Nonlinear and Adaptive Approximation
Foundations and Algorithms

Albert Cohen

Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie, Paris

www.ann.jussieu.fr/~cohen

Vienna, 21-01-2008

Agenda

I. Motivation and basic examples

II. Wavelet bases, smoothness spaces, thresholding

III. Isotropic and anisotropic adaptive finite elements

Basic references

Ron DeVore, “Nonlinear approximation”, Acta Numerica, 1998.

Albert Cohen, “Numerical analysis of wavelet methods”, Elsevier North-Holland , 2003.

Central problem in approximation theory

- X normed space.

- $(\Sigma_N)_{N\geq 0} \subset X$ approximation subspaces ($g \in \Sigma_N$ described by N or $\mathcal{O}(N)$ parameters).

- Best approximation error $\sigma_N(f) := \inf_{g\in\Sigma_N} \|f - g\|_X$.

Problem 1: characterise those functions in $f \in X$ having a certain rate of approximation

$$f \in X^r \Leftrightarrow \sigma_N(f) \lesssim N^{-r}$$

Here $A \lesssim B$ means that $A \leq CB$, where the constant C is independant of the parameters defining A and B.

Examples

Linear approximation: Σ_N space of dimension $\mathcal{O}(N)$

- $\Sigma_N := \Pi_N$ polynomials of degree N in dimension 1

- $\Sigma_N := \{f \in C^r([0,1])\ ;\ f_{|[\frac{k}{N},\frac{k+1}{N}]} \in \Pi_m,\ \ k = 0, \cdots, N-1\}$ with $0 \leq r \leq m$ fixed, splines with uniform knots.

- $\Sigma_N := \mathrm{Vect}(e_1, \cdots, e_N)$ with $(e_k)_{k>0}$ a functional basis.

Nonlinear approximation: $\Sigma_N + \Sigma_N \neq \Sigma_N$

- $\Sigma_N := \{\frac{p}{q},\ p, q \in \Pi_N\}$ rational fractions

- $\Sigma_N := \{f \in C^r([0,1])\ ;\ f_{|[x_k, x_{k+1}]} \in \Pi_m,\ \ 0 = x_0 < \cdots < x_N = 1\}$ with $0 \leq r \leq m$ fixed, free knots splines.

- $\Sigma_N := \{\sum_{\lambda \in E} d_\lambda \psi_\lambda\ ;\ \#(E) \leq N\}$ set of all N-terms combination of a basis (ψ_λ).

Central problem in computational approximation

Problem 2: practical realization of $f \mapsto f_N \in \Sigma_N$ such that

$$\|f - f_N\|_X \lesssim \sigma_N(f).$$

If Σ_N are linear spaces and $P_N : X \to \Sigma_N$ are uniformly bounded projectors $\|P_N\|_{X \to X} \leq C$, then $f_N := P_N f$ is a good choice, since for all $g \in \Sigma_N$,

$$\begin{aligned} \|f - f_N\|_X &\leq \|f - g\|_X + \|g - f_N\|_X \\ &= \|f - g\|_X + \|P_N(g - f))\|_X \\ &\leq (1 + C)\|g - f\|_X, \end{aligned}$$

and therefore $\|f - f_N\|_X \leq (1 + C)\sigma_N(f)$.

What about nonlinear spaces ?

Application 1: signal and image compression

Less information is needed in the homogeneous regions, more information is needed near the edges.

State of the art techniques: combine adaptive discretizations based on wavelets and appropriate encoding strategies.

Application 2: statistical learning theory

Given a set of data (x_i, y_i), $i = 1, 2, \cdots, m$, drawn independently according to a probability law, build a function f such that $|f(x) - y|$ is small in the average ($E(|f(x) - y|^2)$ as small as possible).

Difficulty: build the adaptive grid from uncertain data, update it as more and more samples are received.

Application 3: adaptive numerical simulation of PDE's

Computing on a non-uniform grid is justified for solutions which displays isolated singularities (shocks).

Difficulty: the solution f is unknown. Build the grid which is best adapted to the solution. Use a-posteriori information, gained throughout the numerical computation.

A basic example

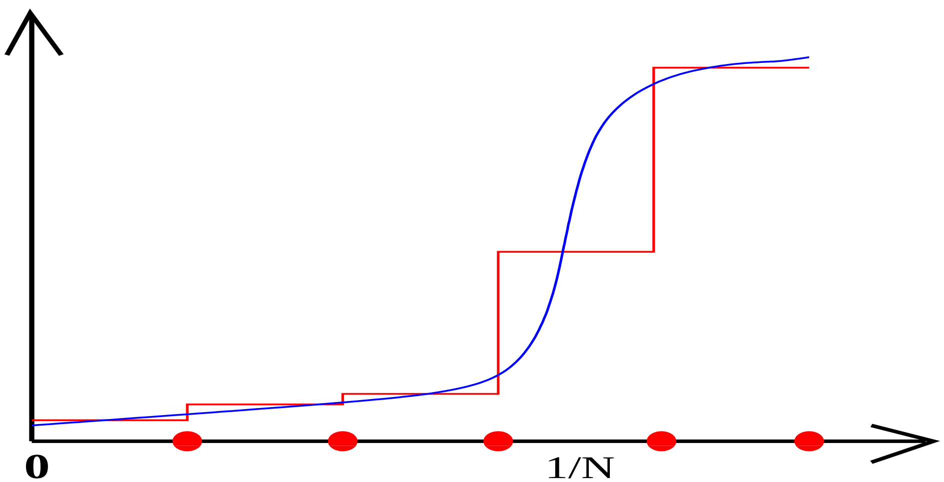
Approximation of $f \in C([0,1])$ by piecewise constant functions on a partition $I_1, \cdots, I_N$, defining

$$f_N(x) = a_k := |I_k|^{-1} \int_{I_k} f, \quad \text{if} \;\; x \in I_k.$$

Local error: $\|f - a_k\|_{L^\infty(I_k)} \leq \max_{x,y \in I_k} |f(x) - f(y)|$

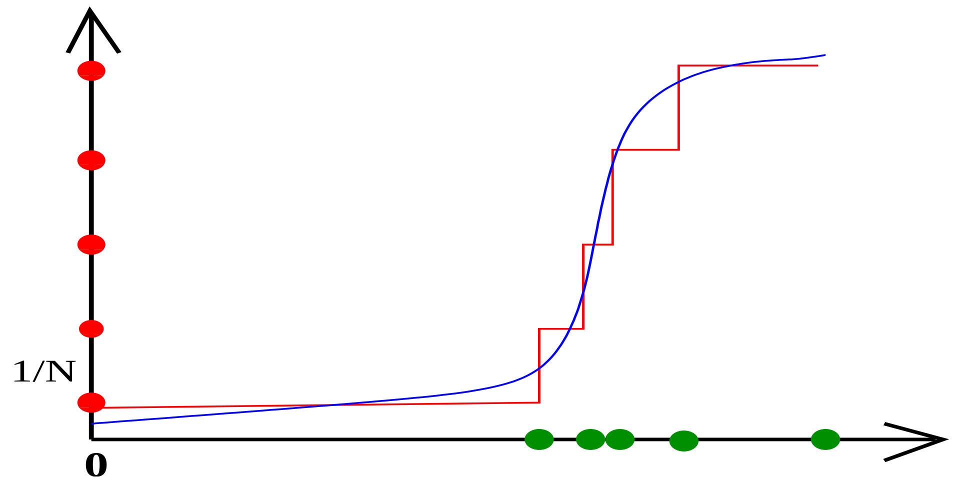
Linear case: $I_k = [\frac{k}{N}, \frac{k+1}{N}]$ uniform partition.

$$f' \in L^\infty \Leftrightarrow \|f - f_N\|_{L^\infty} \leq CN^{-1} \quad (C = \sup |f'|).$$



Nonlinear case: I_k free partition. If $f' \in L^1$, choose the partition such that equilibrates the total variation $\int_{I_k} |f'| = N^{-1} \int_0^1 |f'|$.

$$f' \in L^1 \Leftrightarrow \|f - f_N\|_{L^\infty} \le C N^{-1} \quad (C = \int_0^1 |f'|).$$



Approximation rate governed by differents smoothness spaces !

Example: $f(t) = t^\alpha$ with $0 < \alpha < 1$, then $f'(t) = \alpha t^{\alpha-1}$ is in L^1, not in L^∞. Nonlinear approximation rate N^{-1} outperforms linear approximation rate $N^{-\alpha}$.

Towards an algorithm: equilibrating the error

Fix a tolerance $\varepsilon > 0$ and build a partition $I_1, \cdots, I_N$ such that

$$\varepsilon/2 \leq \|f - a_k\|_{L^\infty(I_k)} \leq \varepsilon.$$

Thus $\|f - f_N\|_{L^\infty} \leq \varepsilon$. If in addition $f' \in L^1$, then

$$\int |f'| \geq \sum_{k=1}^{N} \int_{I_k} |f'| \geq \sum_{k=1}^{N} \|f - a_k\|_{L^\infty(I_k)} \geq N\varepsilon/2,$$

and therefore

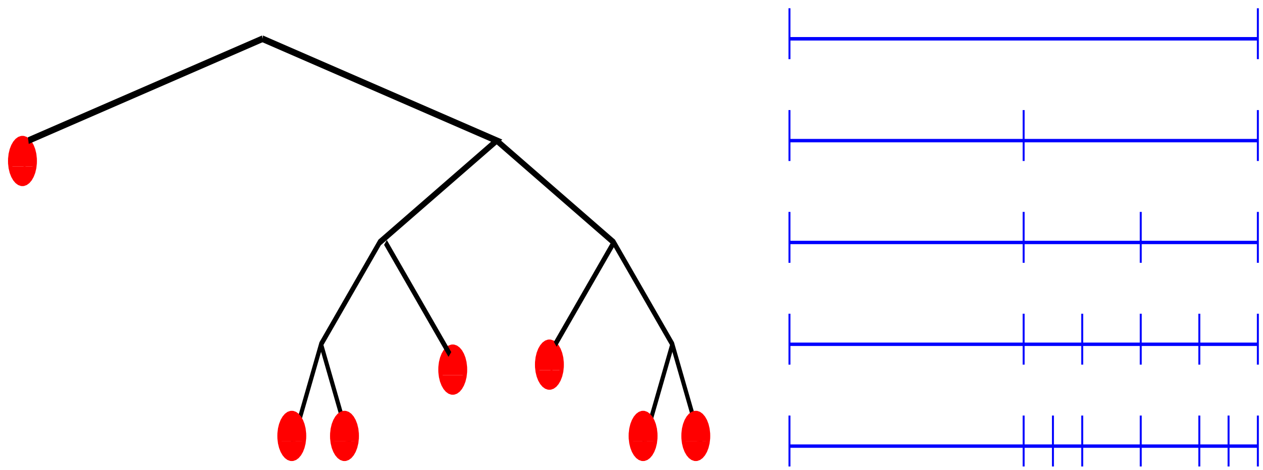
$$\|f - f_N\|_{L^\infty} \leq \varepsilon \leq 2CN^{-1},$$

with $C = \int_0^1 |f'|$.

Can we achieve this in practice by a simple algorithm ?

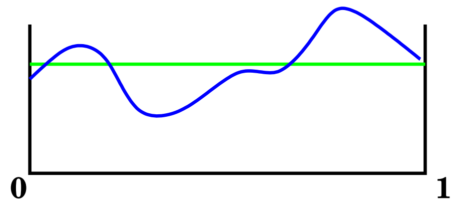
Adaptive greedy splitting

Split intervals I into two equal parts as long as $\|f - a_I\|_{L^\infty(I)} > \varepsilon$, the final adaptive partition is built when $\|f - a_I\|_{L^\infty(I)} \leq \varepsilon$ holds for all intervals (leaves of the decision tree).

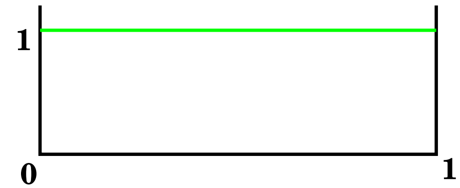


Limitation to dyadic intervals. In turn $f' \in L^1$ is not sufficient to ensure that $\|f - f_N\|_{L^\infty} \lesssim N^{-1}$, but it can be shown that a slightly stronger condition ($f' \in L(\log L)$ or L^p for any $p > 1$) suffices.

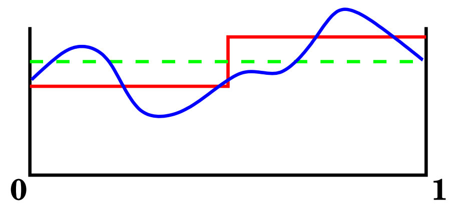
Multiscale decompositions into wavelet bases: the Haar system



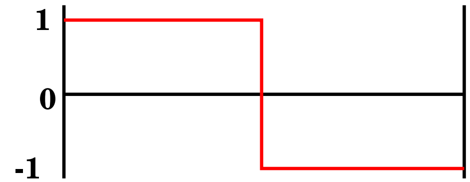
$$\mathbf{f} = \langle \mathbf{f}, \mathbf{e}_0 \rangle \, \mathbf{e}_0$$



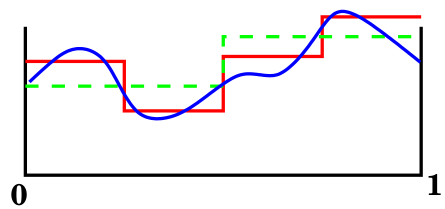
$\mathbf{e}_0$



$$+ \langle \mathbf{f}, \mathbf{e}_1 \rangle \, \mathbf{e}_1$$



$\mathbf{e}_1$



$$+ \langle \mathbf{f}, \mathbf{e}_2 \rangle \, \mathbf{e}_2 + \langle \mathbf{f}, \mathbf{e}_3 \rangle \, \mathbf{e}_3$$

$$\cdots = \sum_{\lambda} \mathbf{f}_\lambda \psi_\lambda \qquad \mathbf{f}_\lambda := \langle \mathbf{f}, \psi_\lambda \rangle$$

$$\psi_\lambda(x) := 2^{j/2}\psi(2^j x - k), \quad \lambda = (j,k), \; j \geq 0, \; k \in \mathbb{Z}, \quad |\lambda| = \mathrm{j} = \mathrm{j}(\lambda).$$

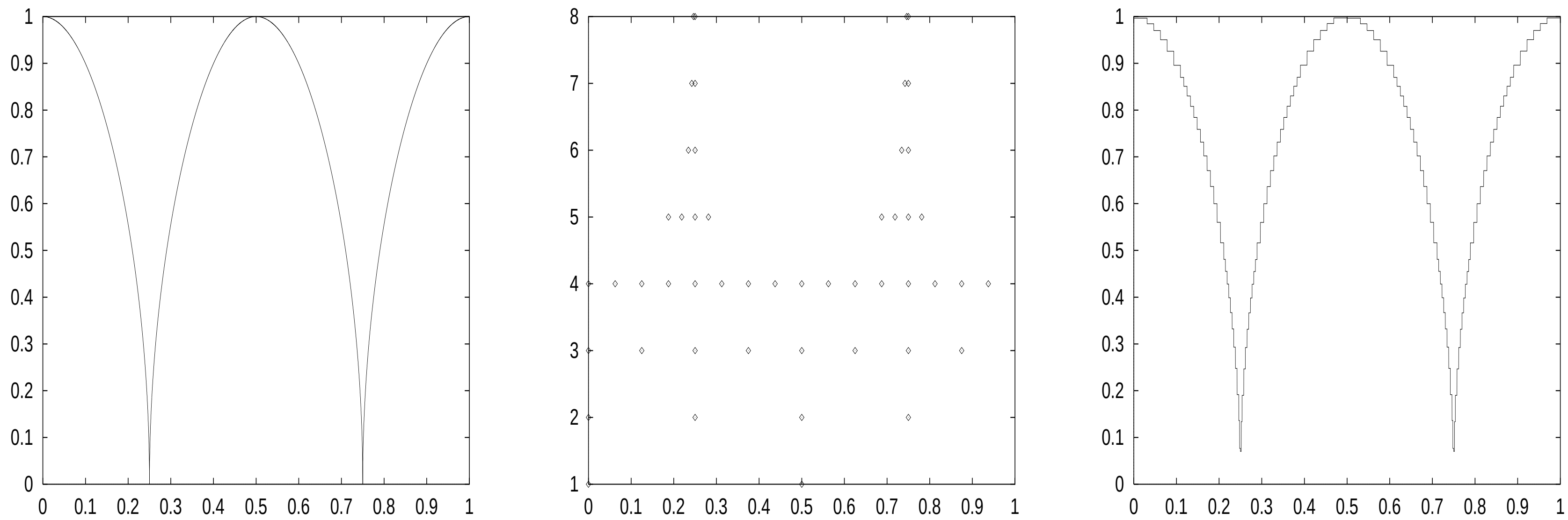
More general wavelets are constructed from similar multiscale approximation processes, using smoother functions such as splines, finite elements... In d dimension $\psi_\lambda(x) := 2^{dj/2}\psi(2^j x - k)$, $k \in \mathbb{Z}^{\mathrm{d}}$.

Approximating functions by wavelet bases

- Linear (uniform) approximation at resolution level j by taking the truncated sum $f \mapsto P_j f := \sum_{|\lambda|<j} f_\lambda \psi_\lambda$.

- Nonlinear (adaptive) approximation obtained by thresholding

$$f \mapsto \mathcal{T}_\Lambda f := \sum_{\lambda \in \Lambda} f_\lambda \psi_\lambda, \quad \Lambda = \Lambda(\eta) = \{\lambda \ \text{s.t.} \ |f_\lambda| \geq \eta\}.$$



Wavelet thresholding applied to an image

Decomposition and reconstruction with 4096 largest coefficients.

Sparse representations (significant coefficients concentrated near the edges) $\Rightarrow$ adaptive approximation by thresholding. Results in important applications in image processing (compression, denoising).

Wavelet analysis of local smoothness

- If f is bounded on $S_\lambda := \mathrm{Supp}(\psi_\lambda)$, an obvious estimate is

$$|f_\lambda| = |\langle f, \psi_\lambda \rangle| \le \sup_{t \in S_\lambda} |f(t)| \int |\psi_\lambda| = 2^{-|\lambda|/2} \sup_{t \in S_\lambda} |f(t)|.$$

- If f is C^1 on S_λ, a finer estimate is

$$\begin{aligned} |f_\lambda| &= \inf_{c \in \mathbb{R}} |\langle f - c, \psi_\lambda \rangle| \\ &\le \inf_{c \in \mathbb{R}} \|f - c\|_{L^\infty(S_\lambda)} \|\psi_\lambda\|_{L^1} \\ &\le 2^{-3|\lambda|/2} \sup_{t \in S_\lambda} |f'(t)|. \end{aligned}$$

- If f is Hölder continuous of exponent α on S_λ, i.e. $|f(x) - f(y)| \le C|x - y|^\alpha$, for some $\alpha \in (0, 1]$, we have the intermediate estimate $|f_\lambda| \le C 2^{-|\lambda|(\alpha+1/2)}$.

Decay of wavelet coefficients influenced by local smoothness.

Fourier analysis of global smoothness

Decomposition of a (1-periodic) function in Fourier series
$f(t) = \sum_{n=\in \mathbb{Z}} c_n(f) e^{i2\pi nt}$, with $c_n(f) := \int_0^1 f(t) e^{-i2\pi nt} dt$.

If $f, f', \cdots, f^{(m)}$ are continuous over $\mathbb{R}$, we can apply n times the integration by part to obtain

$$\begin{aligned} |c_n(f)| &= |(i2\pi n)^{-1} c_n(f')| \\ &= \cdots |(i2\pi n)^{-m} c_n(f^{(m)})| \\ &\leq |i2\pi n|^{-m} \int_0^1 |f^{(m)}| \lesssim n^{-m}. \end{aligned}$$

$\Rightarrow$ Fast decay if f is smooth.

However, if f is smooth everywhere except at some discontinuity point $x \in [0,1]$, we cannot hope better than $|c_n(f)| \lesssim n^{-1}$

Decay of Fourier coefficients influenced by global smoothness.

Wavelet representations are thus more appropriate for piecewise smooth functions.

Summary

- Adaptive methods relate to nonlinear approximation

- Adaptive partitions requires less smoothness than uniform partitions for a given convergence rate

- Adaptive splitting algorithm: aims to equilibrate the local error

- Wavelet thresholding builds an adaptive partition

- Thresholding might not be effective in other bases

A general framework for wavelet bases

Mallat and Meyer (1986): a multiresolution approximation (MRA) is a sequence of nested spaces $V_j \subset V_{j+1} \subset \cdots$ of $L^2(\mathbb{R}^d)$, such that:

- $\overline{\cup V_j} = L^2$, i.e. $\lim_{j\to+\infty} \|f - P_j f\|_{L^2} = 0$ for all $f \in L^2$ where P_j is the L^2-orthogonal projector.

- There exists a scaling function $\varphi \in V_0$ such that

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k), \;\; k \in \mathbb{Z}^d,$$

constitute a Riesz basis of V_j (Riesz basis in Hilbert spaces: basis (e_n) such that $\|(x_n)\|_{\ell^2} \sim \|\sum x_n e_n\|_H$).

For piecewise constant functions we had $\varphi = \chi_{[0,1]}$. In this case

$$\|f - P_j f\|_{L^p} \leq 2^{-j}\|f'\|_{L^p},$$

but no better rate such as $2^{-mj}\|f^{(m)}\|_p$ (first order accuracy).

Raising the accuracy: V_j should contain higher order polynomials.
Example : B-spline of degree N

$$\varphi(x) = \chi_{[0,1]} * \cdots * \chi_{[0,1]} = (*)^{N+1}\chi_{[0,1]},$$

Remark: except for $N = 0$, the functions $\varphi_{j,k}$ are not orthogonal. In turn the orthogonal projector P_j is not local. New difficulties:

- Define numerically simple projectors P_j onto V_j.

- Construct wavelet bases (ψ_λ) which characterize the difference between two successive levels of projection so that

$$f = P_0 f + \sum_{j\geq 0} Q_j f, \;\; Q_j f := P_{j+1}f - P_j f = \sum_{|\lambda|=j} f_\lambda \psi_\lambda$$

Recall that $\psi_\lambda(x) = 2^{dj/2}\psi(2^j x - k)$ and $|\lambda| := j$ when $\lambda = (j,k)$.

Several approaches: orthogonal wavelets, biorthogonal wavelets, finite element wavelets...

Wavelet characterizations of functions spaces

Let $f = \sum f_\lambda \psi_\lambda$, $f_\lambda = \langle f, \tilde{\psi}_\lambda \rangle$.

- L^2 characterized by $\|f\|_{L^2}^2 \sim \|P_0 f\|_{L^2}^2 + \sum_{j\geq 0} \|Q_j f\|_{L^2}^2 \sim \sum |f_\lambda|^2$.

- Sobolev space $H^t = W^{t,2}$ characterized by

$$\|f\|_{H^t}^2 \sim \|P_0 f\|_{L^2}^2 + \sum_{j\geq 0} 2^{2tj} \|Q_j f\|_{L^2}^2 \sim \sum 2^{2t|\lambda|} |f_\lambda|^2 \sim \sum \|f_\lambda \psi_\lambda\|_{H^t}^2.$$

Hints: (i) $\psi_\lambda^{(t)}(x) = 2^{t|\lambda|}(\psi^{(t)})_\lambda(x)$, (ii) $\|f\|_{H^t}^2 \sim \int (1 + |\omega|^{2t}) |\hat{f}(\omega)|^2$

- Besov-Sobolev space $B_{p,p}^t$ characterized by

$$\begin{aligned} \|f\|_{B_{p,p}^t}^p &\sim \|P_0 f\|_{L^p}^p + \sum_{j\geq 0} 2^{ptj} \|Q_j f\|_{L^p}^p \sim \sum 2^{pt|\lambda|} \|f_\lambda \psi_\lambda\|_{L^p}^p \\ &\sim \sum 2^{pt|\lambda|} 2^{pd(1/2-1/p)|\lambda|} |f_\lambda|^p \sim \sum \|f_\lambda \psi_\lambda\|_{B_{p,p}^t}^p. \end{aligned}$$

Remark: $B_{p,p}^t = W^{t,p}$ if $t \notin \mathbb{N}$ or $p = 2$ and $B_{\infty,\infty}^t = C^t$ if $t \notin \mathbb{N}$.

All this holds provided that ψ_λ has enough smoothness

Linear multiscale approximation

From the characterization of H^t, we get $\|Q_j f\|_{L^2} \lesssim 2^{-jt}\|f\|_{H^t}$ and therefore

$$f \in H^t \Rightarrow \|f - P_j f\|_{L^2} \leq \sum_{l \geq j} \|Q_l f\|_{L^2} \lesssim 2^{-tj}.$$

and in a similar manner

$$f \in W^{t,p} \Rightarrow \|f - P_j f\|_{L^p} \lesssim 2^{-tj}.$$

We actually have a finer result

$$f \in B^t_{p,q} \Leftrightarrow (2^{tj}\|f - P_j f\|_{L^p})_{j \geq 0} \in \ell^q.$$

Besov spaces are thus characterized from the rate of linear multiscale approximation.

These results are very similar to (uniform) finite element approximation since $V_j \sim V_h$ with $h \sim 2^{-j}$.

Finite element approximation results

- V_h: finite element space based on a uniform discretization of a domain $\Omega \subset \mathbb{R}^d$ with mesh size h.

- $N := \dim(V_h) \sim \mathrm{vol}(\Omega) h^{-d}$

- $W^{s,p} := \{f \in L^p(\Omega) \ \text{s.t.} \ D^\alpha f \in L^p(\Omega), \ |\alpha| \leq s\}$

Classical finite element approximation theory (Bramble-Hilbert, Ciarlet-Raviart, Deny-Lions, Strang-Fix): provides with the classical estimate

$$f \in W^{s+t,p} \Rightarrow \inf_{g \in V_h} \|f - g\|_{W^{s,p}} \leq C h^t \sim C N^{-t/d},$$

assuming that V_h has enough polynomial reproduction and is contained in $W^{s,p}$.

Measuring sparsity in a representation $f = \sum f_\lambda \psi_\lambda$

Intuition: the number of coefficients above a threshold η should not grow too fast as $\eta \to 0$.

Weak spaces: $(f_\lambda) \in w\ell^p$ if and only if

$$\mathrm{Card}\{\lambda \text{ s.t. } |f_\lambda| > \eta\} \leq C\eta^{-p},$$

or equivalently, the decreasing rearrangement $(f_n^*)_{n>0}$ of $(|f_\lambda|)$ satisfies

$$f_n^* \leq C n^{-1/p}.$$

The representation is sparser as $p \to 0$. If $p < 2$ and (ψ_λ) is an orthonormal basis, an equivalent statement is in terms of best N-term approximation: if $f_N := \sum_{N \text{ largest } |f_\lambda|} f_\lambda \psi_\lambda$, then

$$\|f - f_N\|_{L^2} = [\sum_{n \geq N} |f_n^*|^2]^{1/2} \lesssim N^{-s}, \ \ 1/p = s + 1/2.$$

Nonlinear wavelet approximation in L^2

Recall that $B^t_{p,p}$ is characterized by

$$\|f\|^p_{B^t_{p,p}} \sim \sum 2^{pt|\lambda|} 2^{pd(1/2-1/p)|\lambda|} |f_\lambda|^p$$

Assume that $f \in B^t_{p,p}$ with $1/p = 1/2 + t/d$. In this case

$$\|f\|_{B^t_{p,p}} \sim \|(f_\lambda)\|_{\ell^p},$$

and therefore $(f_\lambda) \in w\ell^p$. If $f_N := \sum_{N \text{ largest } |f_\lambda|} f_\lambda \psi_\lambda$, we have

$$\|f - f_N\|_{L^2} \lesssim N^{-t/d}.$$

For linear approximation, the same rate is achieved under the stronger condition $f \in H^t$.

Nonlinear approximation results

N-terms approximations: $\Sigma_N := \{\sum_{\lambda\in\Lambda} d_\lambda\psi_\lambda \; ; \; \#(\Lambda) \leq N\}$.

- Rate of decay governed by weaker smoothness conditions (DeVore): with $1/q = 1/p + t/d$

$$f \in W^{s+t,q} \Rightarrow \inf_{g\in\Sigma_N} \|f-g\|_{W^{s,p}} \leq CN^{-t/d},$$

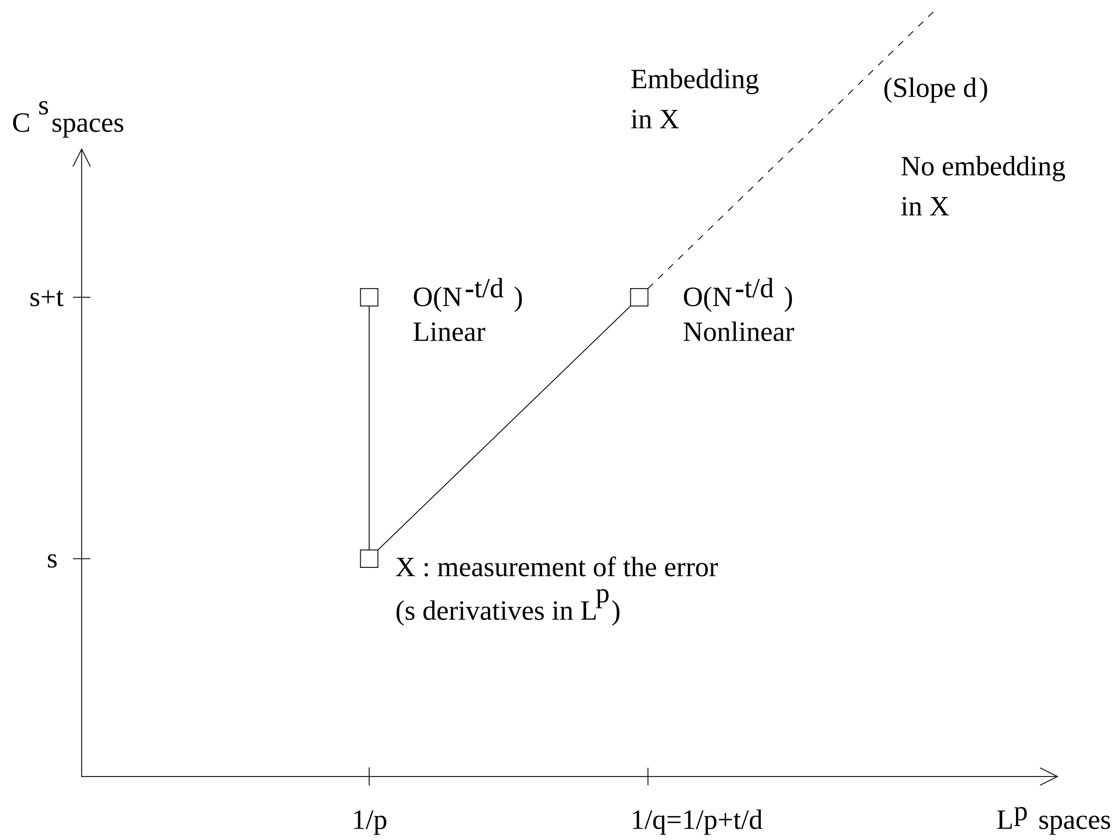
- For most error norm X (e.g. L^p, $W^{s,p}$, $B^s_{p,q}$), a near optimal approximation is obtained by thresholding: if $f = \sum_\lambda f_\lambda\psi_\lambda$, and $f_N := \sum_{N \text{ largest } \|f_\lambda\psi_\lambda\|_X} f_\lambda\psi_\lambda$, we then have

$$\|f - f_N\|_X \leq C \inf_{g\in\Sigma_N} \|f-g\|_X$$

with C independent of f and N.

- Remark: similar theory for adaptive finite element on N triangles with isotropy constraints (minimal angle condition).

Pictorial interpretation of approximation results



Greedy bases

Let (ψ_λ) be a basis in a Banach space X with $\|\psi_\lambda\|_X = 1$ for all λ.

The basis is greedy if and only if for all $f \in X$ and $N > 0$,

$$\|f - \sum_{N \text{ largest } |f_\lambda|} f_\lambda \psi_\lambda\|_X \leq C \inf_{g \in \Sigma_N} \|f - g\|_X.$$

The basis is unconditional if and only there exists $C > 0$ such that

$$|x_\lambda| \leq |y_\lambda| \text{ for all } \lambda \Rightarrow \|\sum x_\lambda \psi_\lambda\|_X \leq C \|\sum y_\lambda \psi_\lambda\|_X.$$

The basis is democratic if and only if there exists $C > 0$ such that

$$\#(E) = \#(F) \Rightarrow \|\sum_{\lambda \in E} \psi_\lambda\|_X \leq C \|\sum_{\lambda \in F} \psi_\lambda\|_X.$$

Two results due to Temlyakov:

1. Greedy $\Leftrightarrow$ unconditional and democratic.
2. Wavelet are democratic in L^p and $W^{m,p}$ when $1 < p < +\infty$.

General program for PDE's

- Theoretical: revisit regularity theory for PDE's. Solutions of certain PDE's might have substantially higher regularity in the scale governing nonlinear approximation than in the scale governing linear approximation. Examples : hyperbolic conservation laws (DeVore and Lucier 1987), elliptic problems on corner domains (Dahlke and DeVore, 1997).

- Numerical: develop for the unknown u of the PDE $\mathcal{F}(u) = 0$ appropriate adaptive resolution strategies which perform essentially as well as thresholding : produce $\tilde{u}_N$ with N terms such that $\|u - \tilde{u}_N\|$ has the same rate of decay N^{-s} as $\|u - u_N\|$ in some prescribed norm, if possible in $\mathcal{O}(N)$ computation.

Remark: similar goals can be formulated for adaptive finite elements with N being the number of elements.

Adaptive finite element approximation theory

In all the following we will work with the error metric

$$X = L^p(\Omega).$$

For simplicity we take

$$\Omega = [0,1]^2,$$

and we only work with piecewise affine finite elements.

We want to discuss the differences in approximation capabilities between :

(i) Uniform and isotropic (shape regular) triangulations

(ii) Adaptive and isotropic triangulations

(iii) Adaptive and anisotropic triangulations

Uniform and isotropic triangulations

If $(\mathcal{T}_h)_{h>0}$ is a family of uniform and isotropic triangulations, and V_h the corresponding piecewise affine finite element space, then

$$f \in W^{s,p} \Rightarrow \inf_{f_h \in V_h} \|f - f_h\|_{L^p} \le C h^{\min\{s,2\}} |f|_{W^{s,p}}.$$

Since $N = \#(\mathcal{T}_h) \sim h^{-2}$, this gives the convergence rate $N^{-\frac{\min\{s,2\}}{2}}$. In particular

$$f \in W^{2,p} \Rightarrow \|f - f_N\|_{L^p} \le C N^{-1} |f|_{W^{2,p}},$$

and

$$f \in C^2 \Rightarrow \|f - f_N\|_{L^\infty} \le C N^{-1} |f|_{C^2}.$$

Remark : almost an "if and only if" result (one needs $B^s_{p,\infty}$ in place of $W^{s,p}$)

Adaptive and isotropic triangulations

Consider here $X = L^\infty$. If R is a reference triangle and I_R the interpolation operator, we have by Sobolev imbedding,

$$\|f - I_R f\|_{L^\infty(R)} \lesssim \|f - I_R f\|_{W^{2,1}(R)},$$

and this by Bramble-Hilbert lemma

$$\|f - I_R f\|_{L^\infty(R)} \lesssim |f|_{W^{2,1}(R)}.$$

The constant in this estimate is invariant by isotropic scaling : for any isotropic triangle T

$$\|f - I_T f\|_{L^\infty(T)} \leq C|f|_{W^{2,1}(T)}.$$

Given $f \in C(\Omega)$, assume that for any prescribed $\varepsilon > 0$, we can find a triangulation $\mathcal{T}_N$, with $N = \#(\mathcal{T}_N) = N(\varepsilon)$ such that the local error is equidistributed in the sense that

$$\varepsilon/2 \le \|f - I_T f\|_{L^\infty(T)} \le \varepsilon, \quad T \in \mathcal{T}_N.$$

Then obviously $\|f - f_N\|_{L^\infty} \le \varepsilon$. Moreover if $f \in W^{2,1}$, we have

$$N\varepsilon/2 \le \sum_{T \in \mathcal{T}_N} \|f - I_T f\|_{L^\infty(T)} \le C \sum_{T \in \mathcal{T}_N} |f|_{W^{2,1}(T)} = C|f|_{W^{2,1}},$$

and therefore

$$f \in W^{2,1} \Rightarrow \|f - f_N\|_{L^\infty} \le CN^{-1}|f|_{W^{2,1}}.$$

The rate of smoothness N^{-1} is governed by weaker smoothness condition than for uniform partition.

For $X = L^p$: equidistributing the local L^p error yields

$$f \in W^{2,q} \Rightarrow \|f - f_N\|_{L^p} \le CN^{-1}|f|_{W^{2,q}}, \quad 1/q = 1/p + 1.$$

A greedy approach to error equidistribution

1) Given $f \in L^p$ and some prescribed $\varepsilon > 0$, we start from an initial coarse triangulation $\mathcal{T}_2$ (split Ω into two triangles).

2) Given $\mathcal{T}_k$ we consider the triangle T where the error $\|f - f_k\|_{L^p(T)}$ is maximal. If it is larger than ε, then split T into four sub-triangles of similar shape using the three midpoints. This gives a new (generally non-conforming) triangulation $\mathcal{T}_{k+3}$.

3) Stop when all triangles have local error less than ε.

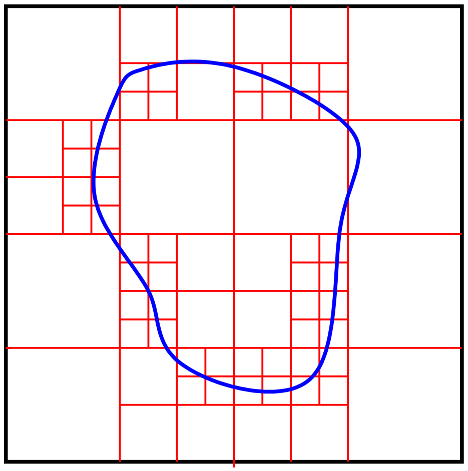
This does not exactly equidistributes the error, but one has for $X = L^p$ and any q such that $1/q > 1/p + 1$

$$f \in W^{2,q} \Rightarrow \|f - f_N\|_{L^p} \leq CN^{-1}|f|_{W^{2,q}}.$$

Remark : the triangulation can be made conforming without changing the convergence rate.

When do we need anisotropy ?

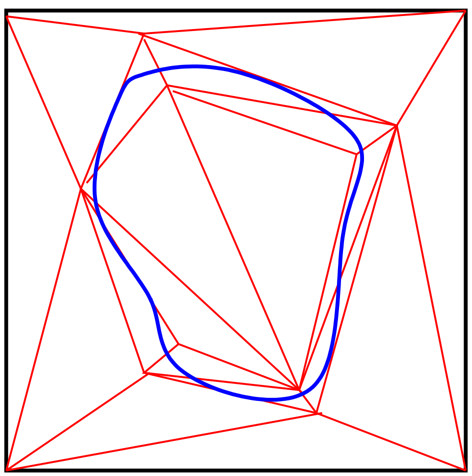
Sharp gradients or jump discontinuities on curved edges : $f = \chi_\Omega$, with $\partial\Omega$ smooth.



f_N = piecewise affine function

on N optimally selected squares

$\Rightarrow \|f - f_N\|_{L^2} \sim N^{-1/2}$



f_N = piecewise affine function

on N optimaly selected triangles

$\Rightarrow \|f - f_N\|_{L^2} \sim N^{-1}$

$C^n - C^m$ models

The function f is $C^n - C^m$ if it is piecewise C^n with jump discontinuities on piecewise C^m curves.

If $f \in C^n - C^m$, then there exists triangulations $(\mathcal{T}_N)_{N>0}$ such that

$$\|f - f_N\|_{L^p} \lesssim N^{-\frac{\min\{n,2\}}{2}} + N^{-m/p}.$$

In particular if $f \in C^2 - C^2$, the rate is N^{-1} in L^p for $p \leq 2$.

Drawbacks of this model:

- lacks a rigourous quantitative definition

- does not describe smooth yet sharp transitions.

- does not lead to a natural algorithm

More quantitative models based on the regularity of level sets (DeVore, Petrova, Wojtaszczyk)

Hessian based models

A very heuristic computation:

A "good" triangle around x has aspect ratio of the ellipsoid

$$E(x) := \{\langle H(x)v, v\rangle \le 1\}$$

where $H(x) = |D^2 f(x)|$, i.e. it is an isotropic triangle with respect to the distorted metric induced H.

For such a triangle T, if (λ_1, λ_2) are the eigenvalues of H and (h_1, h_2) the heights of T in the corresponding directions, we have $h_1/h_2 \approx \sqrt{\lambda_2/\lambda_1}$. Therefore and if $D^2 f$ does not vary too much on T one has

$$\begin{aligned} \|f - I_T f\|_{L^\infty(T)} &\le \lambda_1 h_1^2 + \lambda_2 h_2^2 \\ &\approx h_1 h_2 \sqrt{\lambda_1 \lambda_2} \\ &\approx |T|(\det(H(x)))^{1/2} \approx \int_T \sqrt{\det(H(x))}. \end{aligned}$$

Now, assuming that

$$E(f) := \int_\Omega \sqrt{\det(H(x))} < +\infty$$

and that $\mathcal{T}_N$ is designed such that each triangle has the optimal aspect ratio and $\int_T \sqrt{\det(H(x))} \approx N^{-1}E(f)$, we obtain

$$\|f - f_N\|_{L^\infty} \leq CN^{-1}E(f).$$

By similar heuristics for $X = L^p$, we can obtain adaptive anisotropic triangulations with error estimates of the type

$$\|f - f_N\|_{L^p} \leq CN^{-1}\|\sqrt{\det(H)}\|_{L^q}, \;\; 1/q = 1/p + 1.$$

Non-linear quantities : $E(f+g)$ not controlled by $E(f) + E(g)$.

Making it more rigourous (Shen-Sun-Xu, Babenko)

- $E(f) = 0$ for a degenerate Hessian (univariate f), yet error is non-zero : replace H by a majorant of the type $H + \varepsilon I$.

- Requires enough smoothness on f so that the triangulation $\mathcal{T}_N$ can indeed be constructed for $N > N_0(f, \varepsilon)$

Two drawbacks:

- The construction of the triangulation is based on the Hessian : not robust to noise, does not apply to arbitrary L^p functions.

- The construction is non-hierarchical.

A greedy alternative (Dyn, Hecht, Mirebeau, A.C.): Coarse triangulation $\Rightarrow$ select triangle with largest local L^p error $\Rightarrow$ choose the mid-point bisection that best reduces this error
$\Rightarrow$ split $\Rightarrow$ iterate.... until prescribed accuracy or number of triangles is met.