Krylov subspace methods:
a versatile tool in Scientific Computing

V. Simoncini

Dipartimento di Matematica

Università di Bologna

valeria@dm.unibo.it

http://www.dm.unibo.it/~simoncin

The framework

Given the large $n \times n$ linear system

$$Ax = b$$

Find x_m such that $x_m \approx x$

The framework

Given the large $n \times n$ linear system

$$Ax = b$$

Find x_m such that $x_m \approx x$

x_0 initial guess $\quad r_0 = b - Ax_0 \quad$ (if no info, take $x_0 = 0$)

Krylov subspace approximation: $\quad x_m = x_0 + z_m$

$$z_m \in \mathcal{K}_m(A, r_0) := \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{m-1} r_0\}$$

$\star$ Projection onto a much smaller space $\quad m \ll n$

Basic Idea of Projection

Assume $x_0 = 0$.

Let $\{v_1, \ldots, v_m\}$ be a basis of

$$\mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{m-1} r_0\}$$

and $V_m := [v_1, \ldots, v_m]$

Then

$$x_m = V_m y_m \qquad y_m \in \mathbb{R}^m$$

y_m coefficients of linear combination

Some popular Krylov subspace methods

$$r_m = b - Ax_m = b - AV_m y_m$$

Mostly theoretical (for nonsymmetric A):

- GMRES (Generalized Minimum RESidual)

$$y_m : \qquad \min_{y \in \mathbb{R}^m} \|r_m\|_2$$

- FOM (Full Orthogonalization Method)

$$y_m : \qquad r_m \perp \mathcal{K}_m$$

Note:
for A symmetric pos. def., FOM becomes CG (Conjugate Gradients)

Some popular Krylov subspace methods

More Practical (for nonsymmetric A):

- GMRES(m): Restarted GMRES
- FOM(m): Restarted FOM (far less popular)
- BiCGStab(ℓ): short-term recurrence

Some popular Krylov subspace methods

More Practical (for nonsymmetric A):

- GMRES(m): Restarted GMRES
- FOM(m): Restarted FOM (far less popular)
- BiCGStab(ℓ): short-term recurrence

Restarted Procedure: Given x_0, r_0

do until convergence

* Run m steps of "Method" to get x_m
* Compute $r_m = b - Ax_m$
* Set $x_0 \leftarrow x_m$, $r_0 \leftarrow r_m$

Some popular Krylov subspace methods

More Practical (for nonsymmetric A):

- GMRES(m): Restarted GMRES
- FOM(m): Restarted FOM (far less popular)
- BiCGStab(ℓ): short-term recurrence

Restarted Procedure: Given x_0, r_0

do until convergence

* Run m steps of "Method" to get x_m
* Compute $r_m = b - Ax_m$
* Set $x_0 \leftarrow x_m$, $r_0 \leftarrow r_m$

Characteristics:

⋆ Economy-versions

⋆ "Good" properties are lost or preserved only locally

Outline

Application-driven practical issues:

Outline

Application-driven practical issues:

- Basic considerations on restarted methods

Outline

Application-driven practical issues:

- Basic considerations on restarted methods
- “Quasi-optimal” methods

Outline

Application-driven practical issues:

- Basic considerations on restarted methods
- “Quasi-optimal” methods
- Indefinite inner products

Outline

Application-driven practical issues:

- Basic considerations on restarted methods
- “Quasi-optimal” methods
- Indefinite inner products
- Inexact methods

Outline

Application-driven practical issues:

- Basic considerations on restarted methods
- “Quasi-optimal” methods
- Indefinite inner products
- Inexact methods

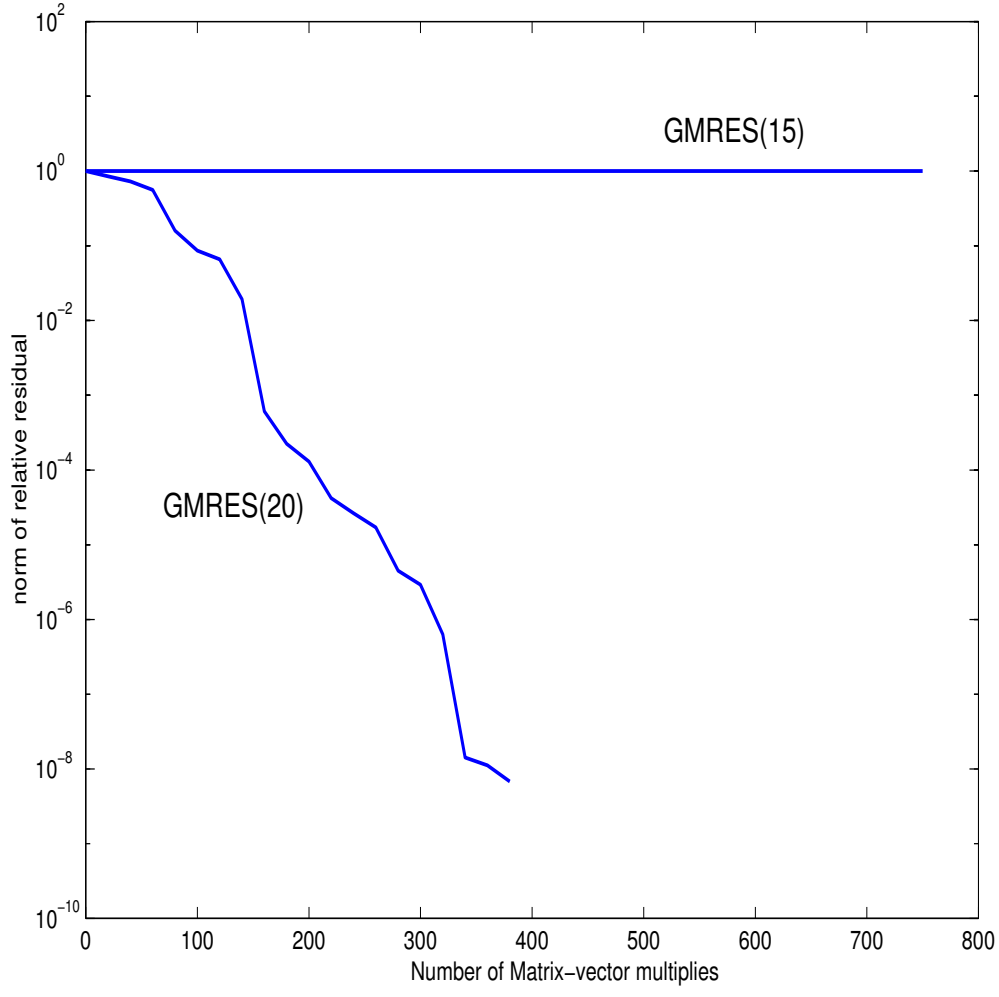
Krylov subspace methods. a versatile Tool for complex problems:

many requirements may be relaxed

Restarted Methods

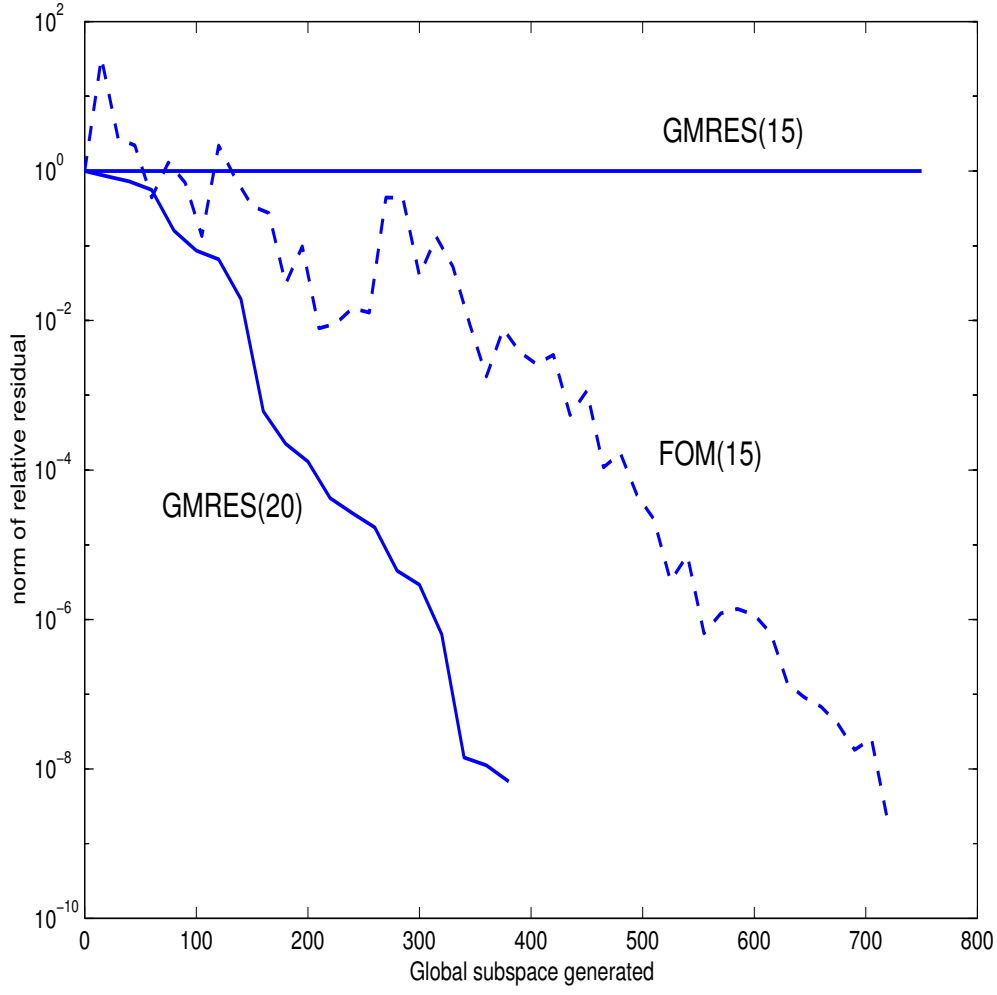
Restarted Methods

Convergence strongly depends on choice of m ...



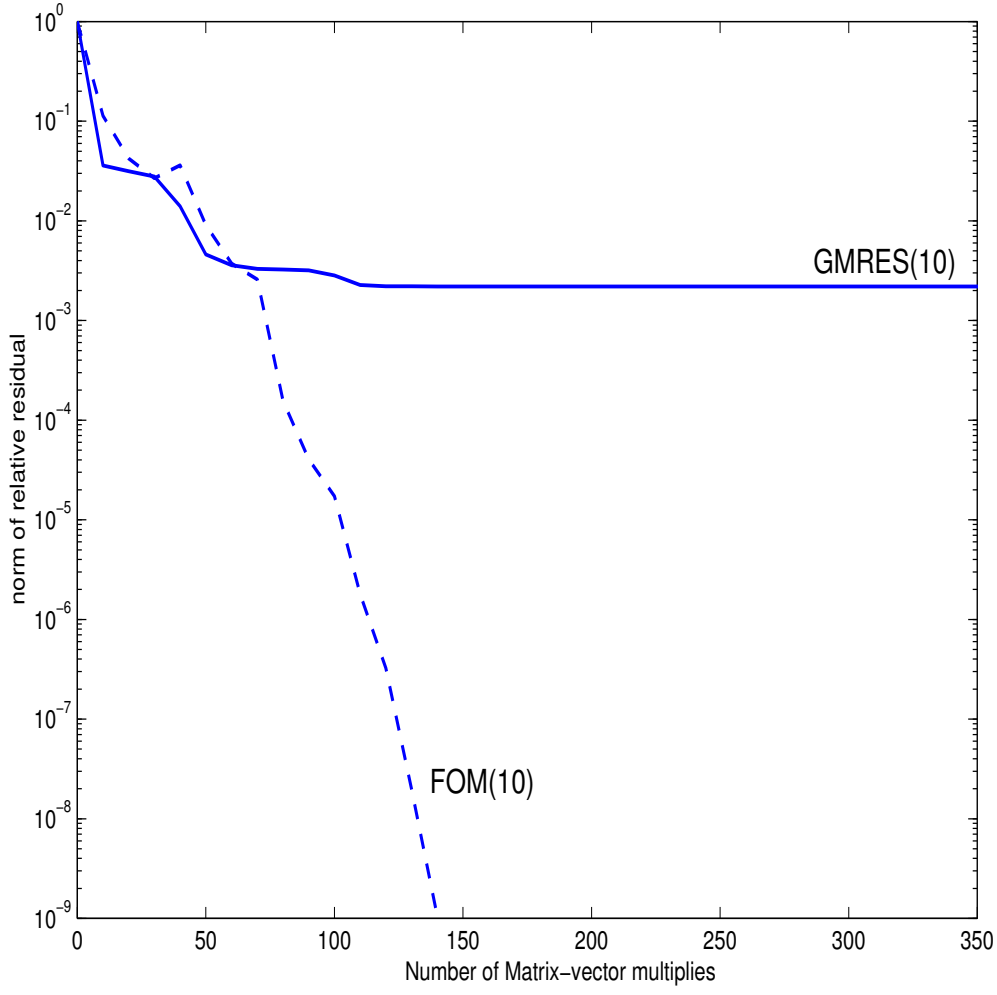
Restarted Methods

Convergence strongly depends on choice of m ... true?



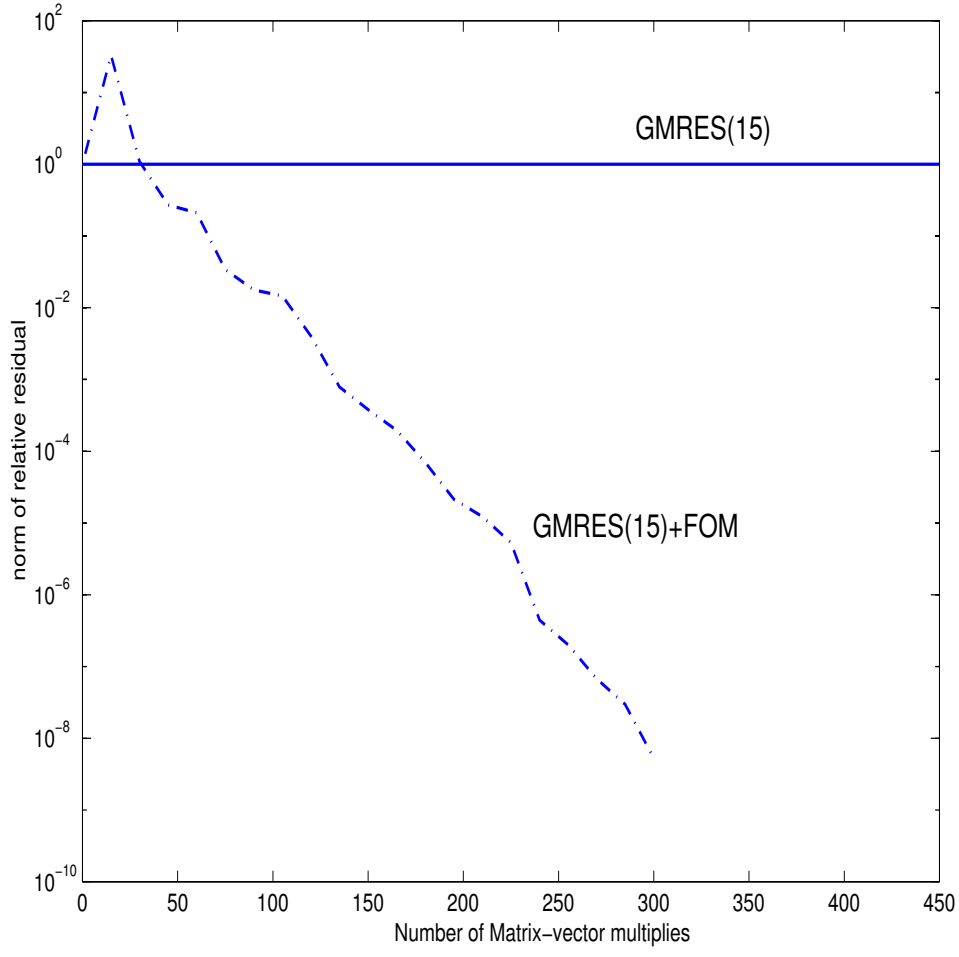
Restarted Methods

Convergence strongly depends on choice of m



Restarted Methods

Switch to FOM residual vector at the very first restart



Pictures from Simoncini, SIMAX 2000.

Outline

Application-driven practical issues:

- Basic considerations on restarted methods
- **“Quasi-optimal” methods**
- Indefinite inner products
- Inexact methods

Enhanced Restarted Methods

Warning: Large m not always means faster convergence

Enhanced Restarted Methods

Warning: Large m not always means faster convergence

Current research:

- Deflated Methods (originally used for A s.p.d.)
  Mansfield, Nicolaides, Erhel etal., Saad etal., Nabben, Vuik, ...

Enhanced Restarted Methods

Warning: Large m not always means faster convergence

Current research:

- Deflated Methods (originally used for A s.p.d.)
  Mansfield, Nicolaides, Erhel etal., Saad etal., Nabben, Vuik, ...

- Augmented Methods (information saved from previous restarts)
  De Sturler, Morgan, Sorensen, Baglama etal, Baker etal, ...

  See also Eiermann, Ernst, Schneider (JCAM 2000)

Enhanced Restarted Methods

Warning: Large m not always means faster convergence

Current research:

- Deflated Methods (originally used for A s.p.d.)
  Mansfield, Nicolaides, Erhel etal., Saad etal., Nabben, Vuik, ...

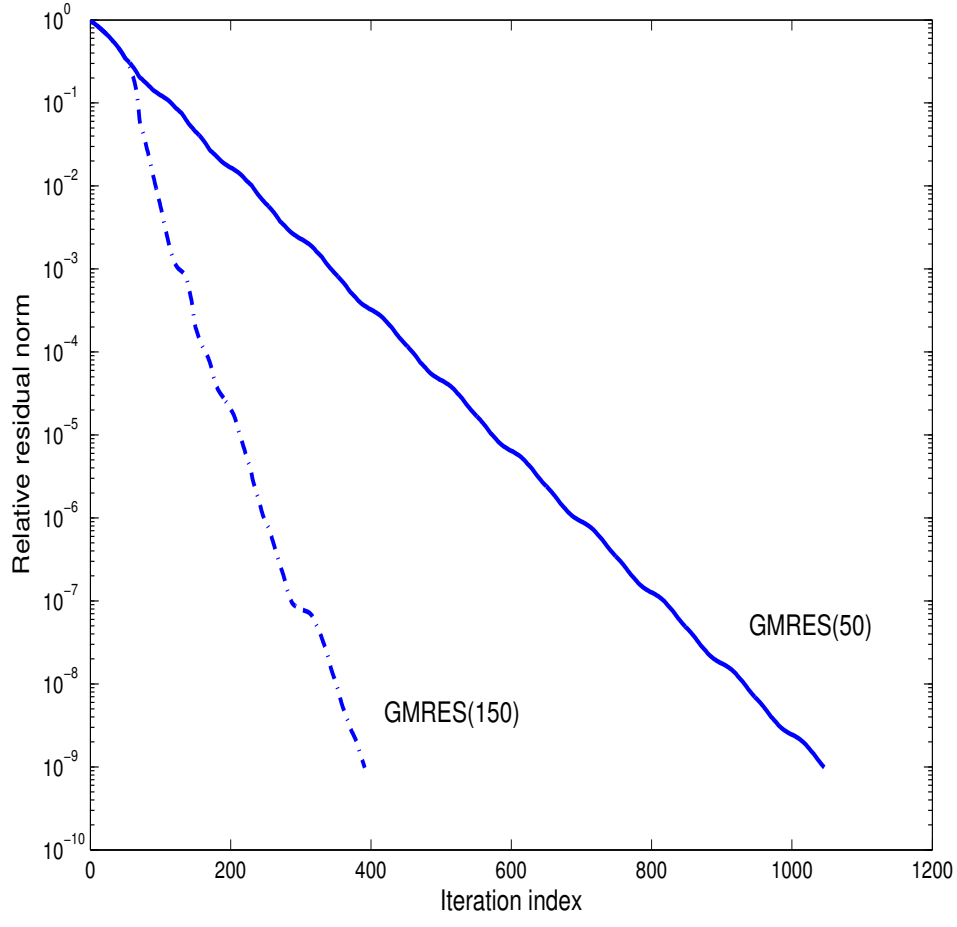
- Augmented Methods (information saved from previous restarts)
  De Sturler, Morgan, Sorensen, Baglama etal, Baker etal, ...

  See also Eiermann, Ernst, Schneider (JCAM 2000)

- Truncated Methods (only local information maintained)
  Golub, Ye, Notay, Szyld, ...

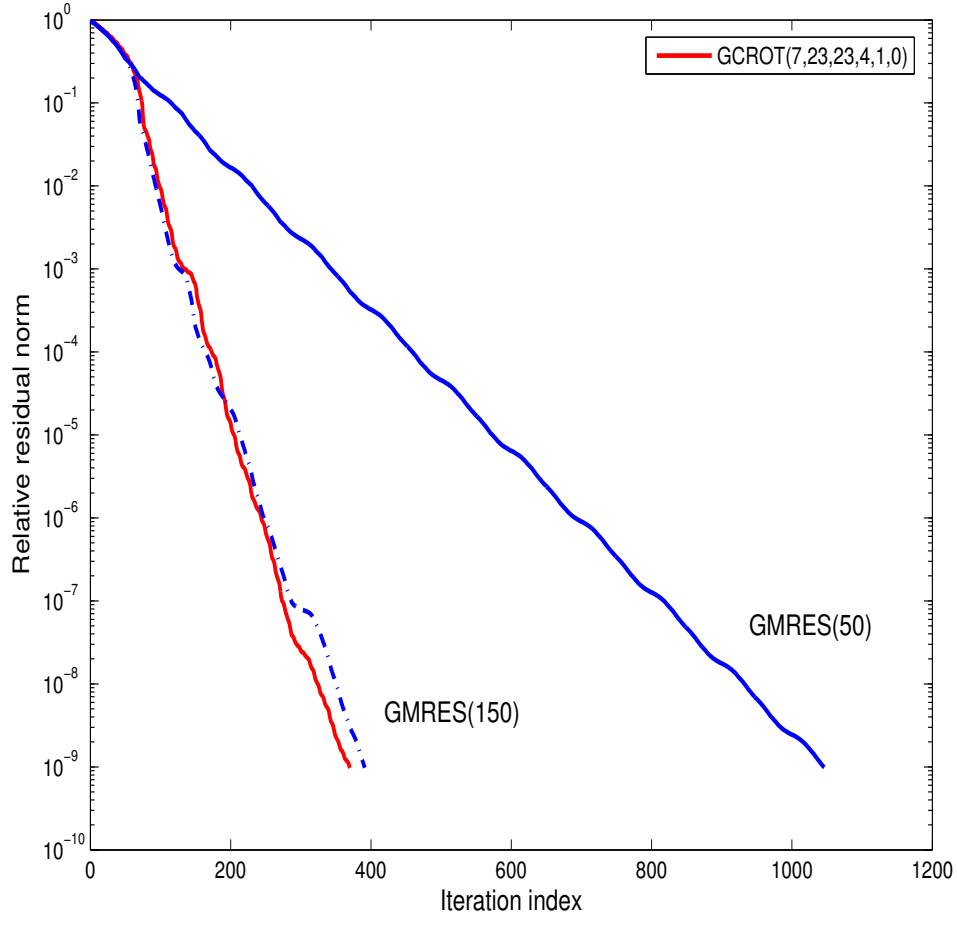
De Sturler's method

Tricky way to enhance approximation space



De Sturler's method

Trickly way to enhance approximation space



Code: courtesy of Oliver Ernst.

Truncated methods and “Quasi-optimality”. I

⋆ A truncated method discards “older” vectors

$$\{v_1, v_2, \ldots, v_{m-k}, \underbrace{v_{m-k+1}, \ldots, v_m}_{\text{orthogonal}}, v_{m+1}, \ldots, \}$$

(local optimality properties)

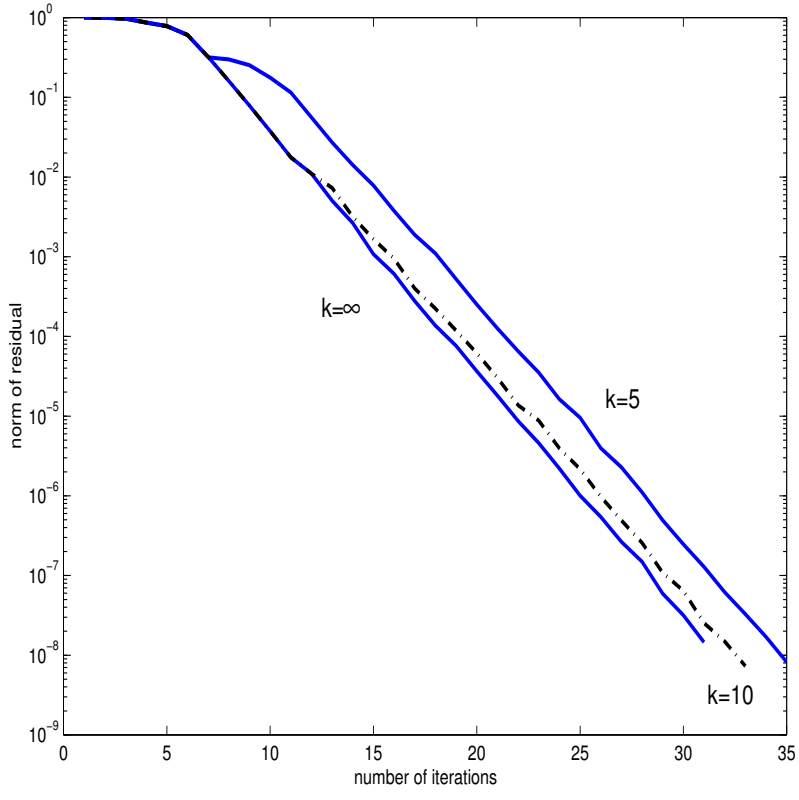
Limited memory requirements

Optimality is lost

How “old” is old?

Truncated methods and "Quasi-optimality'. II

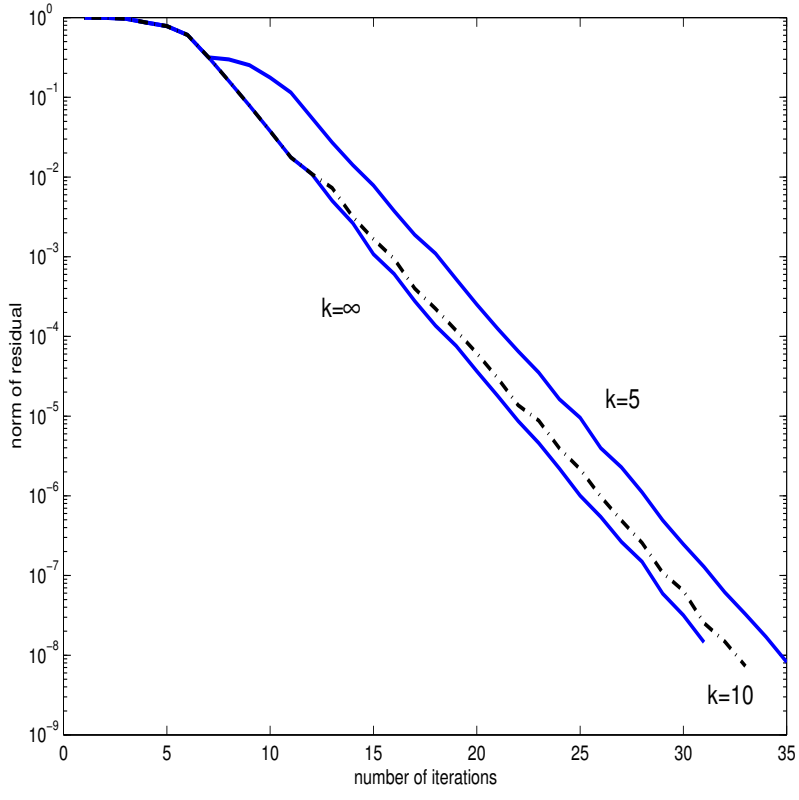
Example: A is non-normal, spectrum on circle $|1 - z| = 0.5$

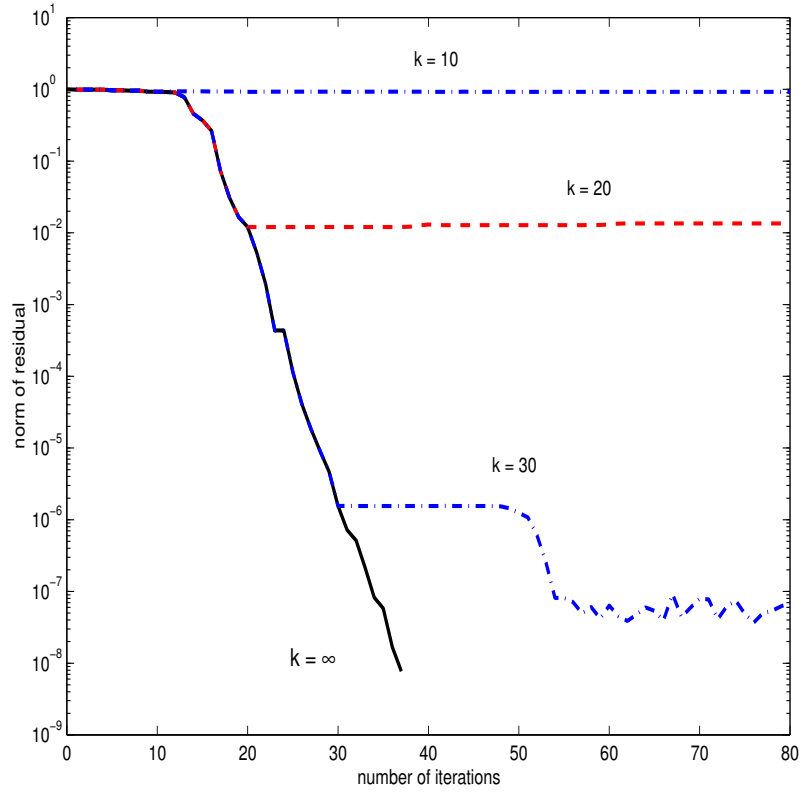


Low eigenvector cond. number

Truncated methods and "Quasi-optimality'. II

Example: A is non-normal, spectrum on circle $|1 - z| = 0.5$





Low eigenvector cond. number

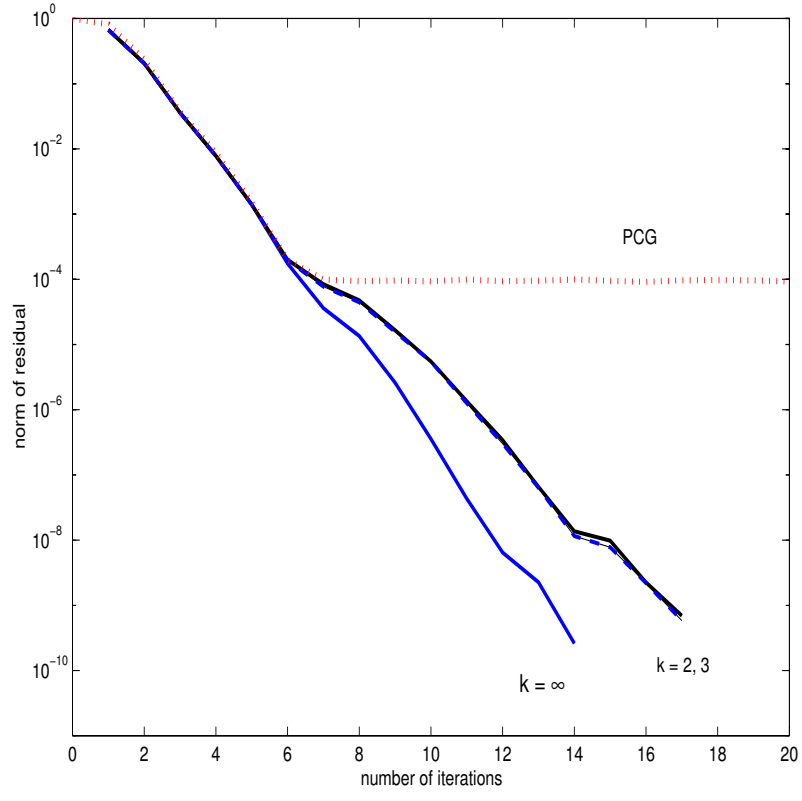
high eigenvector cond. number

Truncated methods and "Quasi-optimality'. III

⋆ If A is nonsymmetric, but harmless modification of a symmetric matrix then short truncation suffices

$$Ax = b \qquad A \text{ symmetric} \quad \Rightarrow \quad \mathcal{P}^{-1}Ax = \mathcal{P}^{-1}b$$

$$\mathcal{P}^{-1}v = L^{-T}L^{-1}v + \epsilon\mathbf{1}, \quad \epsilon = 10^{-5}, \quad L \text{ Incomplete Cholesky of } A$$



Outline

Application-driven practical issues:

- Basic considerations on restarted methods
- “Quasi-optimal” methods
- **Indefinite inner products**
- Inexact methods

Need for a “different” inner product ?

Typical orthogonality: $r_m \perp \mathcal{K}_m$

Common alternative:

$\star$ Given M Hermitian and positive definite,

$$r_m \perp_M \mathcal{K}_m$$

i.e., for Range(V_m)=$\mathcal{K}_m$ it holds $V_m^* M r_m = 0$

may lead to minimization of $\|r_m\|_M$ or $\|e_m\|_M$

$\clubsuit$ In many cases, use of $M^{-\frac{1}{2}} A M^{-\frac{1}{2}}$ hpd

Need for a “different” inner product ?

Typical orthogonality: $r_m \perp \mathcal{K}_m$

Common alternative:

$\star$ Given M Hermitian and positive definite,

$$r_m \perp_M \mathcal{K}_m$$

i.e., for Range(V_m)=$\mathcal{K}_m$ it holds $V_m^* M r_m = 0$

may lead to minimization of $\|r_m\|_M$ or $\|e_m\|_M$

$\clubsuit$ In many cases, use of $M^{-\frac{1}{2}} A M^{-\frac{1}{2}}$ hpd

What about different alternatives?

Motivations for an indefinite inner product

- Exploit inherent properties of the problem. For instance,

  A complex symmetric

Motivations for an indefinite inner product

- Exploit inherent properties of the problem. For instance,

  A complex symmetric

- Exploit matrix structure

$$A = \begin{pmatrix} H & B \\ B^T & 0 \end{pmatrix}$$

(or, say, A Hamiltonian)

Motivations for an indefinite inner product

- Exploit inherent properties of the problem. For instance,

  A complex symmetric

- Exploit matrix structure

$$A = \begin{pmatrix} H & B \\ B^T & 0 \end{pmatrix}$$

  (or, say, A Hamiltonian)

... to gain in efficiency with (hopefully) no loss in reliability

An example. Indefinite (Constraint) Preconditioner

$$Ax = b \qquad A = \begin{pmatrix} H & B \\ B^T & 0 \end{pmatrix}, \qquad H = H^T, H \geq 0$$

Preconditioning: $AP^{-1}\widehat{x} = b$

An example. Indefinite (Constraint) Preconditioner

$$Ax = b \qquad A = \begin{pmatrix} H & B \\ B^T & 0 \end{pmatrix}, \qquad H = H^T, H \geq 0$$

Preconditioning: $AP^{-1}\widehat{x} = b$

Block indefinite Preconditioner:

$$P = \begin{pmatrix} \widetilde{H} & B \\ B^T & 0 \end{pmatrix} \qquad \widetilde{H} \approx H$$

An example. Indefinite (Constraint) Preconditioner

$$Ax = b \qquad A = \begin{pmatrix} H & B \\ B^T & 0 \end{pmatrix}, \qquad H = H^T, H \geq 0$$

Preconditioning: $AP^{-1}\widehat{x} = b$

Block indefinite Preconditioner:

$$P = \begin{pmatrix} \widetilde{H} & B \\ B^T & 0 \end{pmatrix} \qquad \widetilde{H} \approx H$$

* AP^{-1} not symmetrizable!
* However: AP^{-1} is P^{-1}-Hermitian (Hermitian wrto P^{-1})

$\Rightarrow$ Cheap short-term recurrence

(Simplified Lanczos - Freund & Nachtigal '95)

Preconditioner Performance

$$P^{-1} = \begin{pmatrix} \widetilde{H} & B \\ B^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} I & -B \\ O & I \end{pmatrix} \begin{pmatrix} I & O \\ O & -(\mathbf{B^T B})^{-1} \end{pmatrix} \begin{pmatrix} I & O \\ -B^T & I \end{pmatrix}$$

($\widetilde{H} = I$ if prescaling used)

Preconditioner Performance

$$P^{-1} = \begin{pmatrix} \widetilde{H} & B \\ B^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} I & -B \\ O & I \end{pmatrix} \begin{pmatrix} I & O \\ O & -(\mathbf{B^T B})^{-1} \end{pmatrix} \begin{pmatrix} I & O \\ -B^T & I \end{pmatrix}$$

($\widetilde{H} = I$ if prescaling used)

3D Magnetostatic problem. Number of iterations

size	QMR	QMR($P_{\rm def}$)	QMR(P)
1119	2368	40	15
2208	2825	36	13
4371	5191	43	17
8622	>10000	49	16
22675	>10000	81	25

$P_{\rm def}$ = diag $(I, B^T B)$ hpd

Preconditioner Performance

$$P^{-1} = \begin{pmatrix} \widetilde{H} & B \\ B^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} I & -B \\ O & I \end{pmatrix} \begin{pmatrix} I & O \\ O & -(\mathbf{B^T B})^{-1} \end{pmatrix} \begin{pmatrix} I & O \\ -B^T & I \end{pmatrix}$$

($\widetilde{H} = I$ if prescaling used)

3D Magnetostatic problem. Number of iterations

size	QMR	QMR($P_{\rm def}$)	QMR(P)
1119	2368	40	15
2208	2825	36	13
4371	5191	43	17
8622	>10000	49	16
22675	>10000	81	25

$P_{\rm def}$ = diag $(I, B^T B)$ hpd

In practice: $B^T B \approx S$ Incomplete Cholesky fact. $\Rightarrow$ $\widehat{P}$

Outline

Application-driven practical issues:

- Basic considerations on restarted methods
- “Quasi-optimal” methods
- Indefinite inner products
- **Inexact methods**

Inexact methods

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)

Inexact methods

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)

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

⋆ Analysis also possible for eigenproblem

Some typical situations

$\mathcal{A}(v)$ function (linear in v):

- $\mathcal{A}$ result of a complex functional application
- Schur complement: $A = B^T S^{-1} B$ S expensive to invert
- Flexible preconditioned system: $AP^{-1}x = b$, where

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

- etc.

Some typical situations

$\mathcal{A}(v)$ function (linear in v):

- $\mathcal{A}$ result of a complex functional application
- Schur complement: $A = B^T S^{-1} B$ S expensive to invert
- Flexible preconditioned system: $AP^{-1}x = b$, where

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

- etc.

In the eigenvalue context: shift-and-invert strategy

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

Questions

- Do we need to have ϵ small to get good approximation?

  good approximation: $\|r_m\| \leq \varepsilon_0$ (fixed tolerance)

- Do we need to have ϵ fixed throughout?

Questions

- Do we need to have ϵ small to get good approximation?

  good approximation: $\|r_m\| \leq \varepsilon_0$ (fixed tolerance)

- Do we need to have ϵ fixed throughout?

- Do we still converge to a meaningful solution if ϵ varies?

- What happens to convergence rate when ϵ varies?

Assuming A is exact...

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

Arnoldi relation:

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

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

Working with an inaccurate A

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

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

How large is F_m allowed to be?

Working with an inaccurate A

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

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

How large is F_m allowed to be? $\qquad x_m = V_m y_m$

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

where $F_m y_m = \sum_{i=1}^{m} f_i (y_m)_i$

Relaxed methods

$$F_m y_m = \sum_{i=1}^{m} f_i (y_m)_i$$

In fact, for several methods there exists ℓ_m such that

$$| (y_m)_i | \leq \ell_m \| \tilde{r}_{i-1} \|$$

Therefore, $\| f_i \|$ is allowed to be large!

Relaxed methods

$$F_m y_m = \sum_{i=1}^{m} f_i (y_m)_i$$

In fact, for several methods there exists ℓ_m such that

$$| (y_m)_i | \leq \ell_m \|\tilde{r}_{i-1}\|$$

Therefore, $\|f_i\|$ is allowed to be large!

More precisely,

$$\text{If} \quad \|f_i\| \leq \frac{\ell_m}{m} \frac{1}{\|\tilde{r}_{i-1}\|} \varepsilon \qquad i = 1, \ldots, m$$

$$\text{then} \quad \|F_m y_m\| \leq \varepsilon \quad \Rightarrow \qquad \|r_m - \tilde{r}_m\| \leq \varepsilon$$

Bouras, Frayssè, Giraud, Simoncini, Szyld, Sleijpen, Van den Eshof, Gratton ...

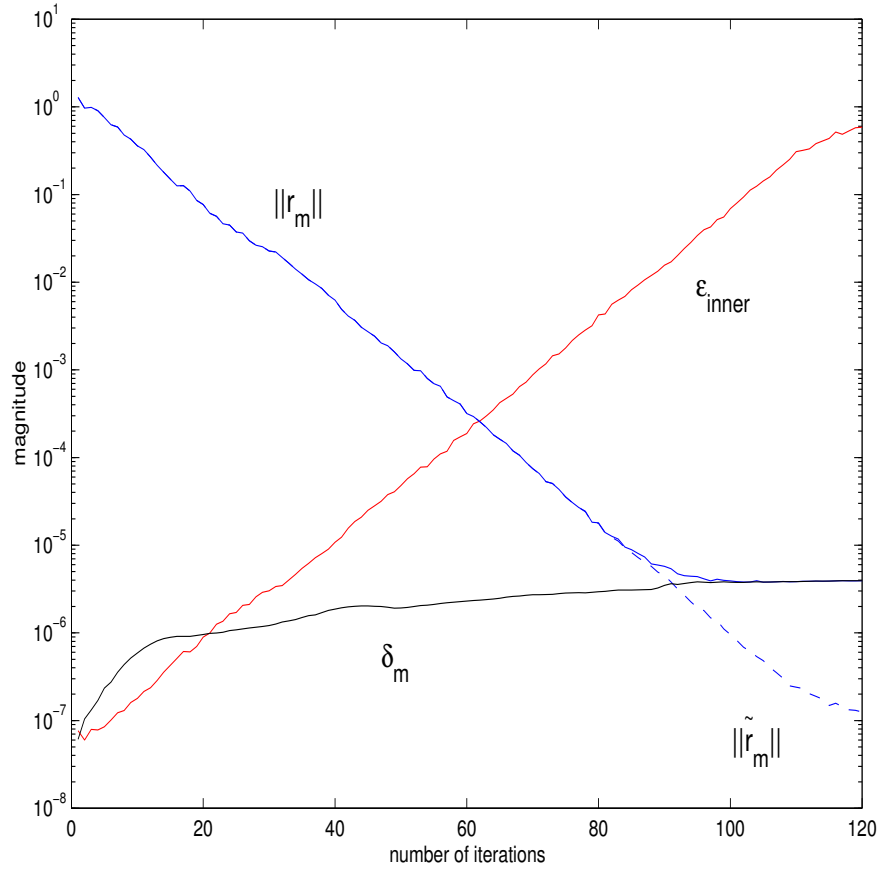
Numerical experiment: Schur complement

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

at each it. i solve $Sw_i = Bv_i$

Inexact FOM

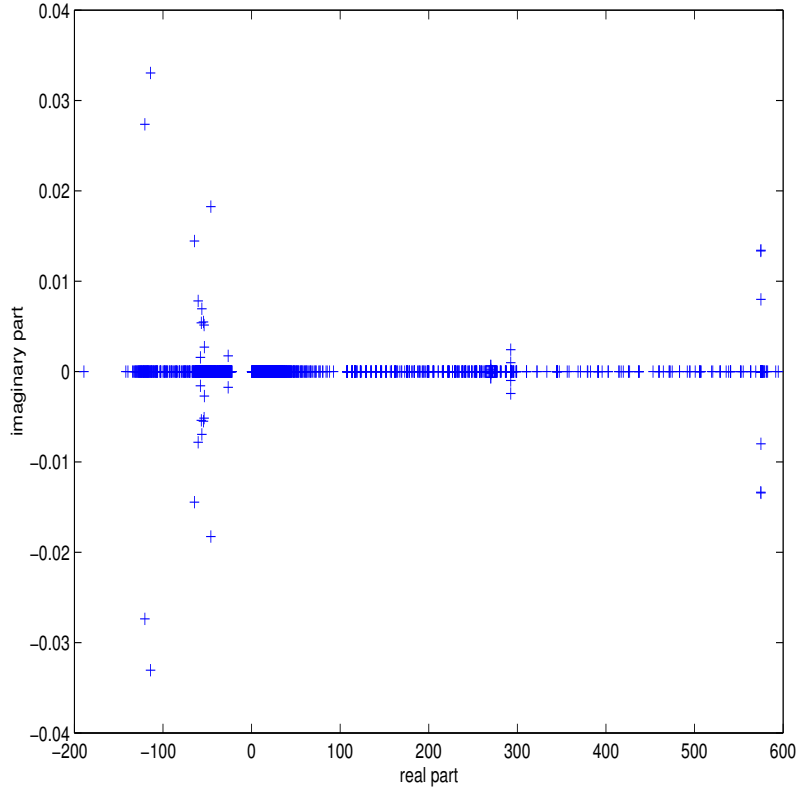
$\delta_m = \|r_m - (b - V_{m+1}\underline{H}_m y_m)\|$

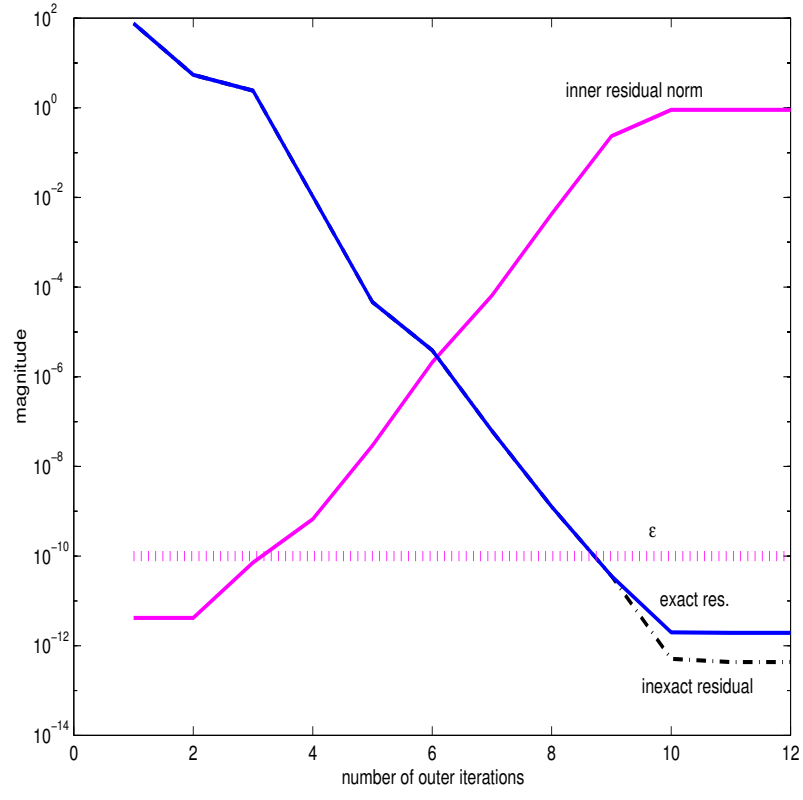


Eigenproblem

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

Matrix SHERMAN5





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

(Preconditioned CG-type iteration for $\mathcal{B}$)

Relaxed procedure

- $\mathcal{A}$ 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

Conclusions

- Often, enough confidence to tailor methods to problems

Conclusions

- Often, enough confidence to tailor methods to problems
- Ability to relax many of the classical requirements

Conclusions

- Often, enough confidence to tailor methods to problems
- Ability to relax many of the classical requirements
- Further enhancements are ahead!

Conclusions

- Often, enough confidence to tailor methods to problems
- Ability to relax many of the classical requirements
- Further enhancements are ahead!

Recent Survey:
Recent computational developments in Krylov Subspace Methods for linear systems
Simoncini & Szyld, 2005
59 pp., 352 references
to appear in Numer. Linear Algebra w/Appl.

http://www.dm.unibo.it/~simoncin