The Success Story of Wavelets: Honoring Yves Meyer's Abel Prize

Gitta Kutyniok and Philipp Petersen

(Technische Universität Berlin)

SFB-Seminar, Berlin, May 11, 2017

Applied Functional Analysis Group
Technische Universität Berlin

1. Introducing Yves Meyer

2. Research on Wavelets

3. Further Developments

Early years:

- Born 19 July 1939 in Paris.
- Growing up in Tunis on the North African coast.

"As a child I was obsessed by the desire of crossing the frontiers between these distinct ethnic groups."

Early years:

- Born 19 July 1939 in Paris.
- Growing up in Tunis on the North African coast.

"As a child I was obsessed by the desire of crossing the frontiers between these distinct ethnic groups."

- 1957: Entering the École normale supérieure in Paris by coming first in the entrance examination.
- 1966: PhD under the supervision of Jean-Pierre Kahane.
- 1960–1963: Teaching at Prytanée national militaire.

"A good teacher needs to be much more methodical and organised than I was."
"To do research is to be ignorant most of the time and often to make mistakes."

Academic positions:

- 1963–1966: Université de Strasbourg (Teaching Assistant).
- 1966–1980: Université Paris-Sud (Professor).
- 1980–1986: École Polytechnique (Professor).
- 1986–1995: Université Paris-Dauphine (Professor).
- 1995–1999: Centre national de la recherche scientifique (Senior Researcher).
- 1999–2003: École Normale Supérieure de Cachan (Professor).
- Since 2003: École Normale Supérieure de Cachan (Professor Emeritus).

"You must dig deeply into your own self in order to do something as difficult as research in mathematics."
"You need to believe that you possess a treasure hidden in the depths of your mind, a treasure which has to be unveiled."

Results in the following areas:

- Ditkin sets.
- Diophantine approximations.

Definition (Meyer; 1972): A set $\Lambda \subset \mathbb{R}^n$ is a model set, if there exist a finite set $F \subset \mathbb{R}^n$ and a constant $C > 0$ such that

(1) $\Lambda - \Lambda \subset \Lambda + F$, and

(2) $\inf_{\lambda \in \Lambda} |x - \lambda| \leq C$ for all $x \in \mathbb{R}^n$.

Results in the following areas:

- Ditkin sets.
- Diophantine approximations.

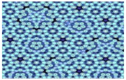
Definition (Meyer; 1972): A set $\Lambda \subset \mathbb{R}^n$ is a model set, if there exist a finite set $F \subset \mathbb{R}^n$ and a constant $C > 0$ such that

(1) $\Lambda - \Lambda \subset \Lambda + F$, and

(2) $\inf_{\lambda \in \Lambda} |x - \lambda| \leq C$ for all $x \in \mathbb{R}^n$.

Theorem (Meyer; 1972):
If Λ is a model set and if $\theta\Lambda \subset \Lambda$, then θ is a Pisot or a Salem number.
(A Pisot number is a real algebraic integer greater than 1 all of whose Galois conjugates are less than 1 in absolute value.)

Impact: Paved the road to the mathematical theory of quasicrystals.
$\rightsquigarrow$ Non-period patterns in quasicrystals can be identified with specific model sets.



Singular Integral Operators (1974–1984)

Definition:

(1) A function $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a Calderón-Zygmund kernel, if it satisfies certain decay and cancellation conditions.

(2) T is a singular integral operator of non-convolution type ass. to K, if

$$\int_{\mathbb{R}^n} g(x)T(f)(x)dx = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} g(x)K(x,y)f(y)dydx,$$

whenever f and g are smooth and have disjoint support.

(3) Such a T is a Calderón–Zygmund operator when it is bounded on L^2.

Singular Integral Operators (1974–1984)

Definition:

(1) A function $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a Calderón-Zygmund kernel, if it satisfies certain decay and cancellation conditions.

(2) T is a singular integral operator of non-convolution type ass. to K, if

$$\int_{\mathbb{R}^n} g(x)T(f)(x)dx = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} g(x)K(x,y)f(y)dydx,$$

whenever f and g are smooth and have disjoint support.

(3) Such a T is a Calderón–Zygmund operator when it is bounded on L^2.

Theorem (Coifman, McIntosh, Meyer; 1982):
Calderón-Zygmund operators are bounded for Lipschitz curves or surfaces.

Impact:

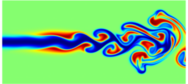
$\rightsquigarrow$ The $T(1)$ theorem (David, Journé; 1984).
$\rightsquigarrow$ Solution of the Dirichlet problem in any Lipschitz domain (Fabes; 1988).
$\rightsquigarrow$ Solution of the Kato's conjecture (Auscher, Tchamitchian; 1998).

...see Philipp's talk...

Switching to yet another area:

Farge, Battle, and Federbush suggested that wavelet transforms might yield better results than previous methods for numerical approximation of turbulent flow.

$\rightsquigarrow$ This got Yves interested in Navier-Stokes equations!



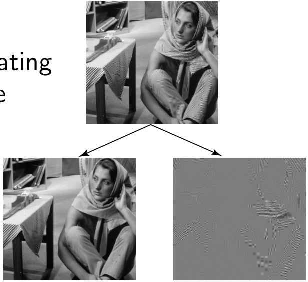
Theorem (Cannone, Meyer, Planchon, Lemarié-Rieusset; 1994):

There exists a global solution to the Navier-Stokes equation in the space $C(\mathbb{R}^+, L^2(\mathbb{R}^3))$, when the initial condition is oscillating in the sense that it belongs to a Besov space of negative order.

Proof: ...uses wavelet-like decompositions.

Current range of topics:

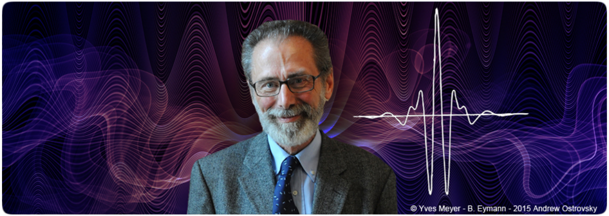
- Imaging Sciences.
  - Theoretical and numerical analysis of separating geometric ("cartoon") content from texture (Meyer; 2006).



- Compressed Sensing
  - Deterministic construction of an optimal sensing system based on the theory of model sets. (Matei, Meyer; 2009).

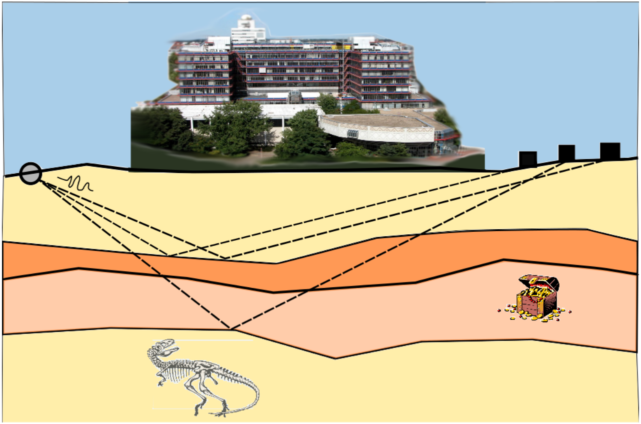
Awards

- 1970/1983/1990: Invited Speaker at ICM.
- 1993: Member of the Académie des Sciences.
- 2010: Carl Friedrich Gauss Prize.
- 2012: Fellow of the American Mathematical Society.
- 2017: Abel Prize for his pivotal role in developing the mathematical theory of wavelets.



Meyer's Motivation

Seismic reflection method:



Problem: Wavelength sometimes much larger than layers.

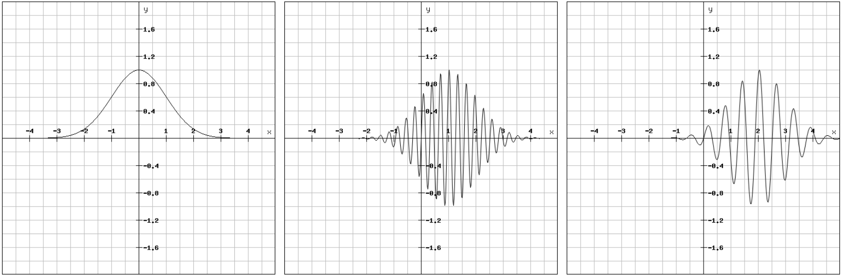
Meyer's Motivation

Gabor transform: At that time standard wave model is given by shifted and modulated Gaussians $g = e^{-x^2/2}$

$$G_{t,v}(x) = e^{itx} g(x - v), \quad t, v, x \in \mathbb{R},$$

which goes back to a work of D. Gabor (1946).

All $G_{t,v}$ have fixed length.



Morlet's idea: Make also the length of these functions variable.

GEOPHYSICS, VOL. 47, NO. 2 (FEBRUARY 1982); P. 203–221, 33 FIGS.

Wave propagation and sampling theory—Part I: Complex signal and scattering in multilayered media

J. Morlet*, G. Arens‡, E. Fourgeau*, and D. Giard‡

ABSTRACT

From experimental studies in digital processing of seismic reflection data, geophysicists know that a seismic signal does vary in amplitude, shape, frequency and phase, versus propagation time. To enhance the resolution of the seismic reflection method, we must investigate these variations in more detail.

We present quantitative results of theoretical studies on propagation of plane waves for normal incidence, through perfectly elastic multilayered media.

As wavelet shapes, we use zero-phase cosine wavelets modulated by a Gaussian envelope and the corresponding complex wavelets. A finite set of such wavelets, for an appropriate sampling of the frequency domain, may be taken as the basic wavelets for a Gabor expansion of any signal or trace in a two-dimensional (2-D) domain (time and frequency). We can then compute the wave propagation using complex functions and thereby obtain quantitative results including energy and phase of the propagating signals. These results appear as complex 2-D functions of time and frequency, i.e., as "instantaneous frequency spectra."

Choosing a constant sampling rate on the logarithmic scale in the frequency domain leads to an appropriate sampling method for phase preservation of the complex signals or traces. For this purpose, we developed a Gabor expansion involving basic wavelets with a constant time duration/mean period ratio.

Three different cases appear for signal scattering, depending upon the value of the ratio wavelength of the

SIAM J. MATH. ANAL.
Vol. 15, No. 4, July 1984

[gsmo84]
120

DECOMPOSITION OF HARDY FUNCTIONS INTO SQUARE INTEGRABLE WAVELETS OF CONSTANT SHAPE*

A. GROSSMANN† AND J. MORLET‡

Abstract. An arbitrary square integrable real-valued function (or, equivalently, the associated Hardy function) can be conveniently analyzed into a suitable family of square integrable wavelets of constant shape, (i.e. obtained by shifts and dilations from any one of them.) The resulting integral transform is isometric and self-reciprocal if the wavelets satisfy an "admissibility condition" given here. Explicit expressions are obtained in the case of a particular analyzing family that plays a role analogous to that of coherent states (Gabor wavelets) in the usual L_2-theory. They are written in terms of a modified Γ-function that is introduced and studied. From the point of view of group theory, this paper is concerned with square integrable coefficients of an irreducible representation of the nonunimodular $ax+b$-group.

1. Introduction.

1.1. It is well known that an arbitrary complex-valued square integrable function $\psi(t)$ admits a representation by Gaussians, shifted in direct and Fourier transformed space. If $g(t)=2^{-1/2}\pi^{-3/4}e^{-t^2/2}$ and t_0, ω_0 are arbitrary real, consider

$$g^{(t_0,\omega_0)}(t)=e^{-i\omega_0 t_0/2}e^{i\omega_0 t}g(t-t_0) \tag{1.1}$$

and form the inner product

Morlet's wavelets:

$$g_{u,v}(x) = 2^u g(2^u x - v),$$

$g \in L^2(\mathbb{R})$ satisfies certain admissibility conditions, usually a shifted Gaussian. These functions have variable length.

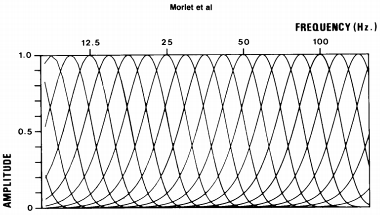
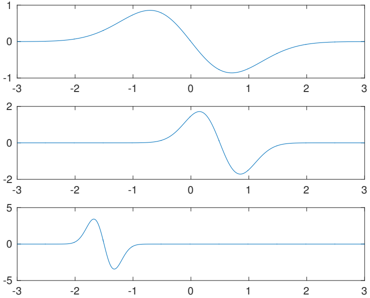


FIG. 2. Set of Gabor basic wavelets used in the Gabor expansion (practical example).



"A 'wavelet' oscillates for 'a little while' like a wave, but is then localized by damping"

Wavelet transform:

$$\mathcal{W}(f)(u,v) = 2^{u/2}\langle f, g_{u,v}\rangle = 2^{u/2}\int_{-\infty}^{\infty} f(x)g(2^u x - v)dx,$$

for $f \in L^2(\mathbb{R})$.

Reconstruction and energy preservation: Under certain assumptions on g:

$$\|f\|_2 = \|\mathcal{W}f\|_2,$$

and f can be uniquely recovered from $\mathcal{W}(f)$ by

$$f(x) = \int\int 2^{u/2} g(2^u x - v)\mathcal{W}(f)(u,v)dudv.$$

Calderon's Reproducing Formula

Meyer and operators: Yves Meyer has been working on singular integral operators until this point.

Calderon's identity: If g satisfies an admissibility condition, then, with $g_u(x) = u^{-1} g(x/u)$, we have for all $f \in L^2(\mathbb{R})$

$$f = \int f * g_u * g_u du/u,$$

where $f * g(x) = \int f(y) g(x-y) dy$.

Compare with the reproducing formula of Morlet's wavelets:

$$\begin{aligned} f(x) &= \int \int 2^{u/2} g(2^u x - v) \mathcal{W}(f)(u,v) dv du \\ &= \int \int 2^{u/2} g(2^u x - v) \int f(y) 2^{u/2} g(2^u y - v) dy dv du. \end{aligned}$$

Discrete Wavelet Systems

Discretization

Recall: Energy conservation:

$$\|f\|_2 = \|\mathcal{W}f\|_2$$

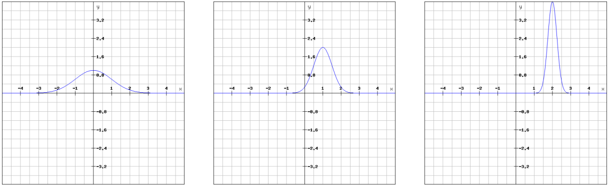
and simple reconstruction like an orthonormal basis.

Discrete wavelets: Can one construct a discrete wavelet system

$$W(\psi) = \{\psi_{j,m} := 2^j \psi(2^j x - m) : (j,m) \in \Lambda\},$$

where $\Lambda \subset \mathbb{R}^2$ is a lattice and $W(\psi)$ is an ONB such that

$$f = \sum_{(j,m)\in\Lambda} \langle f, \psi_{j,m} \rangle \psi_{j,m}?$$

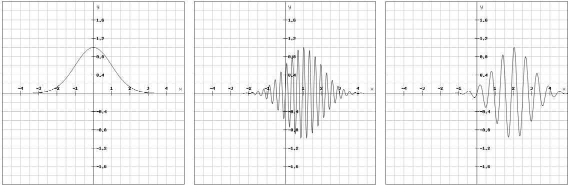


To understand the research on wavelet bases it is impotant to understand what was the standard then.

Gabor systems:

$$\Gamma_{\alpha,\beta}(g) = \{g_{n,m}(x) := e^{2\pi i \alpha n x} g(x - \beta m), \quad m, n \in \mathbb{Z}\},$$

were $\alpha, \beta > 0$, $g \in L^2(\mathbb{R})$.



Typical question: What conditions do we need to impose on α, β, g so that $\Gamma_{\alpha,\beta}(g)$ is an orthonormal basis?

Balian-Low Theorem

Orthonomal bases of Gabor systems: It turns out all orthonormal bases or Gabor systems have bad time-frequency localization.

Theorem: (Balian-Low Theorem)
If $\alpha, \beta > 0$, $g \in L^2(\mathbb{R})$ and $\Gamma_{\alpha,\beta}(g)$ is an orthonormal basis, then

$$\int x^2 |g(x)|^2 dx = \infty \quad \text{or} \quad \int \xi^2 |\hat{g}(\xi)|^2 d\xi = \infty.$$

Meyer wanted to prove the same result for orthonormal wavelet bases...

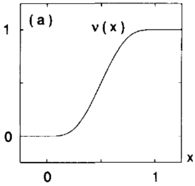
… but he "failed":

Meyer found the following function ψ:

$$\hat{\psi} = \begin{cases} (2\pi)^{-1/2} e^{i\xi/2} \sin\left[\frac{\pi}{2}\vartheta\left(\frac{3}{2\pi}|\xi| - 1\right)\right], & \text{if } \frac{2\pi}{3} \le |\xi| \le \frac{4\pi}{3} \\ (2\pi)^{-1/2} e^{i\xi/2} \cos\left[\frac{\pi}{2}\vartheta\left(\frac{3}{4\pi}|\xi| - 1\right)\right], & \text{if } \frac{4\pi}{3} \le |\xi| \le \frac{8\pi}{3} \\ 0 & \text{otherwise,} \end{cases}$$

where ϑ is a function satisfying

$$\vartheta(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1 & \text{if } x \ge 0, \end{cases}$$



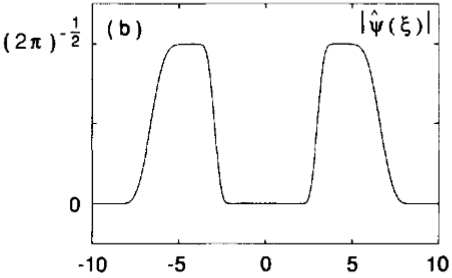
with the additional property

$$\vartheta(x) + \vartheta(1-x) = 1.$$

... but he "failed":

Meyer found the following function ψ:

$$\hat{\psi} = \begin{cases} (2\pi)^{-1/2} e^{i\xi/2} \sin\left[\frac{\pi}{2}\vartheta\left(\frac{3}{2\pi}|\xi| - 1\right)\right], & \text{if } \frac{2\pi}{3} \leq |\xi| \leq \frac{4\pi}{3} \\ (2\pi)^{-1/2} e^{i\xi/2} \cos\left[\frac{\pi}{2}\vartheta\left(\frac{3}{4\pi}|\xi| - 1\right)\right], & \text{if } \frac{4\pi}{3} \leq |\xi| \leq \frac{8\pi}{3} \\ 0 & \text{otherwise,} \end{cases}$$



Using the interplay between the phase of ψ and the assumption

$$\vartheta(x) + \vartheta(1-x) = 1$$

"quasi-miraculous cancellations" (Ingrid Daubechies) yield

1. $\|\psi\| = 1$
2. $\sum_{j,m} |\langle f, \psi_{j,m}\rangle|^2 = \|f\|^2$,

and, hence, that the system

$$\{\psi_{j,m} = 2^j \psi(2^j x - m) : j, m \in \mathbb{Z}\}$$

is an ONB for $L^2(\mathbb{R})$.

Not by Coincidence! $\rightsquigarrow$ multiresolution approximation

Multiresolution Approximation

Multiresolution Approximation

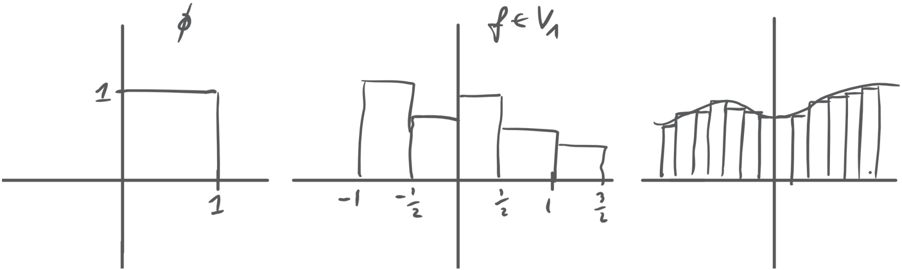
Multiresolution approximation: Formalization of approximation of functions by nested spaces which are connected by dyadic scaling of their elements.

Definition: (Mallat, Meyer; 1988)
A multiresolution approximation (MRA) of $L^2(\mathbb{R})$ is a nested sequence V_j, $j \in \mathbb{Z}$ of closed linear subspaces of $L^2(\mathbb{R})$ with the properties:

- $\bigcap_{-\infty}^{\infty} V_j = \{0\}$, and $\bigcup_{-\infty}^{\infty} V_j$ is dense in $L^2(\mathbb{R})$;
- for all $j \in \mathbb{Z}$ we have that $f \in V_j$ if and only if $f(2\cdot) \in V_{j+1}$;
- for all $m \in \mathbb{Z}$ we have that $f \in V_j$ if and only if $f(\cdot - m) \in V_j$;
- there exists a function $\phi \in V_0$ such that $\{\phi(\cdot - m) : m \in \mathbb{Z}\}$ is an orthonormal basis of the space V_0.

An example: The Haar multiresolution approximation

- Haar scaling function: $\phi = \chi_{[0,1]}$.
- $V_0 = \text{span}\{\phi(\cdot - m) : m \in \mathbb{Z}\}$.
- $V_j = \text{span}\{\phi(2^j \cdot - m) : m \in \mathbb{Z}\}$.



Wavelet spaces: The nestedness $V_j \subset V_{j+1}$ allows us to define the "wavelet spaces" W_j:

$$V_{j+1} = W_j \oplus V_j.$$

As a consequence, we have that

$$L^2(\mathbb{R}) = V_0 \oplus W_1 \oplus W_2 \oplus \dots.$$

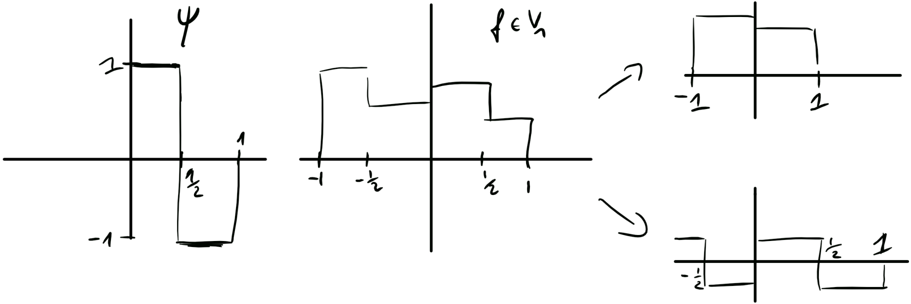
- Assume, that there exists ψ such that $\{\psi(\cdot - m) : m \in \mathbb{Z}\}$ is an ONB for W_0.
- Then, the dyadic scaling relation yields that $\{2^{j/2}\psi(2^j \cdot -m) : m \in \mathbb{Z}\}$ is an ONB for W_j for all $j \in \mathbb{N}$.
- We obtain that $\{2^{j/2}\psi(2^j \cdot -m) : m \in \mathbb{Z}, j \in \mathbb{N}\} \cup \{\phi(\cdot - m) : m \in \mathbb{Z}\}$ is an ONB for $L^2(\mathbb{R})$.

Multiresolution Approximation

Back to our example:

What is W_0 for the Haar multiresolution approximation?

- The Haar wavelet $\psi = \chi_{[0,1/2]} - \chi_{[1/2,1]}$.
- $\langle \phi(\cdot - m), \psi(\cdot - m) \rangle = 0$ is clear.
- Define $W_0 := \text{span}\{\psi(\cdot - m) : m \in \mathbb{Z}\}$.
- $V_1 = V_0 \oplus W_0$ holds.



Question: How do we find ψ?

Theorem: (Daubechies; 1988) Let $\{V_j : j \in \mathbb{Z}\}$ be a multiresolution approximation with scaling function ϕ, i.e. such that

$$\{\phi(\cdot - m) : m \in \mathbb{Z}\}$$

forms an ONB for V_0. Then there exists a 2π-periodic, square integrable function m_0 such that

$$\hat{\phi}(\xi) = m_0(\xi/2)\hat{\phi}(\xi/2).$$

One possibility to construct a wavelet ψ is

$$\hat{\psi} = e^{i\xi/2}\overline{m_0(\xi/2 + \pi)}\hat{\phi}(\xi/2).$$

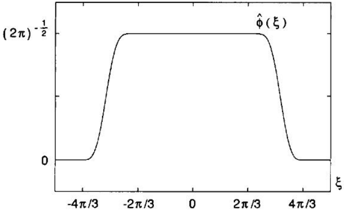
Define

$$\hat{\phi}(\xi) = \begin{cases} (2\pi)^{-1/2} & \text{if } |\xi| \leq \frac{2\pi}{3} \\ (2\pi)^{-1/2} \cos\left[\frac{\pi}{2}\vartheta\left(\frac{3}{2\pi}|\xi| - 1\right)\right], & \text{if } \frac{2\pi}{3} \leq |\xi| \leq \frac{4\pi}{3} \\ 0 & \text{otherwise,} \end{cases}$$

with ϑ such that $\vartheta(x) + \vartheta(1-x) = 1$, then

$$\{\phi(\cdot - m) : m \in \mathbb{Z}\}$$

is an orthonormal system and



Define

$$\hat{\phi}(\xi) = \begin{cases} (2\pi)^{-1/2} & \text{if } |\xi| \le \frac{2\pi}{3} \\ (2\pi)^{-1/2} \cos\left[\frac{\pi}{2}\vartheta\left(\frac{3}{2\pi}|\xi| - 1\right)\right], & \text{if } \frac{2\pi}{3} \le |\xi| \le \frac{4\pi}{3} \\ 0 & \text{otherwise,} \end{cases}$$

with ϑ such that $\vartheta(x) + \vartheta(1-x) = 1$, then

$$\{\phi(\cdot - m) : m \in \mathbb{Z}\}$$

is an orthonormal system and

$$\hat{\psi}(\xi) = \begin{cases} (2\pi)^{-1/2} e^{i\xi/2} \sin\left[\frac{\pi}{2}\vartheta\left(\frac{3}{2\pi}|\xi| - 1\right)\right], & \text{if } \frac{2\pi}{3} \le |\xi| \le \frac{4\pi}{3} \\ (2\pi)^{-1/2} e^{i\xi/2} \cos\left[\frac{\pi}{2}\vartheta\left(\frac{3}{4\pi}|\xi| - 1\right)\right], & \text{if } \frac{4\pi}{3} \le |\xi| \le \frac{8\pi}{3} \\ 0 & \text{otherwise,} \end{cases}$$

$$= e^{i\xi/2}\overline{m_0(\xi/2 + \pi)}\hat{\phi}(\xi/2).$$

$\rightsquigarrow$ Phase factor and interaction with ϑ not coincidental.

The Aftermath

Consequences

The developement of the MRA and a construction of compactly supported wavelets triggered a huge body of further works:

- Image and audio denoising,
- Image compression, e.g. JPEG2000,
- Calderon's programm, T(1) Theorem, (David and Journe; 1984, Meyer; 1986)
- Inverse problems, e.g. MRI (Lustig, Donoho, Pauly; 2007)
- Optimal adaptive discretizations of partial differential equations (Cohen, Dahmen, Devore; 2001),
- Deconvolution of astronomical images e.g. of Hubble telescope,
- Theory of regularity structures (Hairer; 2013),
- Function space characterization (Cohen, Dahmen, Devore, Triebel),
- Observation of gravitational waves (Abbott; 2016, Cornish, Littenberg; 2015)

Consequences

The developement of the MRA and a construction of compactly supported wavelets triggered a huge body of further works:

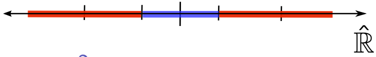
- Image and audio denoising,
- Image compression, e.g. JPEG2000,
- Calderon's programm, T(1) Theorem, (David and Journe; 1984, Meyer; 1986)
- Inverse problems, e.g. MRI (Lustig, Donoho, Pauly; 2007)
- Optimal adaptive discretizations of partial differential equations (Cohen, Dahmen, Devore; 2001),
- Deconvolution of astronomical images e.g. of Hubble telescope,
- Theory of regularity structures (Hairer; 2013),
- Function space characterization (Cohen, Dahmen, Devore, Triebel),
- Observation of gravitational waves (Abbott; 2016, Cornish, Littenberg; 2015)

Tools for Applications

Let $\psi \in L^2(\mathbb{R})$ be a wavelet and $\phi \in L^2(\mathbb{R})$ the corresponding scaling function such that

$$\{\phi(\cdot - m) : m \in \mathbb{Z}\} \cup \{\psi_{j,m} := 2^{j/2}\psi(2^j \cdot - m) : j \geq 0, m \in \mathbb{Z}\}$$

is an ONB for $L^2(\mathbb{R})$.



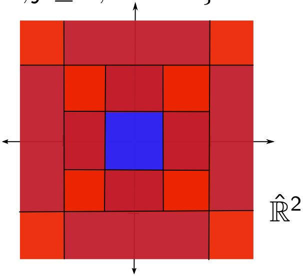
Tensor products yield bases for $\mathbb{R}^2$:

$$\{\phi^{(1)}(\cdot - m) : m \in \mathbb{Z}^2\}$$

$$\cup \{\psi^{(l)}_{j,m} := 2^{j/2}\psi^{(l)}(2^j \cdot - m) : l = 1, 2, 3, j \geq 0, m \in \mathbb{Z}^2\}$$

with

$$\phi^{(1)} = \phi \otimes \phi, \qquad \begin{array}{l} \psi^{(1)} = \phi \otimes \psi \\ \psi^{(2)} = \psi \otimes \phi \\ \psi^{(3)} = \psi \otimes \psi \end{array} .$$



Let

$$\{\phi_m = \phi(\cdot - m) : m \in \mathbb{Z}\} \cup \{\psi_{j,m} := 2^{j/2}\psi(2^j \cdot - m) : j \geq 0, m \in \mathbb{Z}\}$$

be a orthonormal wavelet basis. Then

$$f = \sum_{m \in \mathbb{Z}} \langle \phi_m, f \rangle \phi_m + \sum_{j \geq 0, m \in \mathbb{Z}} \langle \psi_{j,m}, f \rangle \psi_{j,m}.$$

We are very interested in the coefficient sequences $(\langle \phi_m, f \rangle)_{m \in \mathbb{Z}}$ and $(\langle \psi_{j,m}, f \rangle)_{(j,m) \in \mathbb{N} \times \mathbb{Z}}$.

Fast wavelet transform:

Filter based method to compute all coefficients up to a fixed upper scale $J \in \mathbb{N}$, i.e.

$$(\langle \phi_m, f \rangle)_{m \in \mathbb{Z}} \text{ and } (\langle \psi_{j,m}, f \rangle)_{(j,m) \in \mathbb{N} \times \{1,\dots,J\}}$$

Fast Wavelet Transform

Since $W_j, V_j \subset V_{j+1}$ we have that there exist sequences $c = (c_n)_{n\in\mathbb{Z}}$ and $d = (d_n)_{n\in\mathbb{Z}}$:

$$\psi_{j,m} = \sum_{\ell\in\mathbb{Z}} c_\ell \phi_{j+1,m+\ell}, \quad \phi_{j,m} = \sum_{\ell\in\mathbb{Z}} d_\ell \phi_{j+1,m+\ell},$$

holds for all $m \in \mathbb{Z}$. One can compute all coefficients

$$(\langle \phi_m, f\rangle)_{m\in\mathbb{Z}} \text{ and } (\langle \psi_{j,m}, f\rangle)_{(j,m)\in\mathbb{N}\times\{1,\dots,J\}}$$

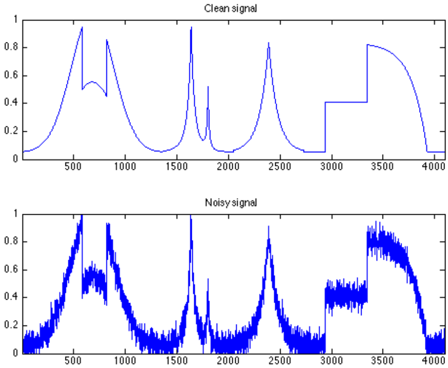
from $(\langle \phi_{J,m}, f\rangle)_{m\in\mathbb{Z}}$:

$$\begin{tikzcd}
(\langle f, \phi_{J,k}\rangle)_{k\in\mathbb{Z}} \arrow[r, "d"] \arrow[dr, "c"'] & (\langle f, \phi_{J-1,k}\rangle)_{k\in\mathbb{Z}} \arrow[r, "d"] \arrow[dr, "c"'] & (\langle f, \phi_{J-2,k}\rangle)_{k\in\mathbb{Z}} \arrow[r] & \dots \\
 & (\langle f, \psi_{J-1,k}\rangle)_{k\in\mathbb{Z}} & (\langle f, \psi_{J-2,k}\rangle)_{k\in\mathbb{Z}} &
\end{tikzcd}$$

Noisy music: Assume we have an audio signal given by a function f such that

$$f = f_0 + n,$$

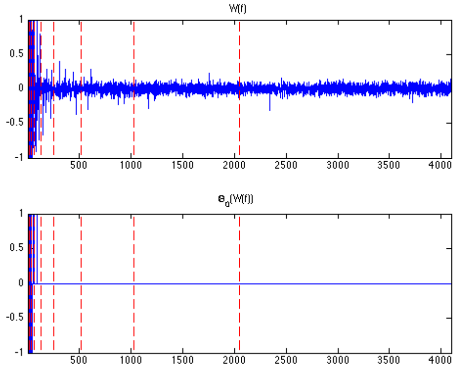
where f_0 is the clean signal and n is some noise.



Wavelet transform of noisy signal:

We take a wavelet transform of the noisy signal and perform Hard threshholding:

$$(T_\lambda(x))_i = \begin{cases} x_i & \text{if } |x| > \lambda, \\ 0 & \text{otherwise.} \end{cases}$$



Reconstruction: Using the inverse wavelet transform we obtain a signal back:

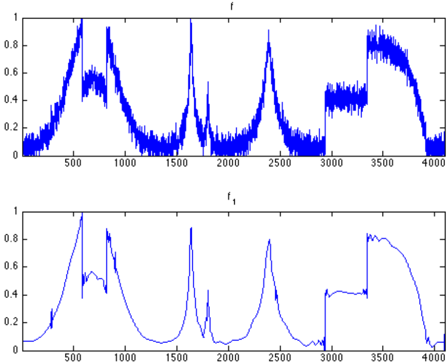


Image compression using sparse wavelet representation:

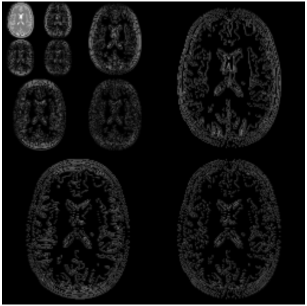
$x \in \mathbb{R}^{256 \times 256}$

Courtesy of Maximilian Maerz

Image Compression

Image compression using sparse wavelet representation:

— fwt ⟶

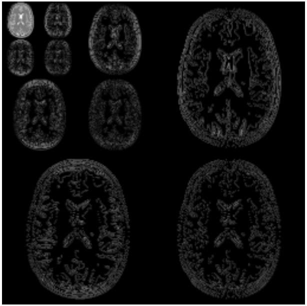


$x \in \mathbb{R}^{256\times 256}$ $\qquad$ $Wx \in \mathbb{R}^{256\times 256}$

Courtesy of Maximilian Maerz

Image compression using sparse wavelet representation:

$\longrightarrow$ fwt $\longrightarrow$



$x \in \mathbb{R}^{256 \times 256}$ $\quad$ $Wx \in \mathbb{R}^{256 \times 256}$

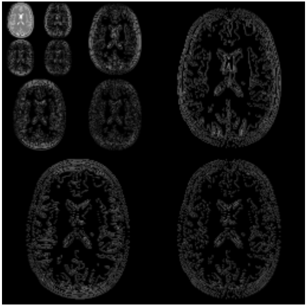
- $Wx \in \mathbb{R}^{256^2}$ is very sparse, i.e., only few coefficients are large.

Courtesy of Maximilian Maerz

Image compression using sparse wavelet representation:

$x_{\text{rec}} \in \mathbb{R}^{256\times 256}$

$\longleftarrow$ `ifwt` $\longrightarrow$



$Wx \in \mathbb{R}^{256\times 256}$

- $Wx \in \mathbb{R}^{256^2}$ is very sparse, i.e., only few coefficients are large.
- We reconstruct using only the largest 16% of the coefficients.
  - Relative error $\|x - x_{\text{rec}}\| / \|x\| < 2\%$

Courtesy of Maximilian Maerz

Function Space Characterization

Norm characterization:

Given an orthonormal (wavelet) basis $(\psi_{j,m})_{(j,m)\in\mathbb{N}\times\mathbb{Z}}$ for $L^2(\mathbb{R})$ we have that

$$\|f\|_{L^2(\mathbb{R})} = \|\langle f, \psi_{j,m}\rangle\|_{\ell_2}, \text{ for all } f \in L^2(\mathbb{R}).$$

Is it possible to replace $L^2(\mathbb{R})$ and ℓ_2 by different spaces in the above statement?

Weighted ℓ_2 spaces:

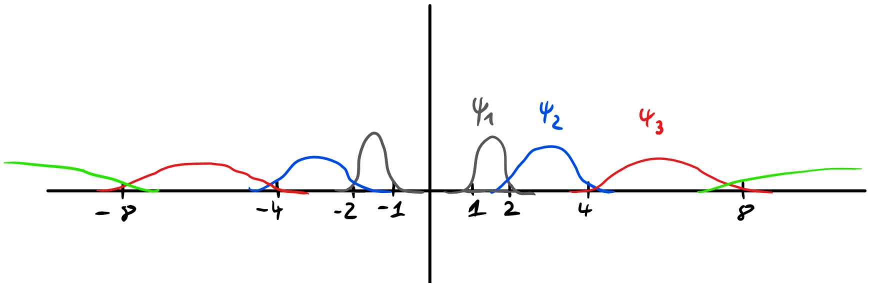
Indeed, under some assumptions on ψ we obtain for example:

$$\|f\|_{H^s(\mathbb{R})} \sim \|2^{sj}\langle f, \psi_{j,m}\rangle\|_{\ell_2} =: \|\langle f, \psi_{j,m}\rangle\|_{\ell_{2,w}}, \text{ for all } f \in H^s(\mathbb{R}).$$

Intuitive explanation: By definition we have that

$$\|f\|_{H^s(\mathbb{R})} = \|(1+|\cdot|^2)^{s/2}\hat{f}\|_{L^2(\mathbb{R})} = \|\langle(1+|\cdot|^2)^{s/2}\hat{f}, \widehat{\psi_{j,m}}\rangle\|_{\ell_2}.$$

Recall shape of wavelets in frequency:



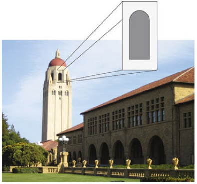
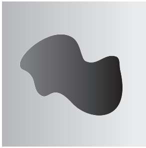
Hence $(1+|\cdot|^2)^{s/2} \sim 2^{sj}$ on $\operatorname{supp}\widehat{\psi_{j,m}}$, i.e.,

$$\|\langle(1+|\cdot|^2)^{s/2}\hat{f}, \widehat{\psi_{j,m}}\rangle\|_{\ell_2} \sim \|2^{js}\langle f, \psi_{j,m}\rangle\|_{\ell_2}$$

Some Shameless Advertisement of our Own Work

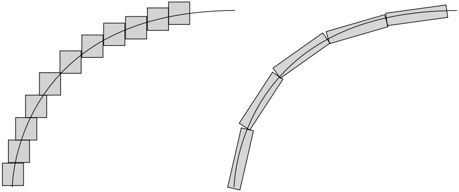
Wavelets and Images

Natural images:


Natural images exhibit jumps along piecewise smooth curves.

Wavelet approximation:



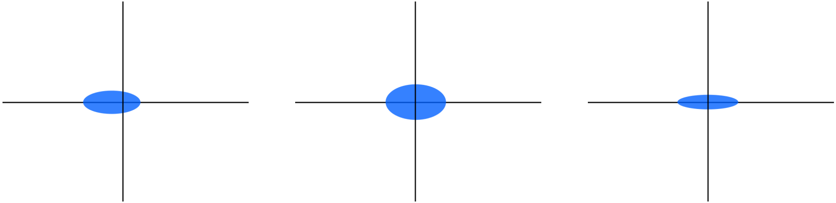
Anisotropic elements could be better than isotropic.

Ingredients: Function $\psi \in L^2(\mathbb{R}^2)$ and scaling and shearing matrices:

$$A_a := \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}, \qquad S_s := \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, a \in \mathbb{R}^+, s \in \mathbb{R}.$$

We define for $j, k \in \mathbb{Z}, m \in \mathbb{Z}^2$:

$$\psi_{j,k,m} := 2^{\frac{3j}{4}} \psi(S_k A_{2^j} \cdot - m)$$

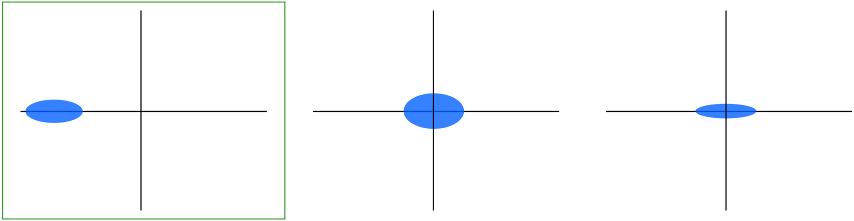


Ingredients: Function $\psi \in L^2(\mathbb{R}^2)$ and scaling and shearing matrices:

$$A_a := \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}, \qquad S_s := \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, a \in \mathbb{R}^+, s \in \mathbb{R}.$$

We define for $j, k \in \mathbb{Z}, m \in \mathbb{Z}^2$:

$$\psi_{j,k,m} := 2^{\frac{3j}{4}} \psi(S_k A_{2^j} \cdot - m)$$

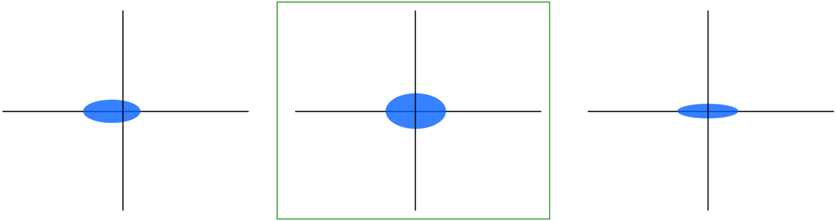


Ingredients: Function $\psi \in L^2(\mathbb{R}^2)$ and scaling and shearing matrices:

$$A_a := \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}, \qquad S_s := \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, a \in \mathbb{R}^+, s \in \mathbb{R}.$$

We define for $j, k \in \mathbb{Z}, m \in \mathbb{Z}^2$:

$$\psi_{j,k,m} := 2^{\frac{3j}{4}} \psi(S_k A_{2^j} \cdot - m)$$

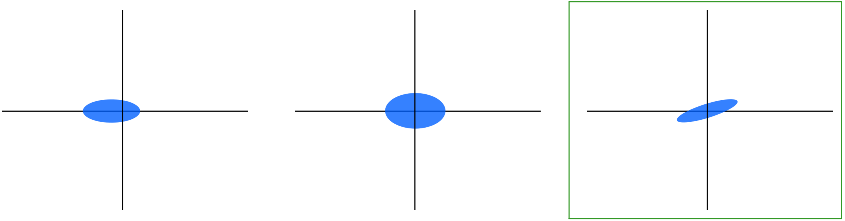


Ingredients: Function $\psi \in L^2(\mathbb{R}^2)$ and scaling and shearing matrices:

$$A_a := \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}, \qquad S_s := \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, a \in \mathbb{R}^+, s \in \mathbb{R}.$$

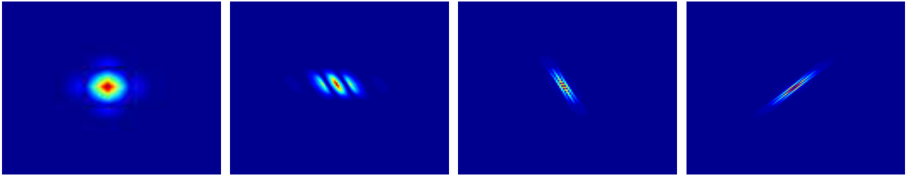
We define for $j, k \in \mathbb{Z}, m \in \mathbb{Z}^2$:

$$\psi_{j,k,m} := 2^{\frac{3j}{4}} \psi(S_k A_{2^j} \cdot - m)$$

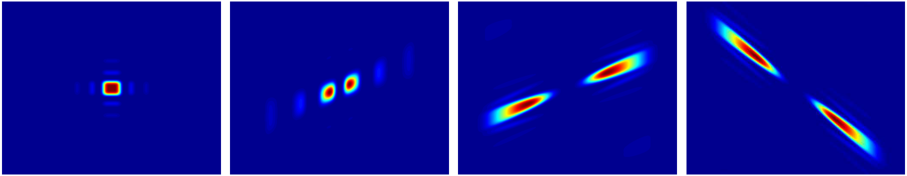


Cone-Adapted Shearlet Systems

Time domain:



Frequency domain:

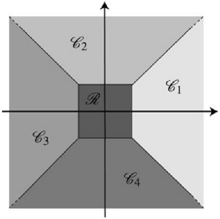


Definition: (Kittipoom, Kutyniok, Lim; 2012)
Let $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$, $c = [c_1, c_2]^T \in \mathbb{R}^2$ with $c_1, c_2 > 0$. Then the cone-adapted shearlet system is defined by $(\psi_{j,k,m,\iota})_{(j,k,m,\iota)\in\Lambda}$, where

$$\Lambda := \{(j,k,m,\iota) : \iota \in \{-1,0,1\}, |\iota| j \geq j \geq 0, |k| \leq |\iota| \lceil 2^{\frac{j}{2}} \rceil, m \in \mathbb{Z}^2\}$$

and

$$\psi_{j,0,m,0} = \phi(\cdot - c_1 m),$$
$$\psi_{j,k,m,1} = 2^{\frac{3j}{4}} \psi(S_k A_j \cdot - M_c m),$$
$$\psi_{j,k,m,-1} = 2^{\frac{3j}{4}} \tilde{\psi}(S_k^T \tilde{A}_j \cdot - M_{\tilde{c}} m),$$



with $M_c := \mathrm{diag}(c_1, c_2)$, $M_{\tilde{c}} = \mathrm{diag}(c_2, c_1)$, and $\tilde{A}_{2^j} = \mathrm{diag}(2^{\frac{j}{2}}, 2^j)$.

Shearlet bases?

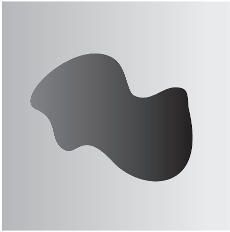
Unlike wavelet systems, the cone-adapted shearlet systems are not orthonormal bases. But we still have:

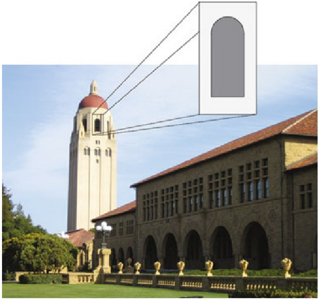
Theorem: (Kittipoom, Kutyniok, Lim; 2012)

Let $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ satisfy some admissibility conditions. Then, $(\psi_{j,k,m,\iota})_{(j,k,m,\iota)\in\Lambda}$ satisfies

$$\|f\|_{L^2(\mathbb{R}^2)} \sim \|(\langle f, \psi_{j,k,m,\iota}\rangle)_{(j,k,m,\iota)\in\Lambda}\|_{\ell_2}.$$

Functions $f = f_1 + \chi_B f_2$ with $B \subset (0,1)^2$ and $\partial B \in C^2$, $f_i \in C^2$, $\operatorname{supp} f_i \subset (0,1)^2$ are called cartoon-like functions.





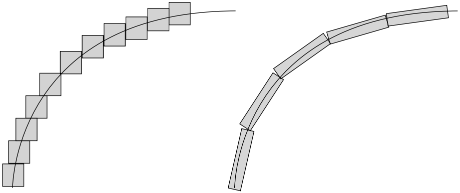
Theorem: (Labate, Guo; 2007/ Kutyniok, Lim; 2011)
The error of the best N-term approximation for cartoon-like images with the shearlet frame decays as $O(N^{-1}(\log(N))^{\frac{3}{2}})$.

This rate is optimal up to the log factors, [Donoho; 2001].

Comparisson to wavelets: Wavelets do not achieve the optimal N-term approximation of shearlets

Approximation rates for cartoon-like functions:

- Shearlets: $N^{-1}\log(N))^{\frac{3}{2}}$,
- Wavelets: $N^{-\frac{1}{2}}$,
- Fourier basis: $N^{-\frac{1}{4}}$.



Applications

Some applications of shearlets:

- Image processing, e.g. denoising, inpainting;
- Resolution of wavefront set;
- Edge detection / classification;
- Regularization of inverse problems, e.g. inverse acoustic scattering, MRI;
- Adaptive discretization and solution of PDEs;
- Characterization of smoothness spaces;
- Theoretical analysis of approximation with neural networks.

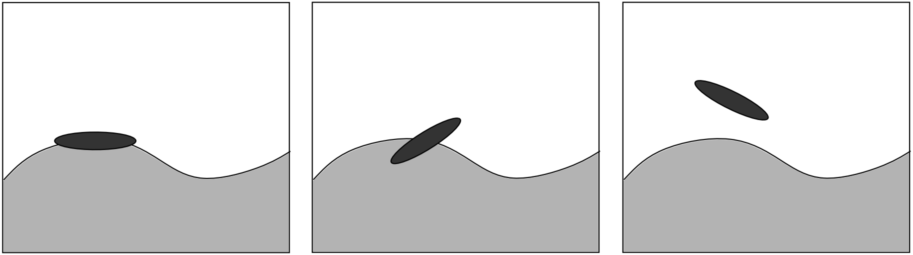
Applications

Some applications of shearlets:

- Image processing, e.g. denoising, inpainting;
- Resolution of wavefront set;
- Edge detection / classification;
- Regularization of inverse problems, e.g. inverse acoustic scattering, MRI;
- Adaptive discretization and solution of PDEs;
- Characterization of smoothness spaces;
- Theoretical analysis of approximation with neural networks.

Observation:

Shearlet coefficients $\langle f, \psi_{j,k,m} \rangle$ react to geometric structures:



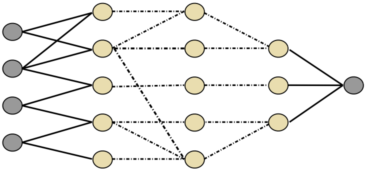
Large coefficients | Small coefficients | Small coefficients

$\rightsquigarrow$ Direction and position of discontinuities can be found by analyzing large shearlet coefficients.

Complex shearlet-based ridge and edge measure (CoShREM) toolbox:

A common model for neural networks is given by

- $d \in \mathbb{N}$: dimension of the input layer,
- L: number of layers,
- N number of neurons,
- $\rho : \mathbb{R} \to \mathbb{R}$ rectifier,
- $W_\ell : \mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}$, $\ell = 1, \dots, L$: Affine linear maps.



Then $\Phi : \mathbb{R}^d \to \mathbb{R}^{N_L}$ given by

$$\Phi(x) = W_L \rho(W_{L-1}\rho(\dots\rho(W_1(x))), \quad x \in \mathbb{R}^d,$$

is called a neural network.

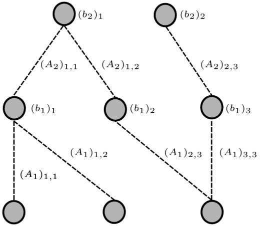
Approximation properties: *How well can an arbitrary function be approximated by a neural network?*

Universal approximation: For any continuous function $f : K \to \mathbb{R}$ and any $\epsilon > 0$ there exists a neural network approximating f up to an error ϵ. [Cybenko; 1989, Hornik et al.; 1991].

Rate of approximation: It is unclear how complex the neural network needs to be.

Approximation by sparsely connected networks:

Sparsely connected networks have few non-zero weights.



$$A_2 = \begin{pmatrix} (A_1)_{1,1} & (A_1)_{1,2} & 0 \\ 0 & 0 & (A_1)_{2,3} \end{pmatrix}$$

$$A_1 = \begin{pmatrix} (A_1)_{1,1} & (A_1)_{1,2} & 0 \\ 0 & 0 & (A_1)_{2,3} \\ 0 & 0 & (A_1)_{3,3} \end{pmatrix}$$

Fundamental lower bound:

There exists a fundamental lower bound on the uniform approximation of function classes by sparsely connected neural networks [Bölcskei, Grohs, Kutyniok, P.; 2017].

Observation towards an upper bound: Let $t \in \mathbb{R}^d$, $A \in GL(d)$, $a \neq 0$, $\psi \in L^2(\mathbb{R}^d)$, then the following are equivalent:

- $a\psi(A(\cdot) - t)$ can be realized by a neural network Φ.
- $a\psi$ can be realized by a neural network $\tilde{\Phi}$.
- ψ can be realized by a neural network $\tilde{\tilde{\Phi}}$.

Moreover, $\tilde{\Phi}$, $\tilde{\Phi}$, Φ all have the same number of edges up to a constant factor.

Proposition:

If a neural network can represent a wavelet ψ, then it can represent sums of N wavelets/shearlets with $O(N)$ edges.

N-term Approximation rates of wavelets and shearlets carry over to N-edge approximations of networks.

$\rightsquigarrow$ Lower bound is sharp for many function classes.

Conclusion

Meyer and wavelets:

- Meyer started his work on wavelets inspired by developements in geophysics;
- He developed one of the first orthogonal wavelet bases;
- He introduced the concept of multiresolution analysis;
- These developements started a wavelet revolution which led to break-through results in various areas.

Thank you for the attention!