Wavelets and Applications

Kévin Polisano
kevin.polisano@univ-grenoble-alpes.fr

M2 MSIAM & Ensimag 3A MMIS

January 21, 2022

LABORATOIRE
JEAN KUNTZMANN
MATHÉMATIQUES APPLIQUÉES - INFORMATIQUE

Linear and nonlinear approximations in wavelet bases

Approximation in an orthonormal basis

Let $\mathcal{B} = \{\psi_m\}_m$ an orthonormal basis of $L^2([0,1]^d)$ for $d = 1$ (signals) or $d = 2$ (images) and the decomposition of f into this basis:

$$f = \sum_{m \in \mathbb{Z}} \langle f, \psi_m \rangle \psi_m$$

Approximation of f keeping a subset $I_M \subset \mathbb{Z}$ of $M = |I_M|$ coefficients and reconstructing from this subset:

$$f_M = \sum_{m \in I_M} \langle f, \psi_m \rangle \psi_m$$

which is the orthogonal projection of f onto $V_M = \text{Span}\{\psi_m, m \in I_M\}$

The approximation error is given by:

$$\|f - f_M\|^2 = \sum_{m \notin I_M} \langle f, \psi_m \rangle^2$$

Linear approximation

When I_M is **fixed once for all** and used the same set of coefficients for all functions, then the mapping $f \mapsto f_M$ is a linear orthogonal projection on V_M satisfying:

$$(f+g)_M = f_M + g_M$$

Examples

- **Fourier basis**

$$I_M = \{-M/2+1, \ldots, M/2\}$$

- **1-D Wavelet basis**

$$I_M = \{m = (j,k) \,|\, j \geqslant j_0\}$$

Linear approximation

- **Orthogonal projection over the space V_J**

$$\begin{array}{rcccl} \mathbf{P}_J & : & L^2(\mathbb{R}) & \longrightarrow & V_J \\ & & f & \longmapsto & \sum_{k\in\mathbb{Z}} \langle f, \varphi_{J,k}\rangle \varphi_{J,k} = \sum_{j\leq J-1} \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k} \end{array}$$

- **Strang-Fix condition** of order N ($x^n \in V_0$):

$$\forall n = 0, \ldots, N-1, \quad x^n = \sum_{k\in\mathbb{Z}} a_k^n \, \varphi(x-k)$$

i.e ψ has N vanishing moments.

- **Projection error**:

$$\|f - \mathbf{P}_J f\|_{L^2}^2 = \sum_{j=J}^{+\infty} \sum_{k\in\mathbb{Z}} |\langle f, \psi_{j,k}\rangle|^2$$

Linear approximation

Theorem

If $f \in H^s(\mathbb{R})$ with $s \leq N$, with

$$\|f - \mathbf{P}_J f\|_{L^2} \leq C\ 2^{-Js}\ \|f\|_{H^s}$$

(*cf.* finite elements with $h = 2^{-J}$) then the following Sobolev-norm equivalence holds:

$$\begin{aligned}
\|f\|_{H^s}^2 \quad &\sim \quad \sum_{j=-\infty}^{+\infty} \sum_{k \in \mathbb{Z}} 2^{2Js} |\langle f, \psi_{j,k} \rangle|^2 \\
&\sim \quad \|\mathbf{P}_0 f\|_{L^2} + \sum_{j=0}^{+\infty} \sum_{k \in \mathbb{Z}} 2^{2Js} |\langle f, \psi_{j,k} \rangle|^2
\end{aligned}$$

Nonlinear approximation

To minimize the approximation error $\|f - f_M\|$ one can choose I_M depending on f, for instance by selecting the M largest coefficients

$$I_M = \{M \text{ largest coefficients } |\langle f, \psi_m \rangle|\}$$

which is equivalent to a T–thresholding

$$I_M = \{m : |\langle f, \psi_m \rangle| > T\}$$

where T depends on the number of coefficients M.
More precisely, by ordering the coefficients $d_m = |\langle f, \psi_m \rangle|$ by decaying order that is $d_m \geqslant d_{m+1}$ then $T = d_M$.

Hard thresholding. The non-linear approximation can be rewritten as:

$$f_M = \sum_{|\langle f, \psi_m \rangle| > T} \langle f, \psi_m \rangle \psi_m = \sum_m S_T^0(\langle f, \psi_m \rangle)\psi_m$$

where

$$S_T^0(x) = \begin{cases} x & \text{if } |x| > T \\ 0 & \text{if } |x| \leqslant T \end{cases}$$

Nonlinear approximation

Proposition

$$d_m = O\left(m^{-\frac{\alpha+1}{2}}\right) \Longleftrightarrow \|f - f_M\| = O(M^{-\alpha})$$

Proof :

$\boxed{\Rightarrow}$ is straightforward by noticing that $\|f - f_M\|^2 = \sum_{m>M} d_m^2$.

$\boxed{\Leftarrow}$ Due to the decaying order we have:

$$\frac{M}{2} \times d_M^2 \leqslant \sum_{m=M/2+1}^{M} d_m^2 \leqslant \sum_{m>M/2} d_m^2 = \|f - f_{M/2}\|^2$$

which proves the result.

Nonlinear approximation

Let $M \in \mathbb{N}$. Let $f \in L^2(\mathbb{R})$, and its wavelet decomposition:

$$f = f_0 + \sum_{j=0}^{+\infty} \sum_{k=-\infty}^{+\infty} d_{j,k} \psi_{j,k}$$

One sorts the wavelet coefficients $d_{j,k}$ in decreasing order:

$$|d_{j_1,k_1}| > |d_{j_2,k_2}| > \cdots > |d_{j_{M-1},k_{M-1}}| > \dots$$

and one obtains the best M-terms non-linear approximation

$$f_M = f_0 + \sum_{i=1}^{M} d_{j_i,k_i} \psi_{j_i,k_i}$$

If $f \in B_q^{s,q}$ with $\frac{1}{q} = \frac{1}{2} + s$, which is equivalent to:

$$\|f\|_{B_q^{s,q}}^q \sim \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |d_{j,k}|^q < +\infty$$

Nonlinear approximation

The non-linear approximation error is $\|f - f_M\|_{L^2}^2 = \sum_{i \geq M+1} |d_{j_i,k_i}|^2$

Proposition

$$\|f - f_M\|_{L^2} \leq C \left(\frac{1}{M}\right)^s \|f\|_{B_q^{s,q}}$$

(in dimension d, s should be replaced by $\frac{s}{d}$).

Proof:

$$m|d_{j_m,k_m}|^q \leq \sum_{i=0}^{m-1} |d_{j_i,k_i}|^q \leq \sum_{i=0}^{+\infty} |d_{j_i,k_i}|^q = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |d_{j,k}|^q \leq C \ \|f\|_{B_q^{s,q}}^q$$

Then

$$|d_{j_m,k_m}| \leq C m^{-1/q} \ \|f\|_{B_q^{s,q}}$$

$$\|f - f_M\|_{L^2} \leq C \|f\|_{B_q^{s,q}} \left(\sum_{m \geq M+1} m^{-\frac{2}{q}} \right)^{1/2} \leq C M^{\frac{1}{2} - \frac{1}{q}} \ \|f\|_{B_q^{s,q}} = C M^{-s} \|f\|_{B_q^{s,q}}$$

Nonlinear approximation

Remarks

- For $M = 2^J$ one obtains the same convergence rate for the linear and nonlinear approximation. But $B_q^{s,q}$ is a space which contains more functions than the space H^s, for instance discontinuous functions for arbitrary large values of s, whenever functions of H^s are necessarily continuous if $s > d/2$ (d space dimension).

- One has also the characterization: if $f \in B_q^{s,q}$

$$\mathrm{Card}\,\{\lambda : |d_\lambda| \geq \varepsilon\} \leq C\varepsilon^{-q}$$

Compression factor of a turbulent 2D vorticity field

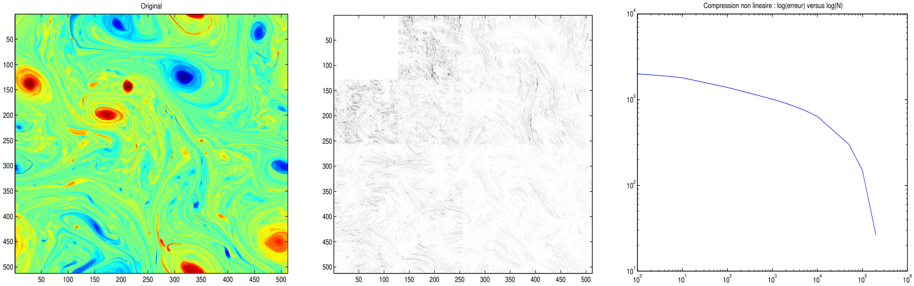


Figure: Analysis of a 2D turbulent field: vorticity field, its wavelet coefficients, and nonlinear approximation error, in terms of the number of retained coefficients

Nonlinear approximation for different class of signal models

Class of signal and image models $f \in \Theta$ where $\Theta \subset L^2([0,1]^d)$

- Uniformly smooth signals and images
- Sobolev smooth signals and images
- Piecewise regular signals and images
- Bounded variation signals and images
- C^α cartoon images

The error decay

$$\forall f \in \Theta,\ \forall M, \quad \|f - f_M\|^2 \leqslant C_f M^{-\alpha}$$

Remark. The power α is independent of f, it depends on the orthogonal basis considered for approximation and on Θ. It should be as large as possible.

Denoising in orthonornal wavelet bases

References: articles of Donoho and Johnstone

Noised data:

$$X[n] = f[n] + W[n], \quad n = 0, \ldots, N-1$$

- X: measured data
- f: (unknown) signal of size N, corrupted by noise
- W: Gaussian *white* noise, with zero mean and variance σ^2

The aim is to provide an estimator $\tilde{F} = \mathcal{D}(X)$ of f minimizing the risk (mean square error):

$$r(\mathcal{D}, f) = \mathbb{E}\left\{\|f - \tilde{F}\|^2\right\} = \sum_{n=0}^{N-1} \mathbb{E}\left\{|f[n] - \tilde{F}[n]|^2\right\}$$

Nonlinear estimators in bases

Let $\mathcal{B} = \{g_k \in \mathbb{R}^N, k = 0, \ldots, N-1\}$ be an orthonormal basis of $\mathbb{R}^N$. One decomposes the noisy signal in $\mathcal{B}$:

$$X[n] = \sum_{k=0}^{N-1} \langle X, g_k \rangle g_k[n]$$

and the inner products satisfy:

$$\langle X, g_k \rangle = \langle f, g_k \rangle + \langle W, g_k \rangle$$

Remarks

- $(\langle W, g_k \rangle)_k$ are independent Gaussian variables of variance σ^2 (since $\mathcal{B}$ is orthonormal).
- $\mathbb{E}\left\{\langle X, g_k \rangle^2\right\} = |\langle f, g_k \rangle|^2 + \sigma^2$

Diagonal operators

A *diagonal operator* $\mathcal{D}$ in the basis $\mathcal{B}$ leads to an estimator of the form:

$$\tilde{F} = \mathcal{D}X = \sum_{k=0}^{N-1} d_k \left(\langle X, g_k \rangle\right) g_k$$

where the d_k are attenuation functions of the noisy coefficients.

Ideal estimator (*i.e.* which minimizes the risk $r(\mathcal{D}, f)$)

$$\tilde{F} = \mathcal{D}X = \sum_{k=0}^{N-1} \langle X, g_k \rangle \ \theta(k) \ g_k$$

with

$$\theta(k) = \begin{cases} 1 & \text{if} \quad |\langle f, g_k \rangle| \geq \sigma \\ 0 & \text{if} \quad |\langle f, g_k \rangle| < \sigma \end{cases}$$

In this case, the operator $\mathcal{D}$ is nonlinear.

Thresholding estimators

A thresholding estimator in the basis $\mathcal{B}$ corresponds to a *diagonal operator* $\mathcal{D}$:

$$\tilde{F} = \mathcal{D}X = \sum_{k=0}^{N-1} d_k\left(\langle X, g_k\rangle\right) g_k$$

where the d_k are thresholding functions (let T be a threshold):

$$d_k(x) = S_T^0(x) = \begin{cases} x & \text{if} \quad |x| > T \\ 0 & \text{if} \quad |x| \leq T \end{cases} \quad (\text{"hard" thresholding})$$

or

$$d_k(x) = S_T^1(x) = \begin{cases} x - T & \text{if} \quad x \geq T \\ x + T & \text{if} \quad x \leq -T \\ 0 & \text{if} \quad |x| \leq T \end{cases} \quad (\text{"soft" thresholding})$$

▷ **Question**: choice of T to approach the risk of the ideal estimator?

Thresholding estimators

Assume that the vector a of coefficients $a[k] = \langle f, g_k \rangle$ is **sparse** (most of them are zero i.e ℓ^0 norm small). We have

$$X = f + W, \quad W \sim \mathcal{N}(0, \sigma)$$
$$A = a + Z, \quad Z \sim \mathcal{N}(0, \sigma)$$

If $\min\limits_{k:a[k]\neq 0} |a[k]|$ is large enough then

$$\| f - \tilde{f} \| = \| a - S_T(\tilde{a}) \|$$

is minimum for

$$T = \tau_N = \max_{0 \leqslant k < N} |z[k]| \sim \sigma\sqrt{2 \ln N} \quad \text{(universal threshold)}$$

Theorem (Donoho-Jonstone)

If $\| f - f_M \| = O(M^{-\alpha})$ then $\mathbb{E}\left\{ \| f - \tilde{F} \| \right\} = O\left(\sigma^{\frac{2\alpha}{\alpha+1}} \right)$

Wavelet thresholding

Consider a (periodic) wavelet basis:

$\mathcal{B} = \{\varphi, \psi_{j,k}\ ;\ 0 \leq j \leq J-1,\ k = 0 : 2^j - 1\}$ ($N = 2^J =$ size of the data)

The thresholding estimator writes:

$$\tilde{F} = S_T(\langle X, \varphi \rangle)\, \varphi + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} S_T\left(\langle X, \psi_{j,k} \rangle\right) \psi_{j,k}$$

Estimation of the noise variance σ^2:
If f is piecewise regular, a robust estimator is given by the *median* of the wavelet coefficients at the finest scale:

- $\{\langle X, \psi_{j,k} \rangle\}_{k=0:2^{J-1}-1}$: $2^{J-1} = \frac{N}{2}$ wavelet coefficients of the noisy data at the finest scale.
- If $\langle f, \psi_{j,k} \rangle$ is small (f is regular on the support of $\psi_{J-1,k}$), one has: $\langle X, \psi_{j,k} \rangle \approx \langle W, \psi_{j,k} \rangle$.

Wavelet thresholding

- If $\langle f, \psi_{j,k} \rangle$ is large, it corresponds to a singularity of f, but for a piecewise regular functions with isolated singularity, only few coefficients $\langle X, \psi_{j,k} \rangle$ are affected at the finest scale.
- Then $\langle X, \psi_{j,k} \rangle$ is a random variable of variance σ^2.

The noise standard deviation σ is estimated by the formula (exact for $P = 2^{J-1}$ independent Gaussian variables, of zero mean, and variance σ^2):

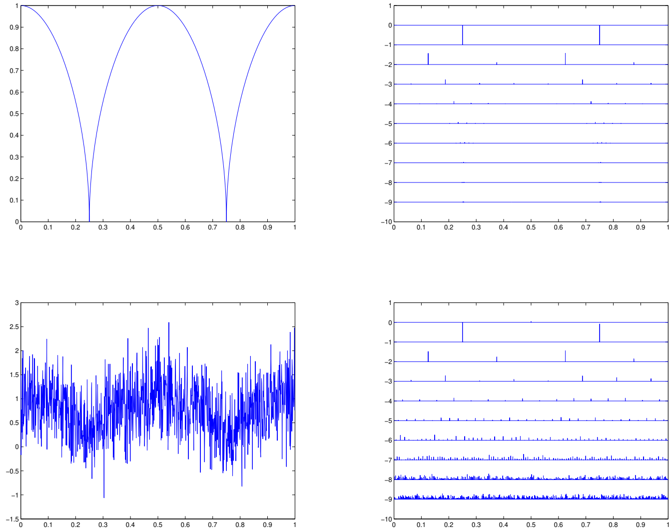
$$\sigma \approx \frac{M_X}{0,6745}$$

where M_X is the median of the coefficients $\{\langle X, \psi_{j,k} \rangle\}_{k=0:2^{J-1}-1}$ at the smallest scale.

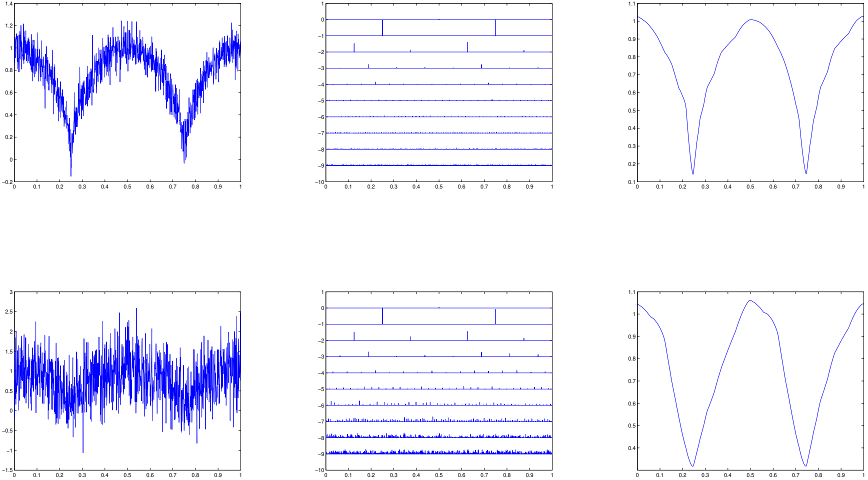
Example:

$f(x) = \sqrt{|\cos 2\pi x|} + \text{noise}$ (discretized on $1024 = 2^{10}$ values)

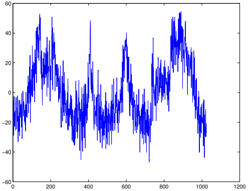
Example: $f(x) = \sqrt{|\cos 2\pi x|} + \text{noise}$

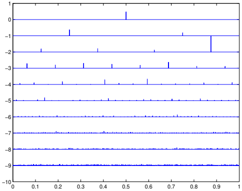


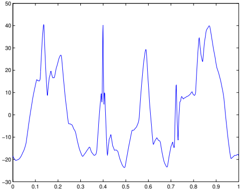
Example: $f(x) = \sqrt{|\cos 2\pi x|} + \text{noise}$

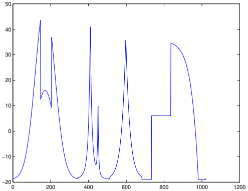


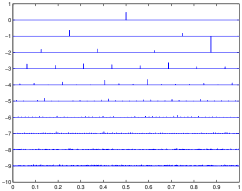
Example: Piece-Regular











Example: WaveLab denoising function

```
%Generation of a signal y
n=1024; dx=1/n; x=(0:n-1)/n;
alpha=0.1 % noise coefficient
y=sqrt(abs(cos(2*pi*x)));
or
y=MakeSignal('Piece-Regular',n);
%
y=y+alpha*randn(size(y)); % add Gaussian noise
plot(x,y) % plot of the noisy signal
% Denoising by hardtresholding on orthonormal wavelet coeff 'Symmlet 4'
out=ThreshWave(y);
plot(x,out) % plot the denoised signal
```

Sparse representation and approximation

Analysis vs. synthesis

- **Analysis**: $\Phi(f) = \{\langle f, \phi_p \rangle\}_{p \in \Gamma}$
- **Synthesis**: $f = \sum_p \langle f, \phi_p \rangle \phi_p$

Suppose that a sparse family of vectors $\{\phi_p\}_{p \in \Lambda}$ has been selected to approximate a signal f. An approximation can be recovered as an orthogonal projection in the space $\mathbf{V}_\Lambda$ generated by these vectors.

1. In a *dual-synthesis* problem, the orthogonal projection f_Λ of f in $\mathbf{V}_\Lambda$ is computed as above from the inner products $\{\langle f, \phi_p \rangle\}$ provided by an analysis operator, whose only a subset of such inner products is selected and possibly thresholded.
2. In a *dual-analysis* problem, the decomposition coefficients of f_Λ must be computed on a family of selected vectors $\{\phi_p\}_{p \in \Lambda}$, by pursuit algorithms which compute approximation supports in highly redundant dictionaries.

Sparse representation and approximation

Approximation in bases vs. redundant dictionaries

1. Choose an **orthogonal basis** $\mathcal{B} = \{\phi_p\}_{p\in\Gamma}$ for which the representation is **not redundant** at all, so we get a representation which is sparse and stable
   - By selecting the first M coefficients *(linear approximation)*
   - By selecting the M largest coefficients *(non-linear approximation)*

   The size support $M = |\Lambda|$ of $f_M \equiv f_\Lambda$ needed to have a good approximation error $\|f - f_M\|$ depends on the regularity of f.
2. Choose a **dictionary** $\mathcal{D} = \{\phi_p\}_{p\in\Gamma}$ which is **highly redundant** in order to obtain a more sparse representation (e.g. natural languages use redundant dictionaries). Identifying *patterns* or *features* consist on finding which vectors (*atoms*, words, ...) to choose to approximate

   $$f \approx f_\Lambda = \sum_{p\in\Lambda} \alpha_p \phi_p$$

   Famous algorithms: Matching pursuit, Orthogonal Matching Pursuit (OMP), Basis pursuit, ...

Sparse representation and approximation

Moving from transforms to dictionaries

- "Xlets" (curvelets, bandlets, contourlet, ...) take advantage of the image geometric regularity
- Redundant dictionaries can improve approximation, compression and denoising
- Optimal approximation finding is NP-hard, only approximated with matching or basis pursuits
- Great impact to inverse problems
  - Compressed sensing
  - Super-resolution
  - Source separation
- Can be used for patter recognition but problems of instabilities
- Deep learning made a breakthrough in classification and pattern recognition (dictionaries are learned i.e linear operators/filters, but need a lot of examples). Increase the level of adaptability.

$\Rightarrow$ have wavelets become *has-been*?