Recent advances in approximation using Krylov subspaces

V. Simoncini

Dipartimento di Matematica, Università di Bologna

valeria@dm.unibo.it

The problem

Solve

$$Ax = b \qquad n \times n \qquad n \gg 1000$$

or

$$Ax = \lambda x, \qquad \|x\| = 1,$$

using Krylov subspace type methods

$$K_m(A, v) = \text{span}\{v, Av, \ldots, A^{m-1}v\} \qquad v = b - Ax_0,$$

when A is either:

- Not known exactly
- Computationally expensive to deal with

Many applications in Scientific Computing

- $Av \equiv F(v)$ function (linear in v)
- Shift-and-Invert procedures for interior eigenvalues
- Schur complement: $A = B^T S^{-1} B \quad S$ expensive to invert
- Preconditioned system: $AP^{-1}x = b$, where

$$P^{-1}v_i \approx P_i^{-1}v_i$$

- etc.

The exact approach

Key relation in Krylov subspace methods:

$$AV_m = V_{m+1}\underline{H}_m \qquad v = V_{m+1}e_1\beta \qquad \underline{H}_m = \begin{bmatrix} H_m \\ h_{m+1,m}e_m^T \end{bmatrix}$$

System:

$$x_m \in \mathrm{Range}(V_m) = K_m(A,b) \quad \Rightarrow \quad x_m = V_m y_m \qquad (x_0 = 0)$$

Eigenproblem:

(θ, z) eigenpair of H_m $\quad \Rightarrow \quad$ $(\theta, V_m z)$ Ritz approximation to (λ, x)

The exact approach. The actual key quantity

System:

For $r_m = b - Ax_m$:

$$r_m = b - AV_m y_m = b - V_{m+1}\underline{H}_m y_m = V_{m+1}(e_1\beta - \underline{H}_m y_m)$$

$$AV_m y_m = V_{m+1}\underline{H}_m y_m$$

Note: all components of y_m may change as m grows

Eigenproblem: (θ, z) eigenpair of H_m:

$$r_m = \theta V_m z - AV_m z = \theta V_m z - V_{m+1}\underline{H}_m z = v_{m+1}h_{m+1,m}e_m^T z$$

The inexact key relation

$$AV_m = V_{m+1}\underline{H}_{m} + \underbrace{F_m}_{[f_1, f_2, \ldots, f_m]}$$

F_m *error* matrix:

- Inexact A (all cases described earlier)
- Finite Precision Computation
- Deflation strategies in block methods

How large can F_m be allowed to be?

$$r_m = b - AV_m y_m = b - V_{m+1}\underline{H}_m y_m - F_m y_m = \underbrace{V_{m+1}(e_1\beta - \underline{H}_m y_m)}_{\text{computed residual}} - F_m y_m$$

$$r_m = \theta V_m z - AV_m z = v_{m+1} h_{m+1,m} e_m^T z - F_m z$$

Size of the error matrix F_m

$$AV_m = V_{m+1} H_{m+1} + \underbrace{F_m}_{[f_1, f_2, \ldots, f_m]}$$

In practice:

$$AV_m y = V_{m+1} H_{m+1} y + F_m y$$

The correct question is: How large can $F_m y$ be allowed to be?

Note: y is given and $\|f_i\|$'s can be controlled

A dynamic setting

$$AV_m y = V_{m+1} H_{m+1} y + F_m y$$

$$F_m y = [f_1, f_2, \ldots, f_m] \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_m \end{bmatrix} = \sum_{i=1}^{m} f_i \eta_i$$

◇ The terms $f_i \eta_i$ need to be small:

$$\|f_i \eta_i\| < \frac{1}{m}\epsilon \quad \forall i \quad \Rightarrow \quad \|F_m y\| < \epsilon$$

◇ η_i small $\quad \Rightarrow \quad$ f_i is allowed to be large

Linear systems: The structure of the solution

$y_m = [\eta_1; \eta_2; \ldots; \eta_m]$ depends on the chosen method, e.g.

- Petrov-Galerkin (e.g. GMRES): $\quad y_m = \mathrm{argmin}_y \| e_1 \beta - \underline{H}_m y \|$,

$$|\eta_i| \le \frac{1}{\sigma_{\min}(\underline{H}_m)} \|\tilde{r}_{i-1}\|$$

$\tilde{r}_{i-1}$: GMRES computed residual at iteration $i-1$.

Simoncini & Szyld, SISC 2003 (see also Sleijpen & van den Eshof, SIMAX 2004)

Analogous result for Galerkin methods (e.g. FOM)

Eigenproblem: The structure of the Ritz pair

Ritz approximation:

$$(\theta, z) \quad \text{eigenpair of} \quad H_m$$

$z = [\eta_1; \eta_2; \ldots; \eta_m]$,

$$|\eta_i| \leq \frac{2}{\delta_{m,i}} \|r_{i-1}\|$$

$\delta_{m,i}$ quantity related to the spectral gap of θ with H_m

r_{i-1}: Computed eigenresidual at iteration $i-1$

Analogous results for Harmonic Ritz values and Lanczos approx.

Simoncini, SINUM To appear

A practical example: Inexact coefficient matrix

At iteration i: $A \cdot v_i$ not performed exactly $\Rightarrow$ $(A + E_i) \cdot v_i$

$\|E_i\|$ (or $\|E_i v_i\|$) can be monitored

(e.g. Schur complement, Multipole methods, Multilevel methods, etc.)

Arnoldi relation: $V_m = [v_1, v_2, \ldots, v_m]$

$$
\begin{aligned}
[(A + E_1)v_1, (A + E_2)v_2, \ldots, (A + E_m)v_m] &= V_{m+1} H_m \\
AV_m + \underbrace{[E_1 v_1, E_2 v_2, \ldots, E_m v_m]}_{-F_m} &= V_{m+1} H_m
\end{aligned}
$$

True vs. computed residuals:

$$
r_m = b - AV_m y_m = V_{m+1}(e_1 \beta - \underline{H}_m y_m) - F_m y_m
$$

Relaxing the inexactness in A

$$r_m = b - AV_m y_m = V_{m+1}(e_1\beta - \underline{H}_m y_m) - F_m y_m$$

with $\quad (A + E_i)v_i \qquad F_m = [E_1 v_1, E_2 v_2, \ldots, E_m v_m]$

GMRES: If (Similar result for FOM)

$$\|E_i\| \le \frac{\sigma_{\min}(\underline{H}_m)}{m} \frac{1}{\|\tilde{r}_{i-1}\|}\varepsilon \quad i = 1, \ldots, m$$

then

$$\|F_m y_m\| \le \varepsilon \quad \Rightarrow \quad \|r_m - V_{m+1}(e_1\beta - \underline{H}_m y_m)\| \le \varepsilon$$

$\tilde{r}_{i-1}$: GMRES computed residual at iteration $i - 1$

An example: Schur complement

$$\underbrace{B^T S^{-1} B}_{A} x = b \qquad\qquad y_i \leftarrow B^T S^{-1} B v_i$$

Inexact matrix-vector product:

$$\begin{cases} \text{Solve } S w_i = B v_i \\ \text{Compute } y_i = B^T w_i \end{cases} \quad \overset{\text{Inexact}}{\Longrightarrow} \quad \begin{cases} \text{Approx solve } S w_i = B v_i \quad \Rightarrow \widehat{w}_i \\ \text{Compute } \widehat{y}_i = B^T \widehat{w}_i \end{cases}$$

$w_i = \widehat{w}_i + \boldsymbol{\epsilon}_i$ $\qquad$ $\boldsymbol{\epsilon}_i$ error in inner solution $\qquad$ so that

$$A v_i \qquad \to \qquad B^T \widehat{w}_i = \underbrace{B^T w_i}_{A v_i} - \underbrace{B^T \boldsymbol{\epsilon}_i}_{-E_i v_i} = (A + E_i) v_i$$

Relaxation strategy for inner stopping criterion

$$Av_i \qquad \to \qquad B^T\hat{w}_i = \underbrace{B^T w_i}_{Av_i} - \underbrace{B^T \boldsymbol{\epsilon}_i}_{-E_i v_i} = (A + E_i)v_i$$

$\|E_i v_i\|$ can be monitored through the inner residual:

$$\|E_i v_i\| \le \|B^T S^{-1}\|\,\|r_k^{\text{inner}}\|, \qquad r_k^{\text{inner}} \text{ inner residual at it. } k$$

This, together with the requirement

$$\|E_i\| \le \frac{\sigma_{\min}(\underline{H}_m)}{m}\frac{1}{\|\tilde{r}_{i-1}\|}\varepsilon \quad i = 1, \dots, m$$

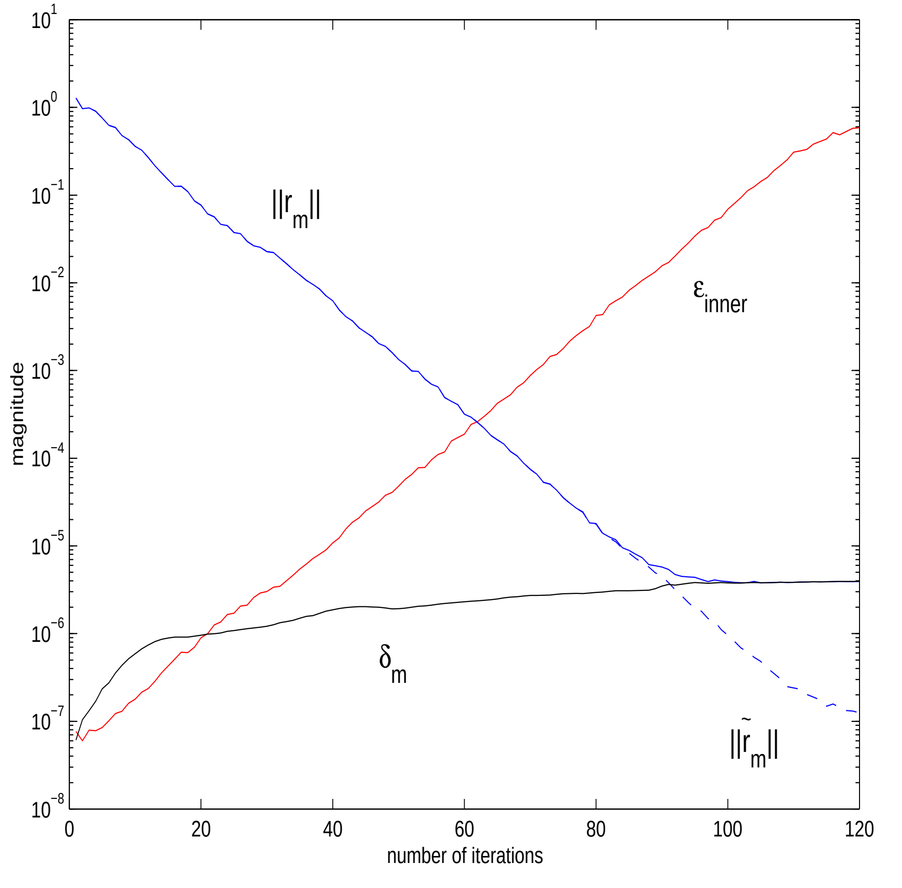
allows to relax the accuracy with which we solve $Sw_i = Bv_i$ at each iteration while outer convergence takes place

Numerical experiment: Schur complement

$$\underbrace{B^T S^{-1} B}_{A} x = b \qquad \text{at each it. } i \text{ solve} \quad S w_i = B v_i$$

Inexact FOM

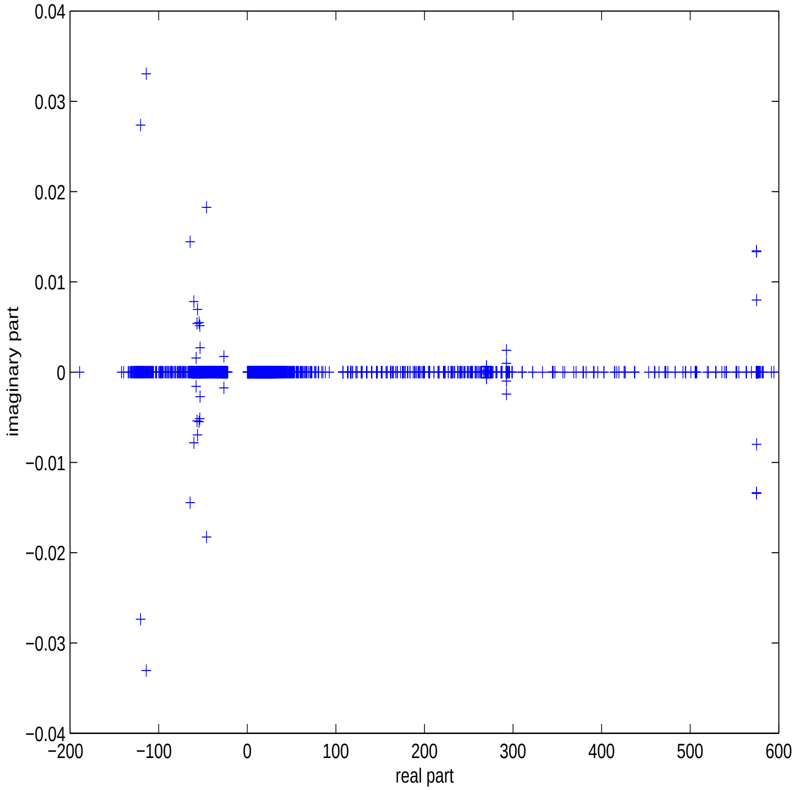
$$\delta_m = \|r_m - (b - V_{m+1} \underline{H}_m y_m)\|$$

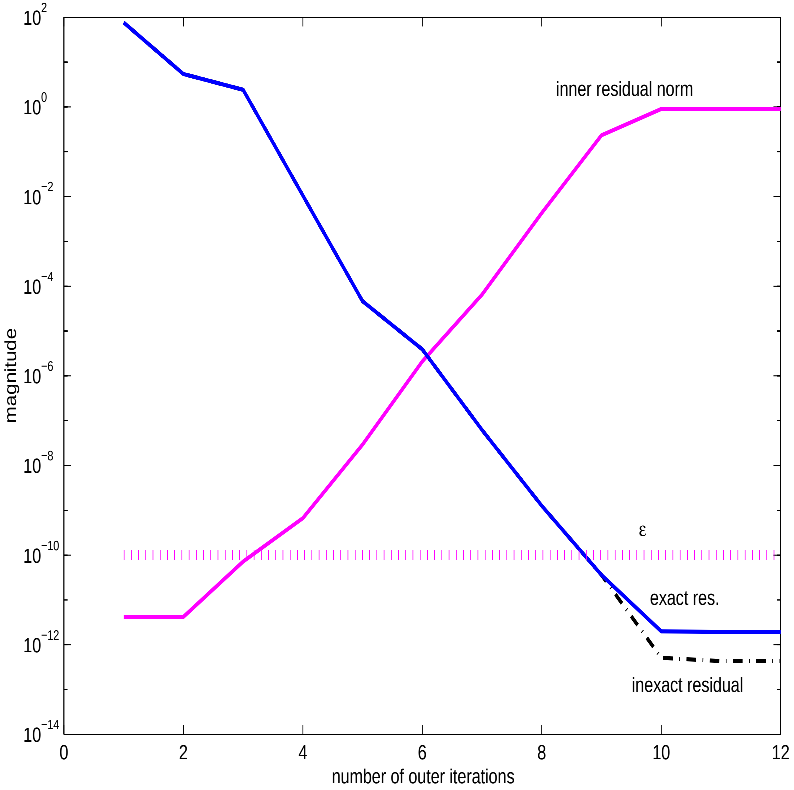


Eigenproblem

Inverted Arnoldi: $Ax = \lambda x$ Find $\min |\lambda|$ $y \leftarrow A^{-1}v$

Matrix SHERMAN5





Problems to be faced

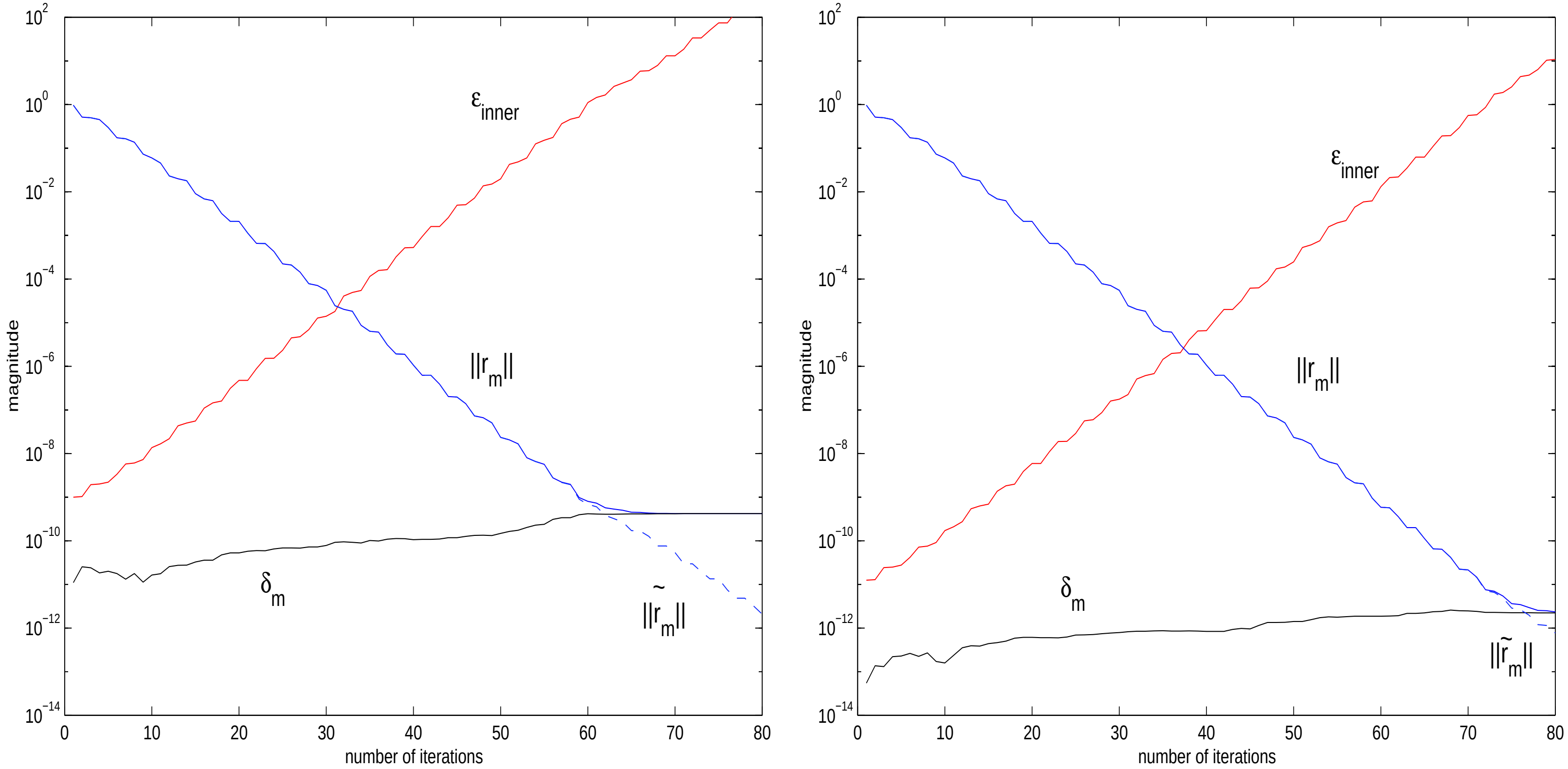
- Make the inexactness criterion practical

$$\|E_i\| \leq \frac{\sigma_{\min}(H_{m_\star})}{m_\star} \frac{1}{\|\tilde{r}_{i-1}\|}\varepsilon \qquad \Rightarrow \qquad \|E_i\| \leq \ell_{m_\star} \frac{1}{\|\tilde{r}_{i-1}\|}\varepsilon$$

(series of CERFACS tr. of Bouras, Frayssé, Giraud, 2000)

- What is the convergence behavior?
- What if original A was symmetric?

Selecting $\ell_{m_\star}$: system $A\mathcal{P}^{-1}x = b$



Left: $\ell_{m_\star} = 1$

Right: estimated $\ell_{m_\star}$

Convergence behavior

Does the **inexact** procedure behave as if $\|E_i\| = 0$?

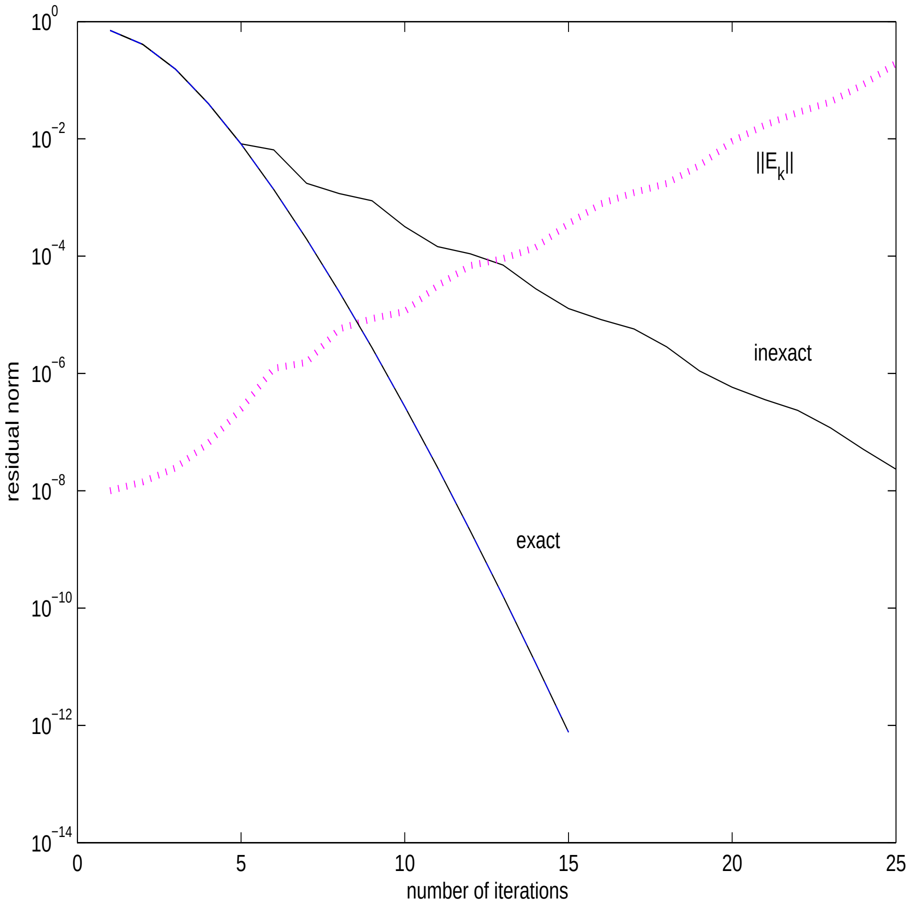
The Sleijpen & van den Eshof's example:

Exact vs. Inexact GMRES

$b = e_1$

E_i random entries

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 2 & 0 & \cdots & 0 \\ 0 & 1 & 3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 100 \end{bmatrix}$$



Inexactness and convergence

$$Av_i \qquad \rightarrow \qquad (A + E_i)v_i$$

For general A and b convergence is the same as exact A

Problems for:

- Sensitive A (highly nonnormal)
- Special starting vector / right-hand side

⋆ Superlinear convergence as for A (Simoncini & Szyld, SIREV 2005)

Flexible preconditioning

$$AP^{-1}\widehat{x} = b \qquad x = P^{-1}\widehat{x}$$

Flexible:

$P^{-1}v_i \;\rightarrow\; P_i^{-1}v_i, \quad \widehat{x}_m \in \mathrm{span}\{v_1, AP_1^{-1}v_1, AP_2^{-1}v_2, \ldots, AP_{m-1}^{-1}v_{m-1}\}$

Directly recover x_m (Saad, 1993):

$$[P_1^{-1}v_1, P_2^{-1}v_2, \ldots, P_m^{-1}v_m] = Z_m \quad \Rightarrow \quad x_m = Z_m y_m$$

$\Rightarrow$ Inexact framework but exact residual

A practical example

$$\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix} \qquad \mathcal{P} = \begin{bmatrix} I & 0 \\ 0 & B^T B \end{bmatrix}$$

Application of $\mathcal{P}^{-1}$ corresponds to solves with $B^T B$

$$\Downarrow$$

$\widetilde{\mathcal{P}} \quad \Rightarrow$ Use CG to solve systems with $B^T B$

Variable inner tolerance: At each outer iteration m,

$$\|r_k^{inner}\| \leq \frac{\ell_{m_\star}}{\|r_{m-1}^{outer}\|}\varepsilon$$

Electromagnetic 2D problem

Outer tolerance: 10^{-8}

$$\|r_k^{inner}\| \leq \frac{\ell_{m_\star}}{\|r_{m-1}^{outer}\|}\varepsilon_0 \equiv \varepsilon$$

Elapsed Time

Pb. Size	Fixed Inner Tol $\varepsilon = 10^{-10}$	Var. Inner Tol. $\varepsilon = 10^{-10}/\|r\|$	Var. Inner Tol. $\varepsilon = 10^{-12}/\|r\|$
3810	17.0 (54)	11.4 (54)	14.7 (54)
9102	82.9 (58)	62.8 (58)	70.7 (58)
14880	198.4 (54)	156.5 (54)	170.1 (54)

Structural Dynamics

$$(\mathcal{A} + \sigma\mathcal{B})x = b$$

Solve for many σ's simultaneously $\quad \Rightarrow \quad (\mathcal{A}\mathcal{B}^{-1} + \sigma I)\widehat{x} = b$

(Perotti & Simoncini 2002)

Inexact solutions with $\mathcal{B}$ at each iteration:

	Prec. Fill-in 5		Prec. Fill-in 10	
	e-time [s]	# outer its	e-time [s]	# outer its
Tol 10^{-6}	14066	296	13344	289
Dynamic Tol	11579	301	11365	293

20 % enhancement with tiny change in the code

Inexactness when A symmetric

A symmetric $\quad\Rightarrow\quad$ $A + E_i$ nonsymmetric

- Assume $V_m^T V_m = I \quad \rightarrow \quad H_m$ upper Hessenberg
- Wise implementation of short-term recurr. /truncated methods (V_m non-orth. $\rightarrow W_m$, $\quad H_m$ tridiag./banded $\rightarrow T_m$)
  - Inexact short-term recurrence system solvers

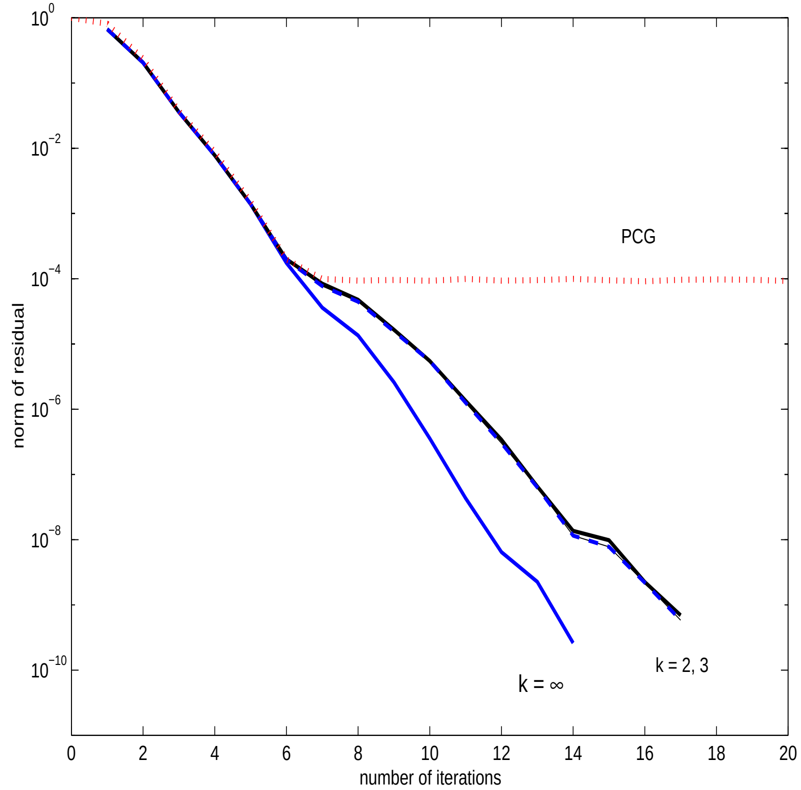
    (Golub & Overton 1988, Golub & Ye 1999, Notay 2000, Sleijpen & van den Eshof tr.2002, . . .)
  - Truncated methods (Simoncini & Szyld, Num. Math. To appear)
  - Inexact symmetric eigensolvers

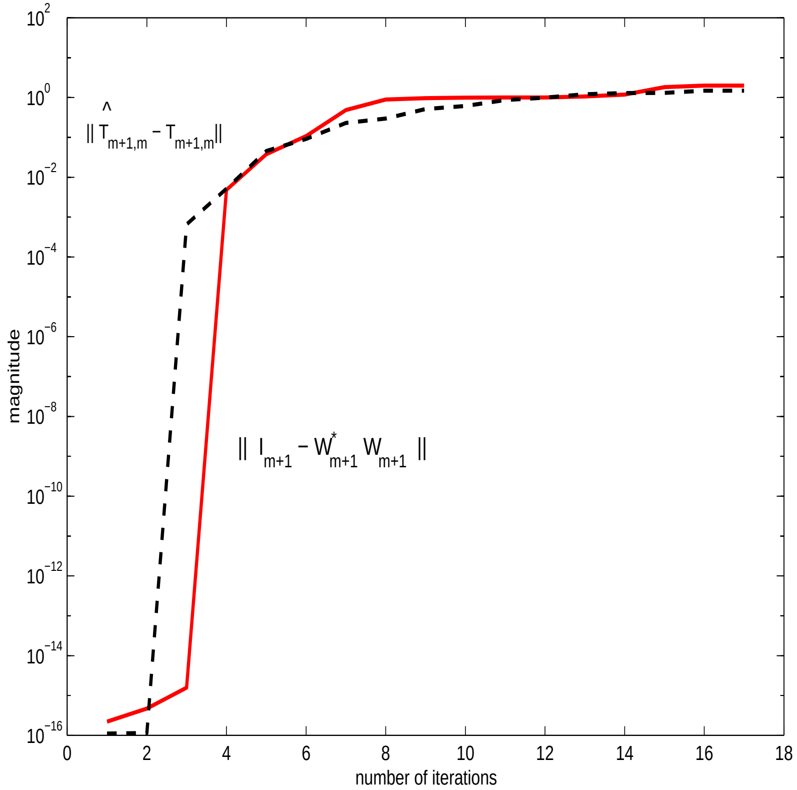
    (Lai, Lin & Lin 1997, Golub & Ye 2000, Golub, Zhang & Zha 2000, Notay 2002, . . .)

$$Ax = b \qquad A \quad \text{sym. (2D Laplacian)}$$

Preconditioner:

$\mathcal{P}$ nonsymmetric perturbation (10^{-5}) of Incomplete Cholesky





Application: Computation of the exponential

A symmetric negative semidefinite (large dimension), v s.t. $\|v\| = 1$,

$$\exp(A)v \approx x_m = V_m \exp(H_m)e_1 \equiv V_m y_m$$

Problem: Find preconditioner for A to speed up convergence

Hochbruck & van den Eshof (SISC To appear):

Determine $x_m \approx \exp(A)v$ as

$$x_m = V_m y_m \in K_m((I - \gamma A)^{-1}, v) \qquad \text{for scalar } \gamma$$

$\Rightarrow$ $y_m = \exp(H_m)e_1$ has a structured decreasing pattern

(Lopez & Simoncini, tr. 2005)

Conclusions

- A may be replaced by $A + E_i$ with $\|E_i\|$ increasing in norm and still converge
- Stable procedure for well conditioned problems

Property inherent of Krylov approximation

$$\Downarrow$$

Many more applications for this general setting

References

1. V. Simoncini and D. B. Szyld. Theory of Inexact Krylov Subspace Methods and Applications to Scientific Computing. *SIAM J. Scientific Comput.* v.25, n.2 (2003), pp. 454-477.

2. F. Perotti and V. Simoncini. Analytical and Numerical Techniques in frequency domain response computation. *Recent Research Developments in Structural Dynamics*, A. Luongo Ed. Research Signpost Pub., 2003, pp.33-54.

3. V. Simoncini and D. B. Szyld. On the occurrence of Superlinear Convergence of Exact and Inexact Krylov Subspace Methods. *SIAM Review 2005*.

4. V. Simoncini. Variable accuracy of matrix-vector products in projection methods for eigencomputation. *SINUM, To appear.*

5. V. Simoncini and D. B. Szyld. The effect of non-optimal bases on the convergence of Krylov Subspace Methods. *Num. Math. To appear.*

6. L. Lopez and V. Simoncini. Analysis of projection-type methods for approximating the matrix exponential operator. March 2005.