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

June 19, 2015

Preconditioning Conference 2015, TU Eindhoven, NL
Motivation (1/3)

Full-waveform inversion

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 $\omega_k$ is given by:

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

where

- $K$, $C$, $M$ are sparse and symmetric,
- $s$ usually models a point source,
- we need to compute the displacement vector $u_k$ for multiple frequencies (shifts) $\omega_k$. 

Motivation (2/3)
The discrete forward model
An FEM discretization of the time-harmonic, inhomogeneous elastic wave equation at multiple frequencies $\omega_k$ is given by:

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

where

- $K, C, M$ are sparse and symmetric,
- $s$ usually models a point source,
- we need to compute the displacement vector $u_k$ for multiple frequencies (shifts) $\omega_k$.  

Motivation (2/3)
The discrete forward model
We can re-formulate the previous problem to:

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

which is of the form:

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

Shift-invariance of Krylov subspaces:

\[K_m(A, r_0) \equiv \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\} = K_m(A - \omega l, r_0)\]

is challenging to preserve when preconditioning!
We can re-formulate the previous problem to:

\[
\begin{pmatrix}
i C & K \\
I & 0
\end{pmatrix} - \omega_k \begin{pmatrix} M & 0 \\
0 & I
\end{pmatrix} \begin{pmatrix} \omega_k u_k \\
u_k
\end{pmatrix} = \begin{pmatrix} s \\
0
\end{pmatrix},
\]

which is of the form:

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

**Shift-invariance** of Krylov subspaces:

\[\mathcal{K}_m(A, r_0) \equiv \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\} = \mathcal{K}_m(A - \omega I, r_0)\]

is challenging to preserve when preconditioning!
Motivation (3/3)
Multi-shift Krylov methods

We can re-formulate the previous problem to:

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

which is of the form:

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

Shift-invariance of Krylov subspaces:

\[K_m(A, r_0) \equiv \text{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\} = K_m(A - \omega I, r_0)\]

is challenging to preserve when preconditioning!
Preconditioners for shifted linear systems

2004 (Complex) shifted Laplace preconditioner:

\[ P = (A - \tau M), \quad \tau \approx \{\omega_1, ..., \omega_N\} \]

2007 Many shifted Laplace preconditioners:

\[ P_j = (A - \tau_j M) \]

2013 Polynomial preconditioners [Plenary talk at PRECON13]

2014 Question: Can we use a Krylov method as preconditioner?
\[ \sim \] Nested Krylov methods
Preconditioners for shifted linear systems

2004 (Complex) shifted Laplace preconditioner:

\[ P = (A - \tau M), \quad \tau \approx \{\omega_1, \ldots, \omega_N\} \]

2007 Many shifted Laplace preconditioners:

\[ P_j = (A - \tau_j M) \]

2013 Polynomial preconditioners [Plenary talk at PRECON13]

2014 Question: Can we use a Krylov method as preconditioner?

\[ \rightsquigarrow \text{Nested Krylov methods} \]
Outline

1. The shifted Laplace preconditioner and its relation to Möbius transformations

2. Inner-outer Krylov methods for shifted linear systems

3. Numerical experiments

4. Conclusion
The shifted Laplacian

The \textit{generalized} shifted Laplace preconditioner,

\[ \mathcal{P} = (\mathcal{A} - \tau \mathcal{M}), \quad \tau \in \mathbb{C}, \]

has two benefits:

1. it transforms our problem to shifted linear systems and, hence, enables the benefits of \textit{shift-invariant Krylov spaces},

2. it maps the original \textit{spectrum} to circles of known center and radius.

Moreover, \((\ast)\) is \textit{easy to apply} because \(\tau \in \mathbb{C}\) leads to a damped problem \(\leadsto\) \textit{multigrid works!}
The shifted Laplacian

The generalized shifted Laplace preconditioner,

\[ \mathcal{P} = (A - \tau M), \quad \tau \in \mathbb{C}, \]  

has two benefits:

1. it transforms our problem to shifted linear systems and, hence, enables the benefits of shift-invariant Krylov spaces,
2. it maps the original spectrum to circles of known center and radius.

Moreover, (\*) is easy to apply because \( \tau \in \mathbb{C} \) leads to a damped problem \( \leadsto \) multigrid works!
The shifted Laplacian

For $\mathcal{P} = (A - \tau M)$, the following relation holds:

$$(A - \omega_k M)\mathcal{P}_k^{-1} = A\mathcal{P}^{-1} - \eta_k(\omega)\mathcal{I}, \quad (***)$$

with

- $\mathcal{P}_k^{-1} = \frac{\tau}{\tau - \omega_k} \mathcal{P}^{-1}$
- $\eta_k = \omega_k / (\omega_k - \tau)$
- $\tau$ is a free parameter (seed shift)

For the spectrum of the RHS in (**), we see:

$$\sigma(AM^{-1}) \ni \lambda \mapsto \frac{\lambda}{\lambda - \tau} - \eta_k$$
The shifted Laplacian

For $\mathcal{P} = (A - \tau M)$, the following relation holds:

$$(A - \omega_k M)\mathcal{P}_k^{-1} = A\mathcal{P}^{-1} - \eta_k(\omega)\mathcal{I}, \quad (**)$$

with

- $\mathcal{P}_k^{-1} = \frac{\tau}{\tau - \omega_k} \mathcal{P}^{-1}$
- $\eta_k = \omega_k / (\omega_k - \tau)$
- $\tau$ is a free parameter (seed shift)

For the spectrum of the RHS in $\text{(**)}$, we see:

$$\sigma(AM^{-1}) \ni \lambda \mapsto \frac{\lambda}{\lambda - \tau} - \eta_k$$
The shifted Laplacian - Spectral analysis

Compare:

\[ \sigma(A - \omega_k M) \quad \text{vs.} \quad \sigma((A - \omega_k M)P_k^{-1}) \]

Open question: What’s the optimal \( \tau \) for equidistantly spaced frequencies \( \omega_1, \ldots, \omega_N \) ???
1. The shifted Laplace preconditioner and its relation to Möbius transformations

2. Inner-outer Krylov methods for shifted linear systems

3. Numerical experiments

4. Conclusion
Nested (inner-outer) multi-shift Krylov methods

Solve the preconditioned shifted problem, $\mathcal{B} := \mathcal{A}\mathcal{P}^{-1}$,

$$(\mathcal{B} - \eta_k I) \mathbf{x}_k = \mathbf{b}, \quad k = 1, \ldots, N,$$

with a nested Krylov method based on

$$\mathcal{K}_k(\mathcal{B}, \mathbf{r}_0) = \mathcal{K}_k(\mathcal{B} - \eta I, \mathbf{r}_0) \quad \forall \eta.$$

Diagram:

- inner method (preconditioner)
- early truncation
- collinear residuals
- outer method
- outer_iter++
Nested (inner-outer) multi-shift Krylov methods

Solve the preconditioned shifted problem, \( B := AP^{-1} \),

\[(B - \eta_k I) x_k = b, \quad k = 1, \ldots, N,\]

with a nested Krylov method based on

\[K_k(B, r_0) = K_k(B - \eta I, r_0) \quad \forall \eta.\]

inner method (preconditioner) \quad \text{early truncation} \quad \text{collinear residuals} \quad \text{outer method}

\text{outer_iter++}
Solve the preconditioned shifted problem, $\mathcal{B} := A\mathcal{P}^{-1},$

$$(\mathcal{B} - \eta_k I)x_k = b, \quad k = 1, \ldots, N,$$

with a nested Krylov method based on

$$\mathcal{K}_k(\mathcal{B}, r_0) = \mathcal{K}_k(\mathcal{B} - \eta I, r_0) \quad \forall \eta.$$
Nested (inner-outer) multi-shift Krylov methods

Solve the preconditioned shifted problem, \( B := AP^{-1} \),

\[(B - \eta_k I)x_k = b, \quad k = 1, \ldots, N,\]

with a nested Krylov method based on

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

Solve the preconditioned shifted problem, $\mathcal{B} := \mathcal{A}\mathcal{P}^{-1}$,

$$(\mathcal{B} - \eta_k I) x_k = b, \quad k = 1, \ldots, N,$$

with a nested Krylov method based on

$$\mathcal{K}_k(\mathcal{B}, r_0) = \mathcal{K}_k(\mathcal{B} - \eta I, r_0) \quad \forall \eta.$$
Nested (inner-outer) multi-shift Krylov methods

Solve the preconditioned shifted problem, $\mathcal{B} := AP^{-1}$,

$$(\mathcal{B} - \eta_k I)x_k = b, \quad k = 1, \ldots, N,$$

with a nested Krylov method based on

$$\mathcal{K}_k(\mathcal{B}, r_0) = \mathcal{K}_k(\mathcal{B} - \eta I, r_0) \quad \forall \eta.$$
Nested (inner-outer) multi-shift Krylov methods

Solve the preconditioned shifted problem, \( B := AP^{-1} \),

\[
(B - \eta_k I) x_k = b, \quad k = 1, \ldots, N,
\]

with a nested Krylov method based on

\[
\mathcal{K}_k(B, r_0) = \mathcal{K}_k(B - \eta I, r_0) \quad \forall \eta.
\]
Inner method: multi-shift FOM

Classical result: In FOM, the residuals are:

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

Thus, for the shifted residuals it holds:

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

Hence, we obtain collinear residuals,

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

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

Reference

Inner method: multi-shift FOM

Classical result: In FOM, the residuals are:

\[ r_j = \mathbf{b} - \mathbf{B} \mathbf{x}_j = ... = -h_{j+1,j} \mathbf{e}_j^T \mathbf{y}_j \mathbf{v}_{j+1}. \]

Thus, for the shifted residuals it holds:

\[ r_j^{(\eta)} = \mathbf{b} - (\mathbf{B} - \eta \mathbf{I}) \mathbf{x}_j^{(\eta)} = ... = -h_{j+1,j}^{(\eta)} \mathbf{e}_j^T \mathbf{y}_j^{(\eta)} \mathbf{v}_{j+1}. \]

Hence, we obtain collinear residuals,

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

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

Reference

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

Use flexible GMRES in the outer loop,

\[(B - \eta I)\hat{V}_m = V_{m+1}H^{(\eta)}_m,\]

where one column yields:

\[(B - \eta I)P(\eta)_j^{-1}v_j = V_{m+1}h^{(\eta)}_j, \quad 1 \leq j \leq m.\]

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

Use flexible GMRES in the outer loop,

\[(\mathcal{B} - \eta I) \hat{V}_m = V_{m+1} H^{(\eta)}_m,\]

where one column yields:

\[(\mathcal{B} - \eta I) P^{(\eta)}_j \mathbf{v}_j = V_{m+1} h^{(\eta)}_j, \quad 1 \leq j \leq m.\]

Recap: The “inner loop” is the truncated solution of \((\mathcal{B} - \eta I)\) with right-hand side \(\mathbf{v}_j\) using msFOM.
The inner residuals are:

\[ r_j^{(\eta)} = v_j - (B - \eta I)P(\eta)_{j}^{-1}v_j, \]
\[ r_j = v_j - BP_j^{-1}v_j. \]

Imposing \( r_j^{(\eta)} = \gamma r_j \) yields:

\[ (B - \eta I)P(\eta)_{j}^{-1}v_j = \gamma BP_j^{-1}v_j - (\gamma - 1)v_j \]
The inner residuals are:

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

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

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

\[(B - \eta I)\mathcal{P}(\eta)^{-1}v_j = V_{m+1}h_j^{(\eta)}\]

\[\gamma B\mathcal{P}_j^{-1}v_j - (\gamma - 1)v_j = V_{m+1}h_j^{(\eta)}\]

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

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

which yields:

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

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

\[(B - \eta I)P(\eta)_{j}^{-1}v_{j} = V_{m+1}h_{j}^{(\eta)}\]

\[\gamma B P_{j}^{-1}v_{j} - (\gamma - 1)v_{j} = V_{m+1}h_{j}^{(\eta)}\]

\[\gamma V_{m+1}h_{j} - V_{m+1} (\gamma - 1) e_{j} = V_{m+1}h_{j}^{(\eta)}\]

\[V_{m+1} (\gamma h_{j} - (\gamma - 1) e_{j}) = V_{m+1}h_{j}^{(\eta)}\]

which yields:

\[H_{m}^{(\eta)} = (H_{m} - I_{m}) \Gamma_{m} + I_{m},\]

with \(\Gamma_{m} := \text{diag}(\gamma_1, ..., \gamma_m).\)
Altogether,

\[(B - \eta I)P(\eta)^{-1}v_j = V_{m+1}h_j^{(\eta)}\]
\[\gamma BP_j^{-1}v_j - (\gamma - 1)v_j = V_{m+1}h_j^{(\eta)}\]
\[\gamma V_{m+1}h_j = V_{m+1}(\gamma - 1)e_j = V_{m+1}h_j^{(\eta)}\]

which yields:

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

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

\[(B - \eta I)P(\eta)_{j-1}v_j = V_{m+1}h^{(\eta)}_j\]
\[\gamma B P_{j-1}v_j - (\gamma - 1)v_j = V_{m+1}h^{(\eta)}_j\]
\[\gamma V_{m+1}h_j - V_{m+1} (\gamma - 1) e_j = V_{m+1}h^{(\eta)}_j\]
\[V_{m+1} (\gamma h_j - (\gamma - 1) e_j) = V_{m+1}h^{(\eta)}_j\]

which yields:

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

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

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

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

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

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

which yields:

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

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

\[
\begin{align*}
x_m^{(\eta)} &= \arg\min_{x \in \hat{V}_m} \|b - (B - \eta I)x\| \\
&= \arg\min_{y \in \mathbb{C}^m} \|b - (B - \eta I)\hat{V}_m y\| \\
&= \arg\min_{y \in \mathbb{C}^m} \|b - V_{m+1}H_m^{(\eta)} y\| \\
&= \arg\min_{y \in \mathbb{C}^m} \left\| \beta e_1 - \left( (H_m - I_m) \Gamma_m^{(\eta)} + I_m \right) y \right\|,
\end{align*}
\]

where the entries of \(\Gamma_m^{(\eta)}\) are collinearity factors of inner FOM.
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: $4n = 40.804$
- $N = 6$ frequencies
- point source at center

Reference
Preconditioned **multi-shift GMRES**:

- simultaneous solve
- linear convergence rates
- \( \tau = (0.7 - 0.7i)\omega_{\text{max}} \)
- 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>rest msGMRES</td>
</tr>
<tr>
<td># inner iterations</td>
<td>103</td>
<td>20</td>
</tr>
<tr>
<td># outer iterations</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>CPU time</td>
<td>17.71s</td>
<td>6.13s</td>
</tr>
</tbody>
</table>
Conclusions and future work

- Inner-outer Krylov methods for $Ax = b$ are widely used
  $\implies$ We present an extension to shifted linear systems
  - The shifted Laplace preconditioner is applied as a first layer
  - Future work: 3D problems
    - discretization using TU/e package nutils (high-order FEM)
    - approximate shifted Laplacian with AGMG
    - nested solver in Fortran90
Thank you for your attention!

Further reading:


*Research funded by Shell.*