Analysis of block GMRES using a new *-algebra-based approach

Kirk M. Soodhalter

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

15.–17. August 2022

with Marie Kubínová (formerly, Czech Acad. Science Ostrava)

Solution of Large-sparse linear system

Goal: approximate the solution to a large, (often) sparse linear system,

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \text{where} \quad \mathbf{A} \in \mathbb{C}^{n \times n} \quad \text{and} \quad n \gg 0$$

- **Sparse** means most of the matrix entries are zeros.
- More generally: matrices which allow for fast application (e.g., FFT-based)

Krylov Subspaces

Given $\mathbf{A}$ and $\mathbf{b}$, the jth Krylov subspace is defined

$$\mathcal{K}_j(\mathbf{A}, \mathbf{b}) = \operatorname{span}\left\{\mathbf{b}, \mathbf{A}\mathbf{b}, \ldots, \mathbf{A}^{j-1}\mathbf{b}\right\}.$$

Thus, $\mathbf{u} \in \mathcal{K}_j(\mathbf{A}, \mathbf{b})$ is such that

$$\mathbf{u} = p(\mathbf{A})\mathbf{b}$$

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

Definition

The basis $\left\{\mathbf{b}, \mathbf{A}\mathbf{b}, \ldots, \mathbf{A}^{j-1}\mathbf{b}\right\}$ is called a **Krylov basis**.

Selecting Approximations from $\mathcal{K}_j(\mathbf{A}, \mathbf{b})$

- In many Krylov subspace methods, we select $\mathbf{x}_j \in \mathcal{K}_j(\mathbf{A}, \mathbf{b})$, so that

$$\mathbf{x}_j = p_j(\mathbf{A})\mathbf{b}$$

  Why?

- The inverse $\mathbf{A}^{-1}$ of any nonsingular matrix $\mathbf{A}$ can be written as

$$\mathbf{A}^{-1} = q(\mathbf{A})$$

  where $q(x)$ is a polynomial of degree less than n.

- We want $p_j(x)$ to be a low-degree "approximation" to $q(x)$...

  $\rightarrow$ only need to approximate action $p_j(\mathbf{A})\mathbf{b} \approx q(\mathbf{A})\mathbf{b}$

Selecting Approximations from $\mathcal{K}_j(\mathbf{A}, \mathbf{b})$

- In many Krylov subspace methods, we select $\mathbf{x}_j \in \mathcal{K}_j(\mathbf{A}, \mathbf{b})$, so that

$$\mathbf{x}_j = p_j(\mathbf{A})\mathbf{b}$$

  Why?

- The inverse $\mathbf{A}^{-1}$ of any nonsingular matrix $\mathbf{A}$ can be written as

$$\mathbf{A}^{-1} = q(\mathbf{A})$$

  where $q(x)$ is a polynomial of degree less than n.

- We want $p_j(x)$ to be a low-degree "approximation" to $q(x)$...

  $\rightarrow$ only need to approximate action $p_j(\mathbf{A})\mathbf{b} \approx q(\mathbf{A})\mathbf{b}$

Selecting Approximations from $\mathcal{K}_j(\mathbf{A}, \mathbf{b})$

- In many Krylov subspace methods, we select $\mathbf{x}_j \in \mathcal{K}_j(\mathbf{A}, \mathbf{b})$, so that

$$\mathbf{x}_j = p_j(\mathbf{A})\mathbf{b}$$

  Why?

- The inverse $\mathbf{A}^{-1}$ of any nonsingular matrix $\mathbf{A}$ can be written as

$$\mathbf{A}^{-1} = q(\mathbf{A})$$

  where $q(x)$ is a polynomial of degree less than n.

- We want $p_j(x)$ to be a low-degree "approximation" to $q(x)$...

  $\rightarrow$ only need to approximate action $p_j(\mathbf{A})\mathbf{b} \approx q(\mathbf{A})\mathbf{b}$

Selecting Approximations from $\mathcal{K}_j(\mathbf{A}, \mathbf{b})$

- In many Krylov subspace methods, we select $\mathbf{x}_j \in \mathcal{K}_j(\mathbf{A}, \mathbf{b})$, so that

$$\mathbf{x}_j = p_j(\mathbf{A})\mathbf{b}$$

  Why?
- The inverse $\mathbf{A}^{-1}$ of any nonsingular matrix $\mathbf{A}$ can be written as

$$\mathbf{A}^{-1} = q(\mathbf{A})$$

  where $q(x)$ is a polynomial of degree less than n.
- We want $p_j(x)$ to be a low-degree "approximation" to $q(x)\ldots$

  $\rightarrow$ only need to approximate action $p_j(\mathbf{A})\mathbf{b} \approx q(\mathbf{A})\mathbf{b}$

A General Linear System

$\mathbf{A}\left(\mathbf{x}_0 + t\right) = \mathbf{b}$ with $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\mathbf{b} \in \mathbb{C}^n$

- For $\mathbf{x}_0$, let $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0 \implies \mathbf{A}\mathbf{t} = \mathbf{r}_0$
- Krylov subspace: $\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0) = \operatorname{span}\left\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{j-1}\mathbf{r}_0\right\}$.
- Choose $\mathbf{x}_j = \mathbf{x}_0 + \mathbf{t}_j$. Let $\mathbf{r}_j = \mathbf{b} - \mathbf{A}\mathbf{x}_j$.
- GMRES - Generalized Minimum Residual Method
- For GMRES, construct $\mathbf{x}_j = \mathbf{x}_0 + \mathbf{t}_j$ where $\mathbf{t}_j \in \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$ such that $\mathbf{t}_j$ minimizes

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

- This is equivalent to $\mathbf{r}_j \perp \mathbf{A}\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$
- Sibling method: **Full Orthogonalization Method (FOM)** – $\mathbf{r}_j \perp \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$

GMRES

A General Linear System

$\mathbf{A}\left(\mathbf{x}_0 + t\right) = \mathbf{b}$ with $\mathbf{A} \in \mathbb{C}^{n\times n}$, $\mathbf{b} \in \mathbb{C}^n$

- For $\mathbf{x}_0$, let $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0 \implies \mathbf{A}\mathbf{t} = \mathbf{r}_0$
- Krylov subspace: $\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0) = \operatorname{span}\left\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{j-1}\mathbf{r}_0\right\}$.
- Choose $\mathbf{x}_j = \mathbf{x}_0 + \mathbf{t}_j$. Let $\mathbf{r}_j = \mathbf{b} - \mathbf{A}\mathbf{x}_j$.
- GMRES - Generalized Minimum Residual Method
- For GMRES, construct $\mathbf{x}_j = \mathbf{x}_0 + \mathbf{t}_j$ where $\mathbf{t}_j \in \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$ such that $\mathbf{t}_j$ minimizes

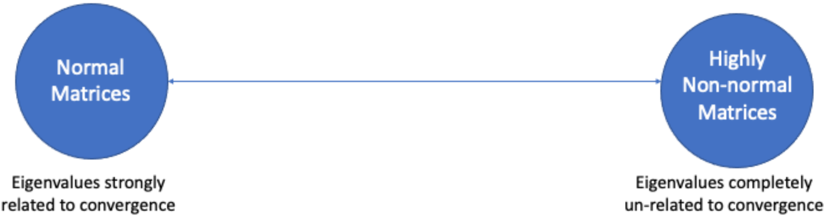
$$\min_{\mathbf{t}\in\mathcal{K}_j(\mathbf{A},\mathbf{r}_0)} \left\|\mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \mathbf{t})\right\|$$

- This is equivalent to $\mathbf{r}_j \perp \mathbf{A}\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$
- Sibling method: **Full Orthogonalization Method (FOM)** – $\mathbf{r}_j \perp \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$

Role of eigenvalues in residual convergence

GMRES polynomial minimization problem

$$
\begin{aligned}
\|\mathbf{r}_j\| &= \min_{\substack{q \in \Pi_j \\ q(0)=1}} \|q(\mathbf{A})\mathbf{r}_0\| \\
&\leq \mathcal{K}_2(\mathbf{X}) \min_{\substack{q \in \Pi_j \\ q(0)=1}} \max_{\lambda \in \sigma(\mathbf{A})} |q(\lambda)| \, \|\mathbf{r}_0\|
\end{aligned}
$$



Properties of the $\mathbf{A}$ determining residual convergence

Theorem (Greenbaum, Ptàk, and Strakoš 1996)

Given any non-increasing sequence

$$f(0) \geq f(1) \geq \cdots \geq f(n-1) > 0,$$

there exists matrices $\mathbf{A} \in \mathbb{C}^{n \times n}$ *and vectors* $\mathbf{r}_0$, $\|\mathbf{r}_0\| = f(0)$ *such that GMRES applied to* $\mathbf{A}\mathbf{t} = \mathbf{r}_0$ *produces residuals* $\mathbf{r}_k$, $\|\mathbf{r}_k\| = f(k)$ *for all* k.

An $\mathbf{A}$ *can be constructed to have any eigenvalues.*

Selected previous work analyzing GMRES/FOM

The relationship between GMRES and FOM

- Relationship of FOM/GMRES convergence: *[Walker '95], [Zhou and Walker '94], [Brown '91], [Saad '03]*
- Galerkin/norm minimizing pairs of methods (e.g., BiCG/QMR): *[Cullum '95], [Cullum and Greenbaum '96]*
- Geometric analysis: *[Eiermann and Ernst '01]*

Constructing matrices with predetermined GMRES convergence

- Any nonincreasing convergence curve is possible for GMRES: *[Greenbaum et al, 1996]*
- Parameterization of the pairs $(\mathbf{A}, \mathbf{b})$ producing specific convergence: *[Arioli et al, 1998]*
- Any Admissible Ritz/harmonic Ritz values: *[Du et al, 2017], [Tebbens and Meurant, 2012]*
- Any admissible CG convergence possible (cannot also specify eigenvalues) *[Meurant 2022]*

What happens if one has multiple right-hand sides?

Block Krylov subspaces

- Consider: $\mathbf{A}\mathbf{X} = \mathbf{B} \in \mathbb{C}^{n\times s}$, $s > 1$
- Let $\mathbf{X}_0 \in \mathbb{C}^{n\times s}$ and

$$\mathbf{F}_0 = \mathbf{B} - \mathbf{A}\mathbf{X}_0 = \begin{bmatrix} \mathbf{f}_0^{(1)} & \mathbf{f}_0^{(2)} & \mathbf{f}_0^{(3)} & \cdots & \mathbf{f}_0^{(s)} \end{bmatrix} \in \mathbb{C}^{n\times s}.$$

- Then we have the **block Krylov subspace**

$$\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0) = \mathcal{K}_j(\mathbf{A}, \mathbf{f}_0^{(1)}) + \mathcal{K}_j(\mathbf{A}, \mathbf{f}_0^{(2)}) + \cdots + \mathcal{K}_j(\mathbf{A}, \mathbf{f}_0^{(s)}).$$

- Assumption: $\dim \mathbb{K}_j(\mathbf{A}, \mathbf{F}_0) = js$

Block Arnoldi process

- Let $\mathbf{F}_0 = \mathbf{V}_1 \mathbf{S}_0$ be a skinny QR-factorization.
- At step j we get $\mathbf{V}_{j+1} \in \mathbb{C}^{n \times s}$ with orthonormal columns
- $\mathbf{W}_j = \begin{bmatrix} \mathbf{V}_1, & \ldots, & \mathbf{V}_j \end{bmatrix} \in \mathbb{C}^{n \times js}$ is basis of $\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)$
- Arnoldi relation: $\mathbf{A}\mathbf{W}_j = \mathbf{W}_{j+1}\overline{\mathbf{H}}_j$, $\overline{\mathbf{H}}_j$
- $\overline{\mathbf{H}}_j = (\mathbf{H}_{ik})_{ik} \in \mathbb{C}^{(j+1)s \times js}$ is block upper Hessenberg
- For $\blacksquare$, ${\color{red}\blacktriangledown} \in \mathbb{C}^{s \times s}$ and ${\color{red}\blacktriangledown}$ upper triangular

$$\overline{\mathbf{H}}_j = \begin{bmatrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare & \cdots & \blacksquare \\ {\color{red}\blacktriangledown} & \blacksquare & \blacksquare & \blacksquare & \cdots & \blacksquare \\ & {\color{red}\blacktriangledown} & \blacksquare & \blacksquare & \cdots & \blacksquare \\ & & {\color{red}\blacktriangledown} & \blacksquare & \cdots & \blacksquare \\ & & & {\color{red}\blacktriangledown} & & \blacksquare \\ & & & & \ddots & \vdots \\ & & & & & {\color{red}\blacktriangledown} \end{bmatrix} \in \mathbb{C}^{(j+1)s \times js}$$

From scalars to $s \times s$ matrices

- Orthogonalization:

$$\mathbf{v} \leftarrow \mathbf{v} - \underbrace{(\mathbf{q}^*\mathbf{v})}_{\in\mathbb{C}}\,\mathbf{q} \qquad \text{becomes} \qquad \mathbf{V} \leftarrow \mathbf{V} - \mathbf{Q}\,\underbrace{(\mathbf{Q}^*\mathbf{V})}_{\in\mathbb{C}^{s\times s}}$$

- Linear combinations:

$$\mathbf{u} = \sum_{i=1}^{k} \underbrace{\alpha_i}_{\in\mathbb{C}} \underbrace{\mathbf{v}_i}_{\in\mathbb{C}^n} \qquad \text{becomes} \qquad \mathbf{U} = \sum_{i=1}^{k} \underbrace{\mathbf{V}_i}_{\in\mathbb{C}^{n\times s}} \underbrace{\boldsymbol{\alpha}_i}_{\in\mathbb{C}^{s\times s}}$$

Block GMRES and Block FOM

Block GMRES and Block FOM valid for all $s \geq 1$

- Build an orthonormal basis for $\mathbb{K}_m(\mathbf{A}, \mathbf{F}_0)$
- For block GMRES

  Compute $\mathbf{Y}_m^{(G)} = \underset{Y \in \mathbb{C}^{ms \times s}}{\operatorname{argmin}} \left\| \overline{\mathbf{H}}_m \mathbf{Y} - \mathbf{E}_1^{(m+1)} \mathbf{S}_0 \right\|_F$ [a]

  Set $\mathbf{X}_m^{(G)} = \mathbf{X}_0 + \mathbf{W}_m \mathbf{Y}_m^{(G)}$, $\mathbf{R}_m^{(G)} = \mathbf{B} - \mathbf{A}\mathbf{X}_m^{(G)}$
- For block FOM

  Compute $\mathbf{Y}_m^{(F)} = \mathbf{H}_m^{-1} \mathbf{E}_1^{[m]} \mathbf{S}_0$ [b]

  Set $\mathbf{X}_m^{(F)} = \mathbf{X}_0 + \mathbf{W}_m \mathbf{Y}_m^{(F)}$, $\mathbf{R}_m^{(F)} = \mathbf{B} - \mathbf{A}\mathbf{X}_m^{(F)}$

---

[a] $\mathbf{E}_1^{(m+1)} \in \mathbb{C}^{(m+1)s \times s}$ has appropriate columns of an identity matrix

[b] $\mathbf{E}_1^{[m]} \in \mathbb{C}^{ms \times s}$ has appropriate columns of an identity matrix

Pros and cons of block Krylov methods

Pros

- Constraining residuals over larger subspaces
  $\rightarrow$ Leads to convergence in fewer iterations
- Block matrix-vector product has more efficient data movement characteristics–i.e., computational intensity

Cons

- More operations per iteration
- Increased operation cost thought to not justify by increase in convergence rate
- Interactions between systems makes analysis more difficult

Renewed interest in block methods in HPC setting necessitates new analysis to extend existing non-block results to block Krylov subspace case

Pros and cons of block Krylov methods

Pros

- Constraining residuals over larger subspaces
  $\rightarrow$ Leads to convergence in fewer iterations
- Block matrix-vector product has more efficient data movement characteristics–i.e., computational intensity

Cons

- More operations per iteration
- Increased operation cost thought to not justify by increase in convergence rate
- Interactions between systems makes analysis more difficult

Renewed interest in block methods in HPC setting necessitates new analysis to extend existing non-block results to block Krylov subspace case

Pros and cons of block Krylov methods

Pros

- Constraining residuals over larger subspaces
  - $\rightarrow$ Leads to convergence in fewer iterations
- Block matrix-vector product has more efficient data movement characteristics–i.e., computational intensity

Cons

- More operations per iteration
- Increased operation cost thought to not justify by increase in convergence rate
- Interactions between systems makes analysis more difficult

Renewed interest in block methods in HPC setting necessitates new analysis to extend existing non-block results to block Krylov subspace case

Selected previous work on analysis of block GMRES

- Convergence analysis: [Simoncini and Gallopoulos; 1997]
- Block Grade: [Gutknecht and Schmelzer; 2009]
- Relationship to block FOM and characterization of stagnation [S.; 2017]
- *-algebra framework [Frommer, Lund, Szyld; 2017]

The *-algebra framework

We follow [Frommer et al 2017] and consider the problem over *-algebra $\mathbb{S}$ of complex $s \times s$ matrices. We define a framework of corresponding objects and operations over $\mathbb{C}$ and over $\mathbb{S}$.

- $\mathbf{A} \in \mathbb{C}^{ns \times ns} \to \mathbf{A} \in \mathbb{S}^{n \times n}$
- $\mathbf{B} \in \mathbb{C}^{ns} \to \mathbf{B} \in \mathbb{S}^{n}$
- $\mathbb{K}_j(\mathbf{A}, \mathbf{B}) = \text{blockspan}\{\mathbf{B}, \mathbf{AB}, \ldots, \mathbf{A}^{j-1}\mathbf{B}\}$
- $\sum_{i=1}^{j} \mathbf{V}_i \, \mathbf{D}_i, \quad \mathbf{D}_i \in \mathbb{C}^{s \times s}$ is a block linear combination
- $\{\mathbf{V}_1, \ldots, \mathbf{V}_j\}$ is the basis of this subspace

The *-algebra framework - definitions

standard	block
$\mathbb{C}$	$\mathbb{S} = \mathbb{C}^{s\times s}$
$\mathbb{R}^+$	$\mathbb{S}^+ \ldots$ upper-Δ with positive diag. entries
$\mathbb{R}_0^+$	$\mathbb{S}_0^+ \ldots$ upper-Δ with nonnegative diag. entries
0	singular $s \times s$ matrix
1	I

The *-algebra framework - definitions

standard	block
$\mathbb{C}$	$\mathbb{S} = \mathbb{C}^{s\times s}$
$\mathbb{R}^+$	$\mathbb{S}^+ \ldots$ upper-Δ with positive diag. entries
$\mathbb{R}_0^+$	$\mathbb{S}_0^+ \ldots$ upper-Δ with nonnegative diag. entries
0	singular $s \times s$ matrix (zero divisors!)
1	I

The *-algebra framework - properties I

standard	block
$a, b \in \mathbb{C}$	$\mathbf{A}, \mathbf{B} \in \mathbb{S}$
$\lvert a \rvert = \sqrt{a^* a} \in \mathbb{R}_0^+$	$\lvert \mathbf{A} \rvert = \sqrt{\mathbf{A}^* \mathbf{A}} \equiv \mathrm{cholUT}(\mathbf{A}^* \mathbf{A}) \in \mathbb{S}_0^+$
$\lvert a \rvert \in \mathbb{R}^+ \Longleftrightarrow a \neq 0$	$\lvert \mathbf{A} \rvert \in \mathbb{S}^+ \Longleftrightarrow \mathbf{A}$ nonsingular

The *-algebra framework - properties II

standard	block
$\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$	$\mathbf{X}, \mathbf{Y} \in \mathbb{S}^n (= \mathbb{C}^{ns \times s})$
$\langle \mathbf{x}, \mathbf{y} \rangle \equiv \mathbf{y}^* \mathbf{x} \in \mathbb{C}$	$\langle\langle \mathbf{X}, \mathbf{Y} \rangle\rangle \equiv \mathbf{Y}^* \mathbf{X} \in \mathbb{S}$
$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^*$	$\langle\langle \mathbf{X}, \mathbf{Y} \rangle\rangle = \langle\langle \mathbf{Y}, \mathbf{X} \rangle\rangle^*$
$\langle \mathbf{x}a, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle a$	$\langle\langle \mathbf{X}\mathbf{A}, \mathbf{Y} \rangle\rangle = \langle\langle \mathbf{X}, \mathbf{Y} \rangle\rangle \mathbf{A}$
$\langle \mathbf{x}, \mathbf{y}a \rangle = a^* \langle \mathbf{x}, \mathbf{y} \rangle$	$\langle\langle \mathbf{X}, \mathbf{Y}\mathbf{A} \rangle\rangle = \mathbf{A}^* \langle\langle \mathbf{X}, \mathbf{Y} \rangle\rangle$
$\Vert \mathbf{x} \Vert \equiv \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \in \mathbb{R}_0^+$	$\vert\vert\vert \mathbf{X} \vert\vert\vert \equiv \sqrt{\langle\langle \mathbf{X}, \mathbf{X} \rangle\rangle} \in \mathbb{S}_0^+$
$\langle \mathbf{x}, \mathbf{y} \rangle = \Vert \mathbf{x} \Vert \, \Vert \mathbf{y} \Vert \cos \theta_{\mathbf{x},\mathbf{y}}$	$\langle\langle \mathbf{X}, \mathbf{Y} \rangle\rangle = \vert\vert\vert \mathbf{Y} \vert\vert\vert^* \mathbf{U} \, \mathrm{diag}(c_i) \mathbf{V}^* \vert\vert\vert \mathbf{X} \vert\vert\vert$

Block Arnoldi revisited

- Let $\mathbf{F}_0 = \mathbf{V}_1\mathbf{S}_0$; $\mathbf{V}_1 \in \mathbb{S}^n$ and $\mathbf{S}_0 = |||\mathbf{F}_0||| \in \mathbb{S}^+$
- The **block Arnoldi process** is generally performed in terms of $\langle\langle\cdot,\cdot\rangle\rangle$
- $\mathbf{W}_j = \begin{bmatrix}\mathbf{V}_1, & \ldots, & \mathbf{V}_j\end{bmatrix} \in \mathbb{S}^{n\times j}$ has orthonormal columns
- Arnoldi relation: $\mathbf{A}\mathbf{W}_j = \mathbf{W}_{j+1}\overline{\mathbf{H}}_j$
- $\overline{\mathbf{H}}_j = (\mathbf{H}_{ik})_{ik} \in \mathbb{S}^{(j+1)\times j}$ is upper Hessenberg
- For $\blacksquare \in \mathbb{S}$ and $\blacktriangledown \in \mathbb{S}^+$

$$\overline{\mathbf{H}}_j = \begin{bmatrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare & \cdots & \blacksquare \\ \blacktriangledown & \blacksquare & \blacksquare & \blacksquare & \cdots & \blacksquare \\ & \blacktriangledown & \blacksquare & \blacksquare & \cdots & \blacksquare \\ & & \blacktriangledown & \blacksquare & \cdots & \blacksquare \\ & & & \blacktriangledown & & \blacksquare \\ & & & & \ddots & \vdots \\ & & & & & \blacktriangledown \end{bmatrix}$$

Peak-plateau relationship between blFOM and blGMRES

Proposition (Kubínová and S. 2020)

The blGMRES and blFOM residuals satisfy:

$$\langle\langle \mathbf{R}_k^F, \mathbf{R}_k^F \rangle\rangle^{-1} = \langle\langle \mathbf{R}_k^G, \mathbf{R}_k^G \rangle\rangle^{-1} - \langle\langle \mathbf{R}_{k-1}^G, \mathbf{R}_{k-1}^G \rangle\rangle^{-1}.$$

Applying this relation recursively, we obtain

$$\langle\langle \mathbf{R}_k^G, \mathbf{R}_k^G \rangle\rangle^{-1} = \sum_{i=0}^{k} \langle\langle \mathbf{R}_i^F, \mathbf{R}_i^F \rangle\rangle^{-1}.$$

Generalization of the ordering of $\mathbb{R}_0^+$

Generalize the ordering of nonnegative real numbers $\mathbb{R}_0^+$ to upper triangular matrices with nonnegative diagonal entries $\mathbb{S}_0^+$ as follows:

$$|\mathbf{A}| \prec |\mathbf{B}| \iff \mathbf{A}^*\mathbf{A} \overset{\text{Löwner}}{\prec} \mathbf{B}^*\mathbf{B},$$
$$|\mathbf{A}| \preceq |\mathbf{B}| \iff \mathbf{A}^*\mathbf{A} \overset{\text{Löwner}}{\preceq} \mathbf{B}^*\mathbf{B}.$$

Peak-plateau result has some nontrivial consequences for the convergence behavior of blGMRES. In particular, the ordering of the residual norms

bIGMRES residual norm ordering

Theorem (Kubínová and S. 2020)

The blGMRES residuals satisfy

$$|||\mathbf{R}_0||| \succeq |||\mathbf{R}_1^G||| \succeq \cdots \succeq |||\mathbf{R}_{n-1}^G||| \succeq 0.$$

Definition (Admissible convergence sequence)

Any sequence $\{\mathbf{F}_k\}_{k=0}^{n-1} \subset \mathbb{S}^+$ that satisfies

$$\mathbf{F}_0 \succeq \mathbf{F}_1 \succeq \cdots \succeq \mathbf{F}_{n-1} \succ 0$$

is called an admissible convergence sequence.

Note: One can construct non-trivial examples of inadmissible sequences where the individual column norms decrease monotonically

blGMRES residual norm ordering

Theorem (Kubínová and S. 2020)

The blGMRES residuals satisfy

$$|||\mathbf{R}_0||| \succeq |||\mathbf{R}_1^G||| \succeq \cdots \succeq |||\mathbf{R}_{n-1}^G||| \succeq 0.$$

Definition (Admissible convergence sequence)

Any sequence $\{\mathbf{F}_k\}_{k=0}^{n-1} \subset \mathbb{S}^+$ that satisfies

$$\mathbf{F}_0 \succeq \mathbf{F}_1 \succeq \cdots \succeq \mathbf{F}_{n-1} \succ 0$$

is called an admissible convergence sequence.

Note: One can construct non-trivial examples of inadmissible sequences where the individual column norms decrease monotonically

Prescribing convergence of blGMRES

Theorem (Kubínová and S. 2020)

Let $\{\mathbf{F}_k\}_{k=0}^{n-1} \subset \mathbb{S}^+$ *be an admissible convergence sequence. The following are equivalent:*

- *Residuals of blGMRES(*$\mathbf{A}$*,*$\mathbf{B}$*) satisfy* $|||\mathbf{R}_k^G||| = \mathbf{F}_k \,\forall\, k$
- *The* $\mathbf{A}$ *and* $\mathbf{B}$ *satisfy*

$$\mathbf{A} = \mathbf{W}\hat{\mathbf{R}}\hat{\mathbf{H}}\mathbf{W}^* \quad \textit{and} \quad \mathbf{B} = \mathbf{W}\mathbf{G},$$

where $\mathbf{W}$ *is unitary,* $\hat{\mathbf{R}} \in \mathbb{S}^{n\times n}$ *nonsing., upper block* Δ*,*

$$\hat{\mathbf{H}} = \begin{pmatrix} 0 & & & \langle\langle \mathbf{B}, \mathbf{W}_n \rangle\rangle^{-1} \\ I & \ddots & & -\langle\langle \mathbf{B}, \mathbf{W}_1 \rangle\rangle \langle\langle \mathbf{B}, \mathbf{W}_n \rangle\rangle^{-1} \\ & \ddots & 0 & \vdots \\ & & I & -\langle\langle \mathbf{B}, \mathbf{W}_{n-1} \rangle\rangle \langle\langle \mathbf{B}, \mathbf{W}_n \rangle\rangle^{-1} \end{pmatrix},$$

and the blocks of $\mathbf{G}$ *are* $\sqrt{\langle\langle \mathbf{F}_{k-1}, \mathbf{F}_{k-1} \rangle\rangle - \langle\langle \mathbf{F}_k, \mathbf{F}_k \rangle\rangle}$

All solvents are possible

Choosing $\hat{\mathbf{R}}$ as

$$\hat{\mathbf{R}} \equiv \hat{\mathbf{H}}^{-1}\mathbf{C}.$$

we can make $\mathbf{A}$ similar to any block companion matrix $\mathbf{C}$.

Lemma (Kubínová and S. 2020)

Assume that $\mathbf{A}$ is of the form $\mathbf{A} = \mathbf{W}\hat{\mathbf{R}}\hat{\mathbf{H}}\mathbf{W}^*$. Then, for any sequence $\mathbf{C}_0, \ldots, \mathbf{C}_n$, $\mathbf{C}_k \in \mathbb{S}$, $k = 0, \ldots, n-1$, $\mathbf{C}_0$ nonsingular, there exists $\hat{\mathbf{R}}$, such that $\mathbf{A}$ is similar to

$$\mathbf{C} = \begin{pmatrix} \mathbf{0} & & & \mathbf{C}_0 \\ \mathbf{I} & \ddots & & \mathbf{C}_1 \\ & \ddots & 0 & \vdots \\ & & \mathbf{I} & \mathbf{C}_{n-1} \end{pmatrix}.$$

Specifying solvents (i.e., "block eigenvalues")

- $\mathbf{C}$ is the block companion matrix to

$$\mathbf{M}(\lambda) = \mathbf{I}\lambda^n - \sum_{j=0}^{n-1} \mathbf{C}_k \lambda^k = \prod_{i=1}^{n} (\mathbf{I}\lambda - \mathbf{S}_k)$$

- "Block eigenvalues" $\mathbf{S}_k \in \mathbb{S}$ are called *solvents.*
- eigenvalues of the solvents are also the eigenvalues of $\mathbf{C}$
- Thus, eigenvalues of the solvents $\mathbf{S}_k \in \mathbb{S}$, $k = 1, \ldots, n$, are also the eigenvalues of $\mathbf{C}$
- Prescribing solvents is however stronger than prescribing just the scalar eigenvalues,
  - $\rightarrow$ since there are multiple block companion matrices similar to each other
  - $\rightarrow$ more right-hand sides reduces the predictive value of the eigenvalues

.

Specifying Ritz solvents

We can in addition specify the Ritz solvents $C_k^{(j)}$ (solvents of Hessenberg matrices at each step).

$$\text{Let} \quad \mathbf{U} = \begin{bmatrix} I & -C_0^{(1)} & -C_0^{(2)} & \cdots & -C_0^{(n-1)} \\ & I & -C_1^{(2)} & \cdots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & I & -C_{n-2}^{(n-1)} \\ & & & & I \end{bmatrix}^{-1}$$

and

$$\mathbf{D}_\Sigma = \operatorname{diag}(I, \Sigma_1, \Sigma_1\Sigma_2, \ldots, \prod_{k=1}^{n-1} \Sigma_k) \in \left(\mathbb{S}^+\right)^{n\times n}.$$

Then $\mathbf{A} = \mathbf{W}\mathbf{D}_\Sigma\mathbf{U}\mathbf{C}\mathbf{U}^{-1}\mathbf{D}_\Sigma^{-1}\mathbf{W}^*$ has the specified solvents, produces the specified Ritz solvents during block Arnoldi, and $\mathbf{W}\mathbf{E}_1 = \mathbf{V}_1$ should be our chosen starting vector (normalized)

Conclusions

We provided:

- an explicit peak-plateau relation for blFOM and blGMRES;
- an explicit characterization of admissible convergence behavior of blGMRES;

and showed that:

- any admissible convergence behavior is also attainable by blGMRES;
- arbitrary spectral properties of $\mathbf{A}$ can be enforced, while preserving the convergence behavior.

Conclusion: the *-algebra framework is the correct way to analyse block Krylov subspace method behavior.

Future work

- handling of linear dependence
  - $\rightarrow$ $\mathbf{V}_{j+1}$ is rank-deficient $\iff$ $|||\mathbf{V}_{j+1}|||$ is singular
  - $\rightarrow$ Zero-divisors complicate the analysis
- analysis of restarted block GMRES
- iterative methods for systems over *-algebras.
- analyze other block-level structural characteristics of matrices and matrix algorithms
  - $\rightarrow$ Understanding of "geometric" relationships of elements of the *-algebra as well as of vectors and systems built from them

References

Results in this talk are available in two papers

- **Kubínová and S.** *Prescribing convergence behavior of block Arnoldi and GMRES*, SIMAX, 2020
- **S.** *Stagnation of block GMRES and its relationship to block FOM*, ETNA, 2017

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

Thank you! Questions?

Bonus Slides!

What does this mean? **residual** convergence need not be connected to the eigenvalues. Meurant observed, however, that error convergence will still be connected to eigenvalues. **This result is an indication that we are perhaps not measuring residual in the correct norm**.

Residual norm versus error norm

What does this mean? **residual** convergence need not be connected to the eigenvalues. Meurant observed, however, that error convergence will still be connected to eigenvalues. **This result is an indication that we are perhaps not measuring residual in the correct norm**.

Geometry in $\mathbb{S}^n$

What is the geometric interpretation of the block vector?

- The span of the columns of $\mathbf{V} \in \mathbb{S}^n$ is generally an s-dimensional subspace.
- $\mathbf{V}$ represents a specific parallelotope[1] living in $\mathcal{R}(\mathbf{V})$.
- Compressed QR-factorization $\mathbf{V} = \mathbf{QR}$ decomposes $\mathbf{V}$ into its "orientation" $\mathbf{Q} \in \mathbb{S}^n$ and its "n orm" $\mathbf{R} \in \mathbb{S}_0^+$
- $\det \mathbf{R}$ is the volume of the parallelotope defined by $\mathbf{V}$

Theorem (Carson et al 2021)

Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbb{S}^n$ be full rank, with $\mathbf{X} = \mathbf{Y} + \mathbf{Z}$ and $\mathbf{Y} \perp \mathbf{Z}$. Then we have the block Pythagorean identity

$$|||\mathbf{X}|||^* |||\mathbf{X}||| = |||\mathbf{Y}|||^* |||\mathbf{Y}||| + |||\mathbf{Z}|||^* |||\mathbf{Z}|||.$$

[1] generalization of a parallelogram

Geometry in $\mathbb{S}^n$

What is the geometric interpretation of the block vector?

- The span of the columns of $\mathbf{V} \in \mathbb{S}^n$ is generally an s-dimensional subspace.
- $\mathbf{V}$ represents a specific parallelotope[1] living in $\mathcal{R}(\mathbf{V})$.
- Compressed QR-factorization $\mathbf{V} = \mathbf{Q}\mathbf{R}$ decomposes $\mathbf{V}$ into its "orientation" $\mathbf{Q} \in \mathbb{S}^n$ and its "n orm" $\mathbf{R} \in \mathbb{S}_0^+$
- $\det \mathbf{R}$ is the volume of the parallelotope defined by $\mathbf{V}$

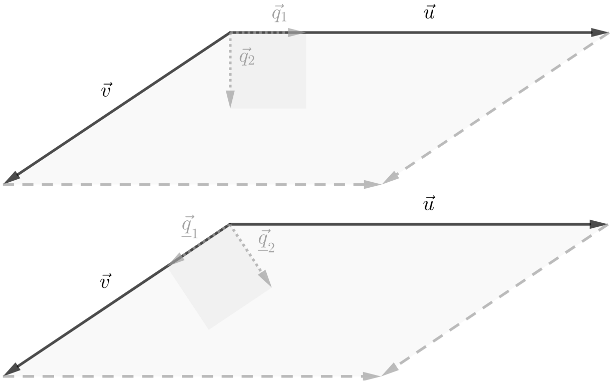
Theorem (Carson et al 2021)

Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbb{S}^n$ be full rank, with $\mathbf{X} = \mathbf{Y} + \mathbf{Z}$ and $\mathbf{Y} \perp \mathbf{Z}$. Then we have the block Pythagorean identity

$$|||\mathbf{X}|||^* |||\mathbf{X}||| = |||\mathbf{Y}|||^* |||\mathbf{Y}||| + |||\mathbf{Z}|||^* |||\mathbf{Z}|||.$$

[1] generalization of a parallelogram

Figure: The two-dimensional parallelogram formed by two vectors, $\mathbf{u}$ and $\mathbf{v}$ along with the normalized cube parallelotope induced by the QR-factorization. The ordering of the vectors changes the normalized orientation of the square parallelogram associated to the $\mathcal{Q}$ factor.



Non-admissible convergence behavior

Ordering of blGMRES residual norms:

- implies **monotonic convergence** of the size of the **individual residuals**
- takes into account the **relationship between the residuals**

Example (of non-admissible convergence behavior, $s = 2$)

- initial residuals of size one and almost linearly dependent:

$$\langle\langle \mathbf{R}_0, \mathbf{R}_0 \rangle\rangle \equiv \begin{pmatrix} 1 & 1-\epsilon \\ 1-\epsilon & 1 \end{pmatrix}, \quad \epsilon = 0.01,$$

- let first residual be decreased to ϵ and the second one to $1-\epsilon$:

$$\langle\langle \mathbf{R}_1, \mathbf{R}_1 \rangle\rangle \equiv \begin{pmatrix} \epsilon & p \\ p & 1-\epsilon \end{pmatrix}, \quad p \text{ unknown},$$

- there is no p such that:

$$\langle\langle \mathbf{R}_0, \mathbf{R}_0 \rangle\rangle \overset{\text{Löwner}}{\succeq} \langle\langle \mathbf{R}_1, \mathbf{R}_1 \rangle\rangle \overset{\text{Löwner}}{\succeq} 0.$$

Non-admissible convergence behavior

Ordering of blGMRES residual norms:

- implies **monotonic convergence** of the size of the **individual residuals**
- takes into account the **relationship between the residuals**

Example (of non-admissible convergence behavior, $s = 2$)

- initial residuals of size one and almost linearly dependent:

$$\langle\langle \mathbf{R}_0, \mathbf{R}_0 \rangle\rangle \equiv \begin{pmatrix} 1 & 1-\epsilon \\ 1-\epsilon & 1 \end{pmatrix}, \quad \epsilon = 0.01,$$

- let first residual be decreased to ϵ and the second one to $1-\epsilon$:

$$\langle\langle \mathbf{R}_1, \mathbf{R}_1 \rangle\rangle \equiv \begin{pmatrix} \epsilon & p \\ p & 1-\epsilon \end{pmatrix}, \quad p \text{ unknown},$$

- there is no p such that:

$$\langle\langle \mathbf{R}_0, \mathbf{R}_0 \rangle\rangle \overset{\text{Löwner}}{\succeq} \langle\langle \mathbf{R}_1, \mathbf{R}_1 \rangle\rangle \overset{\text{Löwner}}{\succeq} 0.$$