Graph partitioning with matrix coefficients for symmetric positive definite linear systems

Eugene Vecharynski — Lawrence Berkeley National Laboratory
Yousef Saad — University of Minnesota
Masha Sosonkina — Old Dominion University

Parallel linear solvers

Consider a parallel (distributed memory) solution of a linear system

$$Ax = b.$$

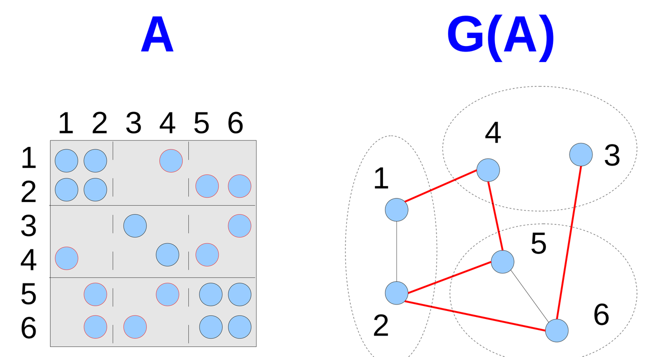
- A is very large and sparse with strongly varying coefficients.
- A is symmetric positive definite.
- Use a preconditioned Krylov subspace linear solver (PCG).

Major challenges:
- Minimizing the number of interprocessor communications $\Rightarrow$ increase parallelism.
- Constructing robust parallel preconditioners $\Rightarrow$ decrease iteration counts.

Partitioning

A performance-critical operation is the matrix-vector multiplication $y = Av$ (matvec). The associated communications are reduced by partitioning the linear system.

Standard goal: Partition vertices of the adjacency graph $G(A)$ into subsets V_k, so that
- The edge cut is minimized.
- Each subgraph has approximately the same number of vertices, $|V_i| \approx |V_j|$.



Available software packages: Chaco, JOSTLE, MeTiS, SCOTCH, etc

Partitioning and preconditioners

Parallel preconditioners:
- Strongly influence the performance of the overall computational scheme.
- Robustness depends on the given partition.
- Built upon partitions that minimize communications in matvecs. Partitions that maximize the quality of preconditioning may be different!

Our goal: replace the requirement on minimizing the number of communications by an objective of improving the preconditioning quality.
- Consider nonoverlapping Additive Schwarz (AS) preconditioning.
- Relate partitioning to convergence theory of PCG with block-diagonal preconditioners.

AS preconditioner

Input: A, r, $\{V_k\}_{k=1}^{s}$. Output: $w = T^{-1}r$.

1. For $k = 1, \ldots, s$, Do
2. Set $A_k := A(V_k, V_k)$, $r_k := r(V_k)$, and $w_k = \mathbf{0} \in \mathbb{R}^n$.
3. Solve $A_k \delta = r_k$.
4. Set $w_k(V_k) := \delta$.
5. EndDo
6. $w = w_1 + \ldots + w_s$.

Bipartitioning and Cauchy-Bunyakowski-Schwarz (CBS) constants

Let $V = I \cup J$, $I \cap J = \emptyset$, $|I| = |J| = n/2$. Reorder A and T with respect to this bipartition

$$C = PAP^T = \begin{pmatrix} A_I & A_{IJ} \\ A_{IJ}^* & A_J \end{pmatrix}, \qquad B = PTP^T = \begin{pmatrix} A_I & 0 \\ 0 & A_J \end{pmatrix}.$$

- Note that $\kappa(T^{-1}A) = \kappa(B^{-1}C)$.
- Use condition number bound based on the CBS constant:

$$\kappa(T^{-1}A) \le \frac{1+\gamma_{IJ}}{1-\gamma_{IJ}}, \quad 0 \le \gamma_{IJ} < 1, \text{ where}$$

$$\gamma_{IJ} = \max_{u \in W_I, v \in W_J} \frac{|(u, Av)|}{(u, Au)^{1/2}(v, Av)^{1/2}};$$

$$W_I = \{u \in \mathbb{R}^n : u(J) = \mathbf{0}\}, \quad W_J = \{v \in \mathbb{R}^n : v(I) = \mathbf{0}\}.$$

Optimal bipartitions

Define an optimal bipartition as

$$\{I_{opt}, J_{opt}\} = \underset{\substack{I,J \subset V = \{1,\ldots,n\}, \\ |I|=|J|=\frac{n}{2}, J = V \setminus I}}{\operatorname{argmin}} \underbrace{\max_{u \in W_I, v \in W_J} \frac{|(u, Av)|}{(u, Au)^{1/2}(v, Av)^{1/2}}}_{\gamma_{IJ}}.$$

- The choice $\{I_{opt}, J_{opt}\}$ provides the least upper bound on the PCG convergence rate.
- Computing $\{I_{opt}, J_{opt}\}$ may not be feasible in practice.

Our approach: replace γ_{IJ} by a related "simpler" quantity, such that the resulting bipartition captures $\{I_{opt}, J_{opt}\}$.

Acut: the minimal averaged cut

Consider a set $S = \left\{(e_i, e_j) \; : \; i \in I, j \in J, a_{ij} \neq 0\right\} \subset W_I \times W_J$; e_i is the i-th unit vector.

Evaluate $\frac{|(u,Av)|}{(u,Au)^{1/2}(v,Av)^{1/2}}$ for all pairs in S and find the mean:

$$\tilde{\gamma}_{IJ} = \frac{1}{|S|} \sum_{i \in I, j \in J} \frac{|a_{ij}|}{\sqrt{a_{ii}a_{jj}}} = \frac{w(I,J)}{\text{cut}(I,J)},$$

where $w(I,J) = \sum_{i \in I, j \in J} w_{ij}$, $w_{ij} = |a_{ij}|/\sqrt{a_{ii}a_{jj}}$, and $\text{cut}(I,J) = |S|$.

Instead of $\{I_{opt}, J_{opt}\}$, we construct bipartitions

$$\{\tilde{I}_{opt}, \tilde{J}_{opt}\} = \underset{\substack{I,J \subset V = \{1,\ldots,n\}, \\ |I|=|J|=\frac{n}{2}, J = V \setminus I}}{\operatorname{argmin}} \underbrace{\frac{w(I,J)}{\text{cut}(I,J)}}_{\tilde{\gamma}_{IJ}}.$$

- In graph terms, we target an edge cut that has, in average, the smallest weight.
- We call such a cut the minimal averaged cut, or Acut.

Spectral Acut computations

The Acut problem: find a minimizer of

$$\min_{\substack{I,J \subset V = \{1,\ldots,n\}, \\ |I|=|J|=\frac{n}{2}, J = V \setminus I}} \frac{w(I,J)}{\text{cut}(I,J)}.$$

The problem can be reformulated as a bilinear form minimization

$$\min_p \frac{p^T L_W p}{p^T L p}, \qquad p^T z_l = 0, \qquad l = 1, \ldots, q.$$

- The minimizer is searched in the space of indicator vectors, $p(k) = \begin{cases} 1, & k \in I, \\ -1, & k \in J. \end{cases}$
- $L_W = D_W - W$ is the weighted graph Laplacian; D_W is the weighted degree matrix and $W = (w_{ij})$ is the weighted adjacency matrix.
- $L = D - Q$ is the standard unweighted graph Laplacian; D is the degree matrix and Q is the adjacency matrix.
- q is the number of connected components.
- $\text{span}\{z_1, \ldots, z_q\} = \text{null}(L_W) = \text{null}(L)$, where $z_l(k) = \begin{cases} 1, & k \in V_l, \\ 0, & k \notin V_l. \end{cases}$

The "relaxed" Acut problem: find the minimizer of

$$\min_v \frac{v^T L_W v}{v^T L v}, \qquad v \in \text{span}\{z_1, \ldots, z_q\}^{\perp}, \qquad v \in \mathbb{R}^n.$$

The minimum is achieved on the eigenvector associated with smallest eigenvalue of the symmetric generalized eigenvalue problem

$$L_W v = \lambda L v, \qquad v \in \text{span}\{z_1, \ldots, z_q\}^{\perp}.$$

- The resulting $\{I, J\}$ is formed by assigning the indices of the smallest $\lceil n/2 \rceil$ components to I and the rest to J; n is the problem size.
- Each bipartition can have several connected components, even for connected graphs.
- Orthogonalization against $\text{span}\{z_1, \ldots, z_q\}$ is computationally inexpensive.
- The approach can be recursively extended to construct a larger number of partitions.

We refer to the resulting algorithm as Acut-RSB.

Acut-RSB

Input: A. Output: Partition $\{V_i\}$.

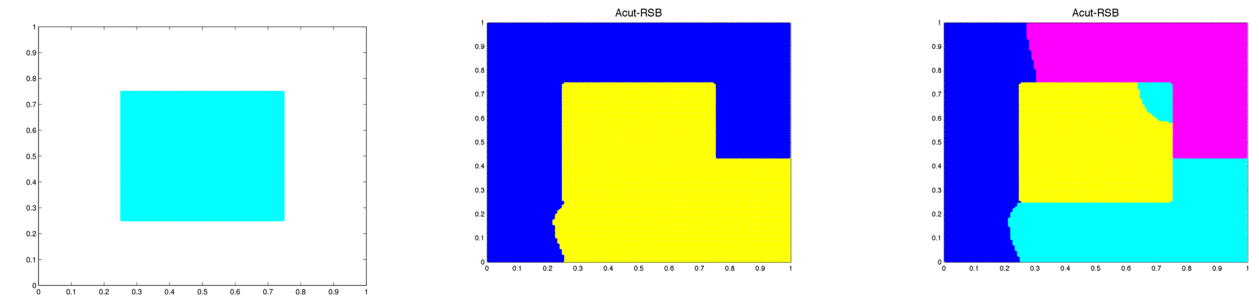
1. Form $G(A)$. Assign weights $w_{ij} = |a_{ij}|/\sqrt{a_{ii}a_{jj}}$; $L_W = D_W - W$ and $L = D - Q$.
2. Find connected components $\{E_l, V_l\}$. Define z_l.
3. Find the smallest eigenpair of $L_W v = \lambda L v$, $v \in \text{span}\{z_1, \ldots, z_q\}^{\perp}$. Define $\{I, J\}$.
4. If needed, apply Acut-RSB recursively to $A(I,I)$ and $A(J,J)$. Otherwise, return $\{I, J\}$.

Numerical experiments

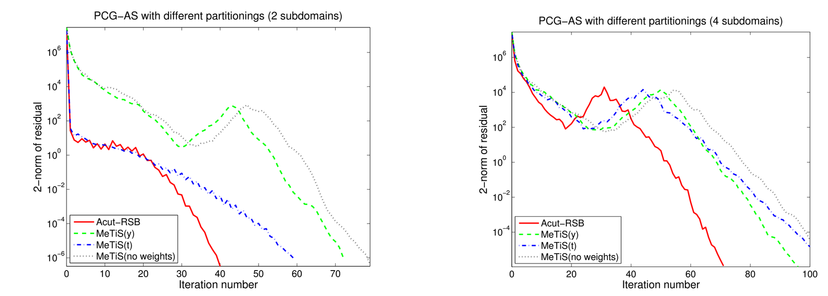
- Compare Acut-RSB to MeTiS with weights $y_{ij} = \lceil \gamma |a_{ij}|/\sqrt{a_{ii}a_{jj}} \rceil$ and $t_{ij} = \lceil \delta |a_{ij}| \rceil$.
- Use resulting partitions to construct the AS preconditioners for PCG (PCG-AS).
- Report relcut $=$ (cut size/nnz) $\times$ 100%, relcoef $= \left(\sum_{(i,j) \in \text{cut}} |a_{ij}|\right) / \left(\sum_{i,j} |a_{ij}|\right) \times 100\%$.
- Use LOBPCG with IC preconditioner as an eigensolver.

2D diffusion equation

FD discretization of the diffusion equation on a unit square with zero Dirichlet boundary conditions. The coefficients are 10^5 on $0.25 < x, y < 0.75$ and 1 elsewhere.



From left to right: problem geometry, Acut-RSB partitioning in 2 and 4 subdomains.

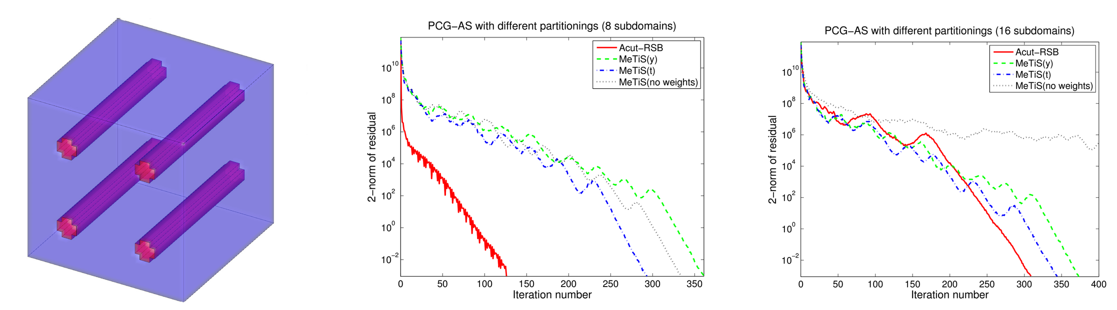


Convergence of PCG-AS with different partitions; $n = 16,384$.

	2 subdomains		4 subdomains	
Partitioning	relcut	relcoef	relcut	relcoef
Acut-RSB	0.76	3×10^{-5}	1.25	0.42
MeTiS(y)	0.44	0.85	0.86	1.62
MeTiS(t)	1.06	4×10^{-5}	1.67	0.88
MeTiS(no w.)	0.48	1.07	1.01	2.13

3D linear elasticity

FE discretization of the linear elasticity problem in a cube with material parameters $E = 10^6$ Pa and $\nu = 0.45$, penetrated by 4 bars with $E = 2.1 \times 10^{11}$ Pa and $\nu = 0.3$.



From left to right: problem geometry, convergence of PCG-AS with 8 and 16 subdomains; $n = 107,811$.

	8 subdomains		16 subdomains	
Partitioning	relcut	relcoef	relcut	relcoef
Acut-RSB	19.98	0.01	25.39	1.55
MeTiS(y)	8.03	2.51	11.14	2.51
MeTiS(t)	7.30	1.50	11.99	1.51
MeTiS(no w.)	6.77	2.24	10.61	10.28

Future work

- Address the trade-off between preconditioning quality and parallel efficiency.
- Establish rigorous conditions under which Acut is a preferable partitioning method.
- Search for possible applications of Acut in other areas of science (e.g., data clustering).

Eugene Vecharynski, Lawrence Berkeley National Laboratory
Email: evecharynski@lbl.gov