Inexact Krylov Subspace Methods for PDEs and Control Problems

Daniel B. Szyld
Temple University, Philadelphia

Collaborators:
Xiuhong Du, Temple University
Eldad Haber, Emory University
Maria Karampataki, Emory University
Marcus Sarkis, WPI, and IMPA, Rio de Janeiro
Christian Schaerer, IMPA, Rio de Janeiro
Valeria Simoncini, Università di Bologna

DOE ASCR Applied Math, Livermore, 23 May 2007

Problem Statement

Solve a system $Hx = b$, H Hermitian or non-Hermitian using Krylov subspace iterative methods

$$\mathcal{K}_m(H, r_0) = \text{span}\{r_0, Hr_0, H^2 r_0, \ldots, H^{m-1} r_0\}.$$

Given x_0, $r_0 = b - Hx_0$, find approximation

$$x_m \in x_0 + \mathcal{K}_m(H, r_0),$$

satisfying some property:

Petrov-Galerkin, e.g., GMRES, MINRES:

$$x_m = \text{arg min}\{\|b - Hx\|_2\}, \quad x \in x_0 + \mathcal{K}_m(H, r_0)$$

Galerkin, e.g., FOM, CG: $\quad b - Hx_m \perp \mathcal{K}_m(H, r_0)$

Krylov subspace methods (cont.)

- Methods work by suitably choosing a basis of $\mathcal{K}_m(H, r_0)$
- Let $v_1, v_2, \ldots, v_m$ be such a basis, chosen to be orthonormal.
- With $V_m = [v_1, v_2, \ldots, v_m]$, obtain Arnoldi relation:

$$HV_m = V_{m+1} H_{m+1,m} = V_m H_m + h_{m+1,m} v_{m+1} e_m^T$$

  $H_{m+1,m}$ is $(m+1) \times m$ upper Hessenberg
- Each method finds y_m so that $x_m = x_0 + V_m y_m$
- Main costs:
  1. Matrix-vector product: Hv_k
  2. Orthogonalization
  3. Storage (if there is no recursion)

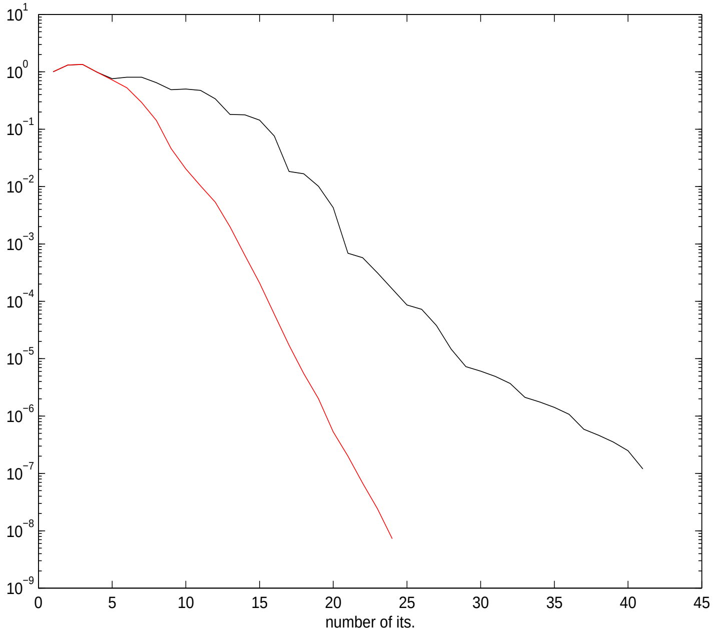
This Talk

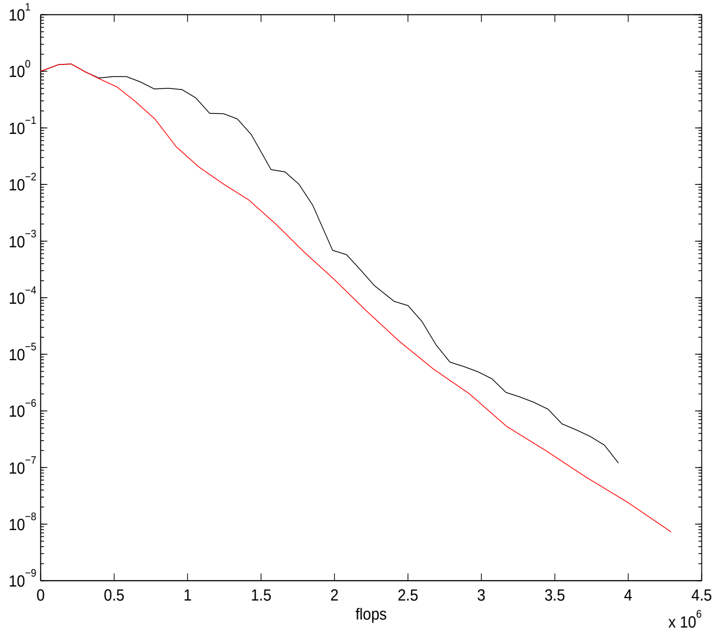
- Consider the case when one does not fully orthogonalize: Truncated methods.
- Reduce the cost of matrix-vector product when H is either
  - Not known exactly
  - Computationally expensive (e.g., Schur complement, reduced Hessian)
  - Preconditioned with variable matrix (i.e., iteration dependent)

Truncated Krylov subspace methods

- Only orthogonalize with respect to some fixed number k of previous vectors [Saad, 1983, 1996].
- $H_{m+1,m}$ banded with upper semiband $k-2$.
  Matrix with basis vectors V_m not orthogonal.
  Can be implemented so that only O(k) vectors are stored.
- Extreme case, $k = 3$, $H_{m+1,m}$ tridiagonal.
  If H is SPD, FOM reduces to CG (and V_m automatically orthogonal).
- Theory for "non-optimal methods" [Simoncini and Szyld, 2005]

Example: $L(u) = -u_{xx} + -u_{yy} + 100(x+y)u_x + 100(x+y)u_y$, on $[0,1]^2$, Dirichlet b.c., centered 5 pts. discretization, $n = 2500$.





GMRES, Truncated $k = 3$.

Inexact Krylov subspace methods

- At the kth iteration of the Krylov space method use

$$(H + D_k)v_{k-1} \text{ instead of } Hv_{k-1},$$

  where $\|D_k\|$ can be monitored

- [Bouras, Frayssé, and Giraud, CERFACS reports 2000, SIMAX 2005] show experimentally that as k progresses $\|D_k\|$ can be allowed to be larger; see also [Sleijpen and van der Eshof, 2004]

Inexact Krylov (cont.)

We repeat: $\|D_k\|$ small at first, $\|D_k\|$ can be big later.
Convergence is maintained!

- Instead of $\quad HV_m = V_{m+1}H_{m+1,m} \quad$ we have now

$$[(H+D_1)v_1, (H+D_2)v_2, \ldots, (H+D_m)v_m] \quad = \quad V_{m+1}H_{m+1,m}$$

- Subspace spanned by $v_1, v_2, \ldots, v_m$ is not a Krylov subspace, but V_m orthogonal (in the full case)

Theorem for Inexact FOM

[Simoninci and Szyld, 2003]

True residual: $r_m = b - Hx_m = r_0 - HV_m y_m$

Computed residual(e.g.): $\tilde{r}_m = r_0 - V_{m+1} H_{m+1,m} y_m = r_0 - W_m y_m$

Let $\varepsilon > 0$. If for every $k \le m$,

$$\|D_k\| \le \frac{\sigma_{min}(H_{m_*})}{m_*} \frac{1}{\|\tilde{r}_{k-1}\|} \varepsilon \equiv \ell_m^F \frac{1}{\|\tilde{r}_{k-1}\|} \varepsilon \ ,$$

then $\|V_m^T r_m\| \le \varepsilon$ and $\|r_m - \tilde{r}_m\| \le \varepsilon$.

m_* being the maximum number of iterations allowed

(Similar results for inexact GMRES)

Theorem for Inexact Truncated FOM

$$\|D_k\| \le \frac{\sigma_{min}(H_{m_*})\sigma_{min}(V_m)}{m_*} \frac{1}{\|\tilde{r}_{k-1}\|}\varepsilon \equiv \ell_m^{TF} \frac{1}{\|\tilde{r}_{k-1}\|}\varepsilon \ ,$$

implies $\|V_m^T r_m\| \le \varepsilon$ and $\delta_m = \|r_m - \tilde{r}_m\| \le \varepsilon$.

Notes:

- This result applies in particular to Inexact CG
  Better criterion than above for ICG [Du, 2007]
- ℓ_m can be estimated from problem, if information is available.

First Experiment

$H = \text{diag}([10^{-4}, 2, 3, \cdots, 100])$ $D_k = \text{symm}\ [\alpha_k \text{randn}(100, 100)]$
$b = \text{randn}(100, 1)$ We chose $\varepsilon = 10^{-8}$

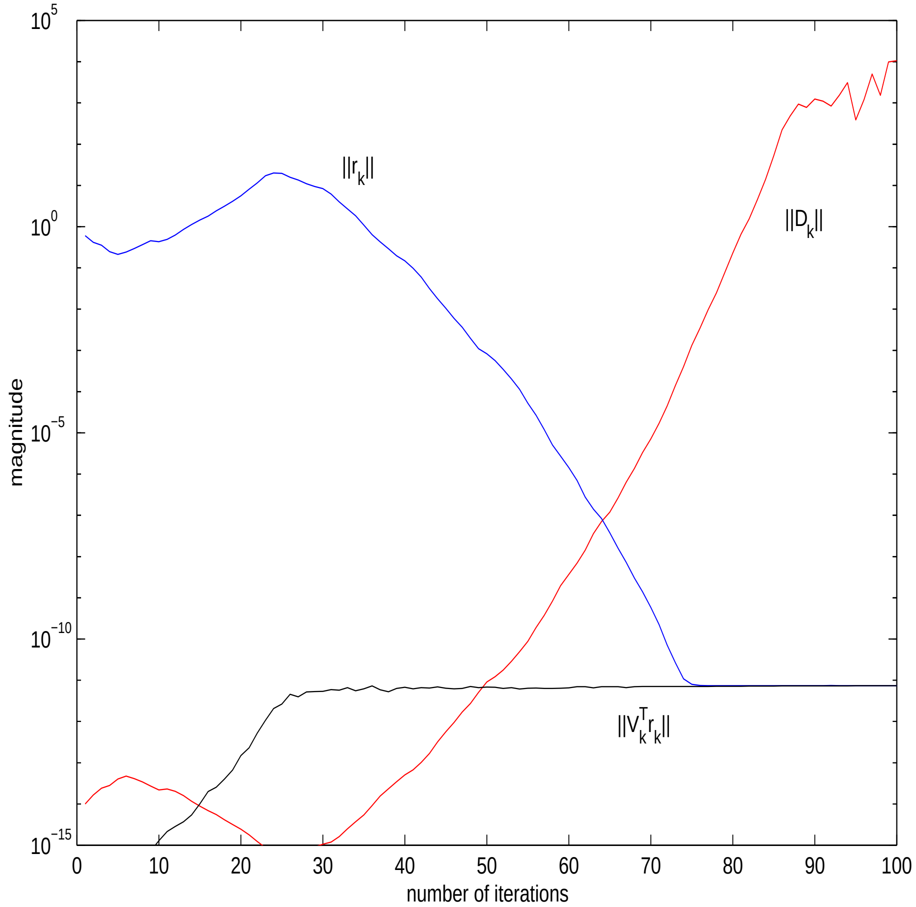
- Our condition (e.g. for FOM)

$$\|D_k\| \leq \frac{\sigma_{min}(H)}{m_*} \frac{1}{\|\tilde{r}_{k-1}\|} \varepsilon$$

is very conservative. In most cases it is too strict.

However, $\sigma_{min}(H)$ does play a role.

CG: condition $\|D_k\| \le \frac{\sigma_{min}(H)}{m_*} \frac{1}{\|\tilde{r}_{k-1}\|} \varepsilon$



$\|V_m^T r_m\| \ll \varepsilon$

Applications:
I. Schur complement systems

$$\begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} w \\ x \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix},$$

$$\begin{aligned} B^T A^{-1} B x &= B^T A^{-1} f; \qquad Aw = f - Bx \\ Hx &= b \end{aligned}$$

A^{-1} not exactly (use Krylov method).

Applications: I. Schur complement systems (cont.)

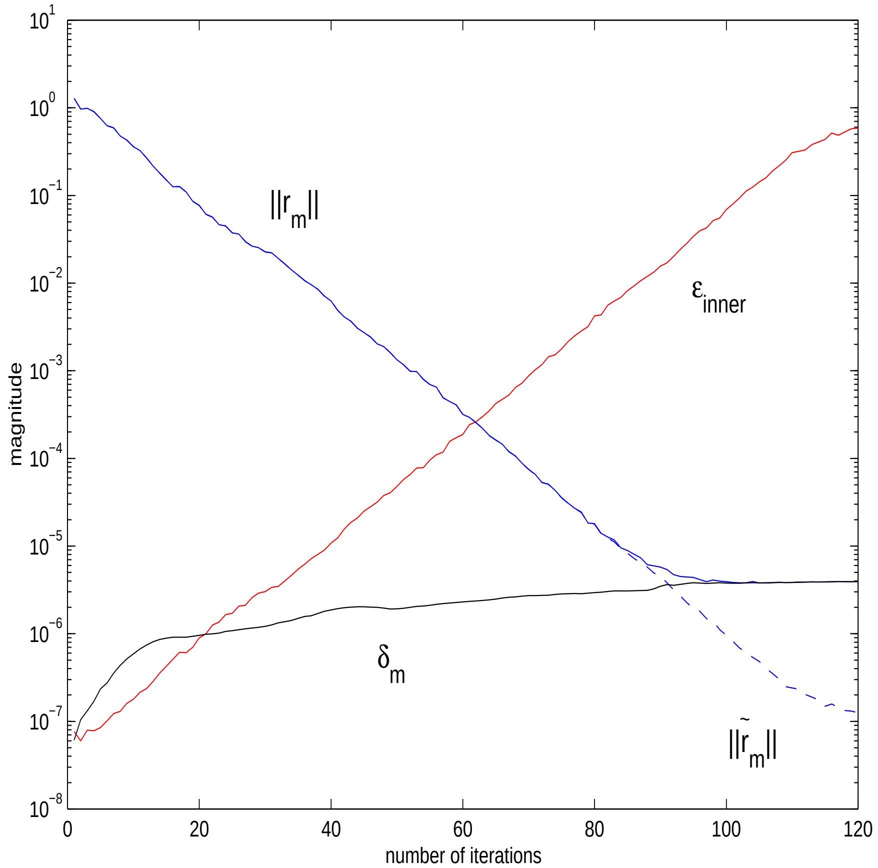
- A^{-1} not exactly (use Krylov method).
- Replace Hv with $\mathcal{H}v = B^T z_j^{(k)}$, where $z_j^{(k)}$ is the approximation obtained at the jth (inner) iteration of the solution to the equation

$$Az = Bv$$

- Question is then: How many inner iterations?
  i.e., at what value of j stop?
  "Translate" conditions on $\|D_k\|$ to conditions on norm of inner residual.

Let $r_k^{inner} = Az_j^{(k)} - Bv$ be the inner residual

Take $$\|r_k^{inner}\| < \frac{\sigma_{m_\star}(H_{m_\star})}{\|B^T A^{-1}\| m_\star} \frac{1}{\|\tilde{r}_{k-1}^{fom}\|} \varepsilon \equiv \varepsilon_{\text{inner}}$$



- Two-dim. saddle point magnetostatic problem from [Perugia, Simoncini, Arioli, 1999], A is 1272×1272
- Inexact FOM, $m_\star = 120$, $\varepsilon = 10^{-4}$

Applications:
II. Inexact Preconditioning

$$Hx = b \quad \longrightarrow \quad H\mathcal{P}^{-1}\bar{x} = b, \quad x = \mathcal{P}^{-1}\bar{x}$$

$\mathcal{P}^{-1}$ not performed exactly (use Krylov method)

$H\mathcal{P}^{-1}v_k$ replaced with $H\tilde{z}_k, \quad \tilde{z}_k \approx \mathcal{P}^{-1}v_k$

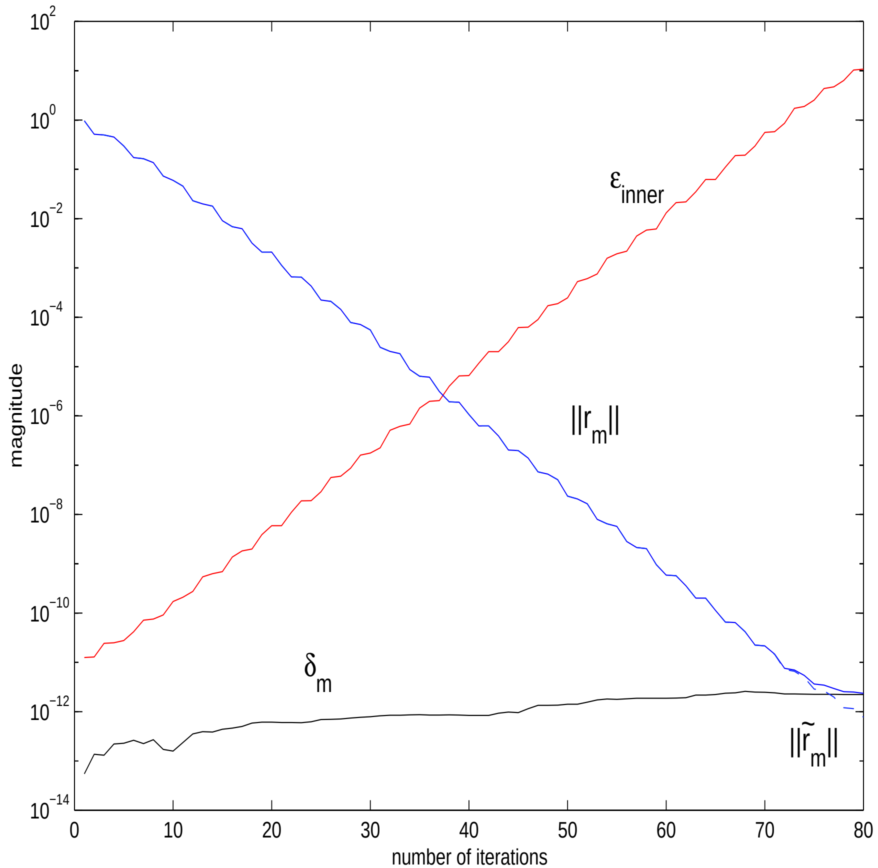
Arnoldi relation $\qquad H\mathcal{P}^{-1}V_m = V_{m+1}H_{m+1,m} \qquad$ is transformed into

$$H[\tilde{z}_1, \cdots, \tilde{z}_m] = V_{m+1}H_{m+1,m}.$$

Use Flexible Krylov subspace method

$r_k^{inner} = v_k - \mathcal{P}\tilde{z}_k$ inner residual

$$\|r_k^{inner}\| \le \frac{\sigma_{m_\star}(H_{m_\star})}{\|H\mathcal{P}^{-1}\|m_\star} \frac{1}{\|\tilde{r}_{k-1}^{gm}\|}\varepsilon \equiv \varepsilon_{inner}$$



For same 2D saddle point, use $\mathcal{P} = \begin{bmatrix} I & 0 \\ 0 & B^T B \end{bmatrix}$. Solve $B^T B p_k = rhs$ iteratively, $m_\star = 80$, $\varepsilon = 10^{-9}$, tolerance ε_{inner}

Some CPU Times: Same Magnetostatic 2D Problem

Outer tolerance: 10^{-8}

Elapsed Time

CPU in seconds of a Sun Enterprise 4500 (Fortran code)
(4 CPU 400MHertz, 2GBytes RAM) CG iterations.

Problem Size	Fixed Inner Tol $\varepsilon_{inner} = 10^{-10}$	Var. Inner Tol. $10^{-10}/\|r\|$	Var. Inner Tol. $10^{-12}/\|r\|$
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)

Applications:
III. Parabolic Control Problems (W i P)
First Example

Inverse problem: Recover control $u(x)$ based on field (state) $z(x)$ related by the forward problem (3D):

$$\triangle z = z_t, \qquad x \epsilon \Omega$$

$$z = u, \qquad x \epsilon \partial\Omega$$

$$z = z_0, \qquad x \epsilon \Omega/\partial\Omega, \qquad \text{for} \;\; t = 0$$

Discretized forward problem (FD)

$$E\mathbf{z} - \delta t N u = c.$$

$$\underbrace{\begin{bmatrix} B & & & & \\ -I & B & & & \\ & -I & B & & \\ & & \ddots & \ddots & \\ & & & -I & B \end{bmatrix}}_{E} \underbrace{\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ \\ z_s \end{bmatrix}}_{\mathbf{z}} - \delta t \underbrace{\begin{bmatrix} M \\ M \\ \vdots \\ \\ M \end{bmatrix}}_{N} u = \underbrace{\begin{bmatrix} z_0 \\ 0 \\ \vdots \\ \\ 0 \end{bmatrix}}_{c}$$

where $z_i \approx z(t_i)$, $B = (I + \delta t A_h)$, with A_h discretization of $\triangle$.

Optimization problem

$$\begin{aligned} \min \qquad & \phi = \frac{1}{2}\|Q\mathbf{z} - d^{obs}\|^2 \\ \text{subject to} \qquad & E\mathbf{z} - \delta t N u = c. \end{aligned}$$

Lagrangian $\quad L(\mathbf{z}, u, \lambda) = \frac{1}{2}\|Q\mathbf{z} - d^{obs}\|^2 + \lambda^T(E\mathbf{z} - \delta t N u - c)$

Linearize to obtain

$$\begin{bmatrix} Q^TQ & 0 & E^T \\ 0 & 0 & N^T \\ E & N & 0 \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ u \\ \lambda \end{bmatrix} = - \begin{bmatrix} L_u \\ L_m \\ L_\lambda \end{bmatrix}$$

Reduced Hessian

After elimination one has $Hu = -p$

$$H\ u = N^T E^{-T} Q^T Q E^{-1} N\ u = -p.$$

Use, e.g., with inexact CG, approximating each of the the systems with E and E^T with CG with varying (increasing) tolerance.

MVP Hv

1. Multiply Nv
2. Solve $Ez = Nv$ by solving $Ez = Nv$ with an inner tolerance ϵ_{in_1}
3. Multiply Qz
4. Multiply $Q^T Qz$
5. Solve $E^T w = Q^T Qz$ by solving with an inner tolerance ϵ_{in_2}
6. Compute $N^T w$

Experiments

$16 \times 16 \times 16$ grid. control u of order 3375, 10 time steps.

fixed 10^{-14}	fixed 10^{-7}	decreasing $10^{-3} \cdot \|\tilde{r}_{k-1}\|$	increasing $10^{-8}/\|\tilde{r}_{k-1}\|$
$35/23812$	$41/15250$	$48/18982$	$47/8689$

Outer iterations / total inners = total matvecs with Laplacian.

Outer $\varepsilon = 10^{-7}$

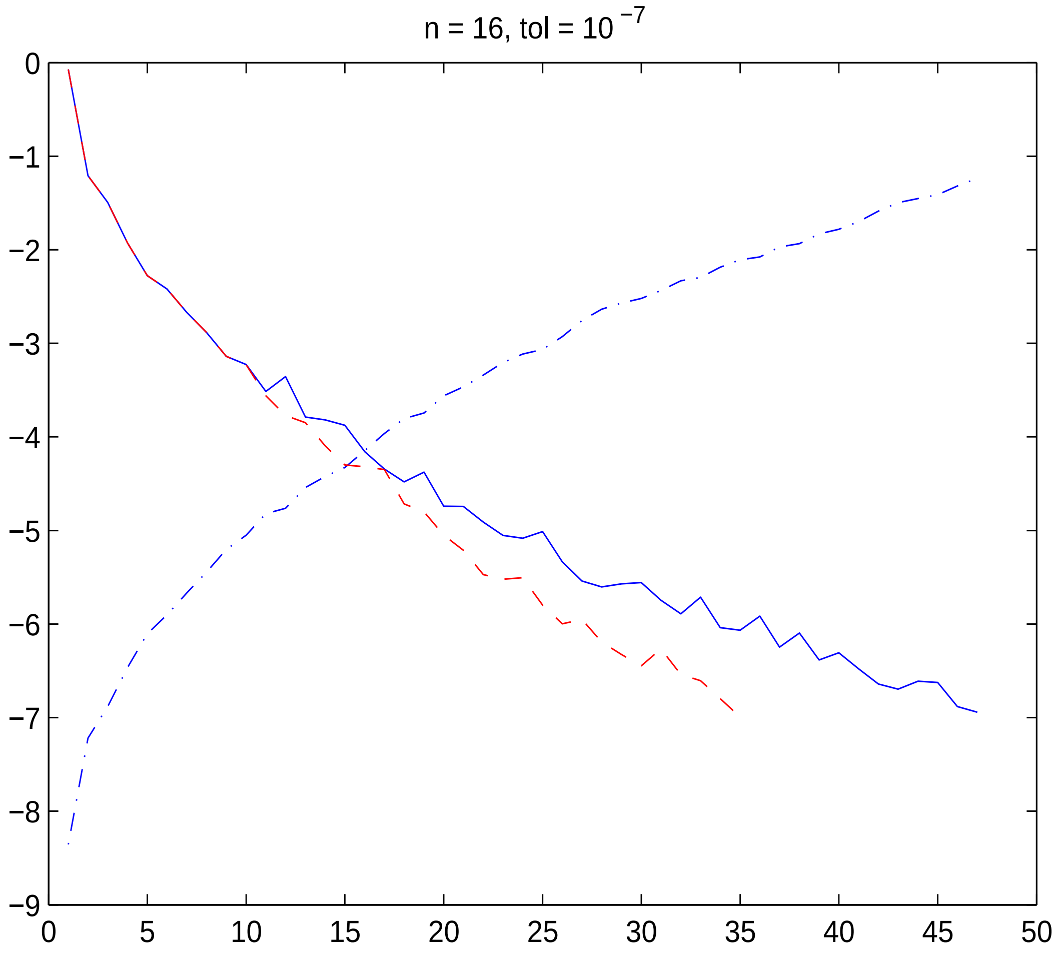
There is a "delay"

12 more outer iter. than "exact", 6 more than fixed

but savings of 64%, and 43%

Illustration of “delay”, cheaper by a factor of about THREE

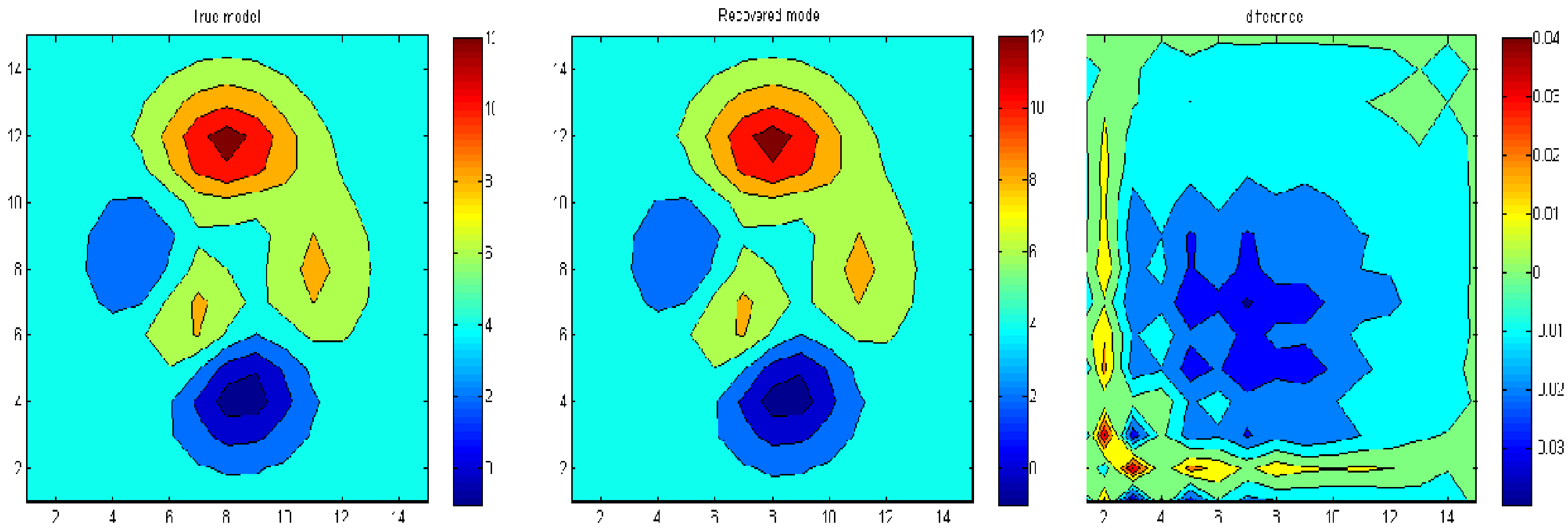
- - - - exact CG, —— inexact CG, — · — · — ε_{inner}



One surface of true and recovered model,

and their difference

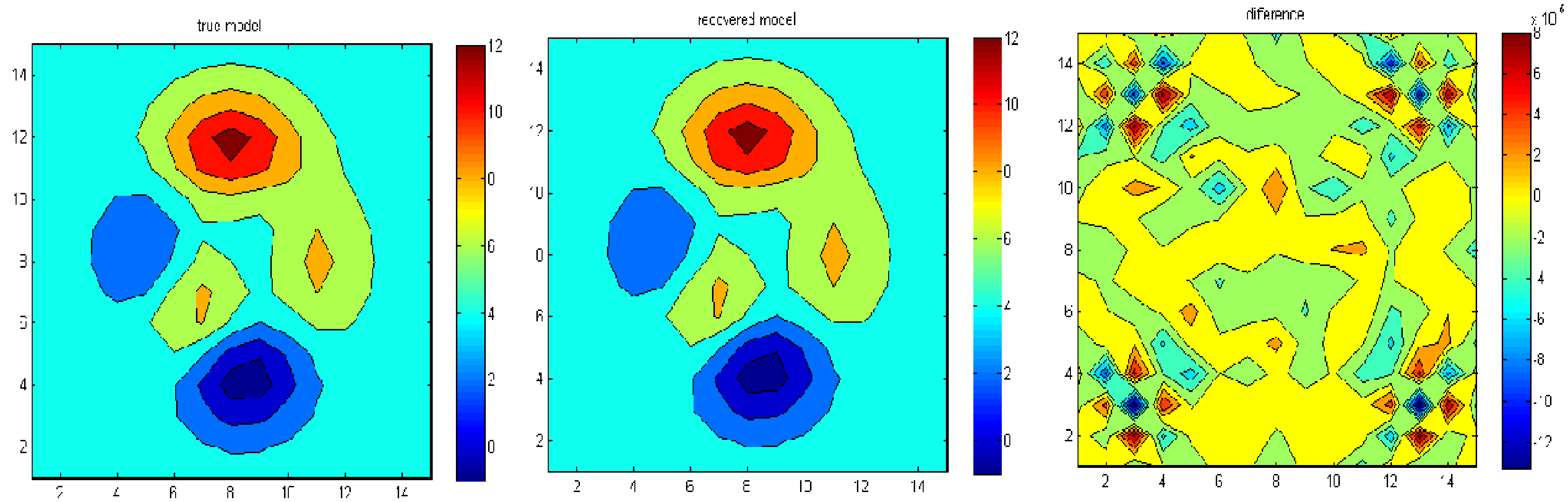
decreasing $\epsilon_{inner} = 10^{-3} \cdot \|\tilde{r}_{k-1}\|$



error O(10^{-3})

One surface of true and recovered model, and their difference

increasing $\epsilon_{inner} = 10^{-8}/\|\tilde{r}_{k-1}\|$



error $\mathrm{O}(10^{-6})$

Parabolic Control Problems, Second Example

General Lagrangian (using FEM)

$$\mathcal{L}_h(\mathbf{z},\mathbf{u},\mathbf{p}) = \frac{1}{2}(\mathbf{e}^T\mathbf{K}\mathbf{e}^T + \mathbf{u}^T\mathbf{G}\mathbf{u}) + \mathbf{p}^T(\mathbf{E}\mathbf{z} + \mathbf{N}\mathbf{u} - \mathbf{f})$$

Reduced system: $\mathbf{H}\mathbf{u} := (\mathbf{G} + \mathbf{N}^T\mathbf{E}^{-T}\mathbf{K}\mathbf{E}^{-1}\mathbf{N})\mathbf{u} = \mathbf{b}_u$

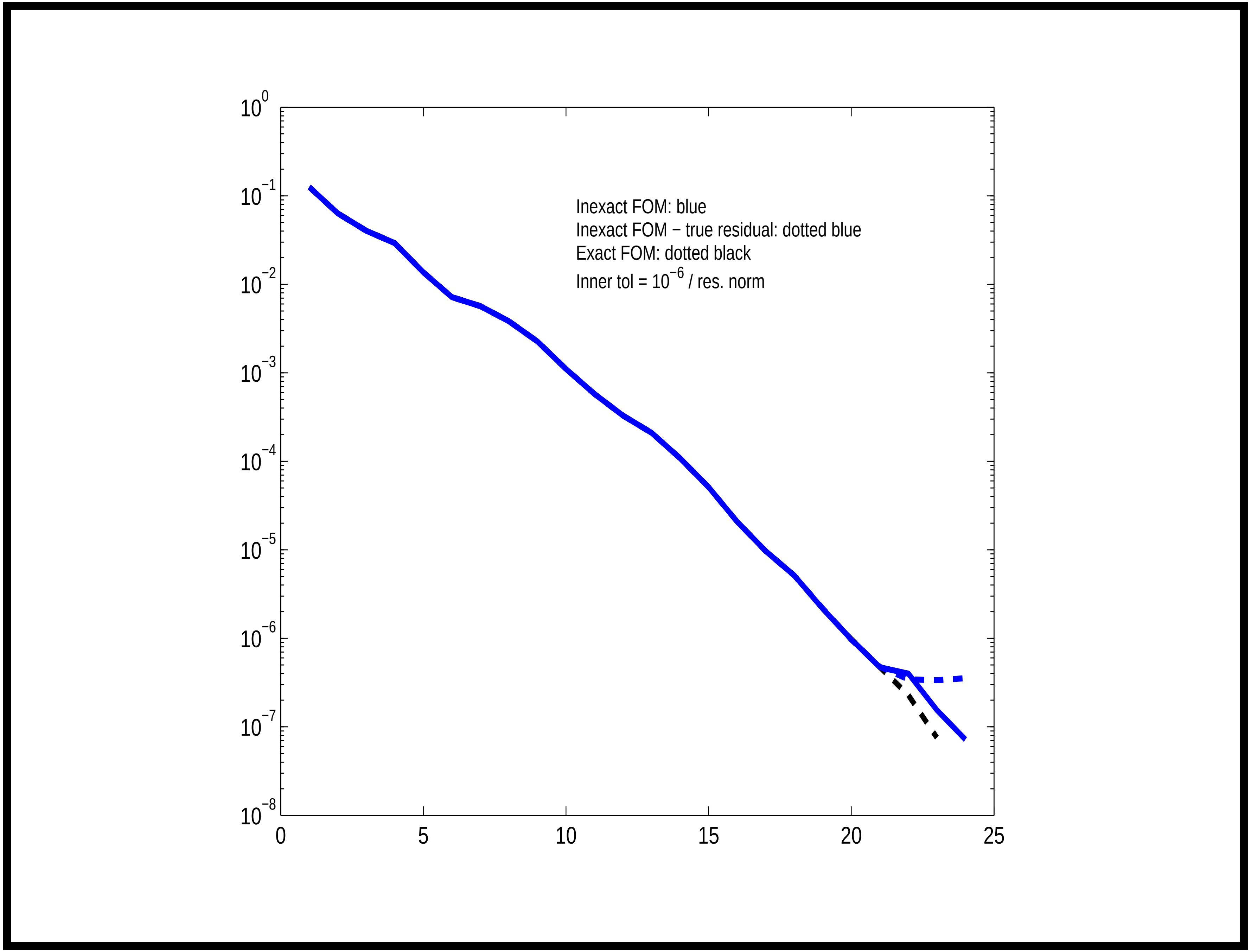
$$\mathbf{E} = \begin{bmatrix} F_h & & & \\ -M_h & F_h & & \\ & \ddots & \ddots & \\ & & -M_h & F_h \end{bmatrix}$$

$F_h = M_h + \delta t A_h$

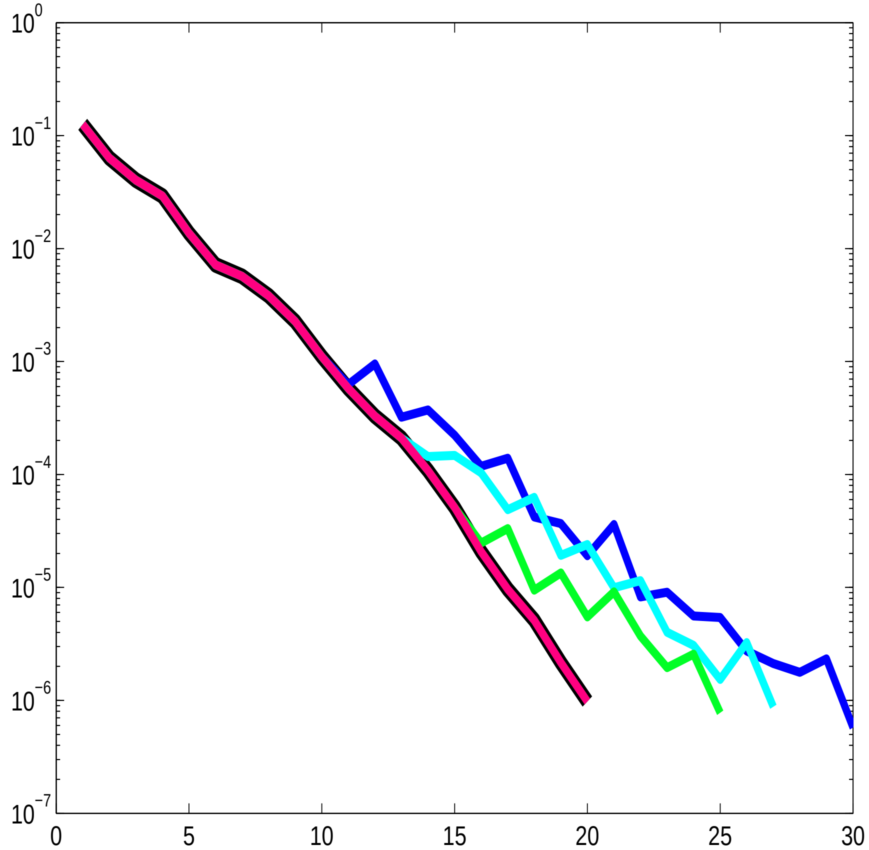
Here we approximate $\mathbf{E}$ with $\mathbf{E}_n$, n sweeps of the Parareal Algorithm

We use our theory to find ε_{inner} which determine how many sweeps we use.

Example. Find u so that z is closest to z_*, subject to $z_t - z_{xx} = u$, $0 < x < 1$, $t > 0$. with initial and boundary data.
Discretize $\delta x = 1/16$ and $\delta t = 1/64$. System size 1024.



Inexact FOM: blue
Inexact FOM − true residual: dotted blue
Exact FOM: dotted black
Inner tol = 10^−6 / res. norm
10^0
10^−1
10^−2
10^−3
10^−4
10^−5
10^−6
10^−7
10^−8
0
5
10
15
20
25



	sweeps (outer it.)
IFOM	161 (20)
$m = 12$	188 (25)
$m = 8$	201 (27)
$m = 1$	220 (30)

Computed residual: Inexact truncated FOM, semiband $m = 20$, $m = 8$ and $m = 1$ (ICG) (blue).

For the stopping criteria we use $\ell_n^{(1)} = \ell_n^{(2)} = 1 \; (10^{-6}\|r_0\|/\|r_{m-1}\|)$

Conclusions

- Inexact matrix-vector product (or inexact preconditioning) might be worth trying for your problem
- Truncated methods might be worth trying for your problem

With Valeria Simoncini:

Theory of Inexact Krylov Subspace Methods and Applications to Scientific Computing
SIAM J. Scientific Computing, v. 25 (2003) 454–477.

On the Occurrence of Superlinear Convergence of Exact and Inexact Krylov Subspace Methods
SIAM Review, v. 47 (2005) 247–272.

The Effect of Non-Optimal Bases on the Convergence of Krylov Subspace Methods
Numerische Mathematik, v. 100 (2005) 711-733.

Recent computational developments in Krylov Subspace Methods for linear systems
Numerical Linear Algebra with Applications, v. 14 (2007) 1-59.

All available at: `http://www.math.temple.edu/~szyld`

Watch for forthcoming reports on the control problems.