Adaptive Wavelet-Galerkin Method

Rob Stevenson
Korteweg-de Vries Institute for Mathematics
University of Amsterdam

Contents

- Well-posed operator eqs
- Approximation classes
- Equiv. formulation as well-posed bi-infinite MV eq using Riesz bases
- Adaptive wavelet-Galerkin method, optimal comput. compl.

Well-posed operator equations

For (real) Hilbert spaces H and K, let $A : H \to K'$ linear and boundedly invertible (i.e., $A \in \mathcal{B}(H, K')$, $A^{-1} \in \mathcal{B}(K', H)$).

Given $f \in K'$, to solve

$$Au = f.$$

Examples include variational formulations of boundary value problems or integral equations.

Ex 1. • $(Au)(v) = \int_\Omega \nabla u \cdot \nabla v$, $H = K = H_0^1(\Omega)$ (Poisson problem),

- $(A(\vec{u}, p))(\vec{v}, q) = \int_\Omega \nabla \vec{u} : \nabla \vec{v} - \int_\Omega p \operatorname{div} \vec{v} - \int_\Omega q \operatorname{div} \vec{u}$, $H = K = H_0^1(\Omega)^n \times L_{2,0}(\Omega)$ for a domain $\Omega \subset \mathbb{R}^n$ (Stokes problem),

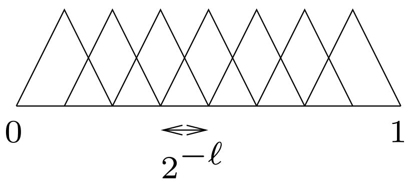
- $(Au)(v) = \frac{1}{4\pi} \int_{\partial\Omega} \int_{\partial\Omega} \frac{(u(y)-u(x))(v(y)-v(x))}{|x-y|^3} dx dy$, $\Omega \subset \mathbb{R}^3$, $H = K = H^{\frac{1}{2}}(\partial\Omega)/\mathbb{R}$ (hypersingular boundary integral equation).

- Parabolic problems. $H \neq K$.

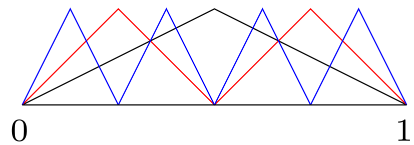
Dictionary and approximation classes

Let $\Psi \subset H$ be a *dictionary*, i.e. $\overline{\operatorname{span}\Psi} = H$.

Ex 2. Let $H = H_0^1(0,1)$.

1. all "hats" w.r.t. to any dyadic partition, i.e. union over $\ell \in \mathbb{N}_0$ of 

(*h-fem*)

2. the "hierarchical basis" 

3. all continuous, piecewise pols that vanish at $\{0,1\}$, w.r.t. any dyadic partition (*hp-fem*).

4. all polynomials on $(0,1)$ that vanish at $\{0,1\}$ (*p-fem*).

5. all continuous, piecewise pols of fixed order d, i.e., degree $d-1$, that vanish at $\{0,1\}$, w.r.t. any dyadic partition (*h-fem*).

6. $\{x \mapsto \sqrt{2}\sin(k\pi x) : k \in \mathbb{N}\}$ (*Fourier*).

For approximation of a general, sufficiently smooth function by a linear combination of N atoms, error at best either

$$\sim N^{-s_{\max}} \quad (\textit{algebraic rate})$$

(methods 1, 2, 5) or

$$\sim N^{-s_{\max}} e^{-\eta N^{\alpha_{\max}}} \quad (\textit{exponential rate})$$

(methods 3, 4, 6).

We restrict ourselves to dictionaries that give algebraic rates, i.e., apply fixed polynomial degrees $d-1$ (optimality results for adaptive methods that give exponential rates are currently subject of research).

For $s > 0$, define the *approximation class* $\mathcal{A}^s = \mathcal{A}^s(H, \Psi)$ as

$$\{u \in H : \sup_{N \in \mathbb{N}_0} (N+1)^s \inf_{\Psi_N \subset \Psi,\, \#\Psi_N = N} \inf_{v \in \operatorname{span} \Psi_N} \|u - v\|_H < \infty\}.$$

Def 1. Num. method for solving op. eq. from $\operatorname{span} \Psi$ is said to converge with the *optimal rate* if whenever $u \in \mathcal{A}^s$ for some $s > 0$, then the num. approx. from the span of N atoms has an error $\mathcal{O}(N^{-s})$, or equivalently, to produce an approximation within $\varepsilon > 0$ it needs $\mathcal{O}(\varepsilon^{-1/s})$ atoms.

Method is said to be of *optimal comput. complexity*, if whenever $u \in \mathcal{A}^s$ for some $s > 0$, then to produce an approximation within $\varepsilon > 0$ it needs $\mathcal{O}(\varepsilon^{-1/s})$ operations.

.

Ex 3. $H = H^1_0(0,1)$, Ψ collection 1 (all "hats" over all dyadic partitions). For $s > 1$, $\mathcal{A}^s$ is "essentially" empty.
On the other hand, $H^2(0,1) \cap H^1_0(0,1) \subset \mathcal{A}^1$ (uniform refinements suffice).
So $s_{\max} = 1$.

For $s \in (0,1]$, $\mathcal{A}^s$ is much larger than $H^s(0,1) \cap H^1_0(0,1)$; for $s \in (0,1)$, $\mathcal{A}^s \subset W^{1+s}_\tau(0,1) \cap H^1_0(0,1)$ where $\tau = (\frac{1}{2} + s)^{-1}$.

Riesz bases

We will assume that $\Psi = \{\psi_\lambda : \lambda \in \nabla\}$ is linearly independent (2 and 6 are, but 1, 3, 4, and 5 aren't), and even a *Riesz basis*. That is, the *synthesis* operator

$$\mathcal{F}' : \ell_2(\nabla) \to H : \mathbf{c} \mapsto \mathbf{c}^\top \Psi := \sum_{\lambda \in \nabla} c_\lambda \psi_\lambda \quad \text{is boundedly invertible,}$$

and so is its adjoint, known as the *analysis* operator,

$$\mathcal{F} : H' \to \ell_2(\nabla) : g \mapsto g(\Psi) := [g(\psi_\lambda)]_{\lambda \in \nabla}$$

[$L \in \mathcal{B}(X, Y)$, then adjoint $L' \in \mathcal{B}(Y', X')$ is defined by $(L'f)(x) := f(Lx)$]

[Properly scaled, 2 and 6 are even *orthonormal* bases for $H_0^1(0,1)$, equipped with $\langle u, v\rangle = \int_0^1 u'v'$]

$Au = f$ reform. as a bi-infinite matrix vector eq. $\mathbf{A}\mathbf{u} = \mathbf{f}$

Riesz bases $\Psi^H = \{\psi_\lambda^H : \lambda \in \nabla_H\}$, $\Psi^K = \{\psi_\lambda^K : \lambda \in \nabla_K\}$ for H, K. Analysis operators $\mathcal{F}_H$, $\mathcal{F}_K$.

Write $u = \mathcal{F}_H'\mathbf{u}$ $(= \mathbf{u}^\top \Psi^H)$. Then

$$Au = f \Longleftrightarrow \underbrace{\mathcal{F}_K A \mathcal{F}_H'}_{\mathbf{A}} \mathbf{u} = \underbrace{\mathcal{F}_K f}_{\mathbf{f}},$$

where $\mathbf{A}_{\lambda\mu} = (A\psi_\mu^H)(\psi_\lambda^K)$, $\mathbf{f}_\lambda = f(\psi_\lambda^K)$.

$$\|\mathbf{A}\|_{\ell_2(\nabla_H)\to\ell_2(\nabla_K)} \leq \|\mathcal{F}_K\|_{K'\to\ell_2(\nabla_K)}\|A\|_{H\to K'}\|\mathcal{F}_H'\|_{\ell_2(\nabla_H)\to H}$$

$$\|\mathbf{A}^{-1}\|_{\ell_2(\nabla_K)\to\ell_2(\nabla_H)} \leq \|\mathcal{F}_H'^{-1}\|_{H\to\ell_2(\nabla_H)}\|A^{-1}\|_{K'\to H}\|\mathcal{F}_K^{-1}\|_{\ell_2(\nabla_K)\to K'}$$

With $K \ni v = \mathcal{F}_K'\mathbf{v}$,

$$(Au)(v) = (A\mathcal{F}_H'\mathbf{u})(\mathcal{F}_K'\mathbf{v}) = \langle \mathbf{A}\mathbf{u}, \mathbf{v}\rangle_{\ell_2(\nabla_K)\times\ell_2(\nabla_K)}.$$

So if $H = K$ and $\Psi_H = \Psi_K$, then $A = A'$ iff $\mathbf{A} = \mathbf{A}^\top$, and A coercive (i.e., $(Av)(v) \gtrsim \|v\|_H^2$) iff $\mathbf{A} \gtrsim \mathrm{Id}$.

$$u \in \mathcal{A}^s \Longleftrightarrow \begin{cases} \sup_{N\in\mathbb{N}_0}(N+1)^s\|\mathbf{u}-\mathbf{u}_N\|_{\ell_2(\nabla_H)} < \infty \\ \text{where } \mathbf{u}_N \text{ is a best } N\text{-term approx. for } \mathbf{u}. \end{cases}$$

Adaptive wavelet-Galerkin method for solving $\mathrm{Au} = \mathrm{f}$

Let $\nabla = \nabla_H = \nabla_K$, and $\mathbf{A} = \mathbf{A}^\top \eqsim \mathrm{Id}$.

Set $\langle\cdot,\cdot\rangle := \langle\cdot,\cdot\rangle_{\ell_2(\nabla)\times\ell_2(\nabla)}$, $\|\cdot\| := \|\cdot\|_{\ell_2(\nabla)}$, energy-norm $|||\cdot||| := \langle\mathbf{A}\cdot,\cdot\rangle^{\frac{1}{2}}$ [so $|||\mathbf{u}||| = (Au)(u)^{\frac{1}{2}}$ if $\mathbf{A}$ is stiffness.]

For $\Lambda \subset \nabla$, $\ell_2(\Lambda)$ is subspace of $\ell_2(\nabla)$ of vectors supported on Λ.
$\mathbf{I}_\Lambda : \ell_2(\Lambda) \to \ell_2(\nabla)$ trivial embedding, $\mathbf{P}_\Lambda := \mathbf{I}_\Lambda^\top$ sets all coefficients outside Λ to zero.

$\mathbf{A}_\Lambda := \mathbf{P}_\Lambda\mathbf{A}\mathbf{I}_\Lambda$, $\mathbf{u}_\Lambda := \mathbf{A}_\Lambda^{-1}\mathbf{P}_\Lambda\mathbf{f}$ is Galerkin solution.
[For $\mathbf{w}_\Lambda \in \ell_2(\Lambda)$, $\langle\mathbf{A}(\mathbf{u}-\mathbf{u}_\Lambda),\mathbf{w}_\Lambda\rangle = \langle\mathbf{f}-\mathbf{P}_\Lambda\mathbf{f},\mathbf{w}_\Lambda\rangle = 0$, so $\mathbf{u}_\Lambda$ is the best approximation to $\mathbf{u}$ from $\ell_2(\Lambda)$ w.r.t. the energy norm.]
[$u_\Lambda := \mathbf{u}_\Lambda^\top\Psi$ is the best approximation from $\mathrm{span}\{\psi_\lambda : \lambda \in \Lambda\}$ to u w.r.t. $(A\cdot)(\cdot)^{\frac{1}{2}}$.]

We have

$$\|\mathbf{A}_\Lambda^{-1}\|^{-\frac{1}{2}}\|\cdot\| \leq \quad |||\cdot||| \leq \|\mathbf{A}_\Lambda\|^{\frac{1}{2}}\|\cdot\| \quad \text{on } \ell_2(\Lambda),$$

$$\|\mathbf{A}_\Lambda^{-1}\|^{-\frac{1}{2}}|||\cdot||| \leq \|\mathbf{A}_\Lambda\cdot\| \leq \|\mathbf{A}_\Lambda\|^{\frac{1}{2}}|||\cdot||| \quad \text{on } \ell_2(\Lambda),$$

as well as $\|\mathbf{A}_\Lambda\| \leq \|\mathbf{A}\|$ and $\|\mathbf{A}_\Lambda^{-1}\| \leq \|\mathbf{A}^{-1}\|$.

Observation: $|||\mathbf{u}-\mathbf{w}||| \eqsim \|\mathbf{f}-\mathbf{A}\mathbf{w}\| \rightsquigarrow$ use residual as **a posteriori error estimator.**

Prop 1. *Let* $\theta \in (0,1]$, $\Lambda \subset \Xi \subset \nabla$, *s.t.*

$$\|\mathbf{P}_\Xi(\mathbf{f} - \mathbf{A}\mathbf{u}_\Lambda)\| \geq \theta \|\mathbf{f} - \mathbf{A}\mathbf{u}_\Lambda\|. \tag{1}$$

Then $|||\mathbf{u} - \mathbf{u}_\Xi||| \leq [1 - \kappa(\mathbf{A})^{-1}\theta^2]^{\frac{1}{2}} |||\mathbf{u} - \mathbf{u}_\Lambda|||$.

Proof.

$$\begin{aligned} |||\mathbf{u}_\Xi - \mathbf{u}_\Lambda||| &\geq \|\mathbf{A}\|^{-\frac{1}{2}} \|\mathbf{A}_\Xi(\mathbf{u}_\Xi - \mathbf{u}_\Lambda)\| = \|\mathbf{A}\|^{-\frac{1}{2}} \|\mathbf{P}_\Xi(\mathbf{f} - \mathbf{A}\mathbf{u}_\Lambda)\| \\ &\geq \|\mathbf{A}\|^{-\frac{1}{2}} \theta \|\mathbf{f} - \mathbf{A}\mathbf{u}_\Lambda\| \geq \theta \kappa(\mathbf{A})^{-\frac{1}{2}} |||\mathbf{u} - \mathbf{u}_\Lambda||| \end{aligned}$$

and Galerkin orthogonality $|||\mathbf{u} - \mathbf{u}_\Lambda|||^2 = |||\mathbf{u} - \mathbf{u}_\Xi|||^2 + |||\mathbf{u}_\Xi - \mathbf{u}_\Lambda|||^2$. □

Remark: Argument similar to the one that guarantees convergence of adaptive finite element methods using the bulk criterion (Dörfler marking).

Prop 2. *If in* (1), $\theta < \kappa(\mathbf{A})^{-\frac{1}{2}}$ *and* Ξ *is the* smallest *set satisfying* (1), *then* $\#(\Xi \backslash \Lambda) \leq N$ *for smallest* N *s.t.*

$$|||\mathbf{u} - \mathbf{u}_N||| \leq [1 - \theta^2 \kappa(\mathbf{A})]^{\frac{1}{2}} |||\mathbf{u} - \mathbf{u}_\Lambda||| \tag{2}$$

Proof. For an N as in (2), let $\Sigma := \Lambda \bigcup \operatorname{supp} \mathbf{u}_N$. Then, for the solution of $\mathbf{A}_\Sigma \mathbf{u}_\Sigma = \mathbf{P}_\Sigma \mathbf{f}$, we have $|||\mathbf{u} - \mathbf{u}_\Sigma||| \leq |||\mathbf{u} - \mathbf{u}_N|||$, and so by (2) and Galerkin orthogonality

$$|||\mathbf{u}_\Sigma - \mathbf{u}_\Lambda||| \geq \theta \kappa(\mathbf{A})^{\frac{1}{2}} |||\mathbf{u} - \mathbf{u}_\Lambda|||,$$

giving

$$\begin{aligned} \|\mathbf{P}_\Sigma(\mathbf{f} - \mathbf{A}\mathbf{u}_\Lambda)\| &= \|\mathbf{A}_\Sigma(\mathbf{u}_\Sigma - \mathbf{u}_\Lambda)\| \geq \|\mathbf{A}^{-1}\|^{-\frac{1}{2}} |||\mathbf{u}_\Sigma - \mathbf{u}_\Lambda||| \\ &\geq \|\mathbf{A}^{-1}\|^{-\frac{1}{2}} \theta \kappa(\mathbf{A})^{\frac{1}{2}} |||\mathbf{u} - \mathbf{u}_\Lambda||| \geq \theta \|\mathbf{f} - \mathbf{A}\mathbf{u}_\Lambda\|. \end{aligned}$$

By our assumption on Ξ, we conclude that $\#(\Xi \backslash \Lambda) \leq \#(\Sigma \backslash \Lambda) \leq N$. □

Corol 1. *If* $u \in \mathcal{A}^s$, $\#(\Xi \backslash \Lambda) \lesssim |||\mathbf{u} - \mathbf{u}_\Lambda|||^{-1/s}$.

Remark: Argument as in **Prop 2** was used later to prove optimality of adaptive finite element methods

(Idealized) adaptive wavelet-Galerkin method

% Let $\theta \in (0, \kappa(\mathbf{A})^{-\frac{1}{2}})$ *a constant.*

$i := 0$, $\Lambda_i := \emptyset$, $\mathbf{u}_{\Lambda_i} := 0$.
`for` $i = 0, 1, \ldots$ `do`
  Determine smallest $\Lambda_{i+1} \supset \Lambda_i$ *with* $\|\mathbf{P}_{\Lambda_{i+1}}(\mathbf{f} - \mathbf{A}\mathbf{u}_{\Lambda_i})\| \geq \theta \|\mathbf{f} - \mathbf{A}\mathbf{u}_{\Lambda_i}\|$
  Solve $\mathbf{A}_{\Lambda_{i+1}} \mathbf{u}_{\Lambda_{i+1}} = \mathbf{P}_{\Lambda_{i+1}} \mathbf{f}$
`endfor`

Thm 1. *If* $\mathbf{u} \in \mathcal{A}^s$, *then* $\sup_\ell (\#\Lambda_\ell)^s |||\mathbf{u} - \mathbf{u}_{\Lambda_\ell}||| < \infty$ **(optimal rate)**.

Proof.

$$\#\Lambda_\ell = \sum_{k=1}^{\ell} \#(\Lambda_k \backslash \Lambda_{k-1}) \overset{\textbf{Corol1}}{\lesssim} \sum_{k=1}^{\ell} |||\mathbf{u} - \mathbf{u}_{\Lambda_{k-1}}|||^{-1/s}$$

$$\overset{\textbf{Prop1}}{\lesssim} |||\mathbf{u} - \mathbf{u}_{\Lambda_{\ell-1}}|||^{-1/s} \overset{\textbf{Prop1}}{\leq} |||\mathbf{u} - \mathbf{u}_{\Lambda_\ell}|||^{-1/s}. \qquad \square$$

Practical adaptive wavelet-Galerkin method (awgm)

% Let $0 < \mu_0 \leq \mu_1 < 1$, $\delta \in (0,1)$, $\gamma > 0$ *constants.*

$i := 0$, $\Lambda_i := \emptyset$, $\mathbf{w}_{\Lambda_i} := 0$.
for $i = 0, 1, \ldots$ do

Compute $\mathbf{r}^{(i)}$ *with* $\frac{\|\mathbf{r}^{(i)} - (\mathbf{f} - \mathbf{A}\mathbf{w}_{\Lambda_i})\|}{\|\mathbf{f} - \mathbf{A}\mathbf{w}_{\Lambda_i}\|} \leq \delta$.

Determine $\Lambda_{i+1} \supset \Lambda_i$ *with* $\|\mathbf{P}_{\Lambda_{i+1}}\mathbf{r}^{(i)}\| \geq \mu_0 \|\mathbf{r}^{(i)}\|$ *and* $\#(\Lambda_{i+1} \setminus \Lambda_i) \lesssim \#(\tilde{\Lambda} \setminus \Lambda_i)$ *for any* $\tilde{\Lambda} \supset \Lambda_i$ *with* $\|\mathbf{P}_{\tilde{\Lambda}}\mathbf{r}^{(i)}\| \geq \mu_1 \|\mathbf{r}^{(i)}\|$.

Compute $\mathbf{w}_{\Lambda_{i+1}} \in \ell_2(\Lambda_{i+1})$ *with* $\|\mathbf{w}_{\Lambda_{i+1}} - \mathbf{A}_{\Lambda_{i+1}}^{-1}\mathbf{P}_{\Lambda_{i+1}}\mathbf{f}\| \leq \gamma \|\mathbf{r}^{(i)}\|$.

endfor

[$\|\mathbf{r}^{(i)}\|$ provides a reliable and efficient stopping criterium.]

Thm 2. *If* μ_1, δ, γ *are sufficiently small, then, whenever* $\mathbf{u} \in \mathcal{A}^s$, $\sup_i (\#\Lambda_i)^s \|\mathbf{u} - \mathbf{w}_{\Lambda_i}\| < \infty$.

The approximate residual computation can be implemented as

let $\zeta \eqsim \|\mathbf{r}^{(i-1)}\|$
do $\zeta := \zeta/2$. *Compute* $\mathbf{r}^{(i)}$ *with* $\|\mathbf{r}^{(i)} - (\mathbf{f} - \mathbf{A}\mathbf{w}_{\Lambda_i})\| \leq \zeta$.
until $\zeta \leq \frac{\delta}{1+\delta}\|\mathbf{r}^{(i)}\|$

Cost

Let C_i $\#$ arith. ops. to produce $\mathbf{w}_{\Lambda_i}$.

Prop 3. *Let* $\mathbf{u} \in \mathcal{A}^s$*. If cost of approximating* $\mathbf{f} - \mathbf{A}\mathbf{w}_{\Lambda_i}$ *within* ε *is*

$$\lesssim \varepsilon^{-1/s} + \#\Lambda_i,$$

then $\sup_i C_i^s \|\mathbf{u} - \mathbf{w}_{\Lambda_i}\| < \infty$. **(opt. comput. compl)**

Proof. [sketch] Known $\#\Lambda_i \lesssim \|\mathbf{u} - \mathbf{w}_{\Lambda_i}\|^{-1/s}$. Tol ε_i for approximating $\mathbf{f} - \mathbf{A}\mathbf{w}_{\Lambda_i}$ is $\gtrsim \|\mathbf{u} - \mathbf{w}_{\Lambda_i}\|$.

So cost of all approximate res evals up to ith one is $\lesssim \sum_{\ell=0}^{i} \|\mathbf{u} - \mathbf{w}_{\Lambda_\ell}\|^{-1/s} \lesssim \|\mathbf{u} - \mathbf{w}_{\Lambda_i}\|^{-1/s}$ by the linear decrease of the errors.

Galerkin matrices uniformly well-conditioned. So starting from previous iterand, a fixed number of inexact, say, Rich. iteration suffices. The cost of the inexact (finite!) residual evals can be dealt with as above. $\square$

Discussion

Are going to approx. both $\mathbf{f}$ and $\mathbf{A}\mathbf{w}$ within tol $\varepsilon/2$. Assuming $\mathbf{u} \in \mathcal{A}^s$, needed cost $\lesssim \varepsilon^{-1/s} + \#\operatorname{supp}\mathbf{w}$. Focus on second task.

Generally, each column of $\mathbf{A}$ has infinitely many non-zeros, but sizes decay away from "diagonal" ($\mathbf{A}$ is near-sparse). Obvious approach, given ε, replace $\mathbf{A}$ by sufficiently large "band". Yields cost $C(\varepsilon) \times \#\operatorname{supp}\mathbf{w}$ with $C(\varepsilon) \uparrow \infty$.

Idea: approx columns corresponding to large entries of $\mathbf{w}$ more accurately than those that correspond to small ones. Only helpful when the entries of $\mathbf{w}$, ordered by their modulus, exhibit some decay, i.e., when $\mathbf{w}$ is near-sparse.

Near-sparsity of iterands inside awgm will follow from near-sparsity of $\mathbf{u}$, given by $\mathbf{u} \in \mathcal{A}^s$.

Def 2. On $\mathcal{A}^s$, $\|\mathbf{u}\|_{\mathcal{A}^s} := \sup_{N\in\mathbb{N}_0}(N+1)^s\|\mathbf{u}-\mathbf{u}_N\|$ $\quad [\geq \|\mathbf{u}\|.]$

Lem 1. *Let* $\mathbf{u} \in \mathcal{A}^s$ *and* $\mathbf{w} \in \ell_0(\nabla)$ *(i.e. finitely supported). Then* $\|\mathbf{w}\|_{\mathcal{A}^s} \lesssim \|\mathbf{u}\|_{\mathcal{A}^s} + (\#\operatorname{supp}\mathbf{w})^s\|\mathbf{u}-\mathbf{w}\|$.

Proof. For $N < \#\operatorname{supp}\mathbf{w}$, $N^s\|\mathbf{w}-\mathbf{w}_N\| \leq N^s\|\mathbf{w}-\mathbf{u}_N\|$
$\leq N^s(\|\mathbf{u}-\mathbf{u}_N\| + \|\mathbf{w}-\mathbf{u}\|)$. □

Corol 2. *If* $\mathbf{u} \in \mathcal{A}^s$, *then the iterands produced by* awgm *satisfy* $\sup_i \|\mathbf{w}_{\Lambda_i}\|_{\mathcal{A}^s} < \infty$.

Thm 3. *Let* $\mathbf{u} \in \mathcal{A}^s$. *If for* $\mathbf{w} \in \ell_0(\nabla)$ *cost of evaluating* $\mathbf{f} - \mathbf{A}\mathbf{w}$ *within tol* ε *is*

$$\lesssim \varepsilon^{-1/s}(1 + \|\mathbf{w}\|_{\mathcal{A}^s}^{1/s}) + \#\operatorname{supp}\mathbf{w},$$

then for $(\mathbf{w}_{\Lambda_i})_i$ *produced by* awgm, $\sup_i C_i^s\|\mathbf{u}-\mathbf{w}_{\Lambda_i}\| < \infty$. **(opt. comput. compl)**.

Routine `apply`

Def 3. $\mathbf{A}$ is called s^*-computable when $\exists$ pos, summable (α_j), (β_j), and for all j, $\exists \mathbf{A}_j$ having $2^j\alpha_j$ non-zeros in each column (whose computation takes $\mathcal{O}(2^j\alpha_j)$ ops) with $\|\mathbf{A}-\mathbf{A}_j\| \leq 2^{-js^*}\beta_j$.

For $\mathbf{w} \in \ell_2(\nabla)$, let $\mathbf{w}_{[k]}$ denotes its (near) best 2^k approx. Approx $\mathbf{A}\mathbf{w}$ by

$$\mathbf{z}_j := \mathbf{A}_j\mathbf{w}_{[0]} + \mathbf{A}_{j-1}(\mathbf{w}_{[1]} - \mathbf{w}_{[0]}) + \cdots + \mathbf{A}_0(\mathbf{w}_{[j]} - \mathbf{w}_{[j-1]}) \quad (\texttt{apply})$$

so that

$$\mathbf{A}\mathbf{w} - \mathbf{z}_j = \mathbf{A}(\mathbf{w} - \mathbf{w}_{[j]}) + (\mathbf{A}-\mathbf{A}_0)(\mathbf{w}_{[j]} - \mathbf{w}_{[j-1]}) + \cdots + (\mathbf{A}-\mathbf{A}_j)\mathbf{w}_{[0]}$$

Then

$$\#\,\mathrm{supp}\,\mathbf{z}_j \leq \alpha_j 2^j + \alpha_{j-1}2^{j-1}(2-1) + \cdots + \alpha_0(2^j - 2^{j-1}) \lesssim 2^j,$$

and, for $\mathbf{w} \in \ell_0(\nabla)$, the cost to computing it is $\lesssim 2^j + \#\,\mathrm{supp}\,\mathbf{w}$, and

$$\begin{aligned}\|\mathbf{A}\mathbf{w} - \mathbf{z}_j\| &\leq \|\mathbf{A}\|2^{-js}\|\mathbf{w}\|_{\mathcal{A}^s} + \beta_0(2^{-js}+2^{-(j-1)s})\|\mathbf{w}\|_{\mathcal{A}^s} + \cdots + 2^{-js^*}\beta_j\|\mathbf{w}\|_{\mathcal{A}^s}\\ &\lesssim 2^{-sj}\|\mathbf{w}\|_{\mathcal{A}^s}.\end{aligned}$$

[For $s \leq s^*$, $\|\mathbf{A}\mathbf{w} - \mathbf{z}_j\| \lesssim 2^{-sj}\|\mathbf{w}\|_{\mathcal{A}^s}$; $\#\operatorname{supp}\mathbf{z}_j \lesssim 2^j$; cost $\lesssim 2^j + \#\operatorname{supp}\mathbf{w}$]

First consequence: With $N := \#\operatorname{supp}\mathbf{z}_j$,

$$N^s\|\mathbf{A}\mathbf{w} - (\mathbf{A}\mathbf{w})_N\| \leq N^s\|\mathbf{A}\mathbf{w} - \mathbf{z}_j\| \lesssim \|\mathbf{w}\|_{\mathcal{A}^s},$$

and so ($\|\mathbf{A}\mathbf{w}\| \lesssim \|\mathbf{w}\| \leq \|\mathbf{w}\|_{\mathcal{A}^s}$),

$$\|\mathbf{A}\mathbf{w}\|_{\mathcal{A}^s} = \sup_{N \in \mathbb{N}_0} (N+1)^s\|\mathbf{A}\mathbf{w} - (\mathbf{A}\mathbf{w})_N\| \lesssim \|\mathbf{w}\|_{\mathcal{A}^s}.$$

So if $\mathbf{u} \in \mathcal{A}^s$ for some $s \leq s^*$, then $\mathbf{f} \in \mathcal{A}^s$, and so it *can* be approximated at rate s. In other words, given ε, $\exists\tilde{\mathbf{f}}$ with $\|\mathbf{f} - \tilde{\mathbf{f}}\| \leq \varepsilon/2$ and $\#\operatorname{supp}\tilde{\mathbf{f}} \lesssim \varepsilon^{-1/s}$. Cost ...

Second consequence: Setting $\varepsilon/2 = C2^{-sj}\|\mathbf{w}\|_{\mathcal{A}^s}$, cost $\lesssim \varepsilon^{-1/s}\|\mathbf{w}\|_{\mathcal{A}^s}^{1/s} + \#\operatorname{supp}\mathbf{w}$.

Conclusion: `awgm` has opt. comput. compl. when $\mathbf{A}$ is s^*-computable for some $s^* \geq s_{\max}$.

Nonlinear eqs

Let $A : H \supset \mathrm{dom}(H) \to H'$, and $A(u) = 0$. Assume

- A is continuously Fréchet differentiable in a neighborhood of u,
- $DA(u) = DA(u)'$ is coercive

Writing $u = \mathcal{F}'\mathbf{u}$, with $\mathbf{A} := \mathcal{F} A \mathcal{F}'$,

$$A(u) = 0 \Longleftrightarrow \mathbf{A}(\mathbf{u}) = 0$$

`awgm` applies and converges with the **optimal rate** when started sufficiently close to the solution. Opt. *comput* compl. verified for some cases.

Nonelliptic problems or problems with nonelliptic Fréchet derivative

Let $A \in \mathcal{B}(H, K')$ with $\|A \cdot\|_{K'} \eqsim \|\cdot\|_H$. Then

$$u = \underset{w \in H}{\operatorname{argmin}} \tfrac{1}{2}\|Aw - f\|^2_{K'} \Longleftrightarrow \mathbf{A}^\top \mathbf{A}\mathbf{u} = \mathbf{A}^\top \mathbf{f}$$

($\mathbf{A} = \mathcal{F}_K A \mathcal{F}'_H$, $u = \mathcal{F}'_H \mathbf{u}$). `awgm` applies.

Let $A : H \supset \mathrm{dom}(H) \to K'$, and $A(u) = 0$. Assume

- A is twice continuously Fréchet differentiable in a neighborhood of u,
- $\|DA(u)(\cdot)\|_{K'} \eqsim \|\cdot\|_H$

Then

$$u = \underset{w \in \mathrm{dom}(A)}{\operatorname{argmin}} \tfrac{1}{2}\|A(w)\|^2_{K'} \Longleftrightarrow (\mathcal{F}_K DA(\mathcal{F}'_H \mathbf{u})\mathcal{F}'_H)^\top \mathcal{F}_K A(\mathcal{F}'_H \mathbf{u}) = 0$$

`awgm` applies.

Ex 4. ODE (model for PDE written as first order system)

$$\begin{cases} u' + \alpha u^3 & = & f \quad \text{on } (0,1), \\ u(0) & = & u_0, \end{cases}$$

where $\alpha = \begin{cases} x \mapsto 1 + 3x & \text{on } (0, \frac{1}{3}), \\ x \mapsto 5 - 3x & \text{on } (\frac{1}{3}, 1). \end{cases}$

$H := H^1(0,1)$, $K := L_2(0,1) \times \mathbb{R}$,

$$A = (A_1, A_2) : H \to K' : w \mapsto (w' + \alpha w^3 - f, w(0) - u_0),$$

and so

$$DA(w) : H \to K' : v \mapsto (v' + \alpha 3 w^2 v, v(0)) \quad \text{is b.i}$$

Equip H with CDF(2,2) wavelets (adapted to interval). Since K is $L_2(0,1)$ it can be avoided to equip it with a basis.

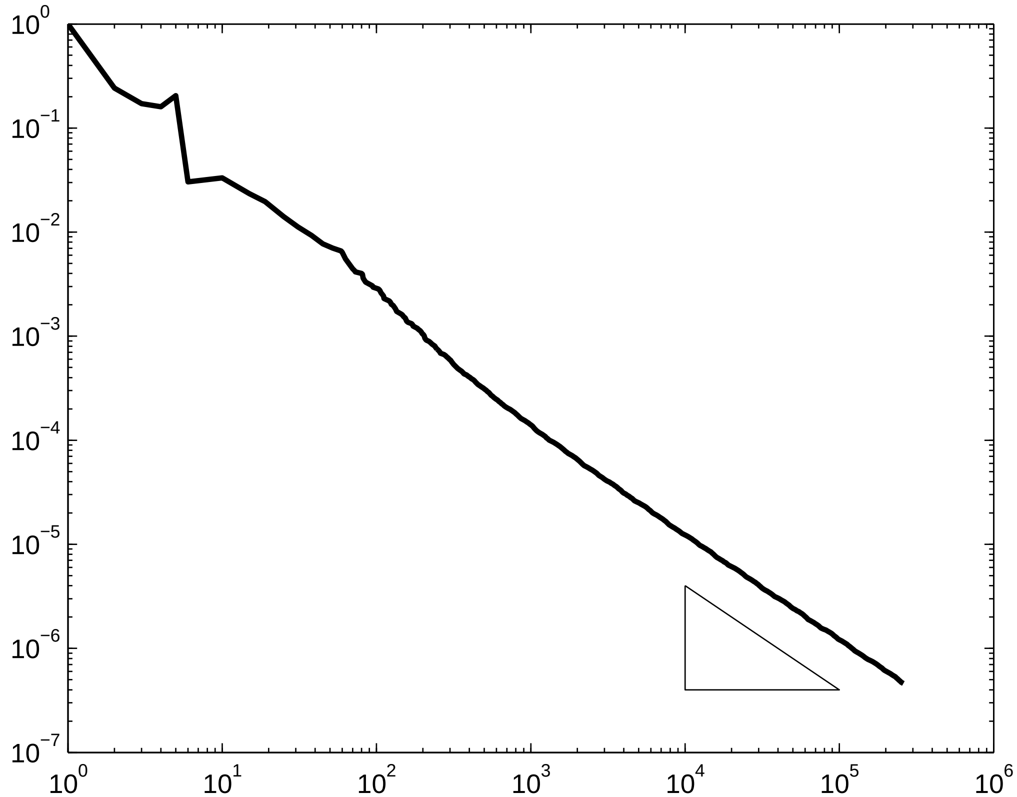


Figure 1: $\#\Lambda$ vs. $\|\mathbf{F}(\mathbf{w}_\Lambda)|_{\bar{\Lambda}}\|_{\ell_2(\bar{\Lambda})}$. The hypothenuse has slope -1.

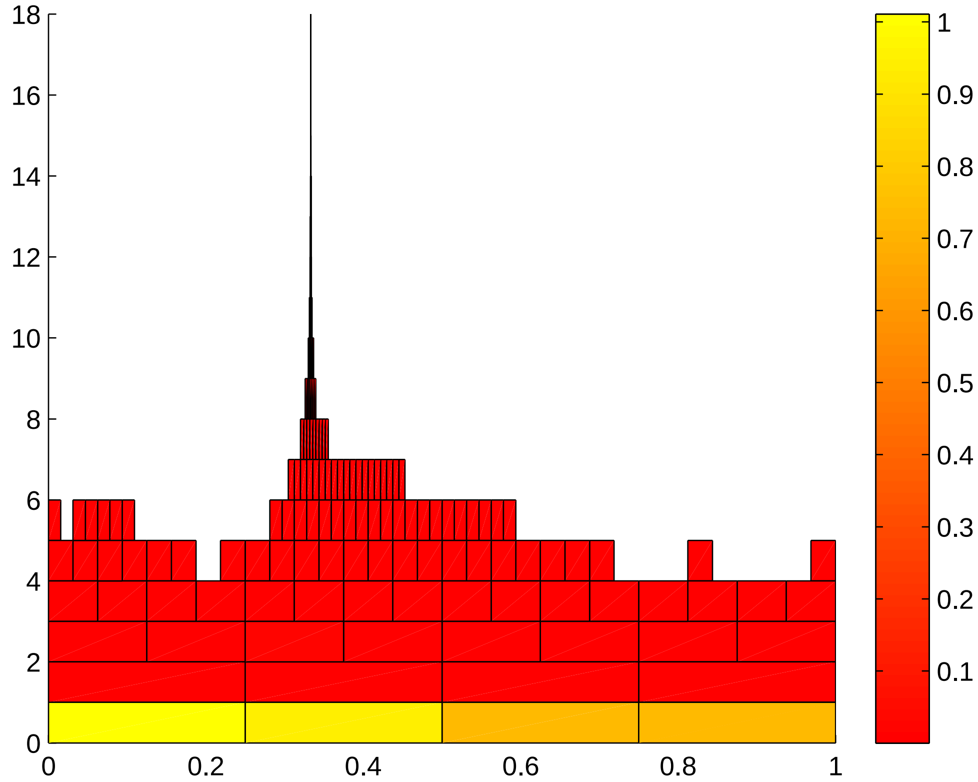


Figure 2: Distribution of the wavelet coefficients over levels and locations for a support size $\#\Lambda = 169$.

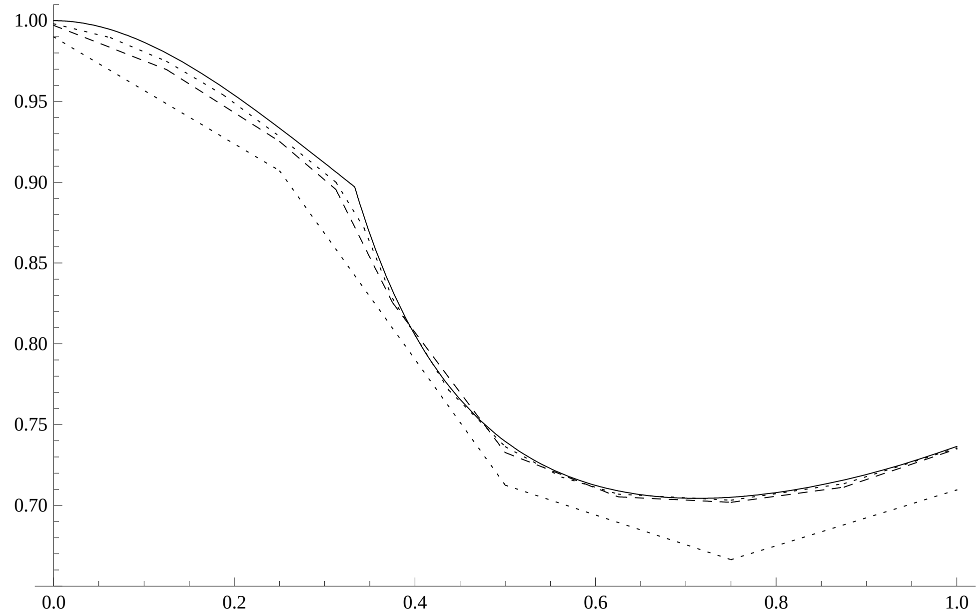


Figure 3: Approximate solutions being linear combinations of 5, 10 and 15 wavelets. The exact solution is indicated with the solid line.

An alternative adaptive wavelet algorithm

Let $\mathbf{A} = \mathbf{A}^\top \eqsim \mathrm{Id}$, then for suitable α, damped Richardson

$$\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + \alpha(\mathbf{f} - \mathbf{A}\mathbf{u}^{(i)})$$

converges linearly.
Still valid with *inexact* evals of the residual (`apply`-routine) for properly decreasing tolerances.
Generalizes to bounded $\mathbf{A}$, when $\mathbf{A}_s := \frac{1}{2}(\mathbf{A} + \mathbf{A}^\top) \eqsim \mathrm{Id}$.
Optimal choice of α by Steepest Decent or MINRES.

No reason to believe why *rates* are *optimal*, and in practice they *aren't*.
Idea: Extend iteration with a recurrent application of *coarsening*.

$\texttt{coarse}[\mathbf{w}, \varepsilon] \to \mathbf{w}_\varepsilon$*:*
% In: $\mathbf{w} \in \ell_0$ *and* $\varepsilon > 0$. *Out:* $\mathbf{w}_\varepsilon \in \ell_0$ *with*

$$\|\mathbf{w} - \mathbf{w}_\varepsilon\| \leq \varepsilon \quad \textit{and} \quad \#\operatorname{supp} \mathbf{w}_\varepsilon \lesssim \min\{N : \|\mathbf{w} - \mathbf{w}_N\| \leq \varepsilon\}. \qquad (3)$$

Prop 4. *Let* $\zeta > 1$, $s > 0$. *Then* $\forall \varepsilon > 0$, $\mathbf{v} \in \mathcal{A}^s$ *and* $\mathbf{w} \in \ell_0$ *with*

$$\|\mathbf{v} - \mathbf{w}\| \leq \varepsilon,$$

for $\mathbf{w}_{\zeta\varepsilon} := \texttt{coarse}[\zeta\varepsilon, \mathbf{w}]$, *one has* $\|\mathbf{v} - \mathbf{w}_{\zeta\varepsilon}\| \leq (1+\zeta)\varepsilon$,

$$\#\operatorname{supp} \mathbf{w}_{\zeta\varepsilon} \lesssim \varepsilon^{-1/s} \|\mathbf{v}\|_{\mathcal{A}^s}^{1/s}, \quad \|\mathbf{w}_{\zeta\varepsilon}\|_{\mathcal{A}^s} \lesssim \|\mathbf{v}\|_{\mathcal{A}^s}.$$

Proof. Consider smallest $N \in \mathbb{N}_0$ with $\|\mathbf{v} - \mathbf{v}_N\| \leq (\zeta - 1)\varepsilon$. Δ-ineq. shows $\|\mathbf{w} - \mathbf{v}_N\| \leq \zeta\varepsilon$, and so $\#\operatorname{supp} \mathbf{w}_{\zeta\varepsilon} \lesssim N \lesssim ((\zeta-1)\varepsilon)^{-1/s} \|\mathbf{v}\|_{\mathcal{A}^s}^{1/s}$. □

Resulting adaptive wavelet method (CDD2) has **opt. comput. complexity**, but turns out quantitatively worse than `awgm` (CDD1).
Explanation: Sawtooth behaviour. Once you have constructed a proper $\Lambda \subset \nabla$, better to find the best approximation from $\operatorname{span}\{\psi_\lambda : \lambda \in \Lambda\}$ (Galerkin).

Summary

$A \in \mathcal{B}(H, K')$ b.i., Riesz bases for H and K.

$$Au = f \Longleftrightarrow \mathbf{A}\mathbf{u} = \mathbf{f}.$$

Let $\mathbf{A} = \mathbf{A}^\top \eqsim \mathrm{Id}$ (otherwise normal eqs)

`for` $i = 0, 1, \ldots$ do
    solve $\mathbf{A}_\Lambda \mathbf{u}_\Lambda = \mathbf{P}_\Lambda \mathbf{f}$ *(Galerkin system)*
    pick (near) smallest $\Lambda_{i+1} \supset \Lambda_i$ *s.t.* $\|\mathbf{P}_{\Lambda_{i+1}}(\mathbf{f} - \mathbf{A}\mathbf{u}_{\Lambda_i})\| \geq \theta \|\mathbf{f} - \mathbf{A}\mathbf{u}_{\Lambda_i}\|$
`endfor`

Linearly convergent, and for suff. small θ, only dependent on $\kappa(\mathbf{A})$, with **opt. rate** (i.e. s, whenever $\mathbf{u} \in \mathcal{A}^s$).
Inexact version has **opt. comput. compl.** when $\mathbf{A}$ is $s_{\max}$-computable (suff. when for some $s^* > s_{\max}$, $\forall M$, $\exists$ approx. $\mathbf{A}_M$ with M nonzeros per column, that can be computed in $\mathcal{O}(M)$ ops, and $\|\mathbf{A} - \mathbf{A}_M\| \lesssim M^{-s^*}$).

References

[1] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations – Convergence rates. *Math. Comp*, 70:27–75, 2001.

[2] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods II - Beyond the elliptic case. *Found. Comput. Math.*, 2(3):203–245, 2002.

[3] R. DeVore. Nonlinear approximation. *Acta Numer.*, 7:51–150, 1998.

[4] T. Gantumur, H. Harbrecht, and R.P. Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comp.*, 76:615–629, 2007.

[5] R.P. Stevenson. Adaptive wavelet methods for solving operator equations: An overview. In R.A. DeVore and A. Kunoth, editors, *Multiscale, Nonlinear and Adaptive Approximation: Dedicated to Wolfgang Dahmen on the Occasion of his 60th Birthday*, pages 543–598. Springer, Berlin, 2009.

[6] R.P. Stevenson. Adaptive wavelet methods for linear and nonlinear least squares problems. Technical report, KdVI, UvA Amsterdam, November 2011. Submitted.

[7] Y. Xu and Q. Zou. Adaptive wavelet methods for elliptic operator equations with nonlinear terms. *Adv. Comput. Math.*, 19(1-3):99–146, 2003. Challenges in computational mathematics (Pohang, 2001).