Near-sparsity of PDOs in wavelet coordinates

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

PDO

Recall: Sufficient for `awgm` to have **opt. comput. compl.** is that $\exists s^* > s_{\max}$, $\forall M$, $\exists \mathbf{A}_M$ with M nonzeros per column, that can be computed in $\mathcal{O}(M)$ ops, and $\|\mathbf{A} - \mathbf{A}_M\| \lesssim M^{-s^*}$.

Let's refer to property of having M nonzeros per column, and error $\lesssim M^{-s^*}$, as s^***-compressibility**. If, additionally, these columns can be computed in $\mathcal{O}(M)$ operations, we speak about s^***-computability**.

Consider repr. of $A : H_0^m(\Omega) \to H^{-m}(\Omega)$ defined by

$$(Au)(v) = \sum_{|\alpha|,|\beta| \leq m} \int_\Omega a_{\alpha\beta} \partial^\alpha u \partial^\beta v \quad (u, v \in H_0^m(\Omega)),$$

with $a_{\alpha\beta}$'s *sufficiently smooth*, w.r.t.

$$\Psi = \{\psi_\lambda : \lambda \in \nabla\} \subset H_0^m(\Omega),$$

normalized w.r.t. $L_2(\Omega)$, i.e.,

$$\mathbf{A} := \left[2^{-(|\mu|+|\lambda|)m} (A\psi_\mu)(\psi_\lambda)\right]_{\lambda,\mu \in \nabla}.$$

Suff. to consider *one* non-zero $a_{\alpha\beta}$ for $|\alpha| = |\beta| = m$.

Wavelet assumptions

Wavelets are **local**, and **piecewise smooth** w.r.t. nested, uniformly shape regular partitions. I.e., $\exists(\{\Omega_j^{(\nu)} : \nu \in \mathcal{O}_j\})_j$, $\overline{\Omega} = \cup_{\nu\in\mathcal{O}_j}\overline{\Omega}_j^{(\nu)}$, $\operatorname{diam}(\Omega_j^{(\nu)}) \eqsim 2^{-j}$, s.t. $\operatorname{supp}\psi_\lambda$ included in the the union of a unif. bounded number of $\bar{\Omega}_{|\lambda|}^{(\nu)}$, and

$$\sup_{x\in\Omega_{|\lambda|}^{(\nu)}} |\partial^\gamma \psi_\lambda(x)| \lesssim 2^{|\lambda|(\frac{n}{2}+|\gamma|)} \quad (\gamma \in I\!N_0^n).$$

Global smoothness: For some $I\!N_0 \cup \{-1\} \ni r \geq m-1$,

$$\|\psi_\lambda\|_{W_\infty^t(\Omega)} \lesssim 2^{|\lambda|(\frac{n}{2}+t)} \quad (t \in [0, r+1]).$$

Cancellation properties of order $\tilde{d} \in I\!N_0$,

$$\left|\int_\Omega u\psi_\lambda\right| \lesssim 2^{-|\lambda|\tilde{d}}\|u\|_{W_\infty^{\tilde{d}}(\operatorname{supp}\psi_\lambda)}\|\psi_\lambda\|_{L_1(\Omega)}.$$

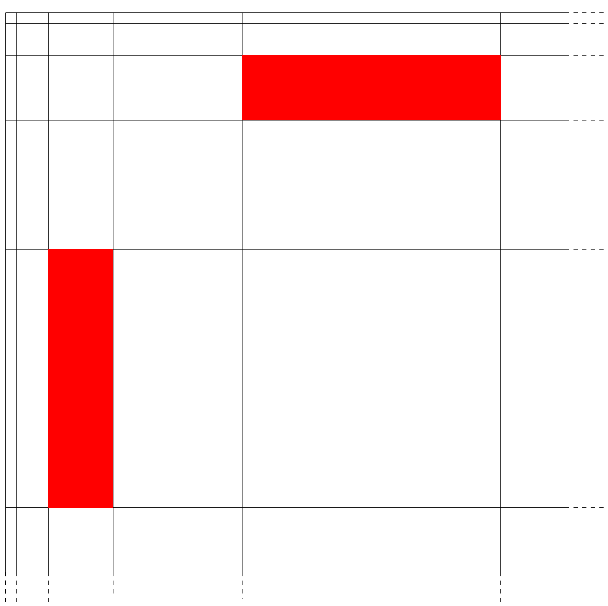
Regular and singular entries

We split

$$\mathbf{A} = \mathbf{A}^{(r)} + \mathbf{A}^{(s)},$$

where $\mathbf{A}^{(r)}_{\lambda,\mu} = \mathbf{A}_{\lambda,\mu}$ when either $|\lambda| > |\mu|$ and $\operatorname{supp}\psi_\lambda \subset \bar{\Omega}^{(\nu)}_{|\mu|}$ for some $\nu \in \mathcal{O}_{|\mu|}$ or $|\lambda| < |\mu|$ and $\operatorname{supp}\psi_\mu \subset \bar{\Omega}^{(\nu)}_{|\lambda|}$ for some $\nu \in \mathcal{O}_{|\lambda|}$, and $\mathbf{A}^{(r)}_{\lambda,\mu}$ is zero otherwise.

We write $\mathbf{A}^{(r)} = (\mathbf{A}^{(r)}_{j,\ell})_{j,\ell \in \mathbb{N}_0}$, where $\mathbf{A}^{(r)}_{j,\ell} = (\mathbf{A}^{(r)}_{\lambda,\mu})_{|\lambda|=j,|\mu|=\ell}$ and similarly $\mathbf{A}^{(s)} = (\mathbf{A}^{(s)}_{j,\ell})_{j,\ell \in \mathbb{N}_0}$.



non-zeros in each row of $\mathbf{A}^{(r)}_{j,\ell}$ or $\mathbf{A}^{(s)}_{j,\ell}$, or column of $\mathbf{A}^{(r)}_{\ell,j}$ or $\mathbf{A}^{(s)}_{\ell,j}$, is bounded by an absolute multiple of $2^{\max(\ell-j,0)n}$ or $2^{\max(\ell-j,0)(n-1)}$.

So one can effort to keep more blocks of $\mathbf{A}^{(s)}$, which is needed since the singular entries are larger.

Construction of sparse approximations

Prop 1. *Let* $\mathbf{C} = (\mathbf{C}_{j,\ell})_{j,\ell \in I\!N_0}$ *with* $\mathbf{C}_{j,\ell} = (\mathbf{C}_{\lambda,\mu})_{|\lambda|=j,|\mu|=\ell}$ *s.t. for some* $q \in I\!N$*, #non-zeros per row of* $\mathbf{C}_{j,\ell}$*, or column of* $\mathbf{C}_{\ell,j}$*, is* $\mathcal{O}(2^{\max(\ell-j,0)q})$*, and for some* $\rho > 0$*,*

$$|\mathbf{C}_{\lambda,\mu}| \lesssim 2^{-\left||\lambda|-|\mu|\right|(\frac{q}{2}+\rho)}.$$

Construct $\mathbf{C}^{(k)}$ *by dropping* $\mathbf{C}_{j,\ell}$ *when* $|j-\ell| > k/\rho$*. Then*

$$\|\mathbf{C} - \mathbf{C}^{(k)}\| \lesssim 2^{-k},$$

and #non-zeros per row and column of $\mathbf{C}^{(k)}$ *is* $\mathcal{O}(2^{qk/\rho})$*.* **(ρ/q-compressible)**

Let, for some $\xi > \rho/q$*,* $\forall \lambda, \mu \in \nabla$*,* $\exists \mathbf{C}^{(N)}_{\lambda,\mu}$*, computable in* $\mathcal{O}(N)$*, with*

$$|\mathbf{C}_{\lambda,\mu} - \mathbf{C}^{(N)}_{\lambda,\mu}| \lesssim N^{-\rho/q} 2^{-\left||\lambda|-|\mu|\right|(\frac{q}{2}+\xi q)}.$$

For some $\sigma \in (1, \xi q/\rho)$*, construct* $\tilde{\mathbf{C}}^{(k)}$ *by approximating each non-zero entry* $\mathbf{C}_{\lambda,\mu}$ *of* $\mathbf{C}^{(k)}$ *by taking*

$$N_{k,\lambda,\mu} \eqsim \max\left(1, 2^{qk/\rho - \left||\lambda|-|\mu|\right|\sigma q}\right).$$

Then each row or column of $\tilde{\mathbf{C}}^{(k)}$ *is computed in* $\mathcal{O}(2^{qk/\rho})$*, and*

$$\|\mathbf{C}^{(k)} - \tilde{\mathbf{C}}^{(k)}\| \lesssim 2^{-k}. \quad \textbf{(}\rho/q\textbf{-computable)}$$

Proof. First statement:

$$\|\mathbf{C}_{j,\ell}\|^2 \leq \max_{|\lambda|=j} \sum_{|\mu|=\ell} |\mathbf{C}_{\lambda,\mu}| \cdot \max_{|\mu|=\ell} \sum_{|\lambda|=j} |\mathbf{C}_{\lambda,\mu}| \lesssim 4^{-(j-m)\rho},$$

$$\|\mathbf{C} - \mathbf{C}^{(k)}\|^2 \leq \max_{j} \sum_{\{\ell:|j-\ell|>k/\rho\}} \|\mathbf{C}_{j,\ell}\| \cdot \max_{\ell} \sum_{\{j:|j-\ell|>k/\rho\}} \|\mathbf{C}_{j,\ell}\| \lesssim 4^{-k}.$$

Proof second statement similar. $\square$

Verification of the decay and quadrature assumptions

Prop 2. *With* $\rho_r := \tilde{d} + m$, $\rho_s := \frac{1}{2} + \min(\tilde{d} + m, r + 1 - m)$,

$$|\mathbf{A}^{(r)}_{\lambda,\mu}| \lesssim 2^{-(\left||\lambda|-|\mu|\right|)(\frac{n}{2}+\rho_r)}, \quad |\mathbf{A}^{(s)}_{\lambda,\mu}| \lesssim 2^{-(\left||\lambda|-|\mu|\right|)(\frac{n-1}{2}+\rho_s)}.$$

Quadrature assumption can be verified by the application of composite quadrature of fixed rank, assuming the wavelets are piecewise polynomials.

Corol 1. $\mathbf{A}^{(r)}$ *is* ρ_r/n*-computable, and* $\mathbf{A}^{(r)}$ *is* $\rho_s/(n-1)$*-computable.*

Now let d be the **order** of the wavelets. Then $s_{\max} = \frac{d-m}{n}$.

So $s^* > s_{\max}$ when $\tilde{d} > d - 2m$ and $\frac{r+\frac{3}{2}-m}{n-1} > \frac{d-m}{n}$.
For $r = d - 2$ (spline wavelets), valid when $\tilde{d} > d - 2m$ and $\frac{d-m}{n} > \frac{1}{2}$.

Proof proposition

When $r+1 \leq 2m$, select a $\gamma \leq \beta$ with $|\alpha+\gamma| = r+1$ and so $|\beta-\gamma| = 2m-(r+1)$. Using integration by parts, $\mathrm{vol}(\operatorname{supp}\psi_\lambda) \lesssim 2^{-|\lambda|n}$ and smoothness wavelets show that for $|\mu| \leq |\lambda|$,

$$
\begin{aligned}
|\mathbf{A}_{\lambda,\mu}| &= 2^{-(|\mu|+|\lambda|)m} \Big| \int_{\operatorname{supp}\psi_\lambda} (-1)^{|\gamma|} \partial^\gamma (a_{\alpha\beta} \partial^\alpha \psi_\mu) \partial^{\beta-\gamma} \psi_\lambda \Big| \\
&\lesssim 2^{-(|\mu|+|\lambda|)m} \|a_{\alpha\beta}\|_{W_\infty^{r+1-|\alpha|}(\Omega)} 2^{-|\lambda|n} 2^{|\mu|(\frac{n}{2}+r+1)} 2^{|\lambda|(\frac{n}{2}+2m-(r+1))} \\
&\lesssim 2^{-(|\lambda|-|\mu|)(\frac{n}{2}+r+1-m)}.
\end{aligned}
$$

For $r+1 > 2m$ by additionally using that the ψ_λ have $\tilde{d}$ vanishing moments,

$$
\begin{aligned}
|\mathbf{A}_{\lambda,\mu}| &= 2^{-(|\mu|+|\lambda|)m} \Big| \int_{\operatorname{supp}\psi_\lambda} (-1)^m \partial^\beta (a_{\alpha\beta} \partial^\alpha \psi_\mu) \psi_\lambda \Big| \\
&\lesssim 2^{-(|\mu|+|\lambda|)m} 2^{-|\lambda| \min(\tilde{d}, r+1-2m)} \\
&\qquad \times \|\partial^\beta (a_{\alpha\beta} \partial^\alpha \psi_\mu)\|_{W_\infty^{\min(\tilde{d}, r+1-2m)}(\operatorname{supp}\psi_\lambda)} \|\psi_\lambda\|_{L_1(\Omega)} \\
&\lesssim 2^{-(|\lambda|-|\mu|)(\frac{n}{2}+\min(\tilde{d}+m, r+1-m))},
\end{aligned}
$$

which completes the proof for the singular entries.
Second estimate can always be used for the regular entries.

Singular integral operators

s^*-computability with $s^* > s_{\max} = \frac{d-m}{n}$, and so opt. comput. compl. of `awgm`, has also been demonstrated for classes of singular integral operators.

$\tilde{d} > d - 2m$ needed, so orthogonal wavelets aren't suited for $m < 0$.

Trees

Let $A \in \mathcal{B}(H, K')$ b.i., say with $H = K$.
Let Ψ be a wavelet Riesz basis for H. Then

$$\mathbf{A} = (A\Psi)(\Psi) = a(\Psi, \Psi) \quad (a(u,v) := (Au)(v)).$$

For $\Lambda \subset \nabla$, $\mathbf{A}_\Lambda = \mathbf{P}_\Lambda \mathbf{A} \mathbf{I}_\Lambda$ generally *not* sparse, even when A is local.

If $\Lambda = \{\psi_\lambda : |\lambda| \leq j\}$, then $\exists \mathbf{T}_j$, with $\Psi|_\Lambda^\top = \Phi_j^\top \mathbf{T}_j$, and so

$$\mathbf{A}_\Lambda = a(\Psi|_\Lambda, \Psi|_\Lambda) = a(\mathbf{T}_j^\top \Phi_j, \mathbf{T}_j^\top \Phi_j) = \mathbf{T}_j^\top a(\Phi_j, \Phi_j) \mathbf{T}_j.$$

All three matrices can be applied exactly in $\mathcal{O}(\#\Lambda)$ operations.

Doesn't work for general $\Lambda \subset \nabla$, but it does for trees:

Def 1. $\Lambda \subset \nabla$ is called a **tree** when for $\lambda \in \Lambda$ with $|\lambda| > 0$, $\operatorname{supp} \psi_\lambda$ is covered by $\operatorname{supp} \psi_\mu$ for some $\mu \in \Lambda$ with $|\mu| = |\lambda| - 1$.

For Λ a tree, $\exists \Phi = \Phi(\Lambda) \subset \cup_j \Phi_j$ which is unif. locally finite (i.e. the support of any atom intersects the supports of at most a unif. bounded number of others), s.t. $\operatorname{span} \Psi|_\Lambda \subset \operatorname{span} \Phi$ and $\# \Phi \lesssim \# \Lambda$.

$\mathbf{T} = \mathbf{T}(\Lambda)$ (generally non-square) with $\Psi|_\Lambda^\top = \Phi^\top \mathbf{T}$ can be applied in $\mathcal{O}(\#\Lambda)$ ops, and so can

$$\mathbf{A}_\Lambda = a(\mathbf{T}^\top \Phi, \mathbf{T}^\top \Phi) = \mathbf{T}^\top a(\Phi, \Phi) \mathbf{T}.$$

Immediate consequence for solving the Galerkin systems inside `awgm` when the subsets Λ are *restricted to trees*.
The classes $\mathcal{A}^s$ become smaller than with unconstrained approximation, but only "slightly".

Now the approximate computation of the infinite residuals (`apply`-routine):

For $\mathbf{w} \in \ell_0(\nabla)$, recall approx $\mathbf{z}_j$ of $\mathbf{Aw}$ defined by

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

where $\mathbf{w}_{[j]}$ is a (near) best 2^j approx. Replace it by a (near) best 2^j-**tree** approx. Then for $s < s^*$, $\|\mathbf{Aw} - \mathbf{z}_j\| \lesssim 2^{-sj}\|\mathbf{w}\|_{\mathcal{A}^s_{\text{tree}}}$ and $\operatorname{supp} \mathbf{z}_j \lesssim 2^j$.

Moreover, $\Lambda := \operatorname{supp} \mathbf{z}_j$ is a tree, and one deduces that also

$$\|\mathbf{Aw} - (\mathbf{Aw})|_\Lambda\| \lesssim 2^{-sj}\|\mathbf{w}\|_{\mathcal{A}^s_{\text{tree}}}.$$

So, alternatively, apply $\mathbf{A}|_{\Lambda \times \Lambda}$ using multiscale transforms, and the application of the stiffness w.r.t. locally single-scale coordinates. Advantages: more efficient, and easier quadrature.

[Under some conditions, it is sufficient to approximate $\mathbf{f} - \mathbf{Aw}$ by $(\mathbf{f} - \mathbf{Aw})|_\Lambda$, where Λ is constructed from the tree $\operatorname{supp} \mathbf{w}$ by adding a fixed number of "layers" (in practice one layer suffices). W.I.P.]

Summary

With a suitable wavelet Riesz basis, PDO in wavelet coordinates is sufficiently close to computable sparse matrices so that the `awgm` has opt. comput. compl.

Same is true for classes of singular integral operators.

The restriction to tree approximation often results in quantitatively more efficient implementations (for nonlinear operators it seems indispensable).

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] T. Gantumur and R.P. Stevenson. Computation of differential operators in wavelet coordinates. *Math. Comp.*, 75:697–709, 2006.

[3] T. Gantumur and R.P. Stevenson. Computation of singular integral operators in wavelet coordinates. *Computing*, 76:77–107, 2006.

[4] R. Schneider. *Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur Lösung großer vollbesetzter Gleigungssysteme.* Habilitationsschrift, 1995. Advances in Numerical Mathematics. Teubner, Stuttgart, 1998.

[5] R.P. Stevenson. On the compressibility of operators in wavelet coordinates. *SIAM J. Math. Anal.*, 35(5):1110–1132, 2004.

[6] R.P. Stevenson and M. Werner. Computation of differential operators in aggregated wavelet frame coordinates. *IMA J. Numer. Anal.*, 28(2):354–381, 2008.