Nested Krylov methods for shifted linear systems

4th IMA-NLAO Conference

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

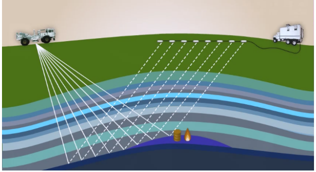
September 3, 2014

Motivation

Full-waveform inversion

PDE-constrained optimization:

$$\min_{\rho(\mathbf{x}),\dots} \left\|\mathbf{u}_{sim} - \mathbf{u}_{meas}\right\|,$$



where in our application:

- $\mathbf{u}_{sim}$ is the (numerical) solution of the elastic wave equation,
- $\mathbf{u}_{meas}$ is obtained from measurements,
- $\rho(\mathbf{x})$ is the density of the earth layers we are interested in.

The modelling is done in frequency domain...

Motivation

Modelling in frequency domain

Frequency domain approach:

The time-harmonic elastic wave equation

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

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

together with absorbing or reflecting boundary conditions.

Inverse (discrete) Fourier transform:

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

Motivation

Shifted linear systems

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

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

which can be re-arranged as,

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

The latter is of the form:

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

Motivation

Shifted linear systems

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

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

which can be re-arranged as,

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

The latter is of the form:

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

Outline

TUDelft

What's a shifted linear system?

Definition

Shifted linear systems are of the form

$$(A - \omega I)\mathbf{x}^{(\omega)} = \mathbf{b},$$

where $\omega \in \mathbb{C}$ is the *shift*.

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

$$(A - \omega I)\mathbf{x}^{(\omega)} = \mathbf{b},$$

where $\omega \in \mathbb{C}$ is the *shift*.

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

$$(A - \omega I)\mathbf{x}^{(\omega)} = \mathbf{b},$$

where $\omega \in \mathbb{C}$ is the *shift*.

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

For example, multi-shift GMRES

After m steps of Arnoldi, we have,

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

and the approximate solution yields:

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

For shifted systems, we get

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

and, therefore,

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

TUDelft

For example, 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,

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

TUDelft

For example, multi-shift GMRES

After m steps of Arnoldi, we have,

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

and the approximate solution yields:

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

For shifted systems, we get

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

and, therefore,

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

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

Y. A. Erlangga, *A robust and efficient iterative method for the numerical solution of the Helmholtz equation.* PhD thesis, TU Delft, 2005.

TU Delft

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

Y. A. Erlangga, *A robust and efficient iterative method for the numerical solution of the Helmholtz equation.* PhD thesis, TU Delft, 2005.

Preconditioning is a problem

... or has been a problem?

Short historical overview:

- (Single) shift-and-invert preconditioner:

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

- Many shift-and-invert preconditioners:

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

- Polynomial preconditioners:

$$p_n(A) \approx \mathcal{P}^{-1}, \quad p_n^{\omega}(A) \approx \mathcal{P}_{\omega}^{-1}$$

- Today's talk: nested Krylov methods

TUDelft

Nested Krylov methods for shifted linear systems

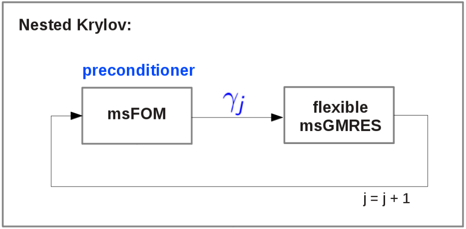
Our approach:

- Use a *flexible* preconditioner, $\mathcal{P}(\omega)_j^{-1}$
- this flexible preconditioner is a truncated multi-shift Krylov method itself ("inner" method)
- we require the inner method to produce **collinear residuals**, i.e. $\mathbf{r}_j^{(\omega)} = \gamma \mathbf{r}_j$. This is the case for:
  - multi-shift GMRES [1998]
  - multi-shift FOM [2003]
  - multi-shift BiCG [2003]
  - multi-shift IDR(s) [2014]
- using γ, we can **preserve the shift-invariance** in the "outer" Krylov method

TUDelft

Nested Krylov methods for shifted linear systems

Overview of one possible combination:



Flexible multi-shift GMRES

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 e.g. `msFOM`.

Flexible multi-shift GMRES

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 e.g. `msFOM`.

TU Delft

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

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

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:

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

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:

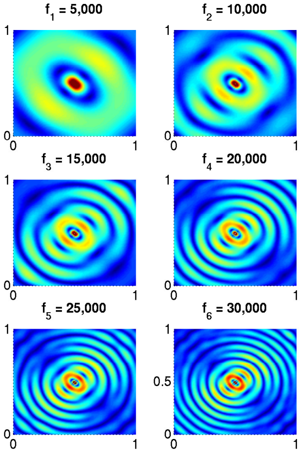
$$\underline{\mathrm{H}}_m^{(\omega)} = (\underline{\mathrm{H}}_m-\underline{\mathrm{I}}_m)\,\Gamma_m+\underline{\mathrm{I}}_m,$$

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

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



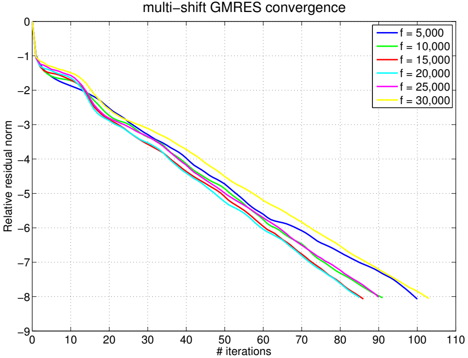
Reference

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

TUDelft

A first example - Convergence behavior 1/2

Preconditioned **multi-shift GMRES**:



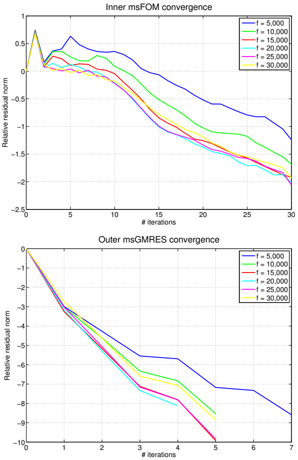
We observe:

- simultaneous solve
- CPU time: 17.71s

TUDelft

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

A first example - More nested methods

We were running the same setting with different (nested) multi-shift Krylov methods:

	multi-shift Krylov methods			
	msGMRES	rest_msGMRES	QMRIDR(4)	msIDR(4)
# inner iterations	-	20	-	-
# outer iterations	103	7	136	134
seed shift τ	0.7-0.7i	0.7-0.7i	0.7-0.7i	0.7-0.7i
CPU time	17.71s	6.13s	22.35s	22.58s
	nested multi-shift Krylov methods			
	FOM-FGMRES	IDR(4)-FGMRES	FOM-FQMRIDR(4)	IDR(4)-FQMRIDR(4)
# inner iterations	30	25	30	30
# outer iterations	7	9	5	15
seed shift τ	0.7-0.7i	0.7-0.7i	0.7-0.7i	0.7-0.7i
CPU time	9.62s	32.99s	8.14s	58.36s

TU Delft

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: recycling, deflation, ...

TUDelft

Thank you for your attention!

Further reading:

M. Baumann and M. B. van Gijzen. *Nested Krylov methods for shifted linear systems.* DIAM technical report 14-01, 2014.

Further coding:

https://bitbucket.org/ManuelMBaumann/nestedkrylov