Recent advances in approximation using Krylov subspaces

V. Simoncini

Dipartimento di Matematica, Università di Bologna
and CIRSA, Ravenna, Italy
valeria@dm.unibo.it

The framework

It is given an operator $v \to \mathcal{A}_\epsilon(v)$.

Efficiently solve the given problem in the approximation space

$$\mathcal{K}_m = \mathrm{span}\{v, \mathcal{A}_{\epsilon_1}(v), \mathcal{A}_{\epsilon_2}(\mathcal{A}_{\epsilon_1}(v)), \ldots\}, \quad v \in \mathbb{C}^n$$

with $\dim(\mathcal{K}_m) = m$, where $\mathcal{A}_\epsilon \to \mathcal{A}$ for $\epsilon \to 0$ (ϵ may be tuned)

$\star$ for $\mathcal{A} = A$, $\epsilon = 0 \Rightarrow \quad \mathcal{K}_m = \mathrm{span}\{v, Av, A^2v, \ldots, A^{m-1}v\}$

Examples of $\mathcal{A}$:

- Solution of (preconditioned) large linear systems,

$$Ax = b \qquad n \times n \qquad \mathcal{A} = A$$

- Shift-and-Invert eigensolvers

$$Ax = \lambda Mx, \qquad \|x\| = 1, \qquad \mathcal{A} = (\sigma M - A)^{-1}$$

- Spectral Transformation for the exponential

$$x = \exp(A)v, \qquad \mathcal{A} = (\gamma I - A)^{-1}$$

- ...

Goal: Achieve approximation x_m to x within a fixed tolerance, by using $\mathcal{A}_\epsilon$ (and not $\mathcal{A}$), with variable ϵ

Many applications in Scientific Computing

$\mathcal{A}(v)$ 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.

$$\mathcal{K}_m = \mathrm{span}\{v, \mathcal{A}(v), \mathcal{A}(\mathcal{A}(v)), \ldots\}, \quad v \in \mathbb{C}^n$$

The exact approach

To focus our attention: $\mathcal{A} = A$.

Krylov subspace:

$$\mathcal{K}_m = \mathrm{span}\{v, Av, A^2v, \ldots, A^{m-1}v\}$$

$V_m = [v_1, \ldots, v_m]$ orthogonal basis

---

1. $\widehat{v}_{m+1} = Av_m$
2. $v_{m+1} \leftarrow$orthogonalize $\widehat{v}_{m+1}$ w.r.to $\{v_i\}$

Key relation in Krylov subspace methods:

$$AV_m = V_m H_m + v_{m+1} h_{m+1,m} e_m^T \qquad v = V_m e_1 \|v\|$$

The exact approach. cont'd

$\mathcal{K}_m$ Krylov subspace $\quad V_m = [v_1, \ldots, v_m]$ orthogonal basis

$$AV_m = V_m H_m + v_{m+1} h_{m+1,m} e_m^T = V_m \underline{H}_m$$

with $v = V_m e_1 \|v\|$

System: $\quad x_m \in \mathcal{K}_m \quad \Rightarrow \quad x_m = V_m y_m \qquad (x_0 = 0)$

Residual $r_m = v - Ax_m = V_{m+1}(e_1\|v\| - \underline{H}_m y_m)$

Eigenpb: (θ, y) eigenpair of $H_m \quad \Rightarrow \quad (\theta, V_m y)$ Ritz pair for (λ, x)

Residual: $r_m = \theta V_m y - AV_m y = v_{m+1} h_{m+1,m} e_m^T y$

The inexact key relation

$$\mathcal{A} = A \quad \longrightarrow \quad \mathcal{A}_\epsilon(v) = Av + f$$

$$AV_m = V_{m+1}\underline{H}_m + \underbrace{F_m}_{[f_1, f_2, \ldots, f_m]} \qquad F_m \text{ error matrix, } \|f_j\| = O(\epsilon_j)$$

How large is F_m allowed to be?

system:

$$\begin{aligned} r_m &= b - AV_m y_m = b - V_{m+1}\underline{H}_m y_m - F_m y_m \\ &= \underbrace{V_{m+1}(e_1\beta - \underline{H}_m y_m)}_{\text{computed residual} =: \tilde{r}_m} - F_m y_m \end{aligned}$$

eigenproblem: $\qquad (\theta, V_m y)$

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

A dynamic setting

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

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

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

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

Linear systems: The solution pattern

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

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

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

$\tilde{r}_{i-1}$: 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:

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

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

$$|\eta_i| \le \frac{2}{\delta_{m,i}} \|\tilde{r}_{i-1}\|$$

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

$\tilde{r}_{i-1}$: Computed eigenresidual at iteration $i-1$

Analogous results for Harmonic Ritz values and Lanczos approx.

Simoncini, SINUM 2005

Relaxing the inexactness in A

$A \cdot v_i$ not performed exactly $\quad \Rightarrow \quad (A + E_i) \cdot v_i$

True (unobservable) vs. computed residuals:

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

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

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

At each Krylov subspace iteration:

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

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

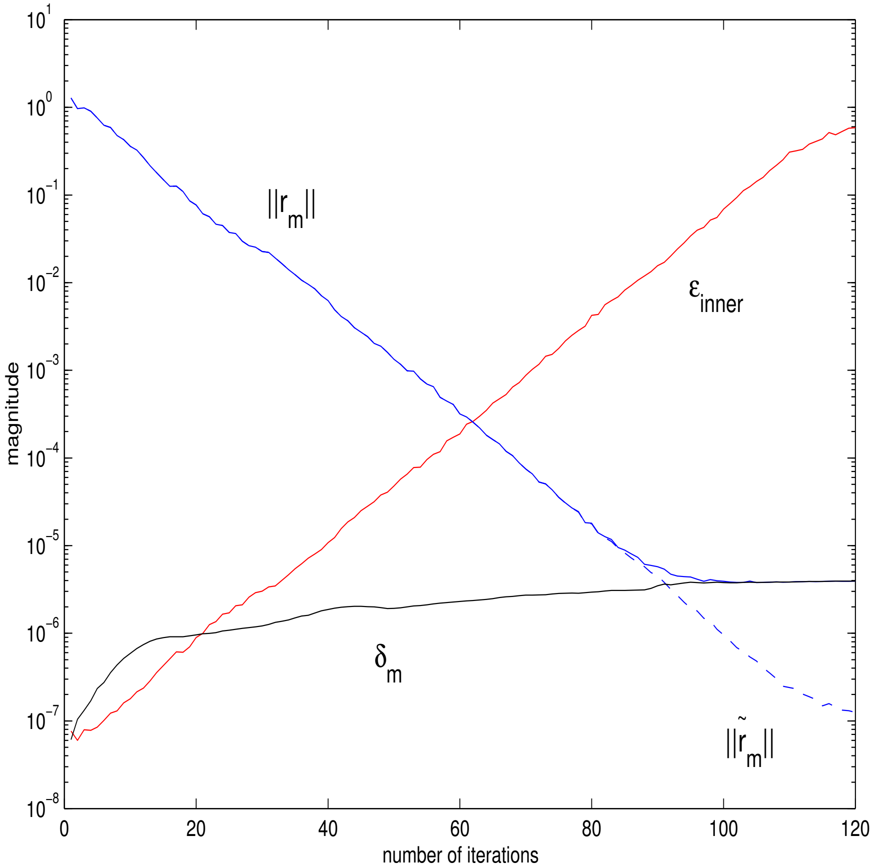
$$A v_i \quad \to \quad B^T \widehat{w}_i = \underbrace{B^T w_i}_{A v_i} - \underbrace{B^T \boldsymbol{\epsilon}_i}_{-E_i v_i} = (A + E_i) v_i$$

Numerical experiment

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

Inexact FOM

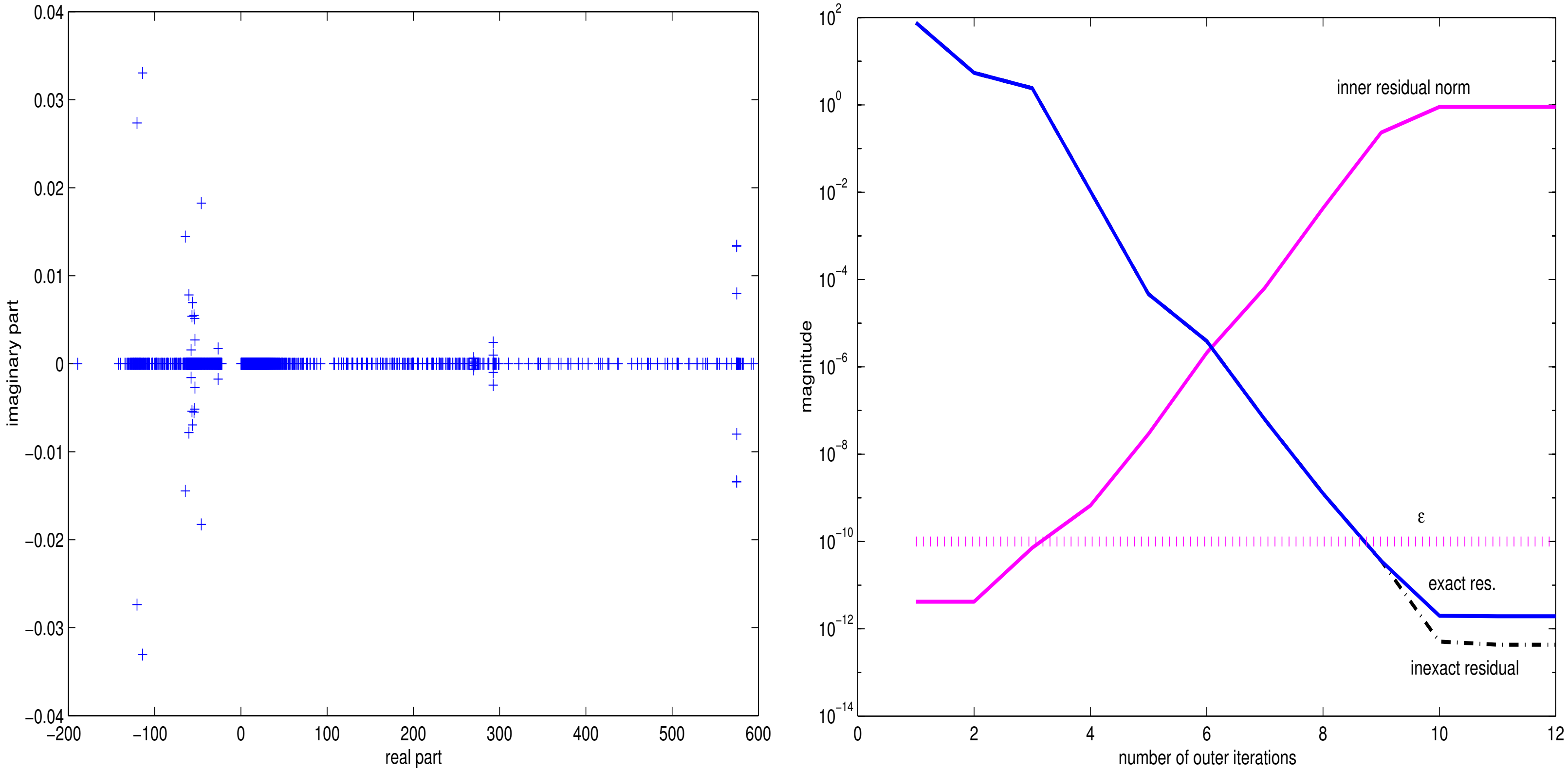
$$\delta_m = \|r_m - (b - V_{m+1}\underline{H}_m y_m)\|$$



Eigenproblem

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

Matrix SHERMAN5



Problems to be faced

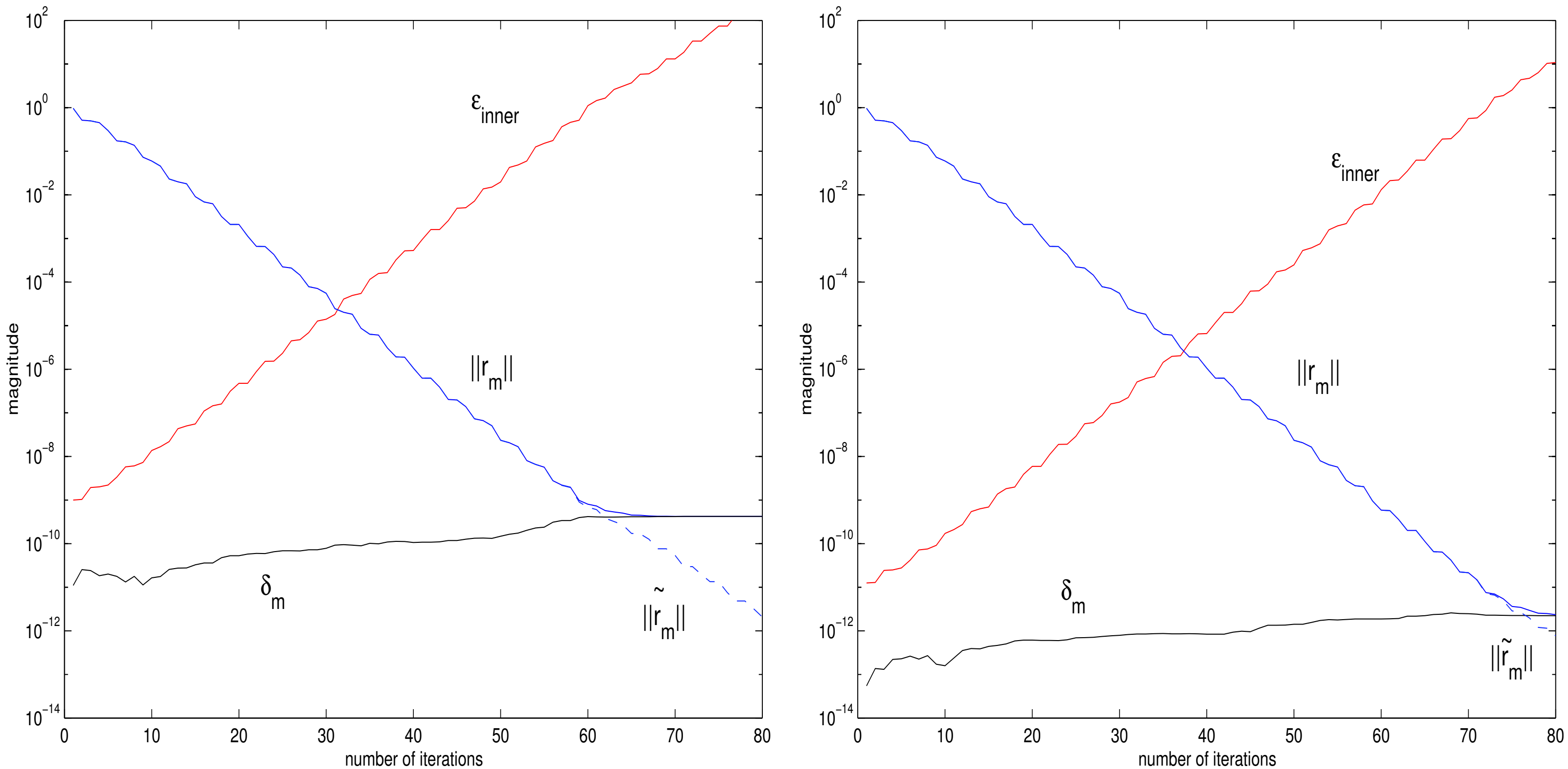
- Make the inexactness criterion practical

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

(CERFACS tr's of Bouras, Frayssè, Giraud, 2000, Bouras, Frayssè SIMAX05)

- What is the convergence behavior?
- What if original $\mathcal{A}$ was symmetric?

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



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

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

Final Requested Tolerance: 10^{-10}

Convergence behavior

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

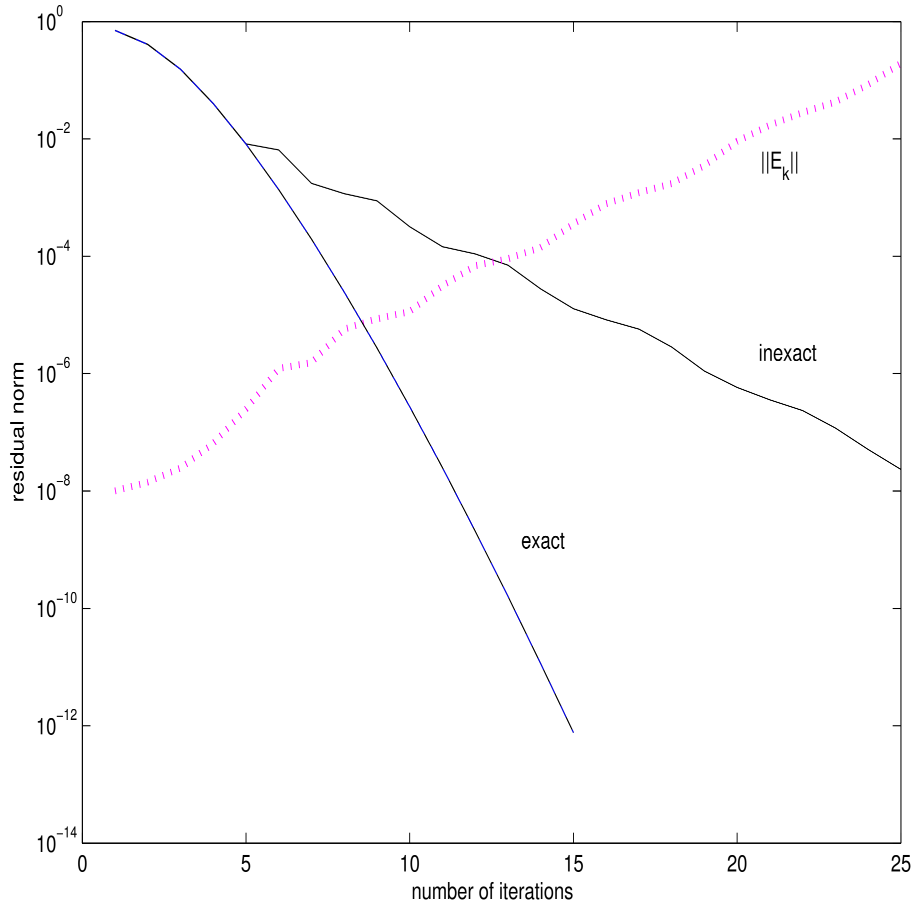
The Sleijpen & van den Eshof's example:

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 (linear systems)

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

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

Problems for:

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

$\star$ Superlinear convergence as for A (Simoncini & Szyld, SIREV 2005)

Flexible preconditioning

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

Flexible:

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

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 $\Rightarrow$ $(\mathcal{A}\mathcal{B}^{-1} + \sigma I)\widehat{x} = b$

(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

(see also van den Eshof, Sleijpen and van Gijzen, JCAM 2005)

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 '88, Golub-Ye '99, Notay '00, Sleijpen-van den Eshof '04, ...)

  - Inexact symmetric eigensolvers

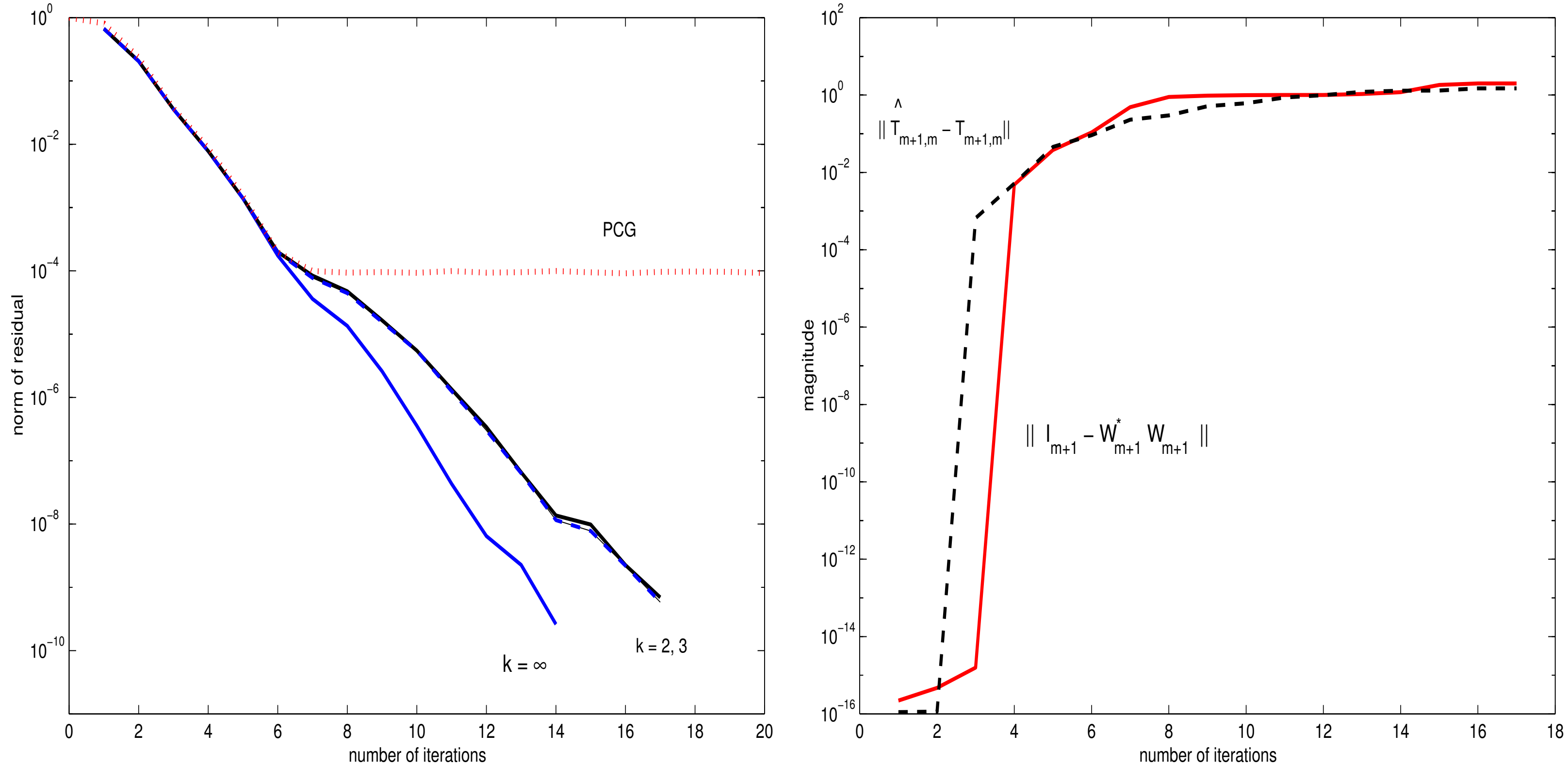
  (Lai-Lin-Lin 1997, Golub-Ye 2000, Golub-Zhang-Zha 2000, Notay 2002, ...)

  - Truncated methods (Simoncini - Szyld, Numer.Math. 2005)

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

Preconditioner:

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



One more application: Approximation of the exponential

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

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

Problem: Find acceleration process for A to speed up convergence

Hochbruck & van den Eshof (SISC 2006):

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

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

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

(Lopez & Simoncini, SINUM 2006)

Conclusions

- $\mathcal{A}$ may be replaced by $\mathcal{A}_{\epsilon_i}$ with increasing ϵ_i and still converge
- Stable procedure for not too sensitive (e.g. non-normal) problems

Property inherent of Krylov approximation

$$\Downarrow$$

Many more applications for this general setting

References at: `www.dm.unibo.it/~simoncin`