Electronic Transactions on Numerical Analysis.
Volume 46, pp. 162–189, 2017.

ISSN 1068–9613.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

STAGNATION OF BLOCK GMRES AND ITS RELATIONSHIP TO BLOCK FOM*

KIRK M. SOODHALTER†

Abstract. We analyze the convergence behavior of block GMRES and characterize the phenomenon of stagnation which is then related to the behavior of the block FOM method. We generalize the block FOM method to generate well-defined approximations in the case that block FOM would normally break down, and these generalized solutions are used in our analysis. This behavior is also related to the principal angles between the column-space of the previous block GMRES residual and the current minimum residual constraint space. At iteration j, it is shown that the proper generalization of GMRES stagnation to the block setting relates to the column space of the jth block Arnoldi vector. Our analysis covers both the cases of normal iterations as well as block Arnoldi breakdown wherein dependent basis vectors are replaced with random ones. Numerical examples are given to illustrate what we have proven, including one built from a small application problem to demonstrate the validity of the analysis in a less pathological case.

Key words. block Krylov subspace methods, GMRES, FOM, stagnation

AMS subject classifications. 65F10, 65F50, 65F08

1. Introduction.

The Generalized Minimum Residual Method (GMRES) [35] and the Full Orthogonalization Method (FOM) [33] are two Krylov subspace methods for solving linear systems with non-Hermitian coefficient matrices and one right-hand side, i.e.,

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \text{with} \quad \mathbf{A} \in \mathbb{C}^{n\times n} \quad \text{and} \quad \mathbf{b} \in \mathbb{C}^{n}. \tag{1.1}$$

The convergence behavior of these two methods is closely related, and this relationship was characterized by Brown [5], other related results can be found in [7, 8, 42], and a related detailed geometric analysis of projection methods was presented in [11]. A nice description can also be found in [34, Section 6.5.5]. Krylov subspace methods have been generalized to treat the situation in which we have multiple right-hand sides, i.e., we are solving

$$\mathbf{A}\mathbf{X} = \mathbf{B} \quad \text{with} \quad \mathbf{B} \in \mathbb{C}^{n\times L}. \tag{1.2}$$

In particular, block GMRES and block FOM [34, Section 6.12] have been proposed for solving (1.2). However, to our knowledge, a similar full analysis of block GMRES, the connection between stagnation and block FOM convergence and accompanying geometric considerations have yet to be described in the literature. Therefore, in this work we analyze the stagnation behavior of block GMRES and characterize its relationship to the behavior of the block FOM method. Similar analytic tools as in [5] and [11] are used, but the behavior of block methods is a bit more complicated to describe. The key result is the proper generalization of GMRES stagnation to the block setting. The analog of stagnation for block GMRES is not simply stagnation of some columns of the iterate. Rather, at iteration j, it is associated to the dimension of the intersection between the column space of the jth block Arnoldi vector and the jth block GMRES correction. Stagnation of some columns of the iterate is shown to be a special case thereof. This then allows analogs of many of the results on stagnation of GMRES and the relationship between GMRES and FOM to be proven in the block setting. As block methods can suffer from partial or full stagnation of the iteration and breakdowns due to linear dependence of the block residual, additional analysis is needed to fully characterize the stagnation in these settings. Here we consider the case that dependent basis vectors are

*Received June 27, 2016, Accepted March 22, 2017. Published online on May 24, 2017. Recommended by H. Sadok.
†Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria (kirk.soodhalter@ricam.oeaw.ac.at), http://math.soodhalter.com.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

replaced with random ones (as in [3, 6, 30, 40]). One could similarly consider the case that dependent vectors are removed and the block size reduced; see, e.g., [1, 22, 32, 28].

The rest of this paper proceeds as follows. In the next section, we review Krylov subspace methods, focusing in particular on block GMRES and block FOM. We also review existing analysis relating GMRES- and FOM-like methods. The type of relationship illuminated in [5] has been extended to many other pairs of methods. In Section 3, we present our main results which characterize the relationship between block GMRES and block FOM. In Section 4, we construct numerical examples which demonstrate what has been revealed by our analysis. We offer some discussion and conclusions in Section 5.

In this paper, we adopt the convention that $\mathbf{I}$ is the identity matrix, where context determines the appropriate dimension. When needed, we specify the dimension $\mathbf{I}_J \in \mathbb{R}^{J\times J}$. Similarly, $\mathbf{0}$ denotes the matrix of zeros, with dimension determined by context. We denote $\mathbf{0}_J \in \mathbb{R}^{J\times J}$ to be a square matrix of zeros and $\mathbf{0}_{J_1\times J_2} \in \mathbb{R}^{J_1\times J_2}$ with $J_1 \neq J_2$ to be a rectangular matrix of zeros. Furthermore, exact arithmetic is assumed for all the analysis in this paper. Breakdown in the case of inexact arithmetic is considered in [32].

2. Background.

In this section, we review the basics about Krylov subspace methods and focus on the block version, designed to solve, e.g., (1.2). We describe everything in terms of block Krylov subspace methods, and discuss the simplifications in the case that the block size $L = 1$. We then review existing results relating the iterates of pairs of methods (many times derived from Galerkin and minimum residual projections, respectively), e.g., FOM and GMRES [5] and BiCG and QMR [15] as well as subsequent works which expand upon and offer additional perspective on these pair-wise relationships, e.g., [7, 8, 18, 31, 42].

2.1. Single-vector and block Krylov subspaces.

In the case that we are solving the system (1.2) with multiple right-hand sides (a block right-hand side), block Krylov subspace methods are an effective family of methods for generating high-quality approximate solutions to (1.2) at relatively low cost. Let $\mathbf{X}_0$ be an initial approximate solution to (1.2) with block initial residual $\mathbf{F}_0 = \mathbf{B} - \mathbf{A}\mathbf{X}_0$. We can define the jth block Krylov subspace as

$$\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0) = \mathrm{colspan}\begin{bmatrix}\mathbf{F}_0 & \mathbf{A}\mathbf{F}_0 & \mathbf{A}^2\mathbf{F}_0 & \dots & \mathbf{A}^{j-1}\mathbf{F}_0\end{bmatrix},$$

where the span of a collection of block vectors is understood to be the span of all their columns. When $L = 1$ ($\mathbf{B}$, $\mathbf{X}_0 \in \mathbb{C}^n$), this definition reduces to the single-vector Krylov subspace, denoted $\mathcal{K}_j(\mathbf{A}, \mathbf{F}_0)$. In the case $L > 1$, it is straightforward to show that

$$\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0) = \mathcal{K}_j(\mathbf{A}, \mathbf{F}_0(:,1)) + \mathcal{K}_j(\mathbf{A}, \mathbf{F}_0(:,2)) + \cdots + \mathcal{K}_j(\mathbf{A}, \mathbf{F}_0(:,L)),$$

where we use the MATLAB style indexing notation $\mathbf{F}(:,i)$ to denote the ith column of a matrix $\mathbf{F} \in \mathbb{C}^{I\times J}$ such that $J \geq i$; see, e.g., [20].

Let $\mathbf{W}_j = \begin{bmatrix}\mathbf{V}_1 & \mathbf{V}_2 & \dots\mathbf{V}_j\end{bmatrix} \in \mathbb{C}^{n\times jL}$ be the matrix with columns spanning $\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)$ with $\mathbf{V}_i \in \mathbb{C}^{n\times L}$ having orthonormal columns and $\mathbf{V}_i^*\mathbf{V}_j = \mathbf{0}$ for $i \neq j$. These orthonormal blocks can be generated one block at a time by an iterative orthogonalization process called the block Arnoldi process, which is a natural generalization of the Arnoldi process for the single-vector case. We have the block Arnoldi relation

$$\mathbf{A}\mathbf{W}_j = \mathbf{W}_{j+1}\overline{\mathbf{H}}_j^{(B)}, \tag{2.1}$$

where $\overline{\mathbf{H}}_j^{(B)} = (\mathbf{H}_{i,j}) \in \mathbb{C}^{(j+1)L\times jL}$ is block upper Hessenberg with $\mathbf{H}_{ij} \in \mathbb{C}^{L\times L}$ and $\mathbf{H}_{j+1,j}$ upper triangular.

We can derive block FOM and block GMRES methods through Galerkin and minimization constraints. We have for each column of the jth block residual the constraints

$$\mathbf{F}_j(:,i) \perp \mathbb{K}_j(\mathbf{A},\mathbf{F}_0) \quad \text{or} \tag{2.2}$$

$$\mathbf{F}_j(:,i) \perp \mathbf{A}\mathbb{K}_j(\mathbf{A},\mathbf{F}_0), \tag{2.3}$$

which lead to the block FOM and block GMRES methods, respectively. For both methods, approximations can be computed for all columns simultaneously. Let $\mathbf{X}_j^{(F)}$ and $\mathbf{X}_j^{(G)}$ denote the jth block FOM and block GMRES approximate solutions for (1.2). Furthermore, let $\mathbf{E}_L^{[I]} \in \mathbb{R}^{I\times L}$ have as columns the first L columns of the $I \times I$ identity matrix, and let $\mathbf{F}_0 = \mathbf{V}_1\mathbf{S}_0$ be the reduced QR factorization with $\mathbf{S}_0 \in \mathbb{C}^{L\times L}$ upper triangular. Using (2.1), block FOM can be derived from (2.2) which leads to the formulation

$$\mathbf{X}_j^{(F)} = \mathbf{X}_0 + \mathbf{T}_j^{(F)} \qquad \text{where} \quad \mathbf{T}_j^{(F)} = \mathbf{W}_j\mathbf{Y}_j^{(F)} \quad \text{and} \quad \mathbf{H}_j^{(B)}\mathbf{Y}_j^{(F)} = \mathbf{E}_L^{[jL]}\mathbf{S}_0,$$

where $\mathbf{H}_j^{(B)} \in \mathbb{C}^{jL\times jL}$ is defined as the matrix containing the first jL rows of $\overline{\mathbf{H}}_j^{(B)}$. Similarly for block GMRES, we can use (2.1), combined with (2.3) to yield a formulation

$$\begin{aligned} \mathbf{X}_j^{(G)} = \mathbf{X}_0 + \mathbf{T}_j^{(G)} \qquad & \text{where} \quad \mathbf{T}_j^{(G)} = \mathbf{W}_j\mathbf{Y}_j^{(G)} \\ & \text{and} \quad \mathbf{Y}_j^{(G)} = \operatorname*{argmin}_{\mathbf{Y}\in\mathbb{C}^{jL\times L}} \left\| \overline{\mathbf{H}}_j^{(B)}\mathbf{Y} - \mathbf{E}_L^{[(j+1)L]}\mathbf{S}_0 \right\|_F, \end{aligned} \tag{2.4}$$

where $\|\cdot\|_F$ is the Frobenius norm. Updates such as $\mathbf{T}_j^{(G)}$ and $\mathbf{T}_j^{(F)}$ are often called *corrections* and the subspaces from which they are drawn are called *correction subspaces*. There has been a great deal of research on the convergence properties of block methods such as block GMRES; see, e.g., [17, 23, 38].

In the case $L = 1$, block Krylov methods reduce to the well-described single-vector Krylov subspace methods; see, e.g., [34, Section 6.3] and [39]. In this case, we drop the superscript (B) and write $\overline{\mathbf{H}}_j := \overline{\mathbf{H}}_j^{(B)}$. The block Arnoldi method simplifies to a simpler Gram-Schmidt process in which the block entries $\mathbf{H}_{i,j}$ of $\overline{\mathbf{H}}_j$ reduce to scalars, now denoted with lower-case $h_{ij} \in \mathbb{C}$. Then using the scalar version of (2.1), single-vector FOM can be derived from (2.2) which leads to the formulation

$$\mathbf{x}_j^{(F)} = \mathbf{x}_0 + \mathbf{t}_j^{(F)} \qquad \text{where} \quad \mathbf{t}_j^{(F)} = \mathbf{V}_j\mathbf{y}_j^{(F)} \quad \text{and} \quad \mathbf{H}_j\mathbf{y}_j^{(F)} = \beta\mathbf{e}_1^{[j]},$$

where $\beta = \|\mathbf{F}_0\|$ is the 2-norm of the single-vector residual, and $\mathbf{e}_J^{[I]} \in \mathbb{C}^I$ is the Jth Cartesian basis vector in $\mathbb{C}^I$. Similarly for single-vector GMRES, we can use (2.1) combined with (2.3) to yield the formulation

$$\mathbf{x}_j^{(G)} = \mathbf{x}_0 + \mathbf{t}_j^{(G)} \qquad \text{where} \quad \mathbf{t}_j^{(G)} = \mathbf{V}_j\mathbf{y}_j^{(G)} \quad \text{and} \quad \mathbf{y}_j^{(G)} = \operatorname*{argmin}_{\mathbf{y}\in\mathbb{C}^j} \left\| \overline{\mathbf{H}}_j\mathbf{y} - \beta\mathbf{e}_1^{[j+1]} \right\|.$$

In the case $L = 1$, if at some iteration j we have $\mathcal{K}_{j-1}(\mathbf{A},\mathbf{F}_0) = \mathcal{K}_j(\mathbf{A},\mathbf{F}_0)$ (i.e., $\dim \mathcal{K}_j(\mathbf{A},\mathbf{F}_0) = j - 1 < j$), then we have reached an invariant subspace, and both GMRES and FOM will produce an exact solution at that iteration. In this case, $j - 1$ is called the *grade* of the pair $(\mathbf{A},\mathbf{F}_0)$, denoted $\nu(\mathbf{A},\mathbf{F}_0)$. This notion of grade has been extended to the case $L > 1$ [20]; however, the situation is a bit more complicated. It can occur that $\dim \mathbb{K}_j(\mathbf{A},\mathbf{F}_0) < jL$ without convergence for all right-hand sides (in other words, without having reached the block grade of $\mathbf{A}$ and $\mathbf{F}_0$, the iteration at which we reach an invariant subspace). It may be that we have convergence for some or no right-hand sides. In this case, dependent block Arnoldi vectors are generated and there must be some procedure in place to gracefully handle this situation for reasons of stability. The dependence of block Arnoldi vectors and methods for handling this dependence have been discussed extensively in the literature (see, e.g., [3, 14, 17, 20, 24, 28, 36, 40]), and general convergence analysis of block methods has been presented in, e.g., [12, 23, 38, 37]. In this paper, we consider only the case that dependent basis vectors are replaced with random vectors.

2.2. Relationships between pairs of projection methods. Pairs of methods such FOM and GMRES which are derived from a Galerkin and minimum residual projection, respectively, over the same space are closely related. The analysis of Brown [5] characterized this relationship in the case of FOM and GMRES when $L = 1$. We state here a theorem encapsulating the results relevant to this work. First, though, note that in FOM at iteration j, we must solve a linear system involving $\mathbf{H}_j$. Thus, if $\mathbf{H}_j$ is singular, the jth FOM iterate does not exist. We define $\widetilde{\mathbf{x}}_j^{(F)}$ to be the *generalized FOM approximation* through

$$\widetilde{\mathbf{x}}_j^{(F)} = \mathbf{x}_0 + \widetilde{\mathbf{t}}_j^{(F)} \quad \text{where} \quad \widetilde{\mathbf{t}}_j^{(F)} = \mathbf{V}_j \widetilde{\mathbf{y}}_j^{(F)} \quad \text{and} \quad \widetilde{\mathbf{y}}_j^{(F)} = \mathbf{H}_j^{\dagger} \left(\beta \mathbf{e}_1^{[j]} \right), \tag{2.5}$$

where $\mathbf{H}_j^{\dagger}$ is the Moore-Penrose pseudoinverse of $\mathbf{H}_j$. In the case that $\mathbf{H}_j$ is nonsingular, we have that $\widetilde{\mathbf{x}}_j^{(F)} = \mathbf{x}_j^{(F)}$, but $\widetilde{\mathbf{x}}_j^{(F)}$ is well-defined in the case that $\mathbf{x}_j^{(F)}$ does not exist. In this case $\widetilde{\mathbf{y}}_j^{(F)}$ minimizes $\left\| \mathbf{H}_j \mathbf{y} - \beta \mathbf{e}_1^{[j]} \right\|$ and has minimum norm of all possible minimizers. The following theorem combines two results proven by Brown in [5].

THEOREM 2.1. *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 $\widetilde{\mathbf{x}}_j^{(F)} = \mathbf{x}_j^{(G)}$.*[1]

Thus in the GMRES stagnation case, it is shown that the two methods are "equivalent", if we consider the generalized formulation of FOM. However, the relationship persists in the case that $\mathbf{H}_j$ is nonsingular as show in, e.g., [34]. In the same text, the following proposition is also shown.

PROPOSITION 2.2. *Let $\mathbf{x}_j^{(G)}$ and $\mathbf{x}_j^{(F)}$ be the the jth GMRES and FOM approximations to the solution of (1.1) over the correction subspace $\mathcal{K}_j(\mathbf{A}, \mathbf{F}_0)$. Then we can write $\mathbf{x}_j^{(G)}$ as the following convex combination,*

$$\mathbf{x}_j^{(G)} = c_j^2 \mathbf{x}_j^{(F)} + s_j^2 \mathbf{x}_{j-1}^{(G)}, \tag{2.6}$$

where s_j and c_j are the jth Givens sine and cosine, respectively, obtained from annihilating the entry $h_{j+1,j}$ while forming the QR factorization of $\overline{\mathbf{H}}_j$.

One proves this by studying the differences between the QR factorizations of the rectangular $\overline{\mathbf{H}}_j \in \mathbb{C}^{(j+1)\times j}$ and square $\mathbf{H}_j \in \mathbb{C}^{j\times j}$ generated by the single-vector Arnoldi process. The relation (2.6) reveals information about GMRES stagnation and its relationship to FOM. If $\mathbf{x}_j^{(G)} = \mathbf{x}_{j-1}^{(G)}$, then we have that $c_j = 0$ which implies that $\mathbf{H}_j$ is singular and $\mathbf{x}_j^{(F)}$ does not exist. In this case, (2.6) can be thought of as still valid, in the sense that $s_j = 1$, and (2.6) reduces to $\mathbf{x}_j^{(G)} = \mathbf{x}_{j-1}^{(G)}$ if we replace $\mathbf{x}_j^{(F)}$ with $\widetilde{\mathbf{x}}_j^{(F)}$.

This characterization of the relationship is not only important for understanding how these two methods behave at each iteration. They also reveal that FOM can suffer from stability issues when GMRES is close to stagnation as the matrix $\mathbf{H}_j$ is nearly singular

[1] Note that Brown in [5] did not use the expression "generalized FOM approximation". He calls it the least squares solution and proves it's equivalence to the stagnated $\mathbf{x}_j^{(G)}$.

(poorly conditioned) in this case. Whereas the residual curve of GMRES is monotonically nonincreasing, we see spikes in the FOM residual norm corresponding to periods of near stagnation in the GMRES method. These so-called "peaks" of residual norms of FOM and their relation to "plateaus" of the residual norms of GMRES have been previously studied; see, e.g., [7, 8, 42, 43]. Of particular interest is the observation by Walker [42] that the GMRES method can be seen as the result of a "residual smoothing" of the FOM residual. Similar observations extend to other pairings, such as QMR and BiCG, and more recently the relationship was studied more generally for such pairs when nonorthonormal bases were used [10]. It should also be noted that there has been work done characterizing GMRES stagnation in various circumstances [25, 26].

3. Main results. When $L > 1$, block GMRES and block FOM also fit into the framework of a Galerkin/minimization pairing. Thus, it is natural that stagnation of block GMRES and behavior of the block FOM algorithm would exhibit the same relationship, using a generalized block FOM iterate defined similar by (2.5). However, this interaction is more complicated for a block method. There are interactions between the different approximations to individual systems. As such, the generalization of stagnation to the block GMRES setting must be done correctly. We introduce two definitions.

DEFINITION 3.1. *At iteration j, we call the situation in which $\mathbf{X}_j^{(G)} = \mathbf{X}_{j-1}^{(G)}$* total stagnation. *We call the situation in which some columns of the block GMRES approximation have stagnated but not all columns* partial stagnation. *Let $\mathbb{I}$ denote an indexing set such that $\mathbb{I} \subsetneq \{1, 2, \ldots, L\}$, and let $\overline{\mathbb{I}} = \{1, 2, \ldots, L\} \setminus \mathbb{I}$. For $\mathbf{F} \in \mathbb{C}^{J\times L}$, let $\mathbf{F}(:, \mathbb{I}) \in \mathbb{C}^{J\times|\mathbb{I}|}$ have as columns those from $\mathbf{F}$ corresponding to indices in $\mathbb{I}$. Then partial stagnation refers to the situation in which we have*

$$\mathbf{X}_j^{(G)}(:, \mathbb{I}) = \mathbf{X}_{j-1}^{(G)}(:, \mathbb{I}) \quad \textit{but} \quad \mathbf{X}_j^{(G)}(:, i) \neq \mathbf{X}_{j-1}^{(G)}(:, i) \quad \textit{for each} \quad i \in \overline{\mathbb{I}}. \tag{3.1}$$

Total stagnation is analogous to stagnation of GMRES in the single-vector case, as characterized in [5], but partial stagnation has no single-vector analog. Both total and partial stagnation can occur for multiple reasons. Total block GMRES stagnation can occur when block GMRES has converged, i.e., $\mathbf{X}_j^{(G)} = \mathbf{X}$, implying (if j is the first iteration for which this occurs) from [20, Theorem 9], that we have that $j = \nu(\mathbf{A}, \mathbf{F}_0)$ and $\dim \mathbb{K}_{j+k}(\mathbf{A}, \mathbf{F}_0) = \dim \mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)$ for all $k > 0$. This case is trivial and will not be considered. If there is no breakdown of the block Arnoldi process (the rank of the block residual is L), then an occurrence of total stagnation is the block analog of single-vector GMRES stagnation. We prove in this case that Theorem 2.1 has a block analog; cf., Corollary 3.17 and Corollary 3.20.

Partial stagnation has no direct analog to the single-vector case. Partial stagnation can occur when for column i, the system is exactly solved with $\mathbf{X}_j^{(G)}(:, i) = \mathbf{X}(:, i)$. This implies that $\mathbf{F}_0(:, i) - \mathbf{A}\mathbf{W}_j\mathbf{Y}_j^{(G)}(:, i) = 0$, which implies that

$$\dim\left(\mathcal{R}(\mathbf{F}_0) \cap \mathbf{A}\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)\right) = 1$$

(see, e.g., [32]) and that a dependent Arnoldi vector has been produced. In this case, one can treat this with one of the referenced strategies; see, e.g., [28, 3, 2, 4, 14, 41].

This is a specific instance of block Arnoldi process breakdown. At iteration j, the process breaks down when the matrix $\begin{bmatrix}\mathbf{F}_0 & \mathbf{A}\mathbf{F}_0 & \cdots & \mathbf{A}^{j-1}\mathbf{F}_0\end{bmatrix}$ is rank deficient which is equivalent to saying $\dim\left(\mathcal{R}(\mathbf{X}) \cap \mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)\right) = \dim \mathcal{N}(\mathbf{R}_j) > 0$. In this case, $\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)$ contains a linear combination of the columns of $\mathbf{X}$ [27, 32]. It has also been observed [32] that a dependent Arnoldi vector can be generated without the convergence of any of the columns.

In the case that there has been no breakdown of the block Arnoldi process we show that partial stagnation is actually a special case of a more general situation in which a part of the

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Krylov subspace does not contribute to the GMRES minimization process and the dimension of this subspace corresponds to the dimension of the null space of the rank-deficient FOM matrix $\mathbf{H}_j^{(B)}$; cf., Theorem 3.16 and Theorem 3.19 below.

We derive a relationship for block GMRES and block FOM which is a generalization of (2.6) and is valid even in the case that $\mathbf{H}_j^{(B)}$ is singular. Thus, as in (2.5), we generalize the definition of the block FOM approximation to be compatible with a singular $\mathbf{H}_j^{(B)}$, i.e.,

$$\widetilde{\mathbf{X}}_j^{(F)} = \mathbf{X}_0 + \widetilde{\mathbf{T}}_j^{(F)} \quad \text{where} \quad \widetilde{\mathbf{T}}_j^{(F)} = \mathbf{W}_j \widetilde{\mathbf{Y}}_j^{(F)}$$

$$\text{and} \quad \widetilde{\mathbf{Y}}_j^{(F)} = \left(\mathbf{H}_j^{(B)}\right)^{\dagger} \left(\mathbf{E}_L^{[jL]} \mathbf{S}_0\right), \tag{3.2}$$

where $\left(\mathbf{H}_j^{(B)}\right)^{\dagger}$ is the Moore-Penrose pseudoinverse of $\mathbf{H}_j^{(B)}$. In the case that $\mathbf{H}_j^{(B)}$ is nonsingular, we have that $\widetilde{\mathbf{X}}_j^{(F)} = \mathbf{X}_j^{(F)}$, but $\widetilde{\mathbf{X}}_j^{(F)}$ is well-defined in the case that $\mathbf{X}_j^{(F)}$ does not exist. In this case $\widetilde{\mathbf{Y}}_j^{(F)}$ minimizes $\left\|\mathbf{H}_j^{(B)}\mathbf{Y} - \mathbf{E}_L^{[jL]}\mathbf{S}_0\right\|_F$ and has minimum norm of all possible minimizers. As in (2.5), this definition reduces to the standard formulation of the FOM approximation in the case that $\mathbf{H}_j^{(B)}$ is nonsingular. In the single-vector case, to prove [34, Lemma 6.1], expressions are derived for the inverses of upper-triangular matrices. We need to obtain similar identities here. However, we want our derivation to be compatible with the case that $\mathbf{H}_j^{(B)}$ is singular.

To characterize both types of stagnation requires us to follow the work in [5], generalizing to the block Krylov subspace case. We also need to generalize (2.6) to the block GMRES/FOM setting. This is quite useful in extending the work in [5] and also of general interest.

3.1. GMRES and FOM from a particular perspective. We discuss briefly the known results for the relationship of single-vector GMRES and (generalized) FOM. This discussion closely relates to the discussion and results on ascent directions in, e.g., [5]. It has been shown that at the jth iteration the approximations $\mathbf{x}_j^{(G)}$ and $\widetilde{\mathbf{x}}_j^{(F)}$ can both be related to the $(j-1)$st, with

$$\mathbf{x}_j^{(G)} = \mathbf{x}_{j-1}^{(G)} + \mathbf{s}_j^{(G)} \quad \text{and} \quad \widetilde{\mathbf{x}}_j^{(F)} = \mathbf{x}_{j-1}^{(G)} + \widetilde{\mathbf{s}}_j^{(F)}, \tag{3.3}$$

$$\text{where} \quad \mathbf{s}_j^{(G)} = \mathbf{V}_j \mathbf{y}_{\mathbf{s}_j}^{(G)} \in \mathcal{K}_j(\mathbf{A}, \mathbf{F}_0) \quad \text{and} \quad \widetilde{\mathbf{s}}_j^{(F)} = \mathbf{V}_j \widetilde{\mathbf{y}}_{\mathbf{s}_j}^{(F)} \in \mathcal{K}_j(\mathbf{A}, \mathbf{F}_0),$$

where $\widetilde{\mathbf{y}}_{\mathbf{s}_j}^{(F)}$ and $\mathbf{y}_{\mathbf{s}_j}^{(G)}$ are representations of the generalized FOM and GMRES *progressive* corrections from $\mathcal{K}_j(\mathbf{A}, \mathbf{F}_0)$. The next proposition follows directly.

PROPOSITION 3.2. *The GMRES and generalized FOM updates $\mathbf{y}_{\mathbf{s}_j}^{(G)}$ and $\widetilde{\mathbf{y}}_{\mathbf{s}_j}^{(F)}$. respectively. satisfy the minimizations*

$$\mathbf{y}_{\mathbf{s}_j}^{(G)} = \underset{\mathbf{y}\in\mathbb{C}^n}{\operatorname{argmin}} \left\| \begin{bmatrix} \beta\mathbf{e}_1^{[j]} - \overline{\mathbf{H}}_{j-1}\mathbf{y}_{j-1}^{(G)} \\ 0 \end{bmatrix} - \overline{\mathbf{H}}_j \mathbf{y} \right\| \quad \textit{and} \tag{3.4}$$

$$\widetilde{\mathbf{y}}_{\mathbf{s}_j}^{(F)} = \underset{\mathbf{y}\in\mathbb{C}^n}{\operatorname{argmin}} \left\| \beta\mathbf{e}_1^{[j]} - \overline{\mathbf{H}}_{j-1}\mathbf{y}_{j-1}^{(G)} - \mathbf{H}_j \mathbf{y} \right\|. \tag{3.5}$$

Proof. To prove (3.4), one simply inserts the expression for $\mathbf{x}_j^{(G)}$ from (3.3) into the residual and applies the GMRES Petrov-Galerkin condition (2.3). To prove (3.5), one begins similarly, by substituting the expression for $\mathbf{x}_j^{(F)}$ from (3.3) into the residual and applying the FOM Galerkin condition (2.2). In this case, if $\mathbf{H}_j$ is nonsingular, then this is equivalent to solving the linear system

$$\mathbf{H}_j \widetilde{\mathbf{y}}^{(F)}_{\mathbf{s}_j} = \beta \mathbf{e}^{[j]}_1 - \overline{\mathbf{H}}_{j-1}\mathbf{y}^{(G)}_{j-1}. \tag{3.6}$$

In the case that $\mathbf{H}_j$ is singular (the jth FOM approximation does not exist), we set

$$\widetilde{\mathbf{y}}^{(F)}_{\mathbf{s}_j} = \mathbf{H}^{\dagger}_j \left(\beta \mathbf{e}^{[j]}_1 - \overline{\mathbf{H}}_{j-1}\mathbf{y}^{(G)}_{j-1}\right). \tag{3.7}$$

In either case, we have that $\widetilde{\mathbf{y}}^{(F)}_{\mathbf{s}_j}$ is the minimizer of (3.5), yielding the result. ∎

The result on FOM is [5, Theorem 3.3] but stated differently. This formulation allows us to discuss GMRES and FOM at iteration j using the $(j-1)$st GMRES minimization. We see that the GMRES method least-squares problem simply grows by one dimension when we go from iteration $j-1$ to j. However, at iteration j, imposing the FOM Galerkin condition (2.2) is equivalent to an augmentation of the $(j-1)$st GMRES least squares matrix. This augmented matrix is square. If it is nonsingular, then the jth FOM approximation exists, and we solve the augmented system (3.6). If the augmented matrix is singular, then the generalized FOM approximation is computed by solving the least squares problem (3.7). In the case of single-vector GMRES and FOM, this is not necessary to characterize their relationship. However, in the case of block GMRES and block FOM, we can better discuss a generalization to the more complicated block Krylov subspace situation.

3.2. The QR factorization of the block upper Hessenberg matrices. We begin by describing the structure of the QR factorizations of the square and rectangular block Hessenberg matrices.

LEMMA 3.3. *Let $\overline{\mathbf{R}}_j \in \mathbb{C}^{(j+1)L\times jL}$ and $\widehat{\mathbf{R}}_j \in \mathbb{C}^{jL\times jL}$ be the R-factors of the respective QR factorizations of $\overline{\mathbf{H}}^{(B)}_j$ and $\mathbf{H}^{(B)}_j$, and let $\mathbf{R}_j$ be the $jL\times jL$ nonzero block of $\overline{\mathbf{R}}_j$. Furthermore, let $j \geq 2$. Then $\overline{\mathbf{R}}_j$ and $\widehat{\mathbf{R}}_j$ both have as their upper left $(j-1)L \times (j-1)L$ block the R-factor of the QR factorization of $\overline{\mathbf{H}}^{(B)}_{j-1}$, i.e., $\mathbf{R}_{j-1}$. Furthermore, the structures of $\overline{\mathbf{R}}_j$ and $\widehat{\mathbf{R}}_j$, respectively, are,*

$$\overline{\mathbf{R}}_j = \begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \mathbf{N}_j \\ & \end{bmatrix} \quad \textit{and} \quad \widehat{\mathbf{R}}_j = \begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \widehat{\mathbf{N}}_j \end{bmatrix}, \tag{3.8}$$

where $\mathbf{Z}_j \in \mathbb{C}^{(j-1)L\times L}$ and $\mathbf{N}_j, \widehat{\mathbf{N}}_j \in \mathbb{C}^{L\times L}$ are upper triangular.

Proof. Let $\mathbf{Q}^{(j+1)}_i \in \mathbb{C}^{(j+1)L\times(j+1)L}$ be the orthogonal transformation which annihilates all subdiagonal entries in columns $(i-1)L+1$ to iL of $\overline{\mathbf{H}}^{(B)}_j$ and effects no other rows so that we can write

$$\mathbf{Q}^{(j+1)}_{j-1}\cdots\mathbf{Q}^{(j+1)}_1 \overline{\mathbf{H}}^{(B)}_j = \begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \widehat{\mathbf{H}}_{j,j} \\ & \mathbf{H}_{j+1,j} \end{bmatrix} \quad \text{and} \quad \mathbf{Q}^{(j)}_{j-1}\cdots\mathbf{Q}^{(j)}_1 \mathbf{H}^{(B)}_j = \begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \widehat{\mathbf{H}}_{j,j} \end{bmatrix}.$$

Let $\widehat{\mathbf{Q}}^{(j)}_j \in \mathbb{C}^{jL\times jL}$ be the orthogonal transformation which annihilates the lower subdiagonal entries of the block $\widehat{\mathbf{H}}_{j,j}$ in $\mathbf{Q}^{(j)}_{j-1}\cdots\mathbf{Q}^{(j)}_1\mathbf{H}^{(B)}_j$ and effects no other rows. Then we have

$$\overline{\mathbf{R}}_j = \mathbf{Q}^{(j+1)}_j \begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \widehat{\mathbf{H}}_{j,j} \\ & \mathbf{H}_{j+1,j} \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{R}}_j = \widehat{\mathbf{Q}}^{(j)}_j \begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \widehat{\mathbf{H}}_{j,j} \end{bmatrix}, \tag{3.9}$$

and the Lemma is proven. ∎

Thus, the two core problems which must be solved at every iteration of block GMRES and block FOM can be written

$$\begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \mathbf{N}_j \end{bmatrix} \mathbf{Y}_j^{(G)} = (\mathbf{Q}_j^{(j+1)} \cdots \mathbf{Q}_1^{(j+1)} \mathbf{E}_L^{[(j+1)L]} \mathbf{S}_0)_{1:jL} \tag{3.10}$$

$$\text{and} \quad \begin{bmatrix} \mathbf{R}_{j-1} & \mathbf{Z}_j \\ & \widehat{\mathbf{N}}_j \end{bmatrix} \mathbf{Y}_j^{(F)} = \widehat{\mathbf{Q}}_j^{(j)} \mathbf{Q}_{j-1}^{(j)} \cdots \mathbf{Q}_1^{(j)} \mathbf{E}_L^{[jL]} \mathbf{S}_0. \tag{3.11}$$

It is also straightforward to show that the block right-hand sides of these core problems are related. If

$$\mathbf{G}_j^{(G)} = (\mathbf{Q}_j^{(j+1)} \cdots \mathbf{Q}_1^{(j+1)} \mathbf{E}_L^{[(j+1)L]} \mathbf{S}_0)_{1:jL} \quad \text{and} \quad \mathbf{G}_j^{(F)} = \widehat{\mathbf{Q}}_j^{(j)} \mathbf{Q}_{j-1}^{(j)} \cdots \mathbf{Q}_1^{(j)} \mathbf{E}_L^{[j)L]} \mathbf{S}_0,$$

then $\mathbf{G}_j^{(G)}$ and $\mathbf{G}_j^{(F)}$ are equal for the first $(j-1)L$ rows, with

$$\mathbf{G}_j^{(G)} = \begin{bmatrix} \mathbf{G}_{j-1}^{(G)} \\ \mathbf{C}_j \end{bmatrix} \quad \text{and} \quad \mathbf{G}_j^{(F)} = \begin{bmatrix} \mathbf{G}_{j-1}^{(G)} \\ \widehat{\mathbf{C}}_j \end{bmatrix}, \tag{3.12}$$

where we have that

$$\mathbf{G}_j^{(G)} = \left(\mathbf{Q}_j^{(j+1)} \begin{bmatrix} \mathbf{G}_{j-1}^{(G)} \\ \widetilde{\mathbf{C}}_j \\ \mathbf{0} \end{bmatrix} \right)_{1:jL} = \left(\begin{bmatrix} \mathbf{G}_{j-1}^{(G)} \\ \mathbf{C}_j \\ \widetilde{\mathbf{C}}_{j+1} \end{bmatrix} \right)_{1:jL}. \tag{3.13}$$

This is a consequence of the structure of the orthogonal transformations used to define these vectors. It is important to pause here for a moment to discuss the $L \times L$ matrices $\mathbf{C}_j$, $\widehat{\mathbf{C}}_j$, and $\widetilde{\mathbf{C}}_j$ and characterize if and when they are full rank. At times for convenience, we refer to these matrices as the "$\mathbf{C}$-matrices". This characterization can be related to the previous block GMRES residual, $\mathbf{F}_{j-1}^{(G)} = \mathbf{B} - \mathbf{A}\mathbf{X}_{j-1}^{(G)}$.

LEMMA 3.4. *We have that* $\operatorname{rank} \widetilde{\mathbf{C}}_j = \operatorname{rank} \mathbf{F}_{j-1}^{(G)}$*; and, in particular, if* $\dim \mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0) = (j-1)L$*, we have that,* $\widetilde{\mathbf{C}}_j$ *is nonsingular.*

Proof. Let $\mathbf{Y}_{j-1}^{(G)}$ be the solution to the block GMRES least squares subproblem (2.4) but for iteration $j-1$. Let

$$\mathbf{F}_{j-1}^{(G)} = \mathbf{B} - \mathbf{A}\mathbf{X}_{j-1}^{(G)} = -\mathbf{W}_j \left(\overline{\mathbf{H}}_{j-1}^{(B)} \mathbf{Y}_{j-1}^{(G)} - \mathbf{E}_L^{[jL]} \mathbf{S}_0 \right).$$

By assumption (3.10) has a solution at iteration $j-1$, and thus

$$\overline{\mathbf{H}}_{j-1}^{(B)} \mathbf{Y}_{j-1}^{(G)} - \mathbf{E}_L^{[jL]} \mathbf{S}_0 = \overline{\mathbf{Q}}_{j-1}^{*} \begin{bmatrix} \mathbf{G}_{j-2}^{(G)} \\ \mathbf{C}_{j-1} \\ \mathbf{0} \end{bmatrix} - \mathbf{E}_L^{[jL]} \mathbf{S}_0,$$

where $\overline{\mathbf{Q}}_{j-1} = \mathbf{Q}_{j-1}^{(j)} \cdots \mathbf{Q}_1^{(j)}$. Since $\mathbf{W}_j$ and $\overline{\mathbf{Q}}_{j-1}$ are both full rank, we have

$$\begin{aligned} \operatorname{rank} \mathbf{F}_{j-1}^{(G)} = \operatorname{rank} \overline{\mathbf{Q}}_{j-1} \mathbf{W}_j^{*} \mathbf{F}_{j-1}^{(G)} = \operatorname{rank} \left(\begin{bmatrix} \mathbf{G}_{j-2}^{(G)} \\ \mathbf{C}_{j-1} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{G}_{j-2}^{(G)} \\ \mathbf{C}_{j-1} \\ \widetilde{\mathbf{C}}_j \end{bmatrix} \right) &= \operatorname{rank} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ -\widetilde{\mathbf{C}}_j \end{bmatrix} \\ &= \operatorname{rank} \widetilde{\mathbf{C}}_j, \end{aligned}$$

where in the third equality in the first line, we use the fact that $\widetilde{\mathbf{C}}_j$ in the last L rows results from multiplication by $\mathbf{Q}_{j-1}^{(j)}$, as is shown in equation (3.13) (but there for iteration j rather than $j-1$). If we assume that the block Arnoldi method has not produced any dependent basis vectors, then we know from [32, Section 2, Corollary 1] that $\mathbf{F}_{j-1}^{(G)}$ is full-rank meaning $\widetilde{\mathbf{C}}_j$ is nonsingular. □

From this, we can similarly characterize the ranks of $\mathbf{C}_j$ and $\widehat{\mathbf{C}}_j$, which are closely related to $\widetilde{\mathbf{C}}_j$.

LEMMA 3.5. *We have that* $\operatorname{rank} \widehat{\mathbf{C}}_j = \operatorname{rank} \widetilde{\mathbf{C}}_j$. *In particular, if* $\dim \mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0) = (j-1)L$, *then we have that* $\widehat{\mathbf{C}}_j$ *is square and nonsingular.*

Proof. Let

$$\widehat{\mathbf{Q}}_j^{(j)} = \begin{bmatrix} \mathbf{I}_{(j-1)L} & \\ & \widehat{\mathbf{Q}}_j^{(b)} \end{bmatrix}$$

where $\widehat{\mathbf{Q}}_j^{(b)} \in \mathbb{C}^{L\times L}$ is the orthogonal transformation such that the second equation of (3.9) holds. Then from (3.13), we have $\widehat{\mathbf{C}}_j = \widehat{\mathbf{Q}}_j^{(b)} \widetilde{\mathbf{C}}_j$. If $\widetilde{\mathbf{C}}_j$ has full rank, the second statement follows. □

We can prove a similar result for $\mathbf{C}_j$, which will be used later to verify the nonsingularity of $\mathbf{C}_j$ under certain conditions.

LEMMA 3.6. *Let*

$$\mathbf{Q}_j^{(j+1)} = \begin{bmatrix} \mathbf{I}_{(j-1)L} & & \\ & \mathbf{Q}_j^{(11)} & \mathbf{Q}_j^{(11)} \\ & \mathbf{Q}_j^{(21)} & \mathbf{Q}_j^{(22)} \end{bmatrix}$$

with $\mathbf{Q}_j^{(11)} \in \mathbb{C}^{L\times L}$, $\mathbf{Q}_j^{(12)} \in \mathbb{C}^{L\times L}$, $\mathbf{Q}_j^{(21)} \in \mathbb{C}^{L\times L}$, *and* $\mathbf{Q}_j^{(22)} \in \mathbb{C}^{L\times L}$. *In general, we have rank* $\mathbf{C}_j \leq \min\left\{\operatorname{rank} \mathbf{Q}_j^{(11)}, \operatorname{rank} \widetilde{\mathbf{C}}_j\right\}$. *If* $\dim \mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0) = (j-1)L$, *then we have* $\mathbf{C}_j$ *is singular if and only if* $\mathbf{Q}_j^{(11)}$ *is singular.*

Proof. From (3.13) we have that $\mathbf{C}_j = \mathbf{Q}_j^{(11)} \widetilde{\mathbf{C}}_j$. The general result comes from basic inequality results for ranks of products of matrices; see, e.g., [21, Chapter 0]. If we assume $\dim \mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0) = (j-1)L$, then we know that $\widetilde{\mathbf{C}}_j$ has full rank, and the second result (in both directions) follows. □

We see that the ranks of $\widetilde{\mathbf{C}}_j$ and $\widehat{\mathbf{C}}_j$ are directly connected to block Arnoldi breakdown at iteration $j-1$. Later in Section 3.3, we assume no breakdown, thus both $\widetilde{\mathbf{C}}_j$ and $\widehat{\mathbf{C}}_j$ are nonsingular. In Section 3.4, we assume that the block Arnoldi process produces dependent vectors at iteration j which are replaced with random vectors. Thus, at iteration j, both $\widetilde{\mathbf{C}}_j$ and $\widehat{\mathbf{C}}_j$ are still nonsingular, and their dimensions do not change at subsequent iterations.

We now turn to solving (3.11) and either solving (3.10) or obtaining the generalized least squares solution if $\widehat{\mathbf{R}}_j$ is singular. Since $\mathbf{R}_j$ is nonsingular, we simply compute the actual inverse while for $\widehat{\mathbf{R}}_j$, we compute the pseudo-inverse. These are both straightforward generalizations of the identities used in the proof of [34, Lemma 6.1], though verifying the structure of the Moore-Penrose pseudo-inverse identity requires a bit of thought. Let us recall briefly the following definition which can be found in, e.g., [13, Section 2.2],

DEFINITION 3.7. *Let* $T : \mathcal{X} \to \mathcal{Y}$ *be a bounded linear operator between Hilbert spaces. Let* $\mathcal{N}(T)$ *denote the null space and* $\mathcal{R}(T)$ *denote the range of* T *and define* $\widetilde{T} : \mathcal{N}(T)^\perp \to \mathcal{R}(T)$ *to be the invertible operator such that* $\widetilde{T}x = Tx$ *for all* $x \in \mathcal{N}(T)^\perp$. *Then we call the operator* $T^\dagger$ *the Moore-Penrose pseudo-inverse if it is the unique operator satisfying*

ETNA
Kent State University and
Johann Radon Institute (RICAM)

1. $T^\dagger\big|_{\mathcal{R}(T)} = \widetilde{T}^{-1}$,
2. $T^\dagger\big|_{\mathcal{R}(T)^\perp} = 0_{op}$,

where 0_{op} *is the zero operator.*

This definition is more general than the matrix-specific definition given in, e.g., [16, Section 5.5.2]. We choose to follow Definition 3.7 as it renders the proof of the following lemma less dependent on many lines of block matrix calculations, but of course the theoretical results would be the same.

LEMMA 3.8. *The inverse and pseudo-inverse, respectively, of* $\mathbf{R}_j$ *and* $\widehat{\mathbf{R}}_j$ *can be directly constructed from the identities* (3.8)*, i.e.,*

$$\mathbf{R}_j^{-1} = \begin{bmatrix} \mathbf{R}_{j-1}^{-1} & -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{N}_j^{-1} \\ & \mathbf{N}_j^{-1} \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{R}}_j^\dagger = \begin{bmatrix} \mathbf{R}_{j-1}^{-1} & -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\widehat{\mathbf{N}}_j^\dagger \\ & \widehat{\mathbf{N}}_j^\dagger \end{bmatrix}, \tag{3.14}$$

where $\widehat{\mathbf{N}}_j^\dagger$ *is the Moore-Penrose pseudo-inverse of* $\widehat{\mathbf{N}}_j$.

Proof. The expression for $\mathbf{R}_j^{-1}$ can be directly verified by left and right multiplication. To verify the expression for $\widehat{\mathbf{R}}_j^\dagger$, we must verify the two conditions listed in Definition 3.7.

To verify condition 1, we first construct a basis for $\mathcal{N}(\widehat{\mathbf{R}}_j)^\perp$. Observe that under our assumption that $\mathbf{R}_{j-1}$ is nonsingular, we have that

$$\dim \mathcal{N}(\widehat{\mathbf{R}}_j) = \dim \mathcal{N}(\widehat{\mathbf{N}}_j) = L - r,$$

where $r = \operatorname{rank}(\widehat{\mathbf{N}}_j)$. Let $\{\mathbf{y}_i\}_{i=1}^r$ be a basis for $\mathcal{N}(\widehat{\mathbf{N}}_j)^\perp$. Furthermore, let $\{\mathbf{m}_i\}_{i=1}^{(j-1)L}$ be a basis for $\mathbb{R}^{(j-1)L}$. Then it follows that

$$\left\{ \begin{bmatrix} \mathbf{R}_{j-1}^{-1}\mathbf{m}_1 \\ \mathbf{0} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{R}_{j-1}^{-1}\mathbf{m}_{j-1} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{y}_1 \\ \mathbf{y}_1 \end{bmatrix}, \ldots, \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{y}_r \\ \mathbf{y}_r \end{bmatrix} \right\}$$

is a basis for $\mathcal{N}(\widehat{\mathbf{R}}_j)^\perp$. For any $\widehat{\mathbf{x}} \in \mathcal{N}(\widehat{\mathbf{R}}_j)^\perp$, we can write

$$\widehat{\mathbf{x}} = \sum_{i=1}^{(j-1)L} \alpha_i \begin{bmatrix} \mathbf{R}_{j-1}^{-1}\mathbf{m}_i \\ \mathbf{0} \end{bmatrix} + \sum_{i=1}^{r} \beta_i \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{y}_i \\ \mathbf{y}_i \end{bmatrix}.$$

By direct calculation, we see that

$$\widehat{\mathbf{R}}_j\widehat{\mathbf{x}} = \sum_{i=1}^{(j-1)L} \alpha_i \begin{bmatrix} \mathbf{m}_i \\ \mathbf{0} \end{bmatrix} + \sum_{i=1}^{r} \beta_i \begin{bmatrix} \mathbf{0} \\ \widehat{\mathbf{N}}_j\mathbf{y}_j \end{bmatrix},$$

and applying our prospective pseudo-inverse yields

$$\widehat{\mathbf{R}}_j^\dagger\widehat{\mathbf{R}}_j\widehat{\mathbf{x}} = \sum_{i=1}^{(j-1)L} \alpha_i \begin{bmatrix} \mathbf{R}_{j-1}^{-1}\mathbf{m}_i \\ \mathbf{0} \end{bmatrix} + \sum_{i=1}^{r} \beta_i \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\widehat{\mathbf{N}}_j^\dagger\widehat{\mathbf{N}}_j\mathbf{y}_i \\ \widehat{\mathbf{N}}_j^\dagger\widehat{\mathbf{N}}_j\mathbf{y}_i \end{bmatrix}.$$

Finally, we observe that since $\{\mathbf{y}_i\}_{i=1}^r$ is a basis for $\mathcal{N}(\widehat{\mathbf{N}}_j)^\perp$, we have from Definition 3.7 that $\widehat{\mathbf{N}}_j^\dagger\widehat{\mathbf{N}}_j\mathbf{y}_i = \mathbf{y}_i$ for all i, and thus $\widehat{\mathbf{R}}_j^\dagger\widehat{\mathbf{R}}_j\widehat{\mathbf{x}} = \widehat{\mathbf{x}}$, verifying condition 1.

To verify condition 2, we first observe that

$$\left\{ \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{0} \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{m}_{(j-1)L} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \widehat{\mathbf{N}}_j\mathbf{y}_1 \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{0} \\ \widehat{\mathbf{N}}_j\mathbf{y}_r \end{bmatrix} \right\}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

is a basis for $\mathcal{R}(\widehat{\mathbf{R}}_j)$. Let $\{\mathbf{c}_i\}_{i=1}^{L-r}$ be a basis for $\mathcal{R}(\widehat{\mathbf{N}}_j)^\perp$. Then it follows that $\left\{\begin{bmatrix}\mathbf{0}\\ \mathbf{c}_i\end{bmatrix}\right\}_{i=1}^{L-r}$ is a basis for $\mathcal{R}(\widehat{\mathbf{R}}_j)^\perp$. Let $\tilde{\mathbf{y}} = \sum_{i=1}^{L-r}\gamma_i \begin{bmatrix}\mathbf{0}\\ \mathbf{c}_i\end{bmatrix}$ be an element of $\mathcal{R}(\widehat{\mathbf{R}}_j)^\perp$. Then we have

$$\widehat{\mathbf{R}}_j^\dagger \tilde{\mathbf{y}} = \sum_{i=1}^{L-r}\gamma_i \begin{bmatrix}-\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\widehat{\mathbf{N}}_j^\dagger \mathbf{c}_i\\ \widehat{\mathbf{N}}_j^\dagger \mathbf{c}_i\end{bmatrix}.$$

It follows directly from the definition in (3.7) that $\widehat{\mathbf{N}}_j^\dagger \mathbf{c}_i = 0$ for all i, and this proves condition 2, thus proving the the lemma. ∎

The following corollary technically follows from Lemma 3.8, though it can easily be proven directly.

COROLLARY 3.9. *If $\mathbf{H}_j^{(B)}$ is nonsingular, then it follows that $\widehat{\mathbf{R}}_j^{-1}$ exists and can be written*

$$\widehat{\mathbf{R}}_j^{-1} = \begin{bmatrix}\mathbf{R}_{j-1}^{-1} & -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\widehat{\mathbf{N}}_j^{-1}\\ & \widehat{\mathbf{N}}_j^{-1}\end{bmatrix}.$$

Now we have all the pieces we need to analyze the relationship between the block GMRES and block FOM approximations, and we can then discuss the implications with respect to stagnation.

3.3. The case of a breakdown-free block Arnoldi process. We begin this section by discussing block GMRES and block FOM from the same perspective as advocated in Section 3.1. We have the block analog of Proposition 3.2, and in this case we explicitly construct the block analogs of $\mathbf{s}_j^{(G)}$ and $\mathbf{s}_j^{(F)}$.

LEMMA 3.10. *Let $\mathbf{S}_j^{(G)} = \mathbf{W}_j\mathbf{Y}_{\mathbf{S}_j}^{(G)}$ and $\widetilde{\mathbf{S}}_j^{(F)} = \mathbf{W}_j\widetilde{\mathbf{Y}}_{\mathbf{S}_j}^{(F)}$ both be in $\mathbb{C}^{n\times L}$ such that they satisfy the block GMRES and FOM progressive update formulas*

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

Then we can write

$$\mathbf{Y}_{\mathbf{S}_j}^{(G)} = \begin{bmatrix}-\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{N}_j^{-1}\mathbf{C}_j\\ \mathbf{N}_j^{-1}\mathbf{C}_j\end{bmatrix} \qquad \text{and} \qquad \widetilde{\mathbf{Y}}_{\mathbf{S}_j}^{(F)} = \begin{bmatrix}-\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\widehat{\mathbf{N}}_j^\dagger\widehat{\mathbf{C}}_j\\ \widehat{\mathbf{N}}_j^\dagger\widehat{\mathbf{C}}_j\end{bmatrix}, \tag{3.15}$$

and these vectors minimize the two residual update equations

$$\mathbf{Y}_{\mathbf{S}_j}^{(G)} = \underset{\mathbf{Y}\in\mathbb{C}^{jL\times L}}{\operatorname{argmin}} \left\|\begin{bmatrix}\mathbf{E}_1^{[jL]}\mathbf{S}_0 - \overline{\mathbf{H}}_{j-1}^{(B)}\mathbf{Y}_{j-1}^{(G)}\\ \mathbf{0}\end{bmatrix} - \overline{\mathbf{H}}_j^{(B)}\mathbf{Y}\right\| \qquad \text{and} \tag{3.16}$$

$$\widetilde{\mathbf{Y}}_{\mathbf{S}_j}^{(F)} = \underset{\mathbf{Y}\in\mathbb{C}^{jL\times L}}{\operatorname{argmin}} \left\|\mathbf{E}_1^{[jL]}\mathbf{S}_0 - \overline{\mathbf{H}}_{j-1}^{(B)}\mathbf{Y}_{j-1}^{(G)} - \mathbf{H}_j^{(B)}\mathbf{Y}\right\|. \tag{3.17}$$

Proof. Combining (3.12) and (3.14) to solve (3.10) and (3.2), we have the following expressions for $\mathbf{Y}_j^{(G)}$ and $\widetilde{\mathbf{Y}}_j^{(F)}$,

$$\mathbf{Y}_j^{(G)} = \begin{bmatrix}\mathbf{R}_{j-1}^{-1}\mathbf{G}_{j-1}^{(G)} - \mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{N}_j^{-1}\mathbf{C}_j\\ \mathbf{N}_j^{-1}\mathbf{C}_j\end{bmatrix} \quad \text{and} \quad \widetilde{\mathbf{Y}}_j^{(F)} = \begin{bmatrix}\mathbf{R}_{j-1}^{-1}\mathbf{G}_{j-1}^{(G)} - \mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\widehat{\mathbf{N}}_j^\dagger\widehat{\mathbf{C}}_j\\ \widehat{\mathbf{N}}_j^\dagger\widehat{\mathbf{C}}_j\end{bmatrix}.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

As it can be appreciated, $\mathbf{R}_{j-1}^{-1}\mathbf{G}_{j-1}^{(G)} = \mathbf{Y}_{j-1}^{(G)}$, and it follows that

$$\mathbf{Y}_j^{(G)} = \begin{bmatrix} \mathbf{Y}_{j-1}^{(G)} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{N}_j^{-1}\mathbf{C}_j \\ \mathbf{N}_j^{-1}\mathbf{C}_j \end{bmatrix} \quad \text{and} \quad \widetilde{\mathbf{Y}}_j^{(F)} = \begin{bmatrix} \mathbf{Y}_{j-1}^{(G)} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\widehat{\mathbf{N}}_j^{\dagger}\widehat{\mathbf{C}}_j \\ \widehat{\mathbf{N}}_j^{\dagger}\widehat{\mathbf{C}}_j \end{bmatrix},$$

which yields (3.15). The proof that these vectors are the minimizers of (3.16) and (3.17) proceeds exactly as in that of Proposition 3.2. □

The behavior of block FOM and GMRES can be divided into three cases.

Case 1 If $\mathbf{H}_j^{(B)}$ is nonsingular (i.e., the block FOM solution exists), then (3.17) is satisfied exactly, and by augmenting with L columns to expand $\overline{\mathbf{H}}_{j-1}^{(B)}$, to $\mathbf{H}_j^{(B)}$, the $(j-1)$st GMRES least squares problem becomes a nonsingular linear system.

Case 2 If $\mathbf{H}_j^{(B)}$ is singular with rank $(j-1)L + r$ with $1 \leq r < L$, then the linear system produced by the augmentation of $\overline{\mathbf{H}}_{j-1}^{(B)}$ produces a better minimizer than $\mathbf{X}_{j-1}^{(G)}$ from (3.17), but it is not exactly solvable. This corresponds to only an r-dimensional subspace of $\mathcal{R}(\mathbf{V}_j)$ contributing to the block GMRES minimization at iteration j.

Case 3 If $\mathbf{H}_j^{(B)}$ is singular with rank $(j-1)L$, then the situation is analogous to that described in Theorem 2.1. We have $\widetilde{\mathbf{X}}_j^{(F)} = \mathbf{X}_j^{G} = \mathbf{X}_{j-1}^{(G)}$, and augmentation of $\overline{\mathbf{H}}_{j-1}^{(B)}$ produces no improvement.

We note that Case 2 is unique to the block setting and represents a block generalization of the concept of GMRES stagnation, where only an r-dimensional subspace of $\mathcal{R}(\mathbf{V}_j)$ (with $r < \operatorname{rank} \mathbf{V}_j = L$) contributes to the minimization of the residual at step j. We direct the reader to the related discussion in [5] about ascent directions, though we omit here such an analysis in the interest of manuscript length. Before proving these results, we prove some intermediate technical results.

Let us begin by discussing the structure of $\mathbf{Q}_j^{(j+1)}$. In this case, as discussed in Lemma 3.6, this matrix has a large $(j-1)L \times (j-1)L$ identity matrix in the upper left-hand corner, and a $2L \times 2L$ nontrivial orthogonal transformation block in the lower right-hand corner, denoted

$$\widehat{\boldsymbol{\mathcal{H}}}_j = \begin{bmatrix} \mathbf{Q}_j^{(11)} & \mathbf{Q}_j^{(12)} \\ \mathbf{Q}_j^{(21)} & \mathbf{Q}_j^{(22)} \end{bmatrix}. \tag{3.18}$$

We note that $\widehat{\boldsymbol{\mathcal{H}}}_j$ is itself a product of elementary orthogonal transformations, and all four blocks are of size $L \times L$. Because $\widehat{\boldsymbol{\mathcal{H}}}_j$ is an orthogonal transformation, it admits a CS-decomposition (see, e.g., [16, Theorem 2.5.3] and more generally for complex matrices [29] and references therein) i.e., there exist unitary matrices $\boldsymbol{\mathcal{U}}_1, \boldsymbol{\mathcal{U}}_2, \boldsymbol{\mathcal{V}}_1, \boldsymbol{\mathcal{V}}_2 \in \mathbb{C}^{L\times L}$ and diagonal matrices $\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{C}} \in \mathbb{R}^{L\times L}$ with $\boldsymbol{\mathcal{S}} = \operatorname{diag}\{s_1, \ldots, s_L\}$ and $\boldsymbol{\mathcal{C}} = \operatorname{diag}\{c_1, \ldots, c_L\}$ such that

$$\begin{aligned} \mathbf{Q}_j^{(11)} &= \boldsymbol{\mathcal{U}}_1\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{V}}_1^*,\ \mathbf{Q}_j^{(12)} = \boldsymbol{\mathcal{U}}_1\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{V}}_2^*,\ \mathbf{Q}_j^{(21)} = \boldsymbol{\mathcal{U}}_2\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{V}}_1^*, \\ &\text{and} \qquad \mathbf{Q}_j^{(22)} = -\boldsymbol{\mathcal{U}}_2\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{V}}_2^*, \end{aligned} \tag{3.19}$$

and for $1 \leq i \leq L$ we have $s_i^2 + c_i^2 = 1$, i.e., the diagonal entries of $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{C}}$ are the sines and cosines of L angles, $\{\theta_1, \ldots, \theta_L\}$. We assume that $c_1 \leq c_2 \leq \cdots \leq c_L$, and it then follows that $s_1 \geq s_2 \geq \cdots \geq s_L$. Note that in the case of the single-vector Krylov methods, $\widehat{\boldsymbol{\mathcal{H}}}_j \in \mathbb{C}^{2\times 2}$, $\boldsymbol{\mathcal{U}}_1 = \boldsymbol{\mathcal{U}}_2 = \boldsymbol{\mathcal{V}}_1 = \boldsymbol{\mathcal{V}}_2 = 1$, and $\boldsymbol{\mathcal{S}} = s_1$ and $\boldsymbol{\mathcal{C}} = c_1$ are the Givens sine and cosine. Thus this CS-decomposition yields a nice generalization of the Givens sine and cosine in the block setting; see, cf. Section 3.5 below . We can characterize some elements of this CS-decomposition by studying the QR factorization of $\overline{\mathbf{H}}_j^{(B)}$ and its relationship to the rank of $\mathbf{H}_j^{(B)}$. The proofs that follow often use generalizations of elements of proofs in [5].

LEMMA 3.11. *Let* $\operatorname{rank} \mathbf{H}_j^{(B)} = (j-1)L + r$ *with* $1 \le r \le L$. *Then we can write*

$$\mathbf{H}_j^{(B)} = \begin{bmatrix} \overline{\mathbf{H}}_{j-1}^{(B)} & \mathbf{L}_j \end{bmatrix} \tag{3.20}$$

with $\mathbf{L}_j \in \mathbb{C}^{jL\times L}$ *such that*

$$\mathbf{L}_j = \overline{\mathbf{H}}_{j-1}^{(B)}\widehat{\mathbf{Y}}_1 + \mathbf{G}_j\widehat{\mathbf{Y}}_2 \tag{3.21}$$

with $\widehat{\mathbf{Y}}_1 \in \mathbb{C}^{(j-1)L\times L}$, $\widehat{\mathbf{Y}}_2 \in \mathbb{C}^{r\times L}$, *and* $\mathbf{G}_j \in \mathbb{C}^{jL\times r}$ *having orthonormal columns which are orthogonal to* $\mathcal{R}\left(\overline{\mathbf{H}}_{j-1}^{(B)}\right)$. *Furthermore, the blocks* $\mathbf{Z}_j$ *and* $\widehat{\mathbf{H}}_{jj}$ *from* (3.9) *have the following representations*

$$\mathbf{Z}_j = \mathbf{R}_{j-1}\widehat{\mathbf{Y}}_1 \qquad \text{and} \qquad \widehat{\mathbf{H}}_{jj} = \widehat{\mathbf{M}}_j\widehat{\mathbf{Y}}_2, \tag{3.22}$$

where $\widehat{\mathbf{M}}_j \in \mathbb{C}^{L\times r}$ *so that* $\widehat{\mathbf{M}}_j\widehat{\mathbf{Y}}_2$ *is a rank-*r *outer product.*

Proof. We begin as in [5] by observing that the square matrix $\mathbf{H}_j^{(B)}$ has the form (3.20) following from its nested structure and rank. Since $\operatorname{rank} \mathbf{H}_j^{(B)} = (j-1)L+r$, we can represent the columns of $\mathbf{L}_j$ as linear combinations of vectors coming from $\mathcal{R}(\overline{\mathbf{H}}_{j-1}^{(B)})$ and vectors coming from a subspace of $\mathcal{R}(\overline{\mathbf{H}}_{j-1}^{(B)})^\perp$, from which (3.21) follows, where $\mathbf{G}_j \in \mathbb{C}^{jL\times r}$ has orthonormal columns such that $\mathcal{R}(\overline{\mathbf{H}}_{j-1}^{(B)}) \perp \mathcal{R}(\mathbf{G}_j)$ and $\mathcal{R}(\mathbf{H}_j^{(B)}) = \mathcal{R}\left(\begin{bmatrix} \overline{\mathbf{H}}_{j-1}^{(B)} & \mathbf{G}_j \end{bmatrix}\right)$. Thus we can write

$$\overline{\mathbf{H}}_j^{(B)} = \begin{bmatrix} \overline{\mathbf{H}}_{j-1}^{(B)} & \mathbf{L}_j \\ & \mathbf{H}_{j+1,j} \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{Q}}_{j-1}\begin{bmatrix}\mathbf{R}_{j-1}\\ \mathbf{0}\end{bmatrix} & \overline{\mathbf{Q}}_{j-1}\begin{bmatrix}\mathbf{R}_{j-1}\\ \mathbf{0}\end{bmatrix}\widehat{\mathbf{Y}}_1 + \mathbf{G}_j\widehat{\mathbf{Y}}_2 \\ & \mathbf{H}_{j+1,j} \end{bmatrix},$$

and we have that

$$\begin{aligned} \overline{\mathbf{R}}_j &= \mathbf{Q}_j^{(j+1)}\overline{\mathbf{Q}}_{j-1}^*\begin{bmatrix} \overline{\mathbf{Q}}_{j-1}\begin{bmatrix}\mathbf{R}_{j-1}\\ \mathbf{0}\end{bmatrix} & \overline{\mathbf{Q}}_{j-1}\begin{bmatrix}\mathbf{R}_{j-1}\\ \mathbf{0}\end{bmatrix}\widehat{\mathbf{Y}}_1 + \mathbf{G}_j\widehat{\mathbf{Y}}_2 \\ & \mathbf{H}_{j+1,j} \end{bmatrix} \\ &= \mathbf{Q}_j^{(j+1)}\begin{bmatrix} \begin{bmatrix}\mathbf{R}_{j-1}\\ \mathbf{0}\end{bmatrix} & \begin{bmatrix}\mathbf{R}_{j-1}\\ \mathbf{0}\end{bmatrix}\widehat{\mathbf{Y}}_1 + \overline{\mathbf{Q}}_{j-1}^*\mathbf{G}_j\widehat{\mathbf{Y}}_2 \\ & \mathbf{H}_{j+1,j} \end{bmatrix}. \end{aligned}$$

Since $\overline{\mathbf{Q}}_{j-1} \in \mathbb{C}^{(j+1)L\times(j+1)L}$, its columns form an orthonormal basis for $\mathbb{C}^{(j+1)L}$. However, from the upper triangular structure of $\overline{\mathbf{R}}_{j-1}$, we know we can partition the columns of $\overline{\mathbf{Q}}_{j-1} \in \mathbb{C}^{(j+1)L\times(j+1)L}$ such that the first $(j-1)L$ columns form a basis of $\mathcal{R}(\overline{\mathbf{H}}_{j-1}^{(B)})$ and the remaining columns form a basis for $\mathcal{R}(\overline{\mathbf{H}}_{j-1}^{(B)})^\perp$, of which $\mathcal{R}(\mathbf{G}_j)$ is a subspace. Thus we can write

$$\overline{\mathbf{Q}}_{j-1}^*\mathbf{G}_j = \begin{bmatrix}\mathbf{0}\\ \widehat{\mathbf{M}}_j\end{bmatrix}$$

with $\widehat{\mathbf{M}}_j \in \mathbb{C}^{L\times r}$, which yields

$$\begin{aligned}\overline{\mathbf{R}}_j &= \mathbf{Q}_j^{(j+1)} \begin{bmatrix} \begin{bmatrix}\mathbf{R}_{j-1}\\ \mathbf{0}\end{bmatrix} & \begin{bmatrix}\mathbf{R}_{j-1}\\ \mathbf{0}\end{bmatrix}\widehat{\mathbf{Y}}_1 + \begin{bmatrix}\mathbf{0}\\ \widehat{\mathbf{M}}_j\end{bmatrix}\widehat{\mathbf{Y}}_2 \\ & \mathbf{H}_{j+1,j}\end{bmatrix}\\ &= \mathbf{Q}_j^{(j+1)} \begin{bmatrix} \begin{bmatrix}\mathbf{R}_{j-1}\\ \mathbf{0}\end{bmatrix} & \begin{bmatrix}\mathbf{R}_{j-1}\widehat{\mathbf{Y}}_1\\ \widehat{\mathbf{M}}_j\widehat{\mathbf{Y}}_2\end{bmatrix} \\ & \mathbf{H}_{j+1,j}\end{bmatrix}.\end{aligned}$$

After some simplifications, both the identities for $\mathbf{Z}_j$ and $\widehat{\mathbf{H}}_{jj}$ have been proven. ∎

COROLLARY 3.12. *The representations in* (3.22) *are not unique, and there always exists one such representation such that* $\widehat{\mathbf{M}}_j$ *has orthonormal columns and* $\widehat{\mathbf{Y}}_2$ *is upper triangular.*

Proof. Let $\widehat{\mathbf{Y}}_2 = \mathbf{Q}_{\widehat{\mathbf{Y}}_2}\mathbf{R}_{\widehat{\mathbf{Y}}_2}$ be the QR factorization. With the updates $\mathbf{G}_j \leftarrow \mathbf{G}_j\mathbf{Q}_{\widehat{\mathbf{Y}}_2}$ and $\widehat{\mathbf{Y}}_2 \leftarrow \mathbf{R}_{\widehat{\mathbf{Y}}_2}$, (3.21) still holds with $\mathbf{G}_j$ still having orthonormal columns. With the updates $\widehat{\mathbf{M}}_j \leftarrow \mathbf{Q}_{\widehat{\mathbf{M}}_j}$ and $\widehat{\mathbf{Y}}_2 \leftarrow \mathbf{R}_{\widehat{\mathbf{M}}_j}\widehat{\mathbf{Y}}_2$, (3.22) still holds. Thus we have demonstrated the non-uniqueness of (3.22) and that $\widehat{\mathbf{M}}_j$ and $\widehat{\mathbf{Y}}_2$ with the structures we sought always exist. ∎

Henceforth, we assume that $\widehat{\mathbf{M}}_j$ has orthonormal columns and that $\widehat{\mathbf{Y}}_2$ is upper triangular. Lemma 3.11 and Corollary 3.12 illuminate various properties of the CS-decomposition of $\widehat{\boldsymbol{\mathcal{H}}}_j$. We note here that for any $1 \le m \le L$ and a matrix $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{L\times m}$ with orthonormal columns, that $\boldsymbol{\mathcal{A}}^\perp \in \mathbb{C}^{L\times(L-r)}$ (a notation we abuse) is some matrix which has orthonormal columns spanning $\mathcal{R}(\boldsymbol{\mathcal{A}})^\perp$ *whose exact structure is determined by the context in which it is used.* Furthermore, let $\boldsymbol{\mathcal{U}}(\cdot)$ refer to the $\mathcal{U}$-factor of the singular value decomposition of the argument.

LEMMA 3.13. *The orthogonal transformation* $\widehat{\boldsymbol{\mathcal{H}}}_j$ *with CS-decomposition described in* (3.19) *has the following properties,*

(I) $\mathbf{Q}_j^{(12)} = \mathbf{N}_j^{-*}\mathbf{H}_{j+1,j}^*$, *and it is lower triangular.*
(II) $\boldsymbol{\mathcal{U}}_1 = \boldsymbol{\mathcal{U}}(\mathbf{N}_j^{-*}\mathbf{H}_{j+1,j}^*)$.
(III) $\operatorname{rank}\mathbf{Q}_j^{(12)} = \operatorname{rank}\mathbf{Q}_j^{(21)} = L$, *i.e., they are nonsingular.*
(IV) $\operatorname{rank}\mathbf{Q}_j^{(11)} = \operatorname{rank}\mathbf{Q}_j^{(22)} = r$.
(V) $\boldsymbol{\mathcal{V}}_1 = \begin{bmatrix}\widehat{\mathbf{M}}_j\boldsymbol{\mathcal{Q}} & \widehat{\mathbf{M}}_j^\perp\end{bmatrix}$ *where* $\boldsymbol{\mathcal{Q}} \in \mathbb{C}^{r\times r}$ *is unitary.*

Proof. Observing that

$$\begin{aligned}&\begin{bmatrix}\mathbf{Q}_j^{(11)} & \mathbf{Q}_j^{(12)}\\ \mathbf{Q}_j^{(21)} & \mathbf{Q}_j^{(22)}\end{bmatrix}\begin{bmatrix}\widehat{\mathbf{H}}_{jj}\\ \mathbf{H}_{j+1,j}\end{bmatrix} = \begin{bmatrix}\mathbf{N}_j\\ \mathbf{0}\end{bmatrix}\\ &\qquad\Longleftrightarrow \begin{bmatrix}\left(\mathbf{Q}_j^{(11)}\right)^* & \left(\mathbf{Q}_j^{(21)}\right)^*\\ \left(\mathbf{Q}_j^{(12)}\right)^* & \left(\mathbf{Q}_j^{(22)}\right)^*\end{bmatrix}\begin{bmatrix}\mathbf{N}_j\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\widehat{\mathbf{H}}_{jj}\\ \mathbf{H}_{j+1,j}\end{bmatrix},\end{aligned} \tag{3.23}$$

the right-hand equation of (3.23) yields

$$\left(\mathbf{Q}_j^{(11)}\right)^*\mathbf{N}_j = \widehat{\mathbf{H}}_{jj} = \widehat{\mathbf{M}}_j\widehat{\mathbf{Y}}_2, \qquad \text{and} \qquad \left(\mathbf{Q}_j^{(12)}\right)^*\mathbf{N}_j = \mathbf{H}_{j+1,j}. \tag{3.24}$$

Since we assume no breakdown of the block Arnoldi method, we know that $\mathbf{N}_j$ is nonsingular and we can see that $\left(\mathbf{Q}_j^{(21)}\right)^* = \mathbf{H}_{j+1,j}\mathbf{N}_j^{-1}$ which yields Property I. This automatically

ETNA
Kent State University and
Johann Radon Institute (RICAM)

proves Property II as well. This also implies that $\mathbf{Q}_j^{(12)}$ is nonsingular (i.e., has rank L). From (3.19), we know $\mathbf{Q}_j^{(21)}$ and $\mathbf{Q}_j^{(12)}$ have the same singular values which completes the proof of Property III. The first equation in (3.24) can be transformed to $\left(\mathbf{Q}_j^{(11)}\right)^* = \widehat{\mathbf{M}}_j\widehat{\mathbf{Y}}_2\mathbf{N}_j^{-1}$ implying that $\mathcal{R}\left(\left(\mathbf{Q}_j^{(11)}\right)^*\right) \subseteq \mathcal{R}(\widehat{\mathbf{M}}_j)$. We know that $\widehat{\mathbf{Y}}_2$ is full rank from how it was constructed, thus $\mathcal{R}\left(\left(\mathbf{Q}_j^{(11)}\right)^*\right) = \mathcal{R}\left(\widehat{\mathbf{M}}_j\right)$. This yields Property IV, since from (3.19) we know that $\mathbf{Q}_j^{(11)}$ and $\mathbf{Q}_j^{(22)}$ also share the same singular values. From (3.19), we know that $\left(\mathbf{Q}_j^{(11)}\right)^* = \boldsymbol{\mathcal{V}}_1\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{U}}_1^*$. This implies Property V due to the assumed ordering of the singular values contained in $\boldsymbol{\mathcal{C}}$. □

Lemma 3.11 also allows us to describe the structure of the orthogonal transformation $\widehat{\mathbf{Q}}_j^{(b)}$, the non-trivial block of $\widehat{\mathbf{Q}}_j^{(j)}$.

LEMMA 3.14. *We have that*

$$\widehat{\mathbf{Q}}_j^{(b)} = \begin{bmatrix}\widehat{\mathbf{M}}_j & \widehat{\mathbf{M}}_j^{\perp}\end{bmatrix}^*,$$

so that we then can write

$$\widehat{\mathbf{N}}_j = \begin{bmatrix}\widehat{\mathbf{Y}}_2 \\ \mathbf{0}_{(L-r)\times L}\end{bmatrix}.$$

Proof. This follows directly from the assumptions on $\widehat{\mathbf{M}}_j$ (orthogonal columns) and $\widehat{\mathbf{Y}}_2$ (upper triangular). □

COROLLARY 3.15. *It follows directly that* $\operatorname{rank} \mathbf{C}_j = \operatorname{rank} \widehat{\mathbf{N}}_j$.

Proof. The combination of Lemma 3.6 with Property IV of Lemma 3.13 yields the result. □

We have now collected sufficient intermediate results to develop our main results. As in the single-vector Krylov method case, the rank of $\mathbf{H}_j^{(B)}$ is intimately related with the solution of the block GMRES least-squares problem (2.4). The following theorem is a generalization of [5, Theorem 3.1], although we frame it a bit differently.

THEOREM 3.16. *The matrix* $\mathbf{H}_j^{(B)}$ *is singular with* $\operatorname{rank} \mathbf{H}_j^{(B)} = (j-1)L + r$ *with* $r < L$ *if and only if the* j*th block GMRES update* $\mathbf{S}_j^{(G)}$ *is such that*

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

Proof. Let us first assume that $\mathbf{H}_j^{(B)}$ is singular with rank $(j-1)L + r$. It follows from Lemma 3.10 that

$$\mathbf{S}_j^{(G)} = \mathbf{W}_j\begin{bmatrix}-\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{N}_j^{-1}\mathbf{C}_j \\ \mathbf{N}_j^{-1}\mathbf{C}_j\end{bmatrix} = \mathbf{W}_{j-1}\left(-\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{N}_j^{-1}\mathbf{C}_j\right) + \mathbf{V}_j\left(\mathbf{N}_j^{-1}\mathbf{C}_j\right). \tag{3.25}$$

From Corollary 3.15 it follows that the rank of $\mathbf{C}_j$ (and thus also $\mathbf{N}_j^{-1}\mathbf{C}_j$) is r. Let

$$\boldsymbol{\mathcal{P}}_j = \begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_r\end{bmatrix} \in \mathbb{C}^{L\times r} \tag{3.26}$$

be the matrix with orthonormal columns spanning $\mathcal{R}(\mathbf{N}_j^{-1}\mathbf{C}_j)$. It follows directly then that the vectors in $\mathcal{R}\left(\mathbf{S}_j^{(G)}\right)$ only have non-trivial intersection with an r-dimensional subspace of $\mathcal{R}\left(\mathbf{V}_j\right)$, namely the subspace $\mathcal{R}(\mathbf{V}_j\boldsymbol{\mathcal{P}}_j)$.

Now assume that at the jth iteration of block GMRES, the span of the columns of the update $\mathbf{S}_j^{(G)}$ has an r-dimensional non-trivial intersection with $\mathcal{R}(\mathbf{V}_j)$. This implies that there exists $\boldsymbol{\mathcal{P}}_j$ of the form (3.26) such that $\mathcal{R}\left(\mathbf{S}_j^{(G)}\right) \cap \mathcal{R}\left(\mathbf{V}_j\right) = \mathcal{R}\left(\mathbf{V}_j \boldsymbol{\mathcal{P}}_j\right)$. It follows again from Lemma 3.10 that $\mathbf{S}_j^{(G)}$ has the form (3.25). However, this then implies that $\operatorname{rank} \mathbf{N}_j^{-1}\mathbf{C}_j = r$. Since $\mathbf{N}_j^{-1}$ is invertible, it follows that $\operatorname{rank} \mathbf{C}_j = r$, and from Corollary 3.15 we then have that $\operatorname{rank} \widehat{\mathbf{N}}_j = r$, and thus $\mathbf{H}_j^{(B)}$ has rank $(j-1)L + r$. ∎

We observe here that Theorem 3.16 and its proof hinge on the structure of $\mathbf{C}_j$. If r is nonzero, it follows that $\mathbf{C}_j$ must be nonzero but singular due to Corollary 3.15. The only case in which we can have total stagnation (i.e., $\mathbf{C}_j = \mathbf{0}$), then, is when $r = 0$. Thus we state the following corollary, which is the block analog of [5, Theorem 3.1].

COROLLARY 3.17. *The matrix* $\mathbf{H}_j^{(B)}$ *is singular with* $\operatorname{rank} \mathbf{H}_j^{(B)} = (j-1)L$ *if and only if block GMRES has totally stagnated with* $\mathbf{X}_j^{(G)} = \mathbf{X}_{j-1}^{(G)}$.

It follows that if there is a nontrivial $\mathbf{S}_j^{(G)}$ whose columns come from $\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)$ yielding a better minimizer, it can be decomposed into a part coming from $\mathcal{R}\left(\mathbf{V}_j\right)$ and a corresponding part from $\mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0)$ which is completely determined by the correction coming from $\mathcal{R}\left(\mathbf{V}_j\right)$.

LEMMA 3.18. *Let* $\mathbf{S}_j^{(\cdot)} = \mathbf{S}_{j,1}^{(\cdot)} + \mathbf{S}_{j,2}^{(\cdot)}$ *where* $\mathbf{S}_{j,1}^{(\cdot)}(:,i) \in \mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0)$, $\mathbf{S}_{j,2}^{(\cdot)}(:,i) \in \mathcal{R}\left(\mathbf{V}_j\right)$ *for* $1 \le i \le L$, *and* $(\cdot)$ *stands for either* (G) *or* (F). *Then* $\mathbf{S}_{j,1}^{(\cdot)} = \mathfrak{N}_j^{(\cdot)} \mathbf{S}_{j,2}^{(\cdot)}$ *where* $\mathfrak{N}_j^{(\cdot)}$ *is a nilpotent operator such that*

$$\mathfrak{N}_j^{(\cdot)} : \mathcal{R}\left(\mathbf{V}_j\right) \to \mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0) \qquad \textit{and} \qquad \mathfrak{N}_j^{(\cdot)} : \mathcal{R}\left(\mathbf{V}_j\right)^{\perp} \to \{\mathbf{0}\},$$

i.e., $\mathcal{R}\left(\mathfrak{N}_j^{(\cdot)}\right) = \mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0)$, and $\mathcal{N}\left(\mathfrak{N}_j^{(\cdot)}\right) = \mathcal{R}\left(\mathbf{V}_j\right)^{\perp}$.

Proof. We prove only for the case $(\cdot) = (G)$, as both proofs proceed in the same way. From (3.25), we see that

$$\begin{aligned}\mathbf{S}_{j,1}^{(G)} &= -\mathbf{W}_{j-1}\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{N}_j^{-1}\mathbf{C}_j = -\mathbf{W}_{j-1}\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{V}_j^{*}\mathbf{V}_j\mathbf{N}_j^{-1}\mathbf{C}_j \\ &= -\mathbf{W}_{j-1}\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\mathbf{V}_j^{*}\mathbf{S}_{j,2}^{(G)} = -\mathbf{W}_{j-1}\widehat{\mathbf{Y}}_1\mathbf{V}_j^{*}\mathbf{S}_{j,2}^{(G)}.\end{aligned}$$

Assigning $\mathfrak{N}_j^{(G)} = -\mathbf{W}_{j-1}\widehat{\mathbf{Y}}_1\mathbf{V}_j^{*}$, one can easily check that it satisfies the statements of the lemma. ∎

The following theorem is a generalization of [5, Theorem 3.3].

THEOREM 3.19. *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* $\mathbf{S}_j^{(G)}$.

Proof. We begin with the assumption that $\mathbf{S}_j^{(G)}$ has this property. We know from Theorem 3.16 that this implies $\operatorname{rank} \mathbf{H}_j^{(B)} = (j-1)L + r$ and that $\operatorname{rank} \widehat{\mathbf{N}}_j = r$. It follows then that $\widehat{\mathbf{N}}_j^{\dagger}$ has a dimension $L - r$ null space.[4] From Lemma 3.10, we can write

$$\widetilde{\mathbf{S}}_j^{(F)} = \mathbf{W}_{j-1}\left(-\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j\widehat{\mathbf{N}}_j^{\dagger}\widehat{\mathbf{C}}_j\right) + \mathbf{V}_j\left(\widehat{\mathbf{N}}_j^{\dagger}\widehat{\mathbf{C}}_j\right).$$

[4]Because we know that $\widehat{\mathbf{N}}_j$ is upper triangular with an $(L-r)\times(L-r)$ zero block in the bottom right-hand corner, it follows that $\mathcal{N}(\widehat{\mathbf{N}}_j^{\dagger}) = \mathcal{R}(\widehat{\mathbf{N}}_j)^{\perp} = \operatorname{span}\left\{\mathbf{e}_{\mathrm{r}+1}^{[\mathrm{L}]}, \mathbf{e}_{\mathrm{r}+2}^{[\mathrm{L}]}, \ldots, \mathbf{e}_{\mathrm{L}}^{[\mathrm{L}]}\right\}$. Thus we can write $\widehat{\mathbf{N}}_j^{\dagger} = \begin{bmatrix} *_{L\times r} & \mathbf{0}_{L\times(L-r)}\end{bmatrix}$.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Since we know that $\widehat{\mathbf{C}}_j$ is nonsingular, it follows that $\operatorname{rank} \widehat{\mathbf{N}}_j^{\dagger}\widehat{\mathbf{C}}_j = r$. Thus, using the same argument used at the end of the proof of Theorem 3.16 it follows that $\widetilde{\mathbf{S}}_j^{(F)}$ only has a non-trivial intersection with an r-dimensional subspace of $\mathcal{R}(\mathbf{V}_j)$. For the other direction, we simply carry out the same steps but in reverse order. ∎

COROLLARY 3.20. *Block GMRES at iteration j totally stagnates with $\mathbf{X}_j^{(G)} = \mathbf{X}_{j-1}^{(G)}$ if and only if $\widetilde{\mathbf{X}}_j^{(F)} = \mathbf{X}_{j-1}^{(G)}$.*

Proof. This corresponds to the case $r = 0$ for Theorem 3.19. ∎

We now show that the case of partial stagnation of block GMRES (as defined at the beginning of Section 3) is actually just a special case of Theorem 3.16, and is not really of special interest with respect to this analysis

THEOREM 3.21. *Block GMRES suffers a partial stagnation at iteration j of the form (3.1) if and only if $0 < \operatorname{rank} \mathbf{C}_j \leq r$ where $r = |\mathbb{I}|$ such that for all $i \in \mathbb{I}$ the ith column of $\mathbf{C}_j$ is the zero vector.*

Proof. Let us first assume that the columns of $\mathbf{C}_j$ corresponding to indices in $\mathbb{I}$ are zero but that $\mathbf{C}_j \neq \mathbf{0}$. Then $\operatorname{rank} \mathbf{C}_j \leq L - |\mathbb{I}|$. Furthermore, since

$$\mathbf{S}_j^{(G)} = \mathbf{W}_j \begin{bmatrix} \mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \mathbf{N}_j^{-1}\mathbf{C}_j,$$

for $i \in \mathbb{I}$, if we look at the ith column of $\mathbf{S}_j^{(G)}$, we see that

$$\mathbf{S}_j^{(G)}\mathbf{e}_i^{[L]} = \mathbf{W}_j \begin{bmatrix} \mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \mathbf{N}_j^{-1}\mathbf{C}_j\mathbf{e}_i^{[L]} = \mathbf{W}_j \begin{bmatrix} \mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \mathbf{N}_j^{-1}\mathbf{0} = \mathbf{0}.$$

The first direction is thus proven.

Now we assume that partial stagnation occurs at the jth iteration where for each $i \in \mathbb{I}$, $\mathbf{X}_{j-1}^{(G)}\mathbf{e}_i^{[L]} = \mathbf{X}_j^{(G)}\mathbf{e}_i^{[L]}$. This implies that $\mathbf{S}_j^{(G)}\mathbf{e}_i^{[L]} = \mathbf{0}$ for all $i \in \mathbb{I}$. Specifically, this implies that $\mathbf{V}_j\mathbf{N}_j^{-1}\mathbf{C}_j\mathbf{e}_i^{[L]} = \mathbf{0}$. Because we assume that $\mathbf{V}_j$ is full rank and $\mathbf{N}_j^{-1}$ is nonsingular, it follows that $\mathbf{C}_j\mathbf{e}_i^{[L]} = \mathbf{0}$, which proves the other direction. ∎

Now we also state the block analog of Proposition 2.2.

THEOREM 3.22. *Suppose that $\mathbf{H}_j^{(B)}$ is nonsingular. Then at iteration j we have the following relationship between the approximations produced by block GMRES and block FOM,*

$$\mathbf{X}_j^{(G)} = \mathbf{X}_j^{(F)} \left(\widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^2\boldsymbol{\mathcal{Q}}^*\widehat{\mathbf{C}}_j\right) + \mathbf{X}_{j-1}^{(G)} \left(\widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{S}}^2\boldsymbol{\mathcal{Q}}^*\widehat{\mathbf{C}}_j\right). \tag{3.27}$$

Proof. Since $\mathbf{H}_j^{(B)}$ is nonsingular, we have that $\widehat{\mathbf{N}}_j$ and $\mathbf{C}_j$ are nonsingular, and the block FOM approximation $\mathbf{X}_j^{(F)}$ (and thus also $\mathbf{Y}_j^{(F)}$) exists. From the proof of Lemma 3.10, we have then that

$$\left(\mathbf{Y}_j^{(F)} - \begin{bmatrix} \mathbf{Y}_{j-1}^{(G)} \\ \mathbf{0} \end{bmatrix}\right) = \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \widehat{\mathbf{N}}_j^{-1}\widehat{\mathbf{C}}_j \quad \text{and}$$

$$\left(\mathbf{Y}_j^{(G)} - \begin{bmatrix} \mathbf{Y}_{j-1}^{(G)} \\ \mathbf{0} \end{bmatrix}\right) = \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \mathbf{N}_j^{-1}\mathbf{C}_j\ .$$

Since everything in this case is invertible, we see that

$$\left(\mathbf{Y}_j^{(F)} - \begin{bmatrix} \mathbf{Y}_{j-1}^{(G)} \\ \mathbf{0} \end{bmatrix}\right) \widehat{\mathbf{C}}_j^{-1}\widehat{\mathbf{N}}_j\mathbf{N}_j^{-1}\mathbf{C}_j = \mathbf{Y}_j^{(G)} - \begin{bmatrix} \mathbf{Y}_{j-1}^{(G)} \\ \mathbf{0} \end{bmatrix}. \tag{3.28}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

We can now simplify $\widehat{\mathbf{C}}_j^{-1}\widehat{\mathbf{N}}_j\mathbf{N}_j^{-1}\mathbf{C}_j$ using Lemmas 3.5, 3.6, and 3.13. We note that in the case that the block FOM approximation exists, we have that $\widehat{\mathbf{N}}_j = \widehat{\mathbf{Y}}_2$ is upper triangular and nonsingular, $\boldsymbol{\mathcal{V}}_1 = \widehat{\mathbf{M}}_j\boldsymbol{\mathcal{Q}}$, and $\widehat{\mathbf{Q}}_j^{(b)} = \widehat{\mathbf{M}}_j^*$. It follows then that

$$\begin{aligned}
\widehat{\mathbf{C}}_j^{-1}\widehat{\mathbf{N}}_j\mathbf{N}_j^{-1}\mathbf{C}_j &= \widehat{\mathbf{C}}_j^{-1}\widehat{\mathbf{Y}}_2\left(\left(\mathbf{Q}_j^{(11)}\right)^{-*}\widehat{\mathbf{M}}_j\widehat{\mathbf{Y}}_2\right)^{-1}\mathbf{Q}_j^{(11)}\widetilde{\mathbf{C}}_j \\
&= \widehat{\mathbf{C}}_j^{-1}\widehat{\mathbf{Y}}_2\left(\widehat{\mathbf{Y}}_2^{-1}\widehat{\mathbf{M}}_j^*\left(\mathbf{Q}_j^{(11)}\right)^*\right)\mathbf{Q}_j^{(11)}\widetilde{\mathbf{C}}_j \\
&= \widehat{\mathbf{C}}_j^{-1}\widehat{\mathbf{M}}_j^*\left(\boldsymbol{\mathcal{U}}_1\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{V}}_1^*\right)^*\boldsymbol{\mathcal{U}}_1\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{V}}_1^*\widetilde{\mathbf{C}}_j \\
&= \widehat{\mathbf{C}}_j^{-1}\widehat{\mathbf{M}}_j^*\left(\widehat{\mathbf{M}}_j\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{U}}_1^*\right)\boldsymbol{\mathcal{U}}_1\boldsymbol{\mathcal{C}}\left(\widehat{\mathbf{M}}_j\boldsymbol{\mathcal{Q}}\right)^*\widetilde{\mathbf{C}}_j \\
&= \widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^2\left(\widehat{\mathbf{M}}_j\boldsymbol{\mathcal{Q}}\right)^*\widetilde{\mathbf{C}}_j \\
&= \widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^2\boldsymbol{\mathcal{Q}}^*\widehat{\mathbf{C}}_j. \qquad (3.29)
\end{aligned}$$

We now insert (3.29) into (3.28), multiply both sides by $\mathbf{W}_j$, and perform some algebraic manipulations to get

$$\mathbf{X}_j^{(G)} = \mathbf{X}_j^{(F)}\left(\widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^2\boldsymbol{\mathcal{Q}}^*\widehat{\mathbf{C}}_j\right) + \mathbf{X}_{j-1}^{(G)}\left(\mathbf{I} - \widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^2\boldsymbol{\mathcal{Q}}^*\widehat{\mathbf{C}}_j\right).$$

Lastly, we observe that $\widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^2\boldsymbol{\mathcal{Q}}^*\widehat{\mathbf{C}}_j$ is an eigendecomposition since $\left(\boldsymbol{\mathcal{Q}}^*\widehat{\mathbf{C}}_j\right)^{-1} = \widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}$, and we thus can write

$$\mathbf{I} - \widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^2\boldsymbol{\mathcal{Q}}^*\widehat{\mathbf{C}}_j = \widehat{\mathbf{C}}_j^{-1}\boldsymbol{\mathcal{Q}}\left(\mathbf{I} - \boldsymbol{\mathcal{C}}^2\right)\boldsymbol{\mathcal{Q}}^*\widehat{\mathbf{C}}_j.$$

The result follows by observing that $\mathbf{I} - \boldsymbol{\mathcal{C}}^2 = \boldsymbol{\mathcal{S}}^2$ which follows from (3.19). □

We will return shortly to understand the meaning of the angles associated to these sines and cosines in Section 3.5.

3.4. The case of breakdown in the block Arnoldi process. Our discussion of the case of breakdown focuses first upon the behaviors of block GMRES and block FOM at the jth iteration in which the block Arnoldi process produces p dependent basis vectors. For simplicity, we assume that no single system has converged but rather that some linear combination of some columns of the solution $\mathbf{X}$ is in $\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)$. Both the block GMRES and block FOM residuals are thus of rank $L - p$. We assume that these p vectors are replaced with p random vectors so that we maintain a block size of L. For the most part, what we have proven thus far holds with little to no alterations, but the reduction of residual rank does have some consequences.

We consider a breakdown at iteration j in which p dependent basis vectors are produced. Various strategies have been suggested for replacing dependent vectors in the interest of maintaining the block size of L. The block Arnoldi process produces from $\mathbf{A}\mathbf{V}_j \in \mathbb{C}^{n\times L}$ the block

$$\mathbf{U}_{j+1} = \mathbf{A}\mathbf{V}_j - \sum_{i=1}^{j}\mathbf{V}_i\mathbf{H}_{ij} \quad \text{with} \quad \mathbf{H}_{ij} \in \mathbb{C}^{L\times L}, \quad \text{and} \quad \operatorname{rank}\mathbf{U}_{j+1} = L - p.$$

Then we have the reduced QR factorization $\mathbf{U}_{j+1} = \ddot{\mathbf{V}}_{j+1}\ddot{\mathbf{H}}_{j+1,j}$, where $\ddot{\mathbf{V}}_{j+1} \in \mathbb{C}^{n\times(L-p)}$ has columns spanning $\mathcal{R}(\mathbf{U}_{j+1})$, and $\ddot{\mathbf{H}}_{j+1,j} \in \mathbb{C}^{(L-p)\times L}$ is upper triangular. To maintain block size, we set $\mathbf{H}_{j+1,j} = \begin{bmatrix}\ddot{\mathbf{H}}_{j+1,j} \\ \mathbf{0}\end{bmatrix} \in \mathbb{C}^{L\times L}$ and $\mathbf{V}_{j+1} = \begin{bmatrix}\ddot{\mathbf{V}}_{j+1} & \boldsymbol{\mathcal{Z}}\end{bmatrix} \in \mathbb{C}^{n\times L}$ where

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$\mathcal{Z} \in \mathbb{C}^{n\times p}$ is a set of orthonormal replacement vectors, which have been orthogonalized against all of the block Arnoldi vectors. Thus the columns of $\mathbf{W}_{j+1}$ no longer span a Krylov subspace, but they do still satisfy the block Arnoldi relation (2.1). The iteration continues unabated. It is observed in, e.g., [32], that at the iteration in which the breakdown occurs, the least squares problem still has a unique solution. From the analysis in this paper, this corresponds to $\mathbf{N}_j$ still being nonsingular. Furthermore, as we assume that iteration j is the first iteration at which there is a block Arnoldi breakdown, the block residual $\mathbf{F}_{j-1}^{(G)}$ is full rank, and, thus, so are the $\mathbf{C}$-matrices. Therefore, at iteration j, if there has been a block Arnoldi breakdown, all of the results we have proven still hold with no alteration. The block residual $\mathbf{F}_j^{(G)}$ has rank $L-p$.

Without loss of generality, let us consider the case that the breakdown at iteration j is the only breakdown. Consider some later iteration $j+k$ with $k>0$. As we have replaced all dependent Arnoldi vectors with linearly independent ones, the GMRES least squares problem still has a unique solution. This implies that $\mathbf{N}_{j+k}$ is still nonsingular. The block GMRES residuals will continue to be rank $L-p$. Thus, the $\mathbf{C}$-matrices will be square (as we maintain block size) and rank-deficient. However, few of the results rely on the invertibility of these matrices. Indeed, the only result not valid in this case is Theorem 3.22. However, we can prove a weaker result in this case.

THEOREM 3.23. *Suppose at step j there has been a block Arnoldi breakdown with p dependent Arnoldi vectors being generated and that there are no further breakdowns. Let these vectors be replaced using the procedure described above. Then at iteration $j+k$, if $\mathbf{H}_{j+k}^{(B)}$ is nonsingular, we have that*

$$\mathbf{X}_j^{(F)} - \mathbf{X}_j^{(G)} = \mathbf{W}_j \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \widehat{\mathbf{Y}}_2^{-1} \mathcal{Q}\mathcal{S}^2\mathcal{Q}^*\widehat{\mathbf{C}}_j.$$

Proof. We show this by substituting many of the identities we have previously proven, which are still valid in this setting.

$$\begin{aligned}
\mathbf{Y}_j^{(F)} - \mathbf{Y}_j^{(G)} &= \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \left(\widehat{\mathbf{N}}_j^{-1}\widehat{\mathbf{Q}}_j^{(b)} - \mathbf{N}_j^{-1}\mathbf{Q}_j^{(11)} \right) \widetilde{\mathbf{C}}_j \\
&= \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \left(\widehat{\mathbf{Y}}_2^{-1}\widehat{\mathbf{M}}_j^* - \widehat{\mathbf{Y}}_2^{-1}\widehat{\mathbf{M}}_j^*\mathbf{Q}_j^{(11)*}\mathbf{Q}_j^{(11)} \right) \widetilde{\mathbf{C}}_j \\
&= \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \left(\widehat{\mathbf{Y}}_2^{-1}\widehat{\mathbf{M}}_j^* - \widehat{\mathbf{Y}}_2^{-1}\widehat{\mathbf{M}}_j^*\mathcal{V}_1\mathcal{C}\mathcal{U}_1^*\mathcal{U}_1\mathcal{C}\mathcal{V}_1^* \right) \widetilde{\mathbf{C}}_j \\
&= \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \left(\widehat{\mathbf{Y}}_2^{-1}\widehat{\mathbf{M}}_j^* - \widehat{\mathbf{Y}}_2^{-1}\widehat{\mathbf{M}}_j^*\widehat{\mathbf{M}}_j\mathcal{Q}\mathcal{C}^2\mathcal{Q}^*\widehat{\mathbf{M}}_j^* \right) \widetilde{\mathbf{C}}_j \\
&= \begin{bmatrix} -\mathbf{R}_{j-1}^{-1}\mathbf{Z}_j \\ \mathbf{I} \end{bmatrix} \widehat{\mathbf{Y}}_2^{-1} \left(\mathbf{I} - \mathcal{Q}\mathcal{C}^2\mathcal{Q}^* \right) \widehat{\mathbf{C}}_j.
\end{aligned}$$

One then performs a bit of algebra and multiplies both sides by $\mathbf{W}_j$ to get the result. ∎

Although this result is less satisfying that Theorem 3.22, as it does not generalize Proposition 2.2, it still yields valuable information about the relationship of the block FOM and block GMRES iterates in the case that breakdown has occurred. We see that if the angles represented by the sines contained on the diagonal of $\mathcal{S}$ are small, this implies that block FOM and block GMRES in this scenario produce iterations which are not far from one another. We must thus now clarify the precise significance of these angles to complete our analysis.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

3.5. Principal angles between the range of $\mathbf{F}_{j-1}^{(G)}$ and $\mathbf{A}\mathbb{K}_j(\mathbf{A},\mathbf{F}_0)$. In this section, we show that the angles represented by the sines and cosines from the CS-decomposition of (3.18) which appear in (3.27) are the principal angles between the columns space of the previous residual and the current residual constraint space.

In [11], many geometric properties of single-vector projection methods were analyzed. In particular, the authors discussed minimum residual projection methods such as GMRES. In that paper, the authors show that the angle represented by the Givens sine and cosine calculated at iteration j of GMRES is actually the principal angle between the $(j-1)$st GMRES residual and the jth constraint space. In essence, the closeness of this angle to zero indicates how much of the $(j-1)$st residual lies in the jth constraint space and will thus be eliminated by the projection at iteration j. If the angle is near $\frac{\pi}{2}$, however, then the Givens cosine c_j is close to 0 and we have near stagnation since almost none of the $(j-1)$st residual lies in the new constraint space, and thus there will not be much improvement from the projection at iteration j.

To illuminate the meaning of these angles in the block setting, we generalize some results from the single-vector GMRES case. Following [11], we represent $\mathbf{A}\mathbb{K}_j(\mathbf{A},\mathbf{F}_0)$ with a specific, useful basis. The columns of $\mathbf{W}_{j+1}$ form an orthonormal basis for $\mathbb{K}_{j+1}(\mathbf{A},\mathbf{F}_0)$, and it follows from the block Arnoldi relation (2.1) that the columns of $\mathbf{W}_{j+1}\overline{\mathbf{H}}_j^{(B)}$ form a non-orthonormal basis for $\mathbf{A}\mathbb{K}_j(\mathbf{A},\mathbf{F}_0)$. Using the QR factorization $\overline{\mathbf{H}}_j^{(B)} = \overline{\mathbf{Q}}_j^*\overline{\mathbf{R}}_j$, we see that the columns of $\mathbf{W}_{j+1}\overline{\mathbf{Q}}_j^*$ form another, orthonormal basis of $\mathbf{A}\mathbb{K}_j(\mathbf{A},\mathbf{F}_0)$. From the equation

$$\mathbf{F}_0 = \mathbf{W}_{j+1}\mathbf{E}_1^{[(j+1)L]}\mathbf{S}_0 = \mathbf{W}_{j+1}\overline{\mathbf{Q}}_j^*\overline{\mathbf{Q}}_j\mathbf{E}_1^{[(j+1)L]}\mathbf{S}_0,$$

we see that $\overline{\mathbf{Q}}_j\mathbf{E}_1^{[(j+1)L]}\mathbf{S}_0$ is a representation of $\mathbf{F}_0$ in that basis. This leads to a generalization of, e.g., [34, Equation 6.48], that the Givens sines can be used to cheaply update the GMRES residual norm. We note that following from the block partitioning of the orthogonal transformation in (3.18), we can write

$$\mathbf{Q}_i^{(j+1)} = \begin{bmatrix} \mathbf{I}_{(i-1)L} & & & \\ & \mathbf{Q}_i^{(11)} & \mathbf{Q}_i^{(12)} & \\ & \mathbf{Q}_i^{(21)} & \mathbf{Q}_i^{(22)} & \\ & & & \mathbf{I}_{(j-i)L} \end{bmatrix}. \tag{3.30}$$

Then we have the following.

LEMMA 3.24. *The representation $\overline{\mathbf{Q}}_j\mathbf{E}_1^{[(j+1)L]}\mathbf{S}_0$ of $\mathbf{F}_0$ has the following structure,*

$$\overline{\mathbf{Q}}_j\mathbf{E}_1^{[(j+1)L]}\mathbf{S}_0 = \begin{bmatrix} \mathbf{Q}_1^{(11)} \\ \mathbf{Q}_2^{(11)}\mathbf{Q}_1^{(21)} \\ \mathbf{Q}_3^{(11)}\mathbf{Q}_2^{(21)}\mathbf{Q}_1^{(21)} \\ \vdots \\ \mathbf{Q}_{j-1}^{(11)}\prod_{i=1}^{j-2}\mathbf{Q}_i^{(21)} \\ \mathbf{Q}_j^{(11)}\prod_{i=1}^{j-1}\mathbf{Q}_i^{(21)} \\ \prod_{i=1}^{j}\mathbf{Q}_i^{(21)} \end{bmatrix}\mathbf{S}_0. \tag{3.31}$$

Proof. This follows from the fact that $\overline{\mathbf{Q}}_j = \prod_{i=1}^{j}\mathbf{Q}_{j-i+1}^{(j+1)}$ and the structure of the orthogonal transformations in (3.30). ∎

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Let us denote with $\measuredangle(\mathcal{U}_1, \mathcal{U}_2)$ the set of principle angles between the subspaces $\mathcal{U}_1$ and $\mathcal{U}_2$. Following [11], we can compute a product of matrices whose singular values are the sines of the principal angles $\measuredangle(\mathcal{R}(\mathbf{F}_0), \mathbf{A}\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0))$.

LEMMA 3.25. *The principal angles $\measuredangle(\mathcal{R}(\mathbf{F}_0), \mathbf{A}\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0))$ are the singular values of the product $\prod_{i=1}^{j} \mathbf{Q}_{j-i+1}^{(21)}$.*

Proof. We have the equalities

$$\begin{aligned}\measuredangle(\mathcal{R}(\mathbf{F}_0), \mathbf{A}\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)) &= \measuredangle\left(\mathcal{R}(\overline{\mathbf{Q}}_j \mathbf{E}_1^{[(j+1)L]} \mathbf{S}_0), \mathcal{R}(\mathbf{W}_{j+1} \overline{\mathbf{Q}}_{j+1}^{*} \overline{\mathbf{R}}_j)\right) \\ &= \measuredangle\left(\mathcal{R}\left(\overline{\mathbf{Q}}_j \mathbf{E}_1^{[(j+1)]}\right), \mathcal{R}\left(\overline{\mathbf{R}}_j\right)\right).\end{aligned}$$

Under the assumptions in this paper for the non-breakdown case, the range $\mathcal{R}\left(\overline{\mathbf{R}}_j\right) = \mathcal{R}\left(\begin{bmatrix}\mathbf{R}_j \\ \mathbf{0}\end{bmatrix}\right)$ is isomorphic with $\mathbb{C}^{(j+1)L}$ (due to the nonsingularity of $\mathbf{R}_j$) with basis $\left\{\mathbf{e}_1^{[(j+1)L]}, \mathbf{e}_2^{[(j+1)L]}, \ldots, \mathbf{e}_{jL}^{[j+1)L]}\right\}$, i.e., the last L coordinates are zero. It is clear that $\overline{\mathbf{Q}}_j \mathbf{E}_1^{[(j+1)L]}$ has orthonormal columns. Let

$$\overline{\mathbf{Q}}_j \mathbf{E}_1^{[(j+1)L]} = \begin{bmatrix}\mathfrak{Q}_1 \\ \mathfrak{Q}_2\end{bmatrix}$$

be a block partitioning with $\mathfrak{Q}_1 \in \mathbb{C}^{jL \times L}$ and $\mathfrak{Q}_2 \in \mathbb{C}^{L \times L}$. With this partitioning, $\overline{\mathbf{Q}}_j \mathbf{E}_1^{[(j+1)]}$ admits a skinny CS-decomposition (see, e.g., [16, Section 2.5.4]) yielding the simultaneous singular value decompositions

$$\mathfrak{C} = \mathfrak{U}_1^{*} \mathfrak{Q}_1 \mathfrak{V} \quad \text{and} \quad \mathfrak{S} = \mathfrak{U}_2^{*} \mathfrak{Q}_2 \mathfrak{V}$$

with $\mathfrak{U}_1 \in \mathbb{C}^{jL \times jL}$, $\mathfrak{U}_2 \in \mathbb{C}^{L \times L}$, $\mathfrak{V} \in \mathbb{C}^{L \times L}$, and

$$\mathfrak{C} = \begin{bmatrix}\mathbf{I}_{(j-1)L} & \\ & \operatorname{diag}_{i=1}^{L}\{c_i\}\end{bmatrix} \in \mathbb{C}^{jL \times jL} \qquad \text{and} \qquad \mathfrak{S} = \operatorname{diag}_{i=1}^{L}\{s_i\} \in \mathbb{C}^{L \times L}.$$

Since $\begin{bmatrix}\mathbf{I}_{jL} \\ \mathbf{0}_{L \times jL}\end{bmatrix}$ has orthonormal columns spanning $\mathcal{R}\left(\overline{\mathbf{R}}_j\right)$, the cosines of the sought-after principal angles are given by the singular values of $\mathfrak{Q}_1 = \begin{bmatrix}\mathbf{I}_{jL} \\ \mathbf{0}_L\end{bmatrix}^{*}\left(\overline{\mathbf{Q}}_j \mathbf{E}_1^{[(j+1)]}\right)$, i.e., the entries of $\mathfrak{C}$. Many of these cosines are equal to 1 (i.e., $\theta_i = 0$ for $i = 1, 2, \ldots (j-1)L$). However, the L nontrivial angles are also represented by their sines in the entries of $\mathfrak{S}$ which are the singular values of $\mathfrak{Q}_2$, and this proves the lemma. □

Using similar techniques, we can prove the following result.

THEOREM 3.26. *The angles represented by the sines and cosines of the CS-decomposition of the jth orthogonal transformation (3.19) are the principal angles between the column space of the previous block GMRES residual $\mathbf{F}_{j-1}^{(G)}$ and the jth constraint space $\mathbf{A}\mathbb{K}_j(\mathbf{A}, \mathbf{F}_0)$.*

Proof. As has already been discussed, the columns of $\mathbf{W}_{i+1} \overline{\mathbf{Q}}_i^{*} \begin{bmatrix}\mathbf{I}_{iL} \\ \mathbf{0}_L\end{bmatrix}$ are an orthonormal basis for $\mathbf{A}\mathbb{K}_i(\mathbf{A}, \mathbf{F}_0)$ for all i. Let $\boldsymbol{\mathcal{P}}_{j-1}$ be the orthogonal projector onto $\mathbf{A}\mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0)$, which means we can write $\mathbf{F}_{j-1}^{(G)} = (\mathbf{I} - \boldsymbol{\mathcal{P}}_{j-1})\mathbf{F}_0$. Using the orthonormal basis of $\mathbf{A}\mathbb{K}_{j-1}(\mathbf{A}, \mathbf{F}_0)$, we can write

$$\boldsymbol{\mathcal{P}}_{j-1} \mathbf{F}_0 = \mathbf{W}_j \overline{\mathbf{Q}}_{j-1}^{*} \begin{bmatrix}\mathbf{I}_{(j-1)L} \\ \mathbf{0}_{L \times (j-1)L}\end{bmatrix} \begin{bmatrix}\mathbf{I}_{(j-1)L} & \mathbf{0}_{(j-1)L \times L}\end{bmatrix} \overline{\mathbf{Q}}_{j-1} \mathbf{E}_1^{[jL]} \mathbf{S}_0.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

It is then straightforward to show that

$$\mathbf{F}_{j-1}^{(G)} = (\mathbf{I}-\boldsymbol{\mathcal{P}}_{j-1})\mathbf{F}_0 = \mathbf{W}_j\overline{\mathbf{Q}}_{j-1}^{*}\left(\mathbf{I}-\begin{bmatrix}\mathbf{I}_{(j-1)L}\\ \mathbf{0}_{L\times(j-1)L}\end{bmatrix}\begin{bmatrix}\mathbf{I}_{(j-1)L} & \mathbf{0}_{(j-1)L\times L}\end{bmatrix}\right)\overline{\mathbf{Q}}_{j-1}\mathbf{E}_1^{[jL]}\mathbf{S}_0.$$

Observe that $\left(\mathbf{I}-\begin{bmatrix}\mathbf{I}_{(j-1)L}\\ \mathbf{0}_{L\times(j-1)L}\end{bmatrix}\begin{bmatrix}\mathbf{I}_{(j-1)L} & \mathbf{0}_{(j-1)L\times L}\end{bmatrix}\right)$ is the orthogonal projector onto the last L coordinate directions, i.e., onto $\operatorname{span}\left\{\mathbf{e}_{(j-1)L+1}^{[jL]},\mathbf{e}_{(j-1)L+2}^{[jL]},\ldots,\mathbf{e}_{jL}^{[jL]}\right\}$. Combining this with (3.31), we can rewrite

$$(\mathbf{I}-\boldsymbol{\mathcal{P}}_{j-1})\mathbf{F}_0 = \mathbf{W}_j\overline{\mathbf{Q}}_{j-1}^{*}\begin{bmatrix}\mathbf{0}_{(j-1)L\times L}\\ \prod_{i=1}^{j-1}\mathbf{Q}_i^{(21)}\end{bmatrix}\mathbf{S}_0.$$

The principal angle calculation can then be simplified,

$$\begin{aligned}
&\measuredangle\left(\mathcal{R}\left(\mathbf{F}_{j-1}^{(G)}\right),\mathbf{A}\mathbb{K}_j(\mathbf{A},\mathbf{F}_0)\right)\\
&\quad=\measuredangle\left(\mathcal{R}\left(\mathbf{W}_j\overline{\mathbf{Q}}_{j-1}^{*}\begin{bmatrix}\mathbf{0}_{(j-1)L\times L}\\ \prod_{i=1}^{j-1}\mathbf{Q}_i^{(21)}\end{bmatrix}\mathbf{S}_0\right),\mathcal{R}\left(\mathbf{W}_{j+1}\overline{\mathbf{Q}}_j^{*}\overline{\mathbf{R}}_j\right)\right)\\
&\quad=\measuredangle\left(\mathcal{R}\left(\mathbf{W}_{j+1}\begin{bmatrix}\mathbf{I}_{jL}\\ \mathbf{0}_{L\times jL}\end{bmatrix}\overline{\mathbf{Q}}_{j-1}^{*}\begin{bmatrix}\mathbf{0}_{(j-1)L\times L}\\ \prod_{i=1}^{j-1}\mathbf{Q}_i^{(21)}\end{bmatrix}\right),\mathcal{R}\left(\mathbf{W}_{j+1}\overline{\mathbf{Q}}_j^{*}\overline{\mathbf{R}}_j\right)\right)\\
&\quad=\measuredangle\left(\mathcal{R}\left(\begin{bmatrix}\overline{\mathbf{Q}}_{j-1}^{*}\\ \mathbf{0}_{L\times jL}\end{bmatrix}\begin{bmatrix}\mathbf{0}_{(j-1)L\times L}\\ \prod_{i=1}^{j-1}\mathbf{Q}_i^{(21)}\end{bmatrix}\right),\mathcal{R}\left(\overline{\mathbf{Q}}_j^{*}\begin{bmatrix}\mathbf{I}_{jL}\\ \mathbf{0}_{L\times jL}\end{bmatrix}\right)\right)\\
&\quad=\measuredangle\left(\mathcal{R}\left(\begin{bmatrix}\overline{\mathbf{Q}}_{j-1}^{*}\begin{bmatrix}\mathbf{0}_{(j-1)L\times L}\\ \prod_{i=1}^{j-1}\mathbf{Q}_i^{(21)}\end{bmatrix}\\ \mathbf{0}_L\end{bmatrix}\right),\mathcal{R}\left(\overline{\mathbf{Q}}_j^{*}\begin{bmatrix}\mathbf{I}_{jL}\\ \mathbf{0}_{L\times jL}\end{bmatrix}\right)\right).
\end{aligned}$$

Observe now that using the structure of $\mathbf{Q}_j^{(j+1)*}$, we can rewrite

$$\begin{aligned}
\overline{\mathbf{Q}}_j^{*}\begin{bmatrix}\mathbf{I}_{jL}\\ \mathbf{0}_{L\times jL}\end{bmatrix} &= \mathbf{Q}_1^{(j+1)*}\cdots\mathbf{Q}_j^{(j+1)*}\begin{bmatrix}\mathbf{I}_L & & & \\ & \mathbf{I}_L & & \\ & & \ddots & \\ & & & \mathbf{I}_L\\ \mathbf{0}_L & \mathbf{0}_L & \cdots & \mathbf{0}_L\end{bmatrix}\\
&= \mathbf{Q}_1^{(j+1)*}\cdots\mathbf{Q}_{j-1}^{(j+1)*}\begin{bmatrix}\mathbf{I}_L & & & & \\ & \mathbf{I}_L & & & \\ & & \ddots & & \\ & & & \mathbf{I}_L & \\ & & & & \mathbf{Q}_j^{(11)*}\\ \mathbf{0}_L & \mathbf{0}_L & \cdots & \mathbf{0}_L & \mathbf{Q}_j^{(12)*}\end{bmatrix}
\end{aligned}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$$
= \begin{bmatrix} \overline{\mathbf{Q}}_{j-1}^{*} & \\ & \mathbf{I}_L \end{bmatrix} \begin{bmatrix} \mathbf{I}_L & & & & \\ & \mathbf{I}_L & & & \\ & & \ddots & & \\ & & & \mathbf{I}_L & \\ & & & & \mathbf{Q}_j^{(11)*} \\ \mathbf{0}_L & \mathbf{0}_L & \cdots & \mathbf{0}_L & \mathbf{Q}_j^{(12)*} \end{bmatrix},
$$

and similarly we have

$$
\begin{bmatrix} \overline{\mathbf{Q}}_{j-1}^{*} \begin{bmatrix} \mathbf{0}_{(j-1)L\times L} \\ \prod_{i=1}^{j-1} \mathbf{Q}_i^{(21)} \end{bmatrix} \\ \mathbf{0}_L \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{Q}}_{j-1}^{*} & \\ & \mathbf{I}_L \end{bmatrix} \begin{bmatrix} \mathbf{0}_{(j-1)L\times L} \\ \prod_{i=1}^{j-1} \mathbf{Q}_i^{(21)} \\ \mathbf{0}_L \end{bmatrix}.
$$

Thus we have

$$
\begin{aligned}
&\measuredangle\left(\mathcal{R}\left(\begin{bmatrix} \overline{\mathbf{Q}}_{j-1}^{*} \begin{bmatrix} \mathbf{0}_{(j-1)L\times L} \\ \prod_{i=1}^{j-1} \mathbf{Q}_i^{(21)} \end{bmatrix} \\ \mathbf{0}_L \end{bmatrix}\right), \mathcal{R}\left(\overline{\mathbf{Q}}_j^{*} \begin{bmatrix} \mathbf{I}_{jL} \\ \mathbf{0}_{L\times jL} \end{bmatrix}\right)\right) \\
&\quad = \measuredangle\left(\mathcal{R}\left(\begin{bmatrix} \mathbf{0}_{(j-1)L\times L} \\ \prod_{i=1}^{j-1} \mathbf{Q}_i^{(21)} \\ \mathbf{0}_L \end{bmatrix}\right), \mathcal{R}\left(\begin{bmatrix} \mathbf{I}_L & & & & \\ & \mathbf{I}_L & & & \\ & & \ddots & & \\ & & & \mathbf{I}_L & \\ & & & & \mathbf{Q}_j^{(11)*} \\ \mathbf{0}_L & \mathbf{0}_L & \cdots & \mathbf{0}_L & \mathbf{Q}_j^{(12)*} \end{bmatrix}\right)\right).
\end{aligned}
$$

We finish the proof by noting that under the assumptions of this paper, we have that $\mathbf{Q}_i^{(21)}$ is nonsingular for $1 \leq i \leq j-1$ and thus

$$
\mathcal{R}\left(\begin{bmatrix} \mathbf{0}_{(j-1)L\times L} \\ \prod_{i=1}^{j-1} \mathbf{Q}_i^{(21)} \\ \mathbf{0}_L \end{bmatrix}\right) = \mathcal{R}\left(\begin{bmatrix} \mathbf{0}_{(j-1)L\times L} \\ \mathbf{I}_L \\ \mathbf{0}_L \end{bmatrix}\right).
$$

Thus the cosines of the principal angles are the singular values of

$$
\begin{bmatrix} \mathbf{0}_{L\times(j-1)L} & \mathbf{I}_L & \mathbf{0}_L \end{bmatrix} \begin{bmatrix} \mathbf{I}_L & & & & \\ & \mathbf{I}_L & & & \\ & & \ddots & & \\ & & & \mathbf{I}_L & \\ & & & & \mathbf{Q}_j^{(11)*} \\ \mathbf{0}_L & \mathbf{0}_L & \cdots & \mathbf{0}_L & \mathbf{Q}_j^{(12)*} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{L\times(j+1)L} & \mathbf{Q}_j^{(11)*} \end{bmatrix},
$$

which are indeed the CS-decomposition cosines, which are the diagonal entries of $\boldsymbol{\mathcal{C}}$ from (3.19), completing the proof. ∎

4. Numerical examples. We constructed two toy examples using a matrix considered, e.g., in [5], to demonstrate stagnation properties. Let $\mathbf{A}_{st} \in \mathbb{R}^{n\times n}$ be defined as the matrix

ETNA
Kent State University and
Johann Radon Institute (RICAM)

which acts upon the Euclidean basis as follows,

$$\mathbf{A}_{st}\mathbf{e}_i^{[n]} = \begin{cases} \mathbf{e}_1^{[n]} & \text{if} \quad i = n \\ \mathbf{e}_{i+1}^{[n]} & \text{otherwise} \end{cases} . \tag{4.1}$$

From this matrix and appropriately chosen right-hand sides, we can generate problems for which block GMRES is guaranteed to have certain stagnation properties.

In order to obtain some example convergence results in a less non-pathological case, we also applied block GMRES and FOM to a block diagonal matrix built from $\mathbf{A}_{st}$ and the `sherman4` matrix from a discretized oil flow problem, downloaded from the University of Florida Sparse Matrix Library [9]. The latter matrix is 1104×1104 and nonsymmetric.

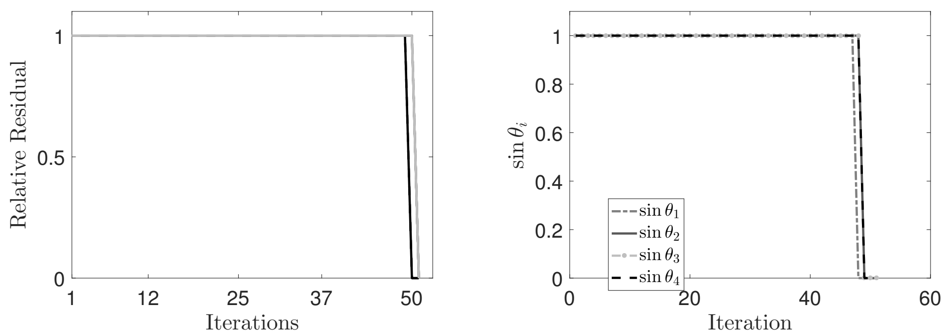

FIG. 4.1. *Left: Relative two-norm residual curves for stagnating block GMRES and block FOM for the* 200×200 *shift matrix for four right-hand sides, namely* $\mathbf{e}_1$*,* $\mathbf{e}_{50}$*,* $\mathbf{e}_{100}$*, and* $\mathbf{e}_{150}$*. The solid and dashed curves correspond respectively to the block FOM and GMRES residuals, with each shade of gray representing a different right-hand side. Similarly, the gray solid and dashed curves, respectively, correspond to the second right-hand side. Right: Sines of principal angles between* $\mathcal{R}\left(\mathbf{F}_{j-1}^{(G)}\right)$ *and* $\mathbf{A}\mathcal{K}_j(\mathbf{A}, \mathbf{F}_0)$ *for each* j*.*

4.1. Total stagnation of block GMRES. Using the shift matrix $\mathbf{A}_{st}$ with $n = 200$, we can construct a problem with four right-hand sides which will stagnate for 50 iterations before converging exactly. Let the four right-hand sides be the canonical basis vectors $\mathbf{e}_1^{[200]}$, $\mathbf{e}_{50}^{[200]}$, $\mathbf{e}_{100}^{[200]}$, and $\mathbf{e}_{150}^{[200]}$. If we let $\mathbf{B} \in \mathbb{R}^{200\times 4}$ be the matrix with these right-hand sides as columns, we know that

$$\mathbf{A}_{st}^{-1}\mathbf{B} = \begin{bmatrix} \mathbf{e}_{200}^{[200]} & \mathbf{e}_{49}^{[200]} & \mathbf{e}_{99}^{[200]} & \mathbf{e}_{149}^{[200]} \end{bmatrix}.$$

Due to the stagnating nature of block GMRES for this problem, we compute the generalized FOM approximation so as to have an iterate at each step. The total stagnation for all four right-hand sides can be seen in Figure 4.1.

If we arrest the iteration at a stagnating step, e.g., the 40th step, we can construct the matrices $\widetilde{\mathbf{C}}_{40}$, $\mathbf{C}_{40}$, $\widehat{\mathbf{C}}_{40}$, $\mathbf{N}_{40}$, and $\widehat{\mathbf{N}}_{40}$ (all of which are 4×4 matrices) to see how such matrices, used to verify theoretical results, actually look for a small problem. For the first three matrices, we have the following,

$$\widehat{\mathbf{C}}_{40} = \widetilde{\mathbf{C}}_{40} = \begin{bmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{C}_{40} = \mathbf{0}_4,$$

and for the last two matrices we have,

$$\mathbf{N}_{40} = -\mathbf{I}_4 \qquad \text{and} \qquad \widehat{\mathbf{N}}_{40} = \mathbf{0}_4.$$

It should be noted that this agrees with what we have proven about block GMRES in the case of total stagnation in Theorem 3.16 and trivially with Theorem 3.22.

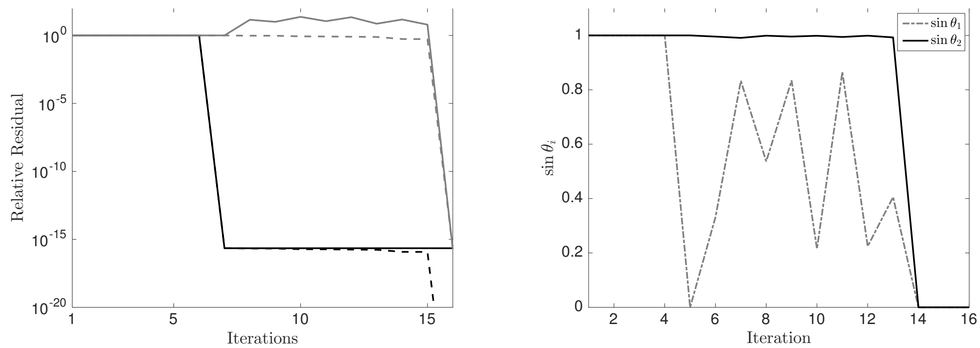

FIG. 4.2. *Left: relative two-norm residual curves for stagnating block GMRES and block FOM for the* 30×30 *shift matrix. The black solid and dashed curves correspond respectively to the block FOM and GMRES residuals for the first right-hand side. Similarly, the gray solid and dashed curves, respectively, correspond to the second right-hand side. Right: sines of principal angles between* $\mathcal{R}\left(\mathbf{F}_{j-1}^{(G)}\right)$ *and* $A\mathcal{K}_j(\mathbf{A}, \mathbf{F}_0)$ *for each* j.

4.2. Partial stagnation/convergence of block GMRES. In Figure 4.2, we demonstrate the behavior of block GMRES and block FOM applied to a linear system for which block GMRES is guaranteed to stagnate but also have earlier convergence for one right-hand side. Here, the coefficient matrix is $\mathbf{A}_{st}$ defined in (4.1) for $n = 30$. The block right-hand side is $\mathbf{B} = \begin{bmatrix} \mathbf{e}_1^{[30]} & \mathbf{e}_{25}^{[30]} \end{bmatrix}$. From the definition of $\mathbf{A}_{st}$, we have that $\mathbf{A}_{st}^{-1}\mathbf{B} = \begin{bmatrix} \mathbf{e}_{30}^{[30]} & \mathbf{e}_{24}^{[30]} \end{bmatrix}$. From this we see that at iteration 5, we will achieve exact convergence for the first right-hand side. In the absence of replacing the dependent Arnoldi vector with a random one, the iteration will not produce any improvement for the second right-hand side until iteration 23, at which point we again have convergence to the exact solution. However, in accordance with our block Arnoldi breakdown strategy, we do replace the the dependent basis vector, meaning we cannot exactly predict stagnation after iteration 5, though we do see near-stagnation until convergence at iteration 15.

Again, at particular iterations, we can inspect various quantities arising which were used in our analysis. We choose three iterations, $j = 5, 6, 11$, to see what happens at breakdown and dependent vector replacement. Indeed we have,

$$\widetilde{\mathbf{C}}_5 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \mathbf{C}_5 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \text{and} \quad \widehat{\mathbf{C}}_5 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$\widetilde{\mathbf{C}}_6 = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}, \qquad \mathbf{C}_6 = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}, \qquad \text{and} \quad \widehat{\mathbf{C}}_6 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$\widetilde{\mathbf{C}}_{11} \approx \begin{bmatrix} 0.97 & 0 \\ -0.22 & 0 \end{bmatrix}, \quad \mathbf{C}_{11} \approx \begin{bmatrix} 4.16 \times 10^{-17} & 0 \\ -2.22 \times 10^{-16} & 0 \end{bmatrix}, \quad \text{and} \quad \widehat{\mathbf{C}}_{11} = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix},$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

and we also have

$$\mathbf{N}_5 \approx \begin{bmatrix} -1.00 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{N}}_5 \approx \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\mathbf{N}_6 \approx \begin{bmatrix} -1.00 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{N}}_6 \approx \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

$$\mathbf{N}_{11} \approx \begin{bmatrix} -1.00 & -0.02 \\ 0 & -0.96 \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{N}}_{11} \approx \begin{bmatrix} -0.30 & 0.79 \\ 0 & -1.11 \times 10^{-16} \end{bmatrix}.$$

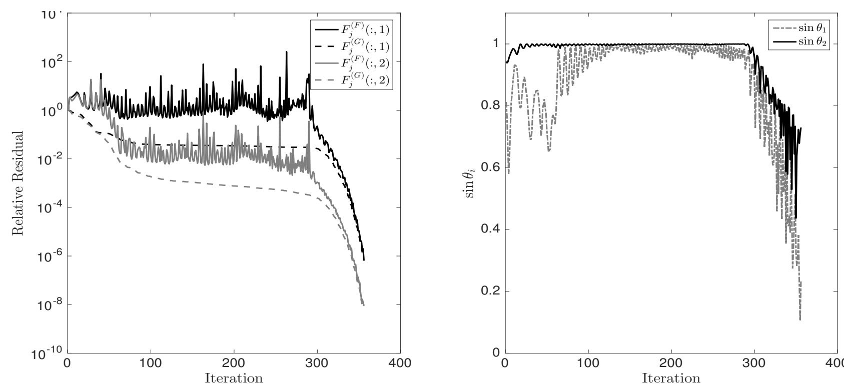

FIG. 4.3. *In the left-hand figure, we have the* 2*-norm residual curves of block GMRES and FOM for a linear system with two right-hand sides using a block diagonal matrix with the* `sherman4` *matrix from [9] as one block and the shift matrix from the other examples as the other block. Right-hand sides are chosen to produce wanted near-stagnation. In the right-hand figure, we have the squares of the sines* $\{s_1^2, s_2^2\}$ *coming from the orthogonal transformations as discussed in our analysis.*

4.3. A less pathological example with sine computation. To stimulate some slightly more interesting near stagnation behavior, we created a block diagonal matrix in which one block is `sherman4` matrix and the other block is the shift matrix $\mathbf{A}_{st}$ used in earlier experiments, this time with $n = 200$. The two right-hand sides are chosen to produce perfect stagnation in the shift-matrix block but convergence in the `sherman4` block. Therefore, in the blocks associated to $\mathbf{A}_{st}$, the subvectors of the right-hand sides were $\mathbf{e}_{50}^{[200]}$ and $\mathbf{e}_{150}^{[200]}$. For the `sherman4` matrix, the subvectors of the right-hand sides were the vector packaged with the matrix and a random vector scaled to have norm on the order of 10^7. The exaggerated scaling was done only to produce visually significant convergence prior to stagnation. In Figure 4.3, we show the individual 2-norm block FOM and block GMRES residual curves as well as the sines from the analysis in Section 3.3.

5. Conclusions. In this paper, we have analyzed the relationship of block GMRES and block FOM and specifically characterized this relation in the case of block GMRES stagnation. These results generalize previous results, particularly those in [5] for single vector GMRES and FOM. We have seen that the relationship can be a bit more complicated for block methods than in the single-vector method case due to interaction between approximations for different right-hand sides and due to block Arnoldi breakdown. We close by noting that one can implement block GMRES so that these sines and cosines are cheaply computable simply by following the strategy advocated in [19] observing that one could implement a version of block GMRES which also cheaply generates the block FOM approximation.

Acknowledgment. We express our gratitude to Mykhaylo Yudytskiy for suggesting the operator approach for the Moore-Penrose pseudo-inverse. We thank Martin Gutknecht for offering many pointers to references on this topic, and Daniel B. Szyld for offering helpful comments on the exposition of this paper. The author also notes that the suggestion to use the CS-decomposition came from Andreas Frommer while the author was visiting TU-Wuppertal. We also thank the two reviewers for offering helpful comments about the language and for sifting through all the mathematical details.

REFERENCES

[1] E. AGULLO, L. GIRAUD, AND Y.-F. JING, *Block GMRES method with inexact breakdowns and deflated restarting*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1625–1651.
[2] J. I. ALIAGA, D. L. BOLEY, R. W. FREUND, AND V. HERNÁNDEZ, *A Lanczos-type method for multiple starting vectors*, Math. Comp., 69 (2000), pp. 1577–1601.
[3] J. BAGLAMA, *Dealing with linear dependence during the iterations of the restarted block Lanczos methods*, Numer. Algorithms, 25 (2000), pp. 23–36.
[4] S. BIRK AND A. FROMMER, *A deflated conjugate gradient method for multiple right hand sides and multiple shifts*, Numer. Algorithms, 67 (2014), pp. 507–529.
[5] P. N. BROWN, *A theoretical comparison of the Arnoldi and GMRES algorithms*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 58–78.
[6] A. T. CHRONOPOULOS AND A. B. KUCHEROV, *Block-s-step Krylov iterative methods*, Numer. Linear Algebra Appl., 17 (2010), pp. 3–15.
[7] J. CULLUM AND A. GREENBAUM, *Relations between Galerkin and norm-minimizing iterative methods for solving linear systems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 223–247.
[8] J. K. CULLUM, *Peaks, plateaus, numerical instabilities in a Galerkin minimal residual pair of methods for solving* $Ax = b$, Appl. Numer. Math., 19 (1995), pp. 255–278.
[9] T. A. DAVIS AND Y. HU, *The University of Florida sparse matrix collection*, ACM Trans. Math. Software, 38 (2011), Art. 1, (25 pages).
[10] J. DUINTJER TEBBENS AND G. MEURANT, *On the convergence of Q-OR and Q-MR Krylov methods for solving nonsymmetric linear systems*, BIT, 56 (2016), pp. 77–97.
[11] M. EIERMANN AND O. G. ERNST, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numer., 10 (2001), pp. 251–312.
[12] L. ELBOUYAHYAOUI, A. MESSAOUDI, AND H. SADOK, *Algebraic properties of the block GMRES and block Arnoldi methods*, Electron. Trans. Numer. Anal., 33 (2008/09), pp. 207–220. http://etna.ricam.oeaw.ac.at/vol.33.2008-2009/pp207-220.dir/pp207-220.pdf
[13] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
[14] R. W. FREUND AND M. MALHOTRA, *A block QMR algorithm for non-Hermitian linear systems with multiple right-hand sides*, Linear Algebra Appl., 254 (1997), pp. 119–157.
[15] R. W. FREUND AND N. M. NACHTIGAL, *QMR: a quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp. 315–339.
[16] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, 2013.
[17] M. H. GUTKNECHT, *Block Krylov space methods for linear systems with multiple right-hand sides: an introduction*, in Modern Mathematical Models, Methods and Algorithms for Real World Systems, A. H. Siddiqi, I. S. Duff, and O. Christensen, eds., Anamaya Publishers, New Delhi, 2007, pp. 420–447.
[18] M. H. GUTKNECHT AND M. ROZLOŽNÍK, *By how much can residual minimization accelerate the convergence of orthogonal residual methods?*, Numer. Algorithms, 27 (2001), pp. 189–213.
[19] M. H. GUTKNECHT AND T. SCHMELZER, *A QR–decomposition of block tridiagonal matrices generated by the block Lanczos process*, in Proceedings of the 17th IMACS World Congress, P. Borne, ed., Ecole Central de Lille, Villeneuve d'Ascq, France, 2005, pp. 1–8.
[20] ———, *The block grade of a block Krylov space*, Linear Algebra Appl., 430 (2009), pp. 174–185.
[21] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
[22] Y.-F. JING, L. GIRAUD, B. CARPENTIERI, AND T.-Z. HUANG, *Recycling block flexible GMRES method with inexact breakdowns for sequences of multiple right-hand side linear systems*, in preparation, 2016.
[23] J. LANGOU, *Iterative Methods for Solving Linear Systems with Multiple Right-Hand Sides*, PhD. Thesis, INSA, Toulouse, 2003.
[24] D. LOHER, *Reliable nonsymmetric block Lanczos algorithms*, PhD. Thesis, Diss. no. 16337, ETH Zürich, Zürich, 2006.
[25] G. MEURANT, *The complete stagnation of GMRES for* $N \leq 4$, Electron. Trans. Numer. Anal., 39 (2012),

pp. 75–101.
http://etna.ricam.oeaw.ac.at/vol.39.2012/pp75-101.dir/pp75-101.pdf

[26] ———, *Necessary and sufficient conditions for GMRES complete and partial stagnation*, Appl. Numer. Math., 75 (2014), pp. 100–107.

[27] A. A. NIKISHIN AND A. Y. YEREMIN, *Variable block CG algorithms for solving large sparse symmetric positive definite linear systems on parallel computers. I. General iterative scheme*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1135–1153.

[28] D. P. O'LEARY, *The block conjugate gradient algorithm and related methods*, Linear Algebra Appl., 29 (1980), pp. 293–322.

[29] C. C. PAIGE AND M. WEI, *History and generality of the* CS *decomposition*, Linear Algebra Appl., 208/209 (1994), pp. 303–326.

[30] M. L. PARKS, K. SOODHALTER, AND D. B. SZYLD, *A block recycled GMRES method with investigations into aspects of solver performance*, Tech. Report 16-04-04, Dept. of Math., Temple University, Philadelphia, 2016.

[31] M. ROBBÉ AND M. SADKANE, *A convergence analysis of GMRES and FOM methods for Sylvester equations*, Numer. Algorithms, 30 (2002), pp. 71–89.

[32] ———, *Exact and inexact breakdowns in the block GMRES method*, Linear Algebra Appl., 419 (2006), pp. 265–285.

[33] Y. SAAD, *Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices*, Linear Algebra Appl., 34 (1980), pp. 269–295.

[34] ———, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.

[35] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

[36] T. SCHMELZER, *Block Krylov Methods for Hermitian Linear Systems*, Master's Thesis, Dept. of Math., University of Kaiserslautern, Kaiserslautern, 2004.

[37] V. SIMONCINI AND E. GALLOPOULOS, *An iterative method for nonsymmetric systems with multiple right-hand sides*, SIAM J. Sci. Comput., 16 (1995), pp. 917–933.

[38] ———, *Convergence properties of block GMRES and matrix polynomials*, Linear Algebra Appl., 247 (1996), pp. 97–119.

[39] V. SIMONCINI AND D. B. SZYLD, *Recent computational developments in Krylov subspace methods for linear systems*, Numer. Linear Algebra Appl., 14 (2007), pp. 1–59.

[40] K. M. SOODHALTER, *A block MINRES algorithm based on the banded Lanczos method*, Numer. Algorithms, 69 (2015), pp. 473–494.

[41] A. STATHOPOULOS AND K. ORGINOS, *Computing and deflating eigenvalues while solving multiple right-hand side linear systems with an application to quantum chromodynamics*, SIAM J. Sci. Comput. 32 (2010), pp. 439–462.

[42] H. F. WALKER, *Residual smoothing and peak/plateau behavior in Krylov subspace methods*, Appl. Numer. Math., 19 (1995), pp. 279–286. 94).

[43] L. ZHOU AND H. F. WALKER, *Residual smoothing techniques for iterative methods*, SIAM J. Sci. Comput., 15 (1994), pp. 297–312.