Nonlinear approximation and adaptive multiscale methods for PDE's

Albert Cohen

Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie
Paris

www.ann.jussieu.fr/~cohen

CEMRACS 2003

Adaptive methods in numerical computation

Multiscale : the discretization is locally refined in the area where the physical process involves fine scales, such as shocks, singularities, high gradients, fast oscillations...

Nonlinear : the local refinement is not fixed a-priori but rather driven by information gained through the computation process.

Theoretical pillar : nonlinear approximation theory.

Numerical tool : wavelet bases.

Agenda

1. Nonlinear approximation and wavelet : basics.
2. Nonlinear approximation and wavelet : advanced.
3. Adaptive space refinement schemes
4. Adaptive postprocessing schemes

Some references

Ingrid Daubechies, “Ten lectures on wavelets”, SIAM, 1992.

Stéphane Mallat, “A wavelet tour of signal processing”, Academic Press, 1998.

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

Wolfgang Dahmen, “Wavelet and multiscale methods for operator equations”, Acta Numerica, 1997.

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

Basic notations

- Lebesgue spaces: $L^p(\Omega) := \{f \; ; \; \int_\Omega |f|^p < \infty\}$ for $p \geq 1$. Banach space when equiped with the norm $\|f\|_p := [\int_\Omega |f|^p]^{1/p}$.

- $L^\infty(\Omega)$ (almost everywhere uniformly bounded functions) and $C(\Omega)$ (continuous functions): Banach spaces when equiped with the norm $\|f\|_\infty := \sup_{x\in\Omega} |f(x)|$.

- Hilbert space in the case $p = 2$: $\|f\|_2 = [\langle f, f\rangle]^{1/2}$ with $\langle f, g\rangle := \int_\Omega f\,\overline{g}$.

- Characteristic functions: $\chi_\Omega(x) = 1$ if $x \in \Omega$, 0 otherwise.

- Change of variable: $f(a\cdot +b) : x \mapsto f(ax + b)$.

- Estimations: $F(p_1, p_2, \cdots) \lesssim G(p_1, p_2, \cdots)$ if there exists $C > 0$ such that $F(p_1, p_2, \cdots) \leq CG(p_1, p_2, \cdots)$ for all $p_1, p_1, \cdots$.

- Equivalences: $F \sim G$ if and only if $F \lesssim G$ and $G \lesssim F$.

Fourier representations

- Analysis: $\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt.$

- Synthesis: $f(t) = (2\pi)^{-1} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{i\omega t} d\omega.$

Representation of f in terms of the pure waves $e_\omega(t) = e^{i\omega t}$, $\omega \in \mathbb{R}$.

For 1-periodic functions:

- Analysis: $c_n(f) = \int_0^1 f(t) e^{-i2\pi nt} dt.$

- Synthesis: $f(t) = \sum_{n \in \mathbb{Z}} c_n(f) e^{i2\pi nt}.$

Discrete Fourier transform: $(x[k])_{k=0,\cdots,N-1}$ and $(\hat{x}[k])_{k=0,\cdots,N-1}$ connected by

$$\hat{x}[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] e^{-i2\pi nk/N} \quad \text{and} \quad x[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \hat{x}[n] e^{i2\pi nk/N}.$$

Implemented in $\mathcal{O}(N \log N)$ operations by FFT.

Fourier representations and computation

Approximation of a (1-periodic) function by its partial sum $S_N f(t) = \sum_{n=-N}^{N} c_n(f) e^{i2\pi nt}$.

Problem: fast convergence ?

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 $|c_n(f)| \lesssim n^{-1}$ (also Gibbs phenomenon for $S_N f$ near the singularity).

Better representations are needed for such functions.

Central problems 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

$$f \in X^r \Leftrightarrow \sigma_N(f) \leq CN^{-r}$$

Problem 2: practical realization of $f \mapsto g \in \Sigma_N$ such that $\|f - g\|_X \lesssim \sigma_N(f)$.

Applications to numerical simulation : in this setting f is not explicitely given but is the unknown of an equation (e.g. PDE) $\mathcal{F}(u) = 0$.

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).

A basic example

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

$$f_N(x) = |I_k|^{-1} \int_{I_k} f, \ \text{ si } \ x \in I_k.$$

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 $\int_{I_k} |f'| = N^{-1} \int_0^1 |f'|$.

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

Approximation rate governed by differents smoothness spaces !

Multiscale approximation : basic 1D example

Approximation of a function $f(t)$, $t \in [0,1]$ by piecewise constant functions on dyadic intervals $I_{j,k} = [2^{-j}k, 2^{-j}(k+1)[$, $k = 0, \cdots, 2^j - 1$,

$$P_j f(t) := a_{j,k} = 2^j \int_{I_{j,k}} f(t)dt, \ \ t \in I_{j,k}.$$

Remark 1: P_j is the L^2-orthogonal projection onto the space V_j of piecewise constant functions on the intervals $I_{j,k}$, $k = 0, \cdots, 2^j - 1$. Indeed an orthonormal basis for this space is provided by

$$\varphi_{j,k} = 2^{j/2}\chi_{I_{j,k}} = 2^{j/2}\varphi(2^j \cdot -k), \ \ k = 0, \cdots, 2^j - 1,$$

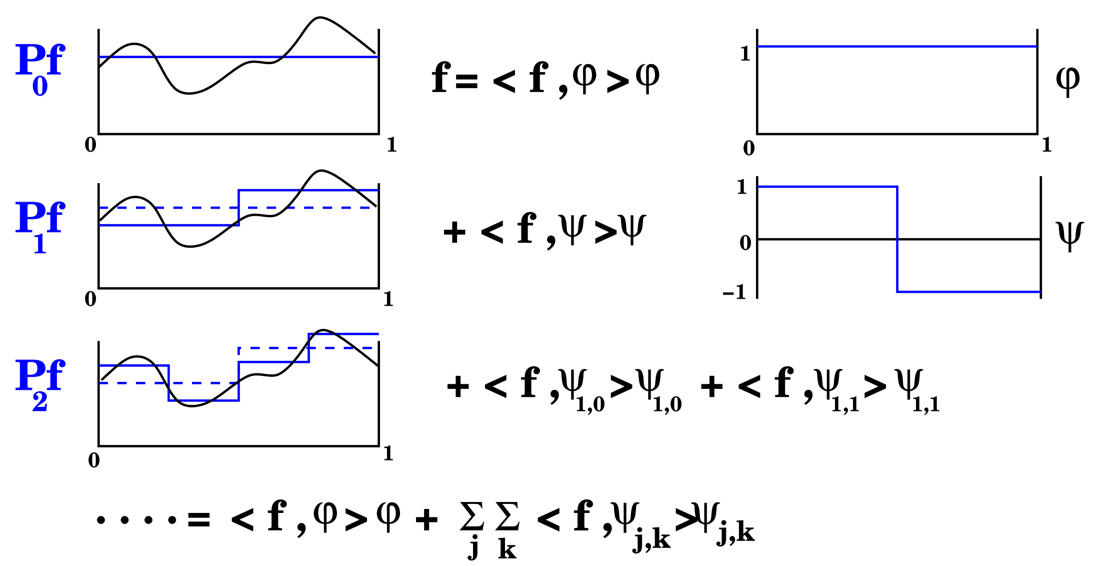
with $\varphi = \chi_{[0,1]}$ and clearly $P_j f = \sum_{k=0}^{2^j-1} \langle f, \varphi_{j,k} \rangle \varphi_{j,k}$.

Remark 2: the spaces V_j are nested i.e. $V_j \subset V_{j+1}$ and $\overline{\cup V_J}^{L^p} = L^p([0,1])$, i.e. $\lim_{J \to +\infty} \|f - P_J f\|_p = 0$ if $f \in L^p([0,1])$.

Multiscale decomposition into the Haar basis

We decompose $P_J f$ into $P_J f = P_0 f + \sum_{j=0}^{J-1} Q_j f$ with $Q_j = P_{j+1} - P_j$ the orthogonal projection onto W_j, the orthogonal complement of V_j into V_{j+1}.

W_j is spanned by $\psi_{j,k} = 2^{j/2}\psi(2^j \cdot -k)$, $k = 0, \cdots, 2^j - 1$, where $\psi = \chi_{[0,1/2]} - \chi_{[1/2,1]}$. Therefore $Q_j f = \sum_{k=0}^{2^j-1} \langle f, \psi_{j,k}\rangle \psi_{j,k}$. Letting $J \to +\infty$, we obtain the decomposition of f in the Haar system



Fast algorithms

Starting point: discretized function at some resolution level J, i.e. $f = \sum_{k=0}^{2^J-1} c_{J,k}\varphi_{J,k} \in V_J$.

Two possible situations: (i) data are directly provided in discrete format $c_J := (c_{J,k})_{k=0,\cdots,2^J-1}$ (e.g. in digital signal or image processing) or (ii) data is a function f with an explicit mathematical expression $\Rightarrow$ compute $c_{J,k} = \langle f, \varphi_{J,k}\rangle$ exactly or approximately.

Problem: fast computation of the coefficients in the multiscale representation $f := c_{0,0}\varphi + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}$.

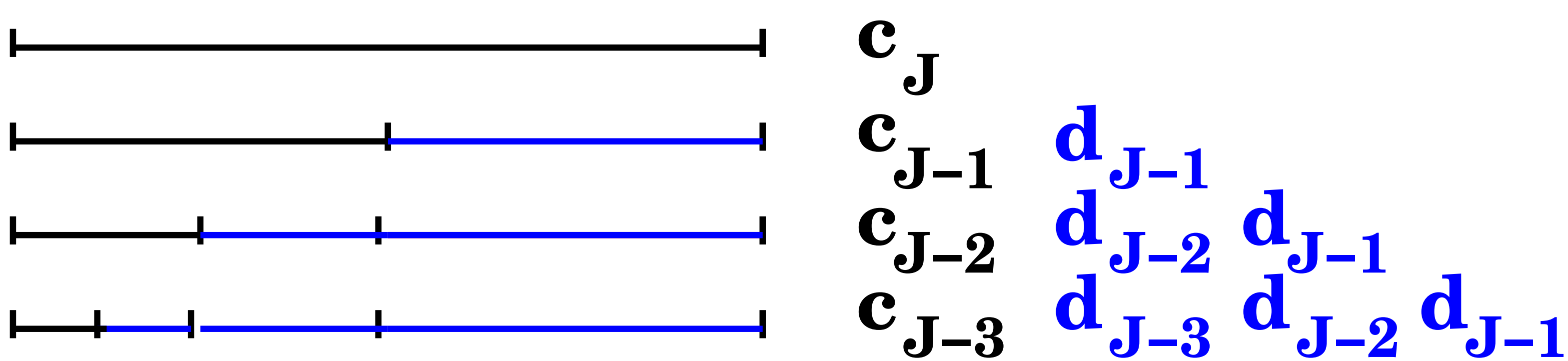
Solution: process hierarchically, using the interscale relations

$$\begin{aligned} c_{j-1,k} &= 2^{-(j-1)/2}a_{j-1,k} = 2^{-(j-1)/2}(a_{j,2k} + a_{j,2k+1})/2 \\ &= (c_{j,2k} + c_{j,2k+1})/\sqrt{2}, \end{aligned}$$

and similarly $d_{j-1,k} = (c_{j,2k} - c_{j,2k+1})/\sqrt{2}$.

This allows the iteration

$c_J \to (c_{J-1}, d_{J-1}) \to \cdots \to (c_0, d_0, d_1, \cdots, d_{J-1})$.



Reconstruction: by inverse formulae

$$c_{j,2k} = (c_{j-1,k} + d_{j-1,k})/\sqrt{2} \text{ and } c_{j,2k+1} = (c_{j-1,k} - d_{j-1,k})/\sqrt{2}.$$

Remark: complexity of step $j \leftrightarrow j-1$ in $\mathcal{O}(2^j) \Rightarrow$ global complexity in $\mathcal{O}(2^J)$, i.e. $\mathcal{O}(N)$ where N is the size of the data.

Compact notations

- Scaling functions and wavelets: $\varphi_{j,k} = \varphi_\lambda$, $\psi_{j,k} = \psi_\lambda$, $\lambda = (j, k)$.

- Scale level: $|\lambda| = j$.

- Coefficients: $c_\lambda = \langle f, \varphi_\lambda \rangle$, $d_\lambda = \langle f, \psi_\lambda \rangle$.

- Projectors: $P_j f = \sum_{|\lambda|=j} c_\lambda \varphi_\lambda = \sum_{|\lambda|<j} d_\lambda \psi_\lambda$ (incorporates the coarse layer of functions φ_λ, $|\lambda| = 0$) and $Q_j f = \sum_{|\lambda|=j} d_\lambda \psi_\lambda$.

Wavelet analysis of local smoothness

- If f is bounded on $I_{j,k}$, an obvious estimate is

$$|d_{j,k}| = |\langle f, \psi_{j,k} \rangle| \leq \sup_{t \in I_{j,k}} |f(t)| \int |\psi_{j,k}| = 2^{-j/2} \sup_{t \in I_{j,k}} |f(t)|.$$

- If f is C^1 on $I_{j,k}$, a finer estimate is

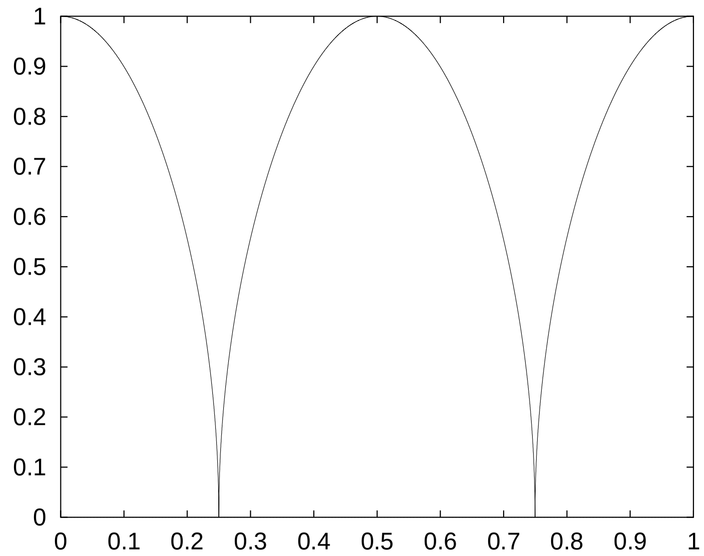
$$\begin{aligned} |d_{j,k}| &= \inf_{c \in \mathbb{R}} |\langle f - c, \psi_{j,k} \rangle| \\ &\leq \inf_{c \in \mathbb{R}} \|f - c\|_{L^\infty(I_{j,k})} \|\psi_{j,k}\|_{L^1} \\ &\leq 2^{-3j/2} \sup_{t \in I_{j,k}} |f'(t)|. \end{aligned}$$

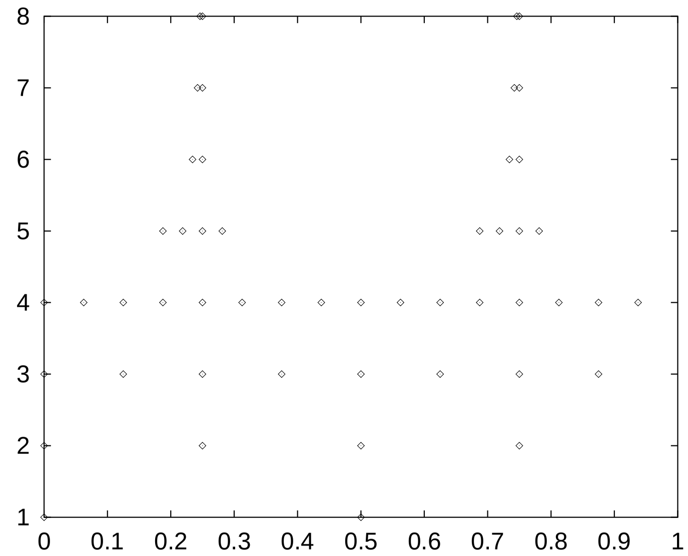
- If f is Hölder continuous of exponent α on $I_{j,k}$, i.e. $|f(x) - f(y)| \leq C|x - y|^\alpha$, for some $\alpha \in]0,1[$, we have the intermediate estimate $|d_{j,k}| \leq C 2^{-j(\alpha+1/2)}$.

Decay of wavelet coefficients influenced by local smoothness.

Example

The function $f(x) = \sqrt{|\cos(2\pi x)|}$ discretized at resolution level $J = 13$ and its coefficients above threshold $\eta = 5 \times 10^{-3}$.





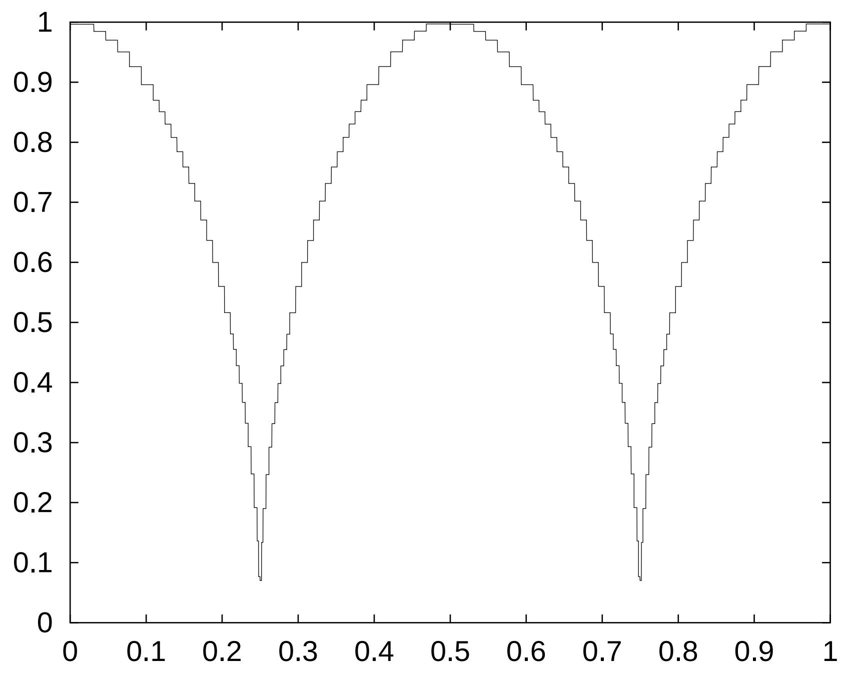
At fine scales, numerically significant coefficients concentrate near the singularities.

Approximating functions by wavelet bases

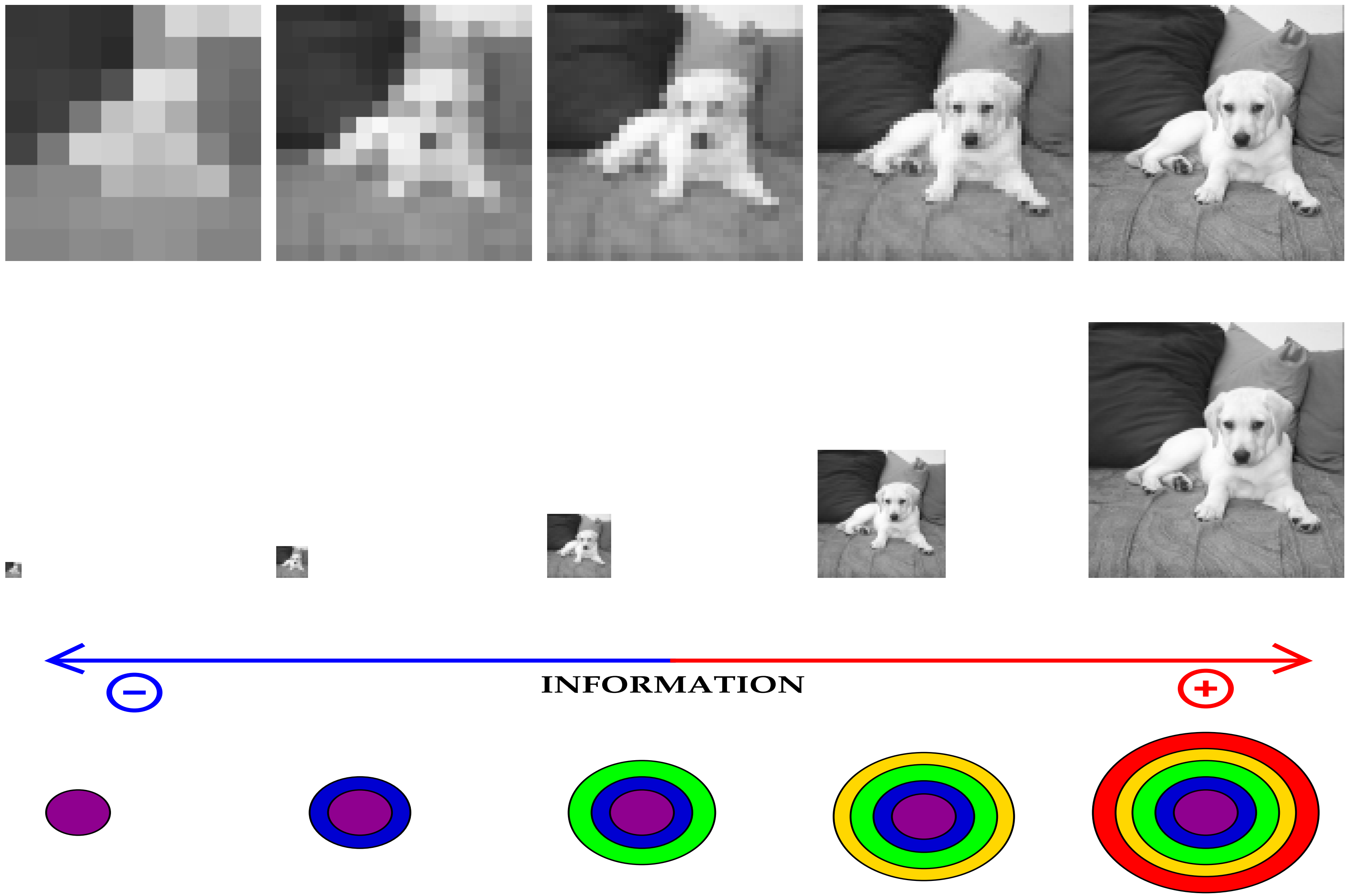
- Linear approximation at resolution level j by taking the truncated sum $f \mapsto P_j f := \sum_{|\lambda|<j} d_\lambda \psi_\lambda$.

- Nonlinear (adaptive) approximation obtained by thresholding

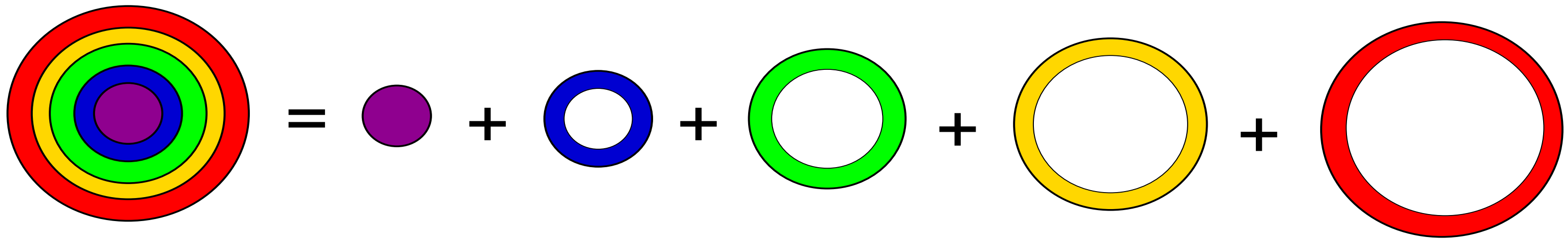
$$f \mapsto \mathcal{T}_\eta f := \sum_{\lambda \in \Lambda} d_\lambda \psi_\lambda, \quad \Lambda = \Lambda(\eta) = \{\lambda \text{ s.t. } |d_\lambda| \geq \eta\}.$$



Hierarchical structure of visual information



Multiscale decomposition



Fine scale image representer by coarse scale approximation + fluctuations at intermediate scales

Decomposition of 2D images

- V_j: piecewise constant on dyadic squares $S_{j,k} = I_{j,k_x} \times I_{j,k_y}$.

- Orthogonal complement spanned by the functions

$$\psi^{\gamma}_{j,k}(x,y) = 2^j \psi^{\gamma}(2^j(x,y) - k), \ \ \gamma = a, b, c,$$

for $k = (k_x, k_y) \in \{0, \cdots, 2^j - 1\}^2$, where $\psi^a(x,y) = \varphi(x)\psi(y)$, $\psi^b(x,y) = \psi(x)\varphi(y)$ and $\psi^c(x,y) = \psi(x)\psi(y)$.

Tensor product algorithm:

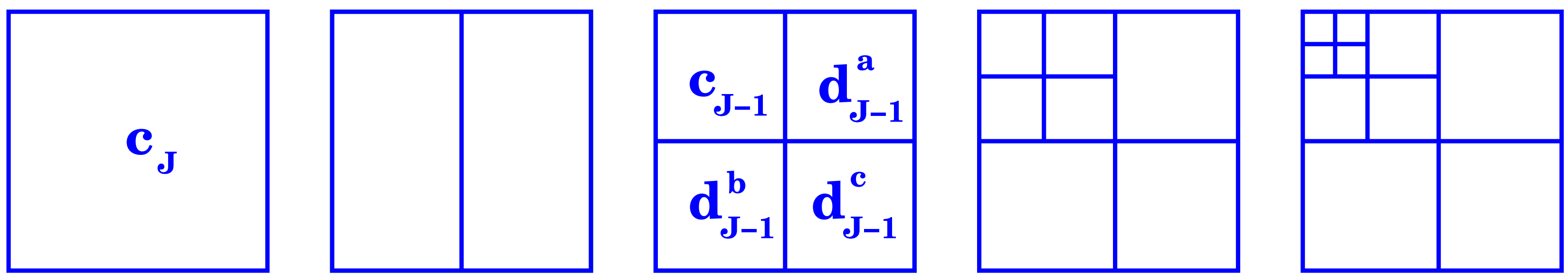


Image $I(k,l) \Rightarrow$ process lines $\Rightarrow$ process columns $\Rightarrow$ Iterate ...

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).

Summary

Important features of the Haar system:

- Multiresolution nested approximation spaces V_j
- Local bases ψ_λ spanning the complement spaces W_j
- Fast $\mathcal{O}(N)$ algorithms
- Decay influenced by local smoothness
- Adaptive approximation by thresholding

Limitations: discontinuous basis functions with poor approximation properties.

A general framework

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

- $\overline{\cup V_j} = L^2$, i.e. $\lim_{j \to +\infty} \|f - P_j f\|_2 = 0$ 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

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

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

Remarks:

- We now work on the whole of $\mathbb{R}$ therefore k runs over $\mathbb{Z}$.

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

Accuracy of MRA spaces

Rate of convergence of $\|f - P_j f\|$ as $j \to +\infty$?

For piecewise constant functions,

$$\|f - P_j f\|_p \leq 2^{-j} \|f'\|_p,$$

but cannot hope for a better rate such as $2^{-mj}\|f^{(m)}\|_p$: piecewise constant approximation is first order accurate.

Raising the accuracy and smoothness: V_j space of piecewise affine functions on the $I_{j,k}$ which are globally C^0. Natural generator: hat function $\varphi = (1 - |x|)_+$.

More generally: splines of degree N, i.e. piecewise polynomials of degree N on the $I_{j,k}$ which are globally C^{N-1}:

Generator: B-spline of degree N

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

New difficulties

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

Two basic questions:

- How to define numerically simple projectors P_j onto V_j ?

- How to construct wavelet bases which characterize the difference between two successive levels of projection ?

Several approaches (by order of generality): orthogonal wavelets, biorthogonal wavelets, generalized wavelets.

A fundamental remark

Scaling function $\varphi \in V_0 \subset V_1$ should satisfy a two scale equation

$$\varphi(t) = \sum_{n \in \mathbb{Z}} h_n \varphi(2t - n)$$

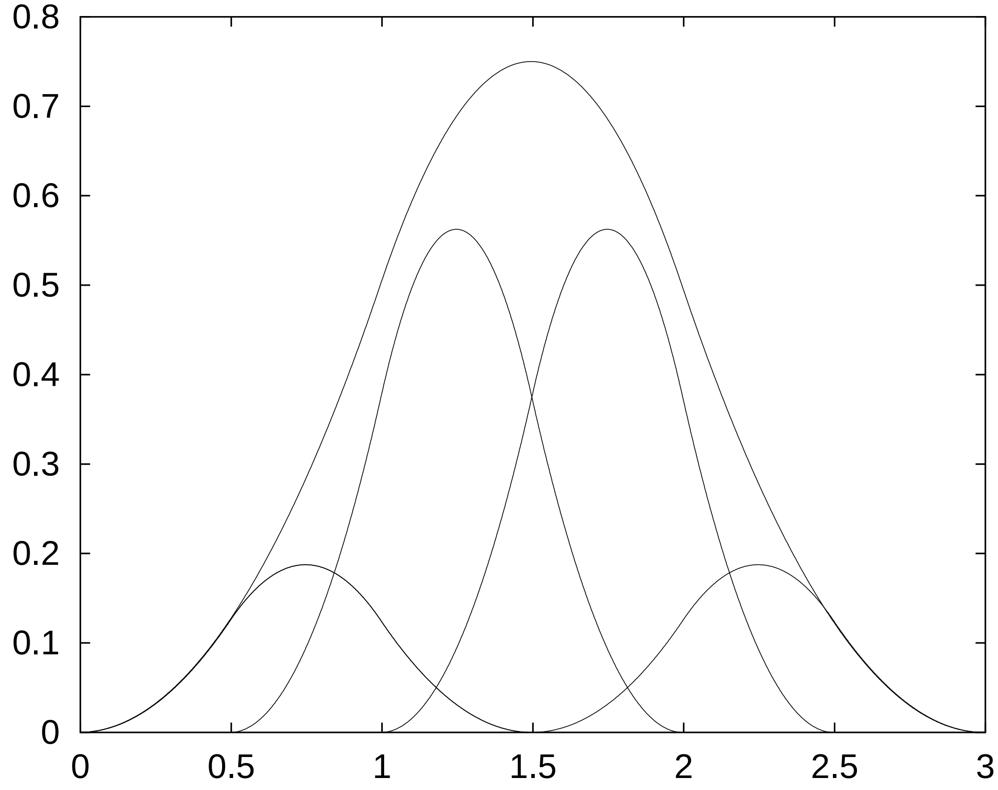
- Expresses that $V_j \subset V_{j+1}$ since by change of variable we obtain $\varphi_{j,k} = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} h_n \varphi_{j+1,2k+n}$.

- Example: $\varphi = \chi_{[0,1]} = \chi_{[0,1/2]} + \chi_{[1/2,1]} = \varphi(2\cdot) + \varphi(2 \cdot -1)$, i.e. $h_0 = h_1 = 1$, $h_n = 0$ otherwise.

- B-splines of order N: $h_n = 2^{-N} \frac{(N+1)!}{n!(N+1-n)!}$ for $n = 0, \cdots, N+1$, and $h_n = 0$ otherwise.

- Support of φ and discrete support of (h_n) have same length.

Example: quadratic B-spline



$$\varphi(x) = \frac{3}{4}(\varphi(2x-1) + \varphi(2x-2)) + \frac{1}{4}(\varphi(2x) + \varphi(2x-3))$$

Orthonormal wavelets

Assuming that φ is such that the $(\varphi_{j,k})_{k\in\mathbb{Z}}$ are an orthonormal basis so that $P_j f = \sum_{k\in\mathbb{Z}}\langle f, \varphi_{j,k}\rangle\varphi_{j,k}$, one builds the wavelet ψ by

$$\psi(t) = \sum_{n\in\mathbb{Z}} g_n \varphi(2t-n)$$

with $g_n = (-1)^n h_{1-n}$. Then $(\psi_{j,k})_{k\in\mathbb{Z}}$ are an orthonormal basis of the orthogonal complement W_j of V_j into V_{j+1} so that

$$Q_j f = (P_{j+1} - P_j) f = \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}.$$

We thus can decompose f in the orthonormal basis of $L^2(\mathbb{R})$

$$\begin{aligned} f &= P_0 f + \textstyle\sum_{j\geq 0} Q_j f \\ &= \textstyle\sum_{k\in\mathbb{Z}}\langle f, \varphi_{0,k}\rangle\varphi_{0,k} + \sum_{j\geq 0}\sum_{k\in\mathbb{Z}}\langle f, \psi_{j,k}\rangle\psi_{j,k}. \end{aligned}$$

Constructing orthonormal scaling functions

Idea: define φ implicitely as a solution of the two scale equation.

Problem: design coefficients h_n such that φ has prescribed properties, e.g orthnormality., compact support, smoothness, high order approximation properties of the corresponding V_j spaces.

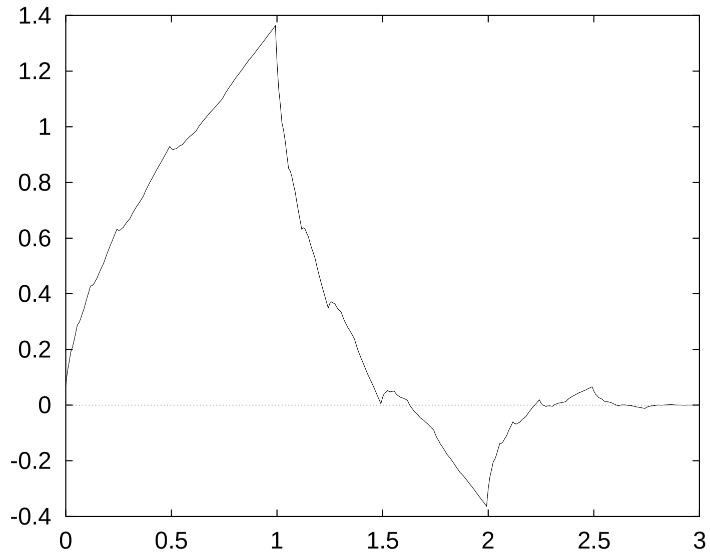
Orthonormality : $\sum_n h_n h_{n+2k} = 2$ if $k = 0$, 0 otherwise.

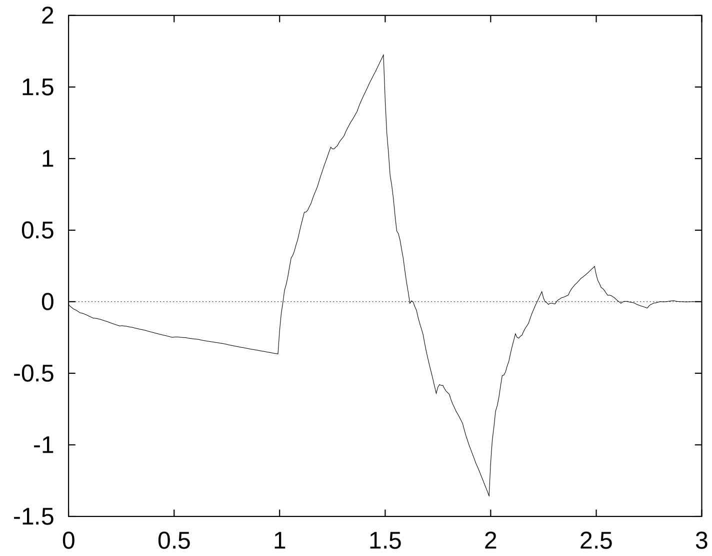
Order N : $\sum_n h_n = 2$ et $\sum_n (-1)^n n^m h_n = 0$, $m = 0, \cdots, N$.

The construction of Daubechies (1988): for each $N > 0$, a sequence (h_n) supported on $\{0, \cdots, 2N-1\}$, such that solution φ_N compactly supported in $[0, 2N-1]$ has orthonormal translates, and resulting V_j spaces have approximation order N. Also $\varphi_N \in C^{s(N)}$ with $s(N) \sim N/5$ as $N \to \infty$. Except $\varphi_1 = \chi_{[0,1]}$, no explicit expression.

Example

Orthonormal scaling function and wavelet in the case $N = 2$ (smoothness $s(2) \approx 0.55$).





Biorthogonal wavelets

Idea: replace orthogonality assumption by a dual scaling function $\tilde{\varphi} = \sum_{n\in\mathbb{Z}} \tilde{h}_n \tilde{\varphi}(2\cdot -n)$ such that $\langle \varphi_{j,k}, \tilde{\varphi}_{j,l}\rangle = 1$ if $k = l$, 0 otherwise.

- Non-orthogonal projector $P_j f = \sum_{k\in\mathbb{Z}} \langle f, \tilde{\varphi}_{j,k}\rangle \varphi_{j,k}$.

- Dual wavelets $\psi = \sum_{n\in\mathbb{Z}} g_n \varphi(2\cdot -n)$ and $\tilde{\psi} = \sum_{n\in\mathbb{Z}} \tilde{g}_n \tilde{\varphi}(2\cdot -n)$, with $g_n = (-1)^n \tilde{h}_{1-n}$ and $\tilde{g}_n = (-1)^n h_{1-n}$.

- Satisfy $\langle \psi_{j,k}, \tilde{\psi}_{j,l}\rangle = 1$ if $k = l$, 0 otherwise and $\langle \varphi_{j,k}, \tilde{\psi}_{j,l}\rangle = \langle \tilde{\varphi}_{j,k}, \psi_{j,l}\rangle = 0$.

- Projector $Q_j f = (P_{j+1} - P_j) f = \sum_{k\in\mathbb{Z}} \langle f, \tilde{\psi}_{j,k}\rangle \psi_{j,k}$ onto non-orthogonal complement $W_j = V_{j+1} \cap \tilde{V}_j^{\perp}$

Results in a decomposition of f in a biorthogonal basis of $L^2(\mathbb{R})$

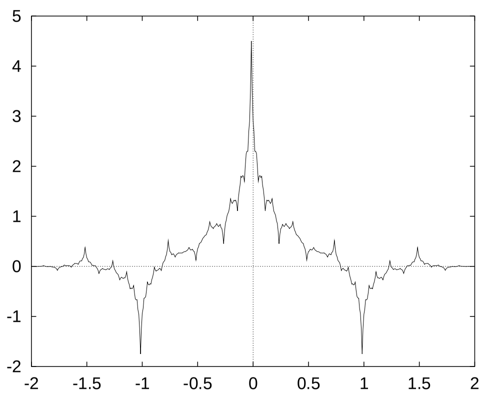
$$\begin{aligned} f &= P_0 f + \sum_{j\geq 0} Q_j f \\ &= \sum_{k\in\mathbb{Z}} \langle f, \tilde{\varphi}_{0,k}\rangle \varphi_{0,k} + \sum_{j\geq 0} \sum_{k\in\mathbb{Z}} \langle f, \tilde{\psi}_{j,k}\rangle \psi_{j,k}. \end{aligned}$$

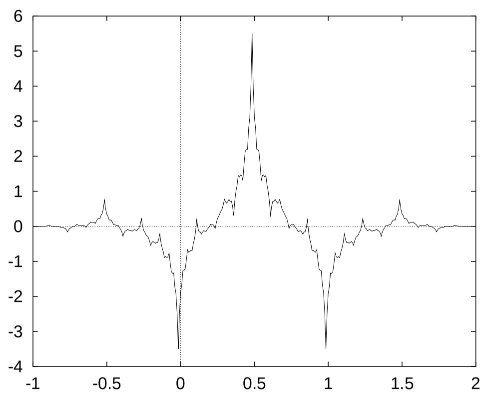
Constructing dual scaling functions

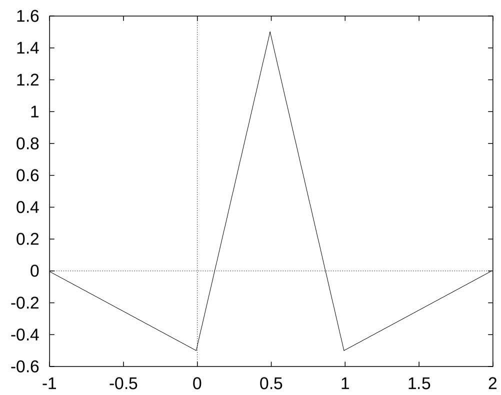
Similar approach as in the orthogonal case: use the refinement equation with proper design of the coefficients h_n and $\tilde{h}_n$.

Duality: $\sum_n \tilde{h}_n h_{n+2k} = 2$ if $k = 0$, 0 otherwise.

More flexible: one can prescribe the h_n and therefore the function φ, and look for coefficients $\tilde{h}_n$ solutions of the resulting linear equations (not unique). In particular, one can obtain dual functions for the B-splines of degree N. Example: $\tilde{\varphi}$, $\tilde{\psi}$ and ψ for linear splines ($N = 1$) with $\tilde{h}_0 = 3/4$, $\tilde{h}_{\pm 1} = 1/4$ and $\tilde{h}_{\pm 2} = -1/8$.







The fast wavelet transform algorithm

Connect standard and multiscale representations

$$f = \sum_{k \in \mathbb{Z}} c_{J,k} \varphi_{J,k} = \sum_{k \in \mathbb{Z}} c_{0,k} \varphi_{0,k} + \sum_{j=0}^{J-1} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k},$$

by the same hierarchical procedure as for the Haar system.

Basic step $c_{j+1} \leftrightarrow (c_j, d_j)$ in the biorthogonal case:

Decomposition: use dual two scale equation

$$\begin{aligned} c_{j,k} &= \langle f, \tilde{\varphi}_{j,k} \rangle = \langle f, \tfrac{1}{\sqrt{2}} \textstyle\sum_{n \in \mathbb{Z}} \tilde{h}_n \tilde{\varphi}_{j+1,2k+n} \rangle \\ &= \tfrac{1}{\sqrt{2}} \textstyle\sum_{n \in \mathbb{Z}} \tilde{h}_n c_{j+1,2k+n} = \tfrac{1}{\sqrt{2}} \textstyle\sum_{n \in \mathbb{Z}} \tilde{h}_{n-2k} c_{j+1,n}. \end{aligned}$$

and similarly for $d_{j,k}$. Therefore

$$c_{j,k} = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} \tilde{h}_{n-2k} c_{j+1,n} \quad \text{and} \quad d_{j,k} = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} \tilde{g}_{n-2k} c_{j+1,n}.$$

Reconstruction: use primal scale equation

$$\begin{aligned} P_{j+1}f &= \textstyle\sum_{k\in\mathbb{Z}} c_{j+1,k}\varphi_{j+1,k} \\ &= \textstyle\sum_{n\in\mathbb{Z}} c_{j,n}\varphi_{j,n} + \sum_{n\in\mathbb{Z}} d_{j,n}\psi_{j,n} \\ &= \textstyle\sum_{n\in\mathbb{Z}} c_{j,n}[\frac{1}{\sqrt{2}}\sum_{k\in\mathbb{Z}} h_{k-2n}\varphi_{j+1,k}] \\ &\quad \textstyle+ \sum_{n\in\mathbb{Z}} d_{j,n}[\frac{1}{\sqrt{2}}\sum_{k\in\mathbb{Z}} g_{k-2n}\varphi_{j+1,k}] \\ &= \textstyle\sum_{k\in\mathbb{Z}} \frac{1}{\sqrt{2}}[\sum_{n\in\mathbb{Z}} c_{j,n}h_{k-2n} + \sum_{n\in\mathbb{Z}} d_{j,n}g_{k-2n}]\varphi_{j+1,k}. \end{aligned}$$

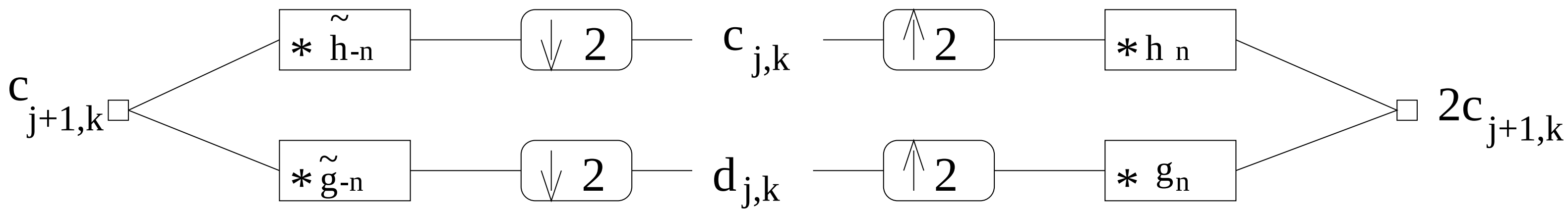
By identification of the coordinates in the first and last expression, we obtain

$$c_{j+1,k} = \frac{1}{\sqrt{2}}\sum_{n\in\mathbb{Z}} c_{j,n}h_{k-2n} + \frac{1}{\sqrt{2}}\sum_{n\in\mathbb{Z}} d_{j,n}g_{k-2n}.$$

Remark: these algorithms only use the coefficients $(h_n, \tilde{h}_n, g_n, \tilde{g}_n)$, not the functions $(\varphi, \psi, \tilde{\varphi}, \tilde{\psi})$.

Signal processing interpretation

Two channel filter bank (low pass and high pass).



Remark: In practice, discretized data have finite support $c_J = (c_{J,k})_{k=0,\cdots,2^J-1} \Rightarrow$ adaptation of the filtering process is needed near the boundary.

Simplest solutions: extension by periodization or symmetrization of the signal, or boundary adapted MRA and wavelets.

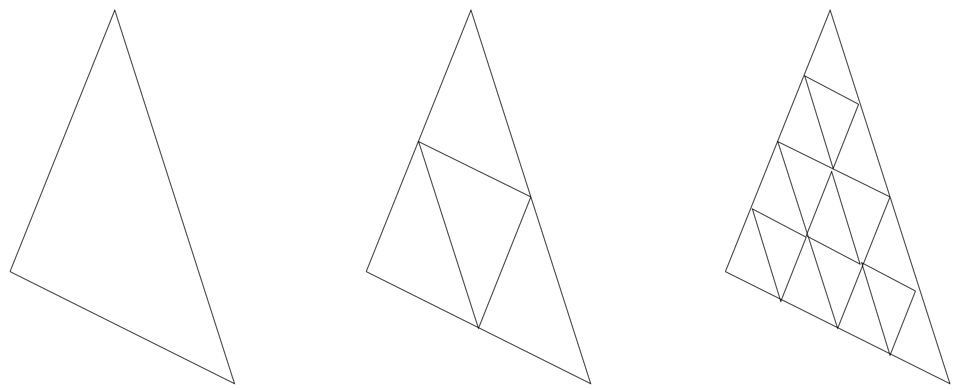
Toward generalized wavelets

Problem: adaptation of MRA and wavelets to general domains $\Omega \in \mathbb{R}^{\mathrm{d}}$, possibly with boundary conditions. Tensor product strategies are not sufficient.

First approach (Canuto, Tabacco, Urban, Dahmen, Schneider): export the tensor product strategy through domain decomposition : Ω partitioned into conforming parametric patches $\Omega_i = \kappa_i([0,1]^d)$. From tensor product wavelets in reference domain, define wavelets inside each patches as $\psi_\lambda^i(\cdot) := \psi_\lambda(\kappa_i^{-1} \cdot)$. At the interfaces, define wavelets by proper "glueing" of functions from both side ensuring at least continuity.

Second approach: hierarchical finite elements

Start from nested triangulations $\mathcal{T}_j$ derived from a coarse triangulation $\mathcal{T}_0$ of Ω by iterative mi-point refinement.



Define associated finite element spaces V_j. Not always nested but nestedness holds e.g. for P_m Lagrange elements (nodal degrees of freedom Γ_j are nested).

Nodal basis $(\varphi_\gamma)_{\gamma\in\Gamma_j}$ and interpolation projector

$$P_j f = \sum_{\gamma\in\Gamma_j} f(\gamma)\varphi_\gamma.$$

Finite element wavelets

Hierarchical basis: $P_{j+1}f - P_j f$ vanishes on $\Gamma_j \Rightarrow$ can be expressed as $\sum_{\lambda\in\nabla_j} d_\lambda\psi_\lambda$, where $\nabla_j = \Gamma_{j+1}\setminus\Gamma_j$ and ψ_λ is the nodal function φ_λ of V_{j+1}. This leads to the hierarchical basis decomposition of $f \in V_J$

$$f = \sum_{\gamma\in\Gamma_0} c_\gamma\varphi_\gamma + \sum_{j=0}^{J-1}\sum_{\lambda\in\nabla_j} d_\lambda\psi_\lambda.$$

Drawback: intrinsic lack of L^2 stability due to the use of the interpolation projector.

Solution: finite element wavelets built by local correction of the ψ_λ, $\lambda \in \nabla_j$, with combinations of the functions $(\varphi_\gamma)_{\gamma\in\Gamma_{j+1}}$ localized around the support of ψ_λ (Oswald, Lorentz, Dahmen, Stevenson).

Linear approximation results

- V_h: finite element space discretizing a domain $\Omega \subset \mathbb{R}^d$.
- $N := \dim(V_h) \sim \mathrm{vol}(\Omega) h^{-d}$
- $W^{s,p} := \{f \in L^p(\Omega) \text{ s.t. } D^\alpha f \in L^p(\Omega),\ |\alpha| \leq s\}$

Classical finite element approximation theory (Bramble-Hilbert, Ciarlet-Raviart, 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_p^s.

Wavelet characterizations of functions spaces

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

- L^2 characterized by $\|f\|_2^2 \sim \|P_0 f\|_2^2 + \sum_{j\geq 0} \|Q_j f\|_2^2 \sim \sum |d_\lambda|^2$.

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

$$\|f\|_{H^s}^2 \sim \|P_0 f\|_2^2 + \sum_{j\geq 0} 2^{2sj} \|Q_j f\|_2^2 \sim \sum 2^{2s|\lambda|} |d_\lambda|^2 \sim \sum \|d_\lambda \psi_\lambda\|_{H^s}^2.$$

- Besov-Sobolev space $B^s_{p,p}$ characterized by

$$\begin{aligned} \|f\|_{B^s_{p,p}}^p &\sim \|P_0 f\|_p^p + \textstyle\sum_{j\geq 0} 2^{psj} \|Q_j f\|_p^p \sim \sum 2^{ps|\lambda|} \|d_\lambda \psi_\lambda\|_p^p \\ &\sim \textstyle\sum 2^{ps|\lambda|} 2^{pd(1/2-1/p)|\lambda|} |d_\lambda|^p \sim \sum \|d_\lambda \psi_\lambda\|_{B^s_{p,p}}^p. \end{aligned}$$

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

All this holds provided that ψ_λ has enough smoothness

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 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 $1/q = 1/p + t/d$

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

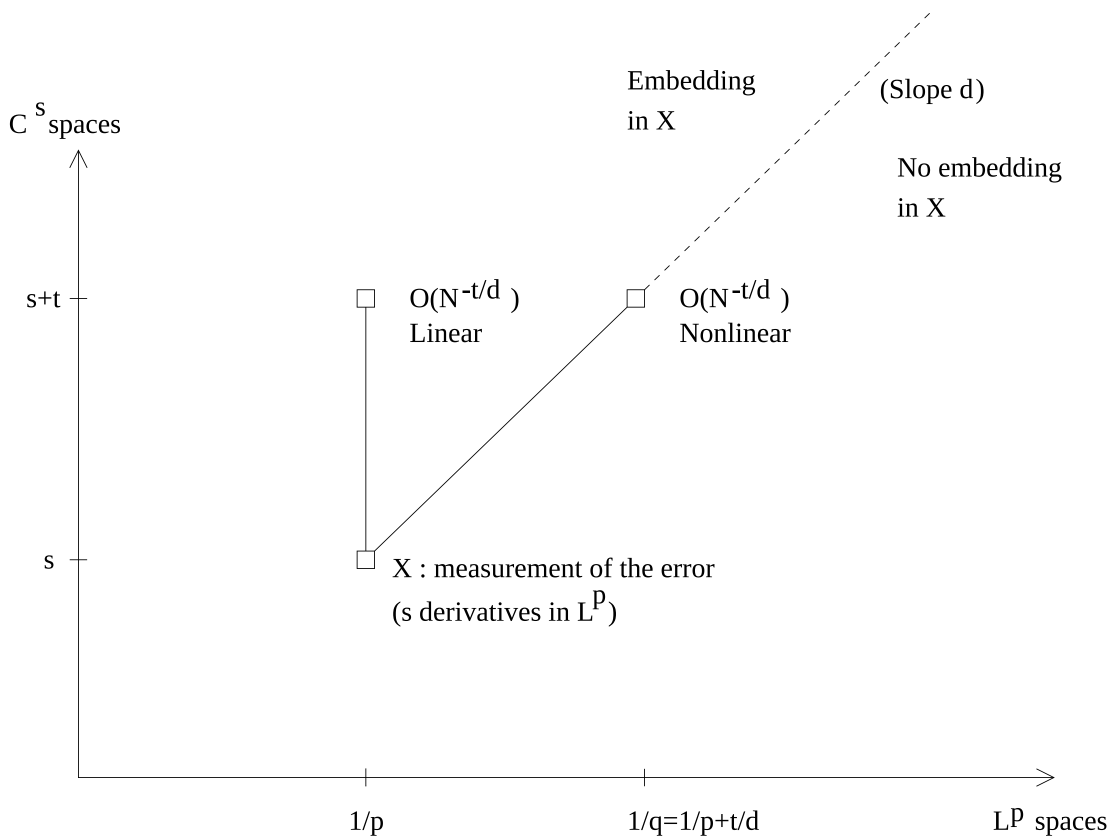
- For most error norm X (e.g. L^p, $W^{s,p}$, $B^s_{p,q}$), a near optimal approximation is obtained by thresholding : if $f = \sum_\lambda d_\lambda\psi_\lambda$, and $f_N := \sum_{N \text{ largest } \|d_\lambda\psi_\lambda\|_X} d_\lambda\psi_\lambda$, we then have

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

with C independent of f and N.

- Remark: a similar theory for piecewise polynomial approximation on N adaptive triangles is still to be completed.

Pictorial interpretation of approximation results



Modeling real images by functions of bounded variation

$I \in BV$ if and only if $I \in L^1$ and ∇I is a finite measure

Prototype: χ_Ω where $\partial\Omega$ has finite length.

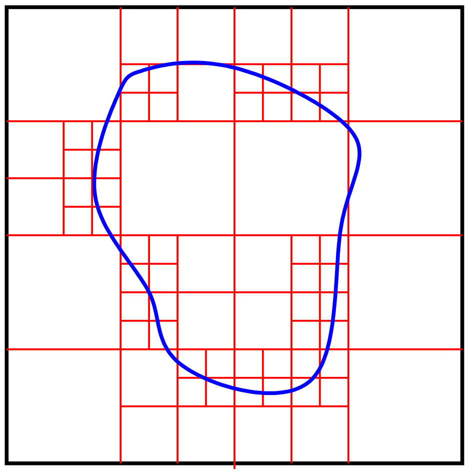
Intuition: Images are “piecewise smooth” and their singularities (edges) have finite total length.

Theorem (1998): $f \in BV([0,1]^2) \Rightarrow (d_\lambda) \in w\ell^1$ i.e. $d_n \leq C/n$.

- BV is “almost characterized” since $(d_\lambda) \in \ell^1 \Rightarrow f \in BV([0,1]^2)$.
- Optimal estimate for wavelets: if $f = \chi_\Omega$ then $d_n \geq c/n$.
- Optimal estimate among all bases

Sparse representations and geometry

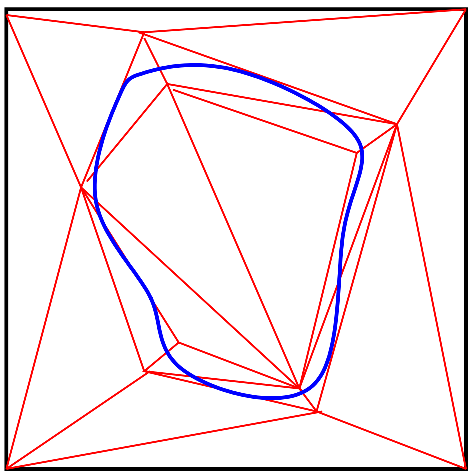
Image: $I = \chi_\Omega$, with $\partial\Omega$ smooth.



I_N = approximation by N largest wavelet coefficients

$\Rightarrow \|I - I_N\|_{L^2} \sim N^{-1/2}$

Problem: imposes isotropic refinement



I_N = piecewise constant approximation on N optimally selected triangles

$\Rightarrow \|I - I_N\|_{L^2} \sim N^{-1}$

Problem: fast hierarchical algorithm ?

Other recent approaches for sparse representation of geometry

- Donoho and Candes: sparse representation based on ridglets/curvelets bases (similar to wavelets with additional directional selectivity). Allows to recover $\|I - I_N\|_{L^2} \sim N^{-1}$ with a thresholding algorithm.

- Mallat: sparse representation based on bandlets (selection of a basis adapted to the edges of the image).

- Arandiga, Donat, A.C.: sparse representation based on nonlinear multiscale decompositions (uses shock capturing techniques introduced by Harten and Osher).

Revisiting 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.

Example: 1D nonlinear conservation law

$$\partial_t u + \partial_x F(u) = 0, \quad u(x,0) = u_0(x),$$

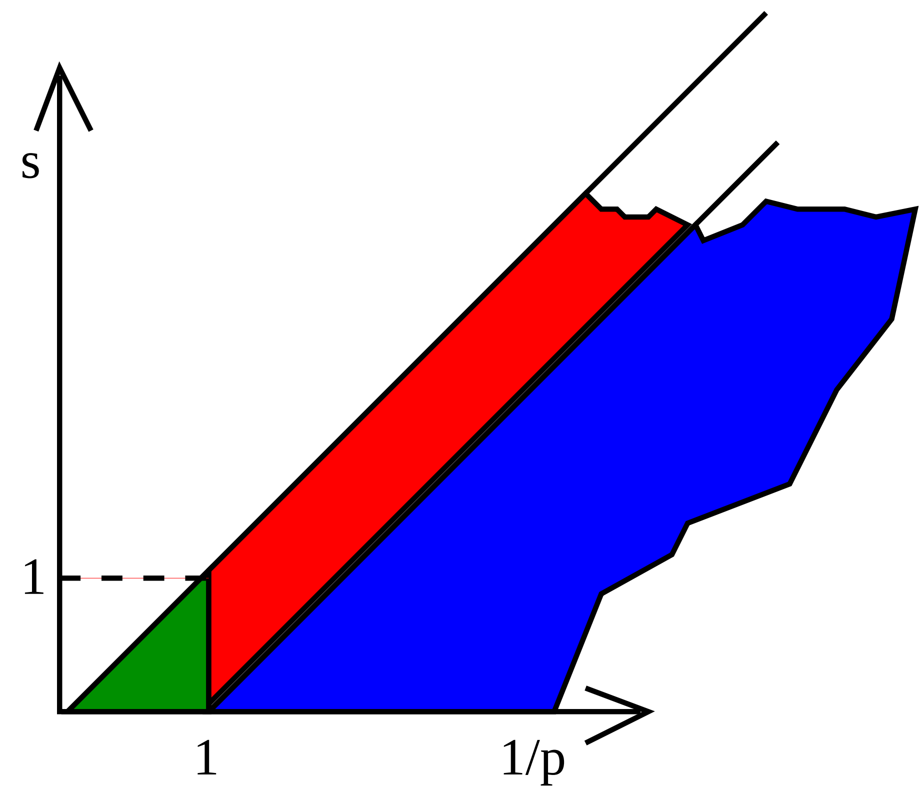
with F smooth and strictly convex (e.g. Burger $F(u) = u^2/2$).

- Smoothness for linear approximation in L^1: for large t, $u(\cdot, t) \in BV$ but not smoother.

- Smoothness for nonlinear approximation (DeVore & Lucier, 1987): for all $s > 0$ and $1/p = 1 + s$, if $u_0 \in B^s_{p,p}$ then $u(\cdot, t) \in B^s_{p,p}$ for all $t > 0$.

Similar results are available for elliptic PDE's on corner domains (DeVore & Dahlke)

Pictorial interpretation



Classical theory : $s < 1/p$ for $s \leq 1$

DeVore-Lucier : $s < 1/p - 1$ for all $s > 0$

Interpolation : $s < 1/p$ for all $s > 0$

Use of adaptive wavelet methods for discretizing PDE's ?

Benchmark for adaptive methods : an optimal adaptive strategy should produce approximate solutions $u_N \in \Sigma_N$ such that $\|u - u_N\|$ behaves asymptotically as good as $\inf_{v \in \Sigma_N} \|u - v\|$, for some norm $\|\cdot\|$ of interest, without the full knowledge of the largest wavelet coefficients of u, and with $\mathcal{O}(N)$ computational cost.

First approach: space refinement techniques to access the appropriate discretization $\{\psi_\lambda\}_{\lambda \in \Lambda_n}$ (Bertoluzza, Perrier, Liandrat, Dahlke, Canuto, Stevenson, Urban, Masson, Dahmen, DeVore, AC). In the case of stationary problems

$$\mathcal{F}(u) = 0,$$

enjoying a suitable variational formulation, this approach has led to optimal strategies in the above sense.

This approach might be difficult to operate for certain types of problems for which certain discretization are doomed to fail.

Second approach: Multiresolution Adaptive Post-processing, i.e. start from a classical and reliable scheme on a uniform grid and use a discrete multiresolution decomposition in order to compress computational time and memory size, while preserving the accuracy of the initial scheme (Harten, Abgrall, Arandiga, Chiavassa, Donat, Dahmen, Mueller, Gottschlich-Mueller, Kaber, Postel, AC). Typically applied to evolution problems

$$\partial_t u = \mathcal{E}(u),$$

with good practical results, but uncomplete convergence analysis.

General variational problems

$\mathcal{H}$ Hilbert space, $\mathcal{F} : \mathcal{H} \to \mathcal{H}'$ continuous mapping, u nonsingular solution of $\mathcal{F}(u) = 0$, i.e. $D\mathcal{F}(u)$ is an isomorphism from $\mathcal{H}$ to $\mathcal{H}'$.

Variational formulation : find $u \in \mathcal{H}$ such that

$$\langle \mathcal{F}(u), v \rangle = 0$$

for all $v \in \mathcal{H}$.

Simple linear examples: $\mathcal{F}(u) = \mathcal{A}u - f$

- Laplace: $\mathcal{A} := -\Delta$ and $\mathcal{H} := H_0^1$
- Stokes: $\mathcal{A}(u, p) := (-\Delta u + \nabla p, -\mathrm{Div}\, u)$ and $\mathcal{H} := (H_0^1)^3 \times L_0^2$.
- Single layer potential $\mathcal{A}u(x) := \int_\Gamma \frac{u(y)}{4\pi|x-y|} dy$ and $\mathcal{H} := H^{-1/2}$.

Standard (FEM) approach to discretisation

1. Well posed problem in infinite dimension $\mathcal{F}(u) = 0$.

2. Finite dimensional discretization $\mathcal{H} \to V_h$ by a Petrov-Galerkin type method ($\langle \mathcal{F}(u_h), v_h \rangle = 0$ for all $v_h \in W_h$).

Difficulties: not always well-posed (compatibility conditions, e.g. LBB for Stokes : $\inf_{p_h \in P_h} \sup_{u_h \in U_h} \frac{\int p_h \mathrm{Div}\, u_h}{\|p_h\|_{L^2} \|u_h\|_{H^1}} \geq \beta_h > 0$).

3. Iterative solver $u_h^0 \to u_h^1 \cdots \to u_h$.

Difficulties: ill-conditionning and dense matrices

4. Adaptivity: derive local error indicators by a-posteriori analysis of residual $\mathcal{F}(u_h)$, and apply local mesh refinement based on these indicators $V_h = V_r^0 \to V_r^1 \to \cdots$, $u_h = u_r^0 \to u_r^1 \to \cdots$

Difficulties: hanging nodes, convergence analysis of such refinement strategies (Dörfler 1996, Morin-Nocetto-Siebert 2000).

Wavelet adaptive discretizations: new paradigm

1. Well posed problem in infinite dimension $\mathcal{F}(u) = 0$.

2. Equivalent discrete problem in infinite dimension by wavelet-Galerkin: find $U = (u_\lambda)_{\lambda\in\nabla}$ such that

$$F(U) := (\langle \mathcal{F}(\sum u_\lambda \psi_\lambda), \psi_\mu\rangle)_{\mu\in\nabla} = 0.$$

Well-posed : $F : \ell^2 \to \ell^2$ if $(\psi_\lambda)_{\lambda\in\nabla}$ is a Riesz basis for $\mathcal{H}$ after renormalization, i.e. $\|u\|_{\mathcal{H}}^2 \sim \sum |u_\lambda|^2$ and $\|u\|_{\mathcal{H}'}^2 \sim \sum |\langle u, \psi_\lambda\rangle|^2$.

3. Converging iteration in infinite dimension $U^0 \to U^1 \to \cdots \to U$.

4. Adaptive approximation of this iteration up to prescribed tolerances in finite dimension: U^n supported by finite wavelet set $\Lambda_n \subset \nabla$.

$\Rightarrow$ allows to establish optimal accuracy and complexity results in the energy $\|u\|_{\mathcal{H}} \sim \|U\|$ norm.

The linear elliptic case

Assume $\mathcal{A}$ is an $\mathcal{H}$-elliptic operator. Equivalent problem :

$$AU = F$$

where A is ℓ^2-elliptic. For a suitable κ the iteration,

$$U^{n+1} = U^n + \kappa[F - AU^n],$$

converges with fixed error reduction rate $\rho < 1$.

Approximate iteration with prescribed tolerance $\varepsilon > 0$,

$$U^{n+1} = U^n + \kappa[\mathrm{APPROX}(F, \varepsilon) - \mathrm{APPROX}(AU^n, \varepsilon)],$$

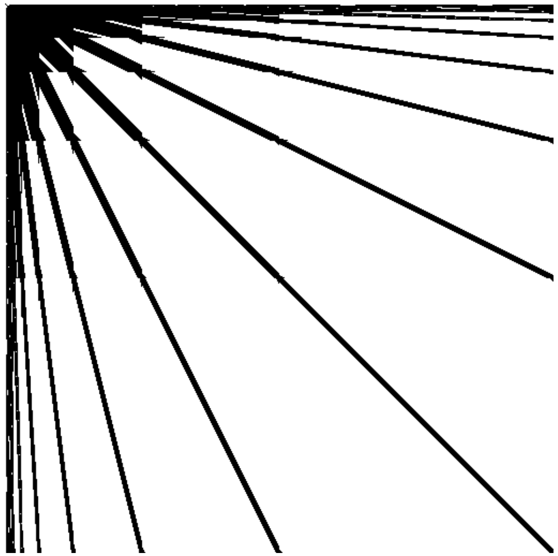
with $\|\mathrm{APPROX}(AU^n, \varepsilon) - AU\| \leq \varepsilon$ and $\|\mathrm{APPROX}(F, \varepsilon) - F\| \leq \varepsilon$. converges with reduction rate ρ until error is of order ε.

The procedure $\mathrm{APPROX}(F, \varepsilon)$ amounts in thresolding F in ℓ^2, or equivalently the data f in the $\mathcal{H}'$ norm.

Matrix-vector approximation

The procedure APPROX(AU^n, ε) is made possible by matrix compression: one can build A_N with N coefficients per rows and colums such that $\|A - A_N\| \lesssim N^{-r}$

(V_1,V_1)	(W_1,V_1)	(W_2,V_1)	(W_3,V_1)
(V_1,W_1)	(W_1,W_1)	(W_2,W_1)	(W_3,W_1)
(V_1,W_2)	(W_1,W_2)	(W_2,W_2)	(W_3,W_2)
(V_1,W_1)	(W_1,W_3)	(W_2,W_3)	(W_3,W_3)



Analysis : based on the Schur lemma, using esimates of the type

$$|\langle \mathcal{A}\psi_\lambda, \psi_\mu\rangle| \lesssim [1 + \mathrm{dist}(\lambda,\mu)]^{-\beta} 2^{-\gamma||\lambda|-|\mu||},$$

derived from the smoothness and vanishing moments of the ψ_λ.

The role of thresholding

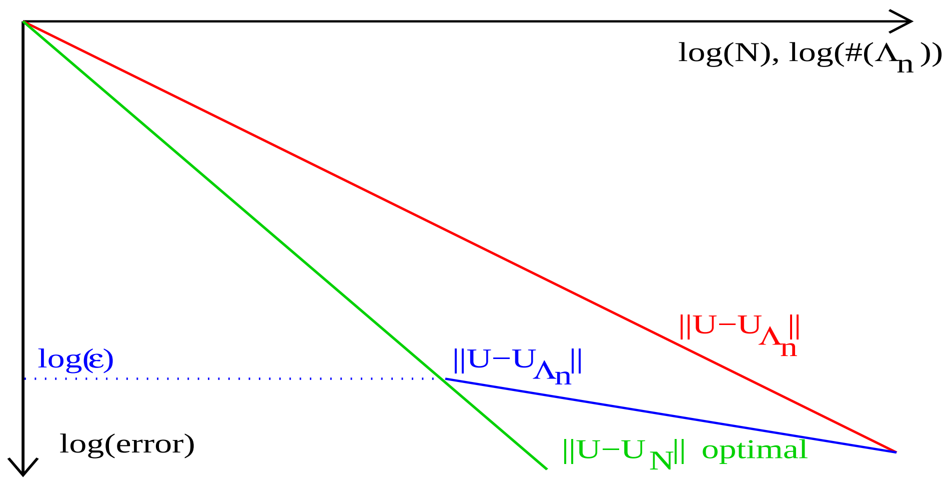
Lemma : if U is such that $\|U - U_N\| \le CN^{-s}$ and V is such that $\|V - U\| \le \varepsilon$. For $a > 1$ define W the smallest subvector of V such that $\|V - W\| \le a\varepsilon$. Then, we have $\|U - W\| \le (1+a)\varepsilon$ and

$$|\mathrm{Supp}(W)| \le C\varepsilon^{-1/s}, \text{ i.e. } \|U - W\| \le C|\mathrm{Supp}(W)|^{-s}$$

Thresholding

ensures

optimality



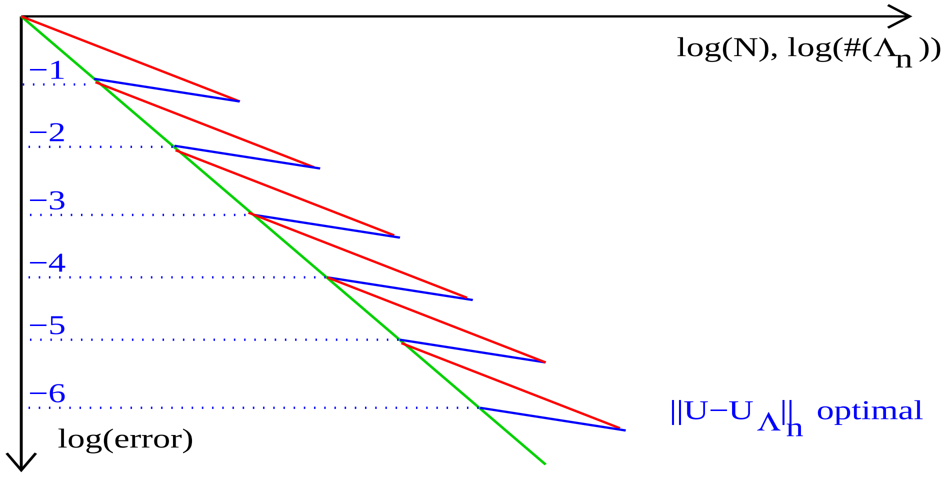
Problem : intermediate memory size and computational time should also be optimal, i.e. $\mathcal{O}(\varepsilon^{-1/s})$.

Geometric tolerances

Idea: decrease

tolerances $\varepsilon_0 = 1$,

$\varepsilon_1 = \frac{1}{2}, \cdots, \varepsilon_j = 2^{-j}$



Fixed number of iteration at each step $j \to j+1$ involving sparse matrix-vector product: $W = \text{APPROX}(AV, \varepsilon)$ obtained by decomposing $V = V_1 + [V_2 - V_1] + [V_4 - V_2] + \cdots$, and taking

$$W := A_{2^J} V_1 + A_{2^{J-1}}[V_2 - V_1] + \cdots + A_1[V_{2^J} - V_{2^{J-1}}]$$

with J large enough such that

$$\|W - AV\| \le \|A\| \|V - V_{2^J}\| + \sum_{j=1}^{J} \|A - A_{2^{J-j}}\| \|V_{2^j} - V_{2^{j-1}}\| \le \varepsilon.$$

Results

Theorem (Dahmen, DeVore, AC - Math. Comp. 2000) : if V is such that $\|V - V_N\| \leq CN^{-s}$, and if $\|A - A_N\| \lesssim N^{-r}$ with $r > s$, then $|\text{Supp}(W)| \lesssim \varepsilon^{-1/s}$ and therefore $\|W - AU\| \lesssim |\text{Supp}(W)|^{-s}$.

Theorem (Dahmen, DeVore, AC - FoCM 2002) : The general strategy for linear operator equations based on the above ingredients (thresholding, adaptive matrix vector multiplication) achieves the ultimate goal, namely production of U^n and $\Lambda_n = \text{Supp}(U^n)$, such that if $\|U - U_N\| \leq CN^{-s}$, then

$$\|U - U^n\| \lesssim \#(\Lambda_n)^{-s},$$

with $\mathcal{O}(\#(\Lambda_n))$ computational cost.

Remarks on practical aspects

All wavelet properties are exploited : Sobolev norm equivalences, smoothness (not always available) and vanishing moments.

Coarsening is not needed in all practical cases studied so far, yet seems necessary in the proof of the optimality theorem ! Similar optimality results recently obtained for adaptive FEM by Binev, Dahmen and DeVore, using the Morin-Nocetto-Siebert algorithm combined with coarsening.

Complexity is dominated by assembling matrix elements, numerical quadratures, addressing the indices in Λ_n (key role of efficient data structures). Practical comparison between adaptive FEM and wavelets based on the same FE spaces : for a given error, wavelets may win for $N_{\text{d.o.f.}}$ but lose (by a factor > 4) for computational cost.

Extension to more general problems

Saddle point problems $AU + B^T P = F$ and $BU = G$, e.g. based on adaptive approximation of the Uzawa iteration (Dahlke, Hochmuth and Urban 1999) :

$$AU^n = F - B^T P^{n-1} \text{ and } P^n = P^{n-1} + \kappa(BU^n - G)$$

No LBB is needed here, adaptivity stabilizes Similar result for adaptive FEM algorithm : Nocetto 2002. Question : do the same concepts apply to convection dominated problems, such as $-\varepsilon\Delta u + a.\nabla u = 0$ with convergence rate independent of ε ?

Extension to nonlinear problems : DeVore, Dahmen, A.C. 2002 (need specific adaptation of fast evaluation of $F(U)$), no available numerical results yet.

Problem dependent tuning seems unavoidable in order to optimize this type of algorithms.

General evolution problems

We are interested in initial value problems

$$\partial_t u = \mathcal{E}(u),$$

which develop singularities in finite time, e.g. hyperbolic systems of conservation laws

$$\partial_t u + \mathrm{Div}_x F(u) = 0, \quad u(x,0) = u_0(x).$$

1. Theoretical difficulties: weak solutions, entropy conditions, ...
2. Numerical difficulties: only few schemes are proved to converge and their convergence rate is limited due to the presence of singularities

Solution to the last difficulty: adaptativity by local mesh refinement ?

Drawback: implementation (singularities are moving $\Rightarrow$ local refinement evolves with time) and convergence analysis.

Multiresolution can help !

Some important contributions (80-90):

- Automatic Mesh Refinement (Berger & Oliger): use hierarchical meshes and locally select the scale of resolution by ad-hoc criterions or error indicators.

- Multiresolution adaptive flux computations (Harten & Abgrall): use discrete multiresolution decomposition to accelerate the numerical flux computations, yet evolution takes place on the uniform finest mesh.

Since 90: attempts to use wavelet discretizations for the multiresolution adaptive solution of PDE's, motivated by their ability for data compression.

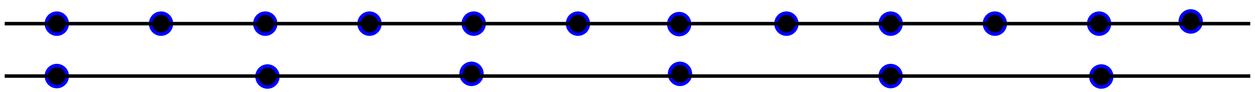
A discrete multiresolution framework

- Γ_j, $j = 0, \cdots, J$: sequence of discretisations at scales 2^{-j}.

- $U_j = (U_j(\gamma))_{\gamma \in \Gamma_j}$ discretisation of a fonction u on Γ_j, i.e. vector of $\mathcal{V}_j := \mathbb{R}^{\Gamma_j}$.

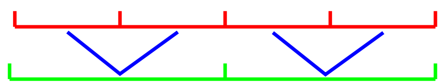
- Restriction operator P_{j-1}^{j} from $\mathcal{V}_j$ onto $\mathcal{V}_{j-1}$: computes coarser discretization $U_{j-1} = P_{j-1}^{j} U_j$ from the next finer.

Basic example 1: point values on nested grids $\Gamma_{j-1} \subset \Gamma_j$, i.e. $U_{j-1}(\gamma) = U_j(\gamma)$ for $\gamma \in \Gamma_{j-1}$.



Basic example 2: cell averages on nested partitions, i.e.

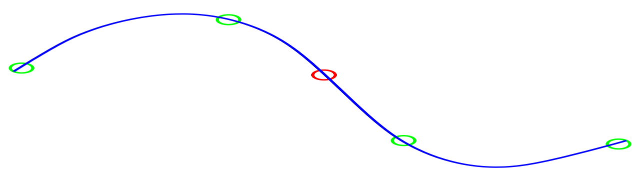
$$U_{j-1}(\gamma) = \mathrm{vol}(\gamma)^{-1} \sum_{\mu \in \Gamma_j, \mu \subset \gamma} \mathrm{vol}(\mu) U_j(\mu)$$



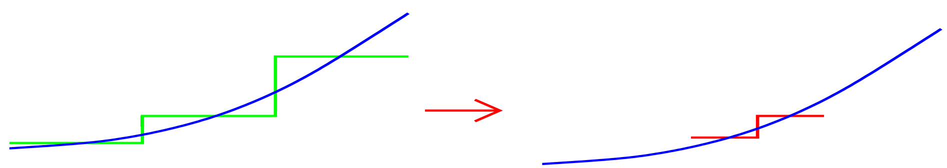
- Prediction operator P_j^{j-1} from $\mathcal{V}_{j-1}$ into $\mathcal{V}_j$: reconstructs an approximation $\hat{U}_j = P_j^{j-1} U_{j-1}$ of U_j.

- Consistancy assumption: $P_{j-1}^{j} P_j^{j-1} = I$

Point value example: $\hat{U}_j(\gamma) = U_{j-1}(\gamma)$ for $\gamma \in \Gamma_{j-1}$, and $\hat{U}_j(\gamma)$ obtained par local interpolation for $\gamma \in \Gamma_j \setminus \Gamma_{j-1}$.



Cell avergage example: $\hat{U}_j(\gamma)$ obtained by “interpolating” the averages in a consistant way, e.g. via polynomial reconstruction.

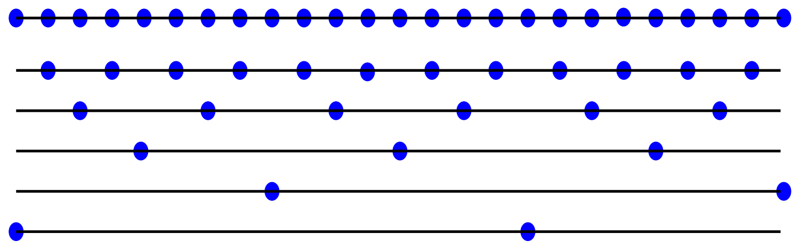


Multiscale decomposition

Prediction error $E_j := U_j - \hat{U}_j \in \mathcal{W}_{j-1} = \mathrm{Ker}(P^j_{j-1})$. Detail vector D_{j-1}: coordinates of E_j in a basis of $\mathcal{W}_{j-1}$.

Point value example: $D_{j-1}(\lambda) = E_j(\lambda)$, $\gamma \in \Gamma_j \setminus \Gamma_{j-1}$ interpolation error at intermediate point. Cell average example: on each coarse cell of Γ_{j-1} the prediction error E_j has null average $\Rightarrow$ define D_j by removing for each coarse cell γ one fine cell $\mu \subset \gamma$.

$$
\begin{aligned}
U_J \quad &\Leftrightarrow (U_{J-1}, D_{J-1}) \Leftrightarrow (U_{J-2}, D_{J-2}, D_{J-1}) \Leftrightarrow \cdots \\
&\Leftrightarrow (U_0, D_0, \cdots, D_{J-1}) = \mathcal{M} U_J = (d_\lambda)_{\lambda \in \nabla_J}
\end{aligned}
$$



Physical grid Γ_J

Multiscale grid (point values) ∇_J

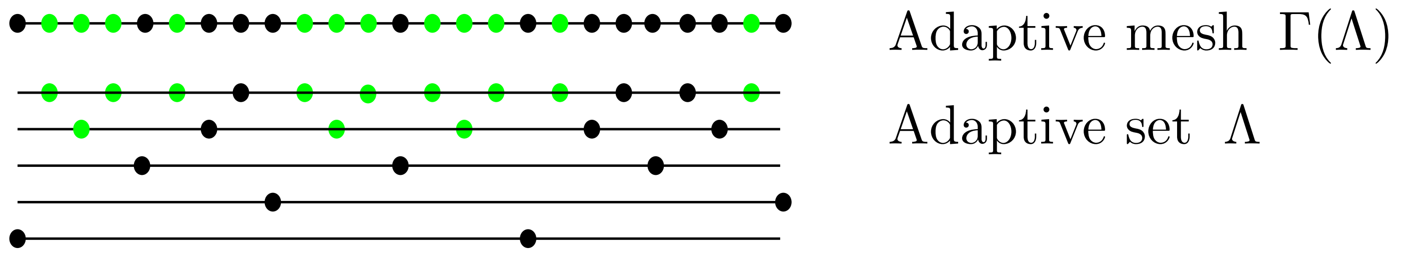
Complexity of $\mathcal{M}$ and $\mathcal{M}^{-1}$: $\mathcal{O}(\mathrm{Card}(\Gamma_J))$

Compression

Thresholding: given a level dependent threshold $\eta = (\eta_0, \cdots, \eta_{J-1})$ set to zero all coefficients $|d_\lambda| \le \eta_{|\lambda|} \Leftrightarrow$ approximation of U_J by

$$\mathcal{T}_\eta U_J = \mathcal{T}_\Lambda U_J = \mathcal{M}^{-1} \mathcal{R}_\Lambda \mathcal{M} U_J,$$

$\mathcal{R}_\Lambda$: restriction of ∇_J to $\Lambda = \Lambda(\eta) = \{\lambda \in \nabla_J \ \text{ t.q. } \ |d_\lambda| \ge \eta_{|\lambda|}\}$.



Compressed representation $(d_\lambda)_{\lambda \in \Lambda}$ to the data of point values or cell averages $U_J(\gamma)$ on an adaptive physical mesh $\Gamma(\Lambda)$ associated to Λ, (need to impose that Λ has a tree structure up to enlarging it).

Complexity of adaptive decomposition and reconstruction algorithms: $\mathcal{O}(\mathrm{Card}(\Lambda))$

Prescriptions

- Prediction operator should have high order accuray: details are very small in the regions where f is smooth ($|d_\lambda| \leq 2^{-s|\lambda|}$ if $f \in C^s$).

- Multiscale reconstruction should be stable: $\|U_J - \mathcal{T}_\Lambda U_J\|$ should be under control for some prescribed norms $\|\cdot\|$. Amounts in analyzing asymptotic behaviour or $P_j^{j-1} P_{j-1}^{j-2} \cdots P_1^0$ as $j \to \infty$ in terms of underlying continuous wavelet systems (ψ_λ)

$$\|U_J - \mathcal{T}_\Lambda U_J\| \leq \sum_{|\lambda| \leq J, \lambda \notin \Lambda} \|d_\lambda \psi_\lambda\|$$

Here ψ_λ is normalized in L^∞. For the control of the L^1 error, the level dependent threshold should be $\eta_j := 2^{dj}\eta_0$ in d dimensions. For the L^∞ norm $\eta_j := \eta_0$. For the BV norm $\eta_j := 2^{(d-1)j}\eta_0$.

Adaptive multiresolution processing

Reference scheme on Γ_J: approximation of $u(x, n\Delta t)$ by $U_J^n = (U_J^n(\gamma))_{\gamma \in \Gamma_J}$ with $U_J^{n+1} = E_J U_J^n$

$$U_J^{n+1}(\gamma) = U_J^n(\gamma) + F(U_J^n(\mu)\ ;\ \mu \in S(\gamma)).$$

$S(\gamma)$: local stencil (excludes implicit schemes).

In the case of FV conservative schemes, F has the form of a balance over the edges surrounding the cell γ

$$U_J^{n+1}(\gamma) = U_J^n(\gamma) + \sum_{\mu \text{ s.t. } |\Gamma_{\gamma,\mu}| \neq 0} F_{\gamma,\mu}^n$$

where $F_{\gamma,\mu}^n = -F_{\mu,\gamma}^n$ is a function of the $U_J^n(\nu)$ for ν in a local stencil surrounding γ and μ.

Adaptive algorithm

Goal: compute approximations of $u(x, n\Delta t)$ by (V_J^n, Λ_η^n), where $V_J^n = (V_J^n(\gamma))_{\gamma\in\Gamma_J}$ is represented by its coefficients $(d_\lambda^n)_{\lambda\in\Lambda_\eta^n}$ or its physical values (point values or cell averages) on the adaptive mesh $(V_J^n(\gamma))_{\gamma\in\Gamma(\Lambda_\eta^n)}$ (we always impose the graded tree structure on Λ_η^n).

Benchmark: an ideal choice would be Λ_η^n the smallest graded tree containing $\{\lambda,\ \ |d_\lambda(U_J^n)| \geq \eta_{|\lambda|}\}$ but it is not be accessible. The adaptive solution V_J^n should still be comparable to $\mathcal{T}_\eta U_J^n$, i.e. the corresponding thresolding operator applied to the exact reference solution.

Initialization: define Λ_η^0 the smallest graded tree containing $\{\lambda,\ \ |d_\lambda(U_J^0)| \geq \eta_{|\lambda|}\}$ and set $V_J^0 := \mathcal{T}_\eta U_J^0$,

Derivation of $(V_J^{n+1}, \Lambda_{n+1})$ from (V_J^n, Λ_n)

Three basic steps:

- Refinement: predict a superset $\Lambda_\eta^n \subset \tilde{\Lambda}_\eta^{n+1}$ adapted to describe the solution at time $n+1$ (ideally such that $|d_\lambda(E_J V_J^n)| < \eta_{|\lambda|}$ if $\lambda \notin \tilde{\Lambda}_\eta^{n+1}$) and extend by $d_\lambda^n = 0$ for $\lambda \in \tilde{\Lambda}_\eta^{n+1} \setminus \Lambda_\eta^n$.

- Evolution: compute the new value $V_J^{n+1}(\gamma)$, for $\gamma \in \Gamma(\tilde{\Lambda}_\eta^{n+1})$ (ideally $V_J^{n+1} = \mathcal{T}_{\tilde{\Lambda}_\eta^{n+1}} E_J V_J^n$).

- Coarsening: apply level dependent thresholding operator to the computed vector $(d_\lambda^n)_{\lambda \in \tilde{\Lambda}_\eta^{n+1}} \Rightarrow$ new set $\Lambda_\eta^{n+1} \subset \tilde{\Lambda}_\eta^{n+1}$ and V_J^{n+1}.

Remark 1: loss of accuracy with respect to the reference scheme is monitored by the threshold η (if $\eta = 0$ the adaptive scheme coincides with the reference scheme). This threshold should be chosen in order to remain within the same accuracy while reducing the CPU time and memory space.

Remark 2: the discretization of the reference scheme and of the multiresolution approximation are decoupled:

- One can apply point value multiresolution on a finite volume scheme based on cell-averages discretizations (however conservativity is violated).

- One can apply multiresolution with high order prediction on a low order scheme. The resulting adaptive scheme should inherit the high order accuracy of the compression process (up to the accuracy of the reference scheme).

Remark 3: loss of accuracy with respect to the reference scheme depends on each of the three steps.

- Coarsening: accuracy is controled by the level of the threshold η and the stability properties of the multiscale reconstruction (existence of underlying continuous wavelet systems).

- Refinement: accuracy is controled by analyzing the action of the discrete evolution operator E_J on the size of the coefficients in the multiscale decomposition.

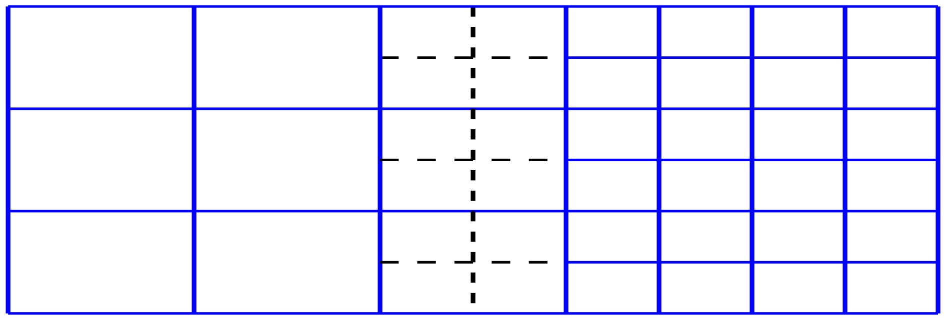
- Evolution: need an accurate evolution step in the compressed form. Two possible approaches:

(i) direct application of the numerical scheme on the adaptive grid $\Gamma(\tilde{\Lambda}_\eta^{n+1})$ (usual AMR approach)

(ii) exact computation of $\mathcal{T}_{\tilde{\Lambda}_\eta^{n+1}} E_J V_J^n$ by local adaptive reconstruction.

Evolution by direct application

Locally the adaptive discretization is uniform, or can be made locally uniform by using the prediction operator.

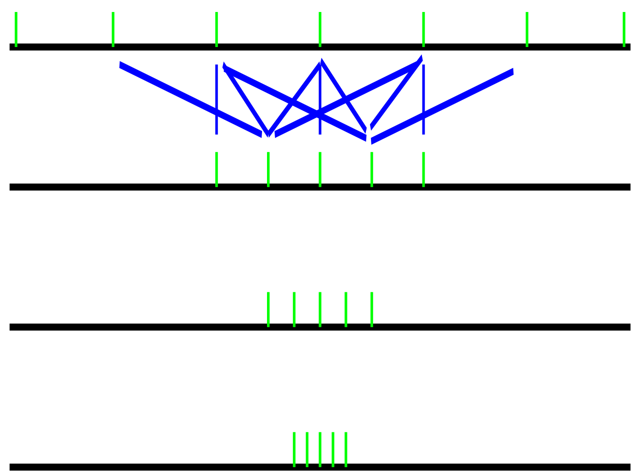


This allows to compute $V_J^{n+1}(\gamma)$, for $\gamma \in \Gamma(\tilde{\Lambda}_\eta^{n+1})$, by direct application of the numerical scheme E_j at the local scale 2^{-j}, therefore with complexity $\mathcal{O}(\#(\tilde{\Lambda}_\eta^{n+1}))$.

This approach gives good results if the reference scheme has high order accuracy comparable to the prediction operator. For a low order reference scheme, it gives significant loss of accuracy in the regions where $\Gamma(\tilde{\Lambda}_\eta^{n+1})$ is coarse.

Evolution by local reconstruction

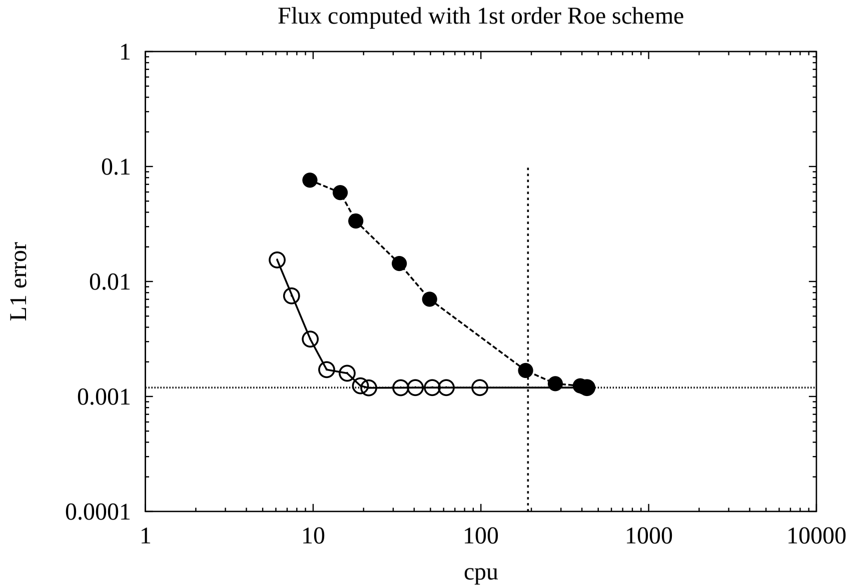
Only reconstruct $V_J^n(\mu)$ for those $\mu \in \Gamma_J$ which are needed to compute exactly $E_J V_J^n$ on the adaptive grid $\Gamma(\tilde{\Lambda}_\eta^{n+1})$. This gives more accurate results, yet with additional computational time.

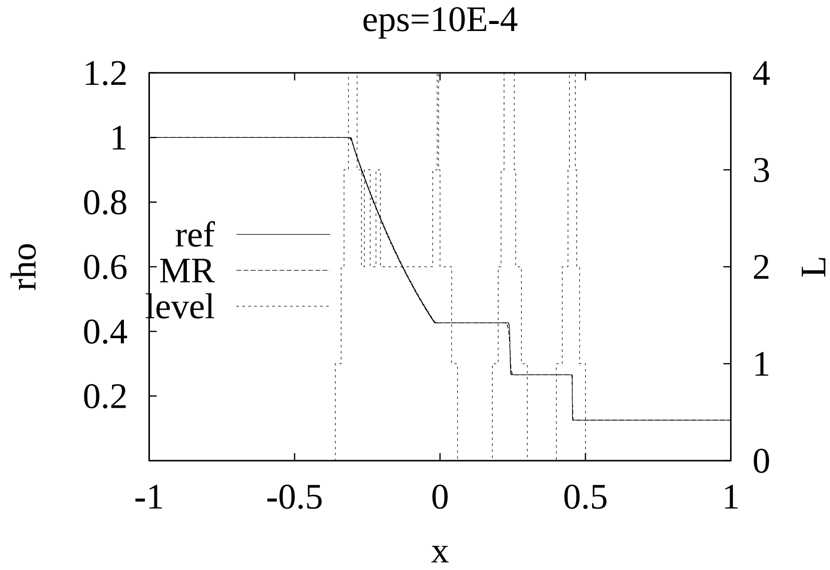


For cell averages in dimension 1 and for point values in any dimension d, the complexity remains in $\mathcal{O}(\#(\tilde{\Lambda}_\eta^{n+1}))$.

Numerical illustration

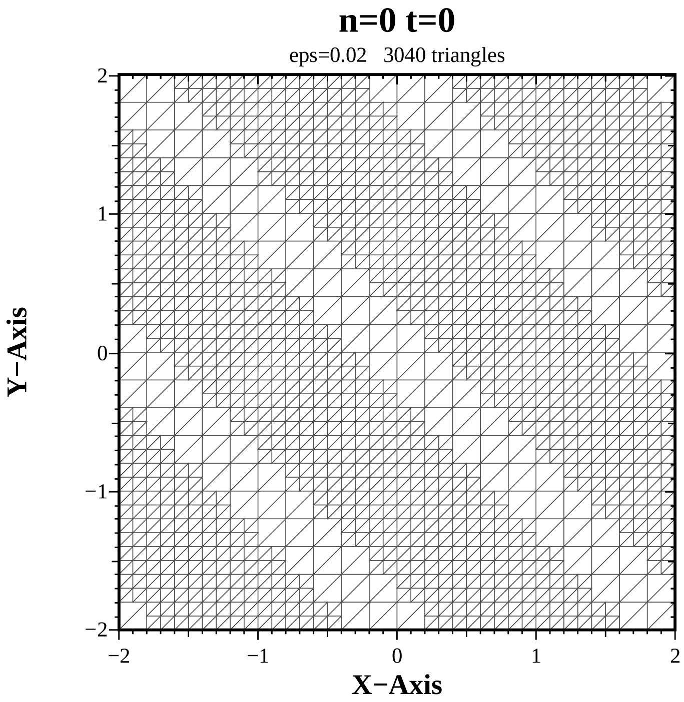
In 1D: comparison of AMR and local reconstruction on Sod tube test.

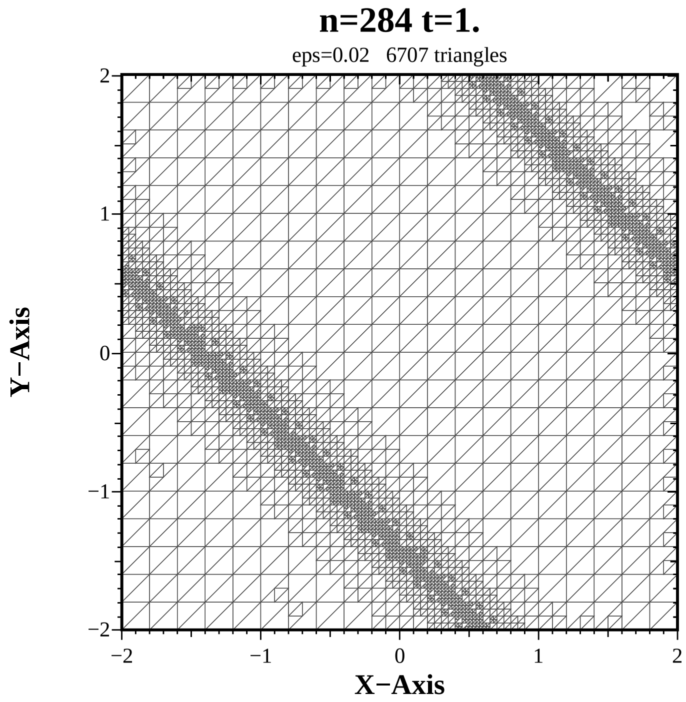




For a low order reference scheme, only local reconstruction preserves the accuracy with a substantial reduction of CPU time and memory space (1/20 at best).

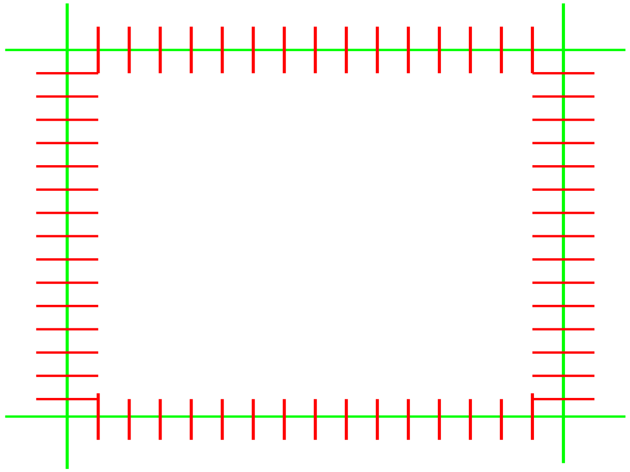
In 2D: Burgers equantion (triangle cell-average multiresolution).





Compression of solution by a factor 40 without loss of accuracy (smaller CPU saving).

Curse of dimensionality for cell-averages



For cell-averages in dimension d, complexity grows like

$$\mathcal{O}(\sum_{j=1}^{J} 2^{(d-1)(J-j)} \#(\tilde{\Lambda}_{\eta}^{n+1} \cap \nabla_j))$$

still less than $\mathcal{O}(\#(\Gamma_J))$.

Error Analysis

Remark: adaptive evolution with local reconstruction is given by

$$V_J^{n+1} = \mathcal{T}_{\Lambda_\eta^{n+1}} \mathcal{T}_{\tilde{\Lambda}_\eta^{n+1}} E_J V_J^n.$$

Compare $U_J^{n+1} = E_J U_J^n$ with $V_J^{n+1} = \mathcal{T}_{\Lambda_{n+1}} \mathcal{T}_{\tilde{\Lambda}_{n+1}} E_J V_J^n$.

Cumulative error analysis between both solutions:

$$\|U_J^{n+1} - V_J^{n+1}\| \le \|E_J U_J^n - E_J V_J^n\| + d_n,$$

with $d_n = \|V_J^{n+1} - E_J V_J^n\| \le t_n + c_n$ where

$t_n := \|\mathcal{T}_{\Lambda_{n+1}} \mathcal{T}_{\tilde{\Lambda}_{n+1}} E_J V_J^n - \mathcal{T}_{\tilde{\Lambda}_{n+1}} E_J V_J^n\|$, $c_n := \|\mathcal{T}_{\tilde{\Lambda}_{n+1}} E_J V_J^n - E_J V_J^n\|$,

denote the thresholding and refinement errors. The analysis of refinement and thresholding strategies should allow to control both terms with a prescribed precision ε.

Controling the thresholding error

Analysis based on underlying continuous wavelet system (ψ_λ):

$$\|U_J - \mathcal{T}_\Lambda U_J\| \leq \sum_{\lambda \notin \Lambda} \|d_\lambda \psi_\lambda\|.$$

For the L^1 norm, this gives $\|U_J - \mathcal{T}_\Lambda U_J\| \leq C \sum_{\lambda \notin \Lambda} 2^{-d|\lambda|} |d_\lambda|$, and therefore with $\eta_j = 2^{dj} \eta_0$,

$$\|U_J - \mathcal{T}_\eta U_J\| \leq C \sum_{2^{-d|\lambda|}|d_\lambda| < \eta_0} 2^{-d|\lambda|} |d_\lambda|$$

- Crudest estimate: $\eta_0 \#(\nabla_J) \sim \eta_0 2^{dJ} \Rightarrow$ take $\eta_0 = \varepsilon 2^{-dJ}$.

- Better estimate: $\eta_0 \#(\tilde{\Lambda}^{n+1}) \Rightarrow$ take $\eta_0 = \varepsilon / \#(\tilde{\Lambda}^{n+1})$.

- Even better: take largest η_0 s.t. $\sum_{2^{-d|\lambda|}|d_\lambda| < \eta_0} 2^{-d|\lambda|} |d_\lambda| \leq \varepsilon$.

Controling the refinement error

Harten's refinement rule for hyperbolic equations (assuming CFL condition for the reference scheme $\Delta t \leq C 2^{-J}$):

- If $|d_\lambda| > \eta_{|\lambda|}$ include in $\tilde{\Lambda}_\eta^{n+1}$ the neighbors of λ at the same level.

- If $|d_\lambda| > 2^{r-1}\eta_{|\lambda|}$ also include the childrens of λ at the finer level.

Here r represents the order of accuracy of the prediction operator. Not sufficient to prove that $|d_\lambda(E_J V_J^n)| < \eta_{|\lambda|}$ if $\lambda \notin \tilde{\Lambda}_\eta^{n+1}$.

This can be proved by a more severe refinement rule: refine of n level if $2^{n(s-1)}\eta_{|\lambda|} \leq |d_\lambda| < 2^{(n+1)(s-1)}\eta_{|\lambda|}$, with s the Hölder smoothness of the underlying wavelet system.

In practice, however, we observe that Harten's rule is sufficient and that the thresolding error dominates the refinement error.

A crude error estimate

Assuming stability in the prescribed norm $\|\cdot\|$ for the reference scheme in the sense that $\|E_J U - E_J V\| \le (1 + c\Delta t)\|U - V\|$, this yields the cumulative (too pessimistic) estimate

$$\|U_J^{n+1} - V_J^{n+1}\| \le (1 + c\Delta t)\|U_J^n - V_J^n\| + \varepsilon \le \cdots \le C(T)n\varepsilon \sim \frac{\varepsilon}{\Delta x}$$

For conservation laws, a natural choice is the L^1 norm. In most practical cases, we observe that thresholding and refinement error does not accumulate linearly.

Moreover, no error bound is available in terms of the number N. of wavelet coefficients which is used to represent the adaptive solution, although we can expect $\varepsilon \sim N^{-s}, \;\; s = \min\{s(\psi), s(u_0)\}$ and therefore when equilibrating with the error estimate $(\Delta x)^\alpha$ for the reference scheme, we would obtain the global L^1 error

$$e_n \lesssim \frac{N^{-s}}{\Delta x} \sim (\Delta x)^\alpha \sim N^{-s\frac{\alpha}{\alpha+1}}$$

The role of oscillations

A simple example : upwind scheme for the 1D linear advection equation $\partial_t u + a\partial_x u = 0$ with $a > 0$:

$$u_k^n = (1-\nu)u_k^{n-1} + \nu u_{k-1}^{n-1}$$

with $\nu = a\frac{\Delta t}{\Delta x} \in]0,1[$. Consistancy error c_k propagated according to

$$c_k^m = (1-\nu)c_k^{m-1} + \nu c_{k-1}^{m-1} = \cdots = \sum_{p=0}^{m} \binom{m}{p} \nu^p (1-\nu)^{m-p} c_{k-p}.$$

Assuming that $c_k = \frac{\tilde{c}_k - \tilde{c}_{k-1}}{\Delta x}$ and that we have a control on $\Delta x \sum_k |\tilde{c}_k| = \|\text{consist}\|_{\text{Lip}'}$, we obtain

$$c_k^m = (\Delta x)^{-1} \sum_{p=-1}^{m} [\binom{m}{p} \nu^p (1-\nu)^{m-p} - \binom{m}{p+1} \nu^{p+1} (1-\nu)^{m-p-1}] \tilde{c}_{k-p},$$

and therefore

$$\begin{aligned}
\|c^m\|_{L^1} &= \Delta x \textstyle\sum_k |c_k^m| \\
&\leq \textstyle\sum_{p=-1}^{m} \left|\binom{m}{p}\nu^p(1-\nu)^{m-p} - \binom{m}{p+1}\nu^{p+1}(1-\nu)^{m-p-1}\right| \sum_k |\tilde{c}_k| \\
&\leq 2\sup_{p=-1}^{m} \left|\binom{m}{p}\nu^p(1-\nu)^{m-p}\right| \textstyle\sum_k |\tilde{c}_k| \sim \frac{1}{\sqrt{m}} \sum_k |\tilde{c}_k|.
\end{aligned}$$

It follows that at time $T \sim n\Delta x$ the L^1 error is controlled by

$$\sqrt{n}\sum_k |\tilde{c}_k| \sim (\Delta x)^{-3/2}\|\text{consist}\|_{\text{Lip}'}$$

This analysis allows to recover the classical $(\Delta x)^{1/2}$ estimate for uniform schemes ($\|\text{consist}\|_{\text{Lip}'} \sim (\Delta x)^2$ assuming BV smoothness).

We can apply it to the adaptive scheme with now a Lip′ thresholding strategy $\eta_j \sim \eta_0 2^{2j}$ ensuring $\|V_J^{n+1} - E_J V_J^n\|_{\text{Lip}'} \leq \varepsilon$, which leads to an L^1 error estimate

$$\|U_J^{n+1} - V_J^{n+1}\|_{L^1} \leq (\Delta x)^{-3/2}\varepsilon$$

We now expect $\varepsilon \sim N^{-(s+1)}, \quad s = \min\{s(\psi), s(u_0)\}$ and therefore when equilibrating with the error estimate $(\Delta x)^\alpha$ for the reference scheme, we would obtain the global L^1 error

$$e_n \lesssim (\Delta x)^{-3/2} N^{-s} \sim (\Delta x)^\alpha \sim N^{-s\frac{\alpha}{\alpha+3/2}}$$

A third possibility is to apply BV thresolding $\eta_j \sim \eta_0$ ensuring $\|V_J^{n+1} - E_J V_J^n\|_{BV} \leq \varepsilon$, for which we now expect an L^1 error estimate in

$$\|U_J^{n+1} - V_J^{n+1}\|_{L^1} \leq \varepsilon \sim N^{-(s-1)}$$

and therefore a similar estimate for e_n.

No clear winner between the three strategies !

Several remaining issues

- Automatic tuning of the thresholding parameter
- Efficient implementation of tree-structured representation
- Problem adapted refinement rules
- Arbitrarily high resolution and time adaptivity
- Prediction operators on unstructured grids
- Nonlinear prediction operators
- Breaking the curse of dimensionality
- More general time-frequency bases
- Adaptation of this approach to non-local implicit schemes
- Comparison to nonlinear approximation benchmark