FLATIRON INSTITUTE

LINEAR PRECONDITIONERS

BARRY SMITH
11/23/2021

(Approximately) Solve

$$Ax^* = b$$

Circumstances for use

- Matrix is too large for direct solvers to be competitive
  - generic dense matrices work is $O(N^3)$
  - generic sparse matrices from an $n \times n \times n = N$ structured grid with nested dissection ordering work is $O\left(N^2 \log(N)\right)$
- Matrix has some/any underlying mathematical structure one can take advantage of (somehow)

Best case scenario work is $O(N) + O(nnz)$. Assuming A is not "approximated" during the entire solution process.

FLATIRON INSTITUTE

Repose as an optimization problem

$$A = A^T > 0$$

$$f(x) = \frac{1}{2} \left|\left|x^* - x\right|\right|_{W=A}^2$$

$$grad_x\, f(x) = -A(x^* - x) = -(b - Ax)$$

$$x_{new} = x_{old} + \tau(b - Ax_{old})$$

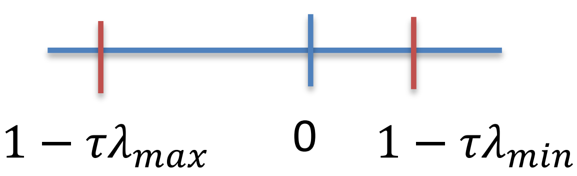
$$e_{new} = (I - \tau A)\, e_{old}$$

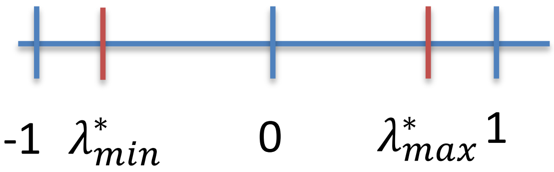
FLATIRON INSTITUTE

Convergence and Optimal τ

$$e_{new} = (I - \tau A)\, e_{old}$$

$$\|e_{new}\|_A \le \rho(I - \tau A)\, \|e_{old}\|_A$$





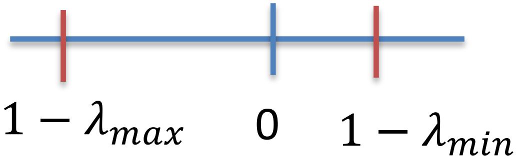
$$\tau^* = \frac{2}{\lambda_{min} + \lambda_{max}}$$

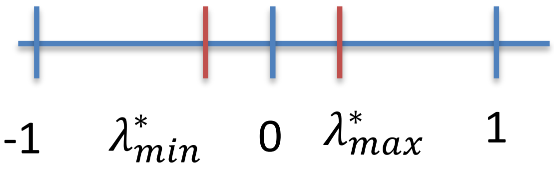
$$-\lambda^*_{min} = \lambda^*_{max} = \frac{1 - \frac{\lambda_{min}}{\lambda_{max}}}{1 + \frac{\lambda_{min}}{\lambda_{max}}}$$

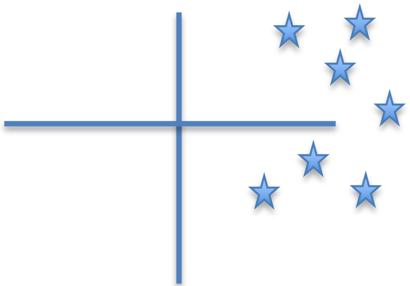
FLATIRON INSTITUTE

Preconditioners

$$e_{new} = (I - BA)\, e_{old}$$







FLATIRON INSTITUTE

Preconditioner in the optimization problem

$$A = A^T > 0;\ B = B^T > 0$$

$$f(y) = \frac{1}{2}\left|\left|y^* - y\right|\right|^2_{W = B^{1/2} A B^{1/2}}$$

$$grad_y\, f(y) = -B^{1/2} A B^{1/2} (y^* - y)$$

$$x_{new} = x_{old} + B(b - A x_{old})$$

$$e_{new} = (I - BA)\, e_{old}$$

FLATIRON INSTITUTE

An aside: Generalization

If $A \neq A^T$

$$f(x) = \frac{1}{2}\left|\left|x^* - x\right|\right|_{W=A^TA}^2$$

$$grad\ f(x) = -A^TA(x^* - x) = A^T(b - Ax)$$

$$x_{new} = x_{old} + \tau A^T(b - Ax_{old})$$

$$e_{new} = (I - \tau A^TA)\, e_{old}$$

FLATIRON INSTITUTE

Chebyshev Iteration

$$x_{new} = x_{old} + \tau^*(b - Ax_{old})$$

$$x_0 = b,\ x_1 = 2b - Ab,\ x_2 = 3b - 3Ab + A^2 b,\ \ \ldots$$

$$e_n = P_n(A)e_0$$

$$||e_n||_A \leq \rho(P_n(A))\,||e_0||_A$$

$$\leq \max_i |P_n(\lambda_i)|\,||e_0||_A$$

$$\leq \max_{\lambda^*_{min} \leq \lambda \leq \lambda^*_{max}} |P_n(\lambda)|\,||e_0||_A$$

(Slightly arbitrary) Division of Labor

$$x_{new} = x_{old} + B(b - Ax_{old})$$

Single-step methods (simple, stationary, or Richardson iteration) (48 in PETSc)

Multi-step methods – step depends on the history of the solution process

- Chebyshev – 3-term recursion, polynomial does not depend on b
- Krylov methods – steps depend **nonlinearly** on b (46 in PETSc)
    - Conjugate gradient – 3-term recursion
    - GMRES – history is the restart size

Linear preconditioners must be **linear** in b

FLATIRON INSTITUTE

Preconditioners complement Krylov methods
Krylov methods accelerate preconditioners

FLATIRON INSTITUTE

Galerkin Solution Process

- Select a subspace S of R^n
- Project the current error onto that subspace

$$\min_{\tilde{x}\in S} ||x^* - x - \tilde{x}||_W^2$$

$$\min_{\hat{x}} ||x^* - x - R^T\hat{x}||_W^2$$

W = 2-norm

$$RR^T\,\hat{x} = R(x^* - x) = R(A^{-1}b - x)$$

$W = A$-norm $(A = A^T > 0)$

$$(RAR^T)\,\hat{x} = RA(x^* - x) = R(b - Ax)$$

$W = A^T A$-norm

$$(RA^T AR^T)\,\hat{x} = RA^T(b - Ax)$$

FLATIRON INSTITUTE

Galerkin Solution Process

$$W = A\text{-norm } (A = A^T > 0)$$

$$(RAR^T)\,\hat{x} = RA(x^* - x) = R(b - Ax)$$

$$x^c = R^T(RAR^T)^{-1}RA(x^* - x)$$

$$e_{new} = (I - R^T(RAR^T)^{-1}RA)e_{old}$$

RAR^T is the operator A restricted to the subspace, represented in a different basis, as such it will in general have different spectral properties than A

In certain cases, like multigrid, it has the same structure

FLATIRON INSTITUTE

Combining Galerkin Solutions

$$R_i, \qquad B_i = R_i^T \left(R_i A R_i^T\right)^{-1} R_i, \qquad P_i = B_i A$$

- Multiplicative – apply the updates sequentially

$$e = (I - P_3)(I - P_2)(I - P_1)e$$

- Symmetric-multiplicative – revisit the subspaces backwards

$$e = (I - P_1)(I - P_2)(I - P_3)(I - P_2)(I - P_1)e$$

- Additive – apply the updates in parallel

$$e = (I - P_1 - P_2 - P_3)e$$

- Coloring – apply independent updates in parallel, dependent updates sequentially

FLATIRON INSTITUTE

Additive Form

$$e_{new} = (I - P_1 - P_2)\, e_{old}$$

$$(x^* - x_{new}) = (I \; - B_1 A \; - B_2 A)(x^* - x_{old})$$

$$x_{new} = x_{old} + (B_1 + B_2)(b \; - A x_{old})$$

Simple iteration for the system

$$(B_1 + B_2) A\, x = (B_1 + B_2) b$$

Multiplicative form

$$(B_2 A B_1 - B_1 - B_2) A x = \; (B_2 A B_1 - B_1 - B_2)\mathrm{b}$$

FLATIRON INSTITUTE

Examples

Jacobi and Gauss-Seidel methods

$$R_i^T = e_i = \{0, \dots, 0,1,0, \dots, 0\}^T$$

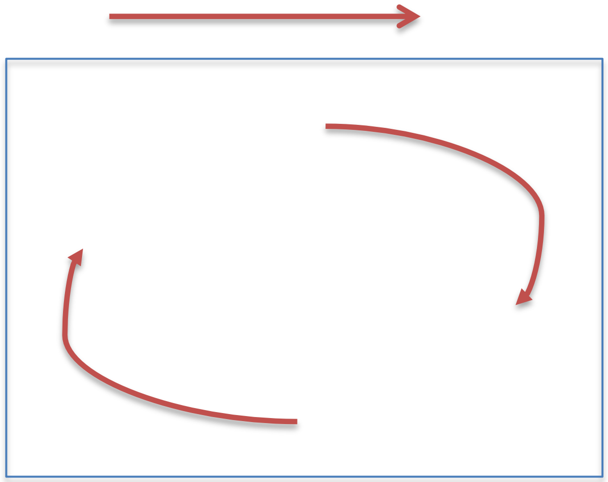
Block Jacobi and block Gauss-Seidel methods

$$R_i^T = \{0, \dots, 0, I, 0, \dots, 0\}^T$$

FLATIRON INSTITUTE

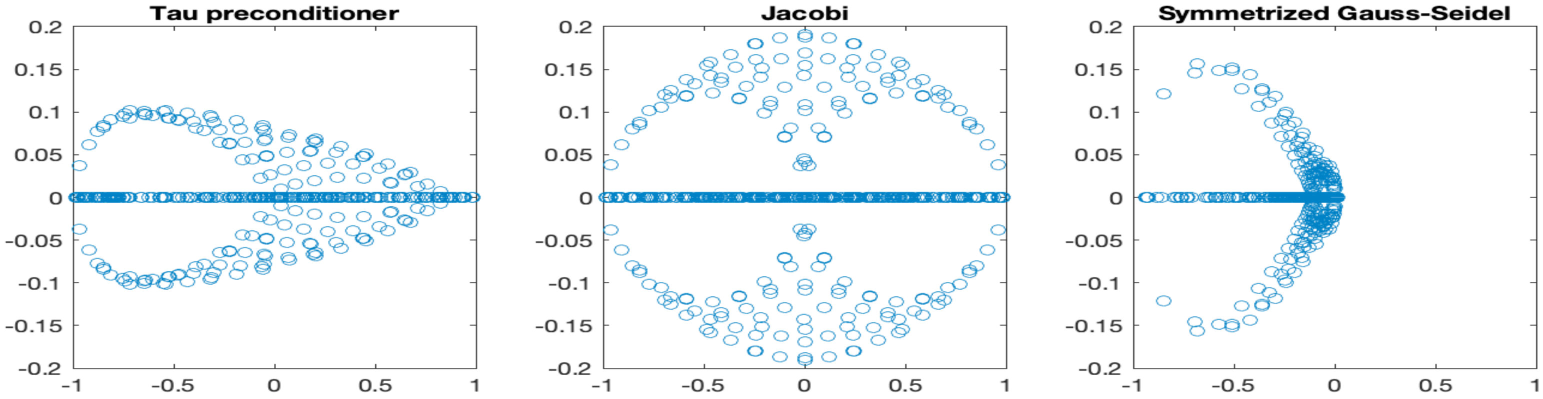
Model Problem

Driven cavity with temperature in two dimensions with velocity-vorticity formulation.

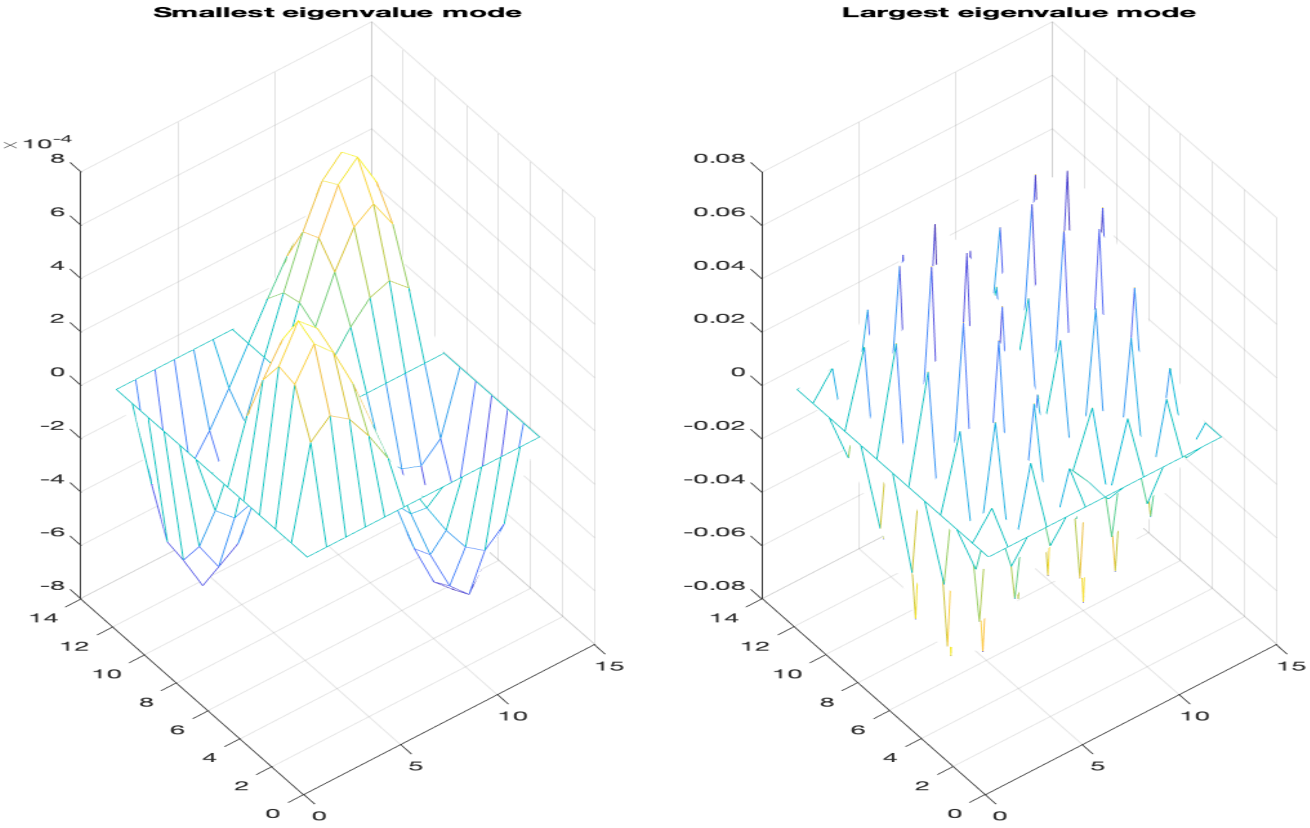


FLATIRON INSTITUTE

Spectrum of Preconditioned Operator

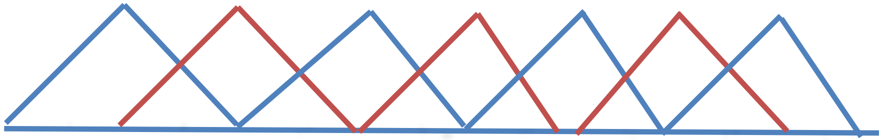


Eigenvectors of x-velocity

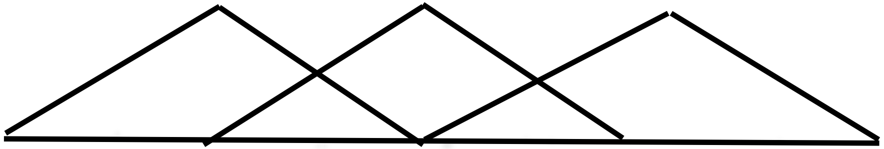


PDE Oriented Preconditioners: Multigrid

Add a space to resolve the smooth components

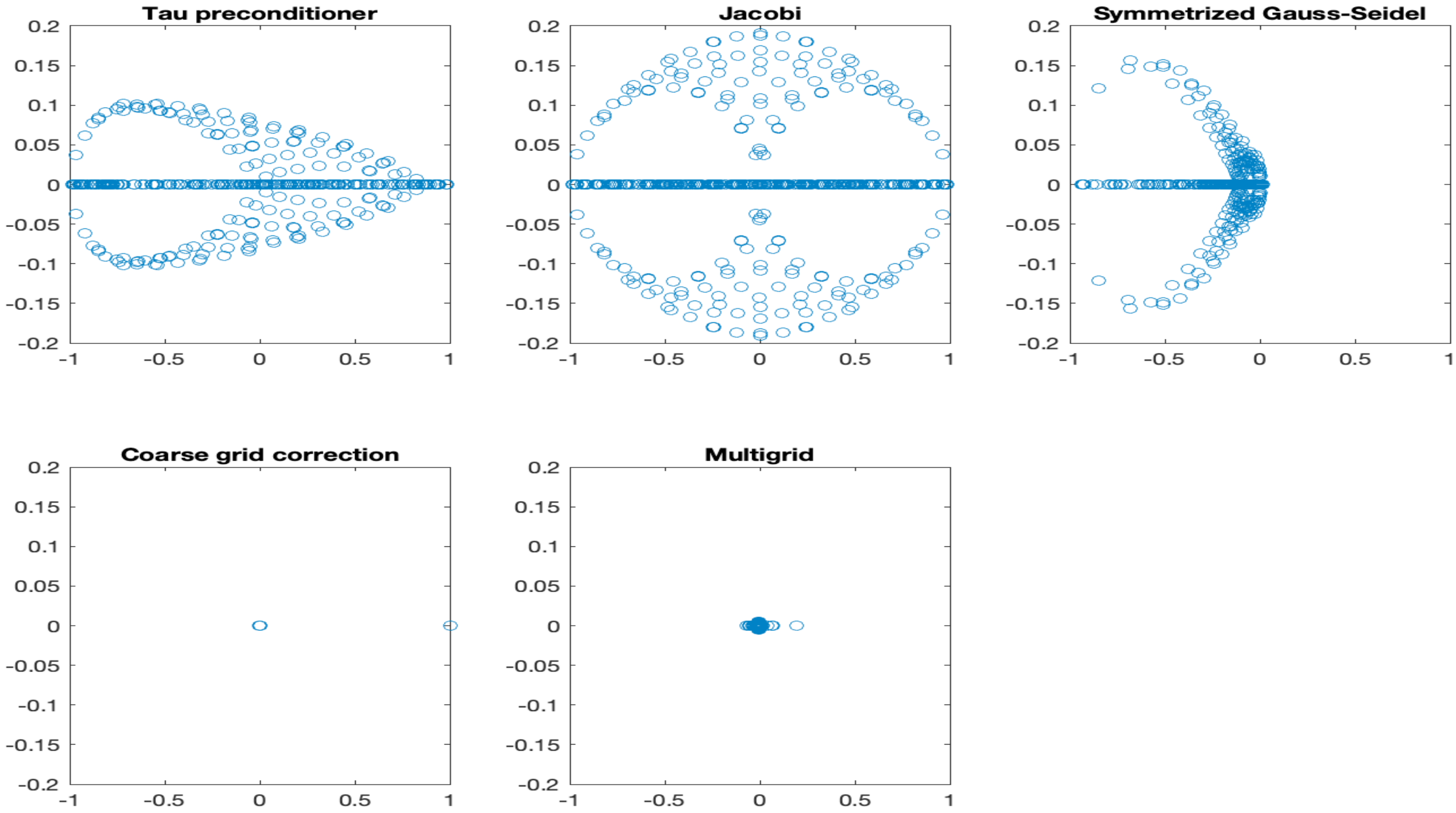


S_i

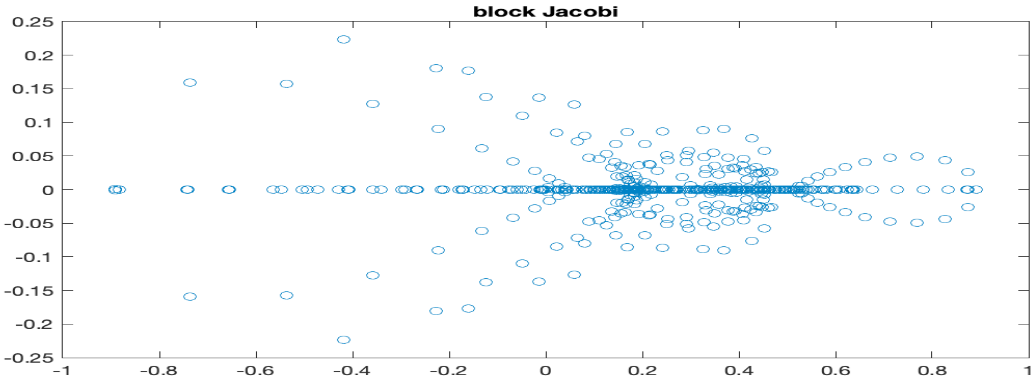


$\hat{S}$

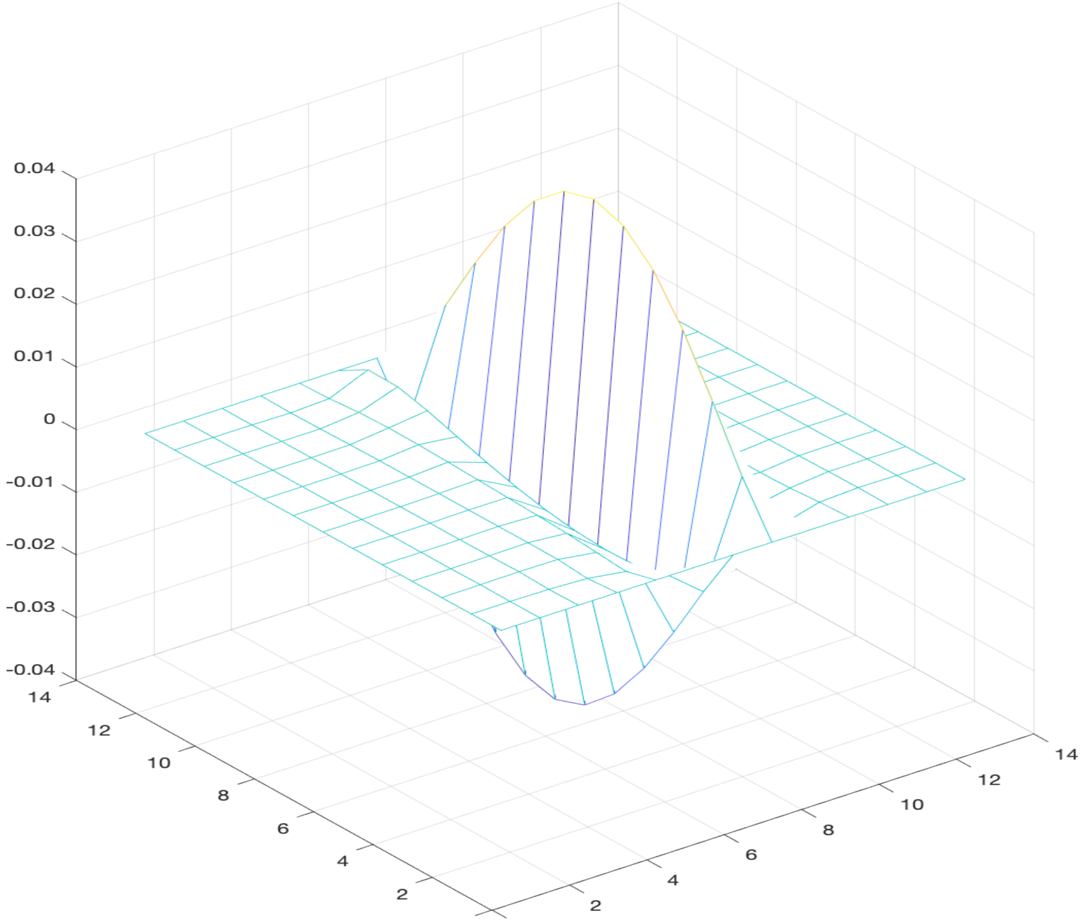
Spectrum of Preconditioned Operator



Spectrum of Preconditioned Operator

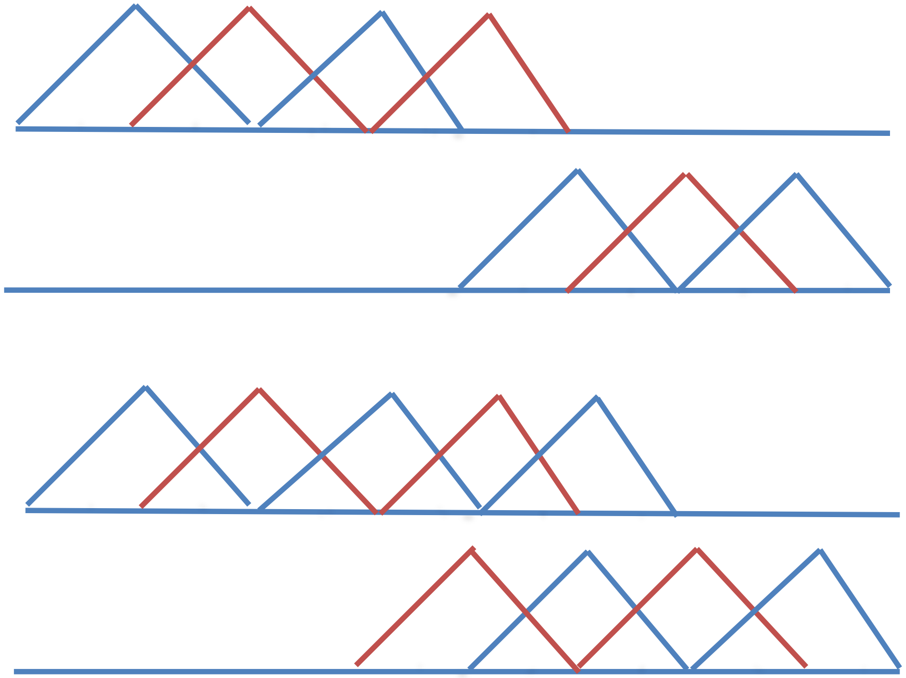


Eigenvector associated with largest eigenvalue



PDE Oriented Preconditioners: Schwarz

Extend the block Jacobi subspaces to larger overlap



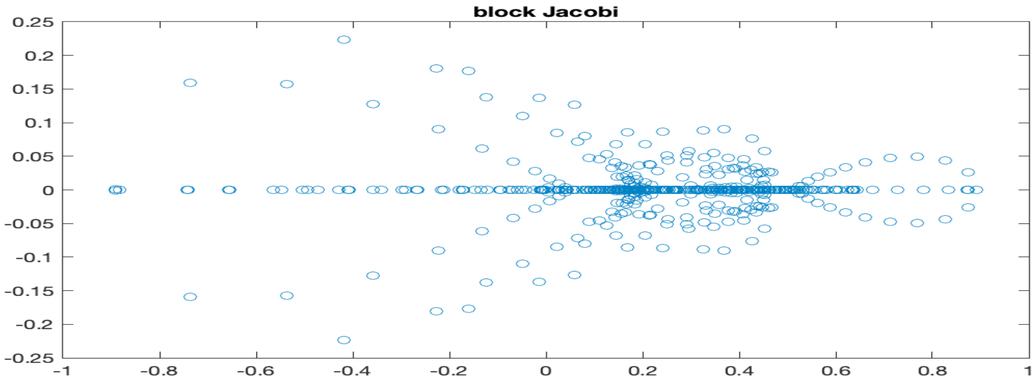
S_1

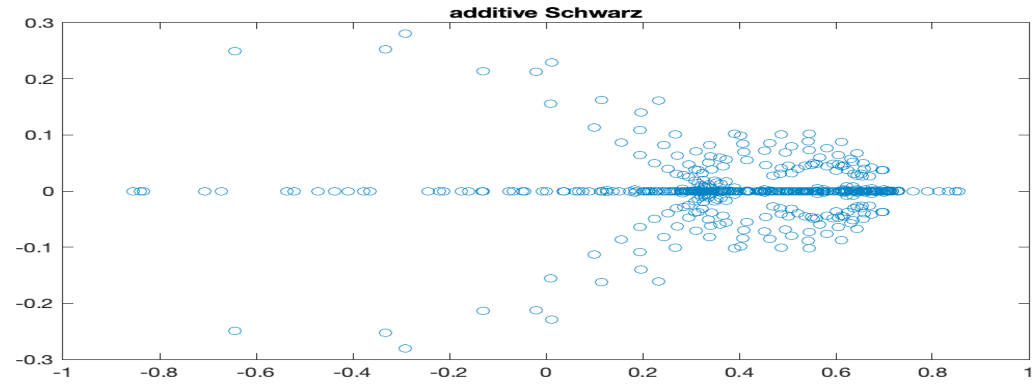
S_2

$\hat{S}_1$

$\hat{S}_2$

Spectrum of Preconditioned Operator





Schur Complement Preconditioners

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix}. \qquad R_1^T = \begin{pmatrix} I \\ 0 \end{pmatrix}, \qquad B_1 = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

Choose R_2^T to be A-orthogonal to R_1^T; $\quad R_1 A R_2^T = 0$

$$\begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix} \begin{pmatrix} -A_{11}^{-1} A_{12} \\ I \end{pmatrix} = 0$$

$$\mathrm{B}_2 = \begin{pmatrix} -A_{11}^{-1} A_{12} \\ I \end{pmatrix} \left(A_{22} - A_{12}^T A_{11}^{-1} A_{12} \right)^{-1} \begin{pmatrix} -A_{12}^T A_{11}^{-1} & I \end{pmatrix}$$

Generalizations

$$R_i, \qquad Q_i, \qquad B_i = Q_i\, \tilde{A}_i^{-1} R_i, \qquad P_i = B_i A$$

Restricted/optimized additive Schwarz method

Either Q_i or R_i zeros or weights the values in the overlapping region

Convergence Theory

Well-developed for multigrid, domain decomposition, etc.

Bounds on the preconditioned spectrum are obtained (usually) by utilizing the underlying continuum spaces (e.g. Sobolov spaces) with finite element discretizations. Use the correct function space (norms)

Results are intuitive: Better convergence with

- more overlap of subspaces (but not too much)
- more orthogonality between the subspaces

"Orthogonality" is expensive to explicitly ensure hence adding more subspace overlap is generally the least expensive way to improve convergence

Chapter 5, **Domain Decomposition Methods,** BF Smith, PE Bjorstad, WD Gropp

Nonlinear analogs

SIAM Review, *Composing scalable nonlinear algebraic solvers*
PR Brune, MG Knepley, BF Smith, and X Tu

FLATIRON INSTITUTE

FLATIRON
INSTITUTE