Modern Krylov subspace methods
(and applications to parabolic control problems)

Daniel B. Szyld
Temple University, Philadelphia

Scientific and Statistical Computing Seminar
University of Chicago
25 October 2012
Collaborators

Xiuhong Du, Alfred University
Marcus Sarkis, Worcester Poly. Inst., and IMPA, Rio de Janeiro
Christian Schaerer, Universidad Nacional de Asunción
Valeria Simoncini, Università di Bologna
Plan for the talk

- Introduction to Truncated Krylov subspace methods
- Inexact Krylov subspace methods: Introduction and Applications
- Special case of parabolic control problems
General Problem Statement

Solve a system

\[ Hx = b, \]

\( H \) Hermitian or non-Hermitian

using Krylov subspace iterative methods
Krylov subspace methods

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

Subspaces are nested: \( \mathcal{K}_m \subset \mathcal{K}_{m+1} \).
Given \( x_0, r_0 = b - Hx_0 \), find approximation

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

satisfying some property.
Krylov subspace methods (cont.)

Conditions:

Galerkin, e.g., FOM, CG:

\[ b - Hx_m \perp \mathcal{K}_m(H, r_0) \]

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

\[ b - Hx_m \perp HK_m(H, r_0) \]

or equivalently

\[ x_m = \arg \min \{ \| b - Hx \|_2 \}, \quad x \in x_0 + \mathcal{K}_m(H, r_0) \]
Krylov subspace methods (cont.)
Some implementation issues

- 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.
- One can of course run Gram-Schmidt on\
  $\{r_0, Hr_0, H^2r_0, \ldots, H^{m-1}r_0\}$ (not advised).

  Instead, Arnoldi[1951] (for general $H$) said:
  $v_1 = r_0$ normalized
  $v_2 = Hv_1 - \langle v_1, Hv_1 \rangle v_1$ normalized,
  $\text{span}\{v_1, v_2\} = \text{span}\{r_0, Hr_0\}$
  $v_3 = Hv_2 - \langle v_2, Hv_2 \rangle v_2 - \langle v_1, Hv_2 \rangle v_1$ normalized,
  $\text{span}\{v_1, v_2, v_3\} = \text{span}\{r_0, Hr_0, H^2r_0\}$
  etc.
Arnoldi method

Let $\beta = \|r_0\|$, and $v_1 = r_0 / \beta$.

For $k = 1, \ldots$

Compute $Hv_k$, then $v_{k+1} h_{k+1,k} = Hv_k - \sum_{j=1}^{k} v_j h_{jk}$,

where $h_{jk} = \langle v_j, Hv_k \rangle$, $j \leq k$,

and $h_{k+1,k}$ is positive and such that $\|v_{k+1}\| = 1$.

In practice: Modified Gram-Schmidt or Householder orthogonalization

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

$$HV_m = V_{m+1} H_{m+1,m}$$

$H_{m+1,m}$ is $(m + 1) \times m$ upper Hessenberg
Arnoldi relation (cont.)

\[
H_{m+1,m} = \begin{bmatrix}
    h_{11} & h_{12} & h_{13} & \cdots & h_{1m} \\
    h_{21} & h_{22} & h_{23} & \cdots & h_{2m} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    \vdots & \vdots & \vdots & \ddots & \ddots \\
    h_{m,m-1} & h_{mm} & h_{m+1,m}
\end{bmatrix} = \begin{bmatrix}
    H_m \\
    h_{m+1,m}e_T^m
\end{bmatrix}
\]

\[
HV_m = V_{m+1}H_{m+1,m} = V_mH_m + h_{m+1,m}v_{m+1}e_T^m
\]
Krylov subspace methods (cont.)

More implementation issues

Element in $\mathcal{K}_m(H, v_1)$ is a linear combination of $v_1, v_2, \ldots, v_m$, i.e., of the form $V_m y$, $y \in \mathbb{R}^m$

Each method finds $y = y_m$ and we have $x_m = V_m y_m$

For FOM, or CG, we have the Galerkin condition $r_m \perp \mathcal{K}_m$, i.e.,

$$0 = V_m^T (b - H x_m) = V_m^T b - V_m^T H V_m y_m = \beta e_1 - H_m y_m$$
**FOM:** Full Orthogonalization Method [Saad, 1978]

Use Arnoldi methods to construct $V_m, H_m$.

Solve $H_m y_m = \beta e_1$

Approximation is $x_m = V_m y_m$.

Test for convergence: Is $\|r_m\| < \varepsilon$?

**Cost:** one matrix-vector product per step,
at step $m$, orthogonalization ($m$ inner-products),
one solution of small $m \times m$ upper Hessenberg matrix (LU with no fill).
($\|r_m\|$ cheap or free - no details here).

**Storage:** at step $m$, $m$ vectors $v_1, v_2, \ldots, v_m$. 
GMRES implementation

We use Arnoldi methods to obtain $V_m$, and as before, element in $K_m(H, v_1)$ is of the form $V_m y$, $y \in \mathbb{R}^m$.

Recall: $\beta = \|r_0\|$, $HV_m = V_m H_{m+1,m}$

\[
b - H x_m = r_0 - HV_m y_m = \beta v_1 - V_{m+1} H_{m+1,m} y = V_{m+1}(\beta e_1 - H_{m+1,m} y)
\]

\[
\|r_m\| = \min_{x \in K_m} \|b - H x\| = \min_{y \in \mathbb{R}^m} \|\beta e_1 - H_{m+1,m} y\|
\]

QR factorization $H_{m+1,m} = Q_{m+1} R_{m+1,m}$, $R_{m+1,m} = \begin{bmatrix} R_m \\ 0 \end{bmatrix}$,

\[
\|r_m\| = \min_{y \in \mathbb{R}^m} \|Q_{m+1}^T \beta e_1 - R_{m+1,m} y\|.
\]
\[ \| r_m \| = \min_{y \in \mathbb{R}^m} \| Q_{m+1}^T \beta e_1 - R_{m+1,m} y \|. \]

\[ Q_{m+1}^T \beta e_1 = \begin{bmatrix} t_m \\ \rho_{m+1} \end{bmatrix} \]

Then, \( y_m = R_m^{-1} t_m \), and \( x_m = V_m y_m \).
Furthermore, \( \| r_m \| = \| Q_{m+1}^T \beta e_1 - R_{m+1,m} y_m \| = |\rho_{m+1}| \)

Use this computed residual for stopping
(may deviate from true residual!)

Use Givens rotations for QR, save rotations from previous steps.
Only two entries per step needed.
GMRES

Cost: one matrix-vector product per step, at step $m$, orthogonalization ($m$ inner-products), rotations for QR, solution of $m \times m$ triangular system.

Storage: at step $m$, $m$ vectors $v_1, v_2, \ldots, v_m$.

Residual norms monotonically nonincreasing, but stagnation possible. Superlinear convergence can be observed.
• **Main costs:**
  1. Matrix-vector product: $Hv_k$
  2. Orthogonalization
  3. Storage (if there is no recursion)

• One popular alternative: Restarted methods. Hit and miss. Little theory. [Saad, 1996], [Simoncini, 2000]

• It may happen than larger value of restart parameter is less efficient! [Embree, 2003]
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)

• Apply all this to Parabolic Control Problems

• Use a Parareal-in-time approximation
Truncated Krylov subspace methods

For the same amount of storage (max \( \ell \) vectors), instead of restarting:
Only orthogonalize with respect of the previous \( \ell \) vectors.
In Arnoldi we have then:

\[
v_{k+1} h_{k+1,k} = H v_k - \sum_{j=\max\{1,k-\ell+1\}}^{k} v_j h_{j,k},
\]

where \( h_{j,k} = \langle v_j, H v_k \rangle, \ j \leq k, \)
and \( h_{k+1,k} \) is positive and such that \( \|v_{k+1}\| = 1. \)

Only need to store these previous \( \ell \) vectors.
Upper Hessenberg matrix is now banded (upper bandwidth $\ell$).

\[ H_{m+1,m} = \begin{bmatrix}
  h_{11} & h_{12} & h_{13} \\
  h_{21} & h_{22} & h_{23} \\
  & & \ddots \ddots \ddots \\
  & & & \ddots & \ddots \ddots \\
  & & & & & h_{m,m-1} & h_{mm} \\
  & & & & & & h_{m+1,m}
\end{bmatrix} \]
Truncated Krylov subspace methods (cont.)

• Truncated GMRES [Saad and Wu, 1996]
  Truncated FOM [Saad, 1981], [Jia, 1996]

• Basis of $K_m(H, r_0), v_1, v_2, \ldots, v_m \ (m > \ell)$ is not orthogonal, but $x_m \in K_m(H, r_0)$, and minimization (or Galerkin condition) is over the whole space.

• $H_{m+1,m}$ banded with upper semiband $\ell - 2$.
  Matrix with basis vectors $V_m$ not orthogonal.
  Can be implemented so that only $O(\ell)$ vectors are stored.

• Extreme case, $\ell = 3$, $H_{m+1,m}$ 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, \( n = 2500 \).

\[
\begin{array}{c}
\text{GMRES, Truncated } \ell = 3.
\end{array}
\]
Inexact Krylov subspace methods

• At the $k$th iteration of the Krylov space method use

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

where $\|D_k\|$ can be monitored

• Two examples now:
  - Schur complement, where the inverse is approximated
  - Inexact preconditioning

• [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 [Golub and Ye, 1999], [Notay, 1999], [Sleijpen and van der Eshof, 2004] and [Simoncini and Eldén, 2002]
Inexact Krylov (cont.)

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

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

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

- 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
[Simoncini and Szyld, 2003]

True residual: \( r_m = b - Hx_m = r_0 - HV_{m,y} \)

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

Let \( \varepsilon > 0 \). If for every \( k \leq m \),

\[
\|D_k\| \leq \frac{\sigma_{\min}(H_{m,*})}{m_*} \frac{1}{\|\tilde{r}_{k-1}\|} \varepsilon \equiv \ell_F^m \frac{1}{\|\tilde{r}_{k-1}\|} \varepsilon ,
\]

then \( \|V_{m}^T r_m\| \leq \varepsilon \) and \( \|r_m - \tilde{r}_m\| \leq \varepsilon \).

\( m_* \) being the maximum number of iterations allowed

Similar results for inexact GMRES
see also [Giraud, Gratton, Langou, 2007]
Theorem for Inexact Truncated FOM

\[ \| D_k \| \leq \frac{\sigma_{\min}(H_{m_\bullet}) \sigma_{\min}(V_m)}{m_\bullet} \frac{1}{\| \tilde{r}_{k-1} \|} \varepsilon \equiv \ell_{m}^{TF} \frac{1}{\| \tilde{r}_{k-1} \|} \varepsilon, \]

implies \[ \| V_{m}^T r_{m} \| \leq \varepsilon \] and \[ \delta_{m} = \| r_{m} - \tilde{r}_{m} \| \leq \varepsilon. \]

Notes:

- This result applies in particular to Inexact CG
- \( \ell_{m} \) can be estimated from problem, if information is available.
First Experiment

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

- Our condition (e.g. for FOM)

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

is very conservative. In most cases it is too strict. However, \(\sigma_{\text{min}}(H)\) does play a role.
CG: condition $\|D_k\| \leq \frac{\sigma_{\text{min}}(H)}{m_*} \frac{1}{\|\tilde{r}_k - 1\|} \varepsilon$

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

I. Schur complement systems

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

\[B^T A^{-1} B x = B^T A^{-1} f; \quad A w = f - B x\]

\[H x = b\]

\(A^{-1}\) 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 $j$th (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

$$\|r_k^{inner}\| < \frac{\sigma_{m_\ast}(H_{m_\ast})}{\|B^T A^{-1}\| m_\ast} \frac{1}{\|\tilde{r}_{k-1}^{fom}\|} \varepsilon \equiv \varepsilon_{inner}$$
• Two-dim. saddle point magnetostatic problem from [Perugia, Simoncini, Arioli, 1999], $A$ is $1272 \times 1272$

• Inexact FOM, $m_\star = 120$, $\varepsilon = 10^{-4}$
Applications:

II. Inexact Preconditioning

\[ Hx = b \quad \rightarrow \quad H\mathcal{P}^{-1}\bar{x} = b, \quad x = \mathcal{P}^{-1}\bar{x} \]

\(\mathcal{P}^{-1}\) 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 \(H\mathcal{P}^{-1}V_m = V_{m+1}H_{m+1,m}\) 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

\[ \|r_k^{inner}\| \leq \frac{\sigma_{m^*}(H_{m^*})}{\|H\mathcal{P}^{-1}\| m^*} \frac{1}{\|\tilde{r}_k^{gm}\|} \varepsilon \equiv \varepsilon^{inner} \]
For same 2D saddle point, use $\mathcal{P} = \begin{bmatrix} I & 0 \\ 0 & B^T B \end{bmatrix}$. Solve

$$B^T B p_k = rhs \text{ iteratively, } m_\ast = 80, \varepsilon = 10^{-9}, \text{ tolerance } \varepsilon_{\text{inner}}.$$
Some CPU Times: Same Magnetostatic 2D Problem

Outer tolerance: $\varepsilon = 10^{-8}$

$$\| r_k^{\text{inner}} \| \leq \frac{c_0}{\| r_{m-1}^{\text{outer}} \|} \varepsilon \equiv \varepsilon_{\text{inner}}$$

$c_0$: Constant estimated a–priori: Here we use $10^{-2}$ and $10^{-4}$.

Elapsed Time

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

<table>
<thead>
<tr>
<th>Problem Size</th>
<th>Fixed Inner Tol $\varepsilon_{\text{inner}} = 10^{-10}$</th>
<th>Var. Inner Tol. $10^{-10}/|r|$</th>
<th>Var. Inner Tol. $10^{-12}/|r|$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3810</td>
<td>17.0 (54)</td>
<td>11.4 (54)</td>
<td>14.7 (54)</td>
</tr>
<tr>
<td>9102</td>
<td>82.9 (58)</td>
<td>62.8 (58)</td>
<td>70.7 (58)</td>
</tr>
<tr>
<td>14880</td>
<td>198.4 (54)</td>
<td>156.5 (54)</td>
<td>170.1 (54)</td>
</tr>
</tbody>
</table>
Applications:

III. Parabolic Control Problems

**Inverse problem:** Recover control $v(x)$ based on field (state) $z(x)$ related by the forward problem

$$z_t + A z = v, \quad x \in \Omega$$

$$z = g, \quad x \in \partial \Omega$$

$$z = z_0, \quad x \in \Omega / \partial \Omega, \quad \text{for} \ t = 0$$

$A$ elliptic, e.g., $A = -\triangle$

This is a distributed control problem. Similar techniques for boundary control problems (control $g$).
Associated variational problem

\[ J(z(v), v) := \frac{\alpha}{2} \int_{t_0}^{t_f} \|z(v)(t, \cdot) - \tilde{y}(t, \cdot)\|^2_{L^2(\Omega)} \]
\[ + \frac{\beta}{2} \|z(v)(t_f, \cdot) - \tilde{y}(t_f, \cdot)\|^2_{L^2(\Omega)} + \frac{\gamma}{2} \int_{t_0}^{t_f} \|v(t, \cdot)\|^2_{L^2(\Omega)}. \]

discretized as

\[ J_{h}^\tau(z, v) = \frac{1}{2} (z - \tilde{y})^T K(z - \tilde{y}) + \frac{1}{2} v^T Gv + (z - \tilde{y})^T g. \]
Discretized forward problem (FD)

\[ \mathbf{E} \mathbf{z} + \mathbf{N} \mathbf{v} = \mathbf{f} \]

\[ \mathbf{E} = \begin{bmatrix} F_1 & \cdot & \cdot & \cdot \\ -F_0 & F_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ -F_0 & F_1 \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} B & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & B \end{bmatrix} \]
Optimization problem

\[
\begin{align*}
\text{min} & \quad J^\tau_h(z, v) \\
\text{subject to} & \quad Ez + Nv = f
\end{align*}
\]

Lagrangian:

\[
J^\tau_h(z, v) + q^T(Ez + Nv - f)
\]
Linearize to obtain
\[
\begin{bmatrix}
K & 0 & E^T \\
0 & G & N^T \\
E & N & 0
\end{bmatrix}
\begin{bmatrix}
z \\
u \\
p
\end{bmatrix}
= \begin{bmatrix}
g \\
0 \\
f
\end{bmatrix}
\]

After elimination one obtains
\[
Hu := (G + N^T E^{-T} K E^{-1} N)u = b
\]

\(H\) being the spd reduced Hessian
\[ Hu = (G + N^T E^{-T} K E^{-1} N) u = b \]

We use inexact FOM, approximating each of the the systems with \( E \) and \( E^T \) with a Parareal method with varying (increasing) tolerance.

MVP Hv

1. Multiply \( Nv \)
2. Solve \( Ez = Nv \) by solving \( Ez = Nv \) with an inner tolerance \( \epsilon_{in_1} \)
3. Multiply \( Kz \)
4. Solve \( E^Tw = Kz \) by solving with an inner tolerance \( \epsilon_{in_2} \)
5. Compute \( N^Tw \)
Approximate solutions of systems with $E$ or $E^T$

\[ Hu = (G + N^T E^{-T} K E^{-1} N) u = b \]

We prove that with $\epsilon_{in_1} = \epsilon_{in_2}$, we have spectral equivalence of the form

\[ (v, Gv) \leq (v, Hv) \leq \mu(v, Gv) \]

We choose FOM, since it reduces to CG when we have full symmetry.
Sample Experiment: 2D heat equation

15 \times 15 grid. \tau = 1/512, control u of order 115200

Same relative tolerance for both Parareal systems, outer \varepsilon = 10^{-6}

<table>
<thead>
<tr>
<th>\ell_m^{(i)} \varepsilon</th>
<th>IFOM</th>
<th>TIFOM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>\begin{array}{cccc} m_T = 2 &amp; m_T = 4 &amp; m_T = 8 &amp; m_T = 12 \end{array}</td>
</tr>
</tbody>
</table>
| \begin{array}{c} 10^{-12} \\
10^{-10} \\
10^{-8} \\
10^{-7} \\
10^{-6} \\
10^{-5} \\
10^{-4} \end{array} | \begin{array}{c} 15(576) \\
15(482) \\
15(388) \\
15(340) \\
15(288) \\
17(238) \\
17(180) \end{array} | \begin{array}{c} 16(610) \\
17(532) \\
17(426) \\
18(394) \\
19(340) \\
24(298) \\
\text{n.c.} \end{array} | \begin{array}{c} 16(608) \\
17(528) \\
17(426) \\
17(374) \\
19(338) \\
21(266) \\
22(210) \end{array} | \begin{array}{c} 15(576) \\
16(504) \\
17(420) \\
17(368) \\
18(320) \\
19(258) \\
22(210) \end{array} | \begin{array}{c} 15(576) \\
15(482) \\
15(388) \\
15(340) \\
16(298) \\
19(242) \\
20(192) \end{array} |
One surface of true and recovered model, and their difference increasing $\varepsilon_{inner} = 10^{-5}/\|\tilde{r}_{k-1}\|$, $m_T = 8$
TIFOM Convergence curves

outer $\varepsilon = 10^{-6}$, inner factor $= 10^{-5}$
Similar results for more non-symmetric

<table>
<thead>
<tr>
<th>$\ell_m^{(1)} \varepsilon$</th>
<th>$\ell_m^{(2)} \varepsilon$</th>
<th>IFOM o-iter. (i-iter)</th>
<th>$m_T = 2$</th>
<th>$m_T = 4$</th>
<th>$m_T = 8$</th>
<th>$m_T = 12$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-6}$</td>
<td>$10^{-7}$</td>
<td>15(288)</td>
<td>18(330)</td>
<td>19(338)</td>
<td>17(310)</td>
<td>16(298)</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>$10^{-6}$</td>
<td>15(288)</td>
<td>19(340)</td>
<td>19(338)</td>
<td>18(320)</td>
<td>16(298)</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>$10^{-5}$</td>
<td>16(300)</td>
<td>23(390)</td>
<td>20(346)</td>
<td>19(334)</td>
<td>17(310)</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>$10^{-4}$</td>
<td>16(300)</td>
<td>51(710)</td>
<td>37(520)</td>
<td>20(352)</td>
<td>17(310)</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>$10^{-7}$</td>
<td>15(234)</td>
<td>19(270)</td>
<td>27(280)</td>
<td>18(254)</td>
<td>21(246)</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>$10^{-6}$</td>
<td>15(234)</td>
<td>19(270)</td>
<td>29(284)</td>
<td>18(254)</td>
<td>19(242)</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>$10^{-5}$</td>
<td>17(238)</td>
<td>24(298)</td>
<td>21(266)</td>
<td>19(258)</td>
<td>19(242)</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>$10^{-4}$</td>
<td>17(240)</td>
<td>30(334)</td>
<td>31(356)</td>
<td>20(270)</td>
<td>19(244)</td>
</tr>
</tbody>
</table>
One surface of true and recovered model, and their difference, outer tolerance $\varepsilon = 10^{-6}$

$\varepsilon_{in1} = 10^{-6}/\|\tilde{r}_{k-1}\|$, $\varepsilon_{in4} = 10^{-4}/\|\tilde{r}_{k-1}\|$, $m_T = 8$

True  Computed  error $O(10^{-6})$
Conclusions

- Inexact matrix-vector product (or inexact preconditioning) might be worth trying for your problem
- Truncated methods might be worth trying for your problem
- New work on inexact and truncated for parabolic control problems (with two inner criteria)
With Valeria Simoncini:

Theory of Inexact Krylov Subspace Methods and Applications to Scientific Computing


On the Occurrence of Superlinear Convergence of Exact and Inexact Krylov Subspace Methods


The Effect of Non-Optimal Bases on the Convergence of Krylov Subspace Methods


Recent computational developments in Krylov Subspace Methods for linear systems


All available at:  http://www.math.temple.edu/szyld
With Xiuhond Du, Marcus Sarkis and Christian E. Schaerer

Inexact and truncated Parareal-in-time Krylov subspace methods for parabolic optimal control problems
Research Report 12-02-06, February 2012

Also available at: http://www.math.temple.edu/szyld