Stagnation of block GMRES and its relationship to block FOM

Kirk M. Soodhalter

Johann Radon Institute for Computational and Applied Mathematics
Transfer Group
Linz, Austria
http://math.soodhalter.com

October 27, 2016

Krylov and block Krylov subspaces

- Consider: $\mathbf{AX} = \mathbf{B} \in \mathbb{C}^{n \times L}$
- For L=1 with $\mathbf{x}_0 \in \mathbb{C}^n$ and $\mathbf{f}_0 = \mathbf{B} \mathbf{A}\mathbf{x}_0$, we have the **Krylov subspace**

$$\mathcal{K}_{j}(\mathbf{A},\mathbf{f}_{0})=\mathsf{span}\left\{\mathbf{f}_{0},\mathbf{A}\mathbf{f}_{0},\ldots,\mathbf{A}^{j-1}\mathbf{f}_{0}
ight\}$$

• For L > 1 let $\mathbf{X}_0 \in \mathbb{C}^{n \times L}$ and

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

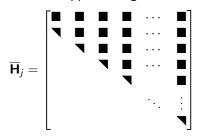
• Then we have the block Krylov subspace

$$\mathbb{K}_{j}(\mathbf{A},\mathbf{F}_{0}) = \mathcal{K}_{j}(\mathbf{A},\mathbf{f}_{0}^{(1)}) + \mathcal{K}_{j}(\mathbf{A},\mathbf{f}_{0}^{(2)}) + \cdots + \mathcal{K}_{j}(\mathbf{A},\mathbf{f}_{0}^{(L)}).$$

- L > 1 is a generalization of the case L = 1.
- dim $\mathbb{K}_i(\mathbf{A}, \mathbf{F}_0) = jL$ except when otherwise noted.

(Block) Krylov Subspace Methods

- Let $\mathbf{F}_0 = \mathbf{V}_1 \mathbf{S}_0$ be a skinny QR-factorization.
- The **block Arnoldi process** at step j produces $\mathbf{V}_{j+1} \in \mathbb{C}^{n \times L}$ with orthonormal columns
- $\mathbf{W}_j = \begin{bmatrix} \mathbf{V}_1, & \dots, & \mathbf{V}_j \end{bmatrix} \in \mathbb{C}^{n \times jL}$ has orthonormal columns spanning $\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)$
- Arnoldi relation: $\mathbf{AW}_j = \mathbf{W}_{j+1}\overline{\mathbf{H}}_j$, $\overline{\mathbf{H}}_j \in \mathbb{C}^{(j+1)L \times jL}$ is (block) upper Hessenberg
- ullet $\overline{\mathbf{H}}_j = (\mathbf{H}_{ik})_{ik} \in \mathbb{C}^{(j+1)L imes jL}$ is block upper Hessenberg
- For \blacksquare , $\P \in \mathbb{C}^{L \times L}$ and \P upper triangular



Restarted Block GMRES and Block FOM

Block GMRES and Block FOM valid for all $L \ge 1$

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

$$ullet$$
 Compute $\mathbf{Y}_m^{(G)} = \operatorname*{argmin}_{\mathbf{Y} \in \mathbb{C}^{mL imes L}} \left\| \overline{\mathbf{H}}_m \mathbf{Y} - \mathbf{E}_1^{(m+1} \mathbf{S}_0
ight\|_F^a$

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

- For block FOM
 - Compute $\mathbf{Y}_{m}^{(F)} = \mathbf{H}_{m}^{-1} \mathbf{E}_{1}^{[m]} \mathbf{S}_{0}^{b}$

• Set
$$X_m^{(F)} = X_0 + W_m Y_m^{(F)}$$
, $F_m^{(F)} = B - AX_m^{(F)}$

- Overwrite \mathbf{X}_0 and \mathbf{F}_0 with the *m*th approximation and residual.
- Repeat until convergence

 $^{{}^{}a}\mathbf{E}_{1}^{(m+1)} \in \mathbb{R}^{(m+1)L \times L}$ has appropriate columns of an identity matrix

 $^{{}^{}b}\mathbf{E}_{1}^{[m]} \in \mathbb{R}^{mL imes L}$ has appropriate columns of an identity matrix

When FOM breaks down (for all L)

If \mathbf{H}_j is singular, then the (block) FOM solution at iteration j does not exist.

The generalized (block) FOM approximation

$$\widetilde{\mathbf{X}}_{j}^{(F)} = \mathbf{X}_{0} + \widetilde{\mathbf{T}}_{j}^{(F)} \quad \text{where} \quad \widetilde{\mathbf{T}}_{j}^{(F)} = \mathbf{V}_{j}\widetilde{\mathbf{Y}}_{j}^{(F)} \quad \text{and} \quad \widetilde{\mathbf{Y}}_{j}^{(F)} = \mathbf{H}_{j}^{\dagger} \left(\mathbf{E}_{1}^{[m]} \mathbf{S}_{0} \right),$$

- Always exists
- $oldsymbol{\widetilde{X}}_{j}^{(F)} = oldsymbol{X}_{j}^{(F)}$ in the case (block) FOM does not breakdown
- Used in [Brown '91] to prove certain results.

The QR-factorizations of $\overline{\mathbf{H}}_j$ and \mathbf{H}_j for L=1

- Since they are upper Hessenberg, we use Givens rotations to annihilate lower subdiagonal.
- Let c_j and s_j be the Givens cosine and sine for column j of $\overline{\mathbf{H}}_j$.
- If \mathbf{H}_j is singular, then $c_j = 0$ and $s_j = 1$.
- The first j-1 rotations triangularize \mathbf{H}_{j} .
- Thus, $\mathbf{R}_j = \mathbf{Q}_j^T \mathbf{H}_j$ and $\overline{\mathbf{R}}_j = \overline{\mathbf{Q}}_j^T \overline{\mathbf{H}}_j$ have the same upper-left $(j-1) \times (j-1)$ submatrices.

Stagnation of GMRES and breakdown of FOM for L=1

Theorem in [Brown '91]

The matrix \mathbf{H}_j is singular (and thus $\mathbf{X}_j^{(F)}$ does not exist) if and only if GMRES stagnates at iteration j with $\mathbf{X}_j^{(G)} = \mathbf{X}_{j-1}^{(G)}$. Furthermore, in the case that \mathbf{H}_j is singular, we have $\widetilde{\mathbf{X}}_j^{(F)} = \mathbf{X}_j^{(G)}$.

Relationship of GMRES and FOM for L=1

Proposition shown in [Saad '03] based on results in [Cullum and Greenbaum '96]

We can write $\mathbf{X}_{j}^{(G)}$ as the following convex combination,

$$\mathbf{X}_{j}^{(G)} = c_{j}^{2} \mathbf{X}_{j}^{(F)} + s_{j}^{2} \mathbf{X}_{j-1}^{(G)}$$

where s_j and c_j are the *j*th Givens sine and cosine, respectively, as defined earlier.

- ullet Derived from relationship between ${f R}_j$ and $\overline{f R}_j$.
- We can replace $\mathbf{X}_{j}^{(F)}$ with $\widetilde{\mathbf{X}}_{j}^{(F)}$ to make this compatible with the FOM breakdown case.
- If FOM breaks down, then $c_j=0$, $s_j=1$, and we recover the stagnation result $\mathbf{X}_j^{(G)}=\mathbf{X}_{j-1}^{(G)}$

Selected previous work analyzing L = 1 and L > 1 cases

The relationship between GMRES and FOM for L=1

- 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]

Analysis of block Krylov methods L > 1

- [O'Leary '80] [Simoncini '96] [Simoncini and Gallopoulos '96],
 [Robbé and Sadkane '06], [Gutknecht and Schmelzer '09],
 [Elbouyahyaoui, Messaoudi, and Sadok '09]
- What we saw this morning: [Frommer, Lund-Nguyen, Szyld '16 (on website of D.B. Szyld in two weeks)]

Generalizing to the case L > 1

Questions:

- How to characterize of stagnation to the block GMRES?
 - Is $\mathbf{X}_{j}^{(G)} = \mathbf{X}_{j-1}^{(G)}$ a generalization of stagnation for L=1? **YES!**
 - If < L columns of $\mathbf{X}_{j}^{(G)}$ stagnate, is this the between-case to analyse? **NO!** only a special case
- How to generalize FOM breakdown to the block case?
 - Is rank $H_j = (j-1)L$ a generalization of stagnation for L = 1? **YES!**
 - What happens when rank $\mathbf{H}_j = (j-1)L + r$ for $0 < r \le L$?
- How to relate $\mathbf{X}_{i}^{(G)}$ and $\mathbf{X}_{i}^{(F)}$ using sines and cosines?

Key observation: we can write

$$\mathbf{X}_{j}^{(G)} = \mathbf{X}_{j-1}^{(G)} + \mathbf{S}_{j}^{(G)} \quad \text{and} \quad \mathbf{X}_{j}^{(F)} = \mathbf{X}_{j-1}^{(G)} + \mathbf{S}_{j}^{(F)};$$

the column spaces of $\mathbf{S}_{j}^{(G)}$ and $\mathbf{S}_{j}^{(F)}$ characterize everything.

FOM and GMRES as twin minimization problems

Let $\widetilde{\mathbf{X}}_j$ be the generalized block FOM approximation. We can write

$$\mathbf{X}_{j}^{(G)} = \mathbf{X}_{j-1}^{(G)} + \mathbf{S}_{j}^{(G)}$$
 and $\widetilde{\mathbf{X}}_{j}^{(F)} = \mathbf{X}_{j-1}^{(G)} + \widetilde{\mathbf{S}}_{j}^{(F)}$

with

$$\mathbf{S}_{j}^{(G)} = \mathbf{W}_{j} \mathbf{Y}_{\mathbf{S}_{j}}^{(G)} \quad \text{and} \quad \widetilde{\mathbf{S}}_{j}^{(F)} = \mathbf{W}_{j} \widetilde{\mathbf{Y}}_{\mathbf{S}_{j}}^{(F)},$$

both with columns in $\mathbb{K}_i(\mathbf{A}, \mathbf{F}_0)$.

Block FOM and GMRES minimizations, similar to [Brown '91]

$$\mathbf{Y}_{\mathbf{S}_{j}}^{(G)} = \underset{\mathbf{Y} \in \mathbb{C}^{n}}{\operatorname{argmin}} \left\| \begin{bmatrix} \mathbf{E}_{1}^{[jL]} \mathbf{S}_{0} - \overline{\mathbf{H}}_{j-1} \mathbf{Y}_{j-1}^{(G)} \\ \mathbf{0} \end{bmatrix} - \overline{\mathbf{H}}_{j} \mathbf{Y} \right\| \quad \text{and}$$

$$\widetilde{\mathbf{Y}}_{\mathbf{S}_{j}}^{(F)} = \underset{\mathbf{Y} \in \mathbb{C}^{n}}{\operatorname{argmin}} \left\| \mathbf{E}_{1}^{[jL]} \mathbf{S}_{0} - \overline{\mathbf{H}}_{j-1} \mathbf{Y}_{j-1}^{(G)} - \mathbf{H}_{j} \mathbf{Y} \right\|.$$

Triangularization of $\overline{\mathbf{H}}_j$ and \mathbf{H}_j

ullet We can progressively triangularize $\overline{f H}_j$ and $f H_j$:

- $\mathbf{R}_{j-1} \in \mathbb{C}^{(j-1)L \times (j-1)L}$ u.pper triangular from QR-factorization at step j-1
- $\mathbf{Q}_i^{(j+1)} \in \mathbb{C}^{(j+1)L \times (j+1)L}$ and $\mathbf{Q}_i^{(j)} \in \mathbb{C}^{jL \times jL}$ use the same transformation in different sized matrices.
- ullet ${f Q}_j^{(j+1)}$ and ${f Q}_j^{(j)}$ complete the triangularizations, respectively

Two core problems

Using the QR-factorizations, we reformulate the block GMRES and FOM subproblems.

$$\begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \mathbf{N}_j \end{bmatrix} \mathbf{Y}_j^{(G)} = \mathbf{G}_j^{(G)} = \begin{bmatrix} \mathbf{G}_{j-1}^{(G)} \\ \mathbf{C}_j \end{bmatrix}$$

and
$$\begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \widehat{\mathbf{N}}_j \end{bmatrix} \mathbf{Y}_j^{(F)} = \mathbf{G}_j^{(F)} = \begin{bmatrix} \mathbf{G}_{j-1}^{(G)} \\ \widehat{\mathbf{C}}_j \end{bmatrix},$$

with N_j and \hat{N}_j upper triangular, where we have that

$$\mathbf{G}_{j}^{(G)} = \left(\mathbf{Q}_{j}^{(j+1)} egin{bmatrix} \mathbf{G}_{j-1}^{(G)} \ \widetilde{\mathbf{C}}_{j} \ \mathbf{0} \end{bmatrix}
ight)_{1:jl} = \left(egin{bmatrix} \mathbf{G}_{j-1}^{(G)} \ \mathbf{C}_{j} \ * \end{bmatrix}
ight)_{1:jl}.$$

Behavior of Block GMRES and $\operatorname{rank}\ \mathbf{H}_{j}$

Theorem

The matrix \mathbf{H}_j is singular with rank $\mathbf{H}_j = (j-1)L + r$ with r < L if and only for the jth block GMRES update $\mathbf{S}_j^{(G)}$,

$$\dim\left(\mathcal{R}\left(\mathbf{S}_{j}^{(G)}\right)\cap\mathcal{R}\left(\mathbf{V}_{j}\right)\right)=r$$

In other words

There exists $\mathcal{Z} \subset \mathcal{R}(\mathbf{V}_j)$ with dim $\mathcal{Z} = L - r$ such that

$$egin{aligned} \mathcal{R}\left(\mathbf{S}_{j}^{(G)}
ight) oldsymbol{\perp} \mathcal{Z} \quad ext{and} \ \mathcal{R}\left(\mathbf{S}_{j}^{(G)}
ight)
ot \not\perp \left(\mathcal{R}(\mathbf{V}_{j}) \setminus \mathcal{Z}
ight). \end{aligned}$$

Full stagnation of Block GMRES and rank H_j

Corollary

The matrix \mathbf{H}_j is singular with rank $\mathbf{H}_j = (j-1)L$ if and only if block GMRES has totally stagnated with $\mathbf{X}_j^{(G)} = \mathbf{X}_{j-1}^{(G)}$.

Relationship of block GMRES and FOM Updates

Theorem

The span of the columns of $\widetilde{\mathbf{S}}_{j}^{(F)}$ has a non-trivial intersection with exactly an r-dimensional subspace of $\mathcal{R}(\mathbf{V}_{j})$ if and only if the same is true of $\mathbf{S}_{j}^{(G)}$. In this case, we also have rank $\mathbf{C}_{j} = r$.

The is the **correct third case** between no stagnation and total block stagnation for L > 1 that one should analyse.

Total stagnation of block GMRES leads to block FOM breakdown

Corollary

Block GMRES at iteration j totally stagnates with $\mathbf{X}_{j}^{(G)} = \mathbf{X}_{j-1}^{(G)}$ if and only if $\widetilde{\mathbf{X}}_{j}^{(F)} = \mathbf{X}_{j-1}^{(G)}$.

Characterization of partial stagnation

Definition

Partial stagnation refers to the case that certain columns of the GMRES iterate stagnate but $\mathbf{X}_{i}^{(G)} \neq \mathbf{X}_{i-1}^{(G)}$.

Theorem

Block GMRES suffers a partial stagnation at iteration j if and only if $0<\mathrm{rank}~\mathbf{C}_j\leq \tilde{r}$ where \tilde{r} is the number of stagnating columns such that if the ith column of $\mathbf{X}_j^{(G)}$ stagnates, then ith column of \mathbf{C}_j is the zero vector.

Block GMRES/FOM connection using CS-decomposition

Theorem

Let \mathbf{H}_i be nonsingular. Then we have,

$$\mathbf{X}_{j}^{(G)} = \mathbf{X}_{j}^{(F)} \left(\widehat{\mathbf{C}}_{j}^{-1} \mathcal{Q} \mathcal{C}^{2} \mathcal{Q}^{*} \widehat{\mathbf{C}}_{j} \right) + \mathbf{X}_{j-1}^{(G)} \left(\widehat{\mathbf{C}}_{j}^{-1} \mathcal{Q} \mathcal{S}^{2} \mathcal{Q}^{*} \widehat{\mathbf{C}}_{j} \right) \quad \textit{where}$$

- $\mathcal{Q} \in \mathbb{C}^{L \times L}$ is a unitary matrix
- ullet C and S are L imes L diagonal matrices with entries ≤ 1
- $C^2 + S^2 = I$ (sines and cosines)
- C, S come from CS-decomposition of non-identity block of

$$\mathbf{Q}_{j}^{(j+1)} = egin{bmatrix} \mathbf{I}_{(j-2)L} & \blacksquare & \blacksquare \ \blacksquare & \blacksquare \end{bmatrix}$$

In case of dependent Arnoldi vectors: stay calm

- Block Arnoldi vectors might not be linearly independent
- For stability, one can, e.g.:
 - reduce block size
 - replace dependent vector with random one
- Block size reduction
 - → less elegant analysis with non-square matrix blocks
 - → weaker versions of many results
 - → some results not valid in this setting
- Block size maintained with random vector replacement
 - → square block structure maintained
 - $\rightarrow \widehat{\mathbf{C}}_i, \widehat{\mathbf{C}}_i, \mathbf{C}_i$ become rank deficient
 - \rightarrow Only results depending on full rank of these matices effected.

Modified relationship between block GMRES and block FOM

Assumptions:

- p dependent basis vectors generated at iteration j.
- No single system has converged.
- No dependent vectors generated subsequently.

Theorem

With the above assumptions, at iteration j + k, if \mathbf{H}_{j+k} is nonsingular, then we have that

$$\widetilde{\mathbf{X}}_{j+k}^{(F)} - \mathbf{X}_{j+k}^{(G)} = \mathbf{W}_{j+k} \begin{bmatrix} \mathbf{R}_{j+k-1}^{-1} \mathbf{Z}_{j+k} \\ \mathbf{I} \end{bmatrix} \widehat{\mathbf{Y}}_{2}^{-1} \mathcal{Q} \mathcal{S}^{2} \mathcal{Q}^{*} \widehat{\mathbf{C}}_{j}.$$

Principal angles between $\mathcal{R}\left(\mathbf{F}_{j-1}^{(G)}\right)$ and $\mathbf{A}\mathbb{K}_{j}(\mathbf{A},\mathbf{F}_{0})$

$\mathsf{Theorem}$

The angles represented by the sines and cosines from the CS-decomposition are the principal angles between $\mathcal{R}\left(\mathbf{F}_{j-1}^{(G)}\right)$ and $\mathbf{A}\mathbb{K}_{j}(\mathbf{A},\mathbf{F}_{0})$.

- The smaller the angles are, the greater the decrease in residual norm from the minimum residual projection at this iteration.
- Using implementation strategy of [Gutknecht and Schmelzer '08] angles available at the cost of two L × L SVD computations per iteration.

Numerical example: partial stagnation of block GMRES

- We use the coordinate shift matrix \mathbf{A}_{st} but with n = 30.
- The right-hand sides are $\mathbf{e}_1^{[30]}$ and $\mathbf{e}_{25}^{[30]}$.
- At iteration 5, we have convergence for the first right-hand side.
- The code used performs a replacement of the dependent vector.
- ullet Near-stagnation of the other system until iteration j=15

Numerical example: partial stagnation of block GMRES

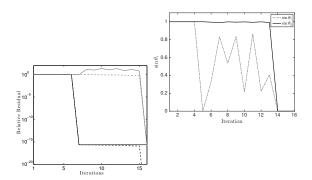


Figure: The **black solid** and **dashed** curves correspond respectively to the block FOM and GMRES residuals for the first right-hand side. Similarly, the **gray solid** and **dashed** curves, respectively, correspond to the second right-hand side.

Numerical example: fast and slow phases of block GMRES convergence

- Build a diagonal matrix with two blocks:
 - the sherman5 matrix1
 - the 200 \times 200 coordinate shift matrix \mathbf{A}_{st}
- Two right-hand sides chosen to produce perfect stagnation in the A_{st} block but convergence in the sherman4 block.
- subvectors of the right-hand sides for \mathbf{A}_{st} block were $\mathbf{e}_{50}^{[200]}$ and $\mathbf{e}_{150}^{[200]}$.
- subvectors of the right-hand sides sherman5 block were a vector packaged with sherman5 and a random vector scaled to have norm on the order of 10⁷.².

¹Univ. Florid Sparse Matrix Library

²to exaggerate convergence prior to stagnation

Numerical example: fast and slow phases of convergence

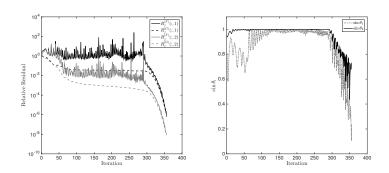


Figure : Left-hand figure: the 2-norm residual curves of block GMRES and FOM. Right-hand figure: squares of the sines $\left\{s_1^2,s_2^2\right\}$ from the CS-decomposition of the orthogonal transformation in the our analysis.

Conclusions and further work

Conclusions:

- $\mathbf{X}_{j}^{(F)}$ may not exist but $\widetilde{\mathbf{X}}_{j}^{(F)}$ is a possible improvement of $\mathbf{X}_{j-1}^{(G)}$.
- Block FOM breakdown and block GMRES "stagnation" are characterized by the intersection of the column spaces of $\mathbf{S}_{j}^{(G)}$ and $\widetilde{\mathbf{S}}_{j}^{(F)}$ with that of \mathbf{V}_{j} .
- "Convex combination" relationship between $\mathbf{X}_{j-1}^{(G)}$ and $\widetilde{\mathbf{X}}_{j}^{(F)}$.
- Connected QR-factorization of $\overline{\mathbf{H}}_j$ with principal angles between $\mathcal{R}\left(\mathbf{F}_{j-1}^{(G)}\right)$ and $\mathbf{A}\mathbb{K}_j(\mathbf{A},\mathbf{F}_0)$.

Future work:

- Extend to block-type methods for Sylvester equations (perhaps other matrix equations or matrix functions?).
- Use the angles for deflation/recycling schemes (see, e.g., [de Sturler '99])

Thank you!

To learn more (about me, my work, my interests,...), visit http://math.soodhalter.com.

- S. Stagnation of block GMRES and its relationship to block FOM. Submitted and available on arXiv (linked from my website).
- 2. S. Block Krylov subspace recycling for shifted systems with unrelated right-hand sides. SISC Volume 38 and available on arXiv (linked from my website).