Shi(f)t happens

- 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

SIAM Student Krylov Day 2015

TUDelft

What's a shifted linear system?

Definition

Shifted linear systems are of the form

$$(A - \omega_k I)\mathbf{x}_k = \mathbf{b},$$

where $\{\omega_k\}_{k=1}^{N} \in \mathbb{C}$ are a sequence of *shifts*.

For the simultaneous solution, **Krylov methods** are well-suited because of the *shift-invariance* property:

$$\mathcal{K}_m(A, \mathbf{b}) \equiv \text{span}\{\mathbf{b}, A\mathbf{b}, ..., A^{m-1}\mathbf{b}\} = \mathcal{K}_m(A - \omega I, \mathbf{b}).$$

"Proof" (shift-invariance)

For $m = 2$: $\mathcal{K}_2(A, \mathbf{b}) = \text{span}\{\mathbf{b}, A\mathbf{b}\}$

$$\mathcal{K}_2(A - \omega I, \mathbf{b}) = \text{span}\{\mathbf{b}, A\mathbf{b} - \omega\mathbf{b}\} = \text{span}\{\mathbf{b}, A\mathbf{b}\}$$

What's a shifted linear system?

Definition

Shifted linear systems are of the form

$$(A - \omega_k I)\mathbf{x}_k = \mathbf{b},$$

where $\{\omega_k\}_{k=1}^N \in \mathbb{C}$ are a sequence of *shifts*.

For the simultaneous solution, **Krylov methods** are well-suited because of the *shift-invariance* property:

$$\mathcal{K}_m(A, \mathbf{b}) \equiv \text{span}\{\mathbf{b}, A\mathbf{b}, ..., A^{m-1}\mathbf{b}\} = \mathcal{K}_m(A - \omega I, \mathbf{b}).$$

"Proof" (shift-invariance)

For $m = 2$: $\mathcal{K}_2(A, \mathbf{b}) = \text{span}\{\mathbf{b}, A\mathbf{b}\}$

$$\mathcal{K}_2(A - \omega I, \mathbf{b}) = \text{span}\{\mathbf{b}, A\mathbf{b} - \omega\mathbf{b}\} = \text{span}\{\mathbf{b}, A\mathbf{b}\}$$

TUDelft

What's a shifted linear system?

Definition

Shifted linear systems are of the form

$$(A - \omega_k I)\mathbf{x}_k = \mathbf{b},$$

where $\{\omega_k\}_{k=1}^N \in \mathbb{C}$ are a sequence of *shifts*.

For the simultaneous solution, **Krylov methods** are well-suited because of the *shift-invariance* property:

$$\mathcal{K}_m(A, \mathbf{b}) \equiv \text{span}\{\mathbf{b}, A\mathbf{b}, \ldots, A^{m-1}\mathbf{b}\} = \mathcal{K}_m(A - \omega I, \mathbf{b}).$$

"Proof" (shift-invariance)

For $m = 2$: $\mathcal{K}_2(A, \mathbf{b}) = \text{span}\{\mathbf{b}, A\mathbf{b}\}$

$$\mathcal{K}_2(A - \omega I, \mathbf{b}) = \text{span}\{\mathbf{b}, A\mathbf{b} - \omega\mathbf{b}\} = \text{span}\{\mathbf{b}, A\mathbf{b}\}$$

TUDelft

Some (semi) open research questions...

$$\boxed{(A - \omega_k I)\mathbf{x}_k = \mathbf{b}}$$

1. Can we allow multiple right-hand sides?
2. Which preconditioners preserve shift-invariance?
3. Can we apply restarting and nested algorithms?
4. Can we benefit from (spectral) deflation?
5. Where do shifted systems occur in practice?

Outline

Multi-shift GMRES

After m steps of Arnoldi, we have,

$$AV_m = V_{m+1}\underline{\mathrm{H}}_m,$$

and the approximate solution yields:

$$\mathbf{x}_m \approx V_m\mathbf{y}_m, \quad \text{where } \mathbf{y}_m = \underset{\mathbf{y}\in\mathbb{C}^m}{\operatorname{argmin}} \left\| \underline{\mathrm{H}}_m\mathbf{y} - \|\mathbf{b}\|\mathbf{e}_1 \right\|.$$

For shifted systems, we get

$$(A - \omega I)V_m = V_{m+1}(\underline{\mathrm{H}}_m - \omega\underline{\mathrm{I}}_m),$$

and, therefore,

$$\mathbf{x}_m^{(\omega)} \approx V_m\mathbf{y}_m^{(\omega)}, \quad \text{where } \mathbf{y}_m^{(\omega)} = \underset{\mathbf{y}\in\mathbb{C}^m}{\operatorname{argmin}} \left\| \underline{\mathrm{H}}_m^{(\omega)}\mathbf{y} - \|\mathbf{b}\|\mathbf{e}_1 \right\|.$$

Multi-shift GMRES

After m steps of Arnoldi, we have,

$$AV_m = V_{m+1}\underline{\mathrm{H}}_m,$$

and the approximate solution yields:

$$\mathbf{x}_m \approx V_m\mathbf{y}_m, \quad \text{where } \mathbf{y}_m = \underset{\mathbf{y}\in\mathbb{C}^m}{\operatorname{argmin}} \left\|\underline{\mathrm{H}}_m\mathbf{y} - \|\mathbf{b}\|\mathbf{e}_1\right\|.$$

For shifted systems, we get

$$(A - \omega I)V_m = V_{m+1}(\underline{\mathrm{H}}_m - \omega\underline{\mathrm{I}}_m),$$

and, therefore,

$$\mathbf{x}_m^{(\omega)} \approx V_m\mathbf{y}_m^{(\omega)}, \quad \text{where } \mathbf{y}_m^{(\omega)} = \underset{\mathbf{y}\in\mathbb{C}^m}{\operatorname{argmin}} \left\|\underline{\mathrm{H}}_m^{(\omega)}\mathbf{y} - \|\mathbf{b}\|\mathbf{e}_1\right\|.$$

Multi-shift GMRES

After m steps of Arnoldi, we have,

$$AV_m = V_{m+1}\underline{\mathrm{H}}_m,$$

and the approximate solution yields:

$$\mathbf{x}_m \approx V_m \mathbf{y}_m, \quad \text{where } \mathbf{y}_m = \underset{\mathbf{y}\in\mathbb{C}^m}{\operatorname{argmin}} \left\| \underline{\mathrm{H}}_m \mathbf{y} - \|\mathbf{b}\| \mathbf{e}_1 \right\| .$$

For shifted systems, we get

$$(A - \omega I)V_m = V_{m+1}(\underline{\mathrm{H}}_m - \omega \underline{\mathrm{I}}_m),$$

and, therefore,

$$\mathbf{x}_m^{(\omega)} \approx V_m \mathbf{y}_m^{(\omega)}, \quad \text{where } \mathbf{y}_m^{(\omega)} = \underset{\mathbf{y}\in\mathbb{C}^m}{\operatorname{argmin}} \left\| \underline{\mathrm{H}}_m^{(\omega)} \mathbf{y} - \|\mathbf{b}\| \mathbf{e}_1 \right\| .$$

Preconditioning is a problem

Main disadvantage:

Preconditioners are in general not easy to apply. For

$$(A - \omega I)\mathcal{P}_\omega^{-1}\mathbf{y}^{(\omega)} = \mathbf{b}, \quad \mathcal{P}_\omega \mathbf{x}^{(\omega)} = \mathbf{y}^{(\omega)}$$

it does not hold:

$$\mathcal{K}_m(A\mathcal{P}^{-1}, \mathbf{b}) \neq \mathcal{K}_m(A\mathcal{P}_\omega^{-1} - \omega\mathcal{P}_\omega^{-1}, \mathbf{b}).$$

However, there are ways...

Reference

B. Jegerlehner, *Krylov space solvers for shifted linear systems.* Published online `arXiv:hep-lat/9612014`, 1996.

Preconditioning is a problem

Main disadvantage:

Preconditioners are in general not easy to apply. For

$$(A - \omega I)\mathcal{P}_\omega^{-1}\mathbf{y}^{(\omega)} = \mathbf{b}, \quad \mathcal{P}_\omega \mathbf{x}^{(\omega)} = \mathbf{y}^{(\omega)}$$

it does not hold:

$$\mathcal{K}_m(A\mathcal{P}^{-1}, \mathbf{b}) \neq \mathcal{K}_m(A\mathcal{P}_\omega^{-1} - \omega\mathcal{P}_\omega^{-1}, \mathbf{b}).$$

However, there are ways...

Reference

B. Jegerlehner, *Krylov space solvers for shifted linear systems.* Published online `arXiv:hep-lat/9612014`, 1996.

Preconditioning is a problem

... or has been a problem ?

Short historical overview:

2002 Shift-and-invert preconditioner:

$$\mathcal{P} = (A - \tau I), \quad \tau \approx \{\omega_1, ..., \omega_N\}$$

2007 Many shift-and-invert preconditioners:

$$\mathcal{P}_j = (A - \tau_j I)$$

2013 Polynomial preconditioners:

$$p_n(A) \approx A^{-1}, \quad p_n^{\omega}(A) \approx (A - \omega I)^{-1}$$

2014 Nested Krylov methods

Nested multi-shift Krylov methods

Methodology:

- Martin knows: Polynomial preconditioners exist
- Question: Can we use a Krylov polynomial?

Nested multi-shift Krylov methods:

- Use an inner multi-shift Krylov method as preconditioner.
- For inner method, require collinear residuals $[\mathbf{r}_j^{(\omega)} = \gamma \mathbf{r}_j]$. This is the case for:
  - multi-shift GMRES [1998]
  - multi-shift FOM [2003]
  - multi-shift BiCG [2003]
  - multi-shift IDR(s) [new!]
- Using γ, we can preserve the shift-invariance in the outer Krylov iteration.

Nested multi-shift Krylov methods

Methodology:

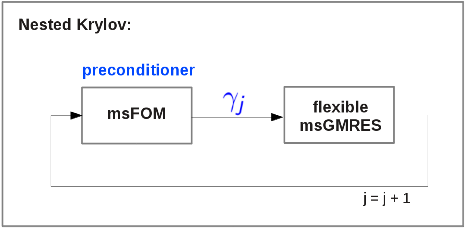
- Martin knows: Polynomial preconditioners exist
- Question: Can we use a Krylov polynomial?

Nested multi-shift Krylov methods:

- Use an inner multi-shift Krylov method as preconditioner.
- For inner method, require collinear residuals $[\mathbf{r}_j^{(\omega)} = \gamma \mathbf{r}_j]$.
  This is the case for:
  - multi-shift GMRES [1998]
  - multi-shift FOM [2003]
  - multi-shift BiCG [2003]
  - multi-shift IDR(s) [new!]
- Using γ, we can preserve the shift-invariance in the outer Krylov iteration.

Nested multi-shift Krylov methods

Overview of one possible combination:



Multi-shift FOM as inner method

Classical result: In FOM, the residuals are

$$\mathbf{r}_j = \mathbf{b} - A\mathbf{x}_j = \ldots = -h_{j+1,j}\mathbf{e}_j^T\mathbf{y}_j\mathbf{v}_{j+1}.$$

Thus, for the shifted residuals it holds:

$$\mathbf{r}_j^{(\omega)} = \mathbf{b} - (A - \omega I)\mathbf{x}_j^{(\omega)} = \ldots = -h_{j+1,j}^{(\omega)}\mathbf{e}_j^T\mathbf{y}_j^{(\omega)}\mathbf{v}_{j+1},$$

which gives $\gamma = y_j^{(\omega)}/y_j$.

Reference

V. Simoncini, *Restarted full orthogonalization method for shifted linear systems.* BIT Numerical Mathematics, 43 (2003).

TUDelft

Flexible multi-shift GMRES as outer method

Use flexible GMRES in the outer loop,

$$(A - \omega I)\widehat{V}_m = V_{m+1}\underline{\mathrm{H}}_m^{(\omega)},$$

where one column yields

$$(A - \omega I)\underbrace{\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j}_{\text{inner loop}} = V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}, \quad 1 \leq j \leq m.$$

The "inner loop" is the truncated solution of $(A - \omega I)$ with right-hand side $\mathbf{v}_j$ using `msFOM`.

Flexible multi-shift GMRES as outer method

Use flexible GMRES in the outer loop,

$$(A-\omega I)\widehat{V}_m = V_{m+1}\underline{\mathrm{H}}_m^{(\omega)},$$

where one column yields

$$(A-\omega I)\underbrace{\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j}_{\text{inner loop}} = V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}, \quad 1 \le j \le m.$$

The “inner loop” is the truncated solution of $(A-\omega I)$ with right-hand side $\mathbf{v}_j$ using `msFOM`.

Flexible multi-shift GMRES

The inner residuals are:

$$\mathbf{r}_j^{(\omega)} = \mathbf{v}_j - (A - \omega I)\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j,$$
$$\mathbf{r}_j = \mathbf{v}_j - A\mathcal{P}_j^{-1}\mathbf{v}_j,$$

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

$$(A - \omega I)\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j = \gamma A\mathcal{P}_j^{-1}\mathbf{v}_j - (\gamma - 1)\mathbf{v}_j \qquad (*)$$

Note that the right-hand side in $(*)$ is a preconditioned shifted system!

TUDelft

Flexible multi-shift GMRES

The inner residuals are:

$$\mathbf{r}_j^{(\omega)} = \mathbf{v}_j - (A - \omega I)\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j,$$
$$\mathbf{r}_j = \mathbf{v}_j - A\mathcal{P}_j^{-1}\mathbf{v}_j,$$

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

$$(A - \omega I)\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j = \gamma A\mathcal{P}_j^{-1}\mathbf{v}_j - (\gamma - 1)\mathbf{v}_j \qquad (*)$$

Note that the right-hand side in $(*)$ is a preconditioned shifted system!

Flexible multi-shift GMRES

Altogether,

$$
\begin{aligned}
(A-\omega I)\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
\gamma A\mathcal{P}_j^{-1}\mathbf{v}_j - (\gamma-1)\mathbf{v}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
\gamma V_{m+1}\underline{\mathbf{h}}_j - V_{m+1}\,(\gamma-1)\,\underline{\mathbf{e}}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
V_{m+1}\left(\gamma\underline{\mathbf{h}}_j - (\gamma-1)\,\underline{\mathbf{e}}_j\right) &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}
\end{aligned}
$$

which yields:

$$\underline{\mathrm{H}}_m^{(\omega)} = (\underline{\mathrm{H}}_m - \underline{\mathrm{I}}_m)\,\Gamma_m + \underline{\mathrm{I}}_m,$$

with $\Gamma_m := diag(\gamma_1, ..., \gamma_m)$.

Flexible multi-shift GMRES

Altogether,

$$(A-\omega I)\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j = V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}$$

$$\gamma A\mathcal{P}_j^{-1}\mathbf{v}_j - (\gamma-1)\mathbf{v}_j = V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}$$

$$\gamma V_{m+1}\underline{\mathbf{h}}_j - V_{m+1}(\gamma-1)\underline{\mathbf{e}}_j = V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}$$

$$V_{m+1}\left(\gamma\underline{\mathbf{h}}_j - (\gamma-1)\underline{\mathbf{e}}_j\right) = V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}$$

which yields:

$$\underline{\mathrm{H}}_m^{(\omega)} = (\underline{\mathrm{H}}_m - \underline{\mathrm{I}}_m)\,\Gamma_m + \underline{\mathrm{I}}_m,$$

with $\Gamma_m := diag(\gamma_1, ..., \gamma_m)$.

Flexible multi-shift GMRES

Altogether,

$$
\begin{aligned}
(A-\omega I)\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
\gamma A\mathcal{P}_j^{-1}\mathbf{v}_j - (\gamma-1)\mathbf{v}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
\gamma V_{m+1}\underline{\mathbf{h}}_j - V_{m+1}\left(\gamma-1\right)\underline{\mathbf{e}}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
V_{m+1}\left(\gamma\underline{\mathbf{h}}_j - \left(\gamma-1\right)\underline{\mathbf{e}}_j\right) &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}
\end{aligned}
$$

which yields:

$$
\underline{\mathrm{H}}_m^{(\omega)} = \left(\underline{\mathrm{H}}_m - \underline{\mathrm{I}}_m\right)\Gamma_m + \underline{\mathrm{I}}_m,
$$

with $\Gamma_m := diag(\gamma_1, ..., \gamma_m)$.

Flexible multi-shift GMRES

Altogether,

$$
\begin{aligned}
(A - \omega I)\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
\gamma A\mathcal{P}_j^{-1}\mathbf{v}_j - (\gamma - 1)\mathbf{v}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
\gamma V_{m+1}\underline{\mathbf{h}}_j - V_{m+1}(\gamma - 1)\,\underline{\mathbf{e}}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
V_{m+1}\left(\gamma\underline{\mathbf{h}}_j - (\gamma - 1)\,\underline{\mathbf{e}}_j\right) &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}
\end{aligned}
$$

which yields:

$$\underline{\mathrm{H}}_m^{(\omega)} = (\underline{\mathrm{H}}_m - \underline{\mathrm{I}}_m)\,\Gamma_m + \underline{\mathrm{I}}_m,$$

with $\Gamma_m := diag(\gamma_1, ..., \gamma_m)$.

Flexible multi-shift GMRES

Altogether,

$$
\begin{aligned}
(A - \omega I)\mathcal{P}(\omega)_j^{-1}\mathbf{v}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
\gamma A\mathcal{P}_j^{-1}\mathbf{v}_j - (\gamma - 1)\mathbf{v}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
\gamma V_{m+1}\underline{\mathbf{h}}_j - V_{m+1}\,(\gamma - 1)\,\underline{\mathbf{e}}_j &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)} \\
V_{m+1}\left(\gamma\underline{\mathbf{h}}_j - (\gamma - 1)\,\underline{\mathbf{e}}_j\right) &= V_{m+1}\underline{\mathbf{h}}_j^{(\omega)}
\end{aligned}
$$

which yields:

$$
\underline{\mathrm{H}}_m^{(\omega)} = \left(\underline{\mathrm{H}}_m - \underline{\mathrm{I}}_m\right)\Gamma_m + \underline{\mathrm{I}}_m,
$$

with $\Gamma_m := diag(\gamma_1, ..., \gamma_m)$.

Geophysical applications

The model

The time-harmonic elastic wave equation

For **many** (angular) frequencies ω_k, we solve

$$-\omega_k^2 \rho(\mathbf{x})\hat{\mathbf{u}} - \nabla \cdot \sigma(\hat{\mathbf{u}}, c_p, c_s) = \hat{\mathbf{s}}, \quad \mathbf{x} \in \Omega \subset \mathbb{R}^{2,3},$$

together with absorbing or reflecting boundary conditions.

Inverse (discrete) Fourier transform:

$$\mathbf{u}(\mathbf{x}, t) = \sum_k \hat{\mathbf{u}}(\mathbf{x}, \omega_k) e^{i\omega_k t}$$

Geophysical applications

Discretization

The **discretized** time-harmonic elastic wave equation is quadratic in ω_k:

$$(K + i\omega_k C - \omega_k^2 M)\underline{\hat{\mathbf{u}}} = \underline{\hat{\mathbf{s}}},$$

which can be re-arranged as,

$$\left[\begin{pmatrix} iM^{-1}C & M^{-1}K \\ I & 0 \end{pmatrix} - \omega_k \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}\right] \begin{pmatrix} \omega_k \underline{\hat{\mathbf{u}}} \\ \underline{\hat{\mathbf{u}}} \end{pmatrix} = \begin{pmatrix} M^{-1}\underline{\hat{\mathbf{s}}} \\ 0 \end{pmatrix}.$$

The latter is of the form:

$$(A - \omega_k I)\mathbf{x}_k = \mathbf{b}, \quad k = 1, ..., N.$$

The **discretized** time-harmonic elastic wave equation is quadratic in ω_k:

$$(K + i\omega_k C - \omega_k^2 M)\underline{\hat{\mathbf{u}}} = \underline{\hat{\mathbf{s}}},$$

which can be re-arranged as,

$$\left[\begin{pmatrix} iM^{-1}C & M^{-1}K \\ I & 0 \end{pmatrix} - \omega_k \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}\right] \begin{pmatrix} \omega_k \underline{\hat{\mathbf{u}}} \\ \underline{\hat{\mathbf{u}}} \end{pmatrix} = \begin{pmatrix} M^{-1}\underline{\hat{\mathbf{s}}} \\ 0 \end{pmatrix}.$$

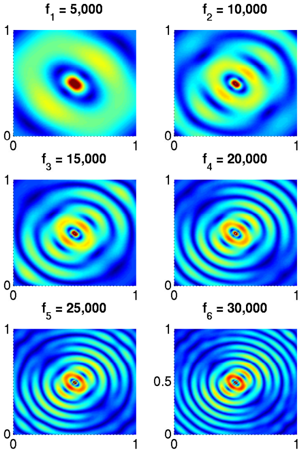
The latter is of the form:

$$(A - \omega_k I)\mathbf{x}_k = \mathbf{b}, \quad k = 1, ..., N.$$

A first example - The setting

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

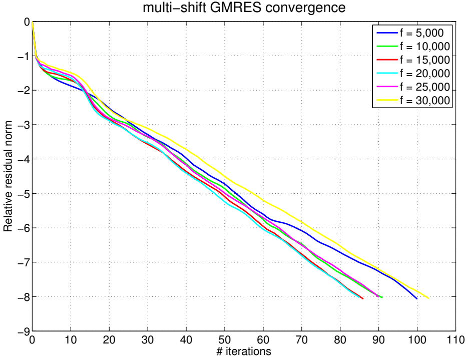


Reference

- T. Airaksinen, A. Pennanen, and J. Toivanen, *A damping preconditioner for time-harmonic wave equations in fluid and elastic material.* Journal of Computational Physics, 2009.

A first example - Convergence behavior (1/2)

Preconditioned **multi-shift GMRES**:

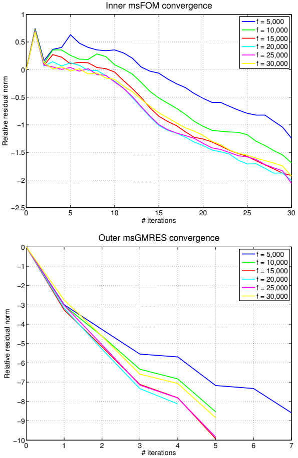


We observe:

- simultaneous solve
- CPU time: 17.71s

A first example - Convergence behavior (2/2)

Preconditioned **nested FOM-FGMRES**:



We observe:

- 30 inner iterations
- truncate when inner residual norm ~ 0.1
- very few outer iterations
- CPU time: 9.62s

Summary

✓ Nested Krylov methods for $A\mathbf{x} = \mathbf{b}$ are widely used
$\hookrightarrow$ extension to shifted linear systems is possible

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

✓ The shift-and-invert preconditioner (or the polynomial preconditioner) can be applied on top

✗ Future work: recylcing, deflation, ...

Thank you for your attention!

Further reading:

M. Baumann and M. B. van Gijzen. *Nested Krylov methods for shifted linear systems.* SISC Copper Mountain Special Section 2014 [Accepted].

Further coding:

https://bitbucket.org/ManuelMBaumann/nestedkrylov