Improving AMG interpolation through energy minimization

Due Giorni di Algebra Lineare Numerica e Applicazioni
14-15 Febbraio 2022, Centro Congressi Federico II, via Partenope, Napoli, Italy

Carlo Janna, Andrea Franceschini and Giovanni Isotton
carlo.janna@unipd.it

February 15, 2022
Outline

1 Motivation
2 Energy minimization
3 Solution algorithms
4 Numerical results
5 Conclusions
AMG is a very popular solution algorithm which is used in several applications:

- it is very flexible as it only requires little information beyond the matrix itself;
- it can be parallelized up to millions of cores;
- as preconditioner for a Krylov method often guarantees performance independent of the grid size;

AMG works almost perfectly for Poisson problems, however, it is less effective in more difficult problems such as those arising in structural mechanics:

- classical or extended interpolation usually give poor results;
- aggregation-based AMG is generally preferred over Classical AMG;
- the near kernel space is larger than the standard constant vector (which is fine for Poisson);
Motivation

In the context of structural problems (or other difficult problems):

- more powerful smoothers than Jacobi or Gauss-Seidel are needed;
- the near kernel is constructed starting from the Rigid Body Modes (RBM);
- it is often necessary to improve the near kernel through smoothing or an eigensolution;
- prolongation is constructed through a least square fit of the test space;

Main issues:

- coarse basis functions with local support are not orthogonal to high frequency modes;
- the prolongation has to be smoothed to allow for fast convergence (helpful in classical AMG as well);
- smoothed prolongation however leads to high complexities;
Energy minimization in prolongation set-up

The main idea is, for a given prolongation pattern $\mathcal{P}$, to compute:

$$P = \arg\min_{P \in \mathcal{P}} \left( \text{tr}(P^T AP) \right)$$

or, equivalently, to minimize the energy of every prolongation column:

$$p_i^T A p_i \rightarrow \min \quad \forall i \in C$$

We consider a **Classical** AMG framework, so that:

$$A = \begin{bmatrix} A_{ff} & A_{fc} \\ A_{fc}^T & A_{cc} \end{bmatrix}, \quad P = \begin{bmatrix} W \\ I \end{bmatrix}, \quad R = \begin{bmatrix} 0 & I \end{bmatrix}, \quad S = \begin{bmatrix} I \\ 0 \end{bmatrix}$$


Why Energy minimization is important

Suppose we have a C/F partition of the unknowns and an SPD $X$ matrix which is spectrally equivalent to the symmetrized smoother $\tilde{M} = M^T(M + M^T - A)^{-1}M$.

Defining $A_s = S^T AS$ and $X_s = S^T X S$ and:

$$k_S \leq \lambda_{\text{min}}(X_s^{-1} A_s) \leq \lambda_{\text{max}}(X_s^{-1} A_s) \leq c_2$$

then, if $\|PR\|_A$ is bounded:

$$K_{TG} \leq \frac{c_2}{k_S} \|PR\|_A^2$$

where $K_{TG}$ is such that $\|E_{TG}\|_A \leq 1 - \frac{1}{K_{TG}}$

In practice, the smaller $\|PR\|_A$, the faster convergence!

The minimum in energy can be simply found by deriving with respect to $W$ and setting to zero:

$$\frac{\partial}{\partial W} \left( \text{tr}(P^T AP) \right) = 0 \Rightarrow A_f f W = -A_f c$$

which is exactly what is defined ideal prolongation $W_{id} = -A_{ff}^{-1}A_{fc}$.

Note that it is also the target that is sought with a 2x2 block FSAI approximation:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix}, \quad G = \begin{bmatrix} I & 0 \\ F & I \end{bmatrix}, \quad GAG^T \simeq I \quad \Rightarrow \quad F = -A_{11}A_{12}$$

Adding constraints to energy minimization

First of all, a dense prolongation is impractical, so we enforce a given non-zero pattern. We consider the pattern $\overline{P}_0$ of a tentative prolongation $P_0$, and we extend it using the strength of connection matrix $S$:

$$\overline{P} = S^k \overline{P}_0$$

for a small power $k$.

Then we minimize the energy of every column of $P$ restricted to the non-zero pattern $\overline{P}$ by solving the following sequence of $n_c$ dense linear systems:

$$A(I_i, I_i)\tilde{p}_i = -A(I_i, i) \quad \forall i \in C$$

with $I_i$ the set of indices of the non-zeroes in the $i$-th column of $\overline{P}$. 
Another important condition to satisfy is to ensure that the prolongation is able to accurately represent some prototypes of the near kernel of $A$. By denoting with $V$ the matrix collecting these prototypes, this condition reads:

$$V \subseteq \text{range}(P), \quad \text{with} \quad V = \begin{bmatrix} V_f \\ V_c \end{bmatrix}$$

This condition becomes:

$$W V_c = V_f \quad \Rightarrow \quad V_c(J_i,:)^T \tilde{w}_i = v_i \quad \forall i \in F$$

where $J_i$ is the set of column indices of the prescribed nonzeros of the $i$-th row of $W$. Note that this is another sequence of $n_f$ small and dense, generally underdetermined, linear systems.
The constrained minimization problem

By collecting all the non-zero entries of $W$ into the vector $w$ and using Lagrange multipliers, the problem can be stated as the solution of the following saddle-point system:

$$
\begin{bmatrix}
K & B \\
B^T & 0
\end{bmatrix}
\begin{bmatrix}
w \\
\lambda
\end{bmatrix}
=
\begin{bmatrix}
f \\
g
\end{bmatrix}
$$

Observations:

- The size of $K$ is typically much larger than the size of $A$ (about 20 times larger in 3D mechanics);
- With a proper unknown ordering both $K$ and $B$ are block diagonal matrices;
- If $w$ is numbered following the columns of $P$ then $K$ is block diagonal;
- If $w$ is numbered following the rows of $P$ then $B$ is block diagonal;
- Unfortunately, they cannot be block diagonal at the same time;
As in the original paper, we solve the minimization problem through a **restricted conjugate gradient**, which is a nullspace method. Define the projection orthogonal to $B$, $\Pi_B = I - B(B^T B)^{-1}B^T$, so that:

\[ B^T y = 0 \quad \forall y = \Pi_B v \]

Then, starting from a tentative prolongation $W_0$ already satisfying the constraint, $B^T w_0 = g$, we apply the **conjugate gradient** method to the system:

\[ \Pi_B K \Pi_B \Delta w = \Pi_B (f - Kw_0) \]

to find a correction $\Delta W$ to be applied to $W_0$:

\[ W = W_0 + \Delta W \]
The solution algorithm

During CG iteration, it is possible to monitor the decrease in energy and stop when a plateau is reached.

Some issues related to the proposed solution method:

- The energy minimization is just a piece of the AMG set-up, so we cannot spend too much effort in it ⇒ it must converge quickly!
- Memory occupation is a big issue. $K$ and $B$ require a significant amount of memory and sometimes it is unavoidable to proceed in matrix-free mode;
- Preconditioning is crucial:
  - projected Jacobi, $\Pi_B D_K \Pi_B$, is enough for many problems;
  - incomplete Cholesky with very low fill-in converges faster but it is more difficult to apply, if $K$ cannot be stored;
Numerical results

**Linear elasticity Benchmark:**
Regular discretization of an elastic cube with tetrahedral FE.

<table>
<thead>
<tr>
<th>Matrix</th>
<th># of rows</th>
<th># of non-zeroes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_4820</td>
<td>14,460</td>
<td>556,686</td>
</tr>
<tr>
<td>C_35199</td>
<td>105,597</td>
<td>4,079,357</td>
</tr>
<tr>
<td>C_246389</td>
<td>739,167</td>
<td>29,610,351</td>
</tr>
<tr>
<td>C_1772489</td>
<td>5,317,443</td>
<td>222,268,213</td>
</tr>
</tbody>
</table>

We compare the following prolongation strategies:

- **EXTI** Extended+i (hypre);
- **BAMG** least squares fit;
- **SBAMG** BAMG + smoothing;
- **EMIN** energy minimization;
Numerical results

<table>
<thead>
<tr>
<th>Matrix</th>
<th>P-type</th>
<th>$C_{gd}/C_{op}$</th>
<th>n_{it}</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_{4820}</td>
<td>EXTI</td>
<td>1.062/1.200</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>BAMG</td>
<td>1.062/1.162</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>SBAMG</td>
<td>1.062/1.404</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>EMIN</td>
<td>1.062/1.383</td>
<td>12</td>
</tr>
<tr>
<td>C_{35199}</td>
<td>EXTI</td>
<td>1.068/1.254</td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>BAMG</td>
<td>1.067/1.211</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>SBAMG</td>
<td>1.066/1.576</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>EMIN</td>
<td>1.066/1.540</td>
<td>15</td>
</tr>
<tr>
<td>C_{246389}</td>
<td>EXTI</td>
<td>1.066/1.214</td>
<td>510</td>
</tr>
<tr>
<td></td>
<td>BAMG</td>
<td>1.064/1.196</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>SBAMG</td>
<td>1.064/1.608</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>EMIN</td>
<td>1.064/1.517</td>
<td>22</td>
</tr>
<tr>
<td>C_{1772481}</td>
<td>EXTI</td>
<td>1.064/1.188</td>
<td>1286</td>
</tr>
<tr>
<td></td>
<td>BAMG</td>
<td>1.062/1.189</td>
<td>270</td>
</tr>
<tr>
<td></td>
<td>SBAMG</td>
<td>1.062/1.590</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>EMIN</td>
<td>1.062/1.460</td>
<td>33</td>
</tr>
</tbody>
</table>

Figure: # of PCG iterations vs. grid size
Numerical results

Figure: Relative residual, $\frac{r_k}{r_0}$, vs. iteration count, $k$
Numerical results

<table>
<thead>
<tr>
<th>Matrix</th>
<th># of rows</th>
<th># of non-zeroes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pflow_{742k}</td>
<td>742,793</td>
<td>37,138,461</td>
</tr>
<tr>
<td>Gear_{167k}</td>
<td>167,460</td>
<td>9,861,437</td>
</tr>
<tr>
<td>S4A_{700k}</td>
<td>631,007</td>
<td>26,148,363</td>
</tr>
<tr>
<td>Mech_{447k}</td>
<td>447,703</td>
<td>18,243,793</td>
</tr>
</tbody>
</table>

Figure: # of PCG iterations.

Figure: Total solution time [s].
Conclusions

- Energy minimizing prolongation is a well-known concept in AMG, which, however has little application due to its cost and its difficult implementation;
- Our main contribution was to make energy minimization feasible by means of preconditioning and an effective implementation;
- Our algorithm has proved very effective on linear systems arising from mechanical problems and also in challenging problems arising from different applications.
- Though requiring a slightly more expensive set-up, this additional cost is largely offset by a fast convergence.

Further work:

- Further improve preconditioning of restricted CG;
- Include energy minimization in the HPC AMG solver Chronos.

Thanks for your attention.

Questions?