Spectral Properties of Saddle Point Linear Systems and Relations to Iterative Solvers

Part II: Iterative solvers

V. Simoncini

Dipartimento di Matematica, Università di Bologna

valeria@dm.unibo.it

Outline of the 3-hour Presentation

- Schematic presentation of certain algebraic preconditioners (Yesterday)
- Iterative solvers. Some (hopefully) helpful considerations... (Today)
- Spectral analysis of nonsymmetric preconditioners (Tomorrow)

The standard solvers

Krylov subspace iterative solvers for $\mathcal{M}x = b$:

- $\mathcal{M}$ symmetric and positive definite $\Rightarrow$ (P)CG
- $\mathcal{M}$ symmetric indefinite $\Rightarrow$ (P)MINRES, (P)SYMLQ
- $\mathcal{M}$ nonsymmetric $\Rightarrow$ (P)GMRES, (P)BiCGSTAB(ℓ)

The standard solvers

Krylov subspace iterative solvers for $\mathcal{M}x = b$:

- $\mathcal{M}$ symmetric and positive definite $\Rightarrow$ (P)CG
- $\mathcal{M}$ symmetric indefinite $\Rightarrow$ (P)MINRES, (P)SYMLQ
- $\mathcal{M}$ nonsymmetric $\Rightarrow$ (P)GMRES, (P)BiCGSTAB(ℓ)

More specific issues:

$\star$ Convergence and clustering

$\star$ Stagnation

$\star$ Symmetry wrto H-inner product (H spd)

$\star$ Symmetry wrto J-inner product (J not spd)

Convergence... CG

CG: minimum error method (in energy norm). For $\mathcal{M}$ spd ($x_0 = 0$)

$$\min_{x \in K_k(\mathcal{M},b)} \|x_\star - x\|_{\mathcal{M}} \quad \leq \quad 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k \|x_\star\|_{\mathcal{M}}$$

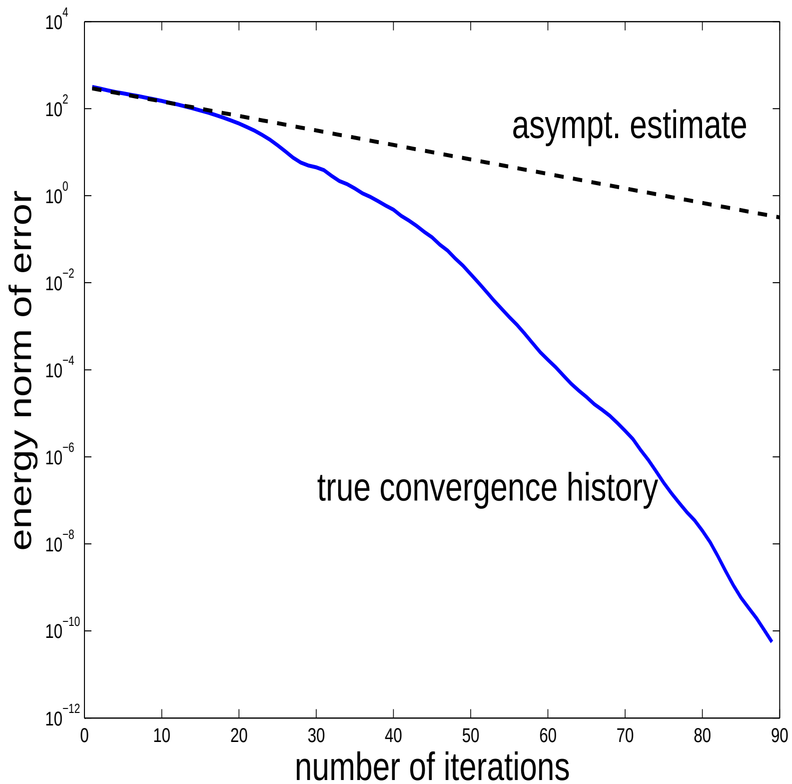
with $\kappa = \lambda_{\max}(\mathcal{M})/\lambda_{\min}(\mathcal{M})$

Convergence... CG

CG: minimum error method (in energy norm). For $\mathcal{M}$ spd ($x_0 = 0$)

$$\min_{x \in K_k(\mathcal{M}, b)} \|x_\star - x\|_{\mathcal{M}} \quad \leq \quad 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \|x_\star\|_{\mathcal{M}}$$

with $\kappa = \lambda_{\max}(\mathcal{M}) / \lambda_{\min}(\mathcal{M})$



Convergence...

GMRES: minimum residual method

$$\min_{x \in K_k(\mathcal{M},b)} \|b - \mathcal{M}x\|, \qquad (x_0 = 0)$$

x_k minimizer.

For $w \in K_k(\mathcal{M}, v)$, , $w = q_{k-1}(\mathcal{M})b$. Then

$$r_k = b - \mathcal{M}x_k = b - \mathcal{M}q_{k-1}(\mathcal{M})b = p_k(\mathcal{M})b$$

Convergence...

GMRES: minimum residual method

$$\min_{x \in K_k(\mathcal{M},b)} \|b - \mathcal{M}x\|, \qquad (x_0 = 0)$$

x_k minimizer.

For $w \in K_k(\mathcal{M}, v)$, $, w = q_{k-1}(\mathcal{M})b$. Then

$$r_k = b - \mathcal{M}x_k = b - \mathcal{M}q_{k-1}(\mathcal{M})b = p_k(\mathcal{M})b$$

Some "intuitive" consequences:

- $\mathcal{M}$ (diag.ble) has few distinct eigs $\Rightarrow$ fast convergence

  (minimal polynomial of $\mathcal{M}$ wrto b has low degree)

Convergence...

GMRES: minimum residual method

$$\min_{x \in K_k(\mathcal{M},b)} \|b - \mathcal{M}x\|, \qquad (x_0 = 0)$$

x_k minimizer.

For $w \in K_k(\mathcal{M}, v)$, $, w = q_{k-1}(\mathcal{M})b$. Then

$$r_k = b - \mathcal{M}x_k = b - \mathcal{M}q_{k-1}(\mathcal{M})b = p_k(\mathcal{M})b$$

Some "intuitive" consequences:

- $\mathcal{M}$ (diag.ble) has few distinct eigs $\Rightarrow$ fast convergence

  (minimal polynomial of $\mathcal{M}$ wrto b has low degree)

- Spectral clustering is beneficial $\Rightarrow$ select appropriate preconditioner

...and clustering

Will any spectral clustering do the job ?

Residual: $r_k = p_k(\mathcal{M})b$ with $r_0 = b$ $\Rightarrow$ $p_k(0) = 1$

$\Rightarrow$ Spectrum away from zero

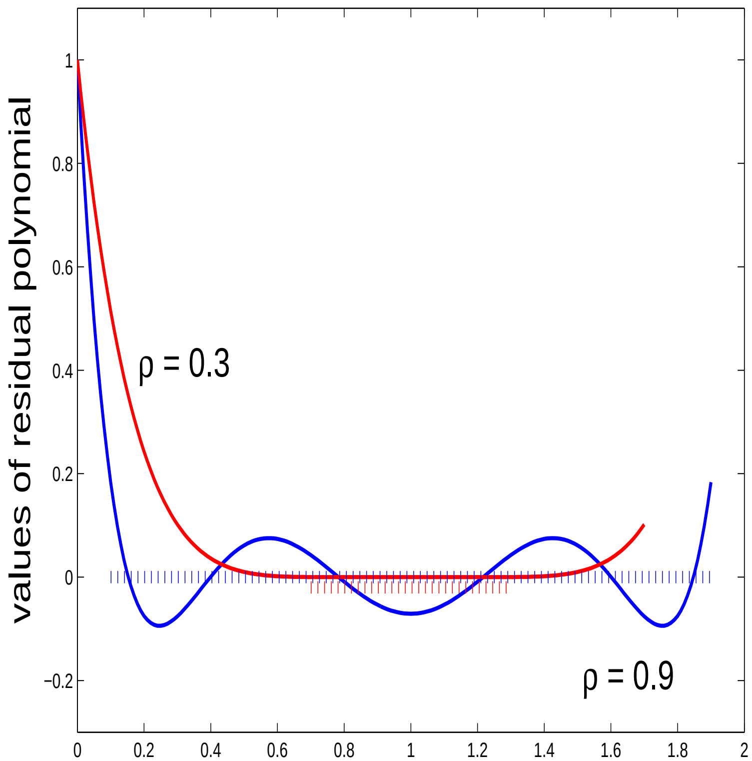
...and clustering

Will any spectral clustering do the job ?

Residual: $r_k = p_k(\mathcal{M})b$ with $r_0 = b$ $\Rightarrow$ $p_k(0) = 1$

$\Rightarrow$ Spectrum away from zero

An example: $\sigma(\mathcal{M}) \subset [1-\rho, 1+\rho]$ $p_k(\lambda)$:



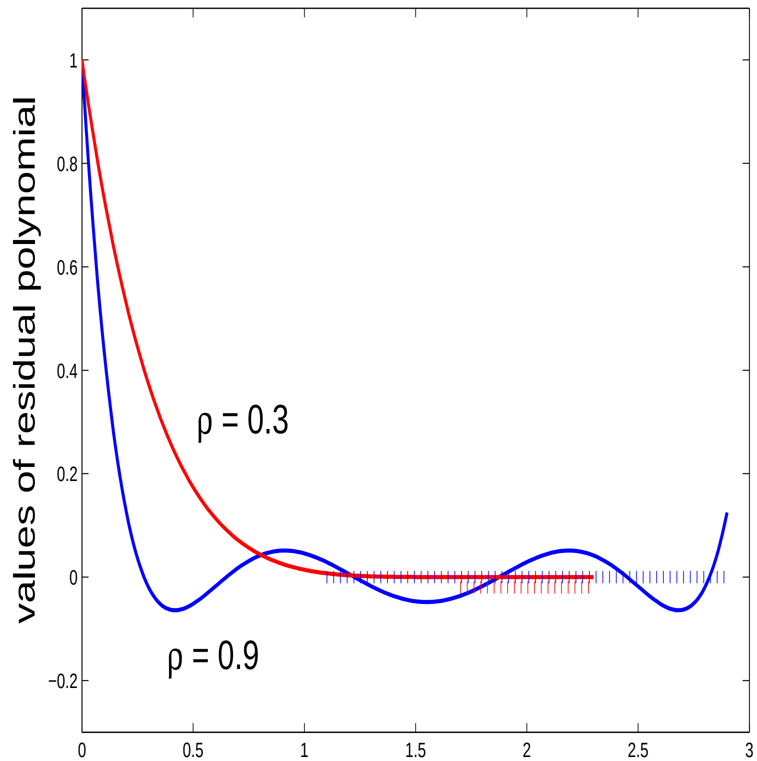
...and clustering

Will any spectral clustering do the job ?

Residual: $r_k = p_k(\mathcal{M})b$ with $r_0 = b$ $\Rightarrow$ $p_k(0) = 1$

$\Rightarrow$ Spectrum away from zero

A second example: $\sigma(\mathcal{M}) \subset [2 - \rho, 2 + \rho]$ $p_k(\lambda)$:



...and a good clustering

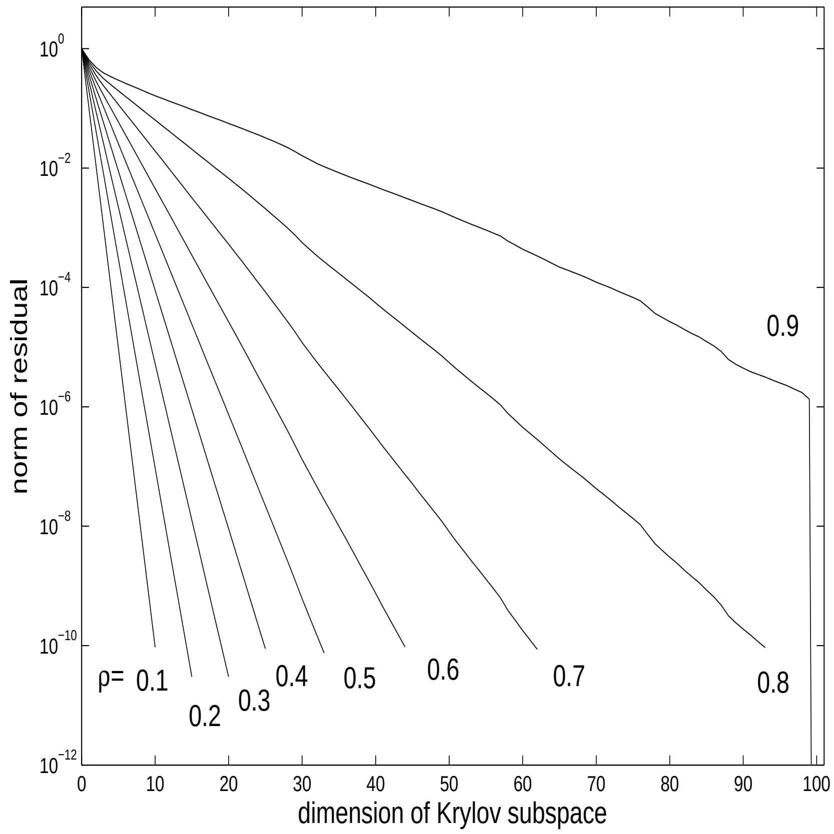
A nonsym example: $\mathcal{M} = I + \rho Q$, Q unitary $(\sigma(\mathcal{M}) \subset D(1, \rho))$

...and a good clustering

A nonsym example: $\mathcal{M} = I + \rho Q$, Q unitary $(\sigma(\mathcal{M}) \subset D(1,\rho))$

GMRES rate: ρ^k

For CG, rate: $\left(\frac{\rho}{1+\sqrt{1-\rho^2}}\right)^k$

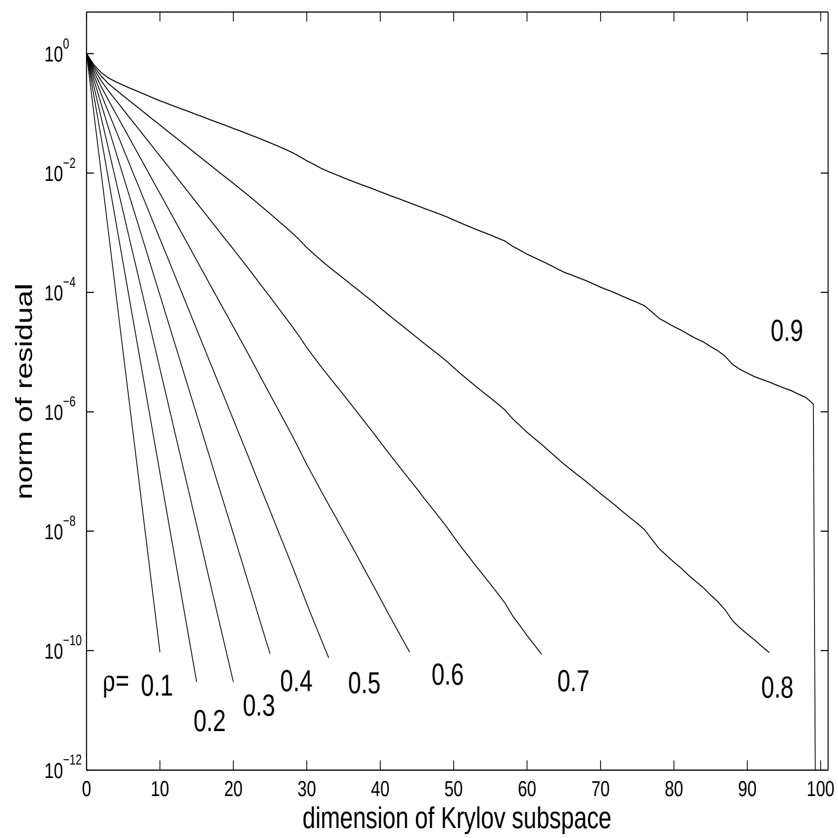


...and a good clustering

A nonsym example: $\mathcal{M} = I + \rho Q$, Q unitary $(\sigma(\mathcal{M}) \subset D(1,\rho))$

GMRES rate: ρ^k

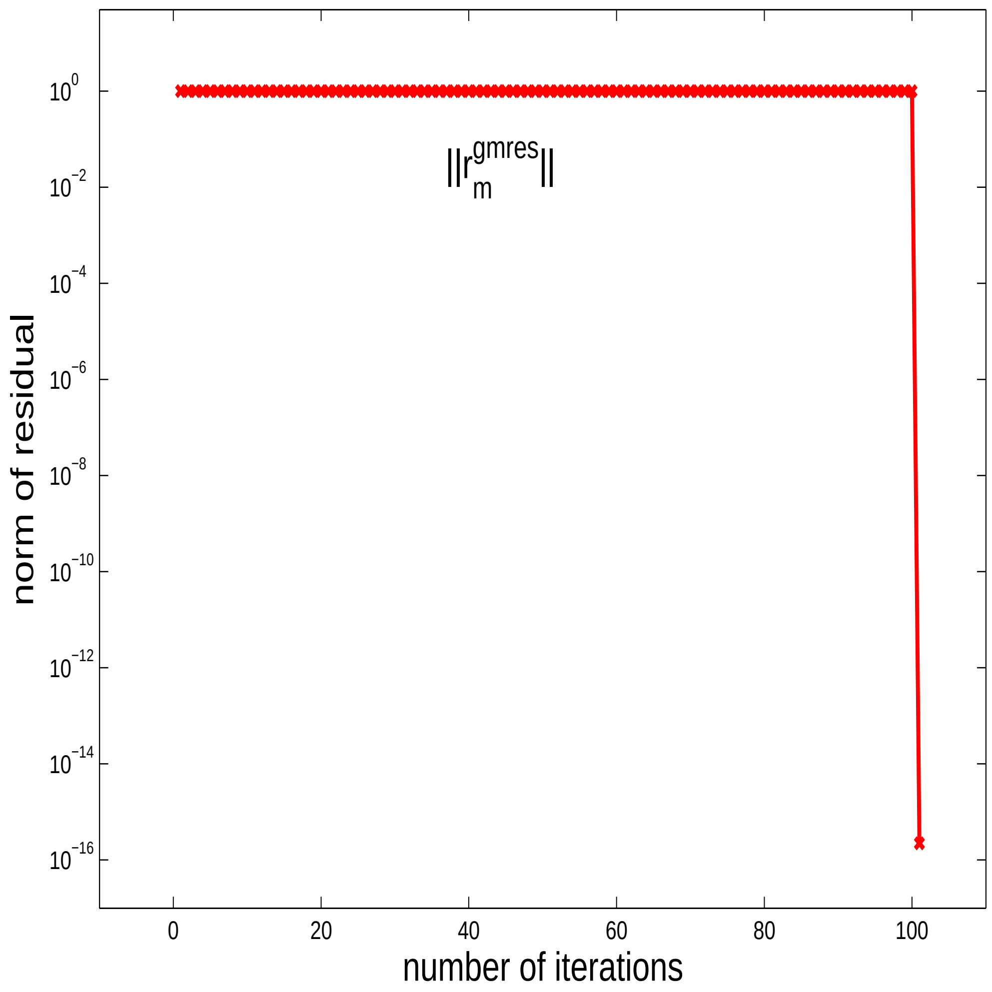
For CG, rate: $\left(\frac{\rho}{1+\sqrt{1-\rho^2}}\right)^k$



If $\sigma(\mathcal{M}) \subset D(2,\rho)$, GMRES has rate $\left(\frac{\rho}{2}\right)^k$

Stagnation of GMRES

A is 100×100



Conditions for non-Stagnation of GMRES

If $\alpha = \lambda_{\min}(\frac{1}{2}(\mathcal{M} + \mathcal{M}^T)) > 0$, then

$$\|r_k\| \leq \left(1 - \frac{\alpha^2}{\|\mathcal{M}\|^2}\right)^{\frac{k}{2}} \|b\| < \|b\|$$

Note: $\mathcal{M}$ must be positive real

Conditions for non-Stagnation of GMRES

If $\alpha = \lambda_{\min}(\frac{1}{2}(\mathcal{M} + \mathcal{M}^T)) > 0$, then

$$\|r_k\| \leq \left(1 - \frac{\alpha^2}{\|\mathcal{M}\|^2}\right)^{\frac{k}{2}} \|b\| < \|b\|$$

Note: $\mathcal{M}$ must be positive real

New condition: Let $H = \frac{1}{2}(\mathcal{M} + \mathcal{M}^T)$, $\quad S = \frac{1}{2}(\mathcal{M} - \mathcal{M}^T)$

If H is nonsingular and $\|SH^{-1}\| < 1$ then there exists (computable) c with $0 < c < 1$ s.t.

$$\|r_2\| \leq c\|b\| < \|b\|$$

(same result for S nonsingular and $\|HS^{-1}\| < 1$)

An additional result

With the same tools:

If $H^2 + S^2$ nonsingular and $\|(HS + SH)(H^2 + S^2)^{-1}\| < 1$ (*)

then there exists c with $0 < c < 1$ s.t.

$$\|r_4\| \leq c\|b\| < \|b\|$$

An additional result

With the same tools:

If $H^2 + S^2$ nonsingular and $\|(HS + SH)(H^2 + S^2)^{-1}\| < 1$ (*)
then there exists c with $0 < c < 1$ s.t.

$$\|r_4\| \leq c\|b\| < \|b\|$$

An example

$$\mathcal{M} = \begin{bmatrix} A & B^T \\ -B & 0 \end{bmatrix}$$

A symmetric

B full rank

Note: $H = \frac{1}{2}(\mathcal{M} + \mathcal{M}^T)$ and $S = \frac{1}{2}(\mathcal{M} - \mathcal{M}^T)$ are singular

Assume $A = \mu I$. If μ s.t. (*) holds, then no full stagnation

Changing the inner product. Occurrence

- Minimize quantity in a different inner product
- Monitor convergence in agreement with the continuous problem
- Exploit "non-canonical" symmetries of the coeff. matrix

Symmetry wrto Euclidean inner product

$$\mathcal{M}x = b, \qquad \mathcal{M} \text{ spd}$$

Classical CG: $\quad (u, v) = u^T v$

Given x_0

$r_0 = b - \mathcal{M}x_0$, $p_0 = r_0$

for $i = 0, 1, \ldots$

$\quad \alpha_i = \frac{(r_i, r_i)}{(p_i, \mathcal{M}p_i)}$

$\quad x_{i+1} = x_i + p_i \alpha_i$

$\quad r_{i+1} = r_i - \mathcal{M}p_i \alpha_i$

$\quad \beta_{i+1} = \frac{(r_{i+1}, \mathcal{M}p_i)}{(p_i, \mathcal{M}p_i)}$

$\quad p_{i+1} = r_i + p_i \beta_{i+1}$

Symmetry wrto H-inner product (H spd)

$$\mathcal{M}x = b$$

Assume there exists H spd such that $H\mathcal{M}$ is also spd

H-sym CG: $(u, v)_H = u^T H v$

Given x_0

$r_0 = b - \mathcal{M}x_0$, $p_0 = r_0$

for $i = 0, 1, \ldots$

$$\alpha_i = \frac{(r_i, r_i)_H}{(p_i, \mathcal{M}p_i)_H}$$

$$x_{i+1} = x_i + p_i \alpha_i$$

$$r_{i+1} = r_i - \mathcal{M}p_i \alpha_i$$

$$\beta_{i+1} = \frac{(r_{i+1}, \mathcal{M}p_i)_H}{(p_i, \mathcal{M}p_i)_H}$$

$$p_{i+1} = r_i + p_i \beta_{i+1}$$

Application to Saddle-point systems. The "minus-signed" matrix

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

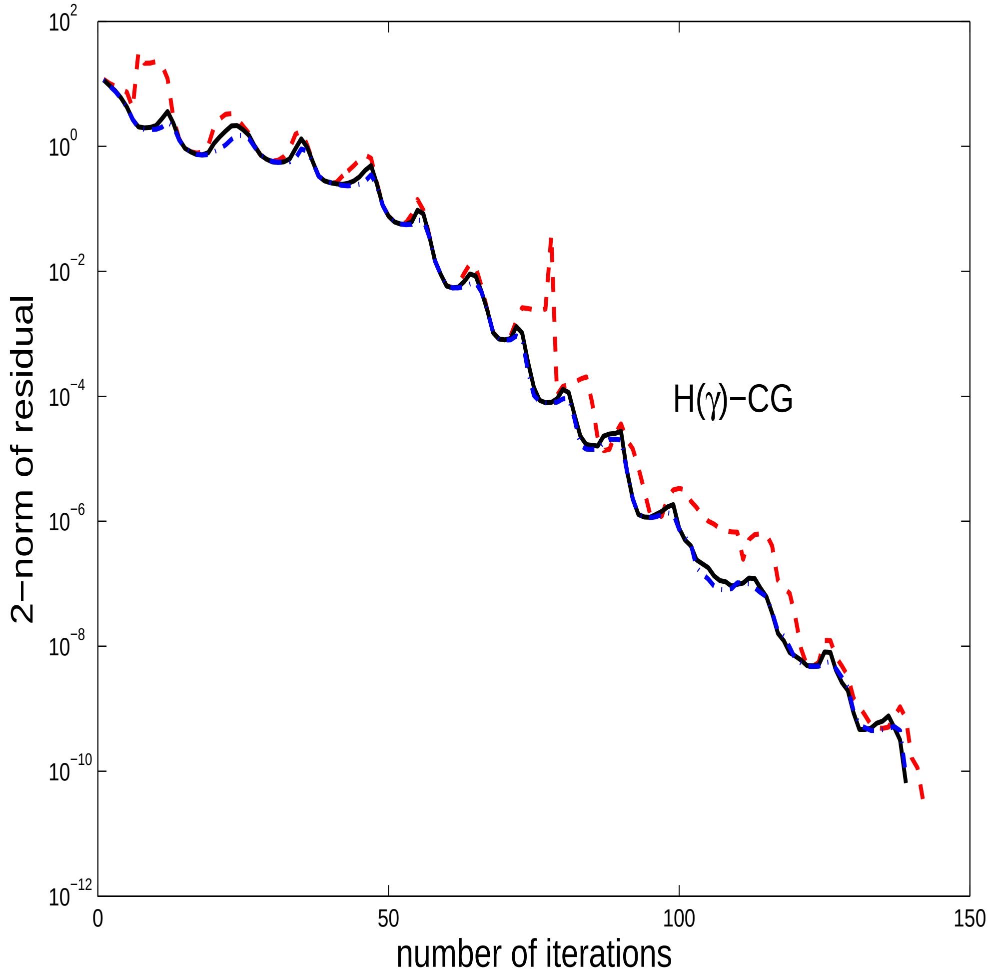
⋆ $\mathcal{M}_-$ is $\mathcal{H}(\gamma)$-symmetric, with

$$\mathcal{H}(\gamma) = \begin{bmatrix} A - \gamma I & B^T \\ B & \gamma I \end{bmatrix}$$

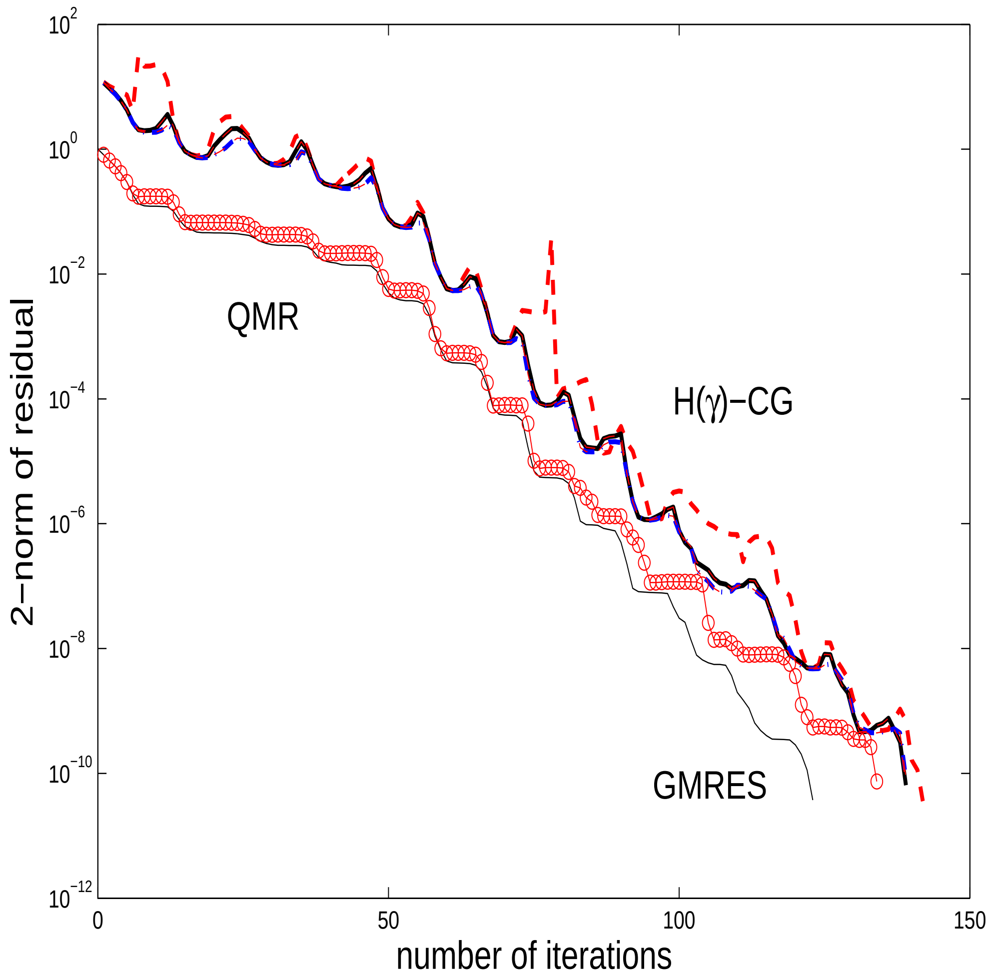
⋆ Let $\gamma_\star = \frac{1}{2}\lambda_{\min}(A)$.
If $\lambda_{\min}(A) > 4\lambda_{\max}(B^T A^{-1} B)$ then $\mathcal{H}(\gamma_\star)$ is spd

⋆ ...and $\mathcal{H}(\gamma_\star)\mathcal{M}$ is also spd

An example. Stokes with mixed b.c. on the unit square



An example. Stokes with mixed b.c. on the unit square



Symmetry wrto an indefinite inner product

Given J symmetric nonsing, $\mathcal{M}$ is J-symmetric if

$$\mathcal{M}^T J = J\mathcal{M}$$

J-inner product: $(x, y)_J = x^T J y$

Example:

$$\mathcal{M}_- = \begin{bmatrix} A & B^T \\ -B & O \end{bmatrix}, \qquad J = \begin{bmatrix} I & O^T \\ O & -I \end{bmatrix},$$

Simplification of Lanczos-type procedure (e.g. QMR):

only one matrix-vector product (by $\mathcal{M}$) per iteration

Another example: Indefinite (Constraint) Preconditioner

$$\mathcal{M} = \begin{bmatrix} A & B^T \\ B & O \end{bmatrix}, \qquad \mathcal{P} = \begin{bmatrix} \widetilde{A} & B^T \\ B & O \end{bmatrix}$$

- ⋆ $\mathcal{M}\mathcal{P}^{-1}$ nonsym (nondiagonalizable!)
- ⋆ $\mathcal{M}\mathcal{P}^{-1}$ is $\mathcal{P}^{-1}$-symmetric $\Rightarrow$ Simplified Lanczos
- ⋆ Applying $\mathcal{P}^{-1}$ may be expensive... Inexact preconditioning
- ⋆ Case $C \neq O$ more challenging $\Rightarrow$ Class of preconditioners

Comparing $H(\gamma)$-CG and Simplified Lanczos

- $H(\gamma)$-CG involves reality condition
- $H(\gamma)$-CG involves estimating γ
- $H(\gamma)$-CG not clear how to precondition
- $H(\gamma)$-CG convergence clear (exact arithm)