ADMISSIBLE AND ATTAINABLE CONVERGENCE BEHAVIOR OF BLOCK ARNOLDI AND GMRES*

MARIE KUBÍNOVÁ† AND KIRK M. SOODHALTER‡

*The second author would like to dedicate this work to the late Richard Timoney, who was very generous with his time in explaining the basics of *-algebras to an uninitiated new colleague*

Abstract. It is well-established that any nonincreasing convergence curve is possible for GMRES and a family of pairs (A, b) can be constructed for which GMRES exhibits a given convergence curve with A having arbitrary spectrum. No analogue of this result has been established for block GMRES, wherein multiple right-hand sides are considered. By reframing the problem as a single linear system over a ring of square matrices, we develop convergence results for block Arnoldi and block GMRES. In particular, we show what convergence behavior is admissible for block GMRES and how the matrices and right-hand sides producing any admissible behavior can be constructed. Moreover, we show that the convergence of the block Arnoldi method for eigenvalue approximation can be almost fully independent of the convergence of block GMRES for the same coefficient matrix and the same starting vectors.

Key words. block Krylov subspace methods, multiple right-hand sides, block GMRES, convergence, spectrum, block companion matrix

AMS subject classifications. 65F10, 65F15

DOI. 10.1137/19M1272469

1. Introduction. The celebrated GMRES algorithm [21] is an effective, widely used iterative method for solving linear systems

$$A\boldsymbol{x} = \boldsymbol{b}, \quad A \in \mathbb{C}^{m \times m}, \quad \boldsymbol{b} \in \mathbb{C}^m,$$

with non-Hermitian coefficient matrices. It has been shown that, in contrast to the case of Hermitian matrices, for non-Hermitian coefficient matrices, we cannot guarantee certain convergence speed based solely on our knowledge of the spectrum. In particular, for any given nonincreasing sequence of positive numbers $f_0 \geq f_1 \geq \cdots \geq f_{n-1}$, one can construct a non-Hermitian matrix with arbitrary spectrum for which GMRES produces residuals whose norms correspond to this sequence [15]. In [19], it was observed this does not mean eigenvalues are meaningless for the convergence of GMRES applied to nonnormal problems. This result simply establishes one extreme of what role eigenvalues can play in the *residual* convergence of GMRES.

If instead we solve multiple systems with the same A (i.e., a system with multiple right-hand sides)

$$AX = B, \quad A \in \mathbb{C}^{m \times m}, \quad B \in \mathbb{C}^{m \times s},$$

a direct generalization of GMRES called block GMRES exists which produces approximate solutions simultaneously for all right-hand sides. We would like to explore if block GMRES residual convergence admits a similar characterization as in [15].

Most often, analysis of block GMRES takes the view of the method as a minimization of each individual residual over a sum of spaces. This enables some basic convergence analysis but such analysis fails to capture the full picture of block GMRES behavior, which is influenced by the interaction between the different right-hand sides. The nature of this interaction is quite difficult to describe when considering block GMRES as a method treating a collection of scalar linear systems.

Indeed, some authors have taken a different approach, discussing these methods in terms of vector blocks in $\mathbb{C}^{m \times s}$, namely, [11] and [24]. We demonstrate here that embracing a totally block view of this iteration greatly simplifies analysis of block GMRES and allows us to obtain clean convergence results.

The aim of the paper is to extend well-known GMRES/Arnoldi convergence results developed in [1, 7, 16, 15] (along with many excellent follow-up papers by subsets of the same authors) to the block case using the framework of [11, 12].

The paper is organized as follows. Section 2 introduces the block methods that will be analyzed and the framework used to perform the analysis. In section 3, we generalize the notion of nonincreasing convergence curve to the block setting. Section 4 provides a characterization of matrices and starting vectors exhibiting prescribed convergence behavior, including discussion of the spectral properties of the obtained coefficient matrices. We summarize our results and formulate open questions in section 5. Throughout the paper, we assume exact arithmetic. We also introduce some specialized notation for this setting in Notation 2.1.

2. Preliminaries. In each step, block Krylov subspace methods look for an approximation of each individual solution in the space

$$(2.1) \qquad \mathcal{K}_k(A, R_0) \equiv \text{colspan}\left\{R_0, AR_0, \dots, A^{k-1}R_0\right\} \quad \text{with} \quad R_0 = B - AX_0.$$

For simplicity of presentation, we assume $X_0 = 0$ leading to $R_0 = B$ throughout the paper.

If the columns of the block vectors $V_1, \dots, V_k$, $V_j \in \mathbb{C}^{m \times s}$, form a basis of the subspace $\mathcal{K}_k(A, B)$, we say that $V_1, \dots, V_k$ is a *block* basis of $\mathcal{K}_k(A, B)$. The block solution X_k can be represented as a *block* linear combination of these block vectors:

$$(2.2) \qquad\qquad X_k = \sum_{i=1}^{k} V_i D_i, \quad D_i \in \mathbb{C}^{s \times s},$$

and we say in this setting that $X_k \in \text{blockspan}\{B, AB, \dots, A^{k-1}B\}$. The particular choice of $\{D_i\}_{i=1}^{k}$ is defined by the conditions which the method imposes on the residual $R_k = B - AX_k$; cf. subsection 2.2. For an overview of block Krylov subspace methods, see, e.g., [20, section 6.12].

Comparing (2.2) with the standard definition of a linear combination of a set of vectors, we see that the $s \times s$ matrices here play the role of complex numbers $\mathbb{C}$. Let $n = \lceil \frac{m}{s} \rceil$. Padding the matrix and the right-hand side by zeros,

$$\widehat{A} \longrightarrow \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \in \mathbb{C}^{ns \times ns}, \quad \widehat{B} \longrightarrow \begin{bmatrix} B \\ 0 \end{bmatrix} \in \mathbb{C}^{ns \times s}, \quad \widehat{X} \longrightarrow \begin{bmatrix} X \\ 0 \end{bmatrix} \in \mathbb{C}^{ns \times s},$$

the unknown X and block GMRES approximations thereof do not change and we can view the new matrix also as an $n \times n$ array of $s \times s$ matrices, and similarly the new right-hand side and the solution as a vector of length n of $s \times s$ matrices. Thus, to

simplify the presentation and without loss of generality (except, perhaps, at the final iteration), we assume hereafter that $m = ns$, and decompose A, B, and X into blocks, which we demonstrate on the following example.

Let $m = 6$ and $s = 2$, then $n = 3$, and we can write

$$
\underbrace{\left[\begin{array}{cc|cc|cc}
a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} & a_{1,6} \\
a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & a_{2,5} & a_{2,6} \\ \hline
a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & a_{3,5} & a_{3,6} \\
a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} & a_{4,6} \\ \hline
a_{5,1} & a_{5,2} & a_{5,3} & a_{5,4} & a_{5,5} & a_{5,6} \\
a_{6,1} & a_{6,2} & a_{6,3} & a_{6,4} & a_{6,5} & a_{6,6}
\end{array}\right]}_{\left[\begin{array}{ccc} A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,1} & A_{3,2} & A_{3,3} \end{array}\right]}
\underbrace{\left[\begin{array}{cc}
x_{1,1} & x_{1,2} \\
x_{2,1} & x_{2,2} \\
x_{3,1} & x_{3,2} \\
x_{4,1} & x_{4,2} \\
x_{5,1} & x_{5,2} \\
x_{6,1} & x_{6,2}
\end{array}\right]}_{\left[\begin{array}{c} X_1 \\ X_2 \\ X_3 \end{array}\right]}
=
\underbrace{\left[\begin{array}{cc}
b_{1,1} & b_{1,2} \\
b_{2,1} & b_{2,2} \\
b_{3,1} & b_{3,2} \\
b_{4,1} & b_{4,2} \\
b_{5,1} & b_{5,2} \\
b_{6,1} & b_{6,2}
\end{array}\right]}_{\left[\begin{array}{c} B_1 \\ B_2 \\ B_3 \end{array}\right]},
$$

where $A_{i,j}$, X_j, and B_i are all 2×2 matrices.

In this paper, we restrict our analysis to the case when $\dim(\mathcal{K}_n(A, B)) = ns$.

2.1. Framework. To generalize the results from standard Krylov subspace methods to block ones, we follow [11, section 2] and replace $\mathbb{C}$ by the noncommutative $*$-algebra $\mathbb{S}$ of complex $s \times s$ matrices.[1] We observe that objects and operations over $\mathbb{C}$ not relying on commutativity have counterparts in $\mathbb{S}$ that are relevant for the analysis of block Krylov subspace methods; some of them are shown in Table 1. We use the notation $\mathbb{S} \simeq \mathbb{C}^{s \times s}$ to say that the elements of $\mathbb{S}$ are from $\mathbb{C}^{s \times s}$ but equipped with additional operations.

We emphasize that to comply with the standard block Krylov subspace method notation, the generalization of positive real numbers denoted by $\mathbb{S}^+$ corresponds to the upper triangular matrices with positive diagonal entries.[2] This choice becomes particularly important for the $*$-algebra formulation of the block Arnoldi algorithm; see subsection 2.2. Note also that in $\mathbb{S}^n$, similarly to $\mathbb{C}^n$, we distinguish between the *block absolute value* $|\cdot|$ and the *block norm* $|||\cdot|||$. Both these operators map to $\mathbb{S}_0^+$; however, the first is defined only for the elements of $\mathbb{S}$, whereas the second one is defined for an arbitrary vector with elements in $\mathbb{S}$; see Table 1.

Notation 2.1. In order to keep our notation clear, we denote matrices over $\mathbb{S}$ with caligraphic letters, vectors over $\mathbb{S}$ by bold capital letters, and elements of $\mathbb{S}$ by unbolded capital letters. Vectors over $\mathbb{C}$ are denoted by lowercase bold letters and scalar elements of $\mathbb{C}$ are denoted by lowercase unbolded letters.

In order to be able to investigate convergence behavior, we generalize the ordering of nonnegative real numbers, denoted by $\mathbb{R}_0^+$, to that of upper triangular complex matrices with real nonnegative diagonal entries, denoted by $\mathbb{S}_0^+$, as follows:

$$(2.3) \quad |A| \overset{\text{Loewner}}{\prec} |B| \iff A^*A \overset{\text{Loewner}}{\prec} B^*B, \text{ and } |A| \overset{\text{Loewner}}{\preceq} |B| \iff A^*A \overset{\text{Loewner}}{\preceq} B^*B,$$

where $\overset{\text{Loewner}}{\prec}$ and $\overset{\text{Loewner}}{\preceq}$ is the Loewner (partial) ordering of Hermitian matrices.[3] Most of the results presented in the upcoming sections can be formulated also in terms of

---

[1] The authors in [11] considered various other subrings of $\mathbb{S}$ to classify different block-type Krylov subspace methods in a common framework.

[2] The standard subalgebra of positive entries is represented by Hermitian positive definite matrices, whose Cholesky factors have a one-to-one correspondence with elements of $\mathbb{S}^+$ used here.

[3] Defined by the relation $A \overset{\text{Loewner}}{\prec} B \iff B - A$ is Hermitian positive definite and $A \overset{\text{Loewner}}{\preceq} B \iff B - A$ is Hermitian positive semidefinite; see [18].

TABLE 1
Entities and operations on $\mathbb{C}$ and their counterparts on $\mathbb{S}$. CholU stands for the upper triangular factor of the Cholesky decomposition.

Standard	Block
$\mathbb{C}$	$\mathbb{S} \simeq \mathbb{C}^{s \times s}$
(commutative *-ring)	(noncommutative *-algebra over $\mathbb{C}$)
$\mathbb{R}^+$	$\mathbb{S}^+ \ldots$ upper-Δ with positive diag. entries
(multiplicative topological group)	(multiplicative topological group)
$\mathbb{R}_0^+$	$\mathbb{S}_0^+ \ldots$ upper-Δ with nonnegative diag. entries
0	singular $s \times s$ matrix
1	$I \in \mathbb{S}$
$a, b \in \mathbb{C}$	$A, B \in \mathbb{S}$
$\|a\| \equiv \sqrt{a^* a} \in \mathbb{R}_0^+$	$\|A\| = \sqrt{A^* A} \equiv \mathrm{cholU}(A^* A) \in \mathbb{S}_0^+$
$\|a\| \in \mathbb{R}^+ \iff a \neq 0$	$\|A\| \in \mathbb{S}^+ \iff A$ nonsingular
$\|a\,\|b\|\| = \|a\|\,\|b\|$	$\|A\,\|B\|\| = \|A\|\,\|B\|$
$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$	$\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{S}^n \simeq \mathbb{C}^{ns \times s}$
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \equiv \boldsymbol{y}^* \boldsymbol{x} \in \mathbb{C}$	$\langle\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\rangle \equiv \boldsymbol{Y}^* \boldsymbol{X} \in \mathbb{S}$
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle^*$	$\langle\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\rangle = \langle\langle \boldsymbol{Y}, \boldsymbol{X} \rangle\rangle^*$
$\langle \boldsymbol{x} a, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle a$	$\langle\langle \boldsymbol{X} A, \boldsymbol{Y} \rangle\rangle = \langle\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\rangle A$
$\langle \boldsymbol{x}, \boldsymbol{y} a \rangle = a^* \langle \boldsymbol{x}, \boldsymbol{y} \rangle$	$\langle\langle \boldsymbol{X}, \boldsymbol{Y} A \rangle\rangle = A^* \langle\langle \boldsymbol{X}, \boldsymbol{Y} \rangle\rangle$
$\|\boldsymbol{x}\| \equiv \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} \in \mathbb{R}_0^+$	$\|\|\boldsymbol{X}\|\| \equiv \sqrt{\langle\langle \boldsymbol{X}, \boldsymbol{X} \rangle\rangle} \in \mathbb{S}_0^+$
$\|\boldsymbol{x}\,\|a\|\| = \|\boldsymbol{x}\|\,\|a\|$	$\|\|\boldsymbol{X}\,\|A\|\|\| = \|\|\boldsymbol{X}\|\|\,\|A\|$
$I_n \in \mathbb{C}^{n \times n} \ldots$ identity matrix	$\boldsymbol{\mathcal{I}}_n \in \mathbb{S}^{n \times n} \simeq \mathbb{C}^{ns \times ns} \ldots$ block identity matrix
$\boldsymbol{e}_k \ldots k$th column of I	$\boldsymbol{E}_k \ldots k$th block column of $\boldsymbol{\mathcal{I}}$
$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ linearly independent	$\{\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n\}$ linearly independent
$\|\sum_{i=1}^n \boldsymbol{v}_i c_i\| = 0 \;\Rightarrow\; c_i = 0 \;\forall i$	$\|\|\sum_{i=1}^n \boldsymbol{V}_i C_i\|\|$ singular $\;\Rightarrow\; C_i$ singular $\forall i$
$\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ orthonormal basis of $\mathbb{C}^n$	$\{\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n\}$ orthonormal basis of $\mathbb{S}^n$
$\langle \boldsymbol{w}_i, \boldsymbol{w}_j \rangle = \delta_{ij}$	$\langle\langle \boldsymbol{W}_i, \boldsymbol{W}_j \rangle\rangle = \delta_{ij} I$
$\boldsymbol{x} = \sum_{k=1}^n \boldsymbol{w}_k \langle \boldsymbol{x}, \boldsymbol{w}_k \rangle \;\; \forall \boldsymbol{x} \in \mathbb{C}^n$	$\boldsymbol{X} = \sum_{k=1}^n \boldsymbol{W}_k \langle\langle \boldsymbol{X}, \boldsymbol{W}_k \rangle\rangle \;\; \forall \boldsymbol{X} \in \mathbb{S}^n$

the decay of the "squares", i.e., the traditional Loewner order. To make the results formally as similar as possible to the classical results for standard GMRES, we have chosen to use the nonincreasing order of the norms rather than their squares.

2.2. Arnoldi-based methods in the new framework. Every Krylov subspace method relies on construction of a well-conditioned (ideally orthonormal) basis of the Krylov subspace. For standard GMRES, this orthonormal basis is computed using the Arnoldi algorithm. For block GMRES, where our aim is to compute an orthonormal block basis (see the last part of Table 1), we use the block Arnoldi

Algorithm 2.1. Block Arnoldi algorithm (run to completion).

Require: $\boldsymbol{\mathcal{A}} \in \mathbb{S}^{n \times n}$, $\boldsymbol{B} \in \mathbb{S}^n$
  $\widetilde{\boldsymbol{V}} = \boldsymbol{B}$
  for $j = 1$ to n **do**
    $H_{j,j-1} = |||\widetilde{\boldsymbol{V}}|||$, $\quad \boldsymbol{V}_j = \widetilde{\boldsymbol{V}} H_{j,j-1}^{-1}$
    $\widetilde{\boldsymbol{V}} = \boldsymbol{\mathcal{A}} \boldsymbol{V}_j$
    for $i = 1$ to j **do**
      $H_{i,j} = \langle\langle \widetilde{\boldsymbol{V}}, \boldsymbol{V}_i \rangle\rangle$
      $\widetilde{\boldsymbol{V}} = \widetilde{\boldsymbol{V}} - \boldsymbol{V}_i H_{i,j}$
    end for
  end for
  return $\boldsymbol{\mathcal{H}} = (H_{i,j})_{i,j=1}^n \in \mathbb{S}^{n \times n}$, $\boldsymbol{\mathcal{V}} = \begin{bmatrix} \boldsymbol{V}_1 & \cdots & \boldsymbol{V}_n \end{bmatrix} \in \mathbb{S}^{n \times n}$

---

algorithm, iteratively producing orthonormal block Arnoldi vectors $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_k$, $k = 1, \ldots, n$, spanning the Krylov subspace (2.1). If we assume no premature breakdown, it can be carried out for n iterations to produce a basis for the space $\mathbb{S}^n$. In Algorithm 2.1, following [11], we provide a pseudocode using the notation introduced in Table 1. There are various strategies for how to handle rank-deficient Arnoldi vectors; see, e.g., [2, 10, 25]. Hereafter, we will however assume that there is no breakdown in Algorithm 2.1.

Taken to the nth iteration, the block Arnoldi algorithm yields the block Arnoldi relation

$$(2.4) \qquad \boldsymbol{\mathcal{A}} \boldsymbol{\mathcal{V}} = \boldsymbol{\mathcal{V}} \boldsymbol{\mathcal{H}}, \qquad \boldsymbol{\mathcal{H}} \in \mathbb{S}^{n \times n} \text{ block upper Hessenberg,}$$

with $\boldsymbol{\mathcal{V}}$ having as columns an orthonormal basis of $\mathbb{S}^n$, and the relation (2.4) represents a full orthogonal Hessenberg factorization of the matrix $\boldsymbol{\mathcal{A}}$. Since the blocks on the subdiagonal of the block Hessenberg matrix $\boldsymbol{\mathcal{H}}$ are from $\mathbb{S}^+$, our definition of $\mathbb{S}^+$ assures that the matrix $\boldsymbol{\mathcal{H}}$ has exactly s subdiagonals and the entries of the last one are positive. Therefore, the block Arnoldi algorithm formulated through the new framework computes the same Arnoldi decomposition as the block Arnoldi algorithm performed in a conventional way; see, e.g., [20, section 6.12]. Terminating Algorithm 2.1 after iteration k produces a basis of the Krylov subspace (2.1).

Using the block Arnoldi method (blArnoldi) for eigenvalue problems, the eigenvalues of $\boldsymbol{\mathcal{A}}$ are in each step k approximated by the Ritz values, i.e., the sk eigenvalues of the kth principal submatrix $\boldsymbol{\mathcal{H}}^{(k)} \in \mathbb{S}^{k \times k}$ of $\boldsymbol{\mathcal{H}}$.[4] For systems of linear algebraic equations, the approximate block solutions $\boldsymbol{X}_k$ are constructed as a linear combination of the computed orthonormal basis of $\mathcal{K}_k(\boldsymbol{\mathcal{A}}, \boldsymbol{B})$, i.e.,

$$\boldsymbol{X}_k = \begin{bmatrix} \boldsymbol{V}_1 & \cdots & \boldsymbol{V}_k \end{bmatrix} \boldsymbol{Y}_k,$$

where the block vector $\boldsymbol{Y}_k \in \mathbb{S}^k$ is obtained from small projected problems. We analyze two different methods: block FOM (blFOM) and block GMRES (blGMRES).

---

[4] By the eigenvalue of a matrix $\boldsymbol{\mathcal{A}}$ we mean the standard eigenvalue, i.e., $\lambda \in \mathbb{C}$ satisfying $\boldsymbol{\mathcal{A}} \boldsymbol{v} = \lambda \boldsymbol{v}$ for some nonzero $\boldsymbol{v} \in \mathbb{C}^{ns}$. We note that considering the linear system over $\mathbb{S}$ induces a notion of a block eigenvalue decomposition, i.e., there exist $\boldsymbol{\Lambda} \in \mathbb{S}$ such that $\boldsymbol{\mathcal{A}} \boldsymbol{V} = \boldsymbol{V} \boldsymbol{\Lambda}$ for some $\boldsymbol{V} \in \mathbb{S}^n$ with linearly independent columns. We do not treat such a notion in this paper as it is highly technical, due in part to the noncommutitivity of multiplication in $\mathbb{S}$. Furthermore, as we show in subsection 4.5, the correct notion of block eigenvalues in this setting is that of *block solvents* of specified λ-matrices.

The blFOM method is a Galerkin method, keeping the individual residuals in each step orthogonal to all previous, i.e.,

$$\boldsymbol{Y}_k^F = \left(\boldsymbol{\mathcal{H}}^{(k)}\right)^{-1} \boldsymbol{E}_1 |||\boldsymbol{R}_0|||.$$

The blGMRES method minimizes the Euclidean norm of each of the individual residuals,[5] i.e.,

$$(2.5) \qquad \boldsymbol{Y}_k^G = \left(\underline{\boldsymbol{\mathcal{H}}}^{(k)}\right)^{\dagger} \boldsymbol{E}_1 |||\boldsymbol{R}_0|||,$$

where $\underline{\boldsymbol{\mathcal{H}}}^{(k)} \in \mathbb{S}^{(k+1) \times k}$ is the upper-left $(k+1) \times k$ block of $\boldsymbol{\mathcal{H}}$ and $\cdot^{\dagger}$ denotes here and hereafter the Moore–Penrose pseudoinverse.

While the blGMRES solution always exists, the blFOM solution is not defined when the matrix $\boldsymbol{\mathcal{H}}^{(k)}$ is singular. In this case, we will consider the generalized blFOM solution and define

$$(2.6) \qquad \boldsymbol{Y}_k^F = \arg\min_{\boldsymbol{Y}} \left\{ \left\|\boldsymbol{E}_{\boldsymbol{k}}^T \boldsymbol{Y}\right\|_F \; ; \; \boldsymbol{Y} = \left(\boldsymbol{\mathcal{H}}^{(k)}\right)^{-} \boldsymbol{E}_1 |||\boldsymbol{R}_0||| \right\},$$

where $\cdot^{-}$ denotes a generalized inverse; see also [26]. Since the true and the generalized FOM solution coincide for nonsingular $\boldsymbol{\mathcal{H}}^{(k)}$, we denote both by $\boldsymbol{X}_k^F$ to simplify the notation.

3. Admissible convergence behavior of blGMRES. In this section, we provide a block definition of admissible convergence behavior of the norm-minimizing method by generalizing some of the well-known relations between the residuals of the norm-minimizing and the Galerkin (residual orthogonalizing) method.

3.1. Block Givens transformation. Each step of standard GMRES requires computation of one new elementary rotation to eliminate the kth subdiagonal entry of $\boldsymbol{\mathcal{H}}$. These rotations facilitate analysis of GMRES convergence and its relationship to FOM [20, section 6.5.7]. For blGMRES, a similar analysis is available using the product of Householder transformations [26], but a block analogue to Givens rotations provides additional and clearer results.

For the elimination in blGMRES, s^2 standard elementary Givens rotations are needed [20, section 6.12]. A product of (elementary) Givens rotations is an orthogonal transformation but loses the properties of rotation, except for the very special case when all the individual rotations commute; see [3, 14, 6]. Therefore, the *block Givens transformation* will not be a true generalization of a Givens rotation but rather a block representation of the product of the s^2 standard elementary Givens rotations.[6,7]

---

[5]The application of the Moore–Penrose pseudoinverse to each column individually gives the least-squares solution for each right-hand side, and applying it to each column individually is equivalent to applying it to the block right-hand side. Further, minimizing the Euclidean norm of individual residuals is also equivalent to minimizing the Frobenius norm of the block residual. This was utilized, for example, in the analysis of blGMRES in [24].

[6]We emphasize that here we are concerned with mathematical properties of such block Givens transformations. In practical computations, the individual Givens sines and cosines are stored, and the product is rarely computed explicitly.

[7]When s is large, the subdiagonal entries of the block Hessenberg matrix are eliminated using Householder reflections. There are two main possible generalizations of Householder reflections to the block case. The first (see [22]) preserves the properties of a reflection but is only able to eliminate the subdiagonal block. The second one (see [23]) is able to eliminate all subdiagonal entries and is a block representation of a product of s standard elementary Householder reflections but does not have properties of a reflection. In any case, they are of little use when generalizing relations from the standard case, where Givens rotations always are performed.

We follow the idea of [17]. Assume first a general orthogonal transformation $\boldsymbol{Q}$ that eliminates an entry of a block vector, i.e.,

$$\boldsymbol{V} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \longrightarrow \boldsymbol{Q}\boldsymbol{V} = \begin{bmatrix} \widetilde{V} \\ 0 \end{bmatrix},$$

where $V_1, V_2 \in \mathbb{S}$ and $|\widetilde{V}| = |||\boldsymbol{V}|||$. If V_2 is invertible, then the unitary matrix $\boldsymbol{Q}$ eliminating the block V_2 can be constructed as

$$(3.1) \qquad \boldsymbol{Q} = \begin{bmatrix} \bar{C} & \bar{S} \\ -S & C \end{bmatrix} := \begin{bmatrix} XZ^* & X \\ -Y & YZ \end{bmatrix},$$

where

$$Z = V_1 V_2^{-1}, \qquad X^*X = (I + Z^*Z)^{-1}, \qquad \text{and} \qquad Y^*Y = (I + ZZ^*)^{-1},$$

which can be verified by simple computation.

Note that there is freedom in the choice of X and Y, since they can be arbitrary right factors of the matrices $(I + Z^*Z)^{-1}$ and $(I + ZZ^*)^{-1}$, respectively. This allows us also to control the nonzero pattern of the matrix $\boldsymbol{Q}$. If we take

$$(3.2) \qquad X = \left(\sqrt{(I + Z^*Z)} \right)^{-*} \qquad \text{and} \qquad Y = \sqrt{(I + ZZ^*)^{-1}},$$

then, due to our definition of $\sqrt{\cdot}$, the matrix $\boldsymbol{Q}$ in (3.1) becomes $(2s+1)$-diagonal. In this case, we can see $\boldsymbol{Q}$ as a product of individual standard Givens rotations applied to $\boldsymbol{V}$ to eliminate the subdiagonal entries. If $V_1, V_2 \in \mathbb{C}$, the choice (3.2) will lead to the standard elementary Givens rotation with $\bar{S}$ and $\bar{C}$ being the complex conjugates of C and S, respectively. This is, however, generally not the case for $V_1, V_2 \in \mathbb{S}$, i.e., $\bar{C} \neq C^*$ and $\bar{S} \neq S^*$. The following proposition explains why the choice (3.2) is important for the solution of (2.5) by factorizing the block upper Hessenberg matrix $\boldsymbol{\mathcal{H}}$ from (2.4).

PROPOSITION 3.1. *Let $\boldsymbol{\mathcal{H}} \in \mathbb{S}^{n \times n}$ be an upper block Hessenberg matrix with blocks $H_{i,j}$, and let further the subdiagonal blocks be upper triangular with positive diagonal entries, i.e., $H_{k+1,k} \in \mathbb{S}^+$, $k = 1, \ldots, n-1$. Define, for $k = 1, \ldots, n-1$, the block Givens transformations $\boldsymbol{\mathcal{G}}^{(k)}$ as*

$$(3.3) \qquad \boldsymbol{\mathcal{G}}^{(k)} = \begin{bmatrix} \boldsymbol{\mathcal{I}}_{k-1} & & & \\ & \bar{C}_k & \bar{S}_k & \\ & -S_k & C_k & \\ & & & \boldsymbol{\mathcal{I}}_{n-k-1} \end{bmatrix},$$

where

$$\begin{bmatrix} \bar{C}_k & \bar{S}_k \\ -S_k & C_k \end{bmatrix} = \begin{bmatrix} \sqrt{(I + Z_k^* Z_k)}^{-*} Z_k^* & \sqrt{(I + Z_k^* Z_k)}^{-*} \\ -\sqrt{(I + Z_k Z_k^*)^{-1}} & \sqrt{(I + Z_k Z_k^*)^{-1}} Z_k \end{bmatrix}, \quad Z_k = H_{k,k}^{(k-1)} H_{k+1,k}^{-1},$$

with $\boldsymbol{\mathcal{H}}^{(0)} = \boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{H}}^{(k)} = \boldsymbol{\mathcal{G}}^{(k)} \boldsymbol{\mathcal{H}}^{(k-1)}$. Define further the unitary matrix $\boldsymbol{Q}$ as

$$\boldsymbol{Q} = \boldsymbol{\mathcal{G}}^{(n-1)} \boldsymbol{\mathcal{G}}^{(n-2)} \cdots \boldsymbol{\mathcal{G}}^{(1)}.$$

Then $\boldsymbol{Q}\boldsymbol{\mathcal{H}}$ is a block upper triangular matrix, with the first $n-1$ block diagonal entries in $\mathbb{S}^+$.

Proof. The unitarity of $\mathcal{Q}$ and the upper block triangular form of $\mathcal{QH}$ follows directly from the construction and discussion above. Further, observe that

$$
\begin{bmatrix} \bar{C}_k & \bar{S}_k \\ -S_k & C_k \end{bmatrix} \begin{bmatrix} H_{k,k}^{(k-1)} \\ H_{k+1,j} \end{bmatrix} = \begin{bmatrix} \Xi \\ 0 \end{bmatrix} \implies \begin{bmatrix} H_{k,k}^{(k-1)} \\ H_{k+1,k} \end{bmatrix} = \begin{bmatrix} \bar{C}_k^* & -S_k^* \\ \bar{S}_k^* & C_k^* \end{bmatrix} \begin{bmatrix} \Xi \\ 0 \end{bmatrix}
$$
$$
\implies H_{k+1,k} = \bar{S}_k^* \Xi
$$
$$
\implies \Xi = \bar{S}_k^{-*} H_{k+1,k}.
$$

Since $\mathbb{S}^+$ is a multiplicative group and both $\bar{S}_k^{-*} \in \mathbb{S}^+$ and $H_{k+1,k} \in \mathbb{S}^+$, also $\Xi \in \mathbb{S}^+$, which yields the desired statement. $\square$

Proposition 3.1 shows that the $n-1$ block Givens transformations $\boldsymbol{\mathcal{G}}_1, \dots, \boldsymbol{\mathcal{G}}_{n-1}$ defined in (3.3), with additional normalization of the last block entry of $\boldsymbol{\mathcal{H}}^{(n-1)}$, enable reduction of the Hessenberg matrix $\boldsymbol{\mathcal{H}}$ to the upper triangular form.

3.2. Residual norms and the peak-plateau relation. In this section, we use the block Givens transformations to generalize some of the well-known results about GMRES and FOM residuals. Note that for the generalized blFOM solution (2.6), the analysis only addresses the component of the true residual that satisfies the desired Galerkin condition, i.e., $\boldsymbol{R}_k^F := \boldsymbol{V}_{k+1}\boldsymbol{V}_{k+1}^*(\boldsymbol{B} - \boldsymbol{\mathcal{A}}\boldsymbol{X}_k^F)$. Since the blGMRES solution always exists, it is defined in the conventional way as $\boldsymbol{R}_k^G := \boldsymbol{B} - \boldsymbol{\mathcal{A}}\boldsymbol{X}_k^G$.

Let $\boldsymbol{R}_0 = \boldsymbol{V}_1 |||\boldsymbol{R}_0|||$ be the initial residual. Note that since we assume no premature breakdown in Algorithm 2.1, it holds that $|||\boldsymbol{R}_k^G|||$ is invertible for $k = 0, \dots, n-1$. We follow [20, section 6.5.7]. Since the kth Givens rotation modifies the right-hand side of the projected problem as

$$
\begin{bmatrix} \bar{C}_k & \bar{S}_k \\ -S_k & C_k \end{bmatrix} \begin{bmatrix} |||\boldsymbol{R}_{k-1}^G||| \\ 0 \end{bmatrix} = \begin{bmatrix} * \\ -S_k|||\boldsymbol{R}_{k-1}^G||| \end{bmatrix}
$$

and since $S_k \in \mathbb{S}^+$, we get that the blGMRES residual satisfies

$$
|||\boldsymbol{R}_k^G||| = S_k|||\boldsymbol{R}_{k-1}^G|||, \qquad \text{which implies} \qquad |||\boldsymbol{R}_k^G||| = S_k \cdots S_1 |||\boldsymbol{R}_0|||.
$$

Using Proposition 3.1, we obtain that the (generalized) blFOM residual satisfies

$$
\text{(3.4)} \qquad |||\boldsymbol{R}_k^F||| = \left| H_{k+1,k} \boldsymbol{E}_k^T \boldsymbol{Y}_k^F \right|
$$
$$
= \left| H_{k+1,k} \left(H_{k,k}^{(k-1)} \right)^\dagger \right| |||\boldsymbol{R}_{k-1}^G|||
$$
$$
= |C_k^\dagger S_k| |||\boldsymbol{R}_{k-1}^G||| = |C_k^\dagger| S_k |||\boldsymbol{R}_{k-1}^G||| = |C_k^\dagger| |||\boldsymbol{R}_k^G|||,
$$

where $C_k^\dagger$ becomes C_k^{-1}, whenever the kth principal submatrix of $\boldsymbol{\mathcal{H}}$, and therefore also $H_{k,k}^{(k-1)}$, is invertible; cf. (2.6).

Since blGMRES and blFOM form a norm-minimizing/Galerkin pair, one expects that they satisfy some form of *peak-plateau relation*; see, e.g., [20, section 6.5.7] or [4]. The following proposition shows that this is indeed the case.

PROPOSITION 3.2. *The residuals of blGMRES and blFOM satisfy*

$$
\text{(3.5)} \qquad \left\langle\left\langle \boldsymbol{R}_k^F, \boldsymbol{R}_k^F \right\rangle\right\rangle^\dagger = \left\langle\left\langle \boldsymbol{R}_k^G, \boldsymbol{R}_k^G \right\rangle\right\rangle^{-1} - \left\langle\left\langle \boldsymbol{R}_{k-1}^G, \boldsymbol{R}_{k-1}^G \right\rangle\right\rangle^{-1}
$$

and

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

Proof. Using (3.4), we have

$$\langle\langle \boldsymbol{R}_k^F, \boldsymbol{R}_k^F \rangle\rangle = |||\boldsymbol{R}_k^G|||^* \, C_k^{\dagger *} \, C_k^{\dagger} \, |||\boldsymbol{R}_k^G||| = |||\boldsymbol{R}_k^G|||^* \, (I - S_k S_k^*)^{\dagger} \, |||\boldsymbol{R}_k^G|||$$
$$= |||\boldsymbol{R}_k^G|||^* \left(I - |||\boldsymbol{R}_k^G||| \, |||\boldsymbol{R}_{k-1}^G|||^{-1} |||\boldsymbol{R}_{k-1}^G|||^{-*} |||\boldsymbol{R}_k^G|||^* \right)^{\dagger} |||\boldsymbol{R}_k^G|||.$$

By taking the pseudoinverse, we obtain

$$\langle\langle \boldsymbol{R}_k^F, \boldsymbol{R}_k^F \rangle\rangle^{\dagger} = |||\boldsymbol{R}_k^G|||^{-1} \left(I - |||\boldsymbol{R}_k^G||| \, |||\boldsymbol{R}_{k-1}^G|||^{-1} |||\boldsymbol{R}_{k-1}^G|||^{-*} |||\boldsymbol{R}_k^G|||^* \right) |||\boldsymbol{R}_k^G|||^{-*}$$
$$= |||\boldsymbol{R}_k^G|||^{-1} |||\boldsymbol{R}_k^G|||^{-*} - |||\boldsymbol{R}_{k-1}^G|||^{-1} |||\boldsymbol{R}_{k-1}^G|||^{-*}$$
$$= \langle\langle \boldsymbol{R}_k^G, \boldsymbol{R}_k^G \rangle\rangle^{-1} - \langle\langle \boldsymbol{R}_{k-1}^G, \boldsymbol{R}_{k-1}^G \rangle\rangle^{-1},$$

which gives (3.5). Applying relation (3.5) recursively gives the second statement. $\quad\square$

Relation (3.5) can be viewed as a generalization of the peak-plateau relation, to which it reduces when $s = 1$. For other relations between blGMRES and blFOM, see also [26].

Also note that $\langle\langle \boldsymbol{R}_k^F, \boldsymbol{R}_k^F \rangle\rangle$ is singular if and only if $|||\boldsymbol{R}_k^F|||$ is singular, i.e., if either the standard FOM iterate does not exist for some of the right-hand sides, or if the individual FOM residuals are linearly dependent. This corresponds to the situation when the residual update in blGMRES is of rank smaller than s; see [26, p. 173]. Such situation will be of interest also in subsection 4.3.

3.3. Admissible convergence behavior of blGMRES. The results of the previous section have some nontrivial consequences for the convergence behavior of blGMRES.

THEOREM 3.3. *The blGMRES residuals satisfy*

$$(3.6) \qquad |||\boldsymbol{R}_0||| \succeq |||\boldsymbol{R}_1^G||| \succeq \cdots \succeq |||\boldsymbol{R}_{n-1}^G||| \succ 0.$$

Proof. Since $\langle\langle \boldsymbol{R}_k^F, \boldsymbol{R}_k^F \rangle\rangle$ is a positive semidefinite matrix, relation (3.5) directly implies that for $k = 1, \ldots, n$

$$\langle\langle \boldsymbol{R}_k^G, \boldsymbol{R}_k^G \rangle\rangle \overset{\text{Loewner}}{\preceq} \langle\langle \boldsymbol{R}_{k-1}^G, \boldsymbol{R}_{k-1}^G \rangle\rangle.$$

Using the generalization of the Loewner ordering from section 2 we can write that

$$(3.7) \qquad |||\boldsymbol{R}_k^G||| \preceq |||\boldsymbol{R}_{k-1}^G|||.$$

Applying this relation recursively, we obtain (3.6). $\quad\square$

We call the sequence $|||\boldsymbol{R}_0|||, |||\boldsymbol{R}_1|||, \ldots, |||\boldsymbol{R}_{n-1}|||$ of elements in $\mathbb{S}^+$ an *admissible convergence behavior* of blGMRES if it satisfies (3.6).

Remark 3.4. Using $\|\boldsymbol{R}_k^G e_j\|^2 = \langle \boldsymbol{R}_k^G e_j, \boldsymbol{R}_k^G e_j \rangle = e_j^T \langle\langle \boldsymbol{R}_k^G, \boldsymbol{R}_k^G \rangle\rangle e_j$, $j = 1, \ldots, s$, relation (3.6) trivially implies monotonic convergence of the size of the individual residuals. However, relation (3.6) is generally stronger, because it takes into account the interresidual relationships, expressed by the off-diagonal entries of the matrices $\langle\langle \boldsymbol{R}_k^G, \boldsymbol{R}_k^G \rangle\rangle$, $k = 0, \ldots, n-1$. If two individual initial residuals (of the same norm) are almost linearly dependent, one cannot expect radically different convergence behaviors for each right-hand side. We will demonstrate this with the following example. Let the initial residuals be almost linearly dependent,

$$\langle\langle \boldsymbol{R}_0^G, \boldsymbol{R}_0^G \rangle\rangle := \begin{bmatrix} 1 & 1 - \varepsilon \\ 1 - \varepsilon & 1 \end{bmatrix}, \quad \varepsilon = 0.01,$$

and let the size of the first residual be decreased to $\sqrt{\varepsilon}$ and of the second one to $\sqrt{1-\varepsilon}$:

$$\langle\langle \boldsymbol{R}_1^G, \boldsymbol{R}_1^G \rangle\rangle := \begin{bmatrix} \varepsilon & p \\ p & 1-\varepsilon \end{bmatrix}, \quad p \text{ unknown.}$$

Then there is no p such that $\langle\langle \boldsymbol{R}_0^G, \boldsymbol{R}_0^G \rangle\rangle \overset{\text{Loewner}}{\succeq} \langle\langle \boldsymbol{R}_1^G, \boldsymbol{R}_1^G \rangle\rangle \overset{\text{Loewner}}{\succ} 0$, and therefore such convergence behavior cannot be exhibited by blGMRES. Conversely, if two initial residuals are orthogonal, (3.7) does not give any further restriction on the convergence curve of the individual residuals.

Remark 3.5. Note that the result of Theorem 3.3 is also very intuitive in the following sense: Any nonincreasing sequence (in the Loewner sense) can be generated through the norms of the orthogonal projections of a (block) vector onto a sequence of embedded subspaces. And vice versa, norms of the orthogonal projections of a (block) vector to any sequence of embedded subspaces will generate a nonincreasing sequence (in the Loewner sense). Since orthogonal projections of the residuals to the residual Krylov subspace are the very essence of the GMRES method, the Loewner ordering of its residual sizes is its inherent property.

The Loewner ordering is also natural to consider when block Krylov methods are formulated via polynomial approximation of the inverse of $\boldsymbol{\mathcal{A}}$. The polynomial with matrix-valued coefficients applied to the right-hand side $\boldsymbol{B}$ to obtain the residual is optimal in the sense of the following theorem. We use the concept of matrix-valued polynomials, which we define in (A.1) and elaborate on in Appendix A, to give the proper generalization of the GMRES residual polynomial minimization.

THEOREM 3.6. *Denote by $\mathcal{P}_k$ the set of all λ-matrices P of degree at most k, with coefficients in $\mathbb{S}$, and satisfying $P(0) = I$; see Appendix* A*. The blGMRES residual norms satisfy*

$$|||\boldsymbol{R}_k^G||| = \operatorname*{argmin}_{P \in \mathcal{P}_k} |||P(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}|||,$$

where $\circ$ is as in (A.5) *and the minimum is considered in the generalized Loewner sense* (2.3).

Proof. From the definition of the blGMRES iterate it follows that $\boldsymbol{R}_k^G = \widehat{P}_k(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}$ for some $\widehat{P}_k \in \mathcal{P}_k$. Therefore, it suffices to show that by changing the polynomial, one can only make the norm of the resulting block vector larger. Any λ-matrix in $\mathcal{P}_k$ can be written as $\widehat{P}_k(\lambda) + \lambda Q_{k-1}(\lambda)$, where Q_{k-1} is a λ-matrix of degree at most $k-1$. Therefore, we can write

$$(3.8) \quad \langle\langle [\widehat{P}_k(\boldsymbol{\mathcal{A}}) + \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}})] \circ \boldsymbol{B}, [\widehat{P}_k(\boldsymbol{\mathcal{A}}) + \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}})] \circ \boldsymbol{B} \rangle\rangle$$
$$= \langle\langle \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}, \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B} \rangle\rangle + \langle\langle \widehat{P}_k(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}, \widehat{P}_k(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B} \rangle\rangle$$
$$+ \langle\langle \widehat{P}_k(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}, \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B} \rangle\rangle + \langle\langle \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}, \widehat{P}_k(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B} \rangle\rangle.$$

Further, we observe that $\boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B} \in \boldsymbol{\mathcal{A}}\mathcal{K}_k(\boldsymbol{\mathcal{A}}, \boldsymbol{B})$. Since the residual $\boldsymbol{R}_k^G$ is orthogonal to this space, we obtain

$$0 = \langle\langle \widehat{P}_k(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}, \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B} \rangle\rangle = \langle\langle \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}, \widehat{P}_k(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B} \rangle\rangle^*.$$

With this, (3.8) becomes

$$\langle\langle[\widehat{P}_k(\boldsymbol{\mathcal{A}}) + \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}})] \circ \boldsymbol{B}, [\widehat{P}_k(\boldsymbol{\mathcal{A}}) + \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}})] \circ \boldsymbol{B}\rangle\rangle$$
$$= \overbrace{\langle\langle\boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}, \boldsymbol{\mathcal{A}}Q_{k-1}(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}\rangle\rangle}^{\succeq 0} + \langle\langle\widehat{P}_k(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}, \widehat{P}_k(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B}\rangle\rangle,$$

which finishes the proof. $\square$

In the following section we show that *any* admissible convergence behavior (3.6) is actually attainable by blGMRES. For better readability, we drop the superscript $\cdot^G$ in the remainder of this paper.

4. Prescribing convergence behavior. In this section, we utilize the new framework for block Krylov subspace methods and generalize some of the results of [1, 7, 16, 15] for standard Arnoldi and GMRES to the block case. We show that, under moderate conditions, we simultaneously can prescribe the residual convergence of blGMRES($\boldsymbol{\mathcal{A}}$,$\boldsymbol{B}$) as well as the spectral properties of $\boldsymbol{\mathcal{A}}$ and the principal submatrices $\boldsymbol{\mathcal{H}}^{(k)}$ of the ultimate upper Hessenberg matrix $\boldsymbol{\mathcal{H}}$ produced by blArnoldi($\boldsymbol{\mathcal{A}}$,$\boldsymbol{B}$). These entities will be prescribed based on the framework introduced in subsection 2.1 and have a slightly different form than in the standard case.

To prescribe the blGMRES residual norms, we require

$$(4.1) \qquad |||\boldsymbol{R}_k||| = F_k, \quad k = 0, \ldots, n-1,$$

for some given sequence $\{F_k\}_{k=0}^{n-1}$, $F_k \in \mathbb{S}^+$, satisfying the admissibility condition, with

$$F_0 \succeq F_1 \succeq F_2 \succeq \cdots \succeq F_{n-1} \succ 0.$$

The spectral properties of $\boldsymbol{\mathcal{A}}$ are in the standard case prescribed through the similarity to the companion matrix corresponding to the characteristic polynomial defined by these eigenvalues; see [15]. We will show that by considering block companion matrices of the form

$$(4.2) \qquad \boldsymbol{\mathcal{C}} = \begin{bmatrix} 0 & & & C_0 \\ I & \ddots & & C_1 \\ & \ddots & 0 & \vdots \\ & & I & C_{n-1} \end{bmatrix},$$

we can proceed analogously in the block case. Note that the matrix $\boldsymbol{\mathcal{C}}$ defined in (4.2) is the block companion matrix corresponding to the λ-matrix

$$(4.3) \qquad M(\lambda) = \lambda^n I - \sum_{k=0}^{n-1} \lambda^k C_k, \quad \lambda \in \mathbb{C}.$$

We will be looking for the matrices $\boldsymbol{\mathcal{A}}$ and the right-hand sides $\boldsymbol{B}$ annihilating the polynomial M, i.e.,

$$(4.4) \qquad M(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{B} = 0,$$

where the operation $\circ$ is as in (A.5). Using $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{V}}^*$ and $\boldsymbol{B} = \boldsymbol{V}_1|||\boldsymbol{R}_0|||$, (4.4) can be rewritten as

$$M(\boldsymbol{\mathcal{H}}) \circ \boldsymbol{E}_1|||\boldsymbol{B}||| = 0.$$

The relation between the spectral properties of $\boldsymbol{A}$ and condition (4.4) is discussed in more detail in subsection 4.5.

Due to [7], it is known that in the standard case, not only the eigenvalues of the matrix $\boldsymbol{A}$ and therefore also the ultimate upper Hessenberg matrix $\boldsymbol{\mathcal{H}}$ can be prescribed, but also the eigenvalues of all the principal block submatrices $\boldsymbol{\mathcal{H}}^{(k)}$ of this Hessenberg matrix, i.e., the Ritz values. In the block case, spectral properties of the submatrices $\boldsymbol{\mathcal{H}}^{(k)}$ will be enforced analogously to the spectral properties of $\boldsymbol{A}$ through

$$(4.5) \qquad M^{(k)}\left(\boldsymbol{\mathcal{H}}^{(k)}\right) \circ \boldsymbol{E}_1 |||\boldsymbol{B}||| = 0, \quad k = 1, \ldots, n-1,$$

where $M^{(k)}$ is defined as

$$(4.6) \qquad M^{(k)}(\lambda) = \lambda^k I - \sum_{j=0}^{k-1} \lambda^j C_j^{(k)}, \quad \lambda \in \mathbb{C}.$$

4.1. Attaining prescribed spectral properties of $\boldsymbol{A}$. Defining the Krylov matrix

$$\boldsymbol{\mathcal{K}} := \begin{bmatrix} \boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \boldsymbol{A}^2\boldsymbol{B} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{B} \end{bmatrix} \in \mathbb{S}^{n \times n},$$

condition (4.4) is equivalent to

$$(4.7) \qquad \boldsymbol{A}\boldsymbol{\mathcal{K}} = \boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{C}},$$

yielding

$$(4.8) \qquad \boldsymbol{A} = \boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{K}}^{-1}, \quad \boldsymbol{B} = \boldsymbol{\mathcal{K}}\boldsymbol{E}_1.$$

Consider the unique QR-decomposition (with separate diagonal scaling matrix $\boldsymbol{D}$) of the matrix $\boldsymbol{\mathcal{K}}$,

$$(4.9) \qquad \boldsymbol{\mathcal{K}} = \boldsymbol{\mathcal{V}}\boldsymbol{D}\boldsymbol{\mathcal{U}},$$

with $\boldsymbol{\mathcal{V}}$ unitary, $\boldsymbol{D}$ nonsingular block diagonal with diagonal entries in $\mathbb{S}^+$, and $\boldsymbol{\mathcal{U}}$ nonsingular upper triangular with unit diagonal entries. With this decomposition, equations (4.8) become

$$(4.10) \qquad \boldsymbol{A} = \boldsymbol{\mathcal{V}}\underbrace{\boldsymbol{D}\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{U}}^{-1}\boldsymbol{D}^{-1}}_{\boldsymbol{\mathcal{H}}}\boldsymbol{\mathcal{V}}^*, \quad \boldsymbol{B} = \boldsymbol{\mathcal{V}}\boldsymbol{E}_1 D_1.$$

Similar to the standard case (see, e.g., [19, section 3]), any product of the form $\boldsymbol{D}\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{U}}^{-1}\boldsymbol{D}^{-1}$ is a block upper Hessenberg matrix with subdiagonal entries in $\mathbb{S}^+$ and vice versa. Moreover, $\boldsymbol{D}\boldsymbol{\mathcal{U}}$ satisfies

$$\boldsymbol{D}\boldsymbol{\mathcal{U}} = \begin{bmatrix} \boldsymbol{E}_1 D_1 & \boldsymbol{\mathcal{H}}\boldsymbol{E}_1 D_1 & \cdots & \boldsymbol{\mathcal{H}}^{n-1}\boldsymbol{E}_1 D_1 \end{bmatrix},$$

where D_1 is the first block entry of $\boldsymbol{D}$.[8]

Using the decomposition (4.10), we provide a complete characterization of matrices and right-hand sides providing prescribed convergence behavior. The convergence

---

[8]Note that contrary to the standard case, for a given upper Hessenberg matrix $\boldsymbol{\mathcal{H}}$, the matrices $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{C}}$ are not defined uniquely, since they depend on the choice of D_1. For the same reason, the scaling by $|||\boldsymbol{B}|||$ cannot be easily omitted in (4.5).

behavior of blArnoldi and blGMRES is unitarily invariant; therefore the choice of $\boldsymbol{\mathcal{V}}$ plays no role in the analysis. Note that from (4.10), we already have that $D_1 = |||\boldsymbol{B}|||$. In the following sections, we determine the $(n-1)(n-2)/2$ block entries of $\boldsymbol{\mathcal{U}}$ and the last $n-1$ diagonal block entries of $\boldsymbol{\mathcal{D}}$ so that the prescribed blGMRES and blArnoldi behavior is met.[9]

4.2. Attaining prescribed blArnoldi convergence. We show that the blArnoldi convergence (4.5) is encoded solely in the upper triangular matrix $\boldsymbol{\mathcal{U}}$.

Using the definition of $M^{(k)}(\lambda)$, condition (4.5) can be rewritten using (4.6) as

$$(4.11) \qquad \left(\boldsymbol{\mathcal{H}}^{(k)}\right)^k \boldsymbol{E}_1 |||\boldsymbol{B}||| = \sum_{j=0}^{k-1} \left(\boldsymbol{\mathcal{H}}^{(k)}\right)^j \boldsymbol{E}_1 |||\boldsymbol{B}||| C_j^{(k)}, \quad k = 1, \ldots, n-1.$$

Since $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{U}}^{-1}\boldsymbol{\mathcal{D}}^{-1}$ and $\boldsymbol{\mathcal{U}}^{-1}\boldsymbol{\mathcal{D}}^{-1}\boldsymbol{E}_1|||\boldsymbol{B}||| = \boldsymbol{E}_1$, we observe that

$$\left(\boldsymbol{\mathcal{H}}^{(k)}\right)^j \boldsymbol{E}_1 |||\boldsymbol{B}||| = \boldsymbol{\mathcal{D}}^{(k)}\boldsymbol{\mathcal{U}}^{(k)} \boldsymbol{E}_{j+1}, \quad j = 0, \ldots, k-1,$$

where $\boldsymbol{\mathcal{D}}^{(k)}$ and $\boldsymbol{\mathcal{U}}^{(k)}$ are the kth principal submatrices of $\boldsymbol{\mathcal{D}}$ and $\boldsymbol{\mathcal{U}}$, respectively. Substituting into (4.11), we have, for $k = 1, \ldots, n-1$,

$$(4.12) \qquad \boldsymbol{\mathcal{H}}^{(k)}\boldsymbol{\mathcal{D}}^{(k)}\boldsymbol{\mathcal{U}}^{(k)} \boldsymbol{E}_k = \boldsymbol{\mathcal{D}}^{(k)}\boldsymbol{\mathcal{U}}^{(k)} \sum_{j=0}^{k-1} \boldsymbol{E}_{j+1} C_j^{(k)}.$$

Using the factorization of $\boldsymbol{\mathcal{H}}$ and the structure of $\boldsymbol{\mathcal{D}}, \boldsymbol{\mathcal{U}}$, and $\boldsymbol{\mathcal{C}}$, the left-hand side of (4.12) can be rewritten as

$$\begin{aligned}
(4.13) \quad \boldsymbol{\mathcal{H}}^{(k)}\boldsymbol{\mathcal{D}}^{(k)}\boldsymbol{\mathcal{U}}^{(k)} \boldsymbol{E}_k &= \left(\begin{bmatrix} \boldsymbol{\mathcal{I}}_k & \boldsymbol{0} \end{bmatrix} \boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{U}}^{-1}\boldsymbol{\mathcal{D}}^{-1} \begin{bmatrix} \boldsymbol{\mathcal{I}}_k \\ \boldsymbol{0} \end{bmatrix}\right) \boldsymbol{\mathcal{D}}^{(k)}\boldsymbol{\mathcal{U}}^{(k)} \boldsymbol{E}_k \\
&= \left(\begin{bmatrix} \boldsymbol{\mathcal{I}}_k & \boldsymbol{0} \end{bmatrix} \boldsymbol{\mathcal{D}}\right) \boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{C}} \left(\boldsymbol{\mathcal{U}}^{-1}\boldsymbol{\mathcal{D}}^{-1} \begin{bmatrix} \boldsymbol{\mathcal{I}}_k \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{\mathcal{D}}^{(k)}\right) \boldsymbol{\mathcal{U}}^{(k)} \boldsymbol{E}_k \\
&= \boldsymbol{\mathcal{D}}^{(k)} \begin{bmatrix} \boldsymbol{\mathcal{I}}_k & \boldsymbol{0} \end{bmatrix} \boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{C}} \begin{bmatrix} \left(\boldsymbol{\mathcal{U}}^{(k)}\right)^{-1} \\ 0 \end{bmatrix} \boldsymbol{\mathcal{U}}^{(k)} \boldsymbol{E}_k \\
&= \boldsymbol{\mathcal{D}}^{(k)} \begin{bmatrix} \boldsymbol{\mathcal{I}}_k & \boldsymbol{0} \end{bmatrix} \boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{C}}\boldsymbol{E}_k \\
&= \boldsymbol{\mathcal{D}}^{(k)} \begin{bmatrix} \boldsymbol{\mathcal{I}}_k & \boldsymbol{0} \end{bmatrix} \boldsymbol{\mathcal{U}}\boldsymbol{E}_{k+1} \\
&= \boldsymbol{\mathcal{D}}^{(k)}\boldsymbol{\mathcal{U}}^{(k)} \begin{bmatrix} \left(\boldsymbol{\mathcal{U}}^{(k)}\right)^{-1} & \boldsymbol{0} \end{bmatrix} \boldsymbol{\mathcal{U}}\boldsymbol{E}_{k+1}.
\end{aligned}$$

Furthermore, using the block inversion formula and the fact that $\boldsymbol{\mathcal{U}}$ is upper triangular with identity matrices on the main diagonal, we have

$$(4.14) \qquad \begin{bmatrix} \left(\boldsymbol{\mathcal{U}}^{(k)}\right)^{-1} & \boldsymbol{0} \end{bmatrix} \boldsymbol{\mathcal{U}}\boldsymbol{E}_{k+1} = \left(- \begin{bmatrix} \boldsymbol{\mathcal{I}}_k & \boldsymbol{0} \end{bmatrix} \left(\boldsymbol{\mathcal{U}}^{(k+1)}\right)^{-1}\right) \boldsymbol{E}_{k+1}.$$

---

[9]In [15, 1], the characterization of matrices and right-hand sides providing prescribed behavior is obtained through the factorization of the matrix corresponding to the Krylov residual subspaces $\boldsymbol{\mathcal{A}}\mathcal{K}_n(\boldsymbol{\mathcal{A}}, \boldsymbol{B})$. This factorization provides an alternative formulation of the results presented in the subsequent sections.

Substituting (4.13) and (4.14) back into (4.12) and premultiplying both sides by the inverse of $\boldsymbol{\mathcal{D}}^{(k)}\boldsymbol{\mathcal{U}}^{(k)}$, we finally obtain

$$(4.15) \qquad -\begin{bmatrix} \boldsymbol{\mathcal{I}}_k & \mathbf{0} \end{bmatrix}\left(\boldsymbol{\mathcal{U}}^{(k+1)}\right)^{-1}\boldsymbol{E}_{k+1} = \sum_{j=0}^{k-1}\boldsymbol{E}_{j+1}C_j^{(k)}, \quad k=1,\dots,n-1.$$

Conditions (4.15) uniquely determine the upper triangular matrix $\boldsymbol{\mathcal{U}}$ as

$$(4.16) \qquad \boldsymbol{\mathcal{U}} = \begin{bmatrix} I & -C_0^{(1)} & -C_0^{(2)} & \cdots & -C_0^{(n-1)} \\ & I & -C_1^{(2)} & \cdots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & I & -C_{n-2}^{(n-1)} \\ & & & & I \end{bmatrix}^{-1}.$$

4.3. Attaining prescribed blGMRES convergence. We show that for a given $\boldsymbol{\mathcal{U}}$, the prescribed blGMRES convergence (4.1) can be achieved by a proper choice of the block diagonal matrix $\boldsymbol{\mathcal{D}}$, which also defines the subdiagonal entries of $\boldsymbol{\mathcal{H}}$.

Let $\boldsymbol{\mathcal{W}}$ be any matrix such that its columns $\boldsymbol{W}_1,\dots,\boldsymbol{W}_k$ form an orthonormal basis of $\boldsymbol{\mathcal{AK}}_k(\boldsymbol{\mathcal{A}},\boldsymbol{B})$, $k=1,\dots,n$. To satisfy (4.1), it has to hold that

$$|\langle\!\langle\boldsymbol{B},\boldsymbol{W}_k\rangle\!\rangle| = \sqrt{\langle\!\langle F_{k-1},F_{k-1}\rangle\!\rangle - \langle\!\langle F_k,F_k\rangle\!\rangle} =: G_k, \quad k=1,\dots,n-1,$$
$$|\langle\!\langle\boldsymbol{B},\boldsymbol{W}_n\rangle\!\rangle| = \sqrt{\langle\!\langle F_{n-1},F_{n-1}\rangle\!\rangle} = F_{n-1} =: G_n.$$

Now we relate the columns of $\boldsymbol{\mathcal{W}}$ to those of $\boldsymbol{\mathcal{V}}$. First, we see that

$$\boldsymbol{\mathcal{K}} = \boldsymbol{\mathcal{W}}\begin{bmatrix} \langle\!\langle\boldsymbol{B},\boldsymbol{W}_1\rangle\!\rangle & I & & \\ \langle\!\langle\boldsymbol{B},\boldsymbol{W}_2\rangle\!\rangle & 0 & \ddots & \\ \vdots & & & I \\ \langle\!\langle\boldsymbol{B},\boldsymbol{W}_n\rangle\!\rangle & & & 0 \end{bmatrix}\begin{bmatrix} I & \\ & \boldsymbol{\mathcal{R}} \end{bmatrix} = \boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{Q}}\begin{bmatrix} G_1 & I & & \\ G_2 & 0 & \ddots & \\ \vdots & & & I \\ G_n & & & 0 \end{bmatrix}\begin{bmatrix} I & \\ & \widehat{\boldsymbol{\mathcal{R}}} \end{bmatrix},$$

where $\boldsymbol{\mathcal{Q}} \in \mathbb{S}^{n\times n}$ is unitary block diagonal and $\boldsymbol{\mathcal{R}},\widehat{\boldsymbol{\mathcal{R}}} \in \mathbb{S}^{(n-1)\times(n-1)}$ are nonsingular upper block triangular matrices. Combining with (4.9), we now have two factorizations of $\boldsymbol{\mathcal{K}}$, i.e.,

$$(4.17) \qquad \boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{Q}}\underbrace{\begin{bmatrix} \widehat{G} & \boldsymbol{\mathcal{I}}_{n-1} \\ G_n & 0 \end{bmatrix}}_{\boldsymbol{\mathcal{G}}}\begin{bmatrix} I & \\ & \widehat{\boldsymbol{\mathcal{R}}} \end{bmatrix} = \boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{U}} \quad \text{with} \quad \widehat{G} := \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_{n-1} \end{bmatrix}.$$

The right-hand side of (4.17) is in the form of the QR-decomposition; we study that of the left-hand side by looking at the structure of the Cholesky factorization of $\boldsymbol{\mathcal{G}}^*\boldsymbol{\mathcal{G}}$. We observe that

$$\begin{bmatrix} \widehat{G}^* & G_n^* \\ \boldsymbol{\mathcal{I}} & 0 \end{bmatrix}\begin{bmatrix} \widehat{G} & \boldsymbol{\mathcal{I}} \\ G_n & 0 \end{bmatrix} = \begin{bmatrix} \langle\!\langle F_0,F_0\rangle\!\rangle & \widehat{G}^* \\ \widehat{G} & \boldsymbol{\mathcal{I}} \end{bmatrix} = \begin{bmatrix} F_0^* & \\ \widehat{G}F_0^{-1} & \boldsymbol{\mathcal{R}}_G^* \end{bmatrix}\begin{bmatrix} F_0 & F_0^{-*}\widehat{G}^* \\ & \boldsymbol{\mathcal{R}}_G \end{bmatrix},$$

where $\mathcal{R}_G$ is the unique upper triangular Cholesky factor of $(\mathcal{I} - \widehat{G} F_0^{-1} F_0^{-*} \widehat{G}^*)$, which is positive definite since $|||\widehat{G} F_0^{-1}||| \preceq I$. There must be equality between the R-factors of the unique QR-decomposition of both sides of (4.17). Note that the matrix $\widehat{\mathcal{R}}$ has arbitrary nonsingular block entries on the diagonal. To obtain the unique R-factor with entries in $\mathbb{S}^+$, further transformation by a block diagonal unitary matrix (here denoted by $\widehat{\Gamma}$) is needed. Using convenient decompositions of the involved matrices, we obtain from (4.17) that

$$\begin{bmatrix} F_0 & F_0^{-*}\widehat{G}^* \\ & \mathcal{R}_G \end{bmatrix} \begin{bmatrix} I & \\ & \widehat{\mathcal{R}} \end{bmatrix} = \begin{bmatrix} I & \\ & \widehat{\Gamma} \end{bmatrix} \mathcal{D}\mathcal{U},$$

$$\begin{bmatrix} F_0 & F_0^{-*}\widehat{G}^* \\ & \mathcal{R}_G \end{bmatrix} \begin{bmatrix} I & \\ & \widehat{\mathcal{R}} \end{bmatrix} = \begin{bmatrix} I & \\ & \widehat{\Gamma} \end{bmatrix} \begin{bmatrix} D_1 & \\ & \widehat{\mathcal{D}} \end{bmatrix} \begin{bmatrix} I & U_{12} \\ & \mathcal{U}_{22} \end{bmatrix},$$

$$\begin{bmatrix} F_0 & F_0^{-*}\widehat{G}^*\widehat{\mathcal{R}} \\ & \mathcal{R}_G\widehat{\mathcal{R}} \end{bmatrix} = \begin{bmatrix} D_1 & D_1 U_{12} \\ & \widehat{\Gamma}\widehat{\mathcal{D}}\mathcal{U}_{22} \end{bmatrix};$$

see also [7, section 3] for the analogue in the standard case.

We proceed by comparing individual block entries. Equality of the first diagonal block entries gives

(4.18) $$D_1 = F_0,$$

which is satisfied trivially. Equality of the second diagonal block entries gives an expression for $\widehat{\mathcal{R}}$

(4.19) $$\widehat{\mathcal{R}} = \mathcal{R}_G^{-1}\widehat{\Gamma}\widehat{\mathcal{D}}\mathcal{U}_{22}.$$

Substituting from (4.18) and (4.19) to the equation given by the off-diagonal entry, we obtain

(4.20) $$F_0^{-*}\widehat{G}^*\mathcal{R}_G^{-1}\widehat{\Gamma}\widehat{\mathcal{D}}\mathcal{U}_{22} = F_0 U_{12},$$

$$\left(\langle\langle F_0, F_0\rangle\rangle^{-1}\widehat{G}^*\mathcal{R}_G^{-1}\right)\left(\widehat{\Gamma}\widehat{\mathcal{D}}\right) = U_{12}\mathcal{U}_{22}^{-1}.$$

We now investigate the objects in (4.20). First, using the block inversion formula on (4.16), we have

$$U_{12}\mathcal{U}_{22}^{-1} = \begin{bmatrix} C_0^{(1)} & \cdots & C_0^{(n-1)} \end{bmatrix}.$$

Second, we apply Lemma B.1 to $\widehat{G} F_0^{-1}$, and together with the definition of $\widehat{G}$, we obtain

$$E_k^T \mathcal{R}_G^{-*}\widehat{G}\langle\langle F_0, F_0\rangle\rangle^{-*} = E_k^T \mathcal{R}_G^{-*}\left(\widehat{G} F_0^{-1}\right)F_0^{-*}$$

$$= Q_k\sqrt{\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}}, \quad |Q_k| = I.$$

This substituting to (4.20) and defining $\widehat{\mathcal{Q}} := \operatorname{diag}(Q_1^*, \ldots, Q_{n-1}^*)$ gives

(4.21)
$$\sqrt{\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}}\,^*\left(E_k^T\widehat{\mathcal{Q}}\widehat{\Gamma}\widehat{\mathcal{D}}E_k\right) = C_0^{(k)}, \quad k = 1, \ldots, n-1.$$

We now have $n-1$ equations for the $n-1$ block entries of the block diagonal matrix $\widehat{\mathcal{Q}}\widehat{\Gamma}\widehat{\mathcal{D}}$. To ensure that a nonsingular $\widehat{\mathcal{Q}}\widehat{\Gamma}\widehat{\mathcal{D}}$ satisfying (4.21) exists, there must be some consistency between $\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}$ and $C_0^{(k)}$. More precisely, it has to hold that, for $k = 1, \ldots, n-1$,

$$\text{Range}\left(\sqrt{\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}}^{\,*}\right) = \text{Range}\left(C_0^{(k)}\right),$$

or alternatively[10]

(4.22) $$\text{Range}\left(\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}\right) = \text{Range}\left(C_0^{(k)}\right).$$

If $\text{Range}(C_0^{(k)}) = \text{Range}(\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}) = \mathbb{C}^s$, then the entry $\boldsymbol{E}_k^T\widehat{\mathcal{D}}\boldsymbol{E}_k$ is defined uniquely as

$$\boldsymbol{E}_k^T\widehat{\mathcal{D}}\boldsymbol{E}_k = \left|\sqrt{\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}}^{-*} C_0^{(k)}\right|$$
$$= \sqrt{\left(C_0^{(k)}\right)^* (\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1})^{-1} C_0^{(k)}}.$$

In other cases satisfying (4.22), there is certain freedom in the components corresponding to the null space of $\sqrt{\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}}^{\,*}$.

In the sense of subsection 3.2, (4.22) implies that the FOM residual norm must satisfy

$$\text{Range}\left(\langle\langle \boldsymbol{R}_k^F, \boldsymbol{R}_k^F\rangle\rangle^\dagger\right) = \text{Range}\left(\langle\langle \boldsymbol{R}_k^F, \boldsymbol{R}_k^F\rangle\rangle\right) = \text{Range}\left(C_0^{(k)}\right).$$

4.4. Final result. In the standard case, (4.22) reduces to the well-known condition that GMRES stagnates if and only if we obtain at least one zero Ritz value. In the block case, (partial) stagnation also appears if and only if $\boldsymbol{\mathcal{H}}^{(k)}$ is singular. But in addition, the rank deficiency of $\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}$, encoding the stagnation, and the rank deficiency of $C_0^{(k)}$, encoding the singularity of $\boldsymbol{\mathcal{H}}^{(k)}$, must have the same structure, i.e., the corresponding matrices must share the same range. If this is the case, we can prescribe the convergence of the block Arnoldi method and the block GMRES method at the same time, as summarized in the following theorem, which generalizes [7, Theorem 3.6].

THEOREM 4.1. *Let* $\left\{M^{(k)}\right\}_{k=1}^n$ *be any sequence of λ-matrices,*

$$M^{(k)}(\lambda) = \lambda^k I - \sum_{j=0}^{k-1} \lambda^j C_j^{(k)},$$

$C_0^{(n)}$ *nonsingular, and let* $\{F_k\}_{k=0}^{n-1}$, $F_k \in \mathbb{S}^+$, *be any sequence satisfying*

$$F_0 \succeq F_1 \succeq F_2 \succeq \cdots \succeq F_{n-1} \succ 0.$$

Under the assumption that the two sequences satisfy the consistency condition

$$\text{Range}\left(\langle\langle F_k, F_k\rangle\rangle^{-1} - \langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1}\right) = \text{Range}\left(C_0^{(k)}\right),$$

---

[10]Using the fact that $\text{Range}(R^*) = \text{Range}(R^*R)$.

the following two assertions are equivalent:
  1. *The residuals of blGMRES($\boldsymbol{\mathcal{A}}$,$\boldsymbol{B}$) satisfy*

$$|||\boldsymbol{R}_k||| = F_k, \quad k = 0, \dots, n-1,$$

  and the principal submatrices of the Hessenberg matrix generated by blArnoldi($\boldsymbol{\mathcal{A}}$,$\boldsymbol{B}$) satisfy

$$M^{(k)}\left(\boldsymbol{\mathcal{H}}^{(k)}\right) \circ \boldsymbol{E}_1 |||\boldsymbol{B}||| = 0, \quad k = 1, \dots, n.$$

  2. *The matrix $\boldsymbol{\mathcal{A}}$ and the starting vector/right-hand side $\boldsymbol{B}$ are of the form*

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{V}}\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{U}}^{-1}\boldsymbol{\mathcal{D}}^{-1}\boldsymbol{\mathcal{V}}^*, \quad \boldsymbol{B} = \boldsymbol{\mathcal{V}}\boldsymbol{E}_1 F_0,$$

  where $\boldsymbol{\mathcal{V}}$ is a unitary matrix, $\boldsymbol{\mathcal{C}}$ is the block companion matrix corresponding to $M^{(n)}$,

$$\boldsymbol{\mathcal{U}} = \begin{bmatrix} I & -C_0^{(1)} & -C_0^{(2)} & \cdots & -C_0^{(n-1)} \\ & I & -C_1^{(2)} & \cdots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & I & -C_{n-2}^{(n-1)} \\ & & & & I \end{bmatrix}^{-1},$$

  and $\boldsymbol{\mathcal{D}}$ is a block diagonal matrix with entries in $\mathbb{S}^+$ satisfying

$$(4.23) \qquad\qquad\qquad\qquad\qquad D_1 = F_0,$$
$$\sqrt{\langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1} - \langle\langle F_{k-2}, F_{k-2}\rangle\rangle^{-1}}^{\,*}(Q_k D_k) = C_0^{(k-1)}, \quad k = 2, \dots, n,$$

  for some $Q_k \in \mathbb{S}, |Q_k| = I$.

  Proof. The proof follows from the construction in subsections 4.1 to 4.3. $\square$

Theorem 4.1 shows that, similar to the standard case, we can prescribe the convergence behavior for blArnoldi independently of the residual convergence of blGMRES, as long as the (partially) stagnating iterations are reflected in the corresponding λ-matrix and vice versa.

Remark 4.2. If we do not prescribe the blGMRES convergence, we only have a condition on $\boldsymbol{\mathcal{U}}$, and $\boldsymbol{\mathcal{D}}$ can be an arbitrary block diagonal matrix with entries in $\mathbb{S}^+$. Similarly, if the blArnoldi convergence is not prescribed, (4.23) only gives a condition on the first row of $(\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{U}})^{-1}$, which has to satisfy

$$\boldsymbol{E}_1^T(\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{U}})^{-1}\boldsymbol{E}_1 = F_0^{-1},$$

$$\boldsymbol{E}_1^T(\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{U}})^{-1}\boldsymbol{E}_k = \sqrt{\langle\langle F_{k-1}, F_{k-1}\rangle\rangle^{-1} - \langle\langle F_{k-2}, F_{k-2}\rangle\rangle^{-1}}^{\,*}Q_k, \quad k = 2, \dots, n;$$

cf. [8, Theorem 1].

Remark 4.3. The Ritz value companion transform $\boldsymbol{\mathcal{U}}$ makes the intermediate λ-matrices $M^{(k)}$ completely independent of the last λ-matrix $M^{(n)}$ and these are all independent of the matrix $\boldsymbol{\mathcal{D}}$ defining the subdiagonal entries of $\boldsymbol{\mathcal{H}}$. In the standard case, this implies that the (often used) residual measure of the Ritz value convergence

$$\left\| A V^{(k)}\boldsymbol{z}_i^{(k)} - V^{(k)}\boldsymbol{z}_i^{(k)}\theta_i^{(k)} \right\| = \left| h_{k+1,k}\boldsymbol{e}_k^T\boldsymbol{z}_i^{(k)} \right|, \quad \left(\theta_i^{(k)}, \boldsymbol{z}_i^{(k)}\right) \text{ an eigenpair of } H^{(k)},$$

may provide little or no useful information about the convergence of the Ritz values to the eigenvalues of A; see [7, pp. 964–965]. With the block generalization of the Jordan form (see, e.g., [13]), a similar conclusion is possible for the block version. Further analysis of this topic is, however, beyond the scope of this paper.

In the next section, we discuss some of the fundamental differences between the results presented in the preceding papers on standard Arnoldi and GMRES and the results regarding their block counterparts presented here.

4.5. The role of polynomials with matrix coefficients. In section 4, the spectral properties of $\boldsymbol{\mathcal{A}}$ and the submatrices $\boldsymbol{\mathcal{H}}^{(k)}$ are prescribed through the λ-matrices $M^{(k)}(\lambda)$. The relation (4.4), equivalent to (4.7), together with the assumption that $\mathcal{K}_n(\boldsymbol{\mathcal{A}}, \boldsymbol{B})$ is of full rank, means that $M(\boldsymbol{\mathcal{A}})$ is zero when evaluated on n linearly independent block vectors, as it satisfies

$$(4.24) \qquad M(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{V}_i = 0 \quad \text{for} \quad \boldsymbol{V}_i = \boldsymbol{\mathcal{A}}^{i-1} \boldsymbol{B}, \quad i = 1, \ldots, n.$$

In the standard case, (4.24) implies $M(A) = 0$, i.e., M is the characteristic polynomial of A, and $M(A) \circ \boldsymbol{v} = 0 \; \forall \boldsymbol{v} \in \mathbb{C}^n$. This can also be seen from

$$M(A) \circ \left(\sum_{i=1}^{n} \boldsymbol{v}_i d_i \right) = \sum_{i=1}^{n} \overbrace{(M(A) \circ \boldsymbol{v}_i)}^{=0} d_i = 0.$$

In other words, requiring $M(A) \circ \boldsymbol{b} = 0$ is equivalent to prescribing the eigenvalues of A.

In the block case, the situation is different. First, $M(\boldsymbol{\mathcal{A}})$ itself cannot be defined, because of the clash of dimensions. Further, despite the fact that any vector $\boldsymbol{V} \in \mathbb{S}^n$ can be written as a block linear combination of $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n$, i.e., $\boldsymbol{V} = \sum_{i=1}^{n} \boldsymbol{V}_i D_i$, (4.24) does *not* imply $M(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{V} = 0 \; \forall \boldsymbol{V} \in \mathbb{S}^n$. This is because

$$M(\boldsymbol{\mathcal{A}}) \circ \left(\sum_{i=1}^{n} \boldsymbol{V}_i D_i \right) = \boldsymbol{\mathcal{A}}^n \left(\sum_{i=1}^{n} \boldsymbol{V}_i D_i \right) - \sum_{k=0}^{n-1} \boldsymbol{\mathcal{A}}^k \left(\sum_{i=1}^{n} \boldsymbol{V}_i D_i \right) C_k$$

$$(4.25) \qquad = \sum_{i=1}^{n} \left(\boldsymbol{\mathcal{A}}^n \boldsymbol{V}_i D_i - \sum_{k=0}^{n-1} \boldsymbol{\mathcal{A}}^k \boldsymbol{V}_i D_i C_k \right)$$

$$(4.26) \qquad \neq \sum_{i=1}^{n} \left(\boldsymbol{\mathcal{A}}^n \boldsymbol{V}_i D_i - \sum_{k=0}^{n-1} \boldsymbol{\mathcal{A}}^k \boldsymbol{V}_i C_k D_i \right) = \sum_{i=1}^{n} \overbrace{(M(\boldsymbol{\mathcal{A}}) \circ \boldsymbol{V}_i)}^{=0} D_i = 0.$$

The transition between (4.25) and (4.26) is possible only when D_i and C_k commute $\forall i$ and k. This makes the standard case different from the block case.

From (4.8), it is clear that the eigenvalues of $\boldsymbol{\mathcal{A}}$ are defined by the eigenvalues of $\boldsymbol{\mathcal{C}}$ and coincide with the latent roots of M; see also Appendix A. For a given block companion matrix $\boldsymbol{\mathcal{C}} \in \mathbb{S}^{n \times n}$, the manifold of block companion matrices similar to $\boldsymbol{\mathcal{C}}$ has dimension $ns^2 - ns$; see, e.g., [9]. Therefore the eigenvalues do not define M uniquely. This can be seen also from the fact that the coefficients $C_0, \ldots, C_{n-1}$ have ns^2 free parameters, while there only are ns eigenvalues.

Instead of focusing on the eigenvalues of $\boldsymbol{\mathcal{A}}$, it is advantageous to remain in the block setting and look at the solvents of the λ-matrix instead; see Appendix A. Prescribing the (chain of) solvents $S_1, \ldots, S_n$ of the λ-matrix (4.3), the coefficients of the λ-matrix M are defined uniquely through (A.4). The eigenvalues of $\boldsymbol{\mathcal{A}}$ are the

eigenvalues of the solvents, but changing the eigenvectors of the individual solvents will have an impact on the eigenvectors of $\mathcal{A}$. To achieve C_0 nonsingular, we only require that each of the solvents $S_1, \ldots, S_n$ of M is nonsingular. A similar result holds for the submatrices $\mathcal{H}^{(k)}$, with the exception that singular solvents are allowed.

Concluding, in the standard case, if the eigenvalues of $\mathcal{A}$ and the convergence behavior of both GMRES and Arnoldi are prescribed, then, in case of no stagnation of GMRES, all matrices and right-hand sides satisfying these conditions will be identical up to a unitary transform. Each step in which GMRES stagnates provides one extra free parameter, represented by the entry of the diagonal matrix $\mathcal{D}$ in (4.23). In the block case, prescribing the eigenvalues of $\mathcal{A}$ and the convergence behavior of both blGMRES and blArnoldi will still give us certain freedom in the choice of the eigenvectors of $\mathcal{A}$, plus again some extra free parameters in the stagnating iterations. From this point of view, the eigenvalues of $\mathcal{A}$ are even less indicative regarding the residual convergence behavior of blArnoldi and blGMRES than they are in the standard case.

5. Conclusions and open questions. The analysis of block Krylov subspace methods has always presented a challenge beyond those encountered with classical nonblock methods. This is due to the interaction between the right-hand sides. We have demonstrated here that the $*$-algebra approach introduced in [11] enables us to cleanly obtain the same sort of results one sees for nonblock Krylov subspace methods, which have previously been unavailable. These results fill a certain gap in the understanding of the convergence behavior of block Krylov methods for non-symmetric matrices.

We have thus obtained block versions of the fundamental results regarding the admissible and attainable convergence of standard Arnoldi and GMRES presented in a series of papers published over the last 25 years. Extending the framework introduced in [11], we were able to keep the formal notation as close as possible to the original results, with the block generalization of the Givens transformation allowing for extension of well-known textbook relations for the residuals of GMRES and FOM to the block case. This framework allows us to see the interdependence of the residual convergence behaviors of the individual systems. We explicitly formulated conditions on the admissible convergence behavior of the residuals of blGMRES using an appropriate block generalization of the norm.

Under the assumption that blGMRES does not converge prematurely, we were then able to completely characterize matrices and right-hand sides producing any prescribed admissible convergence behavior. Furthermore, spectral properties of the matrix can be enforced through the similarity to a block companion matrix. We showed that arbitrary convergence of blArnoldi for the eigenvalue problem is possible and that arbitrary convergence of blGMRES and blArnoldi can be, under moderate assumptions, achieved simultaneously. Combining these results with the theory of block companion matrices, we showed that in a certain sense, increasing the number of right-hand sides reduces the predictive value of the eigenvalues of $\mathcal{A}$.

It should be noted that, as Meurant pointed out in [19], these results all concern *residual* convergence behavior. It is observed that for a class of matrices (each with different spectral properties) constructed to exhibit a specific admissible residual convergence behavior, the actual error convergence behavior may very well exhibit dependence on spectral properties of the matrix. Thus, one interpretation of the results presented in this paper and the work of the last 25 years on this topic is that spectral properties of a non-Hermitian matrix may not a priori tell us much about the behavior of GMRES with respect to *our chosen method* of measuring convergence.

Thus one may connect the results in this and related works to the notion that one should measure (residual) error using an appropriate norm. However, this is beyond the scope of the current paper.

Certain important aspects of blGMRES and blArnoldi convergence also are beyond the scope of this paper. Since the Arnoldi algorithm is based on full orthogonalization, the amount of data that needs to be stored, as well as the computational complexity, grows with each iteration. In practical computations, restarting the orthogonalization process is therefore often unavoidable. For this reason, analysis of the influence of the restarts on the admissible convergence represents another important research direction. Furthermore, the blArnoldi process breaks down when fewer than s linearly independent Arnoldi vectors are generated in the kth step. The situation when no single system has converged but rather a linear combination of the columns of X lies in $\mathcal{K}_k(\mathcal{A}, B)$ is particularly unpleasant. In the considered framework, resolving this situation by reducing the block size is not directly possible, since by this, we change $\mathbb{S}$. To avoid change in the block size, the linearly dependent vectors can be replaced by some auxiliary (random) vectors. This direction will be further explored elsewhere.

Appendix A. Polynomials with matrix coefficients. In this section, we recall some terminology and fundamental results from the theory of polynomials with matrix coefficients. Statements presented in this section are adopted from [5]. For a more comprehensive overview of the topic, we recommend [13].

Let $C_k \in \mathbb{C}^{s \times s}$, $k = 0, \ldots, n-1$. We call

$$(A.1) \qquad M(\lambda) = \lambda^n I - \sum_{k=0}^{n-1} \lambda^k C_k, \quad \lambda \in \mathbb{C},$$

a λ-*matrix* and

$$(A.2) \qquad M(X) = X^n - \sum_{k=0}^{n-1} C_k X^k, \quad X \in \mathbb{C}^{s \times s},$$

a *matrix polynomial*. We call $\lambda \in \mathbb{C}$ a *latent root* of the λ-matrix M if $M(\lambda)$ is singular and $S \in \mathbb{C}^{s \times s}$ a *right solvent* of the matrix polynomial M if $M(S) = 0$. The latent roots and solvents are related as follows.

THEOREM A.1. *If S is a solvent of the matrix polynomial M, then the λ-matrix M can be factorized as*

$$M(\lambda) = Q(\lambda)(I\lambda - S).$$

COROLLARY A.2. *The s eigenvalues of the solvent S of the matrix polynomial M are all latent roots of the λ-matrix M.*

A sequence of matrices $S_1, \ldots, S_n$ forms a *chain of solvents* if

$$(A.3) \qquad M(\lambda) = (I\lambda - S_1)(I\lambda - S_2) \cdots (I\lambda - S_n).$$

It follows directly from (A.3) that

$$\begin{aligned}
C_{n-1} &= S_1 + S_2 + \cdots + S_n, \\
C_{n-2} &= -(S_1 S_2 + S_1 S_3 + \cdots + S_{n-1} S_n), \\
&\;\;\vdots \\
C_0 &= (-1)^{n-1} S_1 S_2 \cdots S_n.
\end{aligned}$$

$(A.4)$

The matrix C, which has the form

$$C = \begin{bmatrix} 0 & & & C_0 \\ I & \ddots & & C_1 \\ & \ddots & 0 & \vdots \\ & & I & C_{n-1} \end{bmatrix},$$

is called the *block companion matrix* associated with the λ-matrix (A.1) or equivalently the matrix polynomial (A.2). Eigenvalues of the block companion matrix C and the latent roots are related as follows.

THEOREM A.3. $\det(C - \lambda I) = (-1)^{ns} \det(I\lambda^n - C_{n-1}\lambda^{n-1} - \cdots - C_0)$.

COROLLARY A.4. *The eigenvalues of the block companion matrix are the latent roots of the associated λ-matrix; therefore M has exactly ns latent roots.*

We define the action of a matrix polynomial on a block vector as

$$(A.5) \qquad M(A) \circ V = A^n V - \sum_{k=0}^{n-1} A^k V C_k, \quad A \in \mathbb{C}^{m \times m}, \quad V \in \mathbb{C}^{m \times s};$$

see [24, p. 108] or [11, p. 107].

Appendix B. Auxiliary lemma. We use MATLAB notation in this lemma. In particular, Z_j denotes the jth block entry and $\boldsymbol{Z}_{1:j}$ denotes the first j block entries of the block vector $\boldsymbol{Z}$.

LEMMA B.1. *Let $\boldsymbol{Z} \in \mathbb{S}^k$ be such that $\boldsymbol{\mathcal{I}} - \boldsymbol{Z}\boldsymbol{Z}^*$ is positive definite, and let $\boldsymbol{\mathcal{R}}_Z$ be the unique upper triangular Cholesky factor of $\boldsymbol{\mathcal{I}} - \boldsymbol{Z}\boldsymbol{Z}^*$. Then*

$$|\boldsymbol{E}_j^T \boldsymbol{\mathcal{R}}_Z^{-*} \boldsymbol{Z}| = \sqrt{(I - \boldsymbol{Z}_{1:j}^* \boldsymbol{Z}_{1:j})^{-1} - (I - \boldsymbol{Z}_{1:j-1}^* \boldsymbol{Z}_{1:j-1})^{-1}}, \quad j = 1, \ldots, k.$$

Proof. Since $\boldsymbol{\mathcal{R}}_Z^{-*}$ is lower triangular, it holds that

$$\boldsymbol{E}_j^T \boldsymbol{\mathcal{R}}_Z^{-*} \boldsymbol{Z} = \boldsymbol{E}_j^T \boldsymbol{\mathcal{R}}_{Z_{1:j}}^{-*} \boldsymbol{Z}_{1:j}.$$

Therefore, it suffices to investigate $\boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z^{-*} \boldsymbol{Z}$ and apply the lemma recursively.

We observe that

$$\begin{aligned} \boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z^{-*} \boldsymbol{Z} = \boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z (\boldsymbol{\mathcal{R}}_Z^* \boldsymbol{\mathcal{R}}_Z)^{-1} \boldsymbol{Z} &= \boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z (\boldsymbol{\mathcal{I}} - \boldsymbol{Z}\boldsymbol{Z}^*)^{-1} \boldsymbol{Z} \\ &= \boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z \boldsymbol{Z} (I - \boldsymbol{Z}^* \boldsymbol{Z})^{-1} \\ &= \boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z \boldsymbol{E}_k \boldsymbol{E}_k^T \boldsymbol{Z} (I - \boldsymbol{Z}^* \boldsymbol{Z})^{-1}, \end{aligned}$$

where we used the push-through identity (Woodbury matrix identity) and the fact that $\boldsymbol{\mathcal{R}}_Z$ is upper triangular.

The block entry $\boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z \boldsymbol{E}_k$, i.e., the (k, k) block entry of the Cholesky factor, can be obtained as

$$\boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z \boldsymbol{E}_k = \sqrt{I - Z_k (I - \boldsymbol{Z}_{1:k-1}^* \boldsymbol{Z}_{1:k-1})^{-1} Z_k^*}.$$

Thus

$$
\langle\langle \boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z^{-*} \boldsymbol{Z}, \boldsymbol{E}_k^T \boldsymbol{\mathcal{R}}_Z^{-*} \boldsymbol{Z}\rangle\rangle
$$
$$
= (I - \boldsymbol{Z}^*\boldsymbol{Z})^{-*} Z_k^* \left(I - Z_k(I - \boldsymbol{Z}_{1:k-1}^*\boldsymbol{Z}_{1:k-1})^{-1}Z_k^*\right) Z_k (I - \boldsymbol{Z}^*\boldsymbol{Z})^{-1}
$$
$$
= (I - \boldsymbol{Z}^*\boldsymbol{Z})^{-1} Z_k^* Z_k \left(I - (I - \boldsymbol{Z}_{1:k-1}^*\boldsymbol{Z}_{1:k-1})^{-1}Z_k^* Z_k\right) (I - \boldsymbol{Z}^*\boldsymbol{Z})^{-1}
$$
$$
= (I - \boldsymbol{Z}^*\boldsymbol{Z})^{-1} Z_k^* Z_k (I - \boldsymbol{Z}_{1:k-1}^*\boldsymbol{Z}_{1:k-1})^{-1} (I - \boldsymbol{Z}^*\boldsymbol{Z}) (I - \boldsymbol{Z}^*\boldsymbol{Z})^{-1}
$$
$$
= (I - \boldsymbol{Z}^*\boldsymbol{Z})^{-1} \left((I - \boldsymbol{Z}_{1:k-1}^*\boldsymbol{Z}_{1:k-1}) - (I - \boldsymbol{Z}^*\boldsymbol{Z})\right) (I - \boldsymbol{Z}_{1:k-1}^*\boldsymbol{Z}_{1:k-1})^{-1}
$$
$$
= (I - \boldsymbol{Z}^*\boldsymbol{Z})^{-1} - (I - \boldsymbol{Z}_{1:k-1}^*\boldsymbol{Z}_{1:k-1})^{-1},
$$

which gives the desired statement.   □

Note that $\boldsymbol{\mathcal{I}} - \boldsymbol{Z}\boldsymbol{Z}^*$ is positive definite if and only if $\boldsymbol{\mathcal{I}} - \boldsymbol{Z}^*\boldsymbol{Z}$ is positive definite.

REFERENCES

[1] M. ARIOLI, V. PTÁK, AND Z. STRAKOŠ, *Krylov sequences of maximal length and convergence of GMRES*, BIT, 38 (1998), pp. 636–643, https://doi.org/10.1007/BF02510405.

[2] 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, https://doi.org/10.1007/s11075-013-9805-9.

[3] Å. BJÖRCK, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996, https://doi.org/10.1137/1.9781611971484.

[4] 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, https://doi.org/10.1137/S0895479893246765.

[5] J. E. DENNIS, JR., J. F. TRAUB, AND R. P. WEBER, *The algebraic theory of matrix polynomials*, SIAM J. Numer. Anal., 13 (1976), pp. 831–845, https://doi.org/10.1137/0713065.

[6] J. DUINTJER TEBBENS, I. HNĚTYNKOVÁ, M. PLEŠINGER, Z. STRAKOŠ, AND P. TICHÝ, *Analýza metod pro maticové výpočty*, MatfyzPress, Prague, 2011.

[7] J. DUINTJER TEBBENS AND G. MEURANT, *Any Ritz value behavior is possible for Arnoldi and for GMRES*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 958–978, https://doi.org/10.1137/110843666.

[8] J. DUINTJER TEBBENS AND G. MEURANT, *Prescribing the behavior of early terminating GMRES and Arnoldi iterations*, Numer. Algorithms, 65 (2014), pp. 69–90, https://doi.org/10.1007/s11075-013-9695-x.

[9] A. EDELMAN AND H. MURAKAMI, *Polynomial roots from companion matrix eigenvalues*, Math. Comp., 64 (1995), pp. 763–776, https://doi.org/10.2307/2153450.

[10] R. W. FREUND AND M. MALHOTRA, *A block QMR algorithm for non-Hermitian linear systems with multiple right-hand sides*, in Proceedings of the Fifth Conference of the International Linear Algebra Society (Atlanta, GA, 1995), Linear Algebra Appl. 254, Springer, New York, 1997, pp. 119–157, https://doi.org/10.1016/S0024-3795(96)00529-0.

[11] A. FROMMER, K. LUND, AND D. B. SZYLD, *Block Krylov subspace methods for functions of matrices*, Electron. Trans. Numer. Anal., 47 (2017), pp. 100–126, https://doi.org/10.1553/etna_vol47s100.

[12] A. FROMMER, K. LUND, AND D. B. SZYLD, *Block Krylov Subspace Methods for Functions of Matrices* II: *Modified Block FOM*, Tech. report, MATHICSE, Écublens, 2019, https://doi.org/10.5075/epfl-MATHICSE-265508.

[13] I. GOHBERG, P. LANCASTER, AND L. RODMAN, *Matrix Polynomials*, Classics in Appl. Math. 58, SIAM, Philadelphia, 2009, https://doi.org/10.1137/1.9780898719024.

[14] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 4th ed., Johns Hopkins Stud. Math. Sci., Johns Hopkins University Press, Baltimore, 2013.

[15] A. Greenbaum, V. Pták, and Z. Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469, https://doi.org/10.1137/S0895479894275030.

[16] A. Greenbaum and Z. Strakoš, *Matrices that generate the same Krylov residual spaces*, in Recent Advances in Iterative Methods, G. Golub, M. Luskin, and A. Greenbaum, eds., IMA Vol. Math. Appl. 60, Springer, New York, 1994, pp. 95–118, https://doi.org/10.1007/978-1-4613-9353-5_7.

[17] J. Halleck, *Block Givens*, http://www.netlib.org/na-digest-html/03/v03n35.html.

[18] K. Löwner, *Über monotone matrixfunktionen*, Math. Z., 38 (1934), pp. 177–216, https://doi.org/10.1007/bf01170633.

[19] G. Meurant, *GMRES and the Arioli, Pták, and Strakoš parametrization*, BIT, 52 (2012), pp. 687–702, https://doi.org/10.1007/s10543-012-0374-4.

[20] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003, https://doi.org/10.1137/1.9780898718003.

[21] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869, https://doi.org/10.1137/0907058.

[22] R. Schreiber and B. Parlett, *Block reflectors: Theory and computation*, SIAM J. Numer. Anal., 25 (1988), pp. 189–205, https://doi.org/10.1137/0725014.

[23] R. Schreiber and C. Van Loan, *A storage-efficient WY representation for products of Householder transformations*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 53–57, https://doi.org/10.1137/0910005.

[24] V. Simoncini and E. Gallopoulos, *Convergence properties of block GMRES and matrix polynomials*, Linear Algebra Appl., 247 (1996), pp. 97–119, https://doi.org/10.1016/0024-3795(95)00093-3.

[25] K. M. Soodhalter, *A block MINRES algorithm based on the band Lanczos method*, Numer. Algorithms, 69 (2015), pp. 473–494, https://doi.org/10.1007/s11075-014-9907-z.

[26] K. M. Soodhalter, *Stagnation of block GMRES and its relationship to block FOM*, Electron. Trans. Numer. Anal., 46 (2017), pp. 162–189.