in J} c_i \varphi_i, \quad J \subset I, \ \#J = N, \ \text{where } N \text{ is small.}$$

If f is not given explicitly, a 'good' representation should (approximately) diagonalize F.

A prominent Example: JPEG 2000

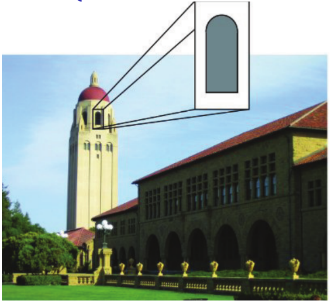


Representation of image in wavelet domain

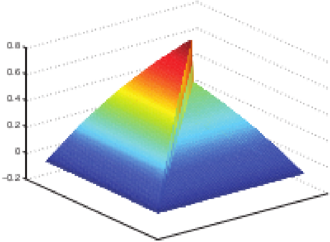
Information has its own architecture. Each data source, whether imagery, sound, text, has an inner architecture which we should attempt to discover and exploit for applications such as noise removal, signal recovery, data compression, and fast computation.

(D. Donoho, ICM Address, 2002)

Anisotropic Data (explicit and implicit)



Images are anisotropic!



Solutions to transport equations ($v \cdot \nabla u + \kappa u = f$) are anisotropic!

Big goal: Find optimal approximation schemes for anisotropic data!

Big goal: Find optimal approximation schemes for anisotropic data!

But What does this mean??

1. What's the best we can do?

A Primer in nonlinear Approximation Theory

Start with a Mathematical Model

Start with a Mathematical Model

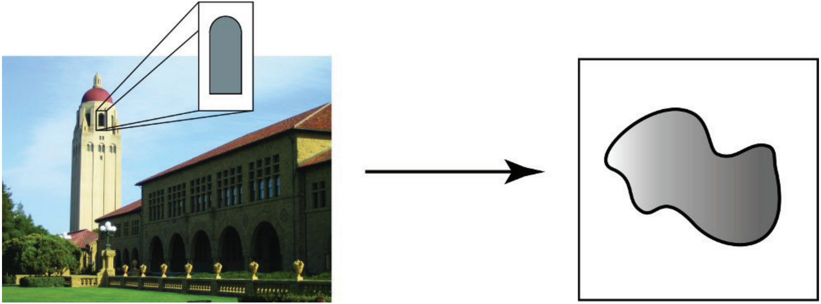
For instance...

Definition (Donoho (2001))

The set $\mathcal{E}$ of cartoon images is given by

$$\mathcal{E} := \{f = f_0 + \chi_B f_1\},$$

where $f_0, f_1 \in C^2([0,1]^2)$ and χ_B is the indicator function of $B \subset [0,1]^2$ with C^2 boundary.



Measuring the Quality of a Dictionary

Measuring the Quality of a Dictionary

- Let $\mathcal{C} \subset L^2(\mathbb{R}^d)$ compact be a 'signal class'

Measuring the Quality of a Dictionary

- Let $\mathcal{C} \subset L^2(\mathbb{R}^d)$ compact be a 'signal class'
- Let $\Phi := \{\varphi_i\}_{i \in I} \subset L^2(\mathbb{R}^d)$ be a dictionary

Measuring the Quality of a Dictionary

- Let $\mathcal{C} \subset L^2(\mathbb{R}^d)$ compact be a 'signal class'
- Let $\Phi := \{\varphi_i\}_{i \in I} \subset L^2(\mathbb{R}^d)$ be a dictionary
- Let $\Sigma_N(\Phi) := \left\{ \sum_{i \in J} c_i \varphi_i : \; J \subset I, \; \#J = N \right\}$

Measuring the Quality of a Dictionary

- Let $\mathcal{C} \subset L^2(\mathbb{R}^d)$ compact be a 'signal class'
- Let $\Phi := \{\varphi_i\}_{i \in I} \subset L^2(\mathbb{R}^d)$ be a dictionary
- Let $\Sigma_N(\Phi) := \left\{\sum_{i \in J} c_i \varphi_i : \ J \subset I, \ \#J = N\right\}$
- Let $\sigma_N(f, \Phi) := \inf_{g \in \Sigma_N} \|f - g\|_2$, the N-term approximation error

Measuring the Quality of a Dictionary

- Let $\mathcal{C} \subset L^2(\mathbb{R}^d)$ compact be a 'signal class'
- Let $\Phi := \{\varphi_i\}_{i \in I} \subset L^2(\mathbb{R}^d)$ be a dictionary
- Let $\Sigma_N(\Phi) := \left\{\sum_{i \in J} c_i \varphi_i : \; J \subset I, \; \#J = N\right\}$
- Let $\sigma_N(f, \Phi) := \inf_{g \in \Sigma_N} \|f - g\|_2$, the N-term approximation error
- Let $\mathcal{A}^s(\Phi) := \left\{f \in L^2(\mathbb{R}^d) : \; \sigma_N(f, \Phi) \leq CN^{-s}\right\}$

Measuring the Quality of a Dictionary

- Let $\mathcal{C} \subset L^2(\mathbb{R}^d)$ compact be a 'signal class'
- Let $\Phi := \{\varphi_i\}_{i \in I} \subset L^2(\mathbb{R}^d)$ be a dictionary
- Let $\Sigma_N(\Phi) := \left\{ \sum_{i \in J} c_i \varphi_i : \; J \subset I, \; \#J = N \right\}$
- Let $\sigma_N(f, \Phi) := \inf_{g \in \Sigma_N} \|f - g\|_2$, the N-term approximation error
- Let $\mathcal{A}^s(\Phi) := \left\{ f \in L^2(\mathbb{R}^d) : \; \sigma_N(f, \Phi) \leq CN^{-s} \right\}$
- Let $s^*(\mathcal{C}, \Phi) := \sup \{ s > 0 : \; \mathcal{C} \subset \mathcal{A}^s \}$

Measuring the Quality of a Dictionary

- Let $\mathcal{C} \subset L^2(\mathbb{R}^d)$ compact be a 'signal class'
- Let $\Phi := \{\varphi_i\}_{i \in I} \subset L^2(\mathbb{R}^d)$ be a dictionary
- Let $\Sigma_N(\Phi) := \left\{\sum_{i \in J} c_i \varphi_i : \ J \subset I, \ \#J = N\right\}$
- Let $\sigma_N(f, \Phi) := \inf_{g \in \Sigma_N} \|f - g\|_2$, the N-term approximation error
- Let $\mathcal{A}^s(\Phi) := \left\{f \in L^2(\mathbb{R}^d) : \ \sigma_N(f, \Phi) \leq CN^{-s}\right\}$
- Let $s^*(\mathcal{C}, \Phi) := \sup\{s > 0 : \ \mathcal{C} \subset \mathcal{A}^s\}$

The bigger $s^*(\mathcal{C}, \Phi)$, the better the dictionary Φ is suitable for the approximation of the signal class $\mathcal{C}$!

A perfect dictionary

A perfect dictionary

- Let $\Phi \subset L^2(\mathbb{R}^d)$ be dense.

A perfect dictionary

- Let $\Phi \subset L^2(\mathbb{R}^d)$ be dense.
- ☺ For every signal class $\mathcal{C}$ we have $s^*(\mathcal{C}, \Phi) = \infty$ (in fact we would have $\sigma_1(f, \Phi) = 0$ for all $f \in L^2(\mathbb{R}^d)$.

A perfect dictionary

- Let $\Phi \subset L^2(\mathbb{R}^d)$ be dense.
- ☺ For every signal class $\mathcal{C}$ we have $s^*(\mathcal{C}, \Phi) = \infty$ (in fact we would have $\sigma_1(f, \Phi) = 0$ for all $f \in L^2(\mathbb{R}^d)$.
- ☹ No way of efficiently computing 1-term approximation of given f. No way of efficiently storing the dictionary Φ.

A PERFECT DICTIONARY

- Let $\Phi \subset L^2(\mathbb{R}^d)$ be dense.
- ☺ For every signal class $\mathcal{C}$ we have $s^*(\mathcal{C}, \Phi) = \infty$ (in fact we would have $\sigma_1(f, \Phi) = 0$ for all $f \in L^2(\mathbb{R}^d)$.
- ☹ No way of efficiently computing 1-term approximation of given f. No way of efficiently storing the dictionary Φ.

$\rightsquigarrow$ need to specify the notion of 'allowed dictionaries'

Polynomial Depth Search

Allow only 'computable' N-term approximations: Identify I with $\mathbb{N}$ and search best N-term approximation, satisfying

$$f_N = \sum_{k=1}^{N} a_{\sigma(k,f)} \varphi_{\sigma(k,f)},$$

where $\sigma(k, f) \leq \pi(k)$ for a polynomial π.

Polynomial Depth Search

Allow only 'computable' N-term approximations: Identify I with $\mathbb{N}$ and search best N-term approximation, satisfying

$$f_N = \sum_{k=1}^{N} a_{\sigma(k,f)} \varphi_{\sigma(k,f)},$$

where $\sigma(k,f) \leq \pi(k)$ for a polynomial π.
Study $s^*_{\text{Poly}}(\mathcal{C},\Phi)$, defined as $s^*(\mathcal{C},\Phi)$ under the constraint that the N-term approximations are computed using polynomial depth search and define

$$s^*(\Phi) := \sup \left\{ s^*_{\text{Poly}}(\mathcal{C},\Phi) : \ \Phi \text{ dictionary} \right\}$$

and use this as benchmark for the complexity of signal class $\mathcal{C}$.

Now we know what 'optimal' means...

Definition

A dictionary Φ is optimal for a signal class $\mathcal{C}$ if

$$s^*(\mathcal{C}, \Phi) = s^*(\mathcal{C}).$$

Now we know what 'optimal' means...

Definition

A dictionary Φ is optimal for a signal class $\mathcal{C}$ if

$$s^*(\mathcal{C}, \Phi) = s^*(\mathcal{C}).$$

...but how can we know $s^*(\mathcal{C})$?

Embedded Hypercubes

Definition (Donoho (2001))

1. A class of functions $\mathcal{C} \subset L^2(\mathbb{R}^d)$ *contains an embedded orthogonal hypercube of dimension m and sidelength* δ if there exist $f_0 \in \mathcal{C}$ and orthogonal functions $\psi_i \in L^2(\mathbb{R}^d)$ for $i = 1, \ldots, m$ with $\|\psi_i\|_2 = \delta$ such that the collection of hypercube vertices

$$\mathfrak{H}(m; f_0, (\psi_i)_i) = \left\{ h = f_0 + \sum_{i=1}^{m} \epsilon_i \psi_i \ : \ \epsilon_i \in \{0, 1\} \right\}$$

is contained in $\mathcal{C}$.

Embedded Hypercubes

Definition (Donoho (2001))

1. A class of functions $\mathcal{C} \subset L^2(\mathbb{R}^d)$ *contains an embedded orthogonal hypercube of dimension* m *and sidelength* δ if there exist $f_0 \in \mathcal{C}$ and orthogonal functions $\psi_i \in L^2(\mathbb{R}^d)$ for $i = 1, ..., m$ with $\|\psi_i\|_2 = \delta$ such that the collection of hypercube vertices

$$\mathfrak{H}(m; f_0, (\psi_i)_i) = \{h = f_0 + \sum_{i=1}^{m} \epsilon_i \psi_i \ : \ \epsilon_i \in \{0, 1\}\}$$

   is contained in $\mathcal{C}$.

2. A class of functions $\mathcal{C}$ *contains a copy of* ℓ_0^p, $p > 0$, if there exists a sequence of orthogonal hypercubes $(\mathfrak{H}_k)_{k \in \mathbb{N}}$, embedded in $\mathcal{C}$, of dimensions m_k and side-lengths δ_k, such that $\delta_k \to 0$ and for $C > 0$

$$m_k \geq C\delta_k^{-p} \quad \text{for all } k \in \mathbb{N}.$$

THEOREM (DONOHO (2001), G-KEIPER-KUTYNIOK-SCHAEFER (2014))

Suppose, that a class of functions $\mathcal{C} \subset L^2(\mathbb{R}^d)$ is uniformly L^2-bounded and contains a copy of ℓ_0^p. Then

$$s^*(\mathcal{C}) \leq (2-p)/2p.$$

proof

Theorem (Donoho (2001), G-Keiper-Kutyniok-Schaefer (2014))

Suppose, that a class of functions $\mathcal{C} \subset L^2(\mathbb{R}^d)$ is uniformly L^2-bounded and contains a copy of ℓ_0^p. Then

$$s^*(\mathcal{C}) \leq (2-p)/2p.$$

proof

Theorem (G-Keiper-Kutyniok-Schaefer (2014))

Let $\mathcal{C}^\alpha = \{f \in C^\alpha[0,1]^d : \|f\|_{C^\alpha} \leq 1\}$. Then $\mathcal{C}$ contains a copy of ℓ_0^p with $p = \frac{2}{2\alpha/d+1}$. Consequently

$$s^*(\mathcal{C}^\alpha) \leq \alpha/d.$$

proof

Theorem (Donoho (2001), G-Keiper-Kutyniok-Schaefer (2014))

Suppose, that a class of functions $\mathcal{C} \subset L^2(\mathbb{R}^d)$ is uniformly L^2-bounded and contains a copy of ℓ_0^p. Then

$$s^*(\mathcal{C}) \leq (2-p)/2p.$$

▸ proof

Theorem (G-Keiper-Kutyniok-Schaefer (2014))

Let $\mathcal{C}^\alpha = \{f \in C^\alpha[0,1]^d : \|f\|_{C^\alpha} \leq 1\}$. Then $\mathcal{C}$ contains a copy of ℓ_0^p with $p = \frac{2}{2\alpha/d+1}$. Consequently

$$s^*(\mathcal{C}^\alpha) \leq \alpha/d.$$

▸ proof

For instance, the best we can do for $\mathcal{E}$ is an N-term approximation rate N^{-1}. We now know a benchmark result! Can we construct simple dictionaries which achieve this best-possible approximation rate??

Frames

We let Φ be a *frame*, i.e.,

$$A\|f\|_2^2 \leq \sum_{i \in I} |\langle f, \varphi_i \rangle|^2 \leq B\|f\|_2^2$$

with dual $\tilde{\Phi} = \{\tilde{\varphi}_i\}_{i \in I}$.

Frames

We let Φ be a *frame*, i.e.,

$$A\|f\|_2^2 \leq \sum_{i\in I} |\langle f, \varphi_i\rangle|^2 \leq B\|f\|_2^2$$

with dual $\tilde{\Phi} = \{\tilde{\varphi}_i\}_{i\in I}$.

Definition

For $(c_i)_{i\in\mathbb{N}}$, define its *weak ℓ^p-norm* by

$$\|(c_i)_{i\in\mathbb{N}}\|_{w\ell^p} := \inf\{C > 0 : \ |c_n^*| \leq Cn^{-1/p}\},$$

where c_n^* is a nondecreasing rearrangement of c_i.

Frames

We let Φ be a *frame*, i.e.,

$$A\|f\|_2^2 \leq \sum_{i\in I} |\langle f, \varphi_i\rangle|^2 \leq B\|f\|_2^2$$

with dual $\tilde{\Phi} = \{\tilde{\varphi}_i\}_{i\in I}$.

Definition

For $(c_i)_{i\in\mathbb{N}}$, define its *weak ℓ^p-norm* by

$$\|(c_i)_{i\in\mathbb{N}}\|_{w\ell^p} := \inf\{C > 0 : \ |c_n^*| \leq Cn^{-1/p}\},$$

where c_n^* is a nondecreasing rearrangement of c_i.

Remark

We have $\|(c_i)\|_{w\ell^p} \leq \|(c_i)\|_{\ell^p}$*:*

Frames

We let Φ be a *frame*, i.e.,

$$A\|f\|_2^2 \leq \sum_{i \in I} |\langle f, \varphi_i \rangle|^2 \leq B\|f\|_2^2$$

with dual $\tilde{\Phi} = \{\tilde{\varphi}_i\}_{i \in I}$.

Definition

For $(c_i)_{i \in \mathbb{N}}$, define its *weak ℓ^p-norm* by

$$\|(c_i)_{i \in \mathbb{N}}\|_{w\ell^p} := \inf\{C > 0 : \; |c_n^*| \leq Cn^{-1/p}\},$$

where c_n^* is a nondecreasing rearrangement of c_i.

Remark

We have $\|(c_i)\|_{w\ell^p} \leq \|(c_i)\|_{\ell^p}$:

$$\sum_{i=0}^{\infty} |c_i|^p = \sum_{i=0}^{\infty} |c_i^*|^p \geq \sum_{i=0}^{n} |c_i^*|^p \geq n|c_n^*|^p.$$

Nonlinear Approximation with Frames

Theorem

Let Φ be frame for L^2. Suppose that

$$\left\|(\langle f, \tilde{\varphi}_i \rangle)_{i \in I}\right\|_{w\ell^p} < \infty.$$

Then, with $s = 1/p - 1/2$ we have

$$f \in \mathcal{A}^s(\Phi)$$

and

$$\|f - \sum_{i \in I_N} \langle f, \tilde{\varphi}_i \rangle \varphi_i\|_2 \lesssim N^{-s},$$

where

$$I_N = \{N \text{ largest framecoefficients}\}.$$

proof

Nonlinear Approximation with Frames

Theorem

Let Φ be frame for L^2. Suppose that

$$\left\|(\langle f, \tilde{\varphi}_i\rangle)_{i\in I}\right\|_{w\ell^p} < \infty.$$

Then, with $s = 1/p - 1/2$ we have

$$f \in \mathcal{A}^s(\Phi)$$

and

$$\|f - \sum_{i\in I_N}\langle f, \tilde{\varphi}_i\rangle\varphi_i\|_2 \lesssim N^{-s},$$

where

$$I_N = \{N \textit{ largest framecoefficients}\}.$$

proof

$\rightsquigarrow$ coefficient sequeces in ℓ^p for small p leads to good compression! Compression works by simple coefficient thresholding!

Literature

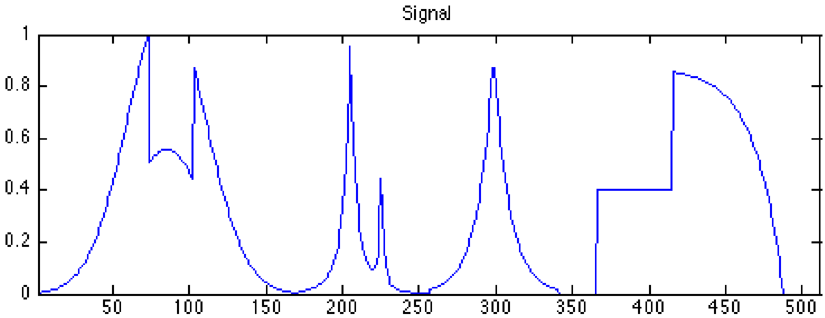
- DeVore. Nonlinear Approximation. *Acta Numerica* 7, 51-150 (1998).
- Donoho. Sparse components of images and optimal atomic decompositions. *Constructive Approximation* 17/3, 353-382 (2001).
- Grohs, Keiper, Kutyniok and Schaefer. Cartoon Approximation with α-Curvelets. preprint (2014), available from `www.math.ethz.ch/~pgrohs/research`.
- Berger. Rate-Distortion Theory. Wiley (1971).

2. Wavelets and Point-Singularities

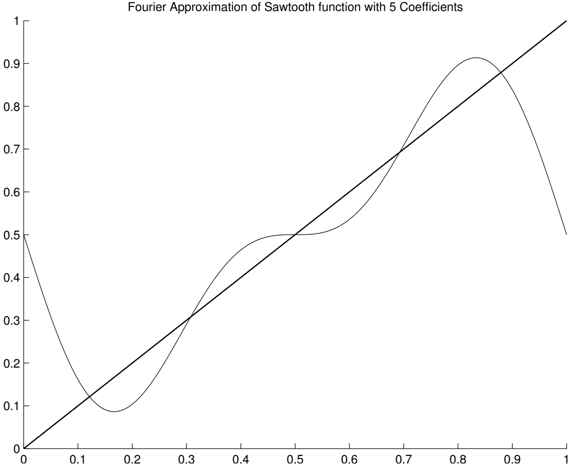
Wavelet Approximation of Point-singularities

Signal class

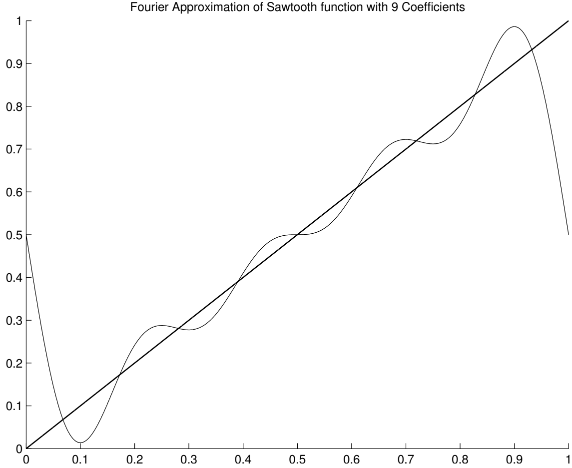
$$\mathcal{D}^\alpha = \{f \in L^2[0,1] : f = f_1\chi_{[0,a]} + \chi_{(a,1]}f_2,\ \|f_i\|_{C^\alpha} \le 1,\ a \in [0,1]\}.$$



Fourier Doesn't Work

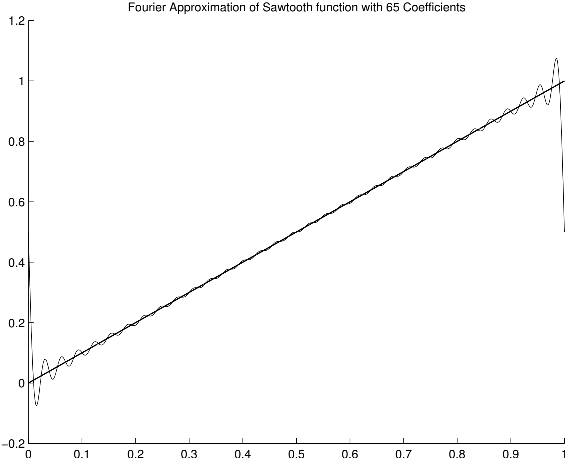


Fourier Doesn't Work

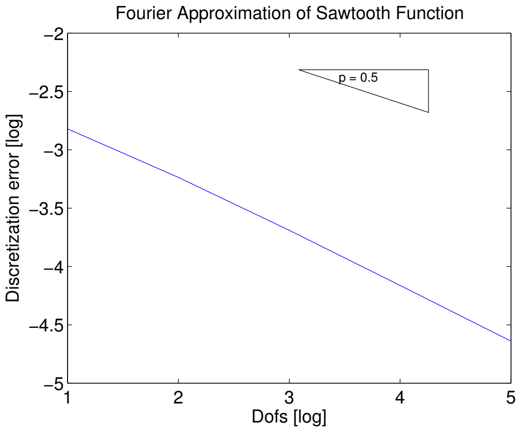


Fourier Doesn't Work

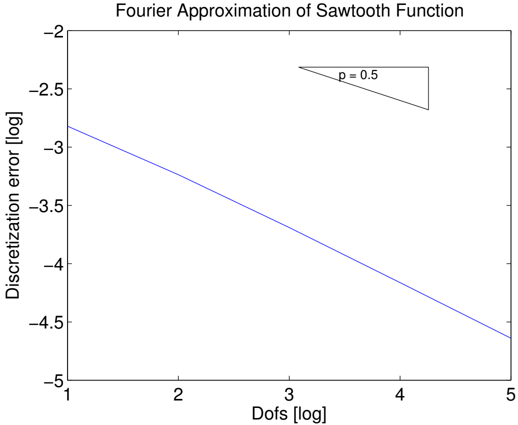
Fourier Doesn't Work



Fourier Doesn't Work



Fourier Doesn't Work



Theorem

We have

$$s^*(\mathcal{D}^\alpha, \mathit{Fourier}) = \frac{1}{2}.$$

proof

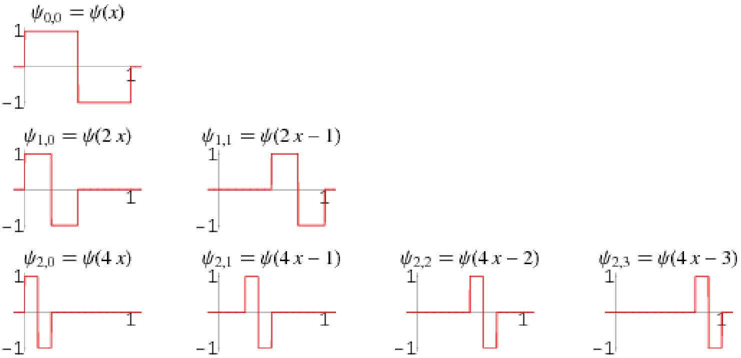
Wavelets

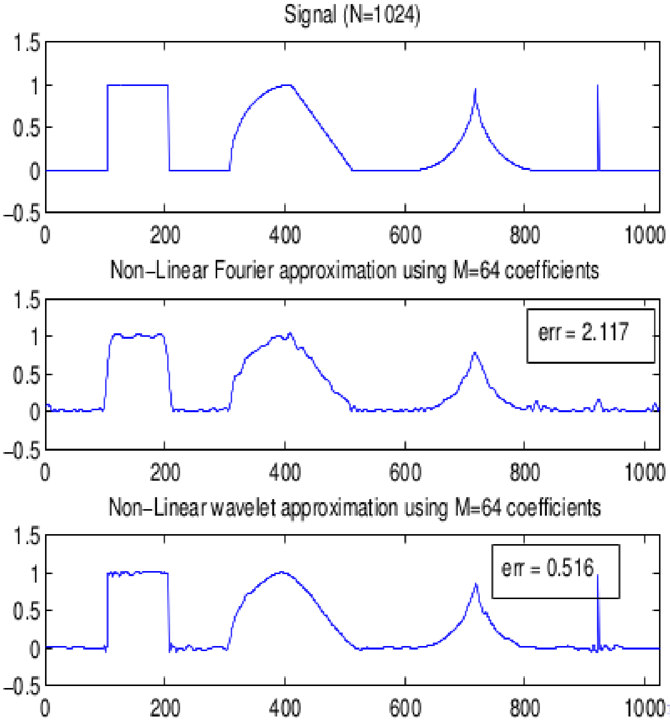
Dictionaries of the form

$$\mathcal{W}(\varphi,\psi,\alpha) := \Big\{\underbrace{\varphi(\cdot-\alpha k)}_{\varphi_k} :\ k\in\mathbb{Z}\Big\} \cup \Big\{\underbrace{2^{j/2}\psi(2^j\cdot-\alpha k)}_{\psi_{j,k}} :\ j\in\mathbb{N},\ k\in\mathbb{Z}\Big\}$$

Example

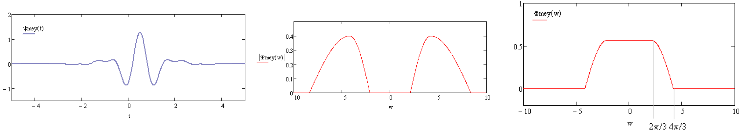
Haar wavelets with $\varphi=\chi_{0,1}$ *and* $\psi=\frac{1}{\sqrt{2}}(\varphi(2\cdot)-\varphi(2\cdot-1))$. $\mathcal{W}(\varphi,\psi,1)$ *constitutes an ONB for* L^2. *Smooth Wavelet ONBs with compactly supported* $\varphi,\ \psi \rightsquigarrow$ *Daubechies wavelets.*


$\psi_{0,0} = \psi(x)$
$\psi_{1,0} = \psi(2\,x)$
$\psi_{1,1} = \psi(2\,x - 1)$
$\psi_{2,0} = \psi(4\,x)$
$\psi_{2,1} = \psi(4\,x - 1)$
$\psi_{2,2} = \psi(4\,x - 2)$
$\psi_{2,3} = \psi(4\,x - 3)$
1
−1



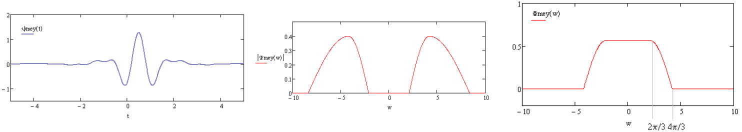
Signal (N=1024)
Non−Linear Fourier approximation using M=64 coefficients
err = 2.117
Non−Linear wavelet approximation using M=64 coefficients
err = 0.516
1.5
1
0.5
0
−0.5
0
200
400
600
800
1000

Meyer Wavelets



Left to right: Meyer Wavelet ψ, Fourier Trafo of Meyer Wavelet, Fourier Trafo of Meyer Scalingfunction φ.

Meyer Wavelets



Left to right: Meyer Wavelet ψ, Fourier Trafo of Meyer Wavelet, Fourier Trafo of Meyer Scalingfunction φ.

$$\hat{\psi}_{j,k}(\xi) = 2^{-j/2}\hat{\psi}(2^{-j}\xi)\exp(2\pi i 2^{-j}k\xi).$$

It follows that

$$\text{supp } \psi_{j,k} \sim 2^{-j}[k-a, k+a], \quad \text{supp } \hat{\psi}_{j,k} \sim [2^{j-1}, 2^{j+1}],$$

and thus $\langle f, \psi_{j,k}\rangle$ extracts frequency information in an annulus around 2^j and spacial information in an interval of width 2^{-j} around $2^{-j}k$.

Wavelets are Optimal for Point-Singularities

Theorem

Suppose that Φ is the ONB of Meyer Wavelets. Then

$$\sup_{f \in \mathcal{D}^\alpha} \|(\langle f, \psi_{j,k})_{j,k}\rangle\|_{w\ell^{2/(1+\alpha)}} < \infty.$$

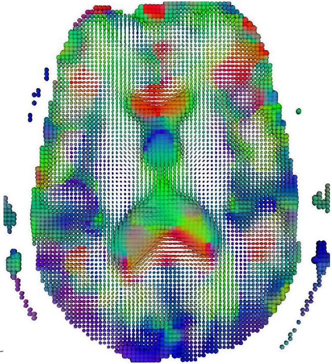
In particular,

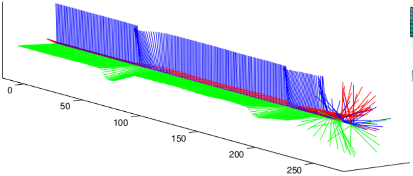
$$s^*(\mathcal{W}(\varphi, \psi, 1), \mathcal{C}^\alpha) = s^*(\mathcal{C}^\alpha) = \alpha.$$

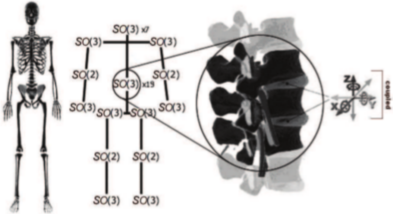
In other words, wavelets provide optimal approximation for functions with point-singularities, namely as efficient as if the singularity wasn't there! proof

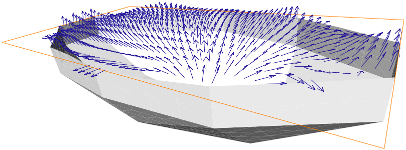
Detour: Wavelets for Manifold-Valued Data

Many data types possess additional structure!







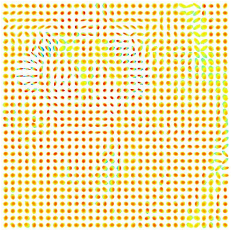


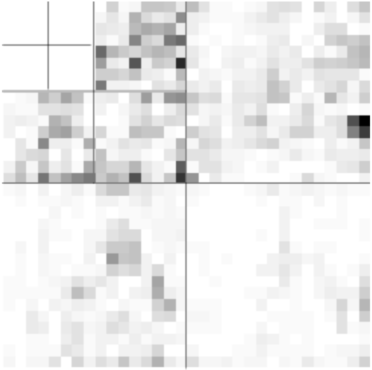
Detour: Wavelets for Manifold-Valued Data

Nonlinear wavelet-type transforms can also be defined for manifold-valued data. Wavelet coefficients are elements of the tangent bundle (UrRahmen et. al., G-Wallner).

Detour: Wavelets for Manifold-Valued Data

Nonlinear wavelet-type transforms can also be defined for manifold-valued data. Wavelet coefficients are elements of the tangent bundle (UrRahmen et. al., G-Wallner).





Detour: Wavelets for Manifold-Valued Data

Theorem (G 2012)

'Manifold-valued wavelets' are optimal for functions with point-singularities.

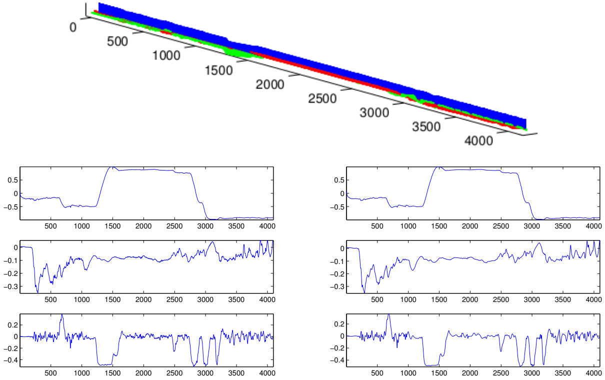
Detour: Wavelets for Manifold-Valued Data

Theorem (G 2012)

'Manifold-valued wavelets' are optimal for functions with point-singularities.

Proof complicated due to loss of linear structure.

Detour: Wavelets for Manifold-Valued Data



Left: Original entries of top 3 matrix entries of *SO*(3)-valued data, Right: 20:1 compression

2-D Wavelets

Dictionaries of the form

$$\mathcal{W}^{2D}(\varphi,\psi,\alpha) := \left\{ \underbrace{\varphi^{2D}(\cdot-\alpha k)}_{\varphi^{2D}_k} : \; k\in\mathbb{Z}^2 \right\} \cup \left\{ \underbrace{2^j\psi^{\varepsilon}(2^j\cdot-\alpha k)}_{\psi^{\varepsilon}_{j,k}} : \; j\in\mathbb{N},\; k\in\mathbb{Z}^2,\; \varepsilon\in\{0,1\}^2\setminus(0,0) \right\},$$

where

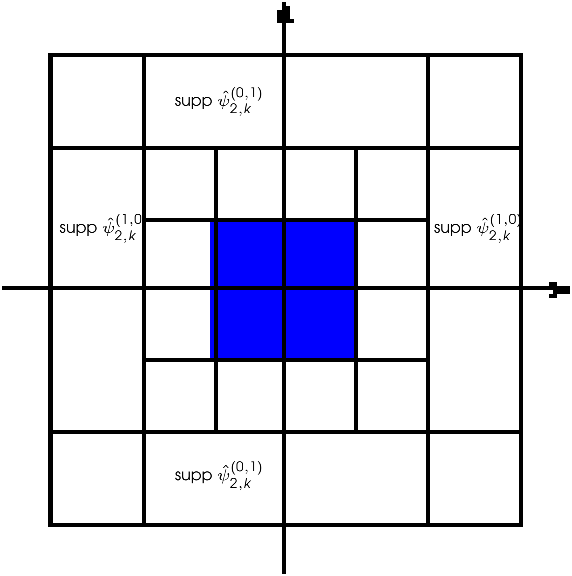
$$\psi^{(1,1)}(x,y)=\psi(x)\psi(y),\;\psi^{(1,0)}(x,y)=\psi(x)\varphi(y),\;\psi^{(0,1)}(x,y)=\psi(y)\varphi(x),\;\varphi^{2D}(x,y)=\varphi(x)\varphi(y),$$

with ψ wavelet and φ scaling function.

Theorem

Suppose $\mathcal{W}(\varphi,\psi,\alpha)$ is ONB or frame for $L^2(\mathbb{R})$. Then $\mathcal{W}^{2D}(\varphi,\psi,\alpha)$ is ONB or frame for $L^2(\mathbb{R}^2)$.

Fourier Partitioning of $2D$ Meyer Wavelets



JPEG2000

The image compression standard JPEG2000 is based on

JPEG2000

The image compression standard JPEG2000 is based on

1. Decomposing a given image f in a wavelet $2D$ wavelet basis,

JPEG2000

The image compression standard JPEG2000 is based on

1. Decomposing a given image f in a wavelet $2D$ wavelet basis,
2. Coefficient thesholding,

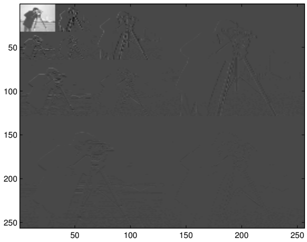
JPEG2000

The image compression standard JPEG2000 is based on

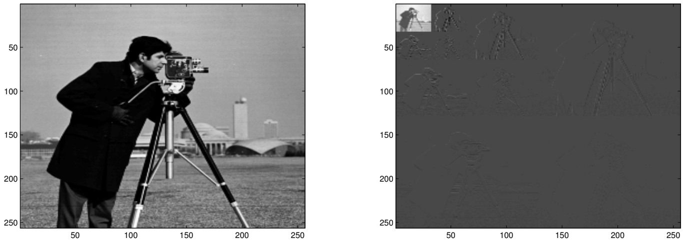
1. Decomposing a given image f in a wavelet $2D$ wavelet basis,
2. Coefficient thesholding,
3. Quantization.

Recall...


50
100
150
200
250
50
100
150
200
250


50
100
150
200
250
50
100
150
200
250

RECALL...



Is this really optimal??

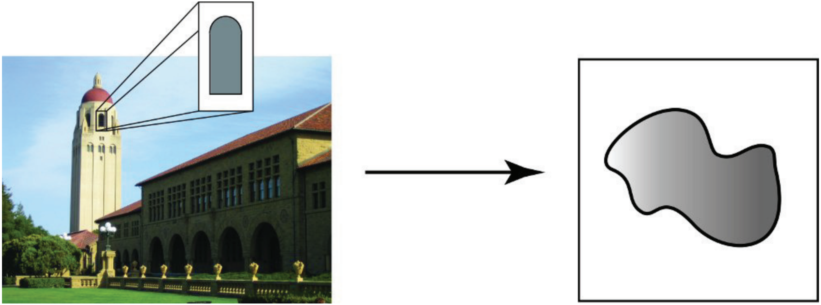
Recall...

Definition (Donoho (2001))

The set $\mathcal{E}$ of cartoon images is given by

$$\mathcal{E} := \{f = f_0 + \chi_B f_1\},$$

where $f_0, f_1 \in C^2([0,1]^2)$ and χ_B is the indicator function of $B \subset [0,1]^2$ with C^2 boundary.



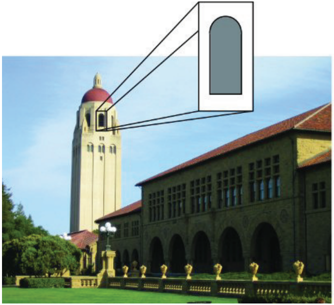
Recall...

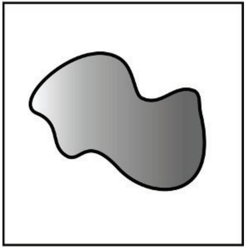
Definition (Donoho (2001))

The set $\mathcal{E}$ of cartoon images is given by

$$\mathcal{E} := \{f = f_0 + \chi_B f_1\},$$

where $f_0, f_1 \in C^2([0,1]^2)$ and χ_B is the indicator function of $B \subset [0,1]^2$ with C^2 boundary.





Recall benchmark approximation rate for $\mathcal{E}$ is N^{-1}!

Limitations of Wavelets

Theorem

We have that

$$s^* \left(\mathcal{E}, \mathcal{W}^{2D}(\varphi, \psi, \alpha) \right) = \frac{1}{2}.$$

This is one magnitude short of the optimal rate N^{-1}*.*

▸ proof

Limitations of Wavelets

Theorem

We have that

$$s^* \left(\mathcal{E}, \mathcal{W}^{2D}(\varphi, \psi, \alpha)\right) = \frac{1}{2}.$$

This is one magnitude short of the optimal rate N^{-1}*.*

proof

Can we find better dictionaries??

Literature

- Daubechies. Ten Lectures on Wavelets. SIAM (1992)
- Mallat. A Wavelet Tour of Signal Processing. Academic Press (2008).

End of Part I

Appendix: Proofs and Additional Material

Proof Sketch: Suppose that $\mathfrak{h}$ is a hypercube of dimension m and sidelength δ.
Every element $h \in \mathfrak{h}$ can be encoded with m bits.
Let $h \in \mathfrak{h}$ arbitrary and $\tilde{h}_N$ its best N-term approximation.
Assume for simplicity that $\tilde{h}_N$ can be encoded with N bits (up to a log-factor this is possible by the polynomial depth search assumption).
Let $\hat{h}_N$ be the orthogonal projection of $\tilde{h}_N$ onto $\mathfrak{H}$.
Then we get the following coding scheme for $\{0,1\}^m$, i.e. a mapping Code_m^N from $\{0,1\}^m$ to $\{0,1\}^N$:

$$(\epsilon_i)_{i=1}^m \mapsto h = f_0 + \sum_{i=1}^m \epsilon_i \psi_i \in \mathfrak{h} \quad \underbrace{\mapsto}_{N\text{-term approximation}} \quad (\tau_i)_{i=1}^N,$$

and corresponding decoding map
$\mathrm{Decode}_m^N \{0,1\}^N \to \{0,1\}^m$

$$(\tau_i)_{i=1}^N \quad \underbrace{\mapsto}_{N\text{-term reconstruction}} \quad \tilde{h}_N \quad \underbrace{\mapsto}_{\text{projection on } \mathfrak{h}} \quad \hat{h}_N.$$

Result from Rate-Distortion-Theory: For $N < m/3$ we have at least one bad $(\epsilon_i)_{i=1}^m$ (corresponding to $h \in \mathfrak{h}$) such that at least order m bits in the reconstruction are false:

$$\mathrm{dist}_{\mathrm{Hamming}}((\epsilon_i)_{i=1}^m, \mathrm{Decode}_m^N \circ \mathrm{Encode}_m^N(\epsilon_i)_{i=1}^m) \gtrsim m.$$

But this implies that

$$\|h - \tilde{h}_N\|_2^2 \geq \|h - \hat{h}_N\|_2^2 \gtrsim \delta^2 m.$$

Let now $N_k := m_k/3$, where m_k is the dimension of our sequence of embedded hypercubes of sidelengths

$$\delta_k \gtrsim m_k^{-1/p} \gtrsim N_k^{-1/p}.$$

Putting this into the above equation we get

$$\|h - \tilde{h}_{N_k}\|_2^2 \gtrsim N_k^{1-2/p}.$$

▸ return to talk

Proof Sketch: Only $d = 1$ (generalization is easy).
We need to construct a family of embedded hypercubes $(\mathfrak{h}_k$ of dimensions m_k and sidelengths δ_k, where

$$m_k \gtrsim \delta_k^{-p}, \quad p = \frac{2}{2\alpha + 1}.$$

Start with $\psi \in \mathcal{C}^\alpha$ and

$$\psi_{i,k} := k^{-\alpha}\psi(k \cdot -i), \quad i, k \in \mathbb{N}.$$

We have

$$\|\psi_{i,k}\|_{C^\alpha} \leq \|\psi\|_{C^\alpha},$$

so $\psi_{i,k} \in \mathcal{C}^\alpha$.
Furthermore, the supports of different $\psi_{i,k}$, $\psi_{i',k}$ are different, so for a fixed k, the system $(\psi_{i,k})_i$ is orthogonal.
A simple computation shows that

$$\|\psi_{i,k}\|_2^2 = k^{-1-2\alpha}\|\psi\|_2^2.$$

So the system

$$\mathfrak{h}_k = \{\sum_{i=1}^{k} \epsilon_i \psi_{i,k}\}$$

is an embedded hypercube of dimension $m_k = k$ with sidelength $\delta_k \sim k^{-(\frac{1}{2}+\alpha)}$. ▸ return to talk

Proof Sketch: Suppose for simplicity that Φ is an ONB (general case not more difficult). Let $c_i = \langle f, \varphi_i \rangle$ and assume that

$$\|(c_i)_i\|_{w\ell^p} < \infty.$$

Let $(c_{i_j})_{j\in\mathbb{N}}$ be a nonincreasing rearrangement of $(c_i)_i$. The best N-term approximation of f is given by

$$f_N = \sum_{j=1}^{N} c_{j_i}\varphi_{i_j}.$$

We need to show that

$$\|f - f_N\|_2 \lesssim N^{-(1/p-1/2)}.$$

Since Φ is an ONB we have

$$\|f - f_N\|_2 = \left(\sum_{j=N+1}^{\infty} c_{i_j}^2\right)^{1/2}.$$

Since $\|(c_i)_i\|_{w\ell^p} < \infty$ we have that

$$\left(\sum_{j=N+1}^{\infty} j^{-2/p}\right)^{1/2} = N^{-1/p+1/2}\left(N^{2/p-1}\sum_{j=N+1}^{\infty} j^{-2/p}\right)^{1/2}.$$

So we need to show that

$$\left(N^{2/p-1}\sum_{j=N+1}^{\infty} j^{-2/p}\right)^{1/2} < \infty.$$

Suppose for simplicity that $N = 2^l$. Then

$$\sum_{j=N+1}^{\infty} j^{-2/p} \le \sum_{m=l}^{\infty} F_m,$$

where

$$F_m := \sum_{i=2^{m-1}}^{2^m} i^{-2/p} \lesssim \sum_{i=2^{m-1}}^{2^m} 2^{-2m/p} \lesssim 2^{-(2/p-1)m}.$$

So

$$\sum_{m=l}^{\infty} F_m \lesssim 2^{-(2/p-1)l}.$$

This implies the desired estimate. return to talk

Proof: Let $f(x) = x$, $x \in [0, 1]$. Then

$$\hat{f}(j) = \int_0^1 x \exp(-2\pi ijx)dx = \left[x \frac{-1}{2\pi ij} \exp(-2\pi ijx)\right]_0^1 + \int_0^1 \frac{1}{2\pi ij} \exp(-2\pi ijx)dx = -\frac{1}{2\pi ij}.$$

So the nonincreasing rearrangement of $c_j := \langle f, \exp(2\pi ij\cdot)\rangle$ is of the form $c_n^* \sim n^{-1}$ which lies in $w\ell^p$ for $p \geq 1$, which corresponds to a best N-term approximation order $\frac{1}{2}$.

▸ return to talk

Proof Sketch: We will show that for any $f \in \mathcal{D}^\alpha$ we have for every $p > \frac{1}{\alpha+1/2}$ that the coefficient sequence $\langle f, \psi_{j,k}\rangle$ lies in ℓ^p. This is almost the result we want.
We suppose that $\frac{d^l}{dt^l}\hat{\psi}(0) = 0$ for all $l = 1, \dots, \lfloor\alpha\rfloor$, which is definitely the case for Meyer wavelets and which is equivalent to

$$\int p(x)\psi(x)dx = 0 \quad \text{for all polynomials } p \text{ of degree} \leq \lfloor\alpha\rfloor. \tag{1}$$

Furthermore we suppose that

$$\text{supp } \psi_{j,l} \subset 2^{-j}[k-a, k+a] \tag{2}$$

for some fixed $a > 0$.
Now fix a scale j. For any bounded f we have the estimate

$$\int f(x)\psi_{j,k}(x)dx = \int_{2^{-j}[k-a,k+a]} f(x)2^{j/2}\psi(2^j x - k)dx \lesssim 2^{-j/2}. \tag{3}$$

For $f \in C^\alpha$ we have with a Taylor polynomial $p_{j,k}(x)$ of f around $2^{-j}k$ that

$$\int f(x)\psi_{j,k}(x)dx = \int_{2^{-j}[k-a,k+a]} \left(p_{j,k}(x) + O(2^{-\alpha j})\right) 2^{j/2}\psi(2^j x - k)dx.$$

Using the vanishing-moment-property (1) and the substitution $y = 2^j x$, we get that

$$\int f(x)\psi_{j,k}(x)dx \lesssim 2^{-(\alpha+1/2)j} \tag{4}$$

for all $f \in C^\alpha$.
Let us split the index set $\mathbb{Z}$ into

$$K_1 = \{k \in \mathbb{Z} : \text{ supp } \psi_{j,k} \text{ intersects the singularity}\}$$

and

$$K_1 = \{k \in \mathbb{Z} : \text{ supp } \psi_{j,k} \text{ intersects the } [0,1] \text{ and not the singularity}$$

By (2) we get that

$$\#K_1 \lesssim 1 \quad \text{and} \quad \#K_2 \lesssim 2^j. \tag{5}$$

For $k \in K_1$ the wavelet coefficients $c_{j,k} := \langle f, \psi_j, k \rangle$ satisfy $|c_{j,k}| \lesssim 2^{-j/2}$ by (3) and for $k \in K_2$ we have, by (4), that $|c_{j,k}| \lesssim 2^{-j(\alpha+1/2)}$.
Hence we get that

$$\sum_k |c_{j,k}|^p \lesssim \sum_{K_1} 2^{-pj/2} + \sum_{K_1} 2^{-pj(\alpha+1/2)} \lesssim 2^{-pj/2} + 2^{-j(p\alpha+p/2-1)},$$

by (5). Now suppose that $p > \frac{1}{\alpha+1/2}$. Then $\tau := \min(p/2, p\alpha + p/2 - 1) > 0$ and

$$\sum_j \sum_k |c_{j,k}|^p \lesssim \sum_j 2^{-\tau j} < \infty.$$

▸ return to talk

Proof Sketch: Fix a scale j. Every wavelet $\psi_{j,k}^{\varepsilon}$ is approximately supported in a square of sidelength $\sim 2^{-j}$.
We need approximately 2^j wavelets to cover the singularity curve.
For those 2^j wavelets the corresponding coefficients $c_{j,l}^{\varepsilon} := \langle f, \psi_{j,l}^{\varepsilon} \rangle$ satisfy

$$|c_{j,l}^{\varepsilon}| \geq 2^{-j}.$$

So at each scale we have $\sim 2^j$ coefficients of magnitude 2^{-j}.
Therefore we have

$$\sum_{\varepsilon,k} |c_{j,l}^{\varepsilon}|^p \geq \sum_{\varepsilon,k} 2^{-pj} \geq 2^{(1-p)j}.$$

So the wavelet coefficients can only lie in ℓ^p if $p > 1$, which implies best N-term rate of order $1/2$. ▸ return to talk