A unifying framework for recycling-based iterative methods

Kirk M. Soodhalter

Trinity College Dublin
The University of Dublin
Ireland
https://math.soodhalter.com
ksoodha@maths.tcd.ie

15.–17. August 2022

Joint work with
Framework: Misha Kilmer (Tufts) and Eric de Sturler (Virginia Tech)
Matrix function evaluation: Liam Burke (TCD), Gustavo Ramirez-Hidalgo (Wuppertal), and Andreas Frommer (Wuppertal)

Outline of this Talk

1. Augmentation of iterative methods
2. A framework describing augmentation methods
3. Unprojected methods
4. Short recurrence schemes
5. Recycling for matrix function evaluation
6. Future work

Iterative methods for linear systems

Consider solving $\mathbf{A}(\mathbf{x}_0 + \boldsymbol{\eta}) = \mathbf{b}$;
We approximate $\boldsymbol{\eta} \approx \mathbf{t}_m \in \mathcal{V}_m$ by constraining the residual

$$\mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \mathbf{t}_m) \perp \widetilde{\mathcal{V}}_m,\ \dim \mathcal{V}_m = \dim \widetilde{\mathcal{V}}_m$$

Examples (with $\mathbf{r}_0 = \mathbf{A}\boldsymbol{\eta} = \mathbf{b} - \mathbf{A}\mathbf{x}_0$):

- GMRES: $\mathcal{V}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ and $\widetilde{\mathcal{V}}_m = \mathbf{A}\mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$
- CG: $\mathbf{A}$ SPD, $\mathcal{V}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ and $\widetilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$
- BiCG: $\mathcal{V}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ and $\widetilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}^T, \mathbf{r}_0)$

...and so on. This formulation works with other Krylov subspace methods, e.g., as well as **gradient descent** and many stationary iterative methods

Iterative methods for linear systems

Consider solving $\mathbf{A}(\mathbf{x}_0 + \boldsymbol{\eta}) = \mathbf{b}$;
We approximate $\boldsymbol{\eta} \approx \mathbf{t}_m \in \mathcal{V}_m$ by constraining the residual

$$\mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \mathbf{t}_m) \perp \widetilde{\mathcal{V}}_m, \ \dim \mathcal{V}_m = \dim \widetilde{\mathcal{V}}_m$$

Examples (with $\mathbf{r}_0 = \mathbf{A}\boldsymbol{\eta} = \mathbf{b} - \mathbf{A}\mathbf{x}_0$):

- GMRES: $\mathcal{V}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ and $\widetilde{\mathcal{V}}_m = \mathbf{A}\mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$
- CG: $\mathbf{A}$ SPD, $\mathcal{V}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ and $\widetilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$
- BiCG: $\mathcal{V}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ and $\widetilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}^T, \mathbf{r}_0)$

...and so on. This formulation works with other Krylov subspace methods, e.g., as well as **gradient descent** and many stationary iterative methods

Krylov Subspace Method - GMRES

Given $\mathbf{A}$ and $\mathbf{r}_0$, the mth Krylov subspace is defined

$$\mathcal{K}_m(\mathbf{A}, \mathbf{r}_0) = \operatorname{span}\left\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{m-1}\mathbf{r}_0\right\}.$$

Thus, $\mathbf{u} \in \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ is such that

$$\mathbf{u} = p(\mathbf{A})\mathbf{r}_0$$

where $p(x)$ is a polynomial of degree less than m.

GMRES

- For an initial approximation $\mathbf{x}_0$, let $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$
- Krylov subspace:
  $\mathcal{K}_m(\mathbf{A}, \mathbf{r}_0) = \text{span}\left\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{m-1}\ \mathbf{r}_0\right\}$.
- Choose $\mathbf{x}_m = \mathbf{x}_0 + \mathbf{t}_m$. Let $\mathbf{r}_m = \mathbf{b} - \mathbf{A}\mathbf{x}_m$.
- For GMRES, construct $\mathbf{x}_m = \mathbf{x}_0 + \mathbf{t}_m$ where $\mathbf{t}_m \in \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ such that $\mathbf{t}_m$ min imizes

$$\min_{\mathbf{t} \in \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \mathbf{t})\|$$

- This is equivalent to $\mathbf{r}_m \perp \mathbf{A}\mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$
- Sibling method: Full Orthogonalization Method (FOM) – $\mathbf{r}_m \perp \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0) \Leftrightarrow$ Conjugate Gradients if $\mathbf{A}$ is Hermitian positive-definite

GMRES

- For an initial approximation $\mathbf{x}_0$, let $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$
- Krylov subspace:
  $\mathcal{K}_m(\mathbf{A}, \mathbf{r}_0) = \text{span}\left\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{m-1}\ \mathbf{r}_0\right\}$.
- Choose $\mathbf{x}_m = \mathbf{x}_0 + \mathbf{t}_m$. Let $\mathbf{r}_m = \mathbf{b} - \mathbf{A}\mathbf{x}_m$.
- For GMRES, construct $\mathbf{x}_m = \mathbf{x}_0 + \mathbf{t}_m$ where $\mathbf{t}_m \in \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ such that $\mathbf{t}_m$ min imizes

$$\min_{\mathbf{t} \in \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \mathbf{t})\|$$

- This is equivalent to $\mathbf{r}_m \perp \mathbf{A}\mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$
- Sibling method: Full Orthogonalization Method (FOM) – $\mathbf{r}_m \perp \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ $\Leftrightarrow$ Conjugate Gradients if $\mathbf{A}$ is Hermitian positive-definite

GMRES

- For an initial approximation $\mathbf{x}_0$, let $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$
- Krylov subspace:
  $\mathcal{K}_m(\mathbf{A}, \mathbf{r}_0) = \text{span}\left\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{m-1}\ \mathbf{r}_0\right\}$.
- Choose $\mathbf{x}_m = \mathbf{x}_0 + \mathbf{t}_m$. Let $\mathbf{r}_m = \mathbf{b} - \mathbf{A}\mathbf{x}_m$.
- For GMRES, construct $\mathbf{x}_m = \mathbf{x}_0 + \mathbf{t}_m$ where $\mathbf{t}_m \in \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$ such that $\mathbf{t}_m$ min imizes

$$\min_{\mathbf{t} \in \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \mathbf{t})\|$$

- This is equivalent to $\mathbf{r}_m \perp \mathbf{A}\mathcal{K}_m(\mathbf{A}, \mathbf{r}_0)$
- Sibling method: Full Orthogonalization Method (FOM) – $\mathbf{r}_m \perp \mathcal{K}_m(\mathbf{A}, \mathbf{r}_0) \Leftrightarrow$ Conjugate Gradients if $\mathbf{A}$ is Hermitian positive-definite

Augmented iterative methods

Consider solving $\mathbf{A}(\mathbf{x}_0 + \boldsymbol{\eta}) = \mathbf{b}$;
We approximate $\boldsymbol{\eta} \approx \mathbf{s}_m + \mathbf{t}_m \in \mathcal{U} + \mathcal{V}_m$ according to a constraint.

- $\mathcal{U}$ is a fixed subspace used to augment
  - $\rightarrow$ many possible choices, may be updated periodically (e.g., at restart/between systems)
- $\mathcal{V}_m$ is an iteratively generated subspace associated to an underlying method
- Enables subspace *recycling* between restarts and multiple linear systems
  - $\rightarrow$ also can append, e.g., approximate solutions, approximate eigenvectors, real-time streaming data

Conceptual Outline: GCRO-type/recycled GMRES cycle

- $\mathcal{C} = \mathbf{A}\mathcal{U}$
- Set $\mathbf{Q}$ to be the orthogonal projector onto $\mathcal{C}$
- Set $\mathbf{P}$ to be the $\mathbf{A}^T\mathbf{A}$-orthogonal projector onto $\mathcal{U}$
- Apply GMRES to $(\mathbf{I}-\mathbf{Q})\mathbf{A}\mathbf{t} = (\mathbf{I}-\mathbf{Q})\mathbf{r}_0$

  $\rightarrow$ at step m: $\mathbf{t} \approx \mathbf{t}_m \in \mathcal{K}_m(\mathbf{I}-\mathbf{Q})\mathbf{A}, (\mathbf{I}-\mathbf{Q})\mathbf{r}_0)$
- $\mathbf{x}_m = \mathbf{x}_0 + \mathbf{P}\boldsymbol{\eta} + (\mathbf{I}-\mathbf{P})\mathbf{t}_m$

  $\rightarrow$ $\mathbf{P}(\boldsymbol{\eta} - \mathbf{t}_m) \in \mathcal{U}$ and $\mathbf{t}_m \in \mathcal{V}_m = \mathcal{K}_m(\mathbf{I}-\mathbf{Q})\mathbf{A}, (\mathbf{I}-\mathbf{Q})\mathbf{r}_0)$
- $\mathcal{U} \leftarrow \mathcal{U}_{new}$ where $\mathcal{U}_{new} \subset \mathcal{U} + \mathcal{V}_m$

See, e.g.,: de Sturler '96, '99, Parks et al '05

GCRO-DR all-at-once approach

- Build $\mathbf{V}_m$ via Arnoldi for $\mathcal{K}_m(\mathbf{I}-\mathbf{Q})\mathbf{A}, (\mathbf{I}-\mathbf{Q})\mathbf{r}_0)$
- Modified Arnoldi

$$\mathbf{A}\begin{bmatrix}\mathbf{U} & \mathbf{V}_m\end{bmatrix} = \begin{bmatrix}\mathbf{C} & \mathbf{V}_{m+1}\end{bmatrix}\underline{\mathbf{G}}_m$$

$$\text{where}\quad \underline{\mathbf{G}}_m = \begin{bmatrix}\mathbf{I}_k & \mathbf{B}_m \\ & \underline{\mathbf{H}}_m\end{bmatrix} \quad \text{and} \quad \mathbf{B}_m = \mathbf{C}^T\mathbf{A}\mathbf{V}_m.$$

- Full Minimization (full residual constraint $\perp \mathbf{A}(\mathcal{U}+\mathcal{V}_m)$)

$$(\mathbf{z}_m, \mathbf{y}_m) = \underset{\substack{\mathbf{u}\in\mathbb{R}^k \\ \mathbf{v}\in\mathbb{R}^j}}{\operatorname{argmin}} \left\| \begin{bmatrix}\mathbf{C} & \mathbf{V}_{m+1}\end{bmatrix}^T \mathbf{r}_0 - \underline{\mathbf{G}}_m \begin{bmatrix}\mathbf{z} \\ \mathbf{y}\end{bmatrix} \right\|_2 .$$

- $\mathbf{x}_m = \mathbf{x}_0 + \mathbf{P}\boldsymbol{\eta}_0 + \mathbf{V}_m\mathbf{y}_m + \mathbf{U}\mathbf{z}_m$

GCRO-DR projected GMRES approach

- Build $\mathbf{V}_m$ via Arnoldi for $\mathcal{K}_m(\mathbf{I}-\mathbf{Q})\mathbf{A},(\mathbf{I}-\mathbf{Q})\mathbf{r}_0)$
- $\mathbf{t}_m = \mathbf{V}_m\mathbf{y}_m$ is the mth GMRES approximation for $(\mathbf{I}-\mathbf{Q})\mathbf{A}\mathbf{t} = (\mathbf{I}-\mathbf{Q})\mathbf{r}_0$
- $\mathbf{x}_m = \mathbf{x}_0 + \mathbf{P}\boldsymbol{\eta}_0 + (\mathbf{I}-\mathbf{P})\mathbf{t}_m$

Deriving augmented iterative methods

One approach has its origins in domain decomposition.[1]

- Choose projectors $\mathbf{P}$ and $\mathbf{Q}$ onto $\mathcal{U}$ and $\mathcal{C} = \mathbf{A}\mathcal{U}$, respectively (orthogonal or oblique)
  - $\rightarrow$ Required: $\mathbf{AP} = \mathbf{QA}$
- $\mathbf{x} = \mathbf{x}_0 + \boldsymbol{\eta} = \mathbf{x}_0 + \mathbf{P}\boldsymbol{\eta} + (\mathbf{I} - \mathbf{P})\boldsymbol{\eta}$
  - $\rightarrow$ $\mathbf{P}\boldsymbol{\eta}$ can be directly computed
  - $\rightarrow$ $(\mathbf{I} - \mathbf{P})\boldsymbol{\eta}$ is approximated by an iterative method

Apply an iterative method to $(\mathbf{I} - \mathbf{Q})\mathbf{A}\mathbf{t} = (\mathbf{I} - \mathbf{Q})\mathbf{r}_0$ to obtain $\mathbf{t}_m$ and approximate

$$(\mathbf{I} - \mathbf{P})\boldsymbol{\eta} \approx (\mathbf{I} - \mathbf{P})\mathbf{t}_m$$

[1] see, e.g., Mandel 1993, Erlangga and Nabben 2008, Dolean et al (SIAM Book) 2015

Residual constraint formulation

Consider solving $\mathbf{A}(\mathbf{x}_0 + \boldsymbol{\eta}) = \mathbf{b}$;
We can also approximate $\boldsymbol{\eta} \approx \mathbf{s}_m + \mathbf{t}_m \in \mathcal{U} + \mathcal{V}_m$ by constraining the residual over a sum of spaces

$$\mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m$$

- GCRO-DR: $\widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m = \mathbf{A}(\mathcal{U} + \mathcal{V}_m)$
- DCG[2] $\widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m = \mathcal{U} + \mathcal{V}_m$
- non-optimal methods have a variety constraint strategies

How to reconcile different derivations and augmentation strategies?

[2]Saad et al '00

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\mathbf{y}_m$ *approx. solves* $\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{A}\mathbf{V}_m\mathbf{v} = \left(\mathbf{I} - \mathbf{Q}\right)\mathbf{r}_0$
- *initial error projection*
- *projection of* $\mathbf{V}_m\mathbf{y}_m$
- *full residual* $\mathbf{r}_m$ *is projected subproblem residual*

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widehat{\mathbf{r}}_m = (\mathbf{I} - \mathbf{Q})\left[\mathbf{r}_0 - \mathbf{A}\mathbf{V}_m\mathbf{y}_m\right] \perp \widetilde{\mathcal{V}}_m$
- *initial error projection*
- *projection of* $\mathbf{V}_m\mathbf{y}_m$
- *full residual* $\mathbf{r}_m$ *is projected subproblem residual*

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widetilde{\mathbf{V}}_m^* \left(\mathbf{I} - \mathbf{Q}\right) \mathbf{A}\mathbf{V}_m\mathbf{y}_m = \widetilde{\mathbf{V}}_m^* \left(\mathbf{I} - \mathbf{Q}\right) \mathbf{r}_0$
- *initial error projection*
- *projection of* $\mathbf{V}_m\mathbf{y}_m$
- *full residual* $\mathbf{r}_m$ *is projected subproblem residual*

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widetilde{\mathbf{V}}_m^* \left(\mathbf{I} - \mathbf{Q}\right) \mathbf{A}\mathbf{V}_m\mathbf{y}_m = \widetilde{\mathbf{V}}_m^* \left(\mathbf{I} - \mathbf{Q}\right) \mathbf{r}_0$
- *initial error projection*
- *projection of* $\mathbf{V}_m\mathbf{y}_m$
- *full residual* $\mathbf{r}_m$ *is projected subproblem residual*

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{A}\mathbf{V}_m\mathbf{y}_m = \widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{r}_0$
- $\mathbf{U}\mathbf{z}^{(1)} = \mathbf{P}\left(\mathbf{x} - \mathbf{x}_0\right)$
- *projection of* $\mathbf{V}_m\mathbf{y}_m$
- *full residual* $\mathbf{r}_m$ *is projected subproblem residual*

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{A}\mathbf{V}_m\mathbf{y}_m = \widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{r}_0$
- $\mathbf{U}\mathbf{z}^{(1)} = \mathbf{P}\left(\mathbf{x} - \mathbf{x}_0\right)$
- *projection of* $\mathbf{V}_m\mathbf{y}_m$
- *full residual* $\mathbf{r}_m$ *is projected subproblem residual*

---

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{A}\mathbf{V}_m\mathbf{y}_m = \widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{r}_0$
- $\mathbf{U}\mathbf{z}^{(1)} = \mathbf{P}\left(\mathbf{x} - \mathbf{x}_0\right)$
- $\mathbf{U}\mathbf{z}_m^{(2)} = -\mathbf{P}\mathbf{V}_m\mathbf{y}_m$
- *full residual* $\mathbf{r}_m$ *is projected subproblem residual*

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{A}\mathbf{V}_m\mathbf{y}_m = \widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{r}_0$
- $\mathbf{U}\mathbf{z}^{(1)} = \mathbf{P}\left(\mathbf{x} - \mathbf{x}_0\right)$
- $\mathbf{U}\mathbf{z}_m^{(2)} = -\mathbf{P}\mathbf{V}_m\mathbf{y}_m$
- *full residual* $\mathbf{r}_m$ *is projected subproblem residual*

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widetilde{\mathbf{V}}_m^* \left(\mathbf{I} - \mathbf{Q}\right) \mathbf{A}\mathbf{V}_m\mathbf{y}_m = \widetilde{\mathbf{V}}_m^* \left(\mathbf{I} - \mathbf{Q}\right) \mathbf{r}_0$
- $\mathbf{U}\mathbf{z}^{(1)} = \mathbf{P}\left(\mathbf{x} - \mathbf{x}_0\right)$
- $\mathbf{U}\mathbf{z}_m^{(2)} = -\mathbf{P}\mathbf{V}_m\mathbf{y}_m$
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m = \left(\mathbf{I} - \mathbf{Q}\right)\left[\mathbf{r}_0 - \mathbf{A}\mathbf{V}_m\mathbf{y}_m\right]$

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widetilde{\mathbf{V}}_m^* \left(\mathbf{I} - \mathbf{Q}\right) \mathbf{A}\mathbf{V}_m\mathbf{y}_m = \widetilde{\mathbf{V}}_m^* \left(\mathbf{I} - \mathbf{Q}\right) \mathbf{r}_0$
- $\mathbf{U}\mathbf{z}^{(1)} = \mathbf{P}\left(\mathbf{x} - \mathbf{x}_0\right)$
- $\mathbf{U}\mathbf{z}_m^{(2)} = -\mathbf{P}\mathbf{V}_m\mathbf{y}_m$
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m = \left(\mathbf{I} - \mathbf{Q}\right)\left[\mathbf{r}_0 - \mathbf{A}\mathbf{V}_m\mathbf{y}_m\right]$

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

How to design an augmented iterative method

Theorem (Kilmer, de Sturler, S. '20)

[a] *The correction* $\underbrace{\mathbf{U}\mathbf{z}^{(1)} + \mathbf{U}\mathbf{z}_m^{(2)}}_{\mathbf{s}_m} + \underbrace{\mathbf{V}_m\mathbf{y}_m}_{\mathbf{t}_m}$ *satisfies*

$\mathbf{b} - \mathbf{A}\left(\mathbf{x}_0 + \mathbf{s}_m + \mathbf{t}_m\right) \perp \widetilde{\mathcal{U}} + \widetilde{\mathcal{V}}_m \iff$

- $\widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{A}\mathbf{V}_m\mathbf{y}_m = \widetilde{\mathbf{V}}_m^*\left(\mathbf{I} - \mathbf{Q}\right)\mathbf{r}_0$
- $\mathbf{U}\mathbf{z}^{(1)} = \mathbf{P}\left(\mathbf{x} - \mathbf{x}_0\right)$
- $\mathbf{U}\mathbf{z}_m^{(2)} = -\mathbf{P}\mathbf{V}_m\mathbf{y}_m$
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m = \left(\mathbf{I} - \mathbf{Q}\right)\left[\mathbf{r}_0 - \mathbf{A}\mathbf{V}_m\mathbf{y}_m\right]$

[a] special cases proven by, e.g., de Sturler '96, Gaul et al '13, Gaul '14, Gutknecht '15, Kahl and Rittich '17

- $\mathbf{P}$, $\mathbf{Q}$ projectors as before with nullspaces determined by $\widetilde{\mathcal{U}}$.
- $\mathbf{r}_m = \widehat{\mathbf{r}}_m \implies$ projected problem determines convergence

Previous work...list not exhaustive

- Existing methods:
  → **Well-posed problems:** [GCRO-DR, Parks et al], [Deflated CG, Saad et al '00; Carlberg '16], [Recycled BiCG, Ahuja '09; Ahuja et al, '12], [Recycled MINRES, Wang et al, '07; Schlömer and Gaul, '14], [GMRES for Shifted Systems, S. '12; S. et al '14; S. '16], [FGMRES-based augmentation, Saad '97]
  → **Ill-posed problems:** [Augmented GMRES, Baglama and Reichel '07], [Augmented CG, Calvetti et al '03], [Renaut et al '12],[Augmented rrGMRES, Dong et al '14],[Augmented LSQR, Jiang et al 2021]
- Analysis/Framework: [Non-optimal augmentation, Saad '97], [Recycling Methods, Gaul Ph.D. Thesis '14],

  [Deflation/Augmentation Framework, Gutknecht '12; Gaul et al '13; Gutknecht '14; **de Sturler, Kilmer, S. '20 and in-progress review/algorithm design paper**]

Separates spaces from projected subproblem

Residual projection over sums of subspaces induces a projected subproblem *independent* from $\mathcal{V}_m$ (and $\widetilde{\mathcal{V}}_m$)

- GCRO-DR and DCG have agreement of projectors (such as in [Gutknecht '14])
- Framework admits *larger* subclass of augmented methods into "recycling" paradigm
- Design of new methods with differing operators in projected subproblem and solution subspace possible
  - $\rightarrow$ need operator compatibility for efficient implementation

rrrGMRES [Dong et al ‘14], [S. ‘22]

Method proposed is a minimum residual method, meaning $\mathbf{r}_m \perp \mathbf{A}(\mathcal{U} + \mathcal{V}_m)$. However, $\mathcal{V}_m = \mathcal{K}_m\left(\mathbf{A}, \mathbf{A}\mathbf{b}^\delta\right)$ is **range restricted** and **unprojected**.

- New modified Arnoldi method
  $$\mathbf{A}\begin{bmatrix}\mathbf{V}_m & \mathbf{U}\end{bmatrix} = \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}\begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}$$
- Solve $\min \left\| \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}^T \mathbf{b}^\delta - \begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}\begin{bmatrix}\mathbf{y} \\ \mathbf{z}\end{bmatrix} \right\|_2$
- Equivalently, $\left(\overline{\mathbf{H}}_m^T \overline{\mathbf{H}} - \overline{\mathbf{H}}_m^T \widehat{\mathbf{B}}_m \widehat{\mathbf{B}}_m^T \overline{\mathbf{H}}_m\right)\mathbf{y}_m = \mathbf{rhs}$
- $\mathbf{t}_m = \mathbf{V}_m \mathbf{y}_m$ and $\mathbf{s}_m = \mathbf{P}\mathbf{e}_0 - \mathbf{P}\mathbf{t}_m$

Unprojected methods

rrrGMRES [Dong et al ‘14], [S. ‘22]

Method proposed is a minimum residual method, meaning $\mathbf{r}_m \perp \mathbf{A}\left(\mathcal{U} + \mathcal{V}_m\right)$. However, $\mathcal{V}_m = \mathcal{K}_m\left(\mathbf{A}, \mathbf{A}\mathbf{b}^\delta\right)$ is **range restricted** and **unprojected**.

- New modified Arnoldi method
  $\mathbf{A}\begin{bmatrix}\mathbf{V}_m & \mathbf{U}\end{bmatrix} = \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}\begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}$
- Solve $\min \left\| \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}^T \mathbf{b}^\delta - \begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}\begin{bmatrix}\mathbf{y} \\ \mathbf{z}\end{bmatrix}\right\|_2$
- Equivalently, $\left(\overline{\mathbf{H}}_m^T\overline{\mathbf{H}} - \overline{\mathbf{H}}_m^T\widehat{\mathbf{B}}_m\widehat{\mathbf{B}}_m^T\overline{\mathbf{H}}_m\right)\mathbf{y}_m = \mathbf{rhs}$
- $\mathbf{t}_m = \mathbf{V}_m\mathbf{y}_m$ and $\mathbf{s}_m = \mathbf{P}\mathbf{e}_0 - \mathbf{P}\mathbf{t}_m$

rrrGMRES [Dong et al '14], [S. '22]

Method proposed is a minimum residual method, meaning $\mathbf{r}_m \perp \mathbf{A}\left(\mathcal{U} + \mathcal{V}_m\right)$. However, $\mathcal{V}_m = \mathcal{K}_m\left(\mathbf{A}, \mathbf{A}\mathbf{b}^\delta\right)$ is **range restricted** and **unprojected**.

- New modified Arnoldi method
  $\mathbf{A}\begin{bmatrix}\mathbf{V}_m & \mathbf{U}\end{bmatrix} = \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}\begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}$
- Solve $\min \left\| \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}^T \mathbf{b}^\delta - \begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}\begin{bmatrix}\mathbf{y} \\ \mathbf{z}\end{bmatrix} \right\|_2$
- Equivalently, $\left(\overline{\mathbf{H}}_m^T \overline{\mathbf{H}} - \overline{\mathbf{H}}_m^T \widehat{\mathbf{B}}_m \widehat{\mathbf{B}}_m^T \overline{\mathbf{H}}_m\right)\mathbf{y}_m = \mathbf{rhs}$
- $\mathbf{t}_m = \mathbf{V}_m \mathbf{y}_m$ and $\mathbf{s}_m = \mathbf{P}\mathbf{e}_0 - \mathbf{P}\mathbf{t}_m$

Unprojected methods

rrrGMRES [Dong et al '14], [S. '22]

Method proposed is a minimum residual method, meaning $\mathbf{r}_m \perp \mathbf{A}\left(\mathcal{U} + \mathcal{V}_m\right)$. However, $\mathcal{V}_m = \mathcal{K}_m\left(\mathbf{A}, \mathbf{A}\mathbf{b}^\delta\right)$ is **range restricted** and **unprojected**.

- New modified Arnoldi method
  $\mathbf{A}\begin{bmatrix}\mathbf{V}_m & \mathbf{U}\end{bmatrix} = \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}\begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}$
- Solve $\min \left\| \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}^T \mathbf{b}^\delta - \begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}\begin{bmatrix}\mathbf{y} \\ \mathbf{z}\end{bmatrix} \right\|_2$
- Equivalently, $\left(\overline{\mathbf{H}}_m^T\overline{\mathbf{H}} - \overline{\mathbf{H}}_m^T\widehat{\mathbf{B}}_m\widehat{\mathbf{B}}_m^T\overline{\mathbf{H}}_m\right)\mathbf{y}_m = \mathbf{rhs}$
- $\mathbf{t}_m = \mathbf{V}_m\mathbf{y}_m$ and $\mathbf{s}_m = \mathbf{P}\mathbf{e}_0 - \mathbf{P}\mathbf{t}_m$

Unprojected methods

rrrGMRES [Dong et al '14], [S. '22]

Method proposed is a minimum residual method, meaning $\mathbf{r}_m \perp \mathbf{A}\left(\mathcal{U} + \mathcal{V}_m\right)$. However, $\mathcal{V}_m = \mathcal{K}_m\left(\mathbf{A}, \mathbf{A}\mathbf{b}^{\delta}\right)$ is **range restricted** and **unprojected**.

- New modified Arnoldi method
  $\mathbf{A}\begin{bmatrix}\mathbf{V}_m & \mathbf{U}\end{bmatrix} = \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}\begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}$
- Solve $\min \left\| \begin{bmatrix}\mathbf{V}_{j+1} & \mathbf{C}_m\end{bmatrix}^T \mathbf{b}^{\delta} - \begin{bmatrix}\overline{\mathbf{H}}_m & \widehat{\mathbf{B}}_m \\ & \mathbf{F}_m\end{bmatrix}\begin{bmatrix}\mathbf{y} \\ \mathbf{z}\end{bmatrix} \right\|_2$
- Equivalently, $\left(\overline{\mathbf{H}}_m^T\overline{\mathbf{H}} - \overline{\mathbf{H}}_m^T\widehat{\mathbf{B}}_m\widehat{\mathbf{B}}_m^T\overline{\mathbf{H}}_m\right)\mathbf{y}_m = \mathbf{rhs}$
- $\mathbf{t}_m = \mathbf{V}_m\mathbf{y}_m$ and $\mathbf{s}_m = \mathbf{P}\mathbf{e}_0 - \mathbf{P}\mathbf{t}_m$

Short recurrence methods

The framework accommodates short recurrence methods

Challenges

- Short-recurrence compatibility must be ensured
  - $\rightarrow$ DCG: projected operator is Hermitian
  - $\rightarrow$ RMINRES: projected operator is Hermitian *on the Krylov subspace*
  - $\rightarrow$ bi-orthog Lanczos: for appropriate projector pairs, bi-orthogonality of bases still holds *on the Krylov subspaces*
  - $\rightarrow$ Golub-Kahan bi-diagonalization: similar
- Must systematically update recycled subspace without storing all vectors
- Stability

Windowed Lanczos for efficient recycled subspace updates

$$
\begin{bmatrix}
\alpha_1 & \delta_2 \\
\beta_2 & \alpha_2 & \delta_3 \\
& \beta_3 & \alpha_3 & \delta_4 \\
& & \beta_4 & \ddots & \ddots \\
& & & \ddots & \alpha_{(\ell-1)p-2} & \delta_{(\ell-1)p-1} \\
& & & & \beta_{(\ell-1)p-1} & \alpha_{(\ell-1)p-1} & \delta_{(\ell-1)p} \\
& & & & & \beta_{(\ell-1)p} & \alpha_{(\ell-1)p} & \delta_{(\ell-1)p+1} \\
& & & & & & \beta_{(\ell-1)p+1} & \alpha_{(\ell-1)p+1} & \ddots \\
& & & & & & & \beta_{(\ell-1)p+2} & \ddots & \delta_{\ell p} \\
& & & & & & & & \ddots & \alpha_{\ell p} \\
& & & & & & & & & \beta_{\ell p+1}
\end{bmatrix}
$$

$\underline{\mathbf{T}}_{[(\ell-1)p+1]}$ $\qquad$ $\underline{\mathbf{T}}$

$$\mathbf{V}_{\ell p} = \left[\mathbf{V}_{(\ell-1)p} \quad \underline{\mathbf{V}}_{\ell}\right]; \qquad \mathbf{A}\underline{\mathbf{V}}_{\ell} = \left[\mathbf{v}_{(\ell-1)p} \quad \underline{\mathbf{V}}_{\ell} \quad \mathbf{v}_{\ell p+1}\right]\underline{\mathbf{T}}$$
$$= \delta_{(\ell-1)p}\mathbf{v}_{(\ell-1)p} + \underline{\mathbf{V}}_{\ell}\mathbf{T} + \beta_{\ell p+1}\mathbf{v}_{\ell p+1}$$

(and similar for biorthogonal bases)[3]

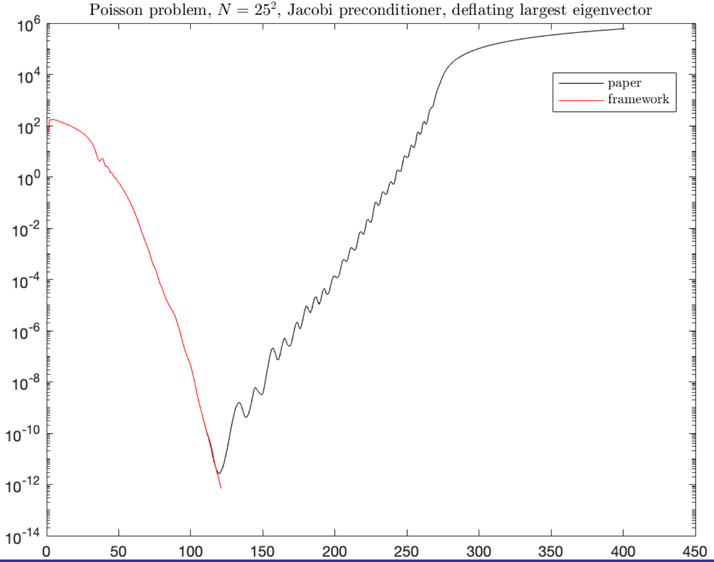
Leads to short recurrence updates for recycled subspace

[3] see, e.g., [Wang et al ‘07; Ahuja et al ‘12; Bolten et al ‘22]

How we exploit short-recurrences can effect stability. For example:

- DCG [Saad et al '00]: all at once approach folds augmenting subspace into search direction construction
- [Kahl & Rittich] first observed the projected problem formulation – apply CG directly to a projected subproblem and get update in $\mathcal{U}$ by projection at convergence
  $\rightarrow$ inherits stability characteristics of CG

DCG - A specific convergence diagram



Recycling for matrix function evaluation

The Arnoldi approximation of $f(\mathbf{A})\mathbf{b}$ Part I

$$\text{Evaluate: } f(\mathbf{A})\mathbf{b} = \frac{1}{2\pi i}\int_\Gamma f(\sigma)(\sigma\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}\, d\sigma$$

- $f(z) = \exp(z)$: the solution of ODEs
- $f(z) = \log(z)$: Markov model analysis
- $f(z) = \mathrm{sign}(z)$: lattice QCD simulations with overlap fermions
- $f(z) = \frac{1}{z}$: standard linear system solution

When $\mathbf{A}$ is large and sparse, matrix-free Krylov subspace methods used to approximate $f(\mathbf{A})\mathbf{b}$

Ingredients:

- $(\sigma\mathbf{I}-\mathbf{A})^{-1}\mathbf{b} \iff (\sigma\mathbf{I}-\mathbf{A})\mathbf{x}(\sigma)=\mathbf{b}$
- Shift Invariance: $\mathcal{K}_m(\sigma\mathbf{I}-\mathbf{A},\mathbf{b})=\mathcal{K}_m(\mathbf{A},\mathbf{b})$
- Shifted FOM Condition: $\mathbf{b}-(\sigma\mathbf{I}-\mathbf{A})(\mathbf{x}_0(\sigma)+\mathbf{V}_m\mathbf{y}_m(\sigma))\perp\mathcal{K}_m(\mathbf{A},\mathbf{b})$
- Shifted FOM Approximation: $\mathbf{y}_m(\sigma)=\|\mathbf{b}\|(\sigma\mathbf{I}-\mathbf{H}_m)^{-1}\mathbf{e}_1\Rightarrow\mathbf{x}(\sigma)\approx\mathbf{V}_m(\sigma\mathbf{I}-\mathbf{H}_m)^{-1}\mathbf{e}_1$
- Matrix function action approximation:

$$f(\mathbf{A})\mathbf{b}\approx\frac{\|\mathbf{b}\|}{2\pi i}\int_\Gamma f(\sigma)\mathbf{V}_m(\sigma\mathbf{I}-\mathbf{H}_m)^{-1}\mathbf{e}_1\,d\sigma=\|\mathbf{b}\|\mathbf{V}_m f(\mathbf{H}_m)\mathbf{e}_1$$

Ingredients:

- $(\sigma\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} \iff (\sigma\mathbf{I} - \mathbf{A})\mathbf{x}(\sigma) = \mathbf{b}$
- Shift Invariance: $\mathcal{K}_m(\sigma\mathbf{I} - \mathbf{A}, \mathbf{b}) = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Shifted FOM Condition:
  $\mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})(\mathbf{x}_0(\sigma) + \mathbf{V}_m\mathbf{y}_m(\sigma)) \perp \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Shifted FOM Approximation:
  $\mathbf{y}_m(\sigma) = \|\mathbf{b}\|(\sigma\mathbf{I} - \mathbf{H}_m)^{-1}\mathbf{e}_1 \Rightarrow \mathbf{x}(\sigma) \approx \mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m)^{-1}\mathbf{e}_1$
- Matrix function action approximation:

$$f(\mathbf{A})\mathbf{b} \approx \frac{\|\mathbf{b}\|}{2\pi i}\int_\Gamma f(\sigma)\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m)^{-1}\mathbf{e}_1 \, d\sigma = \|\mathbf{b}\|\mathbf{V}_m f(\mathbf{H}_m)\mathbf{e}_1$$

The Arnoldi approximation of $f(\mathbf{A})\mathbf{b}$ Part II

Ingredients:

- $(\sigma\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} \iff (\sigma\mathbf{I} - \mathbf{A})\mathbf{x}(\sigma) = \mathbf{b}$
- Shift Invariance: $\mathcal{K}_m(\sigma\mathbf{I} - \mathbf{A}, \mathbf{b}) = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Shifted FOM Condition:
  $\mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})(\mathbf{x}_0(\sigma) + \mathbf{V}_m\mathbf{y}_m(\sigma)) \perp \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Shifted FOM Approximation:
  $\mathbf{y}_m(\sigma) = \|\mathbf{b}\|(\sigma\mathbf{I} - \mathbf{H}_m)^{-1}\mathbf{e}_1 \Rightarrow \mathbf{x}(\sigma) \approx \mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m)^{-1}\mathbf{e}_1$
- Matrix function action approximation:

$$f(\mathbf{A})\mathbf{b} \approx \frac{\|\mathbf{b}\|}{2\pi i}\int_\Gamma f(\sigma)\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m)^{-1}\mathbf{e}_1 \, d\sigma = \|\mathbf{b}\|\mathbf{V}_m f(\mathbf{H}_m)\mathbf{e}_1$$

The Arnoldi approximation of $f(\mathbf{A})\mathbf{b}$ Part II

Ingredients:

- $(\sigma\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} \iff (\sigma\mathbf{I} - \mathbf{A})\mathbf{x}(\sigma) = \mathbf{b}$
- Shift Invariance: $\mathcal{K}_m(\sigma\mathbf{I} - \mathbf{A}, \mathbf{b}) = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Shifted FOM Condition:
  $\mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})(\mathbf{x}_0(\sigma) + \mathbf{V}_m\mathbf{y}_m(\sigma)) \perp \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Shifted FOM Approximation:
  $\mathbf{y}_m(\sigma) = \|\mathbf{b}\|(\sigma\mathbf{I} - \mathbf{H}_m)^{-1}\mathbf{e}_1 \Rightarrow \mathbf{x}(\sigma) \approx \mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m)^{-1}\mathbf{e}_1$
- Matrix function action approximation:

$$f(\mathbf{A})\mathbf{b} \approx \frac{\|\mathbf{b}\|}{2\pi i}\int_\Gamma f(\sigma)\mathbf{V}_m(\sigma\mathbf{I}-\mathbf{H}_m)^{-1}\mathbf{e}_1 \, d\sigma = \|\mathbf{b}\|\mathbf{V}_m f(\mathbf{H}_m)\mathbf{e}_1$$

The Arnoldi approximation of $f(\mathbf{A})\mathbf{b}$ Part II

Ingredients:

- $(\sigma\mathbf{I}-\mathbf{A})^{-1}\mathbf{b} \iff (\sigma\mathbf{I}-\mathbf{A})\mathbf{x}(\sigma)=\mathbf{b}$
- Shift Invariance: $\mathcal{K}_m(\sigma\mathbf{I}-\mathbf{A},\mathbf{b})=\mathcal{K}_m(\mathbf{A},\mathbf{b})$
- Shifted FOM Condition:
  $\mathbf{b}-(\sigma\mathbf{I}-\mathbf{A})(\mathbf{x}_0(\sigma)+\mathbf{V}_m\mathbf{y}_m(\sigma))\perp\mathcal{K}_m(\mathbf{A},\mathbf{b})$
- Shifted FOM Approximation:
  $\mathbf{y}_m(\sigma)=\|\mathbf{b}\|(\sigma\mathbf{I}-\mathbf{H}_m)^{-1}\mathbf{e}_1 \Rightarrow \mathbf{x}(\sigma)\approx\mathbf{V}_m(\sigma\mathbf{I}-\mathbf{H}_m)^{-1}\mathbf{e}_1$
- Matrix function action approximation:

$$f(\mathbf{A})\mathbf{b}\approx\frac{\|\mathbf{b}\|}{2\pi i}\int_\Gamma f(\sigma)\mathbf{V}_m(\sigma\mathbf{I}-\mathbf{H}_m)^{-1}\mathbf{e}_1\,d\sigma=\|\mathbf{b}\|\mathbf{V}_m f(\mathbf{H}_m)\mathbf{e}_1$$

Some related/background works

- Hochbruck and Lubich. *On Krylov subspace approximations to the matrix exponential operator*. SINUM 1997.
- Saad. *Analysis of some Krylov subspace approximations to the matrix exponential operator*. SINUM 1992.
- Eiermann, Ernst, and Güttel. *Deflated restarting for matrix functions*. SIMAX 2011.
- Simoncini. *Restarted full orthogonalization method for shifted linear systems*. BIT 2003.
- Many others. . .

Difficulties extending recycling to approximate $f(\mathbf{A})\mathbf{b}$

Find $\mathbf{t}_m(\sigma) \in \mathcal{V}_m$ corresponding to each σ

$$(\mathbf{I} - \mathbf{Q}_\sigma)(\sigma\mathbf{I} - \mathbf{A})\mathbf{t}(\sigma) = (\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b}$$

such that $\mathbf{r}_m(\sigma) = (\mathbf{I} - \mathbf{Q}_\sigma)(\mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})\mathbf{t}_m(\sigma)) \perp \tilde{\mathcal{V}}_m$

- Augmentation for each shifted system induces its own shift-dependent projected subproblem

$$(\mathbf{I} - \mathbf{Q}_\sigma)(\sigma\mathbf{I} - \mathbf{A})\mathbf{t}(\sigma) = (\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b}$$

- $\mathcal{K}_m((\mathbf{I} - \mathbf{Q}_\sigma)(\sigma\mathbf{I} - \mathbf{A}), (\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b})$ is no longer shift invariant

Alternative augmented projection formulation

Find $\mathbf{t}_m(\sigma) \in \mathcal{V}_m$ corresponding to each σ

$$(\sigma\mathbf{I} - \mathbf{A})\mathbf{t}(\sigma) = \mathbf{b}$$

such that $\mathbf{r}_m(\sigma) = \mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})\mathbf{t}_m(\sigma) \perp (\mathbf{I} - \mathbf{Q}_\sigma)^* \tilde{\mathcal{V}}_m$

- Moves projector from subproblem onto constraint space
- Applying non-projected augmentation enables use of Krylov shift invariance
- FOM-type condition $\tilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Multiple approaches to computing $\mathbf{x}_m(\sigma)$
- $(\mathbf{V}_m^*(\mathbf{I} - \mathbf{Q}_\sigma)[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T])\mathbf{y}_m(\sigma) = \mathbf{V}_{j+1}^*(\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b}$
- $\mathbf{x}_m(\sigma) = \mathbf{V}_m\mathbf{y}_m(\sigma) + \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*\mathbf{b} - \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T]\mathbf{y}_m(\sigma)$

Alternative augmented projection formulation

Find $\mathbf{t}_m(\sigma) \in \mathcal{V}_m$ corresponding to each σ

$$(\sigma\mathbf{I} - \mathbf{A})\mathbf{t}(\sigma) = \mathbf{b}$$

such that $\mathbf{r}_m(\sigma) = \mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})\mathbf{t}_m(\sigma) \perp (\mathbf{I} - \mathbf{Q}_\sigma)^*\tilde{\mathcal{V}}_m$

- Moves projector from subproblem onto constraint space
- Applying non-projected augmentation enables use of Krylov shift invariance
- FOM-type condition $\tilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Multiple approaches to computing $\mathbf{x}_m(\sigma)$
- $(\mathbf{V}_m^*(\mathbf{I} - \mathbf{Q}_\sigma)[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T])\mathbf{y}_m(\sigma) = \mathbf{V}_{j+1}^*(\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b}$
- $\mathbf{x}_m(\sigma) = \mathbf{V}_m\mathbf{y}_m(\sigma) + \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*\mathbf{b} - \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T]\mathbf{y}_m(\sigma)$

Alternative augmented projection formulation

Find $\mathbf{t}_m(\sigma) \in \mathcal{V}_m$ corresponding to each σ

$$(\sigma\mathbf{I} - \mathbf{A})\mathbf{t}(\sigma) = \mathbf{b}$$

such that $\mathbf{r}_m(\sigma) = \mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})\mathbf{t}_m(\sigma) \perp (\mathbf{I} - \mathbf{Q}_\sigma)^* \tilde{\mathcal{V}}_m$

- Moves projector from subproblem onto constraint space
- Applying non-projected augmentation enables use of Krylov shift invariance
- FOM-type condition $\tilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Multiple approaches to computing $\mathbf{x}_m(\sigma)$
- $(\mathbf{V}_m^*(\mathbf{I} - \mathbf{Q}_\sigma)[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T])\mathbf{y}_m(\sigma) = \mathbf{V}_{j+1}^*(\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b}$
- $\mathbf{x}_m(\sigma) = \mathbf{V}_m\mathbf{y}_m(\sigma) + \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*\mathbf{b} - \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T]\mathbf{y}_m(\sigma)$

Alternative augmented projection formulation

Find $\mathbf{t}_m(\sigma) \in \mathcal{V}_m$ corresponding to each σ

$$(\sigma\mathbf{I} - \mathbf{A})\mathbf{t}(\sigma) = \mathbf{b}$$

such that $\mathbf{r}_m(\sigma) = \mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})\mathbf{t}_m(\sigma) \perp (\mathbf{I} - \mathbf{Q}_\sigma)^*\tilde{\mathcal{V}}_m$

- Moves projector from subproblem onto constraint space
- Applying non-projected augmentation enables use of Krylov shift invariance
- FOM-type condition $\tilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Multiple approaches to computing $\mathbf{x}_m(\sigma)$
- $(\mathbf{V}_m^*(\mathbf{I} - \mathbf{Q}_\sigma)[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T])\mathbf{y}_m(\sigma) = \mathbf{V}_{j+1}^*(\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b}$
- $\mathbf{x}_m(\sigma) = \mathbf{V}_m\mathbf{y}_m(\sigma) + \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*\mathbf{b} - \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T]\mathbf{y}_m(\sigma)$

Alternative augmented projection formulation

Find $\mathbf{t}_m(\sigma) \in \mathcal{V}_m$ corresponding to each σ

$$(\sigma\mathbf{I} - \mathbf{A})\mathbf{t}(\sigma) = \mathbf{b}$$

such that $\mathbf{r}_m(\sigma) = \mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})\mathbf{t}_m(\sigma) \perp (\mathbf{I} - \mathbf{Q}_\sigma)^*\tilde{\mathcal{V}}_m$

- Moves projector from subproblem onto constraint space
- Applying non-projected augmentation enables use of Krylov shift invariance
- FOM-type condition $\tilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Multiple approaches to computing $\mathbf{x}_m(\sigma)$
- $(\mathbf{V}_m^*(\mathbf{I} - \mathbf{Q}_\sigma)[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T])\mathbf{y}_m(\sigma) = \mathbf{V}_{j+1}^*(\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b}$
- $\mathbf{x}_m(\sigma) = \mathbf{V}_m\mathbf{y}_m(\sigma) + \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*\mathbf{b} - \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T]\mathbf{y}_m(\sigma)$

Alternative augmented projection formulation

Find $\mathbf{t}_m(\sigma) \in \mathcal{V}_m$ corresponding to each σ

$$(\sigma\mathbf{I} - \mathbf{A})\mathbf{t}(\sigma) = \mathbf{b}$$

such that $\mathbf{r}_m(\sigma) = \mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})\mathbf{t}_m(\sigma) \perp (\mathbf{I} - \mathbf{Q}_\sigma)^*\tilde{\mathcal{V}}_m$

- Moves projector from subproblem onto constraint space
- Applying non-projected augmentation enables use of Krylov shift invariance
- FOM-type condition $\tilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Multiple approaches to computing $\mathbf{x}_m(\sigma)$
- $(\mathbf{V}_m^*(\mathbf{I} - \mathbf{Q}_\sigma)[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T])\mathbf{y}_m(\sigma) = \mathbf{V}_{j+1}^*(\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b}$
- $\mathbf{x}_m(\sigma) = \mathbf{V}_m\mathbf{y}_m(\sigma) + \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*\mathbf{b} - \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T]\mathbf{y}_m(\sigma)$

Alternative augmented projection formulation

Find $\mathbf{t}_m(\sigma) \in \mathcal{V}_m$ corresponding to each σ

$$(\sigma\mathbf{I} - \mathbf{A})\mathbf{t}(\sigma) = \mathbf{b}$$

such that $\mathbf{r}_m(\sigma) = \mathbf{b} - (\sigma\mathbf{I} - \mathbf{A})\mathbf{t}_m(\sigma) \perp (\mathbf{I} - \mathbf{Q}_\sigma)^*\tilde{\mathcal{V}}_m$

- Moves projector from subproblem onto constraint space
- Applying non-projected augmentation enables use of Krylov shift invariance
- FOM-type condition $\tilde{\mathcal{V}}_m = \mathcal{K}_m(\mathbf{A}, \mathbf{b})$
- Multiple approaches to computing $\mathbf{x}_m(\sigma)$
- $(\mathbf{V}_m^*(\mathbf{I} - \mathbf{Q}_\sigma)[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T])\mathbf{y}_m(\sigma) = \mathbf{V}_{j+1}^*(\mathbf{I} - \mathbf{Q}_\sigma)\mathbf{b}$
- $\mathbf{x}_m(\sigma) = \mathbf{V}_m\mathbf{y}_m(\sigma) + \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*\mathbf{b} - \mathbf{U}(\sigma\mathbf{U}^*\mathbf{U} - \mathbf{U}^*\mathbf{C})^{-1}\mathbf{U}^*[\mathbf{V}_m(\sigma\mathbf{I} - \mathbf{H}_m) - h_{j+1,j}\mathbf{v}_{j+1}\mathbf{e}_m^T]\mathbf{y}_m(\sigma)$

Recycled FOM for functions of matrices (rFOM^2)

$$f(\mathbf{A})\mathbf{b} \approx \tilde{\mathbf{f}} = \frac{1}{2\pi i}\int_{\Gamma} f(\sigma)\mathbf{x}_m(\sigma)\, d\sigma$$

- Choose an approach for computing $\mathbf{x}_m(\sigma)$
  - → Decoupled approach (last slide)
  - → All-at-once approach
  - → Matrix-function evaluation plus correction: $f(\mathbf{A})\mathbf{b} \approx \widehat{\mathbf{V}}_m f(\mathbf{G}_m)(\widehat{\mathbf{W}}_m^* \widehat{\mathbf{W}}_m)^{-1}\widehat{\mathbf{W}}_m^*\mathbf{b} - \widehat{\mathbf{V}}_m \mathcal{I} \widehat{\mathbf{W}}_m^* \mathbf{b}$[4][5]
- Choose appropriate contour and some quadrature technique to numerically integrate

---

[4] $\widehat{\mathbf{V}}_m = \begin{bmatrix} \mathbf{U} & \mathbf{V}_m \end{bmatrix}$, $\widehat{\mathbf{W}}_m = \begin{bmatrix} \mathbf{C} & \mathbf{V}_m \end{bmatrix}$

[5] $\mathcal{I} = \frac{1}{2\pi i}\int f(\sigma)((\widehat{\mathbf{W}}_m^*\widehat{\mathbf{W}}_m)(\sigma\mathbf{I} - \mathbf{G}_m))^{-1}(\mathbf{I} + \widehat{\mathbf{W}}_m^*\mathbf{R}_\sigma((\widehat{\mathbf{W}}_m^*\widehat{\mathbf{W}}_m)(\sigma\mathbf{I} - \mathbf{G}_m))^{-1})\widehat{\mathbf{W}}_m^*\mathbf{R}_\sigma((\widehat{\mathbf{W}}_m^*\widehat{\mathbf{W}}_m)(\sigma\mathbf{I} - \mathbf{G}_m))^{-1}\, d\sigma$

Eigenvector/Harmonic Ritz vector recycling

Recycle to project away (approximate) eigenvector directions associated to eigenvalues near singularities of $f(z)$

- Use exact eigenvectors if you have them (rare)
- For a sequence of problems, compute Harmonic Ritz vectors

$$\text{Find} \quad (\mathbf{y}_i, \mu_i) \quad \text{such that} \quad \mathbf{A}^{-1}\mathbf{y}_i - \mu_i\mathbf{y}_i \perp \mathbf{A}(\mathcal{K}_m(\mathbf{A},\mathbf{b}) + \mathcal{U})$$
$$\text{with} \quad \mathbf{y}_m \in \mathbf{A}(\mathcal{K}_m(\mathbf{A},\mathbf{b}) + \mathcal{U})$$

Numerical Results

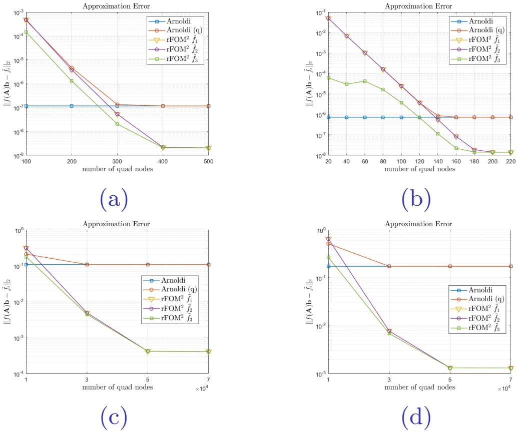
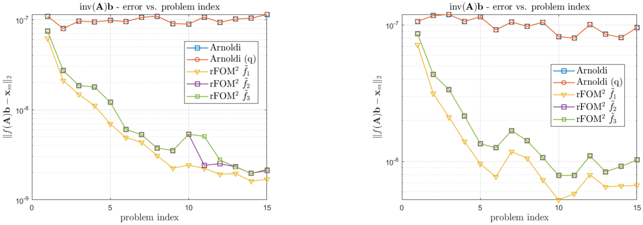


Figure: rFOM^2 approx $f(\mathbf{A})\mathbf{b}$ for (a) $f(z) = \mathrm{sign}(\mathbf{A})$, $\mathbf{A}$ is a Wilson Dirac (4^4 lattice); (b) $f(z) = \frac{1}{z}$ for the same; (c) $f(z) = \log(z)$ for $\mathbf{A}$ is a $10^5 \times 10^5$ chemical potential matrix; (d) $f(z) = \sqrt{z}$ for a $10^5 \times 10^5$ Poisson matrix. Cycle length $m = 40$ and recycle space dim. $k = 20$

Experiment 2: $\mathrm{sign}^{-1}(\mathbf{A} + \varepsilon \mathbf{R}_i)\mathbf{b}_i$



(a) (b)

Figure: Error comparison for a sequence of 15 applications of the inverse of the sign function to 15 random vectors. Cycle length $m = 40$ and recycle space dim. $k = 20$ and 2000 quadrature points. In fig (a) we took $\varepsilon = 0$, and in (b) $\varepsilon = 0.001$.

Experiment 3: $\mathrm{sign}(\mathbf{A} + \varepsilon \mathbf{R}_i)\mathbf{b}_i$

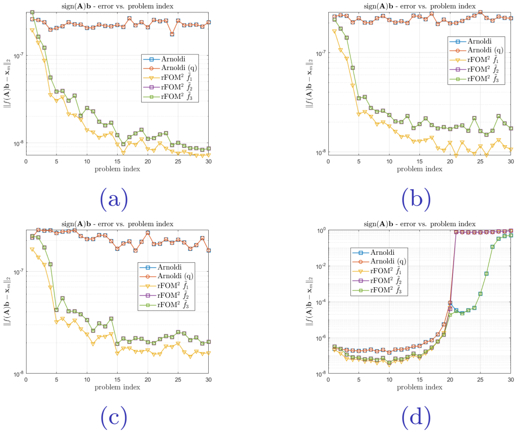


Figure: rFOM^2 30 applications of sign function on for different values of ε. (a) $\varepsilon = 0$; (b) $\varepsilon = 0.0001$; (c) $\varepsilon = 0.001$; (d) $\varepsilon = 0.01$. Cycle length $m = 40$ and recycle space dim. $k = 20$ and 2000 quadrature points. *(d) demonstrates need to modify contour for 20^{th} system.*

Conclusions and future work

- Understanding recycling/augmented iterative methods in a common framework brings tangible benefits
- Straightforward design roadmap for new methods
- Methods inherit convergence and stability properties of the method applied to the projected subproblem
- Framework provides clear path to extending recycling to treating matrix function evaluation
- **Future**: Restarting and error monitoring for recycling for matrix functions; Augmented/recycling methods for iterative solvers for other complicated problems (non-linear, Kronecker/Tensor problems, etc)

References

Results in this talk are drawn from the following manuscripts

- S., Kilmer, de Sturler *A survey of subspace recycling iterative methods.*, GAMM Mitteilungen, 2020.
- S. *A note on augmented unprojected Krylov subspace methods*, ETNA 2022.
- Burke, Frommer, Ramirez-Hidalgo, S. *Krylov subspace recycling for matrix functions*, 2022 (soon on arXiv)
- S., Kilmer, de Sturler *Design and implementation of new recycling methods based on framework.*, 202?.

Some other related works

- Hutterer, Ramlau, and S. *Subspace Recycling-based Regularization Methods*, SIMAX 2021.
- Brennan, Islam, Basquill, S. *Computation of Scattering from Rough Surfaces using Successive Symmetric Over Relaxation and Eigenvalue Deflation*, 16th EuCAP 2022.

For more information: `https://math.soodhalter.com`

References

Results in this talk are drawn from the following manuscripts

- S., Kilmer, de Sturler *A survey of subspace recycling iterative methods.*, GAMM Mitteilungen, 2020.
- S. *A note on augmented unprojected Krylov subspace methods*, ETNA 2022.
- Burke, Frommer, Ramirez-Hidalgo, S. *Krylov subspace recycling for matrix functions*, 2022 (soon on arXiv)
- S., Kilmer, de Sturler *Design and implementation of new recycling methods based on framework.*, 202?.

Some other related works

- Hutterer, Ramlau, and S. *Subspace Recycling-based Regularization Methods*, SIMAX 2021.
- Brennan, Islam, Basquill, S. *Computation of Scattering from Rough Surfaces using Successive Symmetric Over Relaxation and Eigenvalue Deflation*, 16th EuCAP 2022.

For more information: `https://math.soodhalter.com`

Bonus Slides!

Augmented stationary iterations

Framework accommodates deflated stationary methods; see, e.g., [Burrage et al '98; Brennan et al '22]

- Many stationary methods have a residual constraint formulation
- We can build augmented stationary iterative schemes
- Such a method was proposed as a deflation scheme in '98
- Approximately deflate eigenvalues of the iteration matrix > 1
- We applied this to an SSOR technique for scattering of electromagnetic waves from randomly rough surfaces
  - $\rightarrow$ More complicated surface leads to iteration matrix with more "bad" eigenvalues
  - $\rightarrow$ Adaptive: iteration allows for estimation of these eigenvectors

Analysis of methods as a regularization

Can augmentation techniques be used reliably to treat ill-posed problems?

- Plenty of methods proposed already
- General strategy: apply regularization method to

$$(\mathbf{I} - \mathbf{Q})\mathbf{A}\mathbf{t} = (\mathbf{I} - \mathbf{Q})\mathbf{r}_0$$

- Regularization analysis: in infinite-dimensional setting for ill-posed operator $T : \mathcal{X} \to \mathcal{Y}$ mapping between Hilbert spaces

$$Tx = y + e^{\delta}$$

- Recycling methods can be formally posed in this infinite dimensional setting [de Sturler, Kilmer, S.]

Apply any regularizer to $(\mathbf{I} - \mathbf{Q})\mathbf{A}\mathbf{t} = (\mathbf{I} - \mathbf{Q})\mathbf{r}_0$

- We can treat the subproblem with any reg. method
- Noise-based estimates for the error incurred by all parts of the method available
- Formal analysis in the regularization-theory sense as noise-level $\delta \to 0$ possible
- Overall residual behavior determined by residual of projected subproblem

Theorem (Hutterer, Ramlau, and S.)

$Tx = y$ with $T : \mathcal{X} \to \mathcal{Y}$ (Hilbert spaces) with y^δ, $\left\|y - y^\delta\right\|_{\mathcal{Y}} < \delta$. Treating the projected subproblem with any regularization method is itself a regularization method; i.e., as $\delta \to 0$, $x_m^\delta \to x^\dagger = T^\dagger y$.

Apply <u>any</u> regularizer to $(\mathbf{I}-\mathbf{Q})\mathbf{A}\mathbf{t} = (\mathbf{I}-\mathbf{Q})\mathbf{r}_0$

- We can treat the subproblem with any reg. method
- Noise-based estimates for the error incurred by all parts of the method available
- Formal analysis in the regularization-theory sense as noise-level $\delta \to 0$ possible
- Overall residual behavior determined by residual of projected subproblem

Theorem (Hutterer, Ramlau, and S.)

$Tx = y$ with $T : \mathcal{X} \to \mathcal{Y}$ (Hilbert spaces) with y^δ, $\left\|y - y^\delta\right\|_{\mathcal{Y}} < \delta$. Treating the projected subproblem with <u>any</u> regularization method is itself a regularization method; i.e., as $\delta \to 0$, $x_m^\delta \to x^\dagger = T^\dagger y$.

Augmented steepest descent for the normal equations

Given: $\mathbf{U} \in \mathcal{X}^k$ representing $\mathcal{U}$
Set $r_0 = y - Tx_0$
Compute "QR-factorization" (Gram-Schmidt) $T\mathbf{U} = \mathbf{CR}$
$\mathbf{U} \leftarrow \mathbf{UR}^{-1}$
$\mathbf{z}^{(1)} = (r_0, \mathbf{C})_{\mathcal{Y}}$
$x \leftarrow x_0 + \mathbf{Uz}^{(1)}$
$r \leftarrow r_0 - \mathbf{Cz}^{(1)}$
while *STOPPING-CRITERIA* **do**
    $\alpha_i = \frac{\|T^* r_i\|_{\mathcal{Y}}^2}{\|(I_{\mathcal{Y}} - Q)T^*Tr_i\|_{\mathcal{Y}}^2}$
    $\widehat{\mathbf{w}} \leftarrow= (TT^*r, \mathbf{C})_{\mathcal{Y}}$
    $x \leftarrow x + \alpha_i T^* r - \alpha_i \mathbf{U}\widehat{\mathbf{w}}$
    $r \leftarrow r - \alpha_i TT^* r + \alpha_i \mathbf{C}\widehat{\mathbf{w}}$
end

Augmented Landweber for the normal equations

Given: $\mathbf{U} \in \mathcal{X}^k$ representing $\mathcal{U}$, $\alpha > 0$
Set $r_0 = y - Tx_0$
Compute "QR-factorization" (Gram-Schmidt) $T\mathbf{U} = \mathbf{C}\mathbf{R}$
$\mathbf{U} \leftarrow \mathbf{U}\mathbf{R}^{-1}$
$\mathbf{z}^{(1)} = (r_0, \mathbf{C})_{\mathcal{Y}}$
$x \leftarrow x_0 + \mathbf{U}\mathbf{z}^{(1)}$
$r \leftarrow r_0 - \mathbf{C}\mathbf{z}^{(1)}$
while *STOPPING-CRITERIA* **do**
    $\widehat{\mathbf{w}} \leftarrow= (TT^*r, \mathbf{C})_{\mathcal{Y}}$
    $x \leftarrow x + \alpha T^*r - \alpha\mathbf{U}\widehat{\mathbf{w}}$
    $r \leftarrow r - \alpha TT^*r + \alpha\mathbf{C}\widehat{\mathbf{w}}$
end