Nested Krylov methods for shifted linear systems

M. Baumann∗,† and M. B. van Gijzen†

∗Email: M.M.Baumann@tudelft.nl
†Delft Institute of Applied Mathematics
Delft University of Technology
Delft, The Netherlands

October 26, 2015

SIAM Conference on Applied Linear Algebra
Atlanta, Georgia, US
Context of this work (1/3)

**Why are we all here?**

**Definition: Shifted linear systems**

We want to *efficiently* solve

\[(\mathcal{A} - \sigma_k I)x_k = b\]

for multiple shifts \(\{\sigma_1, \ldots, \sigma_{N_\sigma}\}, N_\sigma \in \mathbb{N}\).

The *efficiency* of our numerical method heavily relies on the shift-invariance property of Krylov subspaces,

\[\mathcal{K}_m(\mathcal{A}, b) \equiv \text{span}\{b, \mathcal{A}b, \ldots, \mathcal{A}^{m-1}b\} = \mathcal{K}_m(\mathcal{A} - \sigma I, b).\]
Why are we all here?

Definition: Shifted linear systems

We want to **efficiently** solve

$$(A - \sigma_k I)x_k = b$$

for multiple shifts $\{\sigma_1, \ldots, \sigma_{N_\sigma}\}, N_\sigma \in \mathbb{N}$. 

The **efficiency** of our numerical method heavily relies on the shift-invariance property of Krylov subspaces,

$$\mathcal{K}_m(A, b) \equiv \text{span}\{b, Ab, \ldots, A^{m-1}b\} = \mathcal{K}_m(A - \sigma I, b).$$
Context of this work (2/3)

Where it all began?

References


Their combined message is:

- Many shifted Laplace preconditioners can be used in a flexible multi-shift Krylov method (*Tania et al.*)

- Polynomial preconditioners preserve shift-invariance (*Mian et al.*)
Context of this work (2/3)

Where it all began?

References


Their combined message is:

- Many shifted Laplace preconditioners can be used in a flexible multi-shift Krylov method (*Tania et al.*)
- Polynomial preconditioners preserve shift-invariance (*Mian et al.*)
What does the present work contribute?

We present a nested Krylov solver for shifted linear systems:

1. Use a multi-shift Krylov method as an inner method (the preconditioner),
2. use a flexible multi-shift Krylov method as an outer method,
3. apply a single shifted Laplace preconditioner as a first layer.

In this talk, we will concentrate on 1-2, and comment on 3 when presenting numerical tests.
What does the present work contribute?

We present a nested Krylov solver for shifted linear systems:

1. Use a multi-shift Krylov method as an inner method (the preconditioner),
2. use a flexible multi-shift Krylov method as an outer method,
3. apply a single shifted Laplace preconditioner as a first layer.

In this talk, we will concentrate on 1-2, and comment on 3 when presenting numerical tests.
Nested (inner-outer) multi-shift Krylov methods

Solve the shifted linear systems

\[(A - \sigma_k I) x_k = b, \quad k = 1, ..., N_\sigma,\]

with a nested Krylov method based on

\[K_m(A, r_0) = K_m(A - \sigma I, r_0) \quad \forall \sigma.\]
Nested (inner-outer) multi-shift Krylov methods

Solve the shifted linear systems

\[(A - \sigma_k I) x_k = b, \quad k = 1, ..., N_{\sigma},\]

with a nested Krylov method based on

\[\mathcal{K}_m(A, r_0) = \mathcal{K}_m(A - \sigma I, r_0) \quad \forall \sigma.\]
Nested (inner-outer) multi-shift Krylov methods

Solve the shifted linear systems

\[(A - \sigma_k I)x_k = b, \quad k = 1, ..., N_\sigma,\]

with a nested Krylov method based on

\[\mathcal{K}_m(A, r_0) = \mathcal{K}_m(A - \sigma I, r_0) \quad \forall \sigma.\]
Nested (inner-outer) multi-shift Krylov methods

Solve the shifted linear systems

$$(A - \sigma_k I) x_k = b, \quad k = 1, ..., N_\sigma,$$

with a nested Krylov method based on

$$\mathcal{K}_m(A, r_0) = \mathcal{K}_m(A - \sigma I, r_0) \quad \forall \sigma.$$

msFOM $\xrightarrow{\text{early truncation}}$ collinear residuals $\xrightarrow{\text{msFGMRES}}$ outer_iter++
Nested (inner-outer) multi-shift Krylov methods

Solve the shifted linear systems

\[(A - \sigma_k I) x_k = b, \quad k = 1, ..., N_\sigma,\]

with a nested Krylov method based on

\[\mathcal{K}_m(A, r_0) = \mathcal{K}_m(A - \sigma I, r_0) \quad \forall \sigma.\]
Nested (inner-outer) multi-shift Krylov methods

Solve the shifted linear systems

$$(A - \sigma_k I)x_k = b, \quad k = 1, ..., N_\sigma,$$

with a nested Krylov method based on

$$\mathcal{K}_m(A, r_0) = \mathcal{K}_m(A - \sigma I, r_0) \quad \forall \sigma.$$
Nested (inner-outer) multi-shift Krylov methods

Solve the shifted linear systems

$$(\mathcal{A} - \sigma_k I) x_k = b, \quad k = 1, ..., N_\sigma,$$

with a nested Krylov method based on

$$\mathcal{K}_m(\mathcal{A}, r_0) = \mathcal{K}_m(\mathcal{A} - \sigma I, r_0) \quad \forall \sigma.$$

msFOM \quad \text{early truncation} \quad \text{collinear residuals} \quad \text{msFGMRES} \quad \text{outer_iter++}
Classical result: In FOM, the residuals are:
\[ r_j = b - Ax_j = ... = -h_{j+1,j} e_j^T y_j v_{j+1}. \]

Thus, for the shifted residuals it holds:
\[ r_j^{(\sigma)} = b - (A - \sigma I)x_j^{(\sigma)} = ... = -h_{j+1,j}^{(\sigma)} e_j^{(\sigma)} y_j^{(\sigma)} v_{j+1}. \]

Hence, we obtain collinear residuals,
\[ r_j^{(\sigma)} = \gamma r_j, \]
with factor \( \gamma = y_j^{(\sigma)}/y_j. \)

Reference
Inner method: multi-shift FOM

Classical result: In FOM, the residuals are:

\[ r_j = b - Ax_j = ... = -h_{j+1,j} e_j^T y_j v_{j+1}. \]

Thus, for the shifted residuals it holds:

\[ r_j^{(\sigma)} = b - (A - \sigma I)x_j^{(\sigma)} = ... = -h_{j+1,j}^{(\sigma)} e_j^T y_j^{(\sigma)} v_{j+1}. \]

Hence, we obtain collinear residuals,

\[ r_j^{(\sigma)} = \gamma r_j, \]

with factor \( \gamma = y_j^{(\sigma)}/y_j. \)

Reference

Outer method: flexible multi-shift GMRES (1/3)

Use flexible GMRES in the outer loop,

$$(A - \sigma I)\hat{V}_m = V_{m+1}H^{(\sigma)}_m,$$

where one column yields:

$$(A - \sigma I) P(\sigma)^{-1}v_j = V_{m+1}h^{(\sigma)}_j, \quad 1 \leq j \leq m.$$ 

Recap: The “inner loop” is the truncated solution of $(A - \sigma I)$ with right-hand side $v_j$ using msFOM.
Outer method: flexible multi-shift GMRES (1/3)

Use flexible GMRES in the outer loop,

\[(A - \sigma I) \hat{V}_m = V_{m+1} H_{m}^{(\sigma)},\]

where one column yields:

\[(A - \sigma I) P(\sigma)_j^{-1} v_j = V_{m+1} h_j^{(\sigma)}, \quad 1 \leq j \leq m.\]

Recap: The “inner loop” is the truncated solution of \((A - \sigma I)\) with right-hand side \(v_j\) using msFOM.
The inner residuals are:

\[ r^{(\sigma)}_j = v_j - (A - \sigma I)P(\sigma)^{-1}v_j, \]
\[ r_j = v_j - AP_j^{-1}v_j. \]

Imposing \( r^{(\sigma)}_j = \gamma r_j \) yields:

\[(A - \sigma I)P(\sigma)^{-1}v_j = \gamma AP_j^{-1}v_j - (\gamma - 1)v_j\]
Outer method: flexible multi-shift GMRES (2/3)

The inner residuals are:

\[
\begin{align*}
\mathbf{r}^{(\sigma)}_j &= \mathbf{v}_j - (\mathbf{A} - \sigma \mathbf{I}) \mathcal{P}^{(\sigma)}_j^{-1} \mathbf{v}_j, \\
\mathbf{r}_j &= \mathbf{v}_j - \mathbf{A} \mathcal{P}^{-1}_j \mathbf{v}_j.
\end{align*}
\]

Imposing \( \mathbf{r}^{(\sigma)}_j = \gamma \mathbf{r}_j \) yields:

\[
(\mathbf{A} - \sigma \mathbf{I}) \mathcal{P}^{(\sigma)}_j^{-1} \mathbf{v}_j = \gamma \mathbf{A} \mathcal{P}^{-1}_j \mathbf{v}_j - (\gamma - 1) \mathbf{v}_j
\]
Altogether,

\[(A - \sigma I)P(\sigma)_j^{-1}v_j = V_{m+1}h_j^{(\sigma)}\]

\[\gamma AP_j^{-1}v_j - (\gamma - 1)v_j = V_{m+1}h_j^{(\sigma)}\]

\[\gamma V_{m+1}h_j - V_{m+1}(\gamma - 1)e_j = V_{m+1}h_j^{(\sigma)}\]

\[V_{m+1}(\gamma h_j - (\gamma - 1)e_j) = V_{m+1}h_j^{(\sigma)}\]

which yields:

\[H_m^{(\sigma)} = (H_m - I_m)\Gamma_m + I_m,\]

with \(\Gamma_m := \text{diag}(\gamma_1, \ldots, \gamma_m)\).
Altogether,

\[(A - \sigma I)P(\sigma)^{-1}v_j = V_{m+1}h_j^{(\sigma)}\]

\[\gamma AP_j^{-1}v_j - (\gamma - 1)v_j = V_{m+1}h_j^{(\sigma)}\]

\[\gamma V_{m+1}h_j - V_{m+1}(\gamma - 1)e_j = V_{m+1}h_j^{(\sigma)}\]

\[V_{m+1}(\gamma h_j - (\gamma - 1)e_j) = V_{m+1}h_j^{(\sigma)}\]

which yields:

\[H^{(\sigma)}_m = (H_m - I_m)\Gamma_m + I_m,\]

with \(\Gamma_m := \text{diag}(\gamma_1, \ldots, \gamma_m)\).
Altogether,

\[(A - \sigma I)\mathcal{P}(\sigma)^{-1}v_j = V_{m+1} h^{(\sigma)}_j\]
\[\gamma A\mathcal{P}_j^{-1}v_j - (\gamma - 1)v_j = V_{m+1} h^{(\sigma)}_j\]
\[\gamma V_{m+1} h_j - V_{m+1} (\gamma - 1) e_j = V_{m+1} h^{(\sigma)}_j\]
\[V_{m+1} (\gamma h_j - (\gamma - 1) e_j) = V_{m+1} h^{(\sigma)}_j\]

which yields:

\[H^{(\sigma)}_m = (H_m - I_m) \Gamma_m + I_m,\]

with \(\Gamma_m := \text{diag}(\gamma_1, \ldots, \gamma_m)\).
Altogether,

\[(A - \sigma I)P(\sigma)^{-1}v_j = V_{m+1}h_j^{(\sigma)}\]
\[\gamma A P_j^{-1}v_j - (\gamma - 1)v_j = V_{m+1}h_j^{(\sigma)}\]
\[\gamma V_{m+1}h_j - V_{m+1}(\gamma - 1)e_j = V_{m+1}h_j^{(\sigma)}\]
\[V_{m+1}(\gamma h_j - (\gamma - 1)e_j) = V_{m+1}h_j^{(\sigma)}\]

which yields:

\[H^{(\sigma)}_m = (H_m - I_m)\Gamma_m + I_m,\]

with \(\Gamma_m := \text{diag}(\gamma_1, \ldots, \gamma_m)\).
Altogether,

$$(A - \sigma I)\mathcal{P}(\sigma)_{j}^{-1}v_j = V_{m+1}h_{j}^{(\sigma)}$$

$$\gamma A\mathcal{P}_{j}^{-1}v_j - (\gamma - 1)v_j = V_{m+1}h_{j}^{(\sigma)}$$

$$\gamma V_{m+1}h_{j} - V_{m+1} (\gamma - 1)e_j = V_{m+1}h_{j}^{(\sigma)}$$

$$V_{m+1} (\gamma h_{j} - (\gamma - 1)e_j) = V_{m+1}h_{j}^{(\sigma)}$$

which yields:

$$H_{m}^{(\sigma)} = (H_{m} - I_{m}) \Gamma_{m} + I_{m},$$

with $\Gamma_{m} := \text{diag}(\gamma_1, \ldots, \gamma_m)$.
In nested FOM-FGMRES, we solve the following (small) optimization problem,

\[ \mathbf{x}_m^{(\sigma)} = \arg\min_{\mathbf{y} \in \mathbb{C}^m} \| \mathbf{b} - (\mathbf{A} - \sigma I) \mathbf{y} \| \]

where the entries of \( \Gamma_m^{(\sigma)} \) are collinearity factors of inner FOM.
By the way...

Our key findings,

\[ \mathcal{H}_m^{(\sigma)} = (\mathcal{H}_m - \mathbb{I}_m) \Gamma_m^{(\sigma)} + \mathbb{I}_m, \quad \Gamma_m^{(\sigma)} := \text{diag}(\gamma_1^{(\sigma)}, \ldots, \gamma_m^{(\sigma)}), \]

are closely related to the formula,

\[ \mathcal{H}_m^{(\sigma)} = \mathbb{I}_m - \mathcal{H}_m (\sigma \mathbb{I}_m - \mathcal{T}_m), \quad \mathcal{T}_m := \text{diag}(\tau_1, \ldots, \tau_m), \]

which is taken from [Bakhos et al, 2013] and adapted in notation.
Our key findings,

\[ H_m^{(\sigma)} = (H_m - I_m) \Gamma_m^{(\sigma)} + I_m, \quad \Gamma_m^{(\sigma)} := \text{diag}(\gamma_1^{(\sigma)}, \ldots, \gamma_m^{(\sigma)}) \]

are closely related to the formula,

\[ H_m^{(\sigma)} = I_m - H_m(\sigma I_m - T_m), \quad T_m := \text{diag}(\tau_1, \ldots, \tau_m), \]

which is taken from [Bakhos et al, 2013] and adapted in notation.
Application & numerical tests
PDE-constrained optimization:

$$\min_{\rho(x), c_p(x), c_s(x)} \| u_{sim} - u_{meas} \|,$$

where in our application:

- $u_{sim}$ is the (numerical) solution of the elastic wave equation,
- $u_{meas}$ is obtained from measurements,
- $\rho, c_p, c_s$ are properties of earth layers we are interested in.

The modeling is done in frequency-domain.
An FEM discretization of the time-harmonic, inhomogeneous elastic wave equation at multiple frequencies \( \sigma_k \) is given by:

\[
(K + i\sigma_k C - \sigma_k^2 M)u_k = s, \quad k = 1, \ldots, N_\sigma,
\]

where

- \( M, K \) are mass and stiffness matrix, respectively,
- \( C \) encounters Sommerfeld boundary conditions,
- \( s \) usually models a point source,
- we need to compute the displacement vector \( u_k \) for multiple frequencies (shifts) \( \sigma_k \).
An FEM discretization of the time-harmonic, inhomogeneous elastic wave equation at multiple frequencies $\sigma_k$ is given by:

$$(K + i\sigma_k C - \sigma_k^2 M)u_k = s, \quad k = 1, \ldots, N_\sigma,$$

where

- $M, K$ are mass and stiffness matrix, respectively,
- $C$ encounters Sommerfeld boundary conditions,
- $s$ usually models a point source,
- we need to compute the displacement vector $u_k$ for multiple frequencies (shifts) $\sigma_k$. 
We can reformulate the previous problem to:

\[
\begin{pmatrix}
iC & K \\
I & 0
\end{pmatrix} - \sigma_k \begin{pmatrix} M & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} \sigma_k u_k \\ u_k \end{pmatrix} = \begin{pmatrix} s \\ 0 \end{pmatrix},
\]

which is of the form:

\[(A - \sigma_k M)x_k = b.\]

Apply shifted Laplace preconditioner

\[P = (A - \tau M), \quad \text{with } \Im(\tau) > 0.\]
We can reformulate the previous problem to:

\[
\begin{pmatrix}
iC & K \\
l & 0
\end{pmatrix} - \sigma_k \begin{pmatrix} M & 0 \\ 0 & l \end{pmatrix} \begin{pmatrix} \sigma_k u_k \\ u_k \end{pmatrix} = \begin{pmatrix} s \\ 0 \end{pmatrix},
\]

which is of the form:

\[(A - \sigma_k M)x_k = b.\]

Apply shifted Laplace preconditioner

\[\mathcal{P} = (A - \tau M), \quad \text{with } \Im(\tau) > 0.\]
Numerical experiments (1/4)  
as presented in [B./vG., 2014]

Test case from literature:

- $\Omega = [0, 1] \times [0, 1]$
- $h = 0.01$ implying $n = 10.201$ grid points
- system size: $4n = 40.804$
- $N_\sigma = 6$ frequencies
- point source at center

Reference

Preconditioned **multi-shift GMRES**:

- simultaneous solve
- linear convergence rates
- \( \tau = (0.7 - 0.7i)\sigma_{\text{max}} \)
- CPU time: 17.71s
Preconditioned nested FOM-FGMRES:

- 20 inner iterations
- truncate when inner residual norm $\sim 0.1$
- very few outer iterations
- CPU time: 9.12s
Various combinations of nested algorithms:

<table>
<thead>
<tr>
<th></th>
<th>multi-shift Krylov methods</th>
<th>nested multi-shift Krylov methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>msGMRES</td>
<td>FOM-FGMRES</td>
</tr>
<tr>
<td># inner iterations</td>
<td>rest msGMRES</td>
<td>IDR(4)-FGMRES</td>
</tr>
<tr>
<td># outer iterations</td>
<td>103</td>
<td>20</td>
</tr>
<tr>
<td>CPU time</td>
<td>17.71s</td>
<td>6.13s</td>
</tr>
<tr>
<td></td>
<td>QMRIDR(4)</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>msIDR(4)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.13s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions and future work

✓ Inner-outer Krylov methods for $Ax = b$ are widely used

⇝ We present an extension to shifted linear systems

✓ Multiple combinations of inner-outer methods possible, e.g. FOM-FGMRES, IDR-FQMR, IDR-FQMR, ...

? Future work: 3D problems

▶ discretization using TU/e package nutils (high-order FEM)
▶ approximate shifted Laplacian with MSSS preconditioner
▶ nested solver in a coupled Python - Fortran framework
Thank you for your attention!

Further reading:


*Research funded by Shell.*
Shifted systems appear a lot in practice...

...which is why we meet again for PART II at 3:00 pm.