Adaptive Composite AMG Solvers

with Graph Modularity Coarsening

Austen Nelson [1] and Panayot S. Vassilevski[1]

[1]Portland State University, ajn6@pdx.edu, panayot@pdx.edu

MTH 501 Research Project

April 12, 2024

Overview

Problem Statement

Want to solve the linear matrix system:

$$A\boldsymbol{x} = b$$

Where A is symmetric positive definite (s.p.d.).

Often resulting from discretizations of elliptic PDEs:

$$\begin{cases} -\Delta u = f & \text{, in } \Omega \\ u = 0 & \text{, on } \partial\Omega \end{cases}$$

or with a diffusion coefficient

$$\begin{cases} -\nabla \cdot (\beta \nabla u) = f & \text{, in } \Omega \\ u = 0 & \text{, on } \partial\Omega \end{cases}$$

Stationary Iteration Algorithm

Data: Matrix A, method matrix B, vector b, initial guess x, convergence tolerance ε, maximum iterations *max_iter*

Result: Approximate solution to $Ax = b$

$r \leftarrow b - Ax$ `// Initial residual`

$r_{norm} \leftarrow \|r\|$

$i \leftarrow 0$

while $i < max\_iter$ **do**

  $r \leftarrow b - Ax$ `// Current residual`

  if $\|r\|/r_{norm} < \varepsilon$ **then**

    return x `// Convergence achieved`

  $x \leftarrow x + B^{-1}r$ `// Update step`

  $i \leftarrow i + 1$

return x `// Max iterations reached`

Stationary Iteration Analysis

$$
\begin{aligned}
x_{i+1} &= x_i + B^{-1} r_i \\
&= x_i + B^{-1}(b - Ax_i) \\
&= B^{-1} b + (I - B^{-1} A) x_i
\end{aligned}
$$

We call $E := I - B^{-1}A$ the iteration matrix.

This functional iteration has a closed form of:

$$x_i = E^i x_0 + C(b)$$

Stationary Iteration Analysis

$$x_{i+1} = Ex_i + B^{-1}b$$

The solution x is a fixed point of this functional iteration

$$x = Ex + B^{-1}b$$

Subtracting these two equations gives that,

$$e_{i+1} = Ee_i$$

hence,

$$e_m = E^m e_0$$

Choosing a vector norm and its induced matrix norm,

$$\|e_m\| \leq \|E\|^m \|e_0\|$$

Stationary Iteration Analysis

$$\|e_m\| \leq \|E\|^m \|e_0\|$$

Choosing the L^2 vector norm gives the spectral radius of E,

$$\|E\|_2 = \max |\lambda(E)|$$

which clearly must be less than 1 for the method to be convergent. In the context of iterative methods this is called the **Asymptotic Convergence Factor**.

$$-\log_{10} \|E\|_2$$

is called the **Asymptotic Convergence Rate** and its recipricol is the maximum number of iterations to reduce the error by an order of magnitude.

Simple Example (1d centered finite difference)

Consider $\Omega = (0,1)$ and

$$\begin{cases} -u'' = f & \text{in } \Omega \\ u(0) = u(1) = 0 \end{cases}$$

The classic centered finite difference discretization (n elements of length h) yields the familiar $n-1 \times n-1$ matrix system $Ax = b$:

$$A = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}, \; x = u_h, \; b = h^2 f_h$$

Example adapted from [BHM00]

Simple Example (weighted Jacobi method)

Let D be the diagonal of A. Choose

$$B^{-1} := \frac{2}{3}D^{-1} = \frac{1}{3}$$

as the method matrix. The resulting iteration matrix is

$$E = \left(I - \frac{1}{3}A\right)$$

The kth eigenvalue of E is

$$\lambda_k(E) = 1 - \frac{4}{3}\sin^2\left(\frac{k\pi}{2n}\right), \quad 1 \leq k \leq n-1$$

and the jth component of the associated eigenvector is

$$Q_{j,k} = \sin(x_j k\pi)$$

Notice that as $h \to 0$ we get $\|E\|_2 \to 1$

Simple Example (spectral / convergence analysis)

$$\lambda_k(E) = 1 - \frac{4}{3}\sin^2\left(\frac{k\pi}{2n}\right), \quad 1 \le k \le n-1$$

$$Q_{j,k} = \sin(x_j k\pi)$$

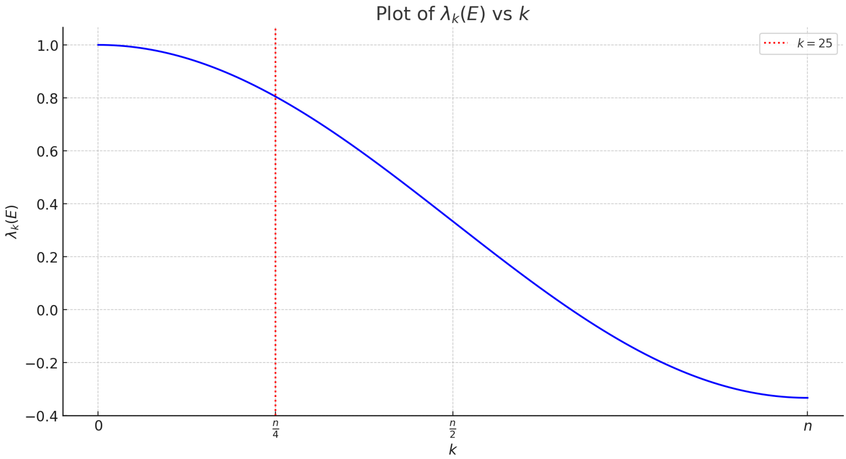
Let w_k be the kth eigenvector (column of Q).

$$e_0 = \sum_{k=1}^{n-1} c_k w_k$$

$$e_m = E^m e_0 = \sum_{k=1}^{n-1} c_k E^m w_k = \sum_{k=1}^{n-1} c_k \lambda_k^m w_k$$

The kth mode of e_0 is reduced by a factor of λ_k^m after m steps

Simple Example (spectrum visualization)



The kth eigenvector is the discretization of $\sin(k\pi)$

Simple Example (geometric multigrid solution)

Discretize with small h and iterate weighted Jacobi i times.

$$
\begin{aligned}
r_i &= b - Ax_i \\
A^{-1} r_i &= A^{-1} b - x_i \\
&= e_i \approx \sum_{k=1}^{n/4} c_k w_k \\
r_i &\approx \sum_{k=1}^{n/4} c_k \lambda_k(A) w_k
\end{aligned}
$$

Main ideas of geometric multigrid:

- If we solve $Ae_i = r_i$, then $b = x_i + e_i$
- r_i and e_i are linear combinations of **smooth eigenvectors**
- Smooth eigenvectors are accurately represented on coarse grids

Anatomy of Multigrid

A multigrid method with ℓ levels has some basic components:

- Hierarchy of vector spaces, operators, and solvers

$$\{V_i\}_{i=1}^{\ell}, \quad \{A_i\}_{i=1}^{\ell}, \quad \{B_i\}_{i=1}^{\ell}$$

- Interpolation (or prolongation) Operators

$$\{P_i\}_{i=1}^{\ell-1}, \quad P_i : V_{i+1} \to V_i$$

- Restriction Operators

$$\{R_i\}_{i=1}^{\ell-1}, \quad R_i : V_i \to V_{i+1}$$

In our case, all operators are matrices:

- $R_i := P_i^T$
- $A_i : V_i \to V_i$
- $A_{i+1} := P_i^T A_i P_i$

V—Cycle Algorithm (recursive definition)

Initial call: $x_{i+1} \leftarrow V(x_i, b, 1)$

Data: Levels ℓ, hierarchy $A = \{A_i\}_{i=1}^{\ell}$, smoothers $B = \{B_i\}_{i=1}^{\ell}$, interpolation operators $P = \{P_i\}_{i=1}^{\ell-1}$, current iterate x, rhs vector b, smoothing steps s, current level k

Result: Next Iterate (or update) $x_{new} \leftarrow V(x, b, k)$

if $k \neq \ell$ **then**
- Relax for s iterations on $A_k x = b$ with B_k^{-1} (stationary algorithm)
- $r \leftarrow b - A_k x$
- $r_c = P_k^T r$
- $k \leftarrow k + 1$
- $c \leftarrow V(\mathbf{0}, r_c, k)$
- $x \leftarrow x + P_k c$

Relax for s iterations on $A_k x = b$ with B_k^{-1}

return x

Algebraic vs Geometric Multigrid

Geometric Multigrid

- Requires a hierarchy of refinements (h or p)
- Interpolation and restriction operators come from this hierarchy
- For 'nice' problems iteration scaling is $\mathcal{O}(1)$
- Analysis is fairly simple and well understood

Algebraic Multigrid

- No knowledge of problem structure/nature required
  - can be utilized for heuristics
- 'Black Box' for the end user with varying levels of tuning
- 'Algebraically' finds R and P matrices from A
- Solver construction and application can be expensive
- General analysis is difficult

Anisotropy

Let $\Omega \subset \mathbb{R}^3$.

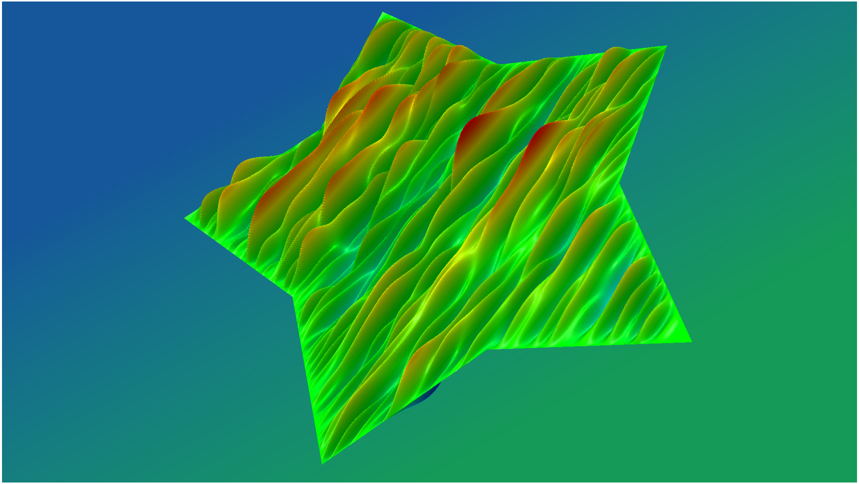
$$\begin{cases} -\nabla \cdot (\beta \nabla u) = f & , \text{ on } \Omega \\ u = 0 & , \text{ on } \partial\Omega \end{cases}$$

$$\beta := \varepsilon I + \boldsymbol{b}\boldsymbol{b}^T$$

for small $\varepsilon > 0$ and

$$\boldsymbol{b} := \begin{bmatrix} \cos\theta\cos\phi \\ \sin\theta\cos\phi \\ \sin\phi \end{bmatrix}$$

Anisotropy — Algebraic Smoothness

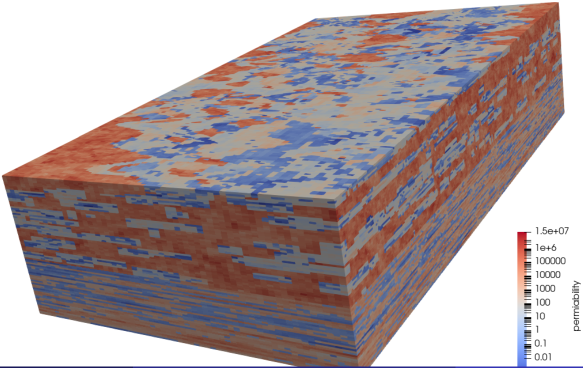


Heterogeneous Coefficients (SPE10)

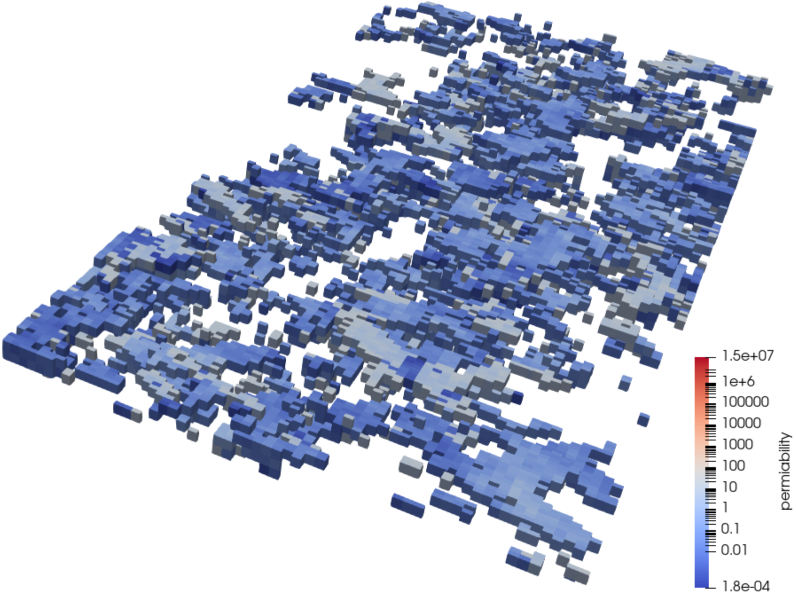
$$\begin{cases} -\nabla \cdot (\beta \nabla u) = f & , \text{ on } \Omega \\ u = 0 & , \text{ on } \partial\Omega \end{cases}$$

In this case, β (called the permiability) is a piecewise constant diagonal matrix coefficient (constant on each element).

$\|\beta\|_2$ **on each element**



SPE10 Clipped Cross Section (High Permeability)



Adaptivity Algorithm

Data: Matrix A, desired convergence factor ρ, max components m, smoother type B

Result: Adaptive Solver $\overline{B}$

$\overline{B} \leftarrow \text{CreateSolver}(B, A)$
$i, cf \leftarrow 1$
while $\rho < cf$ *and* $i < m$ **do**
    $w, cf \leftarrow \text{TestHomogeneous}(A, \overline{B})$
    $w = w / \|w\|_2$
    $B_{new} \leftarrow \text{AdaptiveMLSolver}(B, A, w)$
    $\overline{B} \leftarrow \text{SymmetricComposition}(\overline{B}, B_{new})$
    $i \leftarrow i + 1$

Composition of Solvers

$$I - B^{-1}A = (I - B_1^{-T}A)(I - B_0^{-1}A)(I - B_1^{-1}A). \qquad (1)$$

$$B^{-1} = \overline{B}_1^{-1} + (I - B_1^{-T}A)B_0^{-1}(I - AB_1^{-1}). \qquad (2)$$

$\overline{B}_1$ is a symmetrization of B_1 (if needed)

$$\overline{B}_1 = B_1(B_1 + B_1^T - A)^{-1}B_1^T. \qquad (3)$$

Lemma

If B_0 is s.p.d. and B_0 and B_1 are A-convergent solvers, then their composition defined in (1) or equivalently, in (2), is s.p.d. and B is also A-convergent. Also, if the symmetrized solver $\overline{B}_1$ (see (3)) satisfies $\|\overline{B}_1\| \leq c_0\|A\|$ for some constant $c_0 > 0$, then the same inequality holds for B, i.e., $\|B\| \leq c_0\|A\|$. Finally, if B_1 is s.p.d. and satisfies the inequalities $\mathbf{v}^T B_1 \mathbf{v} \geq \mathbf{v}^T A \mathbf{v}$ and $\|B_1\| \leq c_0\|A\|$, we have $\|B\| \leq \|\overline{B}_1\| \leq \|B_1\| \leq c_0\|A\|$.

Algebraically Smooth Error is Near-nullspace of A

$$A\mathbf{x} = 0, \text{ gives } \quad B(\mathbf{x}_k - \mathbf{x}_{k-1}) = -A\mathbf{x}_{k-1} \tag{4}$$

Theorem

Let B define an s.p.d. A-convergent iterative method such that $\frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T B\mathbf{v}} < 1$ and $\|B\| \simeq \|A\|$, i.e., $\|B\| \leq c_0\|A\|$ for a constant $c_0 \geq 1$. Consider any vector $\mathbf{w}$ such that the iteration process (4) with B stalls for it, i.e.,

$$1 \geq \frac{\|(I - B^{-1}A)\mathbf{w}\|_A^2}{\|\mathbf{w}\|_A^2} \geq 1 - \delta, \tag{5}$$

for some small $\delta \in (0,1)$. Then, the following estimate holds $\|A\mathbf{w}\|^2 \leq c_0\|A\| \; \delta\|\mathbf{w}\|_A^2$.

Strength of Connectivity Graph

Since $A\mathbf{w} \approx 0$ componentwise by construction, we have for each i

$$0 \approx w_i \sum_j a_{ij} w_j,$$

or equivalently

$$0 \leq a_{ii} w_i^2 \approx \sum_{j \neq i} (-w_i a_{ij} w_j).$$

Then, $\overline{A} = (\overline{a}_{ij})$ with non-zero entries $\overline{a}_{ij} = -w_i a_{ij} w_j$, $(i \neq j)$ has positive row-sums.

$\overline{A}$ is the sparse adjacency matrix associated with the connectivity strength graph G.

Modularity Matching (Coarsening) for AMG Hierarchy

Let $\mathbf{1} = (1) \in \mathbb{R}^n$ be the unity constant vector, $\mathbf{r} = A\mathbf{1}$, and $T = \sum_i r_i = \mathbf{1}^T A \mathbf{1}$.

The *Modularity Matrix* [New10]

$$B = A - \frac{1}{T}\mathbf{r}\mathbf{r}^T.$$

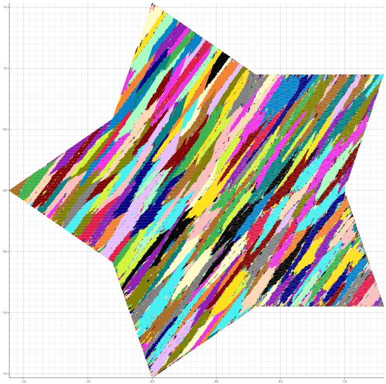
By construction, we have that

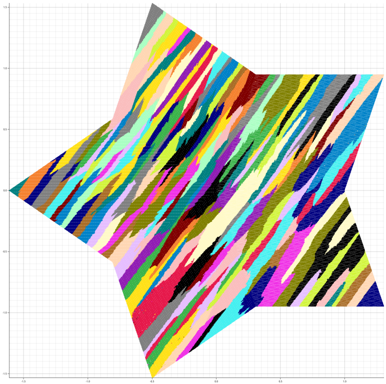
$$B\mathbf{1} = 0. \tag{6}$$

The *Modularity Functional* [QV19]

$$Q = \frac{1}{T}\sum_{\mathcal{A}}\sum_{i,\, j \in \mathcal{A}} b_{ij} = \frac{1}{T}\sum_{\mathcal{A}}\sum_{i,\, j \in \mathcal{A}} \left(a_{ij} - \frac{r_i r_j}{T}\right).$$

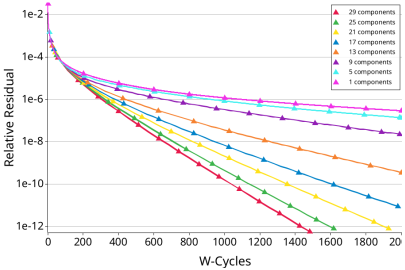
Hierarchy Visualization for 2d Anisotropy



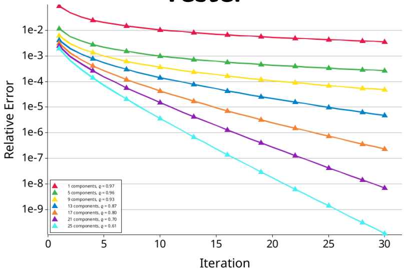


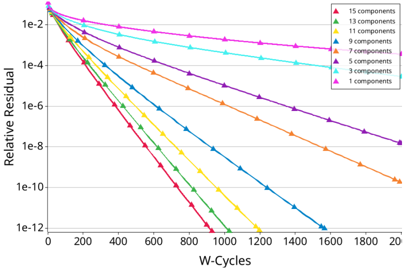
2d Anisotropy (top) and SPE10 (bottom)

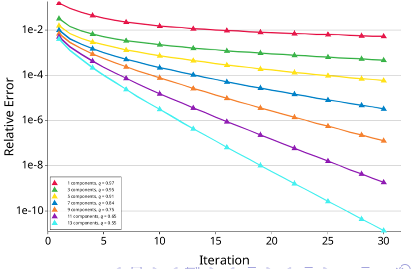
Stationary



Tester

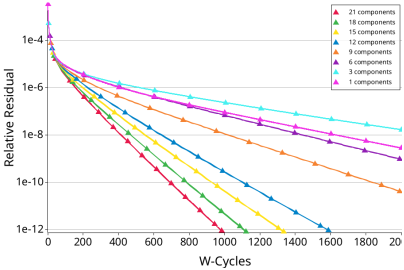


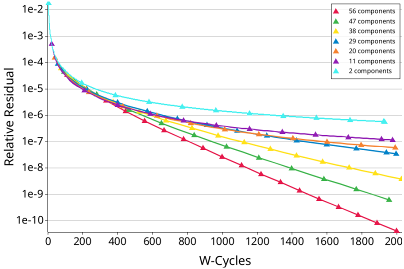




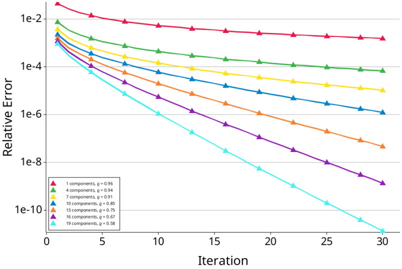
G3-circuit (top) and Janna-Flan (bottom)

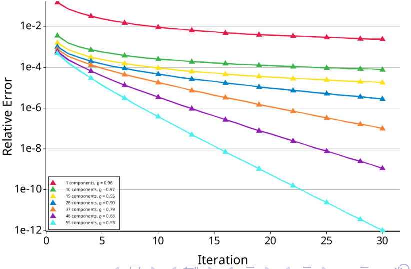
Stationary





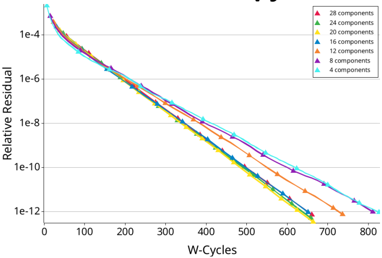
Tester



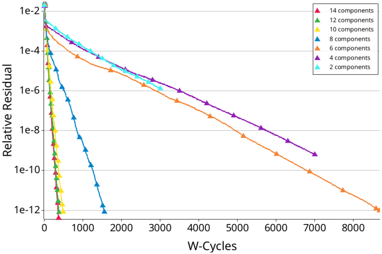


As Preconditioner for Conjugate Gradient

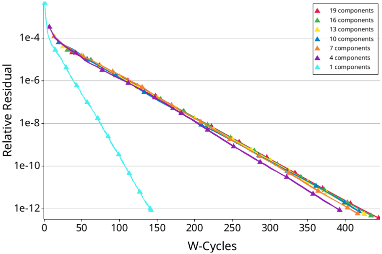
2d Anisotropy



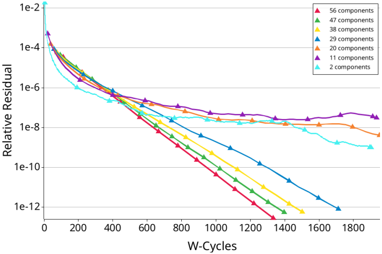
SPE10 (inverted coefficient)



G3-circuit



Janna-Flan



Future Work

- We suspect the interpolation technique is limiting the solver / PC performance
- Study the algorithmic and implementation scalability
- Study more advanced relaxation techniques
- Study applications to eigensolvers

Submitted work to a student paper competition (with presentation) for:

18th Copper Mountain Conference On Iterative Methods
Sunday April 14 - Friday April 19, 2024

References I

- William L. Briggs, Van Emden Henson, and Steve F. McCormick, *A multigrid tutorial, second edition*, second ed., Society for Industrial and Applied Mathematics, 2000.
- M.E.J. Newman, *Networks. an introduction*, Oxford University Press, New York, 2010.
- B.G. Quiring and P.S. Vassilevski, *Properties of the Graph Modularity Matrix and Its Applications*, Tech. report, LLNL-TR-779424, Lawrence Livermore National Laboratory, June 26, 2019.