Exploring (un)conventional preconditioning strategies for large saddle point algebraic linear systems

V. Simoncini

Dipartimento di Matematica, Università di Bologna

valeria.simoncini@unibo.it

The problem

$$\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

- Computational Fluid Dynamics (Elman, Silvester, Wathen 2005)
- Elasticity problems
- Mixed (FE) formulations of II and IV order elliptic PDEs
- Linearly Constrained Programs
- Linear Regression in Statistics
- Image restoration
- ... Survey: Benzi, Golub and Liesen, Acta Num 2005

The problem. Simplifications

$$\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

- Iterative solution by means of Krylov subspace methods

- Structural properties. Focus for this talk:
  - ⋆ A symmetric positive (semi)definite
  - ⋆ B^T tall, possibly rank deficient
  - ⋆ C symmetric positive (semi)definite

Spectral properties

$$\mathcal{M} = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix}$$

$0 < \lambda_n \leq \cdots \leq \lambda_1$ eigs of A

$0 = \sigma_m \leq \cdots \leq \sigma_1$ sing. vals of B

$\lambda_{\max}(C) > 0, \quad BB^T + C$ full rank

$$\mathrm{spec}(\mathcal{M}) \subset [-a, -b] \cup [c, d], \quad a, b, c, d > 0$$

$\Rightarrow$ A large variety of results on the spectrum of $\mathcal{M}$, also for indefinite and singular A

$\Rightarrow$ Search for good preconditioning strategies...

General preconditioning strategy

- Find $\mathcal{P}$ such that

$$\mathcal{M}\mathcal{P}^{-1}\hat{u} = b \qquad \hat{u} = \mathcal{P}u$$

is easier (faster) to solve than $\mathcal{M}u = b$

- A look at efficiency:

- Dealing with $\mathcal{P}$ should be cheap
- Storage requirements for $\mathcal{P}$ should be low

- Properties (algebraic/functional) should be exploited

Mesh/parameter independence

Structure preserving preconditioners

Block diagonal Preconditioner

⋆ A nonsing., $C = 0$:

$$\mathcal{P}_0 = \begin{bmatrix} A & 0 \\ 0 & BA^{-1}B^T \end{bmatrix}$$

$$\Rightarrow \qquad \mathcal{P}_0^{-\frac{1}{2}} \mathcal{M} \mathcal{P}_0^{-\frac{1}{2}} = \begin{bmatrix} I & A^{-\frac{1}{2}} B^T (BA^{-1}B^T)^{-\frac{1}{2}} \\ (BA^{-1}B^T)^{-\frac{1}{2}} BA^{-\frac{1}{2}} & 0 \end{bmatrix}$$

MINRES converges in at most 3 iterations. $\mathrm{spec}(\mathcal{P}_0^{-\frac{1}{2}} \mathcal{M} \mathcal{P}_0^{-\frac{1}{2}}) = \left\{1, \frac{1}{2} \pm \frac{\sqrt{5}}{2}\right\}$

Block diagonal Preconditioner

⋆ A nonsing., $C = 0$:

$$\mathcal{P}_0 = \begin{bmatrix} A & 0 \\ 0 & BA^{-1}B^T \end{bmatrix}$$

$$\Rightarrow \qquad \mathcal{P}_0^{-\frac{1}{2}} \mathcal{M} \mathcal{P}_0^{-\frac{1}{2}} = \begin{bmatrix} I & A^{-\frac{1}{2}} B^T (BA^{-1}B^T)^{-\frac{1}{2}} \\ (BA^{-1}B^T)^{-\frac{1}{2}} BA^{-\frac{1}{2}} & 0 \end{bmatrix}$$

MINRES converges in at most 3 iterations. $\quad \mathrm{spec}(\mathcal{P}_0^{-\frac{1}{2}} \mathcal{M} \mathcal{P}_0^{-\frac{1}{2}}) = \left\{1, \frac{1}{2} \pm \frac{\sqrt{5}}{2}\right\}$

A more practical choice:

$$\mathcal{P} = \begin{bmatrix} \widetilde{A} & 0 \\ 0 & \widetilde{S} \end{bmatrix} \qquad \text{spd.} \quad \widetilde{A} \approx A \qquad \widetilde{S} \approx BA^{-1}B^T$$

eigs of $\mathcal{M}\mathcal{P}^{-1}$ in $\qquad [-a, -b] \cup [c, d], \qquad a, b, c, d > 0$

Still an Indefinite Problem

Giving up symmetry ...

- Change the preconditioner: *Mimic the LU factors*

$$\mathcal{M} = \begin{bmatrix} I & O \\ BA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B^T \\ O & BA^{-1}B^T + C \end{bmatrix} \quad \Rightarrow \mathcal{P} \approx \begin{bmatrix} A & B^T \\ O & BA^{-1}B^T + C \end{bmatrix}$$

Giving up symmetry ...

- Change the preconditioner: *Mimic the LU factors*

$$\mathcal{M} = \begin{bmatrix} I & O \\ BA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B^T \\ O & BA^{-1}B^T + C \end{bmatrix} \quad \Rightarrow \mathcal{P} \approx \begin{bmatrix} A & B^T \\ O & BA^{-1}B^T + C \end{bmatrix}$$

- Change the preconditioner: *Mimic the Structure*

$$\mathcal{M} = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \quad \Rightarrow \mathcal{P} \approx \mathcal{M}$$

Giving up symmetry ...

- Change the preconditioner: *Mimic the LU factors*

$$\mathcal{M} = \begin{bmatrix} I & O \\ BA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B^T \\ O & BA^{-1}B^T + C \end{bmatrix} \quad \Rightarrow \mathcal{P} \approx \begin{bmatrix} A & B^T \\ O & BA^{-1}B^T + C \end{bmatrix}$$

- Change the preconditioner: *Mimic the Structure*

$$\mathcal{M} = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \quad \Rightarrow \mathcal{P} \approx \mathcal{M}$$

- Change the matrix: *Eliminate indef.* $\mathcal{M}_- = \begin{bmatrix} A & B^T \\ -B & C \end{bmatrix}$

Giving up symmetry ...

- Change the preconditioner: *Mimic the LU factors*

$$\mathcal{M} = \begin{bmatrix} I & O \\ BA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B^T \\ O & BA^{-1}B^T + C \end{bmatrix} \quad \Rightarrow \mathcal{P} \approx \begin{bmatrix} A & B^T \\ O & BA^{-1}B^T + C \end{bmatrix}$$

- Change the preconditioner: *Mimic the Structure*

$$\mathcal{M} = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \quad \Rightarrow \mathcal{P} \approx \mathcal{M}$$

- Change the matrix: *Eliminate indef.* $\quad \mathcal{M}_- = \begin{bmatrix} A & B^T \\ -B & C \end{bmatrix}$

- Change the matrix: *Regularize* $(C = 0)$

$$\mathcal{M} \ \Rightarrow \mathcal{M}_\gamma = \begin{bmatrix} A & B^T \\ B & -\gamma W \end{bmatrix} \text{ or } \mathcal{M}_\gamma = \begin{bmatrix} A + \frac{1}{\gamma} B^T W^{-1} B & B^T \\ B & O \end{bmatrix}$$

... But recovering symmetry in disguise

Nonstandard inner product:

Let $\mathcal{W}$ be any of $\mathcal{M}\mathcal{P}^{-1}, \mathcal{M}_-$

For $\mathrm{spec}(\mathcal{W})$ in $\mathbb{R}^+$, find symmetric matrix H such that

$$\mathcal{W}H = H\mathcal{W}^T$$

(that is, $\mathcal{W}$ is H-symmetric)

... But recovering symmetry in disguise

Nonstandard inner product:

Let $\mathcal{W}$ be any of $\mathcal{M}\mathcal{P}^{-1}, \mathcal{M}_-$

For $\mathrm{spec}(\mathcal{W})$ in $\mathbb{R}^+$, find symmetric matrix H such that

$$\mathcal{W}H = H\mathcal{W}^T$$

(that is, $\mathcal{W}$ is H-symmetric)

If H is spd then

- $\mathcal{W}$ is diagonalizable
- Use PCG on $\mathcal{W}$ with H-inner product

Constraint (Indefinite) Preconditioner

$$\mathcal{P} = \begin{bmatrix} \widetilde{A} & B^T \\ B & -C \end{bmatrix} \qquad \mathcal{M}\mathcal{P}^{-1} = \begin{bmatrix} A\widetilde{A}^{-1}(I-\Pi)+\Pi & \star \\ O & I \end{bmatrix}$$

with $\Pi = B(B\widetilde{A}^{-1}B^T + C)^{-1}B\widetilde{A}^{-1}$

- Constraint equation satisfied at each iteration
- If C nonsing $\Rightarrow$ all eigs real and positive
- If $B^TC = 0$ and $BB^T + C > 0 \Rightarrow$ all eigs real and positive

$\Rightarrow$ More general cases, $\widetilde{B} \approx B$, $\widetilde{C} \approx C$

The Stokes problem

Minimize

$$J(u) = \frac{1}{2}\int_\Omega |\nabla u|^2 dx - \int_\Omega f \cdot u dx$$

subject to $\nabla \cdot u = 0$ in Ω

Lagrangian: $\mathcal{L}(u,p) = J(u) + \int_\Omega p \nabla \cdot u dx$

Optimality condition on discretized Lagrangian leads to:

$$\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}$$

A second-order operator, B first-order operator, C zero-order operator

The Stokes problem. Contraint preconditioning

$$\mathcal{P} = \begin{bmatrix} \widetilde{A} & B^T \\ B & B\widetilde{A}^{-1}B^T - S \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ B\widetilde{A}^{-1} & I_m \end{bmatrix} \begin{bmatrix} \widetilde{A} & 0 \\ 0 & -S \end{bmatrix} \begin{bmatrix} I_n & \widetilde{A}^{-1}B^T \\ 0 & I_m \end{bmatrix}$$

with $S \approx B\widetilde{A}^{-1}B^T + C \; spd$

The Stokes problem. Contraint preconditioning

$$\mathcal{P} = \begin{bmatrix} \widetilde{A} & B^T \\ B & B\widetilde{A}^{-1}B^T - S \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ B\widetilde{A}^{-1} & I_m \end{bmatrix} \begin{bmatrix} \widetilde{A} & 0 \\ 0 & -S \end{bmatrix} \begin{bmatrix} I_n & \widetilde{A}^{-1}B^T \\ 0 & I_m \end{bmatrix}$$

with $S \approx B\widetilde{A}^{-1}B^T + C$ *spd*

Selection of $\widetilde{A}$, S: $\widetilde{A} = \text{AMG}(A)$, $S = Q$ (pressure mass matrix)

IFISS 3.1 (Elman, Ramage, Silvester):

Flow over a backward facing step

Stable Q2-Q1 approximation

$(C = 0, B \in \mathbb{R}^{m\times n})$

stopping tolerance: 10^{-6}

non-symmetric solver

n	m	# it.
1538	209	18
5890	769	18
23042	2945	18
91138	11521	17
362498	45569	17

A standard choice: block diagonal preconditioning

$$\mathcal{P} = \begin{bmatrix} \widetilde{A} & 0 \\ 0 & \widetilde{S} \end{bmatrix} \qquad \text{spd.} \quad \widetilde{A} \approx A \qquad \widetilde{S} \approx BA^{-1}B^T$$

spectrum of $\mathcal{M}\mathcal{P}^{-1}$ in $\quad [-a, -b] \cup [c, d], \quad a, b, c, d > 0$

$\Rightarrow$ if $\widetilde{A}, A$ and $\widetilde{S}, BA^{-1}B^T$ spectrally equivalent, then spectrum of$\mathcal{M}\mathcal{P}^{-1}$ is independent of mesh parameter

An example. Stokes problem

$$\begin{bmatrix} -\Delta & -\mathrm{grad} \\ \mathrm{div} & \end{bmatrix} \approx \begin{bmatrix} -\widetilde{\Delta} & \\ & I \end{bmatrix}$$

In algebraic terms:

$I \to$ mass matrix
$-\widetilde{\Delta} \to$ Algebraic MG
(spectrally equivalent matrix)

(cf. K.-A. Mardal & R. Winther
JNLAA 2011)

An example. Stokes problem

$$\begin{bmatrix} -\Delta & -\mathrm{grad} \\ \mathrm{div} & \end{bmatrix} \approx \begin{bmatrix} -\widetilde{\Delta} & \\ & I \end{bmatrix}$$

In algebraic terms:

$I \to$ mass matrix
$-\widetilde{\Delta} \to$ Algebraic MG
(spectrally equivalent matrix)

(cf. K.-A. Mardal & R. Winther JNLAA 2011)

2D. Final residual norm $< 10^{-6}$

size($\mathcal{M}$)	its	Time (secs)
578	26	0.04
217	26	0.14
8450	26	0.50
132098	26	11.17

Next: some unexpected behaviors...

Choice of Schur complement approximation. A quasi-optimal choice

$$\widetilde{S} \approx BA^{-1}B^T$$

For certain operators, $\widetilde{S}$ is quasi-optimal:

spec$(BA^{-1}B^T\widetilde{S}^{-1})$ well clustered except for few eigenvalues

Choice of Schur complement approximation. A quasi-optimal choice

$$\widetilde{S} \approx BA^{-1}B^T$$

For certain operators, $\widetilde{S}$ is quasi-optimal:

spec($BA^{-1}B^T\widetilde{S}^{-1}$) well clustered except for few eigenvalues

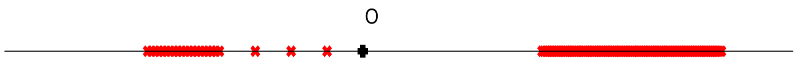
Possibly: well clustered eigs also mesh-independent

The role of $\widetilde{S}$

Claim:

The presence of outliers in $BA^{-1}B^T\widetilde{S}^{-1}$ is accurately inherited by the preconditioned matrix $\mathcal{M}\mathcal{P}^{-1}$ so that $\kappa(\mathcal{M}\mathcal{P}^{-1}) \gg 1$

spec$(BA^{-1}B^T\widetilde{S}^{-1})$



spec$(\mathcal{M}\mathcal{P}^{-1})$

(for a proof, see Olshanskii & Simoncini, SIMAX '10)

Stokes type problem with variable viscosity in $\Omega \subset \mathbb{R}^d$

$$
\begin{aligned}
-\mathbf{div}\nu(\mathbf{x})\mathbf{Du} + \nabla p = \mathbf{f} \quad &\text{in} \quad \Omega, \\
-\mathrm{div}\ \mathbf{u} = 0 \quad &\text{in} \quad \Omega, \\
\mathbf{u} = 0 \quad &\text{on} \quad \partial\Omega,
\end{aligned}
$$

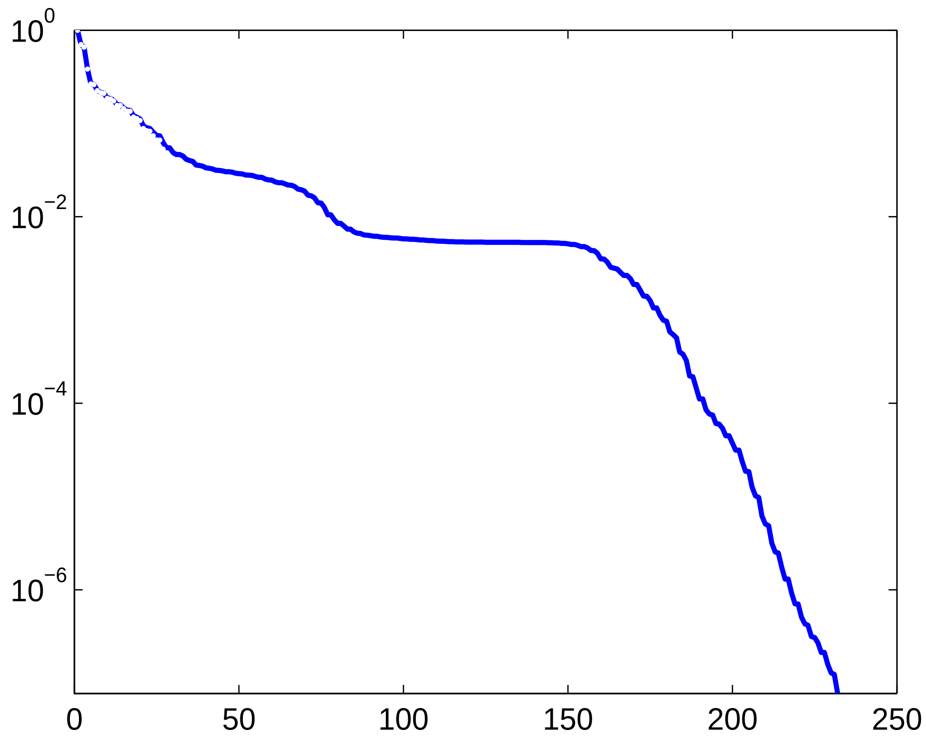
with $0 < \nu_{\min} \leq \nu(\mathbf{x}) \leq \nu_{\max} < \infty$ (Here, $\nu(\mathbf{x}) = 2\mu + \frac{\tau_s}{\sqrt{\varepsilon^2 + |\mathbf{Du}(\mathbf{x})|^2}}$)

$\mathbf{u}$: velocity vector field $\qquad$ p : pressure

$\mathbf{Du} = \frac{1}{2}(\nabla \mathbf{u} + \nabla^T \mathbf{u})$ rate of deformation tensor

Prec. S: pressure mass matrix wrto weighted product $(\nu^{-1}\cdot, \cdot)_{L^2(\Omega)}$

Performance of Krylov subspace solver MINRES

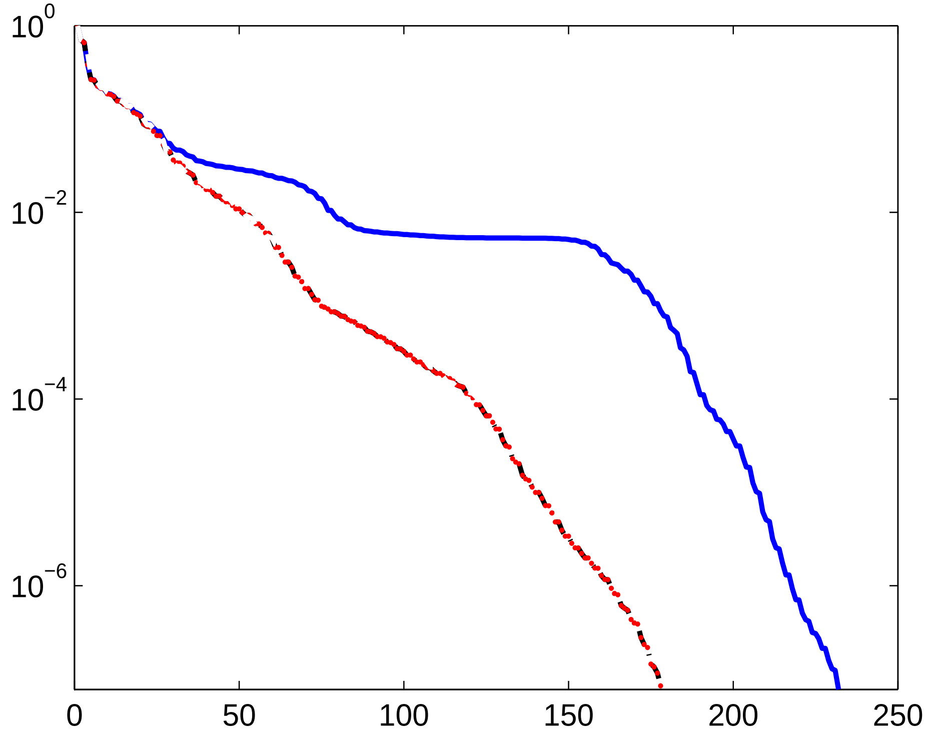


$\widetilde{A} = \mathrm{IC}(A, \delta), \delta = 10^{-2}$ poor approximation

$\Rightarrow$ also one small positive eig

Bercovier-Engelman model of the Bingham viscoplastic fluid

Performance of Krylov subspace solver MINRES



deflation of approximate “bad” eigenvectors

$\widetilde{A} = \mathrm{IC}(A, \delta), \delta = 10^{-2}$ poor approximation

$\Rightarrow$ also one small positive eig

Bercovier-Engelman model of the Bingham viscoplastic fluid

Distributed optimal control for time-periodic parabolic equations

Joint work with W. Zulehner and W. Krendl

$$J(y,u) = \frac{1}{2}\int_0^T \int_\Omega |y(x,t) - y_d(x,t)|^2 \, dx \, dt + \frac{\nu}{2}\int_0^T \int_\Omega |u(x,t)|^2 \, dx \, dt$$

subject to the time-periodic parabolic problem

$$\begin{aligned}
\frac{\partial}{\partial t} y(x,t) - \Delta y(x,t) &= u(x,t) && \text{in } Q_T, \\
y(x,t) &= 0 && \text{on } \Sigma_T, \\
y(x,0) &= y(x,T) && \text{on } \Omega, \\
u(x,0) &= u(x,T) && \text{on } \Omega.
\end{aligned}$$

Here $y_d(x,t)$ is a given target (or desired) state and $\nu > 0$ is a cost or regularization parameter.

Assuming y_d to be time-harmonic (so that there exist y, u time-harmonic), gives the problem:

Minimize

$$\frac{1}{2}\int_{\Omega} |y(x) - y_d(x)|^2 \; dx + \frac{\nu}{2}\int_{\Omega} |u(x)|^2 \; dx$$

subject to

$$\begin{aligned} i\omega\, y(x) - \Delta y(x) &= u(x) \quad in\ \Omega, \\ y(x) &= 0 \qquad\quad on\ \Gamma \end{aligned}$$

Solution using Lagrange multipliers, discretization and elimination of the control, yields:

$$\begin{bmatrix} M & K - i\omega\, M \\ K + i\omega\, M & -\frac{1}{\nu} M \end{bmatrix} \begin{bmatrix} \underline{y} \\ \underline{p} \end{bmatrix} = \begin{bmatrix} M\underline{y}_d \\ 0 \end{bmatrix}$$

Solving the saddle point linear system

After simple scaling,

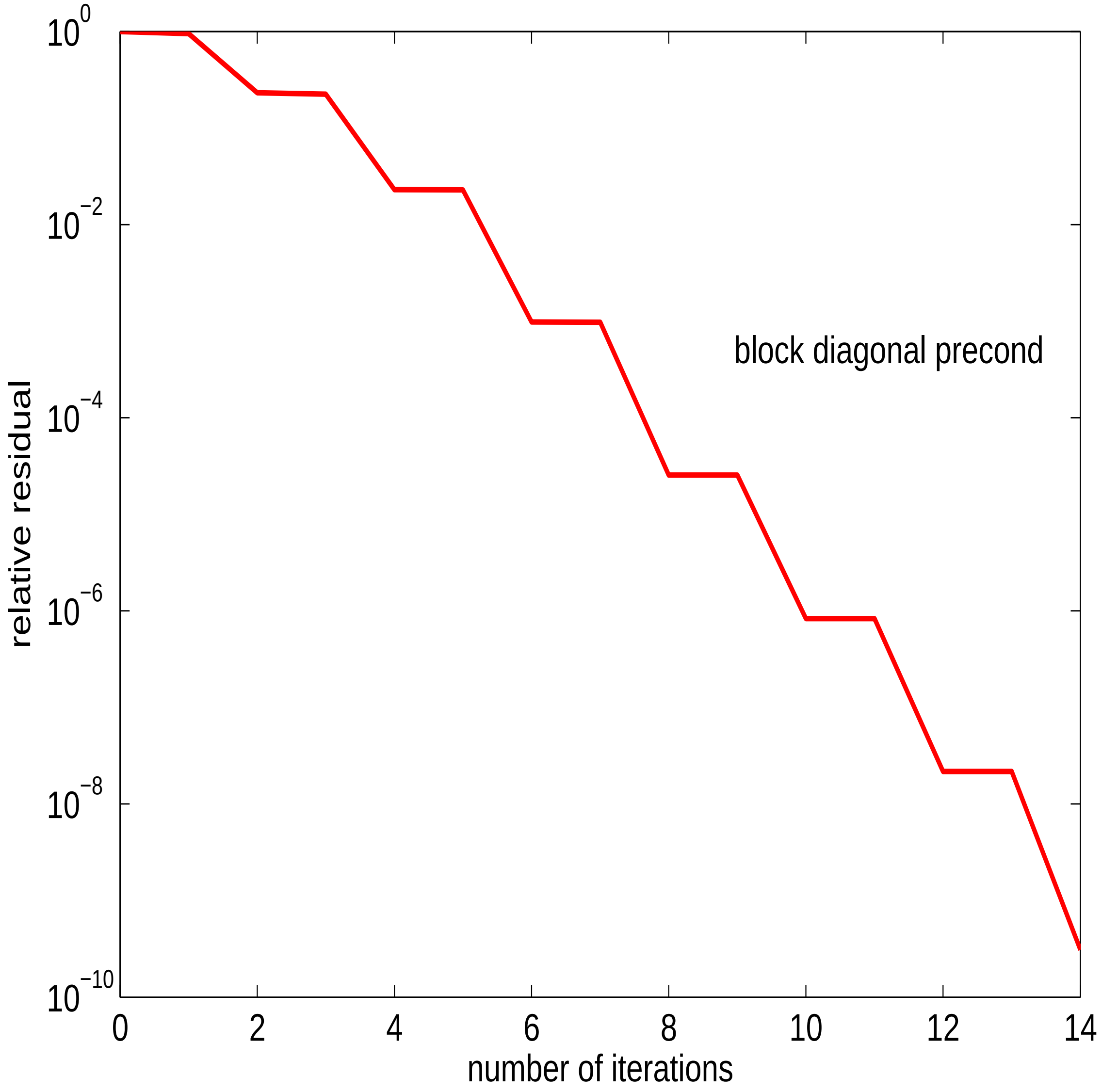
$$\begin{bmatrix} M & \sqrt{\nu}\,(K - i\omega\, M) \\ \sqrt{\nu}\,(K + i\omega\, M) & -M \end{bmatrix} \begin{bmatrix} \underline{y} \\ \frac{1}{\sqrt{\nu}}\,\underline{p} \end{bmatrix} = \begin{bmatrix} M\underline{y}_d \\ 0 \end{bmatrix}$$

Block diagonal Preconditioner:

$$\mathcal{P} = \begin{bmatrix} M + \sqrt{\nu}\,(K + \omega\, M) & 0 \\ 0 & M + \sqrt{\nu}\,(K + \omega\, M) \end{bmatrix}$$

- Accurate estimates for the spectral intervals
- Convergence of MINRES independent of the mesh and regularization parameters

Convergence history. Staircase behavior



Explanation of the Staircase behavior

The previous matrix has the form:

$$\mathcal{M} = \begin{bmatrix} A & B^* \\ B & -A \end{bmatrix} \in \mathbb{C}^{2n\times 2n},$$

with $A \in \mathbb{R}^{n\times n}$ spd, and $B \in \mathbb{C}^{n\times n}$ complex symmetric, i.e., $B = B^T$

THEOREM: Assume that B is nonsingular. Then the eigenvalues μ of $\mathcal{M}$ come in pairs, $(\mu, -\mu)$, with $\mu \in \mathbb{R}$.

$\Rightarrow$ MINRES behaves like CG on a matrix having only the positive eigenvalues, but with twice as many iterations

---

Remark: Similar setting for more complex structures, e.g., for Distributed optimal control for the time-periodic Stokes equations

Convergence history. Staircase behavior

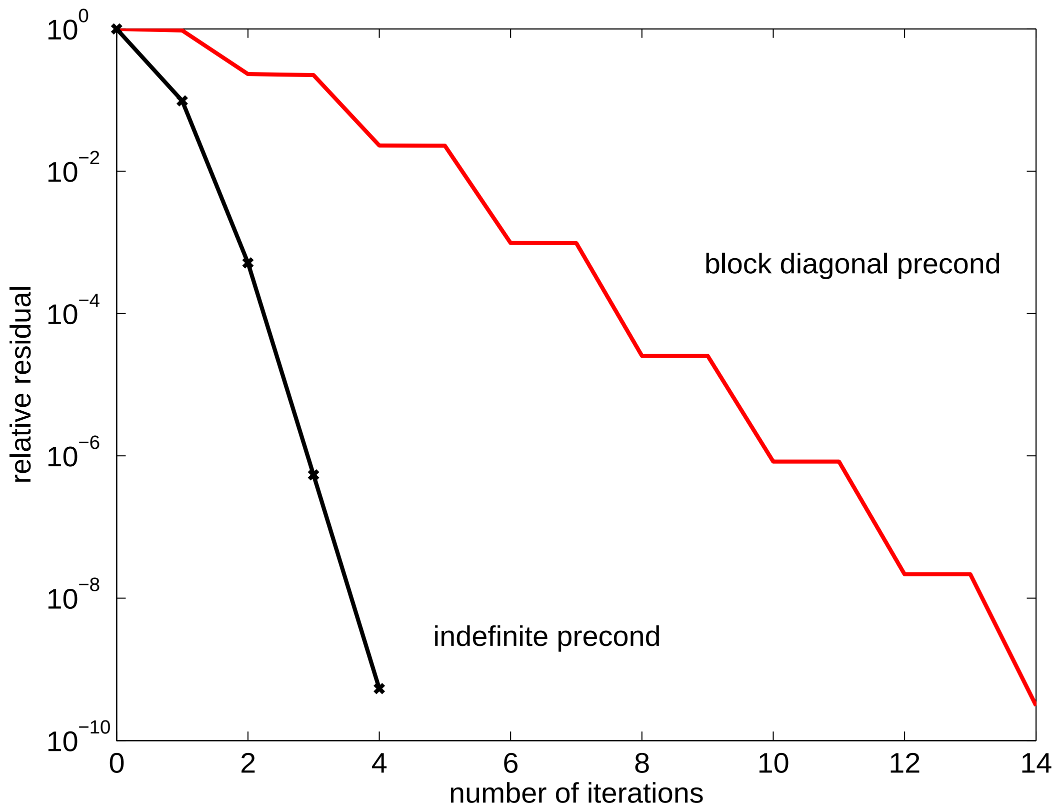
An alternative (indefinite) preconditioner - work in progress:

$$\mathcal{P} = \begin{bmatrix} 0 & K + \omega\, M \\ K + \omega\, M & -\frac{1}{\nu} M \end{bmatrix}$$

Convergence history. Staircase behavior

An alternative (indefinite) preconditioner - work in progress:

$$\mathcal{P} = \begin{bmatrix} 0 & K + \omega\, M \\ K + \omega\, M & -\frac{1}{\nu} M \end{bmatrix}$$



Similar results for the Distributed optimal control for the time-periodic Stokes equations

Final remarks

- Much is known about the behavior of structured preconditioners for well established problems and formulations
- New problems provide new challenges
- Understanding the underlying Linear algebra may be key

References for this talk

W. Krendl, V. Simoncini and W. Zulehner, *Stability Estimates and Structural Spectral Properties of Saddle Point Problems*, submitted, 2012.

M. A. Olshanskii, and V. Simoncini, *Acquired clustering properties and solution of certain saddle point systems*, SIAM. J. Matrix Anal. and Appl. 31(5), 2754-2768 (2010)

M. Benzi, G.H. Golub and J. Liesen, *Numerical Solution of Saddle Point Problems*, Acta Numerica, 14, 1-137 (2005)

K.A. Mardal and R. Winther, *Preconditioning discretizations of systems of partial differential equations*, Numer. Linear Algebra with Appl., 18, 1-40 (2011)