The block rational Arnoldi method

Elsworth, Steven and Güttel, Stefan

2019

MIMS EPrint: **2019.2**

ISSN 1749-9097

THE BLOCK RATIONAL ARNOLDI METHOD

STEVEN ELSWORTH* AND STEFAN GÜTTEL*

Abstract. The block version of the rational Arnoldi method is a widely used procedure for generating an orthonormal basis of a block rational Krylov space. We study block rational Arnoldi decompositions associated with this method and prove an implicit Q theorem. We relate these decompositions to nonlinear eigenvalue problems. We show how to choose parameters to prevent a premature breakdown of the method and improve its numerical stability. We explain how rational matrix-valued functions are encoded in rational Arnoldi decompositions and how they can be evaluated numerically. Two different types of deflation strategies are discussed. Numerical illustrations using the MATLAB Rational Krylov Toolbox are included.

Key words. block rational Krylov, rational matrix-valued function, vector autoregression

AMS subject classifications. 65F25, 65F15, 65F50

1. Introduction.

Block Krylov methods were introduced in the 1970s, starting with the block Lanczos algorithm for linear eigenproblems with repeated eigenvalues [16, 27, 36]. More recently, block Krylov methods have found applications in model order reduction [2, 22, 23], for the solution of matrix equations [6, 19, 31, 33], matrix function approximation [24,37,38,40], including multisource electromagnetic modeling [13,15,42,43], and solving linear systems with multiple right-hand sides [12,14,18,21, 28,41,48]. While the theory of single-vector rational Krylov spaces is well developed [9, 10,44,45,46,47], the block case has only been explored to a limited extent [4,24,29]. A general framework for block polynomial Krylov spaces has recently been proposed by Frommer et al. [24] and Lund [38], exploring the use of block inner products to create different algorithmic variants of block Arnoldi method. Following the terminology introduced by these authors, our focus is on a rational variant of so-called *classical block Krylov spaces*.

Block rational Krylov spaces are closely connected to their polynomial counterparts. To introduce notation, we will include a short review here. Block polynomial Krylov spaces are linear subspaces of $\mathbb{C}^{N\times s}$ built with a matrix $A \in \mathbb{C}^{N\times N}$ and a starting block vector $\boldsymbol{b} = [b_1, \ldots, b_s] \in \mathbb{C}^{N\times s}$ of maximal rank. The associated *(classical) block Krylov space of order* $m+1$ is defined as

$$\begin{aligned}\mathcal{K}^{\square}_{m+1}(A,\boldsymbol{b}) &:= \text{blockspan}\{\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^m\boldsymbol{b}\}\\ &:= \left\{\sum_{k=0}^{m} A^k\boldsymbol{b}C_k : C_k \in \mathbb{C}^{s\times s}\right\}.\end{aligned} \tag{1.1}$$

We have adopted the square superscript notation from [28]. If there is no room for ambiguity, we sometimes omit $(A,\boldsymbol{b})$ for brevity and just write $\mathcal{K}^{\square}_{m+1}$. There exists an integer $M \le N$, called the *invariance index* of $(A,\boldsymbol{b})$ (or *block grade of* $\boldsymbol{b}$ *with respect to* A; see [29]) such that

$$\mathcal{K}^{\square}_{1}(A,\boldsymbol{b}) \subset \mathcal{K}^{\square}_{2}(A,\boldsymbol{b}) \subset \cdots \subset \mathcal{K}^{\square}_{M-1}(A,\boldsymbol{b}) \subset \mathcal{K}^{\square}_{M}(A,\boldsymbol{b}) = \mathcal{K}^{\square}_{M+1}(A,\boldsymbol{b}).$$

* School of Mathematics, The University of Manchester, Alan Turing Building, Oxford Road, M13 9PL Manchester, United Kingdom, steven.elsworth@manchester.ac.uk, stefan.guettel@manchester.ac.uk. This work was supported by Engineering and Physical Sciences Research Council grant EP/N509565/1.

We define the *dimension* d of a block Krylov space $\mathcal{K}_{m+1}^{\square}(A, \boldsymbol{b})$ as the cardinality of a basis over $\mathbb{C}^{N\times s}$, that is, there exist d block vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d \in \mathcal{K}_{m+1}^{\square}$ such that every $\boldsymbol{v} \in \mathcal{K}_{m+1}^{\square}$ can be written in a unique way as $\boldsymbol{v} = \alpha_1 \boldsymbol{b}_1 + \cdots + \alpha_d \boldsymbol{b}_d$ with scalar coefficients $\alpha_k \in \mathbb{C}$. Under the assumption that

$$\text{the } (m+1)s \text{ columns of } [\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^m\boldsymbol{b}] \text{ are linearly independent,} \tag{1.2}$$

the block Krylov space is of dimension $d = (m+1)s^2$ and there is a one-to-one correspondence between any block vector

$$P(A) \circ \boldsymbol{b} := \boldsymbol{b}C_0 + A\boldsymbol{b}C_1 + \cdots + A^m\boldsymbol{b}C_m \in \mathcal{K}_{m+1}^{\square}$$

and the matrix polynomial $P(z) = C_0 + zC_1 + \cdots + z^m C_m$. The circ ($\circ$) notation has been attributed to Gragg in [34].

Given a polynomial $q_m \in \mathcal{P}_m$ such that $q_m(A)$ is nonsingular, we define the (classical) *block rational Krylov space* associated with A, $\boldsymbol{b}$, and q_m as follows:

$$\mathcal{Q}_{m+1}^{\square}(A, \boldsymbol{b}) := q_m(A)^{-1}\mathcal{K}_{m+1}^{\square}(A, \boldsymbol{b}) = \mathcal{K}_{m+1}^{\square}(A, q_m(A)^{-1}\boldsymbol{b}).$$

Note that the polynomial q_m is implicit in our notation $\mathcal{Q}_{m+1}^{\square} = \mathcal{Q}_{m+1}^{\square}(A, \boldsymbol{b})$. The roots $\xi_1, \xi_2, \ldots, \xi_m \in \overline{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ of q_m are referred to as the *poles* of the block rational Krylov space. Both the dimension and the invariance index of a block rational Krylov space equal those of its polynomial counterpart.

The focus of this paper is on theoretical and algorithmic aspects of generating bases of block rational Krylov spaces. A convenient tool in this respect are so-called *block rational Arnoldi decompositions (BRADs)*, which we introduce in Section 2. In Section 3 we provide an implicit Q theorem which characterizes the parameters that uniquely determine a block rational Krylov space. The elements of a block rational Krylov space are closely linked to rational functions: if $\mathcal{Q}_{m+1}^{\square}$ is of full dimension $(m+1)s^2$, then there is a one-to-one correspondence between any vector

$$R(A) \circ \boldsymbol{b} := q_m(A)^{-1}(\boldsymbol{b}C_0 + A\boldsymbol{b}C_1 + \cdots + A^m\boldsymbol{b}C_m) \in \mathcal{Q}_{m+1}^{\square} \tag{1.3}$$

and the *rational matrix-valued function*

$$R(z) = q_m(z)^{-1}(C_0 + zC_1 + \cdots + z^m C_m). \tag{1.4}$$

We refer to such functions as RKFUNBs (block rational Krylov functions), generalizing the RKFUN concept [11] to the block case. In Section 4, we explain how RKFUNBs are uniquely encoded in block rational Arnoldi decompositions and how they can be evaluated efficiently via "rerunning" the decomposition. During the algorithmic construction of a block rational Krylov basis, an appropriate choice of so-called continuation matrices needs to be made, and this is explained in Section 5. Linear dependencies between the block Krylov vectors can lead to breakdowns of the rational Arnoldi method. Possible deflation stategies, which can circumvent such breakdowns, are described in Section 6. Finally, Section 7 contains numerical illustrations.

Notation. Finding an accessible notation for block Krylov methods is a challenge and varying notation is used in the literature. For clarity, we give a short summary of our notation. Throughout this work, scalars are denoted by lowercase Greek letters, e.g., $\alpha, \beta \in \mathbb{C}$. Vectors are denoted by lowercase Latin letters, e.g.,

$b \in \mathbb{C}^N$ and $v_k \in \mathbb{C}^N$. Matrices are denoted by uppercase Latin letters and their elements with the corresponding lowercase letters, e.g., $A = [a_{ij}] \in \mathbb{C}^{N\times N}$. Block vectors are denoted by bold lowercase Latin letters and their columns are vectors, e.g., $\boldsymbol{b} = [b_1, \dots, b_s] \in \mathbb{C}^{N\times s}$. Block matrices are denoted by bold uppercase Latin letters and their elements are matrices, e.g., $\boldsymbol{A} = [A_{ij}]$ and $\underline{\boldsymbol{H}_m} = [H_{ij}]$, or block vectors, e.g., $\boldsymbol{V}_m = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_m]$. This notation has been chosen so that bold symbols imply block structure. An element of a block vector is a vector, whereas an element of a block matrix is a smaller matrix or a block vector. The linear space of scalar polynomials of degree at most m is denoted by $\mathcal{P}_m$. Finally, $\overline{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ denotes the extended complex plane.

2. Block rational Arnoldi decompositions. We aim to establish a correspondence between block rational Krylov spaces and a particular type of block matrix decomposition. A block matrix

$$\underline{\boldsymbol{H}_m} = \begin{bmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,m} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,m} \\ & H_{3,2} & \cdots & H_{3,m} \\ & & \ddots & \vdots \\ & & & H_{m+1,m} \end{bmatrix} \in \mathbb{C}^{(m+1)s\times ms}, \quad H_{ij} \in \mathbb{C}^{s\times s},$$

is called a *block upper-Hessenberg matrix.* A block upper-Hessenberg matrix $\underline{\boldsymbol{H}_m} \in \mathbb{C}^{(m+1)s\times ms}$ is *unreduced* if all its subdiagonal blocks are nonsingular, i.e., $H_{j+1,j}$ is nonsingular for $j = 1, \dots, m$. The notion of being unreduced can be extended to a block upper-Hessenberg pencil as follows.

DEFINITION 2.1. *Let $\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m} \in \mathbb{C}^{(m+1)s\times ms}$ be block upper-Hessenberg matrices. We say that $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ is an* unreduced block upper-Hessenberg pencil *if, for every $j = 1, \dots, m$, at least one of the matrices $H_{j+1,j}$ and $K_{j+1,j}$ is nonsingular.*

We now define block rational Arnoldi decompositions as matrix equations which generalize the decompositions described in [9, 46, 47] to block rational Krylov spaces.

DEFINITION 2.2. *Let $A \in \mathbb{C}^{N\times N}$ be a given matrix. A relation of the form*

$$A\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m} \tag{2.1}$$

is called a block rational Arnoldi decomposition *(BRAD) if the following conditions are satisfied:*

(a) *$\boldsymbol{V}_{m+1} \in \mathbb{C}^{N\times(m+1)s}$ is of full column rank,*
(b) *$(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ is an unreduced block upper-Hessenberg pencil of size $(m+1)s\times ms$,*
(c) *$\mu_j K_{j+1,j} = \nu_j H_{j+1,j}$ with scalars $\mu_j, \nu_j \in \mathbb{C}$ such that $|\mu_j| + |\nu_j| \neq 0$ for $j = 1, \dots, m$, and*
(d) *the numbers $\xi_j = \mu_j/\nu_j$ are outside the spectrum $\Lambda(A)$ for $j = 1, \dots, m$.*

The numbers $\xi_1, \dots, \xi_m \in \overline{\mathbb{C}}$ are called the poles *of the BRAD.*

The block columns of $\boldsymbol{V}_{m+1} = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_{m+1}]$ blockspan the *space* of the decomposition, that is, the linear space of block vectors $\boldsymbol{v} = \sum_{j=1}^{m+1} \boldsymbol{v}_j C_j$ with arbitrary coefficient matrices $C_j \in \mathbb{C}^{s\times s}$. If $\boldsymbol{V}_{m+1}$ has $(m+1)s$ orthonormal columns, we have an *orthonormal BRAD.* Any BRAD can be transformed into an orthonormal BRAD with the same poles ξ_j by using a thin QR factorization $\boldsymbol{V}_{m+1} =: \widehat{\boldsymbol{V}}_{m+1}R$, where $\widehat{\boldsymbol{V}}_{m+1} \in \mathbb{C}^{N\times(m+1)s}$ has orthonormal columns, and $R \in \mathbb{C}^{(m+1)s\times(m+1)s}$ is upper triangular and nonsingular. Setting $\underline{\widehat{\boldsymbol{K}}_m} := R\underline{\boldsymbol{K}_m}$, and $\underline{\widehat{\boldsymbol{H}}_m} := R\underline{\boldsymbol{H}_m}$, we obtain an orthonormal BRAD $A\widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{K}}_m} = \widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{H}}_m}$ that is equivalent to the original BRAD in the following sense.

Definition 2.3. *Two BRADs* (2.1) *with the same matrix* $A \in \mathbb{C}^{N\times N}$ *are* equivalent *if they blockspan the same space and have the same poles* $\xi_1, \ldots, \xi_m$ *(not necessarily in this order).*

This definition of BRAD equivalence generalizes [9, Definition 2.4] from standard rational Krylov spaces to the block case. It justifies that, without loss of generality, one may assume that BRADs are orthonormal.

In order to construct an orthonormal BRAD, Algorithm 2.1 can be used. This algorithm is a natural extension of the single-vector rational Arnoldi method, using thin QR factorizations to ensure orthonormality of the block vectors $\boldsymbol{v}_j$, that is,

$$\boldsymbol{v}_j^*\boldsymbol{v}_k = \begin{cases} I_{s\times s} & \text{if } j = k, \\ O_{s\times s} & \text{if } j \neq k. \end{cases}$$

This type of orthonormality is also enforced in the so-called "classical block method" for polynomial Krylov spaces; see [24, 38]. Our variant of the algorithm starts with the orthogonalization of the initial block vector $\boldsymbol{b} \in \mathbb{C}^{N\times s}$, which we assume to be of maximal rank s. For each of the poles ξ_j, a quadruple $(\nu_j, \mu_j, \rho_j, \eta_j) \in \mathbb{C}^4$ is chosen such that $\mu_j/\nu_j = \xi_j$ and $\mu_j/\nu_j \neq \eta_j/\rho_j$. A simple strategy is to set $(\nu_j, \mu_j) = (1, \xi_j)$ for finite poles ξ_j, and $(\nu_j, \mu_j) = (0, 1)$ when $\xi_j = \infty$. To guarantee that $\mu_j/\nu_j \neq \eta_j/\rho_j$, we set $(\rho_j, \eta_j) = (1, 0)$ when $|\xi_j| > 1$, and $(\rho_j, \eta_j) = (0, 1)$ otherwise. The quotient η_j/ρ_j is referred to as the *continuation root* [10].

The computational core of Algorithm 2.1 is line 5, where a linear system $(\nu_j A - \mu_j I)\boldsymbol{w}_j = (\rho_j A - \eta_j I)\boldsymbol{V}_j T_j$ is solved for $\boldsymbol{w}_j$. The *continuation matrix* $T_j \in \mathbb{C}^{js\times s}$ is used to select an appropriate element from $\mathcal{Q}_j^{\square} = \text{blockspan}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j\}$ to form the right-hand side of the linear system. Different choices of continuation matrices T_j are discussed in Section 5. The block vector $\boldsymbol{w}_j$ is orthogonalized against the previous block vectors. For simplicity of exposition, Algorithm 2.1 uses the classic Gram–Schmidt procedure, but one could equally use the modified Gram–Schmidt procedure; see, e.g., [32, Chapter 19]. Furthermore, we remark that when $s > 1$, there are many possibilities for the implementation of the orthogonalization: for example, one can treat each block vector as a set of individual vectors and orthogonalize on a vector level as opposed to a block orthogonalization. Finally, the newly orthogonalized block vector $\boldsymbol{w}_j$ is made orthonormal by computing a thin QR factorization.

It is straightforward to verify that Algorithm 2.1 produces a BRAD by decomposing the `for` loop. Combining lines 6–8, we have

$$\boldsymbol{w}_j = \boldsymbol{V}_j\boldsymbol{c}_j + \boldsymbol{v}_{j+1}C_{j+1,j} = \boldsymbol{V}_{j+1}\underline{\boldsymbol{c}_j},$$

and by line 5,

$$(\nu_j A - \mu_j I)\,\boldsymbol{V}_{j+1}\underline{\boldsymbol{c}_j} = (\rho_j A - \eta_j I)\,\boldsymbol{V}_j T_j.$$

Separating the terms containing A as a factor, we obtain

$$A\,\boldsymbol{V}_{j+1}(\nu_j\underline{\boldsymbol{c}_j} - \rho_j\underline{T_j}) = \boldsymbol{V}_{j+1}(\mu_j\underline{\boldsymbol{c}_j} - \eta_j\underline{T_j}).$$

Algorithm 2.1 Block rational Arnoldi method RKToolbox: `rat_krylov`

Input: $A \in \mathbb{C}^{N\times N}$, $\boldsymbol{b} \in \mathbb{C}^{N\times s}$ of rank s, poles $\{\xi_j\}_{j=1}^m \subset \overline{\mathbb{C}} \setminus \Lambda(A)$

Output: Orthonormal BRAD $A\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$

1. Compute a thin QR factorization of $\boldsymbol{b} =: \boldsymbol{v}_1 R$
2. **for** $j = 1, \ldots, m$ **do**
3. Choose $(\nu_j, \mu_j, \rho_j, \eta_j) \in \mathbb{C}^4$ such that $\mu_j/\nu_j = \xi_j$ and $\rho_j\mu_j \neq \eta_j\nu_j$
4. Choose a *continuation matrix* $T_j \in \mathbb{C}^{js\times s}$
5. Compute $\boldsymbol{w}_j := (\nu_j A - \mu_j I)^{-1}(\rho_j A - \eta_j I)\,\boldsymbol{V}_j T_j$
6. Project $\boldsymbol{c}_j := \boldsymbol{V}_j^* \boldsymbol{w}_j$
7. Compute $\boldsymbol{w}_j := \boldsymbol{w}_j - \boldsymbol{V}_j \boldsymbol{c}_j$ orthogonal to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j$
8. Compute a thin QR factorization of $\boldsymbol{w}_j =: \boldsymbol{v}_{j+1} C_{j+1,j}$
9. Set $\underline{\boldsymbol{k}_j} := \nu_j \underline{\boldsymbol{c}_j} - \rho_j \underline{T_j}$ and $\underline{\boldsymbol{h}_j} := \mu_j \underline{\boldsymbol{c}_j} - \eta_j \underline{T_j}$, where $\underline{T_j} = [T_j^T \quad O_{s\times s}]^T$ and $\underline{\boldsymbol{c}_j} = [\boldsymbol{c}_j^T, C_{j+1,j}^T]^T$
10. **end for**

Using the definitions of $\underline{\boldsymbol{k}_j}$ and $\underline{\boldsymbol{h}_j}$ in line 9 of the algorithm, we have

$$A\boldsymbol{V}_{j+1}\underline{\boldsymbol{k}_j} = \boldsymbol{V}_{j+1}\underline{\boldsymbol{h}_j}.$$

Concatenating the columns of these relations for all $j = 1, \ldots, m$ gives the orthonormal BRAD $A\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$ returned by the algorithm.

3. The block rational implicit Q theorem.

The implicit Q theorem, to be given in Theorem 3.3, states that an orthonormal BRAD (2.1) is essentially uniquely determined by the starting block vector $\boldsymbol{v}_1$ and the *ordered* poles $\xi_1, \ldots, \xi_m$. To put this result in context, consider a standard ($s = 1$) polynomial Krylov space associated with the matrix $A \in \mathbb{C}^{N\times N}$. Given a matrix $V_m \in \mathbb{C}^{N\times m}$ with orthonormal columns, the standard implicit Q theorem states that if $H_m = V_m^* A V_m$ is an unreduced upper-Hessenberg matrix, then V_m is uniquely determined (up to unimodular column scaling) by its first column; see, e.g., [50, Chapter 2, Theorem 3.3]. This result is extended to standard ($s = 1$) rational Krylov spaces in [9, Theorem 3.2], using the notion of *essentially equal* rational Arnoldi decompositions (RADs); see [9, Definition 3.1]. The concept of essentially equal RADs extends to block rational Krylov spaces as follows.

Definition 3.1. *Two orthonormal BRADs, namely,* $A\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$ *and* $A\widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{K}}_m} = \widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{H}}_m}$, *are* essentially equal *if there exists a unitary block diagonal matrix* $\boldsymbol{D}_{m+1} \in \mathbb{C}^{(m+1)s\times(m+1)s}$, *and a nonsingular block upper-triangular matrix* $\boldsymbol{T}_m \in \mathbb{C}^{ms\times ms}$, *such that* $\widehat{\boldsymbol{V}}_{m+1} = \boldsymbol{V}_{m+1}\boldsymbol{D}_{m+1}$, $\underline{\widehat{\boldsymbol{H}}_m} = \boldsymbol{D}_{m+1}^*\underline{\boldsymbol{H}_m}\boldsymbol{T}_m$, *and* $\underline{\widehat{\boldsymbol{K}}_m} = \boldsymbol{D}_{m+1}^*\underline{\boldsymbol{K}_m}\boldsymbol{T}_m$.

Essentially equal orthonormal BRADs form an equivalence class and we call any of its elements essentially unique.

Two orthonormal BRADs may be equivalent but not essentially equal if the poles are ordered differently. The following lemma collects some useful results on unreduced block upper-Hessenberg matrices and BRADs.

Lemma 3.2.

(i) *If* $\underline{\boldsymbol{H}_m} \in \mathbb{C}^{(m+1)s\times ms}$ *is an unreduced block upper-Hessenberg matrix, then it is of maximal rank* ms.

(ii) *Let $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ be an unreduced block upper-Hessenberg pencil of a BRAD, then each subdiagonal $H_{j+1,j}$ is either nonsingular or the zero matrix.*

(iii) *Let $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ be an unreduced block upper-Hessenberg pencil of a BRAD, then the matrix $\alpha\underline{\boldsymbol{H}_m} - \beta\underline{\boldsymbol{K}_m}$ is of maximal rank ms for all $\alpha, \beta \in \mathbb{C}$ satisfying $|\alpha| + |\beta| \neq 0$.*

(iv) *Let $\underline{\boldsymbol{L}_m} \in \mathbb{C}^{(m+1)s\times ms}$ be of maximal rank ms, then the dimension of the left null space of $\underline{\boldsymbol{L}_m}$ is s. Moreover, if $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^{(m+1)s\times s}$ are of maximal rank s and $\boldsymbol{x}^*\underline{\boldsymbol{L}_m} = O_{s\times ms}$ and $\boldsymbol{y}^*\underline{\boldsymbol{L}_m} = O_{s\times ms}$, then there exists a nonsingular matrix $M \in \mathbb{C}^{s\times s}$ such that $\boldsymbol{x} = \boldsymbol{y}M$.*

Proof.

(i) Removing the first block row of $\underline{\boldsymbol{H}_m}$ yields a block upper-triangular matrix with nonsingular diagonal blocks. Hence $\underline{\boldsymbol{H}_m}$ has full column rank.

(ii) Assume that $H_{j+1,j}$ is singular, in which case $K_{j+1,j}$ must be nonsingular. The pencil satisfies a BRAD and so $\mu_j K_{j+1,j} = \nu_j H_{j+1,j}$ with $|\mu_j| + |\nu_j| \neq 0$. This is only possible when $\mu_j = 0$ and hence $H_{j+1,j} = O_{s\times s}$.

(iii) This proof follows closely that of [10, Lemma 2.5]. Consider auxiliary scalars $\widehat{\alpha} = 1$ and $\widehat{\beta} \in \mathbb{C}$ such that $\widehat{\alpha}\underline{\boldsymbol{H}_m} - \widehat{\beta}\underline{\boldsymbol{K}_m}$ is of rank ms. Multiplying the BRAD (2.1) by $\widehat{\alpha}$ and subtracting $\widehat{\beta}\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m}$ from both sides gives

$$\big(\widehat{\alpha}A - \widehat{\beta}I\big)\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\big(\widehat{\alpha}\underline{\boldsymbol{H}_m} - \widehat{\beta}\underline{\boldsymbol{K}_m}\big). \tag{3.1}$$

Clearly, the right-hand side of this equation is of rank ms, hence so is the left-hand side. As a consequence, $\underline{\boldsymbol{K}_m}$ is of rank ms, which proves the claim for $\alpha = 0$.

If $\alpha \neq 0$, consider $\alpha = \widehat{\alpha}$ and $\beta = \widehat{\beta}$ in (3.1). There are two possible cases: either $\alpha\underline{\boldsymbol{H}_m} - \beta\underline{\boldsymbol{K}_m}$ is unreduced and so by (i) of maximal rank, or $\alpha\underline{\boldsymbol{H}_m} - \beta\underline{\boldsymbol{K}_m}$ is not unreduced, and so there exists an index $j \in \{1, \ldots, m\}$ such that $\alpha H_{j+1,j} - \beta K_{j+1,j}$ is singular. Now either $H_{j+1,j}$ is nonsingular or $H_{j+1,j} = O_{s\times s}$ by (ii). If $H_{j+1,j}$ is nonsingular, then $\alpha H_{j+1,j} - \beta K_{j+1,j}$ being singular is equivalent to

$$H_{j+1,j} - \frac{\beta}{\alpha}K_{j+1,j} = H_{j+1,j}\left(1 - \frac{\nu_j\beta}{\mu_j\alpha}\right)$$

being singular, therefore $\beta/\alpha = \mu_j/\nu_j$, i.e., equal to the jth pole and hence $\alpha A - \beta I$ must be nonsingular. If $H_{j+1,j} = O_{s\times s}$, then we would have $\beta/\alpha = \infty$ and hence $\alpha = 0$, contradicting our assumption. Finally, since $\alpha A - \beta I$ is nonsingular, and $\boldsymbol{V}_{m+1}$ and $\underline{\boldsymbol{K}_m}$ are of full column rank, we conclude that $\alpha\underline{\boldsymbol{H}_m} - \beta\underline{\boldsymbol{K}_m}$ is of full column rank by (3.1).

(iv) The first part follows directly from the rank–nullity theorem (see, e.g., [49, Theorem 3.17]). For the second part, assume that $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^{(m+1)s\times s}$ are of maximal rank s and $\boldsymbol{x}^*\underline{\boldsymbol{L}_m} = O_{s\times ms}$ and $\boldsymbol{y}^*\underline{\boldsymbol{L}_m} = O_{s\times ms}$. Then by definition, the columns of $\boldsymbol{x}, \boldsymbol{y}$ form a basis of the left null space of $\underline{\boldsymbol{L}_m}$, and hence there exists a nonsingular matrix $M \in \mathbb{C}^{s\times s}$ such that $\boldsymbol{x} = \boldsymbol{y}M$. ∎

We are now in the position to state our implicit Q theorem.

THEOREM 3.3 (block implicit Q theorem). *Let $A \in \mathbb{C}^{N\times N}$ satisfy an orthonormal block rational Arnoldi decomposition $A\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$ with poles $\{\mu_j/\nu_j\}_{j=1}^m \subset \overline{\mathbb{C}} \setminus \Lambda(A)$. The orthonormal matrix $\boldsymbol{V}_{m+1}$ and the pencil $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ are* *essentially uniquely determined by* $\boldsymbol{v}_1$ *(the first block column of* $\boldsymbol{V}_{m+1}$*) and the poles* $\mu_1/\nu_1, \ldots, \mu_m/\nu_m$.

Proof. Let $A\widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{K}}_m} = \widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{H}}_m}$ be an orthonormal BRAD with $\widehat{\boldsymbol{V}}_{m+1} = [\boldsymbol{v}_1, \widehat{\boldsymbol{v}}_2, \ldots, \widehat{\boldsymbol{v}}_{m+1}]$ and poles $\mu_1/\nu_1, \ldots, \mu_m/\nu_m$. We want to show that $A\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$ and $A\widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{K}}_m} = \widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{H}}_m}$ are essentially equal.

Without loss of generality, we may assume that $H_{j+1,j}$ is nonsingular and therefore $\mu_j/\nu_j \neq 0$, for $j = 1, \ldots, m$. Otherwise, if $H_{j+1,j}$ is singular for some j, the matrix $K_{j+1,j}$ is nonsingular as $(\underline{\boldsymbol{H}_j}, \underline{\boldsymbol{K}_j})$ is an unreduced block upper-Hessenberg pencil, and so $\mu_j = 0$ as $\nu_j H_{j+1,j} = \mu_j K_{j+1,j}$. In this case $0 = \mu_j/\nu_j \notin \Lambda(A)$ and so A is nonsingular, hence we can consider $\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = A^{-1}\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$ at the jth step, thus interchanging the roles of $\underline{\boldsymbol{H}_m}$ and $\underline{\boldsymbol{K}_m}$.

Given any pole $\mu/\nu \in \overline{\mathbb{C}} \setminus (\Lambda(A) \cup \{0\})$, we subtract $(\nu/\mu)A\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$ from both sides of the relation $A\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$ to give the decomposition

$$A^{(\nu/\mu)}\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{L}_m^{(\nu/\mu)}} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}, \tag{3.2}$$

where $A^{(\nu/\mu)} := (I - A\nu/\mu)^{-1}A$ and $\underline{\boldsymbol{L}_m^{(\nu/\mu)}} := (\underline{\boldsymbol{K}_m} - \underline{\boldsymbol{H}_m}\nu/\mu)$. The matrix $\underline{\boldsymbol{L}_m^{(\nu/\mu)}}$ is of maximal rank by Lemma 3.2 (iii).

By construction, for each pole $\{\mu_j/\nu_j\}_{j=1}^m$, the matrix $\underline{\boldsymbol{L}_m^{(\nu_j/\mu_j)}}$ is block upper-Hessenberg. Furthermore, the jth column of $\underline{\boldsymbol{L}_m^{(\nu_j/\mu_j)}}$ has all but the leading j blocks equal to the zero matrix: consider the $(j+1,j)$th block of $\underline{\boldsymbol{L}_m^{(\nu_j/\mu_j)}}$, then $L_{j+1,j} = K_{j+1,j} - (H_{j+1,j}\nu_j)/\mu_j = O_{s\times s}$. An analogous result holds for $A\widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{K}}_m} = \widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{H}}_m}$.

We now prove by induction on m, one block column at a time, that the two BRADs are essentially equal. Define $D_{1,1} = I_{s\times s}$, as $\widehat{\boldsymbol{v}}_1 = \boldsymbol{v}_1 D_{1,1}$, and consider the first block column of (3.2),

$$A^{(\nu_1/\mu_1)}\boldsymbol{v}_1 L_{1,1}^{(\nu_1/\mu_1)} = \boldsymbol{v}_1 H_{1,1} + \boldsymbol{v}_2 H_{2,1}. \tag{3.3}$$

The block columns of $\boldsymbol{V}_{m+1}$ are block orthonormal, therefore pre-multiplying (3.3) by $\boldsymbol{v}_1^*$ gives

$$\boldsymbol{v}_1^* A^{(\nu_1/\mu_1)}\boldsymbol{v}_1 L_{1,1}^{(\nu_1/\mu_1)} = H_{1,1}. \tag{3.4}$$

Substituting (3.4) into (3.3), rearranging and right-multiplying by $H_{2,1}^{-1}$ yields

$$\boldsymbol{v}_2 = (A^{(\nu_1/\mu_1)}\boldsymbol{v}_1 - \boldsymbol{v}_1\boldsymbol{v}_1^* A^{(\nu_1/\mu_1)}\boldsymbol{v}_1)(L_{1,1}^{(\nu_1/\mu_1)} H_{2,1}^{-1}). \tag{3.5}$$

Analogously, we have $\widehat{H}_{1,1} = \boldsymbol{v}_1^* A^{(\nu_1/\mu_1)}\boldsymbol{v}_1 \widehat{L}_{1,1}^{(\nu_1/\mu_1)}$ and

$$\widehat{\boldsymbol{v}}_2 = (A^{(\nu_1/\mu_1)}\boldsymbol{v}_1 - \boldsymbol{v}_1\boldsymbol{v}_1^* A^{(\nu_1/\mu_1)}\boldsymbol{v}_1)(\widehat{L}_{1,1}^{(\nu_1/\mu_1)} \widehat{H}_{2,1}^{-1}). \tag{3.6}$$

By rearranging (3.5) and (3.6), we find $\widehat{\boldsymbol{v}}_2 = \boldsymbol{v}_2 D_2$ with

$$D_2 := H_{2,1}(L_{1,1}^{(\nu_1/\mu_1)})^{-1}\widehat{L}_{1,1}^{(\nu_1/\mu_1)}\widehat{H}_{2,1}^{-1}.$$

The matrix D_2 is unitary as the block vectors $\boldsymbol{v}_2$ and $\widehat{\boldsymbol{v}}_2$ have orthonormal columns. Defining $\boldsymbol{T}_1 = T_{1,1} = (L_{1,1}^{(\nu_1/\mu_1)})^{-1}(\widehat{L}_{1,1}^{(\nu_1/\mu_1)})$ and $\boldsymbol{D}_2 = \mathrm{diag}(D_1, D_2)$, we can write

$$\boldsymbol{D}_2^*\underline{\boldsymbol{H}_1}\,\boldsymbol{T}_1 = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}^* \begin{pmatrix} H_{1,1} \\ H_{2,1} \end{pmatrix} T_{1,1} = \underline{\widehat{\boldsymbol{H}}_1},$$

as $H_{1,1}(L_{1,1}^{(\nu_1/\mu_1)})^{-1} = \widehat{H}_{1,1}(\widehat{L}_{1,1}^{(\nu_1/\mu_1)})^{-1}$ and $\boldsymbol{v}_2 H_{2,1}(L_{1,1}^{(\nu_1/\mu_1)})^{-1} = \widehat{\boldsymbol{v}}_2\widehat{H}_{2,1}(\widehat{L}_{1,1}^{(\nu_1/\mu_1)})^{-1}$. An analogous calculation shows that $\boldsymbol{D}_2^* \underline{\boldsymbol{K}_1}\, \boldsymbol{T}_1 = \underline{\widehat{\boldsymbol{K}_1}}$.

Assume that for $2 \le j \le m$, there exists a unitary block diagonal matrix $\boldsymbol{D}_j = \mathrm{diag}(D_1, \dots, D_j)$ and a block upper-triangular matrix $\boldsymbol{T}_{j-1}$ such that $\widehat{\boldsymbol{V}_j} = \boldsymbol{V}_j \boldsymbol{D}_j$, $\underline{\widehat{\boldsymbol{H}}_{j-1}} = \boldsymbol{D}_j^* \underline{\boldsymbol{H}_{j-1}}\, \boldsymbol{T}_{j-1}$, and $\underline{\widehat{\boldsymbol{K}}_{j-1}} = \boldsymbol{D}_j^* \underline{\boldsymbol{K}_{j-1}}\, \boldsymbol{T}_{j-1}$.

Reading off the jth block column of (3.2) yields

$$A^{(\nu_j/\mu_j)}\, \boldsymbol{V}_j \boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)} = \boldsymbol{V}_{j+1} \underline{\boldsymbol{H}_{:,j}} = \boldsymbol{V}_j \boldsymbol{H}_{:,j} + \boldsymbol{v}_{j+1} H_{j+1,j}, \tag{3.7}$$

where $\underline{\boldsymbol{H}_{:,j}}$ is the jth block column of the matrix $\underline{\boldsymbol{H}_j}$. The columns of $\boldsymbol{V}_{j+1}$ are block orthonormal, and so $\boldsymbol{V}_j^* \boldsymbol{V}_{j+1} = [I_{js\times js} \mid O_{js\times s}]$. Premultiplying (3.7) by $\boldsymbol{V}_j^*$ gives

$$\boldsymbol{V}_j^* A^{(\nu_j/\mu_j)}\, \boldsymbol{V}_j \boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)} = \boldsymbol{H}_{:,j}. \tag{3.8}$$

Substituting (3.8) into (3.7) and rearranging for $\boldsymbol{v}_{j+1}$ yields

$$\begin{aligned} \boldsymbol{v}_{j+1} H_{j+1,j} &= A^{(\nu_j/\mu_j)}\, \boldsymbol{V}_j \boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)} - \boldsymbol{V}_j \boldsymbol{V}_j^* A^{(\nu_j/\mu_j)}\, \boldsymbol{V}_j \boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)} \\ &= (I - \boldsymbol{V}_j \boldsymbol{V}_j^*) A^{(\nu_j/\mu_j)}\, \boldsymbol{V}_j \boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)}. \end{aligned} \tag{3.9}$$

We can write the block column $\boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)} \in \mathbb{C}^{js\times s}$ as

$$\boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)} = \underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}}\, \boldsymbol{z}_{j-1} + \boldsymbol{q}_j \tag{3.10}$$

with block vectors $\boldsymbol{z}_{j-1} \in \mathbb{C}^{(j-1)s\times s}$ and $\boldsymbol{q}_j \in \mathbb{C}^{js\times s}$, where the columns of $\boldsymbol{q}_j$ are chosen so that they do not lie in the span of the columns of $\underline{\boldsymbol{L}_{j-1}}^{(\nu_j/\mu_j)}$, those components are in $\boldsymbol{z}_{j-1}$, and so $\boldsymbol{q}_j^* \underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}} = O_{s\times(j-1)s}$.

By Lemma 3.2 (iii), the block matrix $\underline{\boldsymbol{L}_j^{(\nu_j/\mu_j)}}$ is of full column rank. Furthermore, the block vector $\boldsymbol{q}_j$ has full column rank by (3.10), as the s columns of $\boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)}$ (the jth column block of $\underline{\boldsymbol{L}_j^{(\nu_j/\mu_j)}}$) are linearly independent, and are not contained in the span of the columns of $\underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}}$.

Substituting (3.10) into (3.9), gives

$$\boldsymbol{v}_{j+1} H_{j+1,j} = (I - \boldsymbol{V}_j \boldsymbol{V}_j^*) A^{(\nu_j/\mu_j)}\, \boldsymbol{V}_j \underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}}\, \boldsymbol{z}_{j-1} + (I - \boldsymbol{V}_j \boldsymbol{V}_j^*) A^{(\nu_j/\mu_j)}\, \boldsymbol{V}_j \boldsymbol{q}_j. \tag{3.11}$$

We show that the first term on the right hand side of (3.11) is $O_{N\times s}$, by substituting the first $j-1$ block columns of (3.2) with $\nu = \nu_j$ and $\mu = \mu_j$:

$$\begin{aligned} (I - \boldsymbol{V}_j \boldsymbol{V}_j^*) A^{(\nu_j/\mu_j)}\, \boldsymbol{V}_j \underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}}\, \boldsymbol{z}_{j-1} &= (I - \boldsymbol{V}_j \boldsymbol{V}_j^*)\, \boldsymbol{V}_j \underline{\boldsymbol{H}_{j-1}}\, \boldsymbol{z}_{j-1}, \\ &= \boldsymbol{V}_j \underline{\boldsymbol{H}_{j-1}}\, \boldsymbol{z}_{j-1} - \boldsymbol{V}_j \boldsymbol{V}_j^* \boldsymbol{V}_j \underline{\boldsymbol{H}_{j-1}}\, \boldsymbol{z}_{j-1}, \\ &= O_{N\times s}. \end{aligned}$$

Equation (3.11) reduces to

$$\boldsymbol{v}_{j+1} H_{j+1,j} = (I - \boldsymbol{V}_j \boldsymbol{V}_j^*) A^{(\nu_j/\mu_j)}\, \boldsymbol{V}_j \boldsymbol{q}_j. \tag{3.12}$$

Analogously, we have $\widehat{\boldsymbol{L}}_{:,j}^{(\nu_j/\mu_j)} = \underline{\widehat{\boldsymbol{L}}_{j-1}^{(\nu_j/\mu_j)}}\widehat{\boldsymbol{z}}_{j-1} + \widehat{\boldsymbol{q}}_j$, where $\widehat{\boldsymbol{q}}_j^* \underline{\widehat{\boldsymbol{L}}_{j-1}^{(\nu_j/\mu_j)}} = O_{s\times(j-1)s}$ and $\widehat{\boldsymbol{v}}_{j+1}\widehat{H}_{j+1,j} = (I - \widehat{\boldsymbol{V}}_j \widehat{\boldsymbol{V}}_j^*)A^{(\nu_j/\mu_j)}\widehat{\boldsymbol{V}}_j \widehat{\boldsymbol{q}}_j$. By the induction hypothesis $\widehat{\boldsymbol{V}}_j = \boldsymbol{V}_j \boldsymbol{D}_j$, and so

$$\widehat{\boldsymbol{v}}_{j+1}\widehat{H}_{j+1,j} = (I - \boldsymbol{V}_j \boldsymbol{V}_j^*)A^{(\nu_j/\mu_j)}\boldsymbol{V}_j \boldsymbol{D}_j \widehat{\boldsymbol{q}}_j. \tag{3.13}$$

Combining $\underline{\widehat{\boldsymbol{L}}_{j-1}^{(\nu_j/\mu_j)}} = \boldsymbol{D}_j^* \underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}}\boldsymbol{T}_{j-1}$ from the induction hypothesis and $\widehat{\boldsymbol{q}}_j^* \underline{\widehat{\boldsymbol{L}}_{j-1}^{(\nu_j/\mu_j)}} = O_{s\times(j-1)s}$, we have $\widehat{\boldsymbol{q}}_j^* \boldsymbol{D}_j^* \underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}}\boldsymbol{T}_{j-1} = O_{s\times(j-1)s}$. Now by construction $\boldsymbol{T}_{j-1}$ is nonsingular, hence $\widehat{\boldsymbol{q}}_j^* \boldsymbol{D}_j^* \underline{\boldsymbol{L}_{j-1}^{(\xi_j)}} = O_{s\times(j-1)s}$.

The block vectors $\boldsymbol{D}_j \widehat{\boldsymbol{q}}_j$, $\boldsymbol{q}_j$ and the block matrix $\underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}}$ are of maximal rank. By Lemma 3.2 (iv), there exists a nonsingular matrix $M \in \mathbb{C}^{s\times s}$ such that $\boldsymbol{D}_j \widehat{\boldsymbol{q}}_j = \boldsymbol{q}_j M$. Combining equations (3.12) and (3.13), we obtain

$$\widehat{\boldsymbol{v}}_{j+1}\widehat{H}_{j+1,j} = \boldsymbol{v}_{j+1} H_{j+1,j} M, \tag{3.14}$$

and so the columns of the block vectors $\boldsymbol{v}_{j+1}$ and $\widehat{\boldsymbol{v}}_{j+1}$ span the same space. Furthermore, $\boldsymbol{v}_{j+1}^*\boldsymbol{v}_{j+1} = \widehat{\boldsymbol{v}}_{j+1}^*\widehat{\boldsymbol{v}}_{j+1} = I_{s\times s}$, so there exists a unitary matrix $D_{j+1,j+1} \in \mathbb{C}^{s\times s}$ such that $\widehat{\boldsymbol{v}}_{j+1} = \boldsymbol{v}_{j+1} D_{j+1,j+1}$. This relation can be written as $\boldsymbol{v}_{j+1}(D_{j+1,j+1}\widehat{H}_{j+1,j} - H_{j+1,j}M) = O_{(j+1)s\times s}$, and so $\widehat{H}_{j+1,j} = D_{j+1,j+1}^* H_{j+1,j} M$.

It remains to find the block column $\boldsymbol{T}_{:,j}$ of the nonsingular matrix $\boldsymbol{T}_j$ such that $\underline{\widehat{\boldsymbol{H}}_j} = \boldsymbol{D}_{j+1}^* \underline{\boldsymbol{H}_j}\boldsymbol{T}_j$, and $\underline{\widehat{\boldsymbol{K}}_j} = \boldsymbol{D}_{j+1}^* \underline{\boldsymbol{K}_j}\boldsymbol{T}_j$. Rearranging and substituting (3.10) and analogous result into $\boldsymbol{D}_j \widehat{\boldsymbol{q}}_j = \boldsymbol{q}_j M$ yields

$$\boldsymbol{D}_j(\widehat{\boldsymbol{L}}_{:,j}^{(\nu_j/\mu_j)} - \underline{\widehat{\boldsymbol{L}}_{j-1}^{(\nu_j/\mu_j)}}\widehat{\boldsymbol{z}}_{j-1}) = (\boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)} - \underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}}\boldsymbol{z}_{j-1})M,$$

which can be rearranged to

$$\widehat{\boldsymbol{L}}_{:,j}^{(\nu_j/\mu_j)} = \boldsymbol{D}_j^* \underline{\boldsymbol{L}_{j-1}^{(\nu_j/\mu_j)}}(\boldsymbol{T}_{j-1}\widehat{\boldsymbol{z}}_{j-1} - \boldsymbol{z}_{j-1}M) + \boldsymbol{D}_j^* \boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)} M = \boldsymbol{D}_j^* \boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)}\boldsymbol{T}_{:,j}$$

where $\boldsymbol{T}_{:,j} = \begin{pmatrix} \boldsymbol{T}_{j-1}\widehat{\boldsymbol{z}}_{j-1} - \boldsymbol{z}_{j-1}M \\ M \end{pmatrix}$.

Finally substituting our results in (3.8), we have

$$\begin{aligned}
\widehat{\boldsymbol{H}}_{:,j} &= \widehat{\boldsymbol{V}}_j^* A^{(\nu_j/\mu_j)}\widehat{\boldsymbol{V}}_j \widehat{\boldsymbol{L}}_{:,j}^{(\nu_j/\mu_j)}, \\
&= (\boldsymbol{V}_j \boldsymbol{D}_j)^* A^{(\nu_j/\mu_j)}\boldsymbol{V}_j \boldsymbol{D}_j \boldsymbol{D}_j^* \boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)}\boldsymbol{T}_{:,j}, \\
&= \boldsymbol{D}_j^* \boldsymbol{V}_j^* A^{(\nu_j/\mu_j)}\boldsymbol{V}_j \boldsymbol{L}_{:,j}^{(\nu_j/\mu_j)}\boldsymbol{T}_{:,j}.
\end{aligned}$$

Substituting (3.7) yields

$$\begin{aligned}
\widehat{\boldsymbol{H}}_{:,j} &= \boldsymbol{D}_j^* \boldsymbol{V}_j^* \boldsymbol{V}_{j+1}\underline{\boldsymbol{H}_{:,j}}\boldsymbol{T}_{:,j}, \\
&= \boldsymbol{D}_j^* \boldsymbol{H}_{:,j}\boldsymbol{T}_{:,j}.
\end{aligned}$$

The subdiagonal elements follow from (3.14), and so $\underline{\widehat{\boldsymbol{H}}_j} = \boldsymbol{D}_{j+1}^* \underline{\boldsymbol{H}_j}\boldsymbol{T}_j$. An analogous result can be found for $\underline{\widehat{\boldsymbol{K}}_j}$. Therefore $\widehat{\boldsymbol{V}}_{j+1} = \boldsymbol{V}_{j+1}\boldsymbol{D}_{j+1}$, $\underline{\widehat{\boldsymbol{H}}_j} = \boldsymbol{D}_{j+1}^* \underline{\boldsymbol{H}_j}\boldsymbol{T}_j$, and

$\underline{\widehat{\boldsymbol{K}}_j} = \boldsymbol{D}_{j+1}^* \underline{\boldsymbol{K}_j} \boldsymbol{T}_j$, so the two BRADs are essentially equal, thereby completeing the induction. ∎

In Section 2 we showed how any BRAD can be transformed into an orthonormal BRAD using a thin QR factorization. By the implicit Q theorem, these two decompositions are essentially equal.

4. Rational matrix-valued functions and RKFUNB objects. Let us recall from the introduction that, under assumption (1.2), there is a one-to-one correspondence between any vector $R(A) \circ \boldsymbol{b} \in \mathcal{Q}_{m+1}^{\square}(A, \boldsymbol{b})$, given in (1.3), and a rational matrix-valued function $R(z)$ given in (1.4). Our aim is to show how rational matrix-valued functions are encoded in BRADs. The following lemma will be useful.

LEMMA 4.1. *Given two rational matrix-valued functions $R(z) = q_m(z)^{-1}(C_0 + zC_1 + \cdots + z^m C_m)$ and $S(z) = q_m(z)^{-1}(D_0 + zD_1 + \cdots + z^m D_m)$ with matrix coefficients $C_0, \ldots, C_m, D_0, \ldots, D_m \in \mathbb{C}^{s \times s}$, and a scalar polynomial $p(z)$. Let $A \in \mathbb{C}^{N \times N}$ and $B \in \mathbb{C}^{N \times s}$, then*

(i) $[p(z)R(z)]\big|_{z=A} \circ B = p(A)(R(A) \circ B) = R(A) \circ (p(A)B)$,
(ii) *Let* $M \in \mathbb{C}^{s \times s}$, *then* $[MR(z)]\big|_{z=A} \circ B = R(A) \circ (BM)$,
(iii) *Let* $M \in \mathbb{C}^{s \times s}$, *then* $[R(z)M]\big|_{z=A} \circ B = (R(A) \circ B)M$,
(iv) $(R \cdot S)(A) \circ B = S(A) \circ (R(A) \circ B)$.

Proof.

(i) This result follows from the definition of the circ operator in (1.3) and the fact that $p(A)$ commutes with $q_m(A)^{-1}$ and all matrix powers of A.

(ii)
$$\begin{aligned}
[MR(z)]\big|_{z=A} \circ B &= q_m(A)^{-1}([MC_0 + zMC_1 + \cdots + z^m MC_m]\big|_{z=A} \circ B) \\
&= q_m(A)^{-1}(MC_0 + ABMC_1 + \cdots + A^m BMC_m) \\
&= R(A) \circ (BM).
\end{aligned}$$

(iii)
$$\begin{aligned}
[R(z)M]\big|_{z=A} \circ B &= q_m(A)^{-1}([C_0 M + zC_1 M + \cdots + z^m C_m M]\big|_{z=A} \circ B) \\
&= q_m(A)^{-1}(C_0 M + ABC_1 M + \cdots + A^m BC_m M) \\
&= (R(A) \circ B)M.
\end{aligned}$$

(iv) This is a generalization of [20, Proposition 1]. The left-hand side is
$$(R \cdot S)(A) \circ B = q_m(A)^{-2} \sum_{k=0}^{2m} A^k B \sum_{i+j=k} C_i D_j.$$

For the right-hand side, we have
$$\begin{aligned}
S(A) \circ (R(A) \circ B) &= S(A) \circ \big(q_m(A)^{-1} \sum_{i=0}^{m} A^i BC_i\big) \\
&= q_m(A)^{-1}\big(S(A) \circ \sum_{i=0}^{m} A^i BC_i\big) \\
&= q_m(A)^{-2} \sum_{j=0}^{m} A^j \sum_{i=0}^{m} A^i BC_i D_j,
\end{aligned}$$

where for the second equality we have used part (i) of the lemma and the fact that there exists a polynomial $p(z)$ such that $q_m(A)^{-1} = p(A)$.

∎

Given a BRAD $A\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$, we can identify each block column of $\boldsymbol{V}_{m+1} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m+1}]$ with a rational matrix-valued function $R_j(z)$ such that $\boldsymbol{v}_{j+1} = R_j(A) \circ \boldsymbol{b}$. In other words, we can write

$$A[R_0(A) \circ \boldsymbol{b}, \ldots, R_m(A) \circ \boldsymbol{b}]\underline{\boldsymbol{K}_m} = [R_0(A) \circ \boldsymbol{b}, \ldots, R_m(A) \circ \boldsymbol{b}]\underline{\boldsymbol{H}_m}.$$

Using Lemma 4.1 (i) and (iii), this is equivalent to

$$z[R_0(z), \ldots, R_m(z)]\underline{\boldsymbol{K}_m}\Big|_{z=A} \circ \boldsymbol{b} = [R_0(z), \ldots, R_m(z)]\underline{\boldsymbol{H}_m}\Big|_{z=A} \circ \boldsymbol{b}.$$

Note that we have formally "isolated" both A and $\boldsymbol{b}$ from the decomposition using the circ operator, hence it is meaningful to consider the functions $R_j(z)$ independently from $(A, \boldsymbol{b})$. Moreover, these functions are fully encoded in the matrix pencil $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$. To give a more insightful characterization of the functions $R_j(z)$, it is helpful to transform a BRAD into an essentially equal BRAD where the subdiagonals of the upper-Hessenberg matrices are multiples of the identity (and therefore commute with all other square matrices of appropriate size).

DEFINITION 4.2. *A BRAD is* normalized *if the subdiagonal blocks of the matrix pencil* $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ *are multiples of the identity.*

By the definition of a BRAD $A\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$, the subdiagonal blocks satisfy $\nu_j H_{j+1,j} = \mu_j K_{j+1,j}$ for $j = 1, \ldots, m$ with at least one of the two matrices being nonsingular. We define $\boldsymbol{Z}_m := \mathrm{diag}(Z_1, \ldots, Z_m)$ with

$$Z_j := \begin{cases} H_{j+1,j}^{-1} & \text{if } H_{j+1,j} \text{ is nonsingular,} \\ K_{j+1,j}^{-1} & \text{otherwise.} \end{cases}$$

Setting $\underline{\widehat{\boldsymbol{K}}_m} := \underline{\boldsymbol{K}_m}\boldsymbol{Z}_m$ and $\underline{\widehat{\boldsymbol{H}}_m} := \underline{\boldsymbol{H}_m}\boldsymbol{Z}_m$, we obtain a normalized BRAD $A\widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{K}}_m} = \widehat{\boldsymbol{V}}_{m+1}\underline{\widehat{\boldsymbol{H}}_m}$ which is essentially equal to the original BRAD (by the implicit Q theorem). The following theorem gives a recursive characterization of the functions $R_j(z)$ associated with a normalized BRAD.

THEOREM 4.3. *Given a normalized BRAD with block upper-Hessenberg pencil* $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ *having subdiagonal blocks* $(\mu_j I_{s\times s}, \nu_j I_{s\times s})$, $j = 1, \ldots, m$, *and a block matrix* $\boldsymbol{V}_{m+1} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m+1}]$. *Then, for* $j = 1, \ldots, m$,

$$\boldsymbol{v}_{j+1} = R_j(A) \circ \boldsymbol{v}_1 = \sum_{i=1}^{j} (\mu_j - z\nu_j)^{-1}(zK_{ij} - H_{ij})\big|_{z=A} \circ \boldsymbol{v}_i. \tag{4.1}$$

Proof. Reading off the jth block column of the normalized BRAD yields

$$(I\mu_j - A\nu_j)\boldsymbol{v}_{j+1} = \sum_{i=1}^{j} \left(A\boldsymbol{v}_i K_{i,j} - \boldsymbol{v}_i H_{i,j}\right), \tag{4.2}$$

for $j = 1, \ldots, m$. Multiplying on the left by $(I\mu_j - A\nu_j)^{-1}$, which is guaranteed to exist by the choice of the pole μ_j/ν_j, gives

$$\boldsymbol{v}_{j+1} = \sum_{i=1}^{j} (I\mu_j - A\nu_j)^{-1} \left(A\boldsymbol{v}_i K_{i,j} - \boldsymbol{v}_i H_{i,j}\right).$$

Making use of the circ operator, we can write this as

$$\boldsymbol{v}_{j+1} = \sum_{i=1}^{j} (\mu_j - z\nu_j)^{-1} (zK_{ij} - H_{ij})\big|_{z=A} \circ \boldsymbol{v}_i.$$

∎

We can also characterize the functions $R_j(z)$ in terms of *block determinants.* Given an $s \times s$ matrix $\boldsymbol{X} = [X]$, the associated block determinant is defined as $\det^{\square}(\boldsymbol{X}) := X$. Further, the block determinant of a block upper-Hessenberg matrix $\boldsymbol{X}$ whose subdiagonal block elements are multiples of the identity,

$$\boldsymbol{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1,n} \\ x_1 I_{s\times s} & X_{22} & \cdots & X_{2,n} \\ & \ddots & \ddots & \vdots \\ & & x_{n-1} I_{s\times s} & X_{n,n} \end{bmatrix}$$

is defined recursively as

$$\det{}^{\square}(\boldsymbol{X}) := \sum_{i=1}^{n} (-1)^{i+n} \det{}^{\square}(\boldsymbol{M}_{i,n}) X_{i,n}, \tag{4.3}$$

where $\boldsymbol{M}_{i,n}$ is obtained by removing the ith block row and nth block column of $\boldsymbol{X}$. This block determinant is a special form of a *quasideterminant*; see, e.g., [25, 26]. In our case, the assumption of the block upper-Hessenberg structure and structured subdiagonal elements simplifies the following derivations significantly and is sufficient for our purposes. We have the following theorem.

THEOREM 4.4. *Given a normalized BRAD with block upper-Hessenberg pencil* $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ *having subdiagonal blocks* $(\mu_j I_{s\times s}, \nu_j I_{s\times s})$, $j = 1, \ldots, m$, *and a block matrix* $\boldsymbol{V}_{m+1} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m+1}]$. *Then, for* $j = 1, \ldots, m$,

$$\boldsymbol{v}_{j+1} = q_j(A)^{-1} \left(\det{}^{\square}(z\boldsymbol{K}_j - \boldsymbol{H}_j)\big|_{z=A} \circ \boldsymbol{v}_1 \right), \tag{4.4}$$

where $q_j(z) = \prod_{\ell=1}^{j} (\mu_\ell - z\nu_\ell)$.

Proof. Multiplying (4.2) on the left by $q_{j-1}(A)$ yields

$$q_j(A)\boldsymbol{v}_{j+1} = q_{j-1}(A) \sum_{i=1}^{j} (A\boldsymbol{v}_i K_{i,j} - \boldsymbol{v}_i H_{i,j}). \tag{4.5}$$

It is clear that (4.5) is the same as (4.4) for $j = 1$. Assume that (4.4) is true for $j = 1, \ldots, n-1$ and consider $j = n$. Substituting every occurence of $\boldsymbol{v}_i$ on the right-hand side of (4.5) by (4.4) yields

$$\begin{aligned} q_n(A)\boldsymbol{v}_{n+1} = q_{n-1}(A) \sum_{i=1}^{n} \Big[& A \left(q_{i-1}(A)^{-1} \det{}^{\square}(z\boldsymbol{K}_{i-1} - \boldsymbol{H}_{i-1})\big|_{z=A} \circ \boldsymbol{v}_1 \right) K_{i,n} \\ & - \left(q_{i-1}(A)^{-1} \det{}^{\square}(z\boldsymbol{K}_{i-1} - \boldsymbol{H}_{i-1})\big|_{z=A} \circ \boldsymbol{v}_1 \right) H_{i,n} \Big]. \end{aligned}$$

Using Lemma 4.1 (i) and (iv) yields

$$\begin{aligned} q_n(A)\boldsymbol{v}_{n+1} &= \sum_{i=1}^{n} \left[q_{n-1}(z) q_{i-1}(z)^{-1} \det{}^{\square}(z\boldsymbol{K}_{i-1} - \boldsymbol{H}_{i-1})(zK_{i,n} - H_{i,n}) \right] \Big|_{z=A} \circ \boldsymbol{v}_1 \\ &= \left[\sum_{i=1}^{n} (-1)^{i+n} \left(\det{}^{\square}(z\boldsymbol{K}_{i-1} - \boldsymbol{H}_{i-1}) (-1)^{n-i} q_{n-1}(z) q_{i-1}(z)^{-1} \right) (zK_{i,n} - H_{i,n}) \right] \Big|_{z=A} \circ \boldsymbol{v}_1. \end{aligned}$$

It remains to show that the term in the last square brackets corresponds to $\det^{\square}(z\boldsymbol{K}_n - \boldsymbol{H}_n)$, from which would then follow the required statement that

$$q_n(A)\boldsymbol{v}_{n+1} = \det{}^{\square}(z\boldsymbol{K}_n - \boldsymbol{H}_n)\Big|_{z=A} \circ \boldsymbol{v}_1.$$

Indeed, the term in square brackets can be identified with the expansion given in (4.3) by setting $X_{i,n} = (zK_{i,n} - H_{i,n})$ and

$$\det{}^{\square}(\boldsymbol{M}_{i,n}) = \det{}^{\square}(z\boldsymbol{K}_{i-1} - \boldsymbol{H}_{i-1})(-1)^{n-i} q_{n-1}(z) q_{i-1}(z)^{-1}.$$

The latter formula follows from the fact that each $\boldsymbol{M}_{i,n}$ has an $(i-1)\times(i-1)$ leading normalized block upper-Hessenberg matrix with block determinant $\det^{\square}(z\boldsymbol{K}_{i-1} - \boldsymbol{H}_{i-1})$, and an $(n-i)\times(n-i)$ block upper-triangular part along the diagonal with block determinant $(-1)^{n-i}q_{n-1}(z)q_{i-1}(z)^{-1}$. □

We have shown that a block matrix pencil $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ encodes a recursion of rational matrix-valued functions. If we fix $R_0 \equiv I_{s\times s}$, then all functions $R_j(z)$ are specified. Given a set of matrix coefficients $D_0, D_1, \ldots, D_m \in \mathbb{C}^{s\times s}$, which we collect in a block matrix $\boldsymbol{D} := [D_0, D_1, \ldots, D_m]$ for convenience, then

$$R(z) := R_0(z)D_0 + R_1(z)D_1 + \cdots + R_m(z)D_m$$

defines a rational matrix-valued function which we refer to as *RKFUNB* (short for *block rational Krylov function*). The function $R(z)$ is represented by the triplet $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m}, \boldsymbol{D})$, a block extension of the RKFUN objects introduced in [7].

The RKFUNB representation $R(z) \equiv (\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m}, \boldsymbol{D})$ allows for an efficient numerical evaluation procedure. Note that evaluating $R(z)$ at a scalar $z \in \mathbb{C}$ such that $q_m(z) \neq 0$ is equivalent to evaluating $R(A)\circ \boldsymbol{b}$ for $A = zI_{s\times s}$ and $\boldsymbol{b} = I_{s\times s}$. More generally, we consider evaluating $R(\widehat{A})\circ\widehat{\boldsymbol{b}}$ for $\widehat{A} \in \mathbb{C}^{\widehat{N}\times\widehat{N}}$ with $\xi_1, \ldots, \xi_m \notin \Lambda(\widehat{A})$ and $\widehat{\boldsymbol{b}} \in \mathbb{C}^{\widehat{N}\times s}$. We start by computing $\{R_j(\widehat{A})\circ\widehat{\boldsymbol{b}}\}_{j=0}^m$ by rerunning the block rational Arnoldi method without computing the orthogonalization coefficients, but reusing the quantities provided in $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$. This is shown in Algorithm 4.2, which constructs a new decomposition

$$\widehat{A}\widehat{\boldsymbol{V}}_{m+1}\underline{\boldsymbol{K}_m} = \widehat{\boldsymbol{V}}_{m+1}\underline{\boldsymbol{H}_m},$$

where $\widehat{\boldsymbol{V}}_{m+1} = [\widehat{\boldsymbol{b}}, R_1(\widehat{A})\circ\widehat{\boldsymbol{b}}, \ldots, R_m(\widehat{A})\circ\widehat{\boldsymbol{b}}]$. Note that this decomposition is not necessarily a BRAD as $\widehat{\boldsymbol{V}}_{m+1}$ might fail to be of full column rank. Finally, the evaluated RKFUNB is obtained by summation

$$R(\widehat{A})\circ\widehat{\boldsymbol{b}} := (R_0(\widehat{A})\circ\widehat{\boldsymbol{b}})D_0 + (R_1(\widehat{A})\circ\widehat{\boldsymbol{b}})D_1 + \cdots + (R_m(\widehat{A})\circ\widehat{\boldsymbol{b}})D_m.$$

This evaluation procedure is implemented in the Rational Krylov Toolbox since version 2.8 as a method of the `rkfunb` class. Given an RKFUNB represented as a MATLAB object `R`, the user can type `R(A,b)` to obtain $R(A)\circ\boldsymbol{b}$. We will show an application of this procedure in Section 7.1.

5. Continuation matrices. As before we work with the rank assumption (1.2). At each iteration $j = 1, 2, \ldots, m$ of the block rational Arnoldi method (Algorithm 9), we have available a BRAD

$$A\boldsymbol{V}_j\underline{\boldsymbol{K}_{j-1}} = \boldsymbol{V}_j\underline{\boldsymbol{H}_{j-1}} \tag{5.1}$$

Algorithm 4.2 Evaluating an RKFUNB

Input: $\widehat{A} \in \mathbb{C}^{\widehat{N}\times\widehat{N}}$, $\widehat{\boldsymbol{b}} \in \mathbb{C}^{\widehat{N}\times s}$, RKFUNB $R(z) \equiv (\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m}, \boldsymbol{D})$

Output: $R(\widehat{A}) \circ \widehat{\boldsymbol{b}}$

1. Set $\widehat{\boldsymbol{v}}_1 := \widehat{\boldsymbol{b}}$, $\widehat{\boldsymbol{V}}_1 := \widehat{\boldsymbol{v}}_1$
2. **for** $j = 1, \ldots, m$ **do**
3. Find $(\nu_j, \mu_j, \rho_j, \eta_j) \in \mathbb{C}^4$ such that $\mu_j K_{j+1,j} = \nu_j H_{j+1,j}$, $|\mu_j| + |\nu_j| \neq 0$, $\rho_j \mu_j \neq \eta_j \nu_j$
4. Compute $T_j := (\mu_j \boldsymbol{k}_j - \nu_j \boldsymbol{h}_j)/(\eta_j \nu_j - \rho_j \mu_j)$
5. Compute $\boldsymbol{c}_j := (\eta_j \boldsymbol{k}_j - \rho_j \boldsymbol{h}_j)/(\eta_j \nu_j - \rho_j \mu_j)$, $C_{j+1,j} := (\eta_j K_{j+1,j} - \rho_j H_{j+1,j})/(\eta_j \nu_j - \rho_j \mu_j)$
6. Compute $\boldsymbol{w}_j := (\nu_j \widehat{A} - \mu_j I)^{-1}(\rho_j \widehat{A} - \eta_j I)\widehat{\boldsymbol{V}}_j T_j$
7. Compute $\boldsymbol{w}_j := \boldsymbol{w}_j - \widehat{\boldsymbol{V}}_j \boldsymbol{c}_j$
8. Compute $\widehat{\boldsymbol{v}}_{j+1} := \boldsymbol{w}_j C_{j+1,j}^{-1}$ and set $\widehat{\boldsymbol{V}}_{j+1} := [\widehat{\boldsymbol{V}}_j, \widehat{\boldsymbol{v}}_{j+1}]$
9. **end for**
10. Compute $R(\widehat{A}) \circ \widehat{\boldsymbol{b}} := \sum_{j=0}^{m} \widehat{\boldsymbol{v}}_{j+1} D_j$

which we would like to extend by an orthonormalized version of the block vector

$$\boldsymbol{w}_j := (\nu_j A - \mu_j I)^{-1}(\rho_j A - \eta_j I) \cdot \boldsymbol{V}_j T_j.$$

To select an appropriate block element $\boldsymbol{V}_j T_j \in \mathcal{Q}_j^{\square}(A, \boldsymbol{v}_1)$ for the extension, a continuation matrix $T_j \in \mathbb{C}^{js\times s}$ of rank s needs to be chosen. By Theorem 4.3, $\boldsymbol{V}_j T_j$ can be written in terms of a rational matrix-valued function

$$\boldsymbol{V}_j T_j = \boldsymbol{v}_1 T_{1,1} + \boldsymbol{v}_2 T_{2,1} + \cdots + \boldsymbol{v}_j T_{j,1} =: R(A) \circ \boldsymbol{v}_1,$$

where $R(z) = q_{j-1}(z)^{-1}(C_0 + zC_1 + \cdots + z^{j-1}C_{j-1})$ for $C_0, C_1, \ldots, C_{j-1} \in \mathbb{C}^{s\times s}$. Let us investigate the case of an *avoidable breakdown* of the orthogonalization procedure with a finite pole $\xi_j = \mu_j/\nu_j$, where there exists a vector $x \in \mathbb{C}^s \setminus \{0\}$ and another function $S(z) = q_{j-1}(z)^{-1}(D_0 + zD_1 + \cdots + z^{j-1}D_{j-1})$ for $D_0, D_1, \ldots, D_{j-1} \in \mathbb{C}^{s\times s}$ such that

$$[(A - \xi_j I)^{-1}(\rho_j A - \eta_j I) \cdot R(A) \circ \boldsymbol{v}_1]x = [S(A) \circ \boldsymbol{v}_1]x. \tag{5.2}$$

This means that the columns of the block vector $\boldsymbol{w}_j$ can be linearly combined into a vector already contained in the span of $\boldsymbol{V}_j$, hence not expanding this space by s new directions. Left-multiplying (5.2) by the nonsingular matrix $q_{j-1}(A)(A - \xi_j I)$, it can be rewritten in the equivalent form

$$(\rho_j z - \eta_j)R(z)x = (z - \xi_j)S(z)x.$$

Comparing the coefficients of the independent variable z in

$$(\rho_j z - \eta_j)(C_0 + zC_1 + \cdots + z^{j-1}C_{j-1})x = (z - \xi_j)(D_0 + zD_1 + \cdots + z^{j-1}D_{j-1})x,$$

results in the following system of equations:

$$\begin{array}{lcl}\rho_j C_{j-1}x & = & D_{j-1}x,\\ (\rho_j C_{j-2}-\eta_j C_{j-1})x & = & (D_{j-2}-\xi_j D_{j-1})x,\\ & \vdots & \\ (\rho_j C_0-\eta_j C_1)x & = & (D_0-\xi_j D_1)x,\\ -\eta_j C_0 x & = & -\xi_j D_0 x.\end{array}$$

This can be rewritten as a polynomial (nonlinear) eigenvalue problem [30, 51]

$$(\rho_j\xi_j-\eta_j)(C_0+\xi_j C_1+\xi_j^2 C_2+\cdots+\xi_j^{j-1}C_{j-1})x=0.$$

The factor $(\rho_j\xi_j-\eta_j)$ is nonzero by our assumption $\xi_j=\mu_j/\nu_j\neq\eta_j/\rho_j$ (see Algorithm 9). Consequently, given a block vector $\boldsymbol{V}_jT_j=R(A)\circ\boldsymbol{v}_1$, an avoidable breakdown will occur if ξ_j is an eigenvalue of the polynomial eigenvalue problem

$$(C_0+zC_1+z^2C_2+\cdots+z^{j-1}C_{j-1})x=0.$$

Throughout the literature, the most commonly used continuation matrix is such that $\boldsymbol{V}_jT_j=\boldsymbol{v}_j$, i.e.,

$$T_j=\begin{bmatrix}O_{s\times s} & \cdots & O_{s\times s} & I_{s\times s}\end{bmatrix}^T. \tag{5.3}$$

With this continuation strategy, which always uses the last computed block basis vector and is henceforth referred to as strategy `'last'`, one can read off the "forbidden poles" (which would lead to an avoidable breakdown) as the eigenvalues of the matrix pencil in (5.1).

THEOREM 5.1. *Given a BRAD* (5.1) *with* $\boldsymbol{V}_j=[\boldsymbol{v}_1,\ldots,\boldsymbol{v}_j]$, *and write* $\boldsymbol{v}_{j-1}=R(A)\circ\boldsymbol{v}_1$. *The eigenvalues* λ *of the nonlinear eigenvalue problem* $x^TR(\lambda)=0^T$ *are contained in the set of generalized eigenvalues of* $(\boldsymbol{H}_{j-1},\boldsymbol{K}_{j-1})$.

Proof. Consider (5.1) in the "scalar" form

$$z[R_0(z),\ldots,R_{j-2}(z),R(z)]\underline{\boldsymbol{K}_{j-1}}=[R_0(z),\ldots,R_{j-2}(z),R(z)]\underline{\boldsymbol{H}_{j-1}},$$

merge the last two columns and rearrange to get

$$[R_0(z),\ldots,R_{j-2}(z)](z\boldsymbol{K}_{j-1}-\boldsymbol{H}_{j-1})=R(z)(H_{j,j-1}-zK_{j,j-1})\boldsymbol{E}_{j-1}^T,$$

where $\boldsymbol{E}_{j-1}^T=[O_{s\times s},\ldots,O_{s\times s},I_{s\times s}]\in\mathbb{R}^{s\times(j-1)s}$. Suppose that $\lambda\in\mathbb{C}$ is an eigenvalue of $R(z)$ with corresponding left eigenvector $x\in\mathbb{C}^s\setminus\{0\}$, i.e., $x^TR(\lambda)=0^T$. Then

$$x^T[R_0(\lambda),\ldots,R_{j-2}(\lambda)](\lambda\boldsymbol{K}_{j-1}-\boldsymbol{H}_{j-1})=0^T,$$

and so $(\lambda,x^T[R_0(\lambda),\ldots,R_{j-2}(\lambda)])$ is a left eigenpair of $(\boldsymbol{H}_{j-1},\boldsymbol{K}_{j-1})$. □

Theorem 5.1 provides an easy way to check whether a given pole ξ_j should be avoided when the continuation strategy `'last'`, according to (5.3), is used. If all poles ξ_j are pairwise distinct, another strategy is to use

$$T_j=\begin{bmatrix}I_{s\times s} & O_{s\times s} & \cdots & O_{s\times s}\end{bmatrix}^T.$$

This strategy, called `'first'`, allows for parallelism in the construction of the block Krylov basis vectors, but even for the case $s = 1$ it can lead to severe numerical instabilities [10].

In practice we are often in the situation where the pole sequence $\xi_1, \xi_2, \ldots, \xi_m$ is given a priori and we wish to have a continuation strategy that avoids breakdowns and numerical instabilities. For the single-vector ($s = 1$) case, Ruhe [47] introduced a continutation vector designed to prevent breakdowns caused by the cancelation of a common root in the numerator and denominator of the rational function underlying a rational Krylov vector. We now extend this strategy, referred to as `'ruhe'`, to the block case: again we assume that at iteration j of Algorithm 2.1 we have a BRAD (5.1) which we would like to extend by a new block vector using the pole $\xi_j = \mu_j/\nu_j \neq \infty$. We start with computing a full QR factorization of

$$\nu_j \underline{\boldsymbol{H}_{j-1}} - \mu_j \underline{\boldsymbol{K}_{j-1}} = Q_j R_{j,j-1}. \tag{5.4}$$

The involved matrices are of size $\nu_j \underline{\boldsymbol{H}_{j-1}} - \mu_j \underline{\boldsymbol{K}_{j-1}} \in \mathbb{C}^{js \times (j-1)s}$, $Q_j \in \mathbb{C}^{js \times js}$, and $R_{j,j-1} \in \mathbb{C}^{js \times (j-1)s}$. Note that the subscripts refer to the iteration index and not the matrix sizes. By Lemma 3.2 (iii) we know that $R_{j-1,j-1}$, the upper $(j-1)s \times (j-1)s$ part of $R_{j,j-1}$, is nonsingular.

Multiplying the BRAD by $\nu_j \eta_j$, subtracting $\mu_j \eta_j \, \boldsymbol{V}_j \underline{\boldsymbol{K}_{j-1}}$ from both sides and rearranging yields

$$(\nu_j A - \mu_j I)\, \boldsymbol{V}_j \eta_j \underline{\boldsymbol{K}_{j-1}} = \eta_j \, \boldsymbol{V}_j (\nu_j \underline{\boldsymbol{H}_{j-1}} - \mu_j \underline{\boldsymbol{K}_{j-1}}). \tag{5.5}$$

Similarly, multiplying the BRAD by $\rho_j \mu_j$, subtracting $\rho_j \nu_j A \, \boldsymbol{V}_j \underline{\boldsymbol{H}_{j-1}}$ from both sides and rearranging yields

$$(\nu_j A - \mu_j I)\, \boldsymbol{V}_j \rho_j \underline{\boldsymbol{H}_{j-1}} = \rho_j A \, \boldsymbol{V}_j (\nu_j \underline{\boldsymbol{H}_{j-1}} - \mu_j \underline{\boldsymbol{K}_{j-1}}). \tag{5.6}$$

Subtracting (5.5) from (5.6) and rearranging gives the decomposition

$$(\nu_j A - \mu_j I)^{-1} (\rho_j A - \eta_j I)\, \boldsymbol{V}_j (\nu_j \underline{\boldsymbol{H}_{j-1}} - \mu_j \underline{\boldsymbol{K}_{j-1}}) = \boldsymbol{V}_j (\rho_j \underline{\boldsymbol{H}_{j-1}} - \eta_j \underline{\boldsymbol{K}_{j-1}}).$$

Substituting the QR factorization (5.4) on the left and inserting identity on the right, we have

$$(\nu_j A - \mu_j I)^{-1} (\rho_j A - \eta_j I)\, \boldsymbol{V}_j Q_j R_{j,j-1} = \boldsymbol{V}_j Q_j Q_j^* (\rho_j \underline{\boldsymbol{H}_{j-1}} - \eta_j \underline{\boldsymbol{K}_{j-1}}).$$

The matrix $R_{j,j-1}$ has s more rows than columns, and so the last s rows must be zero. Multiplying on the right by the inverse of $R_{j-1,j-1}$ leads to

$$(\nu_j A - \mu_j I)^{-1} (\rho_j A - \eta_j I)\, \boldsymbol{W}_{j-1} = \boldsymbol{W}_j F_{j-1}, \tag{5.7}$$

where $\boldsymbol{W}_j := \boldsymbol{V}_j Q_j$ and $F_{j-1} := Q_j^* (\rho_j \underline{\boldsymbol{H}_{j-1}} - \eta_j \underline{\boldsymbol{K}_{j-1}}) R_{j-1,j-1}^{-1}$.

Our aim is to find a continuation combination $\boldsymbol{V}_j T_j$ such that multiplying by $(\nu_j A - \mu_j I)^{-1} (\rho_j A - \eta_j I)$ expands the space. By (5.7) we know that any vector in the span of $\boldsymbol{W}_{j-1}$ does not enlarge the span of $\boldsymbol{W}_j = \boldsymbol{V}_j Q_j$, and so we should choose vectors from the orthogonal complement of $\boldsymbol{W}_{j-1}$ in the span of $\boldsymbol{V}_j$. This can be achieved by selecting the last s columns of Q_j as the continuation matrix. Hence, the `'ruhe'` continuation strategy extended to the block case becomes

$$T_j = \begin{cases} I_{s \times s} & \text{if } j = 1, \\ Q_j(\texttt{:}, \texttt{end} - s + 1 : \texttt{end}) & \text{otherwise.} \end{cases} \tag{5.8}$$

In Section 7.2 we will provide a numerical illustration of the benefits of the continuation strategy `'ruhe'` over the more commonly used strategy `'last'`.

REMARK 5.2. *It is worth noting that, when constructing a rational Krylov basis using Algorithm 2.1 with a pole* $\xi_j = \xi_{j-1}$ *repeated from the previous iteration, the continuation strategies* `'ruhe'` *and* `'last'` *are essentially equivalent. With* $\xi_{j-1} = \mu_{j-1}/\nu_{j-1}$ *we have that* $\mu_{j-1}K_{j,j-1} = \nu_{j-1}H_{j,j-1}$ *in* (5.1)*. Hence, the final* s *rows of the matrix* $\nu_j \underline{\boldsymbol{H}_{j-1}} - \mu_j \underline{\boldsymbol{K}_{j-1}}$ *in* (5.4) *will be zero and therefore the last* s *columns of* Q_j *must be of the form*

$$T_j = \begin{bmatrix} O_{s\times s} & \cdots & O_{s\times s} & U_{s\times s} \end{bmatrix}^T$$

for any unitary matrix $U_{s\times s}$*, which can be chosen as* $I_{s\times s}$*. This is the same continuation matrix as* (5.3)*.*

6. Deflation. So far we have worked with the full rank assumption (1.2), but in practice exact rank deficiencies may occur in the Krylov matrix $[\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^j\boldsymbol{b}]$, or the basis vectors $\boldsymbol{w}_j$ computed in line 5 of Algorithm 2.1 might be such that $[\,\boldsymbol{V}_j, \boldsymbol{w}_j]$ is badly conditioned. If the (nearly) dependent vectors in $\boldsymbol{w}_j$ are not removed, the matrix $C_{j+1,j}$ computed in line 8 of the algorithm is (nearly) singular, resulting in a (near) breakdown of the orthonormalization procedure. The removal of the (nearly) linearly dependent vectors from $\boldsymbol{w}_j$ is known as *deflation*.

Deflation can implemented by a straightforward modification of Algorithm 2.1. Assume that the block vector $\boldsymbol{w}_j$ in line 8 has $s_j \le s$ columns. We can compute a thin SVD or column-pivoted QR factorization which, in both cases, is of the form

$$\boldsymbol{w}_j = Q_j R_j P_j^* = \left[Q_j^{(k)}, Q_j^{(d)}\right] \begin{bmatrix} R_j^{(k)} \\ R_j^{(d)} \end{bmatrix} P_j^*,$$

where $Q_j^{(k)} \in \mathbb{C}^{N\times s_{j+1}}$ and $Q_j^{(d)} \in \mathbb{C}^{N\times(s_j - s_{j+1})}$ have (mutually) orthonormal columns, $R_j^{(k)} \in \mathbb{C}^{s_{j+1}\times s_j}$, $R_j^{(d)} \in \mathbb{C}^{(s_j-s_{j+1})\times s_j}$, and $P_j \in \mathbb{C}^{s_j\times s_j}$ is unitary. Here, $Q_j^{(k)}$ corresponds to the columns to keep, and $Q_j^{(d)}$ to the columns to deflate. The matrix $R_j^{(k)}$ is of upper trapezoidal form and $R_j^{(d)}$ should have small norm. From here, we have two options:

(i) We define the next basis vector $\boldsymbol{v}_{j+1} := Q_j^{(k)}$ and set $C_{j+1,j} := R_j^{(k)} P_j^* \in \mathbb{C}^{s_{j+1}\times s_j}$. When combined into an approximate BRAD

$$A\,\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m} + \boldsymbol{S}_m,$$

the block upper-Hessenberg matrices $\underline{\boldsymbol{H}_m}$ and $\underline{\boldsymbol{K}_m}$ are of the format (only shown for $\underline{\boldsymbol{H}_m}$):

$$\underline{\boldsymbol{H}_m} = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ \\ s_m \\ s_{m+1} \end{array} \!\!\begin{array}{c} \begin{array}{ccccc} s_1 & s_2 & s_3 & & s_m \end{array} \\ \begin{bmatrix} H_{1,1} & H_{1,2} & H_{3,1} & \cdots & H_{m,1} \\ H_{2,1} & H_{2,2} & H_{3,2} & \cdots & H_{m,2} \\ & H_{3,2} & H_{3,3} & \cdots & H_{m,3} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & H_{m,m} \\ & & & & H_{m+1,m} \end{bmatrix} \end{array},$$

and the block columns of the residual matrix $\boldsymbol{S}_m = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m]$ are

$$\boldsymbol{s}_j = -(\nu_j A - \mu_j I) Q_j^{(d)} R_j^{(d)} P_j^*.$$

This type of *fat decomposition,* used by Gutknecht in [28] for block polynomial Krylov spaces, ensures that the diagonal blocks $H_{j,j}$ and $K_{j,j}$ are square. The sequence $s_1, s_2, \ldots, s_{m+1}$ is monotonically nonincreasing and at least one of the subdiagonal blocks $H_{j+1,j}$ or $K_{j+1,j}$ is of maximal rank s_{j+1}. This means that, given s_1, one can infer the complete block structure of the pencil $(\underline{\boldsymbol{H}_m}, \underline{\boldsymbol{K}_m})$ by calculating the ranks of the subdiagonal blocks.

(ii) A *thin decomposition* $A\boldsymbol{V}_{m+1}\underline{\widetilde{\boldsymbol{K}}_m} = \boldsymbol{V}_{m+1}\underline{\widetilde{\boldsymbol{H}}_m}$ can obtained from a fat decomposition $A\boldsymbol{V}_{m+1}\underline{\boldsymbol{K}_m} = \boldsymbol{V}_{m+1}\underline{\boldsymbol{H}_m}$ via right-multiplication by a (rectangular) block-diagonal matrix $\boldsymbol{D}_m = \operatorname{diag}(D_1, \ldots, D_m)$. The matrices $D_j \in \mathbb{C}^{s_j \times s_{j+1}}$ are chosen such that at least one of $\widetilde{H}_{j+1,j} := H_{j+1,j} D_j \in \mathbb{C}^{s_{j+1} \times s_{j+1}}$ and $\widetilde{K}_{j+1,j} := K_{j+1,j} D_j \in \mathbb{C}^{s_{j+1} \times s_{j+1}}$ is nonsingular. This can be achieved, for example, by setting $D_j = [e_1, e_2, \ldots, e_{s_{j+1}}]$, where $e_i \in \mathbb{C}^{s_{j+1}}$ is the ith unit vector. The resulting block upper-Hessenberg matrices $\underline{\widetilde{\boldsymbol{H}}_m}$ and $\underline{\widetilde{\boldsymbol{K}}_m}$ are of the format (only shown for $\underline{\widetilde{\boldsymbol{H}}_m}$):

$$\underline{\widetilde{\boldsymbol{H}}_m} = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ \\ s_m \\ s_{m+1} \end{array} \begin{array}{c} \begin{array}{ccccc} s_2 & s_3 & s_4 & \cdots & s_{m+1} \end{array} \\ \left[\begin{array}{ccccc} H_{1,1} & H_{1,2} & H_{3,1} & \cdots & H_{m,1} \\ H_{2,1} & H_{2,2} & H_{3,2} & \cdots & H_{m,2} \\ & H_{3,2} & H_{3,3} & \cdots & H_{m,3} \\ & & \ddots & \ddots & \vdots \\ & & & & H_{m,m} \\ & & & & H_{m+1,m} \end{array}\right] \end{array}.$$

In this type of thin decomposition, all the subdiagonal blocks $H_{j+1,j}$ and $K_{j+1,j}$ are square.

The MATLAB Rational Krylov Toolbox implements both types of decompositions in the `rat_krylov` function; see [8] for details.

7. Examples.

In this section we provide numerical illustrations using the MATLAB Rational Krylov Toolbox version 2.8[†]. All the experiments were performed with the 64-bit version of MATLAB 2018a on a machine equipped with an Intel I7-6700 processor running at 3.40 GHz.

7.1. Vector autoregression via RKFUNB.

In Section 4, we introduced the RKFUNB framework which provides a Krylov-based representation of rational matrix-valued functions. Matrix-valued functions have applications to modeling multivariate time series, with one particular example being vector autoregression (VAR). Consider the realization of a multivariate time series $\boldsymbol{y} = [y_1, \ldots, y_s] \in \mathbb{R}^{N \times s}$ $(s \ll N)$ such that $\boldsymbol{y}$ is of maximal rank s. Assuming that this times series is generated by a VAR(p) process with mean zero, it can be written as

$$\boldsymbol{y}_t = \boldsymbol{y}_{t-1} C_1 + \cdots + \boldsymbol{y}_{t-p} C_p + \boldsymbol{\varepsilon}_t,$$

[†]Available for download from `http://rktoolbox.org`.

where $\boldsymbol{y}_t$ refers to the t-th row of $\boldsymbol{y}$, $C_1, \ldots, C_p \in \mathbb{R}^{s \times s}$, and $\boldsymbol{\varepsilon}_t$ is a white noise process [39, Chapter 2]. The coefficients $C_1, \ldots, C_p$ can be estimated by solving the least squares problem

$$\min_{C_1,\ldots,C_p} \left\| \begin{pmatrix} \boldsymbol{y}_{p+1} \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} - \begin{pmatrix} \boldsymbol{y}_p \\ \vdots \\ \boldsymbol{y}_{N-1} \end{pmatrix} C_1 - \cdots - \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_{N-p} \end{pmatrix} C_p \right\|_2^2 . \tag{7.1}$$

Let us define the matrices

$$A = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} \in \mathbb{R}^{N \times N}, \quad D = \operatorname{diag}(\underbrace{1, \ldots, 1}_{N-p}, \underbrace{0 \ldots, 0}_{p}) \in \mathbb{R}^{N \times N},$$

and the seminorm $\|\boldsymbol{y}\|_D := \|D\boldsymbol{y}\|_2$. Then (7.1) can be written more concisely as

$$\min_{C_1,\ldots,C_p} \left\| A^p \boldsymbol{y} - A^{p-1}\boldsymbol{y}C_1 - \cdots - A^0 \boldsymbol{y} C_p \right\|_D^2 = \min_{P(z)} \left\| A^p \boldsymbol{y} - P(A) \circ \boldsymbol{y} \right\|_D^2, \tag{7.2}$$

where the minimization on the right-hand side is over all matrix polynomials $P(z) = z^{p-1}C_1 + \cdots + z^0 C_p$. Using the block rational Arnoldi method, we can generate a block basis $\boldsymbol{V}_p = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p]$ which blockspans $\mathcal{K}_p^{\square}(A, \boldsymbol{y})$ and is such that $\boldsymbol{y} = \boldsymbol{v}_1 R$ with nonsingular $R \in \mathbb{R}^{s \times s}$, and $\boldsymbol{V}_p^* D \boldsymbol{V}_p = I_{ps \times ps}$. Using Lemma 4.1 (ii)–(iii), we can as well perform the minimization (7.1) using the block basis $\boldsymbol{V}_p$ as

$$\min_{P(z)} \left\| A^p(\boldsymbol{v}_1 R) - P(A) \circ (\boldsymbol{v}_1 R) \right\|_D^2 = \min_{\widehat{P}(z)} \left\| [A^p \boldsymbol{v}_1 - \widehat{P}(A) \circ \boldsymbol{v}_1] R \right\|_D^2, \tag{7.3}$$

where $\widehat{P}(z) = RP(z)R^{-1}$ and the minimizer $\widehat{P}(A) \circ \boldsymbol{v}_1$ is naturally represented as an RKFUNB object. Single-step predictions can now be obtained by repeatedly applying $\widehat{P}(A)$ to the time series data.

The intimate connection between univariate autoregressive modeling and polynomial Krylov spaces is well known, with [17] being one of the key references on this topic. The connection between the multivariate case and block Krylov methods is natural. Here we want to demonstrate the use of block Krylov spaces on a time series example taken from [39, Example 3.2.3]‡. The $s = 3$ time series under consideration, stored in the matrix $\boldsymbol{y}$, correspond to first-order differences of the logarithms of quarterly seasonally adjusted West German fixed investment, disposable income, and consumption expenditures. The aim is to forecast these time series using a VAR(2) model. Figure 7.1 shows a snippet of MATLAB code performing a one-step VAR(2) prediction using the RKFUNB class of the Rational Krylov Toolbox. The predictions shown in Figure 7.2 are visually identical to those in [39, Fig. 3.3, Example 3.5.4].

7.2. Block continuation strategies. We now consider the INLET problem from the Oberwolfach Model Reduction Benchmark Collection [1], an active control model of a supersonic engine inlet; see also [35]. There are two nonsymmetric matrices $A, E \in \mathbb{R}^{N \times N}$, a block vector $\boldsymbol{b} \in \mathbb{R}^{N \times 2}$, and a row vector $c^T \in \mathbb{R}^{1 \times N}$ with $N =$

‡Data available from `http://www.jmulti.de/data_imtsa.html` under the filename `e1.dat`.

```
s = 3; % number of time series
[V, K, H, out] = rat_krylov(A, y, xi, param);
invR = inv(out.R(1:s, 1:s));
C = param.inner_product(A^2*y, V);
coeffs = {C(1:s, :), C(s+1:2*s, :)};
r = rkfunb(K, H, coeffs); % construct RKFUNB

Ahat = [0, 1; 0, 0];
yhat = y(end-1:end, :);
prediction = r(Ahat, yhat*invR)
```

Fig. 7.1: MATLAB code to construct a VAR(2) model of $\boldsymbol{y}$ and perform a one-step prediction using the RKFUNB class in the Rational Krylov Toolbox.

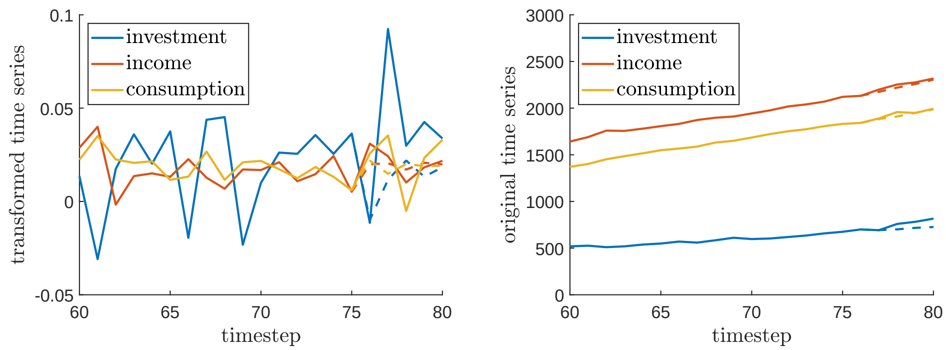

Fig. 7.2: Four-step forecasts of the investment/income/consumption system computed using RKFUNBs. The solid lines correspond to the observed values while the dashed lines are the predictions. These plots replicate the VAR(2) predictions in [39, Fig. 3.3, Example 3.5.4]. Left: Transformed time series which is obtained by taking the logarithm of the original one and then differencing. Right: Original time series and the back-transformed predictions.

11730. We consider the problem of approximating the transfer function $H(s) = c^T(sE - A)^{-1}\boldsymbol{b}$ over a range of frequencies.

We compare two different sequences of poles denoted by $\xi^{(1)}$ and $\xi^{(2)}$. The first sequence $\xi^{(1)}$ is chosen equal to that in [10, Section 5.2], with four poles equidistantly placed on the interval $i[0, 40]$ and cyclically repeated until we have 24 poles in total. The second sequence $\xi^{(2)}$ is identical to $\xi^{(1)}$ except for the 13th pole being changed to $0.996000 - 0.0762000i$, which is close to an eigenvalue $0.996026 - 0.0762341i$ of the matrix pencil $(\boldsymbol{H}_{12}, \boldsymbol{K}_{12})$. Classic Gram–Schmidt orthogonalization is used without reorthogonalization. We construct the rational Krylov space for the matrix pencil (A, E) and starting block vector $\boldsymbol{b}$ using three different continuation strategies. The strategies `'last'` and `'ruhe'` are those defined in Section 5. The strategy `'last_4'` is a modification of `'last'` with parallelization parameter $p = 4$; see [10]. It allows for the parallel computation of four block basis vectors at a time, but is known to be prone to numerical instabilities.

For each pole sequence $\xi^{(1)}$ and $\xi^{(2)}$ and each continuation strategy, we report three different quantities `cond`, `orth`, and `space` in Table 7.1. Here, `cond` $= \kappa(\boldsymbol{W}_{m+1}D)$ is the condition number of the rescaled rational Krylov basis $\boldsymbol{W}_{m+1}D$

before orthogonalization, where $\boldsymbol{W}_{m+1} = [R, \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m]$ with the quantities computed in lines 1 and 5 of Algorithm 2.1, and D is a diagonal matrix chosen such that $\kappa(\boldsymbol{W}_{m+1}D)$ is (approximately) minimized. Furthermore, `orth` $= \| \widetilde{\boldsymbol{V}}_{m+1}^* \widetilde{\boldsymbol{V}}_{m+1} - I_{(m+1)s\times(m+1)s}\|_2$ and `space` $= \| \boldsymbol{V}_{m+1}(\boldsymbol{V}_{m+1}^* \widetilde{\boldsymbol{V}}_{m+1}) - \widetilde{\boldsymbol{V}}_{m+1}\|_2$, where $\widetilde{\boldsymbol{V}}_{m+1}$ is the rational Krylov basis computed using double precision and $\boldsymbol{V}_{m+1}$ is computed using quadruple precision via the Multiprecision Computing Toolbox [3].

Table 7.1 shows that the condition number of the block basis being orthogonalized is smallest with the continuation strategy `'ruhe'`, and the computed block vectors are closer to being orthonormal after a single orthogonalization step. The 13th pole in $\xi^{(2)}$ results in a near-breakdown and an inaccurate basis when the continuation strategy `'last'` is used, while the strategy `'ruhe'` still performs robustly.

Table 7.1: Inlet example

	$\xi^{(1)}$			$\xi^{(2)}$	
	`'last'`	`'ruhe'`	`'last_4'`	`'last'`	`'ruhe'`
`cond`	8.7×10^{5}	4.2×10^{3}	8.7×10^{5}	1.4×10^{9}	4.8×10^{4}
`orth`	3.1×10^{-9}	3.2×10^{-11}	7.4×10^{-10}	1.9×10^{-4}	2.7×10^{-10}
`space`	5.0×10^{-6}	1.5×10^{-11}	5.2×10^{-10}	7.3×10^{-3}	1.9×10^{-10}

8. Conclusions and future work. In view of the favourable numerical stability observed with the continuation strategy `'ruhe'`, which we have extended to the block case in Section 5, we believe this should be the default strategy for the (block) rational Arnoldi method. It now is the default choice in the Rational Krylov Toolbox. While the RKFUNB framework is applicable to vector autoregressive modeling as demonstrated in Section 7.1, we currently do not know how to obtain rational models of vector autoregression with moving averages (VARMA). This would require a block-version of the RKFIT pole relocation strategy developed in [11].

The connections between rational Krylov methods and nonlinear eigenvalues of rational matrix-valued functions might open several research directions. For example, it does not seem to be understood how these nonlinear eigenvalues relate to the eigenvalues of the matrix A in the block case, and a generalization of a convergence theory such as that in [5] is currently lacking. An example illustrating the different convergence behavior of single-vector and block rational Ritz values for a Wilkinson matrix can be found in the example collection of the RKToolbox.

Acknowledgements. We thank the Engineering and Physical Sciences Research Council (EPRSC) and Sabisu for providing Steven Elsworth with a CASE PhD studentship. We are grateful to Kathryn Lund for discussions and constructive comments that improved the presentation.

REFERENCES

[1] *Benchmark Collection. Oberwolfach Model Reduction Benchmark Collection*, 2003.
[2] O. Abidi, M. Hached, and K. Jbilou, *Adaptive rational block Arnoldi methods for model reductions in large-scale MIMO dynamical systems*, New Trends Math. Sci., 4 (2016), pp. 227–239.
[3] Advanpix LLC., *Multiprecision Computing Toolbox for MATLAB*, ver 4.4.7.12739, Tokyo, Japan, 2018. `http://www.advanpix.com/`.

[4] H. Barkouki, A. H. Bentbib, and K. Jbilou, *An adaptive rational block Lanczos-type algorithm for model reduction of large scale dynamical systems*, J. Sci. Comput., 67 (2016), pp. 221–236.
[5] B. Beckermann, S. Güttel, and R. Vandebril, *On the convergence of rational Ritz values*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1740–1774.
[6] A. H. Bentbib, K. Jbilou, and E. Sadek, *On some extended block Krylov based methods for large scale nonsymmetric Stein matrix equations*, Mathematics, 5 (2017), p. 21.
[7] M. Berljafa, *Rational Krylov Decompositions: Theory and Applications*, PhD thesis, The University of Manchester, 2017.
[8] M. Berljafa, S. Elsworth, and S. Güttel, *A Rational Krylov Toolbox for MATLAB*, MIMS EPrint 2014.56, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2014. http://rktoolbox.org/.
[9] M. Berljafa and S. Güttel, *Generalized rational Krylov decompositions with an application to rational approximation*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 894–916.
[10] ———, *Parallelization of the rational Arnoldi algorithm*, SIAM J. Sci. Comput., 39 (2017), pp. S197–S221.
[11] ———, *The RKFIT algorithm for nonlinear rational approximation*, SIAM J. Sci. Comput., 39 (2017), pp. A2049–A2071.
[12] S. Birk, *Deflated Shifted Block Krylov Subspace Methods for Hermitian Positive Definite Matrices*, PhD thesis, Bergische Universität Wuppertal, 2015.
[13] W. E. Boyse and A. A. Seidl, *A block QMR method for computing multiple simultaneous solutions to complex symmetric systems*, SIAM J. Sci. Comput., 17 (1996), pp. 263–274.
[14] H. Calandra, S. Gratton, R. Lago, X. Vasseur, and L. M. Carvalho, *A modified block flexible GMRES method with deflation at each iteration for the solution of non-Hermitian linear systems with multiple right-hand sides*, SIAM J. Sci. Comput., 35 (2013), pp. S345–S367.
[15] H. Calandra, S. Gratton, J. Langou, X. Pinel, and X. Vasseur, *Flexible variants of block restarted GMRES methods with application to geophysics*, SIAM J. Sci. Comput., 34 (2012), pp. A714–A736.
[16] J. Cullum and W. Donath, *A block Lanczos algorithm for computing the q algebraically largest eigenvalues and a corresponding eigenspace of large, sparse, real symmetric matrices*, in IEEE Conference on Decision and Control, vol. 13, 1974, pp. 505–509.
[17] G. Cybenko, *Restrictions of normal operators, Padé approximation and autoregressive time series*, SIAM J. Math. Anal., 15 (1984), pp. 753–767.
[18] L. Du, T. Sogabe, B. Yu, Y. Yamamoto, and S.-L. Zhang, *A block IDR(s) method for nonsymmetric linear systems with multiple right-hand sides*, J. Comput. Appl. Math., 235 (2011), pp. 4095–4106.
[19] A. El Guennouni, K. Jbilou, and A. J. Riquet, *Block Krylov subspace methods for solving large Sylvester equations*, Numer. Algorithms, 29 (2002), pp. 75–96.
[20] A. El Guennouni, K. Jbilou, and H. Sadok, *The block Lanczos method for linear systems with multiple right-hand sides*, Appl. Numer. Math., 51 (2004), pp. 243–256.
[21] 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.
[22] I. M. Elfadel and D. D. Ling, *A block rational Arnoldi algorithm for multipoint passive model-order reduction of multiport RLC networks*, in Proceedings of the 1997 IEEE/ACM International Conference on Computer-Aided Design, IEEE Computer Society, 1997, pp. 66–71.
[23] R. W. Freund, *Krylov-subspace methods for reduced-order modeling in circuit simulation*, J. Comput. Appl. Math., 123 (2000), pp. 395–421.
[24] 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.
[25] I. Gelfand, S. Gelfand, V. Retakh, and R. L. Wilson, *Quasideterminants*, Adv. Math., 193 (2005), pp. 56–141.
[26] I. M. Gelfand and V. S. Retakh, *Determinants of matrices over noncommutative rings*, Funktsional. Anal. i Prilozhen., 25 (1991), pp. 13–25, 96.
[27] G. H. Golub and R. Underwood, *The block Lanczos method for computing eigenvalues*, in Mathematical Software, III (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1977), Academic Press, New York, 1977, pp. 361–377.
[28] 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, Anshan Ltd, 2006/2007, pp. 420–447.
[29] M. H. Gutknecht and T. Schmelzer, *The block grade of a block Krylov space*, Linear Algebra

Appl., 430 (2009), pp. 174–185.
[30] S. Güttel and F. Tisseur, *The nonlinear eigenvalue problem*, Acta Numer., 26 (2017), pp. 1–94.
[31] M. Heyouni and K. Jbilou, *An extended block Arnoldi algorithm for large-scale solutions of the continuous-time algebraic Riccati equation*, Electron. Trans. Numer. Anal., 33 (2008/09), pp. 53–62.
[32] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, second ed., 2002.
[33] K. Jbilou, *Block Krylov subspace methods for large algebraic Riccati equations*, Numer. Algorithms, 34 (2003), pp. 339–353. International Conference on Numerical Algorithms, Vol. II (Marrakesh, 2001).
[34] M. D. Kent, *Chebyshev, Krylov, Lanczos: Matrix Relationships and Computations*, PhD thesis, Stanford University, 1990.
[35] G. Lassaux and K. Willcox, *Model reduction for active control design using multiple-point Arnoldi methods*, AIAA, 616 (2003), pp. 1–11.
[36] J. G. Lewis, *Algorithms for Sparse Matrix Eigenvalue Problems*, PhD thesis, Stanford University, 1977.
[37] L. Lopez and V. Simoncini, *Preserving geometric properties of the exponential matrix by block Krylov subspace methods*, BIT, 46 (2006), pp. 813–830.
[38] K. Lund, *A New Block Krylov Subspace Framework with Applications to Functions of Matrices Acting on Multiple Vectors*, PhD thesis, Temple University, 2018.
[39] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin, 2005.
[40] T. Mach, M. S. Pranić, and R. Vandebril, *Computing approximate (block) rational Krylov subspaces without explicit inversion with extensions to symmetric matrices*, Electron. Trans. Numer. Anal., 43 (2014/15), pp. 100–124.
[41] M. Malhotra, R. W. Freund, and P. M. Pinsky, *Iterative solution of multiple radiation and scattering problems in structural acoustics using a block quasi-minimal residual algorithm*, Comput. Methods Appl. Mech. Engrg., 146 (1997), pp. 173–196.
[42] V. Puzyrev and J. María Cela, *A review of block Krylov subspace methods for multisource electromagnetic modelling*, Geophys. J. Int., 202 (2015), pp. 1241–1252.
[43] C. Qiu, S. Güttel, X. Ren, C. Yin, Y. Liu, B. Zhang, and G. Egbert, *A block rational Krylov method for three-dimensional time-domain marine controlled-source electromagnetic modeling*, MIMS EPrint 2018.22, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2018.
[44] A. Ruhe, *Rational Krylov sequence methods for eigenvalue computation*, Linear Algebra Appl., 58 (1984), pp. 391–405.
[45] ———, *The rational Krylov algorithm for nonsymmetric eigenvalue problems. III. Complex shifts for real matrices*, BIT, 34 (1994), pp. 165–176.
[46] ———, *Rational Krylov algorithms for nonsymmetric eigenvalue problems. II. Matrix pairs*, Linear Algebra Appl., 197/198 (1994), pp. 283–295.
[47] ———, *Rational Krylov: a practical algorithm for large sparse nonsymmetric matrix pencils*, SIAM J. Sci. Comput., 19 (1998), pp. 1535–1551.
[48] T. Schmelzer, *Block Krylov methods for Hermitian Linear Systems*, Master's thesis, University of Kaiserslautern, Germany, 2004.
[49] G. W. Stewart, *Matrix Algorithms. Vol. I Basic Decompositions*, SIAM, 1998.
[50] ———, *Matrix Algorithms. Vol. II Eigensystems*, SIAM, 2001.
[51] H. Voss, *Nonlinear eigenvalue problems*, in Handbook of Linear Algebra, L. Hogben, ed., Chapman and Hall/CRC, Boca Raton, FL, USA, second ed., 2014, pp. 115:1–115:24.