SIAM J. Sci. Comput.
Vol. 46, No. 2, pp. A798–A824

AN EFFICIENT BLOCK RATIONAL KRYLOV SOLVER FOR SYLVESTER EQUATIONS WITH ADAPTIVE POLE SELECTION*

A. CASULLI† AND L. ROBOL‡

Abstract. We present an algorithm for the solution of Sylvester equations with right-hand side of low rank. The method is based on projection onto a block rational Krylov subspace, with two key contributions with respect to the state of the art. First, we show how to maintain the last pole equal to infinity throughout the iteration, by means of pole reordering, which allows for a cheap evaluation of the true residual at every step. Second, we extend the convergence analysis in [B. Beckermann, *SIAM J. Numer. Anal.*, 49 (2011), pp. 2430–2450] to the block case. This extension allows us to link the convergence with the problem of minimizing the norm of a small rational matrix over the spectra or field-of-values of the involved matrices. This is in contrast with the nonblock case, where the minimum problem is scalar, instead of matrix valued. Replacing the norm of the objective function with a more easily evaluated function yields several adaptive pole selection strategies, providing a theoretical analysis for known heuristics, as well as effective novel techniques.

Key words. Sylvester equations, block rational Krylov, adaptive pole selection

MSC codes. 65F45, 15A24

DOI. 10.1137/23M1548463

See reproducibility of computational results at end of the article.

1. Introduction.

We are concerned with the solution of Sylvester equations of the form

$$AX - XB = \mathbf{u}\mathbf{v}^H, \qquad \mathbf{u} \in \mathbb{C}^{n\times b},\ \mathbf{v} \in \mathbb{C}^{m\times b}, \tag{1.1}$$

and A, B are square matrices of sizes $n \times n$ and $m \times m$, respectively. The matrices $\mathbf{u}, \mathbf{v}$ are block vectors, i.e., matrices with a few columns with $b \ll n, m$. If A and B have disjoint spectra, the solution is unique and can be expressed in the integral form

$$X = \frac{1}{2\pi i}\int_\gamma (zI_n - A)^{-1}\mathbf{u}\mathbf{v}^H(zI_m - B)^{-1}dz, \tag{1.2}$$

where γ is a compact contour that encloses once, in positive orientation, the eigenvalues of A, but not the eigenvalues of B [17].

Sylvester equations arise often in control theory [1, 5], and in the solution of 2 dimensional PDEs on tensorized domains [21, 27]. In this setting the matrices involved are often of large size, and exploiting the low-rank structure in the right-hand side is essential. For problems arising from control theory, the rank is linked with the number of inputs and outputs in the system, so b is typically moderate and related to the analysis of MIMO systems [1]. For PDEs, the low-rank property holds in an approximate sense and is related to the regularity of the problem under consideration.

*Submitted to the journal's Numerical Algorithms for Scientific Computing section January 19, 2023; accepted for publication (in revised form) October 27, 2023; published electronically March 6, 2024. The authors are members of the research group INdAM-GNCS.
https://doi.org/10.1137/23M1548463

†Scuola Normale Superiore, Pisa, P.za Cavalieri, 7, 56126, Pisa, Italy (angelo.casulli@sns.it).
‡Department of Mathematics, University of Pisa, L.go B. Pontecorvo, 5, 56127, Pisa, Italy (leonardo.robol@unipi.it).

When the spectra of A and B are well-separated, one can show that the matrix X that solves (1.1) has exponentially decaying singular values [4], and can be approximated as a low-rank matrix [25]. If X is close to a low-rank matrix, i.e., we can write it as $X = UYV^H + E$, where U, V are matrices with a few orthogonal columns, and E is a small error, then the Sylvester equation can be approximately solved by computing the exact solution of the projected equation $(U^H AU)Y - Y(V^H BV) = U^H \mathbf{u}\mathbf{v}^H V$. This is the core idea of projection methods. The main difficulty is identifying good bases U, V to use for projecting the equation.

A common choice is to take U and V as an orthonormal basis of Krylov or rational Krylov subspaces. When $\mathbf{u}, \mathbf{v}$ are vectors, these subspaces contain a basis for $f(A)\mathbf{u}$ or $f(B^H)\mathbf{v}$, where $f(z)$ is a low degree polynomial or rational function with assigned poles. Increasing the degree produces a sequence of subspaces, one contained in the other, and therefore a sequence of approximations. The characterization through polynomials and rational functions allow us to link the convergence of the method with a polynomial (resp., rational) approximation problem, which allows us to state explicit results (at least in the case of normal matrix coefficients) [2, 3]. The rational methods are inherently more complex to analyze because a choice of poles is involved, and the convergence is dependent on the quality of these poles.

When $\mathbf{u}$ and $\mathbf{v}$ are block vectors an analogous construction can be made, by building a basis for the column spans of $f(A)\mathbf{u}$ or $f(B^H)\mathbf{v}$. The results in the literature focus mostly on the nonblock case, and are more scarce for this setting. One of the contributions of this work is to extend the convergence analysis for rational Krylov found in [2] to this more general setting. This is done by exploiting the notation for the characteristic matrix polynomial used in [19] to analyze various block polynomial Krylov methods.

If $X = UYV^H$ with U, V bases of a Krylov subspace of order ℓ, then the residual $AX - XB - C$ belongs to the Krylov subspace of order $\ell + 1$ [25]. This property can be exploited to compute the residual error almost for free at each step. For rational Krylov subspaces, the analogous result tells us that the residual belongs to a larger subspace obtained by adding an infinity pole. However, if infinity poles are periodically injected in the space, we may incur an artificial inflation of the size of the projected problem. In this work, we show how one can exploit the theory of block rational Arnoldi decomposition (BRAD) from [11] and the reordering of the poles in the subspaces to maintain a single infinity pole in the definition of the rational block Krylov subspace, precisely with the aim of checking the residual.

Then, the convergence analysis introduced by extending the results in [2] is used to design an adaptive-pole-selection algorithm. Since the objective function is now matrix valued, instead of scalar, the problem is much richer. In particular, the minimization of its norm is numerically challenging, and it is natural to replace the objective function with a simpler surrogate. We present various options, and we show that one of these leads to the same heuristic proposed by Druskin and Simoncini in [10] generalizing the rank 1 case. Hence, our theory provides a theoretical analysis to the convergence of this choice. Then, we show that other choices for the surrogate function are possible; in particular, we provide an adaptive technique of pole selection that slightly improves the one proposed in [10].

The paper is structured as follows. In section 2 we introduce the notation used in the paper, and then in section 3 we discuss the tools needed from the theory of matrix polynomials and rational functions. Section 4 is devoted to the introduction of rational block Krylov subspaces and the related theory, and section 5 presents the algorithm based on projection on these subspaces for the solution of Sylvester equations. Section 6 discusses the convergence and the adaptive pole selection. Finally, we present some numerical tests in section 7.

2. Notation.

Given a matrix A we denote by $\Lambda(A)$ its spectrum, by $\mathbb{W}(A)$ its field of values, and by $\sigma(A)$ the set of its singular values. We use $\bar{A}$ and A^H to denote the conjugate and the conjugate transpose of A, respectively. For any polynomial $Q(z)$ we use $\bar{Q}(z)$ to denote the polynomial that has as coefficients the conjugate of the coefficients of $Q(z)$. The identity matrix of size s is denoted by I_s. We use bold letters to indicate block vectors, that is, tall and skinny matrices. The size of blocks is denoted by b. The Frobenius norm and the two norm are denoted by $\|\cdot\|_F$ and $\|\cdot\|_2$, respectively. We employ a MATLAB-like notation for submatrices; for instance, given $A \in \mathbb{C}^{m\times n}$ the matrix $A_{i_1:i_2,j_1:j_2}$ is the submatrix obtained selecting only rows from i_1 to i_2 and columns from j_1 to j_2 (extrema included). To simplify the notation we use bold letters also to denote block indices, that is, we use $\mathbf{s}$ to denote the set of indices $b(s-1)+1:bs$. We use the symbols $\otimes$ and $\oplus$ to denote the Kronecker product and the Kronecker sum, respectively, and the symbol vec to denote the operator that transforms a matrix into a vector obtained by stacking the columns of the matrix on top of one another. We denote by $\mathbf{e}_i$ the block vector defined as $e_i \otimes I_b$, where e_i is the ith element of the canonical basis.

3. Matrix polynomials and rational functions.

In this section, we provide some definitions and properties about matrix polynomials that we use in the paper. Matrix polynomials can be equivalently interpreted as polynomials with a scalar variable and matrix coefficients or as a matrix with polynomial entries. Both interpretations can be useful for proving different results. Formally, we will denote by $\mathbb{P}(\mathbb{C}^{b\times b})$ the space of $b\times b$ matrix polynomials, with coefficients in $\mathbb{C}^{b\times b}$. We use the notation $\mathbb{P}_d(\mathbb{C}^{b\times b})$ to denote the set of matrix polynomials of degree less than or equal to d. A matrix polynomial is said to be monic if its leading coefficient is equal to the identity.

We will use the notation $P(z) = \sum_{i=0}^{d} z^i\Gamma_i$ to indicate a generic matrix polynomial of degree less than d with matrix coefficients $\Gamma_i \in \mathbb{C}^{b\times b}$. In order to analyze (block) Krylov methods, we associate a matrix polynomial with a linear operator that acts on block vectors. More precisely, we define an operator $\circ$ as a function from $\mathbb{C}^{n\times n}\times\mathbb{C}^{n\times b}$ to $\mathbb{C}^{n\times b}$ as follows: given two matrices $A \in \mathbb{C}^{n\times n}$ and $\mathbf{v} \in \mathbb{C}^{n\times b}$, we set

$$P(A)\circ \mathbf{v} := \sum_{i=0}^{d} A^i \mathbf{v}\Gamma_i.$$

This notation has already been used in [16, 23, 26], and has been exploited in [19] for the analysis of block Krylov subspaces. If the matrix A is fixed, the map $\mathbf{v}\mapsto P(A)\circ\mathbf{v}$ is a function from $\mathbb{C}^{n\times b}$ to $\mathbb{C}^{n\times b}$. When dealing with rational Krylov method, it will often be useful to apply the inverse of the operator, that is, given a generic vector $\mathbf{v}$, finding another block vector $\mathbf{w}$ such that $P(A)\circ\mathbf{w} = \mathbf{v}$. Since the operator is linear in $\mathbf{w}$, this is equivalent to solving a linear system. A formal definition can be given as follows.

Definition 3.1. *Given a matrix $A \in \mathbb{C}^{n\times n}$, a block vector $\mathbf{v} \in \mathbb{C}^{n\times b}$, and a matrix polynomial $P(z) = \sum_{i=0}^{d} z^i\Gamma_i \in \mathbb{P}(\mathbb{C}^{b\times b})$, such that $\det(P(\lambda)) \neq 0$ for each λ eigenvalue of A, we define $P(A)\circ^{-1}\mathbf{v}$ as the block vector $\mathbf{w} \in \mathbb{C}^{n\times b}$, such that $P(A)\circ\mathbf{w} = \mathbf{v}$.*

Since $\mathbf{w}$ is implicitly defined as the solution of a linear system, we shall check that the system is invertible to ensure that the definition is well-posed.

LEMMA 3.2. *Given a matrix* $A \in \mathbb{C}^{n\times n}$, *a block vector* $\mathbf{v} \in \mathbb{C}^{n\times b}$, *and a matrix polynomial* $P(\lambda)$ *as above such that* $\det(P(\lambda)) \neq 0$ *for* $\lambda \in \Lambda(A)$, *there is a unique* $\mathbf{w} \in \mathbb{C}^{n\times b}$ *verifying* $P(A) \circ \mathbf{w} = \mathbf{v}$.

Proof. The relation $P(A) \circ \mathbf{w} = \mathbf{v}$ can be rewritten as $\mathrm{vec}(P(A) \circ \mathbf{w}) = \mathrm{vec}(\mathbf{v})$; in addition, we note that

$$\mathrm{vec}(P(A) \circ \mathbf{w}) = \left(\sum_{i=0}^{d} \Gamma_i^T \otimes A^i\right) \mathrm{vec}(\mathbf{w}),$$

where $\otimes$ denotes the Kronecker product, and we used the standard Kronecker relation $\mathrm{vec}(AXB) = (B^T \otimes A)\mathrm{vec}(X)$. We now prove that the matrix $\sum_{i=0}^{d} \Gamma_i^T \otimes A^i$ is invertible, which implies the sought claim, since $\mathbf{w}$ can be defined as

$$\mathbf{w} = \mathrm{vec}^{-1}\left(\left(\sum_{i=0}^{d} \Gamma_i^T \otimes A^i\right)^{-1} \mathrm{vec}(\mathbf{v})\right).$$

Let $A = UTU^H$ be a Schur decomposition of A, with T upper triangular; then

$$\sum_{i=0}^{d} \Gamma_i^T \otimes A^i = (I_b \otimes U)\left(\sum_{i=0}^{d} \Gamma_i^T \otimes T^i\right)\left(I_b \otimes U^H\right).$$

There exists a permutation matrix $P \in \mathbb{C}^{nb\times nb}$ (the "perfect shuffle;" see [13]), such that

$$\sum_{i=0}^{d} \Gamma_i^T \otimes T^i = P\left(\sum_{i=0}^{d} T^i \otimes \Gamma_i^T\right) P^H.$$

Hence, it is sufficient to prove the invertibility of $\sum_{i=0}^{d} T^i \otimes \Gamma_i^T$, that is, a block triangular matrix with block diagonal matrices given by $P(\lambda_1)^T, \ldots, P(\lambda_n)^T$, where λ_i are the eigenvalues of A. Therefore, the assumption $\det(P(\lambda)) \neq 0$ for each λ eigenvalue of A yields the claim. ◻

Remark 3.3. The proof of well-posedness of Definition 3.1 also gives us an explicit representation of $P(A)\circ^{-1}$: for any $\mathbf{v} \in \mathbb{C}^{n\times b}$

$$P(A) \circ^{-1} \mathbf{v} = \mathrm{vec}^{-1}\left(\left(\sum_{i=0}^{d} \Gamma_i^T \otimes A^i\right)^{-1} \mathrm{vec}(\mathbf{v})\right).$$

In particular, the hypothesis $\det(P(\lambda)) \neq 0$ for $\lambda \in \Lambda(A)$ is necessary to guarantee the invertibiliy of $\sum_{i=0}^{d} \Gamma_i^T \otimes A^i$.

The previous definitions and results essentially deal with matrix polynomials; for rational Krylov methods, we will need a way to incorporate rational functions into the picture. In practice, it will be sufficient to consider objects of the form $Q(\lambda)^{-1}P(\lambda)$, where $Q(\lambda)$ is a scalar polynomial and $P(\lambda)$ a matrix polynomial. It is immediate to check that any rational matrix (i.e., a matrix with rational entries) can always be written in this form.

The following remark suggests a way to extend the operators $\circ$ and $\circ^{-1}$ to rational matrix polynomials with scalar denominator.

LEMMA 3.4. *Let* $P(z) = \sum_{i=0}^{d} \Gamma_i z^i \in \mathbb{P}_d(\mathbb{C}^{b\times b})$ *and let* $Q(z) \in \mathbb{P}_k(\mathbb{C})$ *be a scalar polynomial. Denoting* $\tilde{P}(z) = Q(z)P(z) = \sum_{i=0}^{d+k} \Delta_i z^i$, *it holds that*

$$Q(A) \cdot (P(A) \circ \mathbf{v}) = \tilde{P}(A) \circ \mathbf{v} \quad \text{and} \quad Q(A)^{-1} \cdot (P(A) \circ^{-1} \mathbf{v}) = \tilde{P}(A) \circ^{-1} \mathbf{v},$$

where in the second equality we assume $\det(\tilde{P}(\lambda)) \neq 0$ *for each* $\lambda \in \Lambda(A)$.

Proof. To derive the first equality it is sufficient to prove the case of $Q(z) = z - \alpha$ for $\alpha \in \mathbb{C}$, since we can factor $Q(z)$ as the product of linear terms. By the definition of $\tilde{P}(z)$,

$$\tilde{P}(z) = (z - \alpha)P(z) = \sum_{i=0}^{d+1} (\Gamma_{i-i} - \alpha\Gamma_i)z^i$$

with the convention that $\Gamma_{-1} = \Gamma_{d+1} = 0$. In particular $\Delta_i = \Gamma_{i-1} - \alpha\Gamma_i$. Hence,

$$\begin{aligned}\tilde{P}(A) \circ \mathbf{v} &= \sum_{i=0}^{d+1} A^i \mathbf{v}\Delta_i = \sum_{i=0}^{d} A^{i+1}\mathbf{v}\Gamma_i - \alpha \sum_{i=0}^{d} A^i \mathbf{v}\Gamma_i \\ &= A \cdot P(A) \circ \mathbf{v} - \alpha P(A) \circ \mathbf{v} = (A - \alpha I_n) \cdot (P(A) \circ \mathbf{v}) = Q(A) \cdot (P(A) \circ \mathbf{v}).\end{aligned}$$

For the second identity it is sufficient to prove that $\tilde{P}(A) \circ (Q(A)^{-1}\mathbf{w}) = \mathbf{v}$, where $\mathbf{w} = P(A) \circ^{-1} \mathbf{v}$. Using the first identity,

$$\tilde{P}(A) \circ (Q(A)^{-1}\mathbf{w}) = Q(A) \cdot (P(A) \circ (Q(A)^{-1}\mathbf{w})) = Q(A) \sum_{i=0}^{d} A^i Q(A)^{-1}\mathbf{w}\Gamma_i.$$

Since $Q(A)$ commutes with the powers of A, this can be reduced to

$$\tilde{P}(A) \circ (Q(A)^{-1}\mathbf{w}) = P(A) \circ \mathbf{w}.$$

By the definition of $\mathbf{w}$ it follows that $P(A) \circ \mathbf{w} = \mathbf{v}$, which concludes the proof. ∎

In view of the previous result, we can extend the action of a matrix polynomial $P(A) \circ \mathbf{v}$ to the case of rational matrices with prescribed poles.

DEFINITION 3.5. *Let $Q(z) \in \mathbb{P}(\mathbb{C})$ and let $R(z) \in \mathbb{P}(\mathbb{C}^{b\times b})/Q(z)$, that is, there exists $P(z) \in \mathbb{P}(\mathbb{C}^{b\times b})$ such that $R(z) = P(z)/Q(z)$. Given $A \in \mathbb{C}^{n\times n}$ such that $Q(A)$ is invertible and $\mathbf{v} \in \mathbb{C}^{n\times b}$, we define*

$$R(A) \circ \mathbf{v} = Q(A)^{-1} (P(A) \circ \mathbf{v}) \quad \text{and} \quad R(A) \circ^{-1} \mathbf{v} = Q(A) \left(P(A) \circ^{-1} \mathbf{v}\right).$$

The expression of a rational matrix in the form $R(z) = P(z)/Q(z)$ is not unique. However, the previous definition does not depend on the representation; indeed if $R(z) = P(z)/Q(z) = \tilde{P}(z)/\tilde{Q}(z)$, then $Q(z)\tilde{P}(z) = \tilde{Q}(z)P(z)$; hence by Lemma 3.4,

$$Q(A) \cdot (\tilde{P}(A) \circ \mathbf{v}) = \tilde{Q}(A) \cdot (P(A) \circ \mathbf{v}) \tag{3.1}$$

and

$$Q(A)^{-1} \cdot (\tilde{P}(A) \circ^{-1} \mathbf{v}) = \tilde{Q}(A)^{-1} \cdot (P(A) \circ^{-1} \mathbf{v}). \tag{3.2}$$

Multiplying both sides of (3.1) on the left by $Q(A)^{-1} \cdot \tilde{Q}(A)^{-1}$ we obtain the well-posedness of the map $\mathbf{v} \mapsto R(A) \circ \mathbf{v}$, and multiplying both sides of (3.2) on the left by $Q(A)\tilde{Q}(A)$ we have the well-posedness of the map $\mathbf{v} \mapsto R(A) \circ^{-1} \mathbf{v}$.

Remark 3.6. If the matrix A is fixed, both operators

$$R(A)\circ : \mathbb{C}^{n\times d} \to \mathbb{C}^{n\times d} \quad \text{and} \quad R(A)\circ^{-1} : \mathbb{C}^{n\times d} \to \mathbb{C}^{n\times d}$$

are linear. As in the polynomial case, the latter is only defined if $R(z)$ is nonsingular over all the eigenvalues of A.

LEMMA 3.7. *If $A, B \in \mathbb{C}^{n\times n}$ commute, then for every rational matrix $R(z) = P(z)/Q(z)$, where $P(z) \in \mathbb{P}(\mathbb{C}^{b\times b})$ and $Q(z) \in \mathbb{P}(\mathbb{C})$,*

$$B \cdot R(A) \circ \mathbf{v} = R(A) \circ (B\mathbf{v});$$

moreover, if $det(P(\lambda)) \neq 0$ for each $\lambda \in \Lambda(A)$,

$$B \cdot R(A) \circ^{-1} \mathbf{v} = R(A) \circ^{-1} (B\mathbf{v}).$$

Proof. Let $P(z) = \sum_{i=1}^{d} z^i \Gamma_i \in \mathbb{P}_d(\mathbb{C}^{b\times b})$ and $Q(z) \in \mathbb{P}(\mathbb{C})$, such that $R(z) = P(z)/Q(z)$. Then

$$B \cdot R(A) \circ \mathbf{v} = BQ(A)^{-1} \sum_{i=1}^{d} A^i \mathbf{v}\Gamma_i = Q(A)^{-1} \sum_{i=1}^{d} A^i B\mathbf{v}\Gamma_i = R(A) \circ (B\mathbf{v}),$$

and therefore

$$\begin{aligned}
&\mathrm{vec}(B \cdot R(A) \circ^{-1} \mathbf{v}) \\
&\quad = (I_b \otimes B)(I_b \otimes Q(A)) \left(\sum_{i=1}^{d} \Gamma_i^T \otimes A^i\right)^{-1} \mathrm{vec}(\mathbf{v}) \\
&\quad = (I_b \otimes Q(A)) \left(\sum_{i=1}^{d} \Gamma_i^T \otimes A^i\right)^{-1} (I_b \otimes B)\mathrm{vec}(\mathbf{v}) = \mathrm{vec}(R(A) \circ^{-1} (B\mathbf{v})). \qquad \square
\end{aligned}$$

Given a matrix polynomial $P(z) = \sum_{i=0}^{d} z^i \Gamma_i$, we denote by $P^H(z)$ the matrix polynomial $P^H(z) := \sum_{i=0}^{d} z^i \Gamma_i^H$. Similarly, we denote by $\bar{P}(z)$ the matrix polynomial with complex conjugate (but not transposed) coefficients. Given a function $R(z) = P(z)/Q(z)$, we denote by $\bar{R}(z)$ and $R^H(z)$ the rational functions $\bar{P}(z)/\bar{Q}(z)$ and $P^H(z)/\bar{Q}(z)$, respectively.

LEMMA 3.8. *Given $\mathbf{v} \in \mathbb{C}^{n\times b}$ and $\mathbf{w} \in \mathbb{C}^{m\times b}$, the following identities hold:*

$$R(zI_n) \circ^{-1} \mathbf{v} = \mathbf{v}(R(z))^{-1} \quad \text{and} \quad R(zI_n) \circ^{-1} \mathbf{v}\mathbf{w}^H = \mathbf{v}(R^H(\bar{z}I_m) \circ^{-1} \mathbf{w})^H.$$

Proof. Let $P(z) = \sum_{i=1}^{d} z^i \Gamma_i \in \mathbb{P}_d(\mathbb{C}^{b\times b})$ and $Q(z) \in \mathbb{P}(\mathbb{C})$, such that $R(z) = P(z)/Q(z)$. It holds that

$$\begin{aligned}
\mathrm{vec}\left(R(zI_n) \circ^{-1} \mathbf{v}\right) &= Q(z) \left(\sum_{i=0}^{d} \Gamma_i^T \otimes z^i I_n\right)^{-1} \mathrm{vec}(\mathbf{v}) \\
&= \left(\left(R^T(z)\right)^{-1} \otimes I_n\right) \mathrm{vec}(\mathbf{v}) = \mathrm{vec}\left(\mathbf{v}(R(z))^{-1}\right)
\end{aligned}$$

from which follows the first equality. For the second identity notice that

$$R(zI_n) \circ^{-1} \mathbf{v}\mathbf{w}^H = \mathbf{v}(R(z))^{-1}\mathbf{w}^H = \mathbf{v}(\mathbf{w}(R^H(\bar{z}))^{-1})^H = \mathbf{v}(R^H(\bar{z}I_m) \circ^{-1} \mathbf{w})^H. \qquad \square$$

The following theorem is a generalization of the Cauchy integral formula to the action of rational matrices.

THEOREM 3.9. *Let $A \in \mathbb{C}^{n\times n}$, $\mathbf{v} \in \mathbb{C}^{n\times b}$, and let γ be a compact contour that encloses once the eigenvalues of A with positive orientation. Then, for any $R(z) \in \mathbb{P}(\mathbb{C}^{b\times b})/Q(z)$, such that $\det(R(z)) \neq 0$ for each z in the compact set enclosed by γ, it holds that*

$$\frac{1}{2\pi i} \int_\gamma R(zI_n) \circ^{-1} \left[(zI_n - A)^{-1}\mathbf{v}\right] dz = R(A) \circ^{-1} \mathbf{v}.$$

Proof. Let $P(z) = \sum_{i=1}^{d} z^i \Gamma_i$ be such that $R(z) = P(z)/Q(z)$. Then

$$\begin{aligned}
&\operatorname{vec}\left(\int_\gamma R(zI_n) \circ^{-1} \left[(zI_n - A)^{-1}\mathbf{v}\right] dz\right) \\
&= \left(\int_\gamma Q(z) \left(\sum_{i=0}^{d} \Gamma_i^T \otimes z^i I_n\right)^{-1} \cdot \left(I_n \otimes (zI_n - A)^{-1}\right) dz\right) \operatorname{vec}(\mathbf{v}) \\
&= \left(\int_\gamma Q(z) \left(\sum_{i=0}^{d} \Gamma_i^T z^i\right)^{-1} \otimes (zI_n - A)^{-1}\, dz\right) \operatorname{vec}(\mathbf{v}).
\end{aligned}$$

For each $s,t \in \{1,\ldots,b\}$, let $f_{s,t}(z)$ be the function that maps z in the entry in position (s,t) of $Q(z)\left(\sum_{i=0}^{d} \Gamma_i^T z^i\right)^{-1}$. Since for each z inside the compact set bounded by γ it holds that $\det(R(z)) \neq 0$, the functions $f_{s,t}(z)$ are holomorphic on such a set. Then for the Cauchy integral formula, we have

$$\frac{1}{2\pi i}\int_\gamma Q(z) \left(\sum_{i=0}^{d} \Gamma_i^T z^i\right)^{-1}_{s,t} \otimes (zI_n - A)^{-1}\, dz = \frac{1}{2\pi i}\int_\gamma f_{s,t}(z) \cdot (zI_n - A)^{-1}\, dz = f_{s,t}(A).$$

Then, if we denote by $F \in \mathbb{C}^{nb\times nb}$ the block matrix for which the block in position (s,t) is defined by $f_{s,t}(A)$, we have the equivalence

$$F = \frac{1}{2\pi i}\int_\gamma Q(z) \left(\sum_{i=0}^{d} \Gamma_i^T z^i\right)^{-1} \otimes (zI_n - A)^{-1}\, dz.$$

We now claim that $F = (I_b \otimes Q(A))\left(\sum_{i=0}^{d} \Gamma_i^T \otimes A^i\right)^{-1}$, which implies the sought results, since

$$\begin{aligned}
&\operatorname{vec}\left(\frac{1}{2\pi i}\int_\gamma R(zI_n) \circ^{-1} (zI_n - A)^{-1}\mathbf{v}\, dz\right) = F \cdot \operatorname{vec}(\mathbf{v}) \\
&\quad = (I_b \otimes Q(A)) \left(\sum_{i=0}^{d} \Gamma_i^T \otimes A^i\right)^{-1} \operatorname{vec}(\mathbf{v}) = \operatorname{vec}\left(R(A) \circ^{-1} \mathbf{v}\right).
\end{aligned}$$

Hence in the following we prove that $\left(\sum_{i=0}^{d} \Gamma_i^T \otimes A^i\right) \cdot F = I_b \otimes Q(A)$.

For any $s,t \in \{1,\ldots,b\}$, let us define $g_{s,t}(z) = \left(\sum_{i=0}^{d} \Gamma_i^T z^i\right)_{s,t}$. Since

$$\left(\sum_{i=0}^{d} \Gamma_i^T z^i\right) \cdot \left[Q(z) \left(\sum_{i=0}^{d} \Gamma_i^T z^i\right)^{-1}\right] = Q(z) I_b,$$

it holds that

$$Q(z)\delta_{s,t} = \sum_{r=1}^{b} g_{s,r}(z) f_{r,t}(z), \tag{3.3}$$

where $\delta_{s,t}$ denotes the Kronecker delta.

To simplify the notation, for any integer $r \in \{1, \dots, b\}$, we define $ix(r)$ as the set of indices $n(r-1)+1 : nr$. For any $s, t \in \{1, \dots, b\}$ we have

$$\left(\left(\sum_{i=0}^{d} \Gamma_i^T \otimes A^i \right) \cdot F \right)_{ix(s),ix(t)} = \sum_{r=1}^{b} \left(\sum_{i=0}^{d} (\Gamma_i^T)_{s,r} \cdot A^i \right) f(A)_{r,t}$$

$$= \sum_{r=1}^{b} g_{s,r}(A) f_{r,t}(A) = \delta_{s,t} Q(A),$$

where the last equality follows from (3.3). □

Let us now recall the concept of divisibility for matrix polynomials and the definition of block characteristic polynomial. We use the term *regular* to identify matrix polynomials whose determinant is not identically zero over $\mathbb{C}$. The following results, including proofs of theorems, can be found in [19, section 2.5] or in the more classical reference [12, section 7.7].

The results extend the familiar concept of Euclidean division to matrix polynomials. Matrix polynomials form a noncommutative ring, so we need to differentiate between left and right divisors. However, the underlying idea of dividing $P(z)$ by $D(z)$ is still the same: we want to write $P(z)$ as a multiple of $D(z)$ plus an additional remainder term, which should be of lower degree than $D(z)$.

DEFINITION 3.10. *Let $P(z), K(z), R(z)$, and $D(z)$ be matrix polynomials, where $P(z)$ has degree d, $D(z)$ is regular with degree less than d, and $R(z)$ has degree less than* $\deg D(z)$. *$K(z)$ is defined as "left quotient" and $R(z)$ as the "left remainder" of $P(z)$ divided by $D(z)$ if*

$$P(z) = D(z)K(z) + R(z).$$

If $R(z) = 0$, we say that $P(z)$ is left divisible by $D(z)$.

A natural question arises: given $P(z)$ and a lower degree polynomial $D(z)$, can we easily check if $D(z)$ divides $P(z)$ (i.e., if the remainder of the left or right division is zero)?

For a scalar polynomial $p(\lambda)$ and a linear divisor $\lambda - s$, this amounts to checking if $p(s) = 0$. A similar result holds for matrix polynomials as well.

THEOREM 3.11 (see [19, Theorem 2.17]). *The matrix polynomial $P(z) \in \mathbb{P}(\mathbb{C}^{b \times b})$ is left divisible by $zI_b - S$, where $S \in \mathbb{C}^{b \times b}$ if and only if $P(S) = 0$.*

DEFINITION 3.12. *Let $P(z)$ be a matrix polynomial. A matrix $S \in \mathbb{C}^{b \times b}$ is called a left solvent of $P(z)$ if $P(S) = 0$.*

In the following, we omit "left" when referring to quotients, divisibility, and solvents.

We remark that solvents are important tools in the analysis of matrix polynomials. They can be used to compute a part of the spectrum [18], and are closely related to the solution of the one-sided matrix equation that arises, for instance, in some Markov chains (see [7] and the references therein).

We now present a possible way to construct a block characteristic polynomial. In the scalar case, we may think of building the characteristic polynomial of a matrix A by computing its eigenvalues $s_1, \dots, s_n$, and then taking the product of the linear factors $p(\lambda) = (\lambda - s_1) \cdots (\lambda - s_n)$. The next theorem presents the extension of this idea to the block case, where the eigenvalues are replaced by blocks in a block diagonal matrix similar to the original one, and solvents play the role of the roots.

DEFINITION 3.13. *Let $A \in C^{db\times db}$ and $\mathbf{v} \in \mathbb{C}^{db\times b}$. A block characteristic polynomial of A with respect to $\mathbf{v}$ is a matrix polynomial $P(z) \in \mathbb{P}_d(\mathbb{C}^{b\times b})$ such that*

$$P(A) \circ \mathbf{v} = 0.$$

THEOREM 3.14 (see [19, Theorem 2.24]). *Let $A \in C^{db\times db}$ and $\mathbf{v} \in \mathbb{C}^{db\times b}$. Let $P(z)$ be a monic block characteristic polynomial of A with respect to $\mathbf{v}$. Assuming that there exists a block diagonal matrix*

$$T = \begin{bmatrix} \Theta_1 & & \\ & \ddots & \\ & & \Theta_d \end{bmatrix}$$

with $\{\Theta_i\}_{i=1:d} \subseteq \mathbb{C}^{b\times b}$ and an invertible matrix $\mathcal{U} \in \mathbb{C}^{db\times db}$ such that

$$A = \mathcal{U} T \mathcal{U}^{-1},$$

and letting $W = [W_1, \ldots, W_d]^T = \mathcal{U}^{-1}\mathbf{v}$, with $\{W_i\}_{i=1}^d \subseteq \mathbb{C}^{b\times b}$, then if W_i is invertible for each i, it holds that

1. *$S_i = W_i^{-1}\Theta_i W_i$ are solvents of $P(z)$;*
2. *if $S_i - S_j$ is nonsingular for each $i \neq j$ then*

$$P(z) = (zI_b - S_1) \cdot \cdots \cdot (zI_b - S_d).$$

4. Block rational Krylov methods.

Given a matrix $A \in \mathbb{C}^{n\times n}$, a block vector $\mathbf{v} \in \mathbb{C}^{n\times b}$, and a sequence of poles $\boldsymbol{\xi}_k = \{\xi_j\}_{j=0}^{k-1} \subseteq \mathbb{C} \cup \{\infty\} \setminus \Lambda(A)$, the kth block rational Krylov space is defined as

$$\mathcal{Q}_k(A, \mathbf{v}, \boldsymbol{\xi}_k) = \left\{ R(A) \circ \mathbf{v} : R(z) = \frac{P(z)}{Q_k(z)} \text{ with } P(z) \in \mathbb{P}_{k-1}(\mathbb{C}^{b\times b}) \right\},$$

where $Q_k(z) = \prod_{\xi_j \in \boldsymbol{\xi}_k, \xi_j \neq \infty}(z - \xi_j)$. For simplicity, we sometimes denote such a space by $\mathcal{Q}_k(A, \mathbf{v})$ omitting poles. Note that when choosing all poles equal to ∞ we recover the classical definition of block rational Krylov subspaces.

It can be proved that $\mathcal{Q}_k(A, \mathbf{v}) \subseteq \mathcal{Q}_{k+1}(A, \mathbf{v})$. In this work, we will assume that the block rational Krylov subspaces are always strictly nested, that is, $\mathcal{Q}_k(A, \mathbf{v}) \subsetneq \mathcal{Q}_{k+1}(A, \mathbf{v})$ and that the dimension of $\mathcal{Q}_k(A, \mathbf{v})$ is equal to kb.

An orthonormal block basis of $\mathcal{Q}_k(A, \mathbf{v})$ (for simplicity, we will often just say "orthonormal basis") is defined as a matrix $V_k = [\mathbf{v}_1, \ldots, \mathbf{v}_k] \in \mathbb{C}^{n\times bk}$ with orthonormal columns, such that every block vector $\mathbf{v} \in \mathcal{Q}_k(A, b)$ can be written as $\mathbf{v} = \sum_{i=1}^k \mathbf{v}_i \Gamma_i$, for $\Gamma_i \in \mathbb{C}^{b\times b}$.

Krylov methods require the computation of the block orthogonal basis and the corresponding projection of the matrix A. If an orthogonal basis V_{k+1} is known, then the projected matrix is given by $A_{k+1} = V_{k+1}^H A V_{k+1}$.

The matrix V_{k+1} can be computed by block rational Arnoldi Algorithm[1] 4.1, which iteratively computes the block columns of V_{k+1} and two matrices $\underline{K_k}, \underline{H_k} \in \mathbb{C}^{b(k+1)\times bk}$ in block upper Hessenberg form such that

$$A V_{k+1} \underline{K_k} = V_{k+1} \underline{H_k}. \tag{4.1}$$

Relation (4.1) completely determines the rational Krylov subspace, and encodes all the information regarding poles and column span of the starting block vector.

[1]For simplicity we describe a version of the algorithm that does not allow poles equal to zero. For a more complete version of the algorithm we refer the reader to [11].

Algorithm 4.1. Block rational Arnoldi.

Require: $A \in \mathbb{C}^{n\times n}, \mathbf{v} \in \mathbb{C}^{n\times b}, \boldsymbol{\xi}_{k+1} = \{\xi_0, \ldots, \xi_k\}$
Ensure: $V_{k+1} \in \mathbb{C}^{n\times b(k+1)}$, $\underline{H}_k, \underline{K}_k \in \mathbb{C}^{b(k+1)\times bk}$
  $\mathbf{w} \leftarrow (I - A/\xi_0)^{-1}\mathbf{v}$ ▷ with the convention $A/\infty = 0$
  $[\mathbf{v}_1, \sim] \leftarrow \mathrm{qr}(\mathbf{w})$ ▷ compute the thin QR decomposition
  for $j = 1, \ldots, k$ **do**
    Compute $\mathbf{w} = (I - A/\xi_j)A\mathbf{v}_j$
    for $i = 1, \ldots, j$ **do**
      $(\underline{H}_k)_{\mathbf{i},\mathbf{j}} \leftarrow \mathbf{v}_i^H \mathbf{w}$ ▷ where **i** and **j** are block indices
      $\mathbf{w} \leftarrow \mathbf{w} - \mathbf{v}_j(\underline{H}_k)_{\mathbf{i},\mathbf{j}}$
    end for
    $[\mathbf{v}_{j+1}, (\underline{H}_k)_{\mathbf{j+1},\mathbf{j}}] \leftarrow \mathrm{qr}(\mathbf{w})$ ▷ compute the thin QR decomposition
    $(\underline{K}_k)_{\mathbf{i},1:\mathbf{j+1}b} \leftarrow (\underline{H}_k)_{\mathbf{i},1:\mathbf{j+1}b}/\xi_j - \mathbf{e}_j,$ ▷ where $\mathbf{e}_j = [0, \ldots, 0, I_b, 0]^T$
  end for
  $V_k \leftarrow [\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}]$

DEFINITION 4.1 (see [11]). *Let $A \in \mathbb{C}^{n\times n}$. A relation of the form*

$$AV_{k+1}\underline{K_k} = V_{k+1}\underline{H_k}$$

is called an orthonormal BRAD if the following conditions are satisfied:
1. *$V_{k+1} \in \mathbb{C}^{n\times b(k+1)}$ has orthonormal columns;*
2. *$\underline{K_k}$ and $\underline{H_k}$ are $b(k+1)\times bk$ block upper Hessenberg matrices such that for each i either $(\underline{K}_k)_{\mathbf{i+1},\mathbf{i}}$ or $(\underline{H}_k)_{\mathbf{i+1},\mathbf{i}}$ (or both) are invertible;*
3. *for any i, there exist two scalars $\mu_i, \nu_i \in \mathbb{C}$, with at least one different from zero, such that $\mu_i(\underline{K}_k)_{\mathbf{i+1},\mathbf{i}} = \nu_i(\underline{H}_k)_{\mathbf{i+1},\mathbf{i}}$;*
4. *the numbers $\xi_i = \mu_i/\nu_i$ above, called poles of the BRAD, are outside the spectrum of A.*

Remark 4.2. The relation (4.1) produced by the block rational Arnoldi algorithm is a BRAD; see [11, section 2].

Remark 4.3. The matrices $\underline{H_k}$ and $\underline{K_k}$ of a BRAD are both full rank. This follows from [11, Lemma 3.2].

The following theorem relates rational Arnoldi decompositions with rational Krylov subspaces.

THEOREM 4.4. *Let $A \in \mathbb{C}^{n\times n}$, $\mathbf{v} \in \mathbb{C}^{n\times b}$, $\boldsymbol{\xi}_{k+1} = \{\xi_0, \ldots \xi_k\}$, and let $\mathcal{Q}_{k+1}(A, \mathbf{v})$ be the block rational Krylov subspace with poles $\boldsymbol{\xi}_{k+1}$. Let*

$$AV_{k+1}\underline{K_k} = V_{k+1}\underline{H_k}$$

be a BRAD with poles $\{\xi_1, \ldots \xi_k\}$, such that the first block column of V_{k+1} is an orthonormal basis of the space spanned by the columns of $(I - A/\xi_0)^{-1}\mathbf{v}$. Then V_{k+1} is an othonormal block basis of $\mathcal{Q}_{k+1}(A, \mathbf{v})$. Moreover, the matrix obtained by taking the first bj columns of V_{k+1} is an orthonormal block basis for $\mathcal{Q}_j(A, \mathbf{v})$ for each $j \le k+1$.

For the proof of the theorem and a more detailed description of BRADs we refer the reader to [11].

Let V_k be the matrix obtained by taking the first bk columns of V_{k+1}. The computation of the projected matrix $A_k = V_k^H AV_k$ by using the formula is usually expensive if the dimension of the matrix A is large. For the case of Hermitian A, several methods that exploit the structure of A_k have been developed to avoid expensive operations for the computation; see, for instance, [8, 20]. In the non-Hermitian case, it is more difficult to exploit a structure of A_k. However, if the last pole of the associated BRAD is equal to infinity the projected matrix can be easily computed as $A_k = H_k K_k^{-1}$, where K_k and H_k are the head $kb \times kb$ principal submatrices of $\underline{K_k}$ and $\underline{H_k}$, respectively. To prove this, notice that if the last pole is equal to infinity then the last block row of $\underline{K_k}$ has to be zero; then since $\underline{K_k}$ is full rank, K_k is invertible; hence multiplying both the terms of the BRAD (4.1) on the left by V_k^H and on the right by K_k^{-1} we obtain $A_k = H_k K_k^{-1}$.

A technique that is often used to compute A_k is to add a pole equal to infinity every time we want to compute a new projected matrix. However, this would significantly increase the size of the block rational Krylov subspace considered by Algorithm 4.1. In the next section, we describe a way to ensure that the last pole is always equal to infinity, avoiding these additional steps.

4.1. Reordering poles. We propose to start the Krylov method with $\xi_1 = \infty$, then after each step transform the BRAD into another one that has the last two poles swapped. Doing this procedure after each step of the block rational Krylov method the last pole is always equal to infinity.

This technique has been already described for the nonblock case in [15]. In the following, we introduce a practical way to swap the last two poles by using unitary transformations.

Let us consider a BRAD

$$A\hat{V}_{k+1}\underline{\hat{K}_k} = \hat{V}_{k+1}\underline{\hat{H}_k} \tag{4.2}$$

with poles $\{\xi_1, \ldots, \xi_{k-2}, \infty, \xi_k\}$. By Definition 4.1, since the second last pole is equal to infinity, the submatrix $(\underline{\hat{K}_k})_{\mathbf{k},\mathbf{k-1}}$ is equal to zero. Moreover, to produce a new BRAD that has the last pole equal to infinity it is sufficient to annihilate the submatrix $(\underline{\hat{K}_k})_{\mathbf{k+1},\mathbf{k}}$, keeping the block Hessenberg structure of the two matrices. This can be done by employing unitary transformations. Let

$$Q_1 R_1 = \begin{bmatrix} (\underline{\hat{K}_k})_{\mathbf{k},\mathbf{k}} \\ (\underline{\hat{K}_k})_{\mathbf{k+1},\mathbf{k}} \end{bmatrix}$$

be a thin QR decomposition and let $R_2 Q_2$ be an RQ decomposition[2] for the last block row of

$$Q_1^H \begin{bmatrix} (\underline{\hat{H}_k})_{\mathbf{k},\mathbf{k-1}} & (\underline{\hat{H}_k})_{\mathbf{k},\mathbf{k}} \\ 0 & (\underline{\hat{H}_k})_{\mathbf{k+1},\mathbf{k}} \end{bmatrix}.$$

Then, the matrices

$$Q_1^H \begin{bmatrix} 0 & (\underline{\hat{K}_k})_{\mathbf{k},\mathbf{k}} \\ 0 & (\underline{\hat{K}_k})_{\mathbf{k+1},\mathbf{k}} \end{bmatrix} Q_2^H \quad \text{and} \quad Q_1^H \begin{bmatrix} (\underline{\hat{H}_k})_{\mathbf{k},\mathbf{k-1}} & (\underline{\hat{H}_k})_{\mathbf{k},\mathbf{k}} \\ 0 & (\underline{\hat{H}_k})_{\mathbf{k+1},\mathbf{k}} \end{bmatrix} Q_2^H$$

are block upper triangular and the last block row of the first one is equal to zero.

[2]An RQ decomposition consists in writing a matrix as the product of an upper triangular matrix times a unitary matrix. It can be computed by using the same techniques involved in the computation of a QR decomposition.

If we let

$$\begin{aligned} V_{k+1} &= \hat{V}_{k+1}(I_{b(k-1)} \oplus Q_1), \\ \underline{K_k} &= (I_{b(k-1)} \oplus Q_1^H)\underline{\hat{K}_k}(I_{b(k-2)} \oplus Q_2^H), \\ \underline{H_k} &= (I_{b(k-1)} \oplus Q_1^H)\underline{\hat{H}_k}(I_{b(k-2)} \otimes Q_2^H), \end{aligned}$$

where $\oplus$ denotes the Kronecker sum, the relation

$$AV_{k+1}\underline{K_k} = V_{k+1}\underline{H_k}$$

is a new BRAD that has infinity as the last pole.

The computational cost of this procedure is $\mathcal{O}(kb^3)$, which is negligible with respect to the computational cost of a step of block rational Arnoldi Algorithm 4.1.

Remark 4.5. When we transform the matrix $\hat{V}_k$ in V_k we only perform a linear combination between the last two block columns. For this reason the top-left principal $b(k-1) \times b(k-1)$ submatrix of A_k is equal to A_{k-1}. Hence, to compute A_k it is sufficient to determine its last block row and column. This can be done using the relation $A_k = H_k K_k^{-1}$ and so

$$A_k \mathbf{e}_k = H_k K_k^{-1}\mathbf{e}_k \quad \text{and} \quad \mathbf{e}_k^T A_k = \mathbf{e}_k^T H_k K_k^{-1}.$$

5. Rational Krylov for Sylvester equation.

Krylov subspace methods are one of the most popular methods for solving the Sylvester equation (1.1) where A, B are large size matrices and $\mathbf{u}, \mathbf{v}$ are tall and skinny. In such a case, the solution can be approximated by a low-rank matrix to avoid storing the complete solution which is prohibitive for large n and m. We refer the reader to [25, section 4.4] for a more complete discussion of the topic.

The technique described in section 4.1 can be used for the resolution of Sylvester equations: let U_{h+1} and V_{k+1} be orthonormal block bases for $\mathcal{Q}_{h+1}(A, \mathbf{u})$ and $\mathcal{Q}_{k+1}(B^H, \mathbf{v})$, respectively, generated by the block rational Arnoldi Algorithm 4.1 and let $U_h \in C^{n\times bh}$ and $V_k \in \mathbb{C}^{m\times bk}$ be the matrices obtained removing from U_{h+1} and V_{k+1} the last b columns. Letting $A_h = U_h^H A U_h$ and $B_k = V_k^H B V_k$, the solution X can be approximated by $X_{h,k} = U_h \hat{X} V_k^H$, where $\hat{X}$ solves the projected equation

$$A_h\hat{X} - \hat{X}B_k = U_h^H\mathbf{u}(V_k^H\mathbf{v})^H. \tag{5.1}$$

For simplicity of notation in the rest of the section, we assume that $\xi_0 = \infty$, that is,

$$U_h^H\mathbf{u} = \|\mathbf{u}\|_2\mathbf{e}_1 \quad \text{and} \quad V_k^H\mathbf{v} = \|\mathbf{v}\|_2\mathbf{e}_1.$$

If U_{h+1} and V_{k+1} are determined as described in section 4.1, the projected matrices A_h and B_k can be easily computed at each step. In the following we show that this choice of poles also allows a cheap computation of the norm of the residual matrix

$$R_{h,k} = AX_{h,k} - X_{h,k}B - \mathbf{u}\mathbf{v}^H.$$

Since the last pole used to generate $\mathcal{Q}_{h+1}(A, u)$ is always equal to infinity, the columns of AU_h belong to $\mathcal{Q}_{h+1}(A, u)$, that is,

$$U_{h+1}U_{h+1}^H AU_h = AU_h.$$

In the same way it holds that

$$V_k^H B V_{k+1} V_{k+1}^H = V_k^H B.$$

Using the last two relations, the definition of $X_{h,k}$, and that the first block columns of U_{h+1} and V_{k+1} are given by the orthonormalization of $\mathbf{u}$ and $\mathbf{v}$, respectively, we can rewrite the residual as

$$\begin{aligned}
R_{h,k} &= U_{h+1} U_{h+1}^H A U_h \hat{X} V_k^H - U_h \hat{X} V_k^H B V_{k+1} V_{k+1}^H - U_{h+1} \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \mathbf{e}_1 \mathbf{e}_1^T V_{k+1}^H \\
&= U_{h+1} \left(U_{h+1}^H A U_h \hat{X} \begin{bmatrix} I_{bh} & 0 \end{bmatrix} - \begin{bmatrix} I_{bk} \\ 0 \end{bmatrix} \hat{X} V_k^H B V_{k+1} - \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \mathbf{e}_1 \mathbf{e}_1^T \right) V_{k+1}^H \\
&= U_{h+1} \begin{bmatrix} U_h^H A U_h \hat{X} - \hat{X} V_k^H B V_k - \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \mathbf{e}_1 \mathbf{e}_1^T & -\hat{X} V_k^H B \mathbf{v}_{k+1} \\ \mathbf{u}_{h+1}^H A U_h \hat{X} & 0 \end{bmatrix} V_{k+1}^H \\
&= U_{h+1} \begin{bmatrix} A_h \hat{X} - \hat{X} B_k - \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \mathbf{e}_1 \mathbf{e}_1^T & -\hat{X} V_k^H B \mathbf{v}_{k+1} \\ \mathbf{u}_{h+1}^H A U_h \hat{X} & 0 \end{bmatrix} V_{k+1}^H \\
&= U_{h+1} \begin{bmatrix} 0 & -\hat{X} V_k^H B \mathbf{v}_{k+1} \\ \mathbf{u}_{h+1}^H A U_h \hat{X} & 0 \end{bmatrix} V_{k+1}^H,
\end{aligned}$$

where $\mathbf{u}_{h+1}$ and $\mathbf{v}_{k+1}$ are the last block columns of U_{h+1} and V_{k+1}, respectively, and the zero matrix in the top-left corner of the block matrix in the last row is given by (5.1).

Since the columns of U_{h+1} and V_{k+1} are orthonormal, the norm of the residual is equal to the norm of the block matrix

$$\begin{bmatrix} 0 & -\hat{X} V_k^H B \mathbf{v}_{k+1} \\ \mathbf{u}_{h+1}^H A U_h \hat{X} & 0 \end{bmatrix}. \tag{5.2}$$

Let us now consider the BRAD

$$A U_{h+1} \underline{K_h}^{(A)} = U_{h+1} \underline{H_h}^{(A)}.$$

Multiplying both the terms of the equations on the right by $\left(K_h{}^{(A)}\right)^{-1}$, where $K_h{}^{(A)}$ is the $bh \times bh$ head principal submatrix of $\underline{K_h}^{(A)}$, noting that the last block row of $\underline{K_h}^{(A)}$ is equal to zero, we have

$$A U_h = U_{h+1} \underline{H_h}^{(A)} \left(K_h{}^{(A)} \right)^{-1}. \tag{5.3}$$

Analogously, if

$$B^H V_{k+1} \underline{K_k}^{(B)} = V_{k+1} \underline{H_k}^{(B)}$$

is a BRAD, we have that

$$B^H V_k = V_{k+1} \underline{H_k}^{(B)} \left(K_k{}^{(B)} \right)^{-1}, \tag{5.4}$$

where $K_k^{(B)}$ is the head $bk \times bk$ principal submatrix of $\underline{K_k}^{(B)}$.

Using (5.3) and (5.4), we can rewrite the matrix (5.2) as

$$\begin{bmatrix} 0 & -\hat{X} \left(K_k^{(B)} \right)^{-H} \left(\underline{H_k}^{(B)} \right)^H V_{k+1}^H \mathbf{v}_{k+1} \\ \mathbf{u}_{h+1}^H U_{h+1} \underline{H_h}^{(A)} \left(K_h^{(A)} \right)^{-1} \hat{X} & 0 \end{bmatrix}; \tag{5.5}$$

exploiting the orthogonality of the columns of U_{k+1} and V_{k+1}, the matrix (5.5) is equal to

$$\begin{bmatrix} 0 & \hat{X}\left(K_k^{(B)}\right)^{-H}\left(\underline{H_k}^{(B)}\right)^H \mathbf{e}_{k+1}^H \\ \mathbf{e}_{h+1}\underline{H_h}^{(A)}\left(K_h^{(A)}\right)^{-1}\hat{X} & 0 \end{bmatrix},$$

where $\mathbf{e}_{h+1} \in \mathbb{C}^{b(h+1)\times b}$ and $\mathbf{e}_{k+1}\mathbb{C}^{b(k+1)\times b}$.

The norm of this matrix can be recovered by the norms of the block vectors

$$\mathbf{e}_{h+1}\underline{H_h}^{(A)}\left(K_h^{(A)}\right)^{-1}\hat{X} \quad \text{and} \quad \hat{X}\left(K_k^{(B)}\right)^{-H}\left(\underline{H_k}^{(B)}\right)^H \mathbf{e}_{k+1}^H.$$

In particular the computation of the norm of the residual does not involve the matrices A and B; hence it can be performed with a computational cost that does not depend on n and m.

6. Residual and pole selection.

The aim of this section is to prove the following theorem.

THEOREM 6.1. *Let $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{m\times m}$, $\mathbf{u} \in \mathbb{C}^{n\times b}$, and $\mathbf{v} \in \mathbb{C}^{m\times b}$. Let $U \in \mathbb{C}^{n\times bh}$ and $V \in \mathbb{C}^{m\times bk}$ be orthonormal block bases for $\mathcal{Q}_h(A, \mathbf{u}, \boldsymbol{\xi}_h^{(A)})$ and $\mathcal{Q}_k(B^H, \mathbf{v}, \boldsymbol{\xi}_k^{(B)})$, respectively, and let $A_h = UAU^H$, $B_k = VBV^H$. Let $X_{h,k} = UY_{h,k}V^H$, where $Y_{h,k}$ is the solution of the Sylvester equation*

$$A_hY_{h,k} - Y_{h,k}B_k = \mathbf{u}^{(h)}(\mathbf{v}^{(k)})^H$$

with $\mathbf{u}^{(h)} = U^H\mathbf{u}$, and $\mathbf{v}^{(k)} = V^H\mathbf{v}$. Let $\chi_A(z) \in \mathbb{P}_h(\mathbb{C}^{b\times b})$ and $\chi_B(z) \in \mathbb{P}_k(\mathbb{C}^{b\times b})$ be monic block characteristic polynomials of A_h with respect to $\mathbf{u}^{(h)}$ and B_k with respect to $\mathbf{v}^{(k)}$, respectively. Define

$$R_A^G(z) = \frac{\chi_A(z)}{Q_A(z)} \quad \textit{and} \quad R_B^G(z) = \frac{\chi_B(z)}{Q_B(z)},$$

where

$$Q_A(z) = \prod_{\xi\in\boldsymbol{\xi}_h^{(A)},\xi\neq\infty}(z-\xi) \quad \textit{and} \quad Q_B(z) = \prod_{\xi\in\boldsymbol{\xi}_k^{(B)},\xi\neq\infty}(z-\xi).$$

Then the residual matrix can be written as $R_{h,k} = \rho_{1,2} + \rho_{2,1} + \rho_{2,2}$, where

$$\begin{aligned}
\rho_{1,2} &= U(R_B^{G^H}(A_h)\circ^{-1}\mathbf{u}^{(h)})(R_B^G(B^H)\circ\mathbf{v})^H,\\
\rho_{2,1} &= \left(R_A^G(A)\circ\mathbf{u}\right)(R_A^{G^H}(B_k)\circ^{-1}\mathbf{v}^{(k)})^HV^H,\\
\rho_{2,2} &= \left(R_A^G(A)\circ\mathbf{u}\left(R_A^G(\infty)\right)^{-1}\right)\left(R_B^G(B^H)\circ\mathbf{v}\left(R_B^G(\infty)\right)^{-1}\right)^H
\end{aligned}$$

with

$$R_A^G(\infty) = \lim_{|\lambda|\to\infty} R_A^G(\lambda) \quad \textit{and} \quad R_B^G(\infty) = \lim_{|\lambda|\to\infty} R_B^G(\lambda).$$

Moreover

$$\|R_{h,k}\|_F^2 = \|\rho_{1,2}\|_F^2 + \|\rho_{2,1}\|_F^2 + \|\rho_{2,2}\|_F^2. \tag{6.1}$$

Remark 6.2. If one of the poles of $\boldsymbol{\xi}_h^{(A)}$ or $\boldsymbol{\xi}_k^{(B)}$ is chosen equal to infinity, then $\rho_{2,2} = 0$.

The representation of the residual matrix given by Theorem 6.1 allows us to provide adaptive techniques for the pole selection for the resolution of Sylvester equations.

6.1. Proof of Theorem 6.1. Theorem 6.1 and the proof we provide in this section are generalizations of the ones provided by Beckermann in [2] for the case of classical rational Krylov methods.

Let us start by introducing a lemma that is needed for the proof of the theorem.

LEMMA 6.3 (block exactness). *For any $R_A(z) \in \mathbb{P}_h(\mathbb{C}^{b\times b})/Q_A(z)$, we have*

$$UU^H R_A(A) \circ \mathbf{u} = U R_A(A_h) \circ \mathbf{u}^{(h)};$$

in particular, if $R_A(z) \in \mathbb{P}_{h-1}(\mathbb{C}^{b\times b})/Q_A(z)$, it holds that

$$R(A) \circ \mathbf{u} = U R_A(A_h) \circ \mathbf{u}^{(h)}.$$

Similarly for any $R_B \in \mathbb{P}_k(\mathbb{C}^{b\times b})/Q_B(z)$, we have that $VV^H R_B(B^H) \circ \mathbf{v} = V R_B(B_k) \circ \mathbf{v}^{(k)}$ and for any $R_B \in \mathbb{P}_{k-1}(\mathbb{C}^{b\times b})/Q_B(z)$, it holds that $R_B(B^H) \circ \mathbf{v} = V R_B(B_k) \circ \mathbf{v}^{(k)}$.

Proof. We only prove the first two identities, since the other claims follow using the same argument. The proof is composed of two parts. First, we suppose that the poles are all equal to infinity, i.e., $Q_A(z) = 1$. Then, we extend the proof for a generic choice of poles.

For the first part, by linearity, it is sufficient to prove the equalities for $R_A(z) = z^j$ with $j \le h$. We proceed by induction on j. If $j = 0$ there is nothing to prove. For $R_A(z) = z^{j+1}$, by the inductive hypothesis we have

$$UU^H A^{j+1}\mathbf{u} = UU^H AA^j\mathbf{u} = UU^H AUA_h^j\mathbf{u}^{(h)} = UA_h^{j+1}\mathbf{u}^{(h)}.$$

Moreover, if $j+1 \le h-1$, $A^{j+1}\mathbf{u} \in \mathcal{Q}_h(A, \mathbf{u}, \infty)$; hence $UU^H A^{j+1}\mathbf{u} = A^{j+1}\mathbf{u}$.

Let now $R_A(z) = P(z)/Q_A(z)$ with $P(z) \in \mathbb{P}(\mathbb{C}^{b\times b})$. Using the commutativity property of Lemma 3.7, we have that

$$R_A(A) \circ \mathbf{u} = Q_A(A)^{-1}P(A) \circ \mathbf{u} = P(A) \circ (Q_A(A)^{-1}\mathbf{u}).$$

Hence, if we let $\mathbf{c} = Q_A(A)^{-1}\mathbf{u}$, from the result of the first step we have

$$UU^H R_A(A) \circ \mathbf{u} = UU^H P(A) \circ \mathbf{c} = UP(A_h) \circ (U^H\mathbf{c})$$

and, if $R_A(A) \in \mathbb{P}_{h-1}(\mathbb{C}^{b\times b})/Q_A(A)$, we have

$$R_A(A) \circ \mathbf{u} = P(A) \circ \mathbf{c} = UP(A_h) \circ (U^H\mathbf{c}).$$

To conclude, it is sufficient to prove that $U^H\mathbf{c} = Q(A_h)^{-1}\mathbf{u}^{(h)}$. Since $\mathbf{u} = Q(A) \circ \mathbf{c}$, by the first step of the proof we have

$$UU^H\mathbf{u} = UU^H Q(A) \circ \mathbf{c} = UQ(A_h) \circ (U^H\mathbf{c}) = UQ(A_h)U^H\mathbf{c}.$$

Since $U^H U = I_{bh}$, multiplying both sides on the left by $Q(A_h)^{-1}U^H$ we get

$$Q(A_h)^{-1}U^H\mathbf{u} = U^H\mathbf{c},$$

which concludes the proof. ∎

COROLLARY 6.4. *Let $\chi_A(z) \in \mathbb{P}_h(\mathbb{C}^{b\times b})$ and $\chi_B(z) \in \mathbb{P}_k(\mathbb{C}^{b\times b})$ be monic block characteristic polynomials for A_h with respect to $\mathbf{u}^{(h)}$ and B_k with respect to $\mathbf{v}^{(k)}$, respectively. Let $R_A^G(z) = \chi_A(z)/Q_A(z)$ and $R_B^G(z) = \chi_B(z)/Q_B(z)$. It holds that*

$$U^H R_A^G(A) \circ \mathbf{u} = 0 \quad \text{and} \quad V^H R_B^G(B) \circ \mathbf{v} = 0.$$

Moreover, $R_A^G(A)\circ\mathbf{u}$ minimizes $\|R(A)\circ\mathbf{u}\|_F$ over all the $R(z)\in\mathbb{P}_h(\mathbb{C}^{b\times b})/Q_A(z)$ such that $R(z)=P(z)/Q_A(z)$, where $P(z)$ is a monic matrix polynomial. Analogously, $R_B^G(B)\circ\mathbf{v}$ minimizes $\|R(B)\circ\mathbf{v}\|_F$ over all the $R(z)\in\mathbb{P}_k(\mathbb{C}^{b\times b})/Q_B(z)$ with monic numerator.

Proof. In the following, we prove the corollary for $R_A^G(A)\circ\mathbf{u}$. The proof for $R_B^G(B)\circ\mathbf{v}$ is the same. By Lemma 6.3 it holds that

$$UU^H R_A^G(A)\circ\mathbf{u} = UR_A^G(A_h)\circ\mathbf{u}^{(h)}=0.$$

Since $U^HU=I_{bh}$, multiplying on the left by U^H we obtain the first equivalence.

The problem of minimizing $\|R(A)\circ\mathbf{u}\|_F$ over all the $R(z)\in\mathbb{P}_h(\mathbb{C}^{b\times b})/Q_A(z)$ with monic numerator can be rewritten as

$$\begin{aligned}
&\min_{\hat{R}(z)\in\mathbb{P}_{h-1}(z)/Q_A(z)} \|Q_A(A)^{-1}A^h\mathbf{u}-\hat{R}(z)\circ\mathbf{u}\|_F\\
&\quad=\min_{\mathbf{y}\in\mathbb{C}^{h\times b}}\|Q_A(A)^{-1}A^h\mathbf{u}-U\mathbf{y}\|_F\\
&\quad=\min_{\mathbf{y}\in\mathbb{C}^{h\times b}}\|(I_b\otimes Q_A(A)^{-1}A^h)\mathrm{vec}(\mathbf{u})-(I_b\otimes U)\mathrm{vec}(\mathbf{y})\|_2.
\end{aligned}$$

The solution of the least square problem is given by the matrix $\mathbf{y}$ such that

$$(I_b\otimes U)^H\left((I_b\otimes Q_A(A)^{-1}A^h)\mathrm{vec}(\mathbf{u})-(I_b\otimes U)\mathrm{vec}(\mathbf{y})\right)=0,$$

which is analogous to asking that $U^H(Q_A(A)^{-1}A^h\mathbf{u}-U\mathbf{y})=0$, that is, the solution of the minimization problem satisfies $U^H(R(A)\circ\mathbf{u})=0$; hence the function $R_A^G(z)$ is the solution. ∎

Lemma 6.3 is usually referred to as the exactness property of rational Krylov spaces. The proof is a generalization of the ones for nonblock rational Krylov methods, which are described in [15, Lemma 4.6].

LEMMA 6.5. *Let $R_A(z)\in\mathbb{P}_h(\mathbb{C}^{b\times b})/Q_A(z)$, and z such that $\det(R_A(z))\neq 0$. Then,*

(6.2)
$$R_A(zI_n)\circ^{-1}\left[R_A(zI_n)\circ\mathbf{x}-R_A(A)\circ\mathbf{x}\right]=UR_A(zI_{bh})\circ^{-1}\left[R_A(zI_{bh})\circ\tilde{\mathbf{x}}-R_A(A_h)\circ\tilde{\mathbf{x}}\right],$$

where $\mathbf{x}:=(zI_n-A)^{-1}\mathbf{u}$ and $\tilde{\mathbf{x}}:=(zI_{bh}-A_h)^{-1}\mathbf{u}^{(h)}$.

Similarly, for any $R_B(z)\in\mathbb{P}_k(\mathbb{C}^{b\times b})/Q_B(z)$ and for each z such that $\det(R_B(z))$ $\neq 0$

$$\begin{aligned}
&R_B(zI_m)\circ^{-1}\left[R_B(zI_m)\circ\mathbf{y}-R_B(B^H)\circ\mathbf{y}\right]\\
&\quad=VR_B(zI_{bk})\circ^{-1}\left[R_B(zI_{bk})\circ\tilde{\mathbf{y}}-R_B(B_k)\circ\tilde{\mathbf{y}}\right],
\end{aligned}$$

where $\mathbf{y}:=(zI_m-B^H)^{-1}\mathbf{v}$ and $\tilde{\mathbf{y}}:=(zI_{bk}-B_k)^{-1}\mathbf{v}^{(k)}$.

Proof. We only derive the first equality; the second follows by an analogous argument. Note that $R_A(zI_n)\circ^{-1}$ is well-defined by the fact that $\det(R_A(z))\neq 0$. By Lemma 3.8 (6.2) is equivalent to

$$[R_A(zI_n)\circ\mathbf{x}-R_A(A)\circ\mathbf{x}](R(z))^{-1}=U\left[R_A(zI_{bh})\circ\tilde{\mathbf{x}}-R_A(A_h)\circ\tilde{\mathbf{x}}\right](R(z))^{-1};$$

hence, multiplying both sides on the right by $R(z)$, it is sufficient to prove that

$$R_A(zI_n)\circ\mathbf{x}-R_A(A)\circ\mathbf{x}=U\left[R_A(zI_{bh})\circ\tilde{\mathbf{x}}-R_A(A_h)\circ\tilde{\mathbf{x}}\right].$$

Since A and $(zI_n - A)^{-1}$ commute and analogously for $(zI_{bh} - A_h)^{-1}$ and A_h, by Lemma 3.7 the claim can be equivalently restated as follows:

$$(6.3)\qquad \begin{aligned} &(zI_n - A)^{-1}\left[R_A(zI_n)\circ \mathbf{u} - R_A(A)\circ \mathbf{u}\right] \\ &\quad = U(zI_{bh} - A_h)^{-1}\left[R_A(zI_{bh})\circ \mathbf{u}^{(h)} - R_A(A_h)\circ \mathbf{u}^{(h)}\right]. \end{aligned}$$

To prove it, we introduce the auxiliary function $G_z(x) := R_A(z) - R_A(x)$. We consider $G_z(x)$ as a function in the variable x, and assume that z is fixed; in particular $G_z(x) = P_z(x)/Q_A(x)$, where $P_z(x)$ is a matrix polynomial of degree h in the variable x. Note that $G_z(A)\circ \mathbf{u} = R_A(zI_n)\circ \mathbf{u} - R_A(A)\circ \mathbf{u}$; indeed, letting $P_z(x) = \sum_{i=0}^{h} \Delta_i x^i \in \mathbb{P}_h(\mathbb{C}^{b\times b})$ and $Q_A(x) = \sum_{i=0}^{h} q_i x^i$, from the definition of $G_z(x)$ we have that

$$\begin{aligned} R_A(x) &= R_A(z) - G_z(x) = (Q_A(x)R_A(z) - P_z(x))/Q_A(x) \\ &= \left[\sum_{i=0}^{h}(q_i R_A(z) - \Delta_i)x^i\right]/Q_A(x). \end{aligned}$$

Hence,

$$\begin{aligned} R_A(A)\circ \mathbf{u} &= Q_A(A)^{-1}\left[\sum_{i=0}^{h} A^i \mathbf{u}(q_i R_A(z) - \Delta_i)\right] \\ &= Q_A(A)^{-1}\left[R_A(z)\sum_{i=0}^{h} q_i A^i \mathbf{u}\right] - Q_A(A)^{-1}\left[\sum_{i=0}^{h} A^i \mathbf{u}\Delta_i\right] \\ &= R_A(z)Q_A(A)^{-1}Q_A(A)\mathbf{u} - G_z(A)\circ \mathbf{u} = R_A(zI_n)\circ \mathbf{u} - G_z(A)\circ \mathbf{u}. \end{aligned}$$

Analogously, it can be proven that $G_z(A_h)\circ \mathbf{u}^{(h)} = R_A(zI_{bh})\circ \mathbf{u}^{(h)} - R_A(A_h)\circ \mathbf{u}^{(h)}$. Using the equivalences introduced before, we may rewrite (6.3) as

$$(6.4)\qquad (zI_n - A)^{-1}G_z(A)\circ \mathbf{u} = U(zI_{bh} - A_h)^{-1}G_z(A_h)\circ \mathbf{u}^{(h)}.$$

By definition, evaluating $G_z(x)$ at $x = zI_b$ yields $G_z(zI_b) = R_A(z) - R_A(zI_b) = 0$. This implies that the linear matrix polynomial $(xI_b - zI_b)$ is a left solvent for $P_z(x)$, and we may write

$$\tilde{G}_z(x) := (z - x)^{-1}G_z(x) = -(xI_b - zI_b)^{-1}G_z(x) \in \mathbb{P}_{h-1}(\mathbb{C}^{b\times b})/Q_A(x).$$

Thanks to the exactness from Lemma 6.3 we have $\tilde{G}_z(A)\circ \mathbf{u} = U\tilde{G}_z(A_h)\circ \mathbf{u}^{(h)}$, which by Lemma 3.4 is equal to (6.4), concluding the proof. ∎

LEMMA 6.6. *Let $\chi_A(z) \in \mathbb{P}_h(\mathbb{C}^{b\times b})$ and $\chi_B(z) \in \mathbb{P}_k(\mathbb{C}^{b\times b})$ be block characteristic polynomials for A_h with respect to $\mathbf{u}^{(h)}$ and B_k with respect to $\mathbf{v}^{(k)}$, respectively. Let $R_A^G(z) = \chi_A(z)/Q_A(z)$ and $R_B^G(z) = \chi_B(z)/Q_B(z)$. We have that*

$$(zI_n - A)^{-1}\mathbf{u} - U(zI_{bh} - A_h)^{-1}\mathbf{u}^{(h)} = R_A^G(zI_n)\circ^{-1} R_A^G(A)\circ (zI_n - A)^{-1}\mathbf{u}$$

and

$$(zI_m - B^H)^{-1}\mathbf{v} - U(zI_{bk} - B_k)^{-1}\mathbf{v}^{(k)} = R_B^G(zI_m)\circ^{-1} R_B^G(B^H)\circ (zI_m - B^H)^{-1}\mathbf{v}.$$

Proof. It follows from Lemma 6.5 observing that $R_A^G(A_h)\mathbf{u}^{(h)} = 0$ and $R_B^G(B_k)\mathbf{v}^{(k)} = 0$. ∎

We are now ready to give the proof of Theorem 6.1.

Proof of Theorem 6.1. To simplify the notation we define $\mathbf{x} = (zI_n - A)^{-1}\mathbf{u}$, $\tilde{\mathbf{x}} = (zI_{bh} - A_h)^{-1}\mathbf{u}^{(h)}$, $\mathbf{y} = (\bar{z}I_m - B^H)^{-1}\mathbf{v}$, and $\tilde{\mathbf{y}} = (\bar{z}I_{bk} - B_k)^{-1}\mathbf{v}^{(k)}$. According to (1.2), letting X be the solution of the Sylvester equation, we have

$$X - X_{h,k} = \frac{1}{2\pi i}\int_{\gamma_A} \mathbf{x}\mathbf{y}^H - U\tilde{\mathbf{x}}\tilde{\mathbf{y}}^H V^H dz,$$

where γ_A is a compact contour with positive orientation that encloses the eigenvalues of A and A_h, but not the eigenvalues of B and B_k. Using Lemma 6.6 we have

$$X - X_{h,k} = \frac{1}{2\pi i}\int_{\gamma_A} \left((\mathbf{x} - U\tilde{\mathbf{x}})\mathbf{y}^H + \mathbf{x}(\mathbf{y} - V\tilde{\mathbf{y}})^H - (\mathbf{x} - U\tilde{\mathbf{x}})(\mathbf{y} - V\tilde{\mathbf{y}})^H\right) dz \tag{6.5}$$

$$= \frac{1}{2\pi i}\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{x}\right)\mathbf{y}^H dz \tag{6.6}$$

$$+ \frac{1}{2\pi i}\int_{\gamma_A} \mathbf{x}\left(R_B^G(\bar{z}I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{y}\right)^H dz \tag{6.7}$$

$$- \frac{1}{2\pi i}\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{x}\right)\left(R_B^G(\bar{z}I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{y}\right)^H dz. \tag{6.8}$$

The residual matrix can be written as $R_{h,k} = A(X - X_{h,k}) - (X - X_{h,k})B$, that is, the sum of the three differences of integrals $A\mathcal{S} - \mathcal{S}B$, where $\mathcal{S}$ is substituted for by (6.6), (6.7), and (6.8). In the following, we study each difference of integrals separately. Concerning (6.6), by Lemma 3.7 we have

$$\frac{1}{2\pi i}A\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{x}\right)\mathbf{y}^H dz - \frac{1}{2\pi i}\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{x}\right)\mathbf{y}^H B dz \tag{6.9}$$

$$= \frac{1}{2\pi i}\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ A\mathbf{x}\right)\mathbf{y}^H dz - \frac{1}{2\pi i}\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{x}\right)(B^H\mathbf{y})^H dz. \tag{6.10}$$

Let now γ_B be a positively oriented compact contour that encloses the eigenvalues of B and B_k, but not the eigenvalues of A and A_h. Since the integrand is $\mathcal{O}(z^{-2})_{z\to\infty}$, we can replace γ_A with γ_B just by changing the sign of the integral. Noting that

$$A\mathbf{x} = (A - zI_n)\mathbf{x} + zI_n\mathbf{x} = -\mathbf{u} + z\mathbf{x} \quad \text{and, analogously,} \quad B^H\mathbf{y} = -\mathbf{v} + \bar{z}\mathbf{y}, \tag{6.11}$$

the sum of integrals in (6.10) can be rewritten as

$$-\frac{1}{2\pi i}\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{u}\right)\mathbf{y}^H dz + \frac{1}{2\pi i}\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{x}\right)\mathbf{v}^H dz.$$

Then, exchanging γ_A with γ_B we obtain

$$\frac{1}{2\pi i}\int_{\gamma_B} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{u}\right)\mathbf{y}^H dz, \tag{6.12}$$

since the integral

$$\frac{1}{2\pi i}\int_{\gamma_B} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{x}\right) \mathbf{v}^H dz$$

vanishes for the residual theorem.

The same technique can be used to write the second difference of integrals as

$$\frac{1}{2\pi i}\int_{\gamma_A} \mathbf{x} \left(R_B^G(\bar{z}I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{v}\right)^H dz. \tag{6.13}$$

Using again the relations in (6.11), the third difference of integrals can be written as $I_{3,1} + I_{3,2}$, where

$$I_{3,1} = \frac{1}{2\pi i}\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{u}\right) \left(R_B^G(\bar{z}I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{y}\right)^H dz$$

and

$$I_{3,2} = -\frac{1}{2\pi i}\int_{\gamma_A} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{x}\right) \left(R_B^G(\bar{z}I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{v}\right)^H dz. \tag{6.14}$$

For a generic choice of poles, it is only guaranteed that the integrand of $I_{3,1}$ is $\mathcal{O}(z^{-1})_{z\to\infty}$; hence, exchanging γ_A with γ_B, we can rewrite $I_{3,1}$ as

$$\left(R_A^G(\infty \cdot I_n) \circ^{-1} R_A^G(A) \circ \mathbf{u}\right) \left(R_B^G(\infty \cdot I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{v}\right)^H \tag{6.15}$$

$$-\frac{1}{2\pi i}\int_{\gamma_B} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{u}\right) \left(R_B^G(\bar{z}I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{y}\right)^H dz. \tag{6.16}$$

Summing (6.12), (6.13), (6.14), (6.15), and (6.16), we obtain

$$\begin{aligned} R_{h,k} = &\left(R_A^G(\infty \cdot I_n) \circ^{-1} R_A^G(A) \circ \mathbf{u}\right) \left(R_B^G(\infty \cdot I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{v}\right)^H \\ &+ \frac{1}{2\pi i}\int_{\gamma_B} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{u}\right) \left(\left(I_m - R_B^G(\bar{z}I_m) \circ^{-1} R_B^G(B^H)\right) \circ \mathbf{y}\right)^H dz \\ &+ \frac{1}{2\pi i}\int_{\gamma_A} \left(\left(I_n - R_A^G(zI_n) \circ^{-1} R_A^G(A)\right) \circ \mathbf{x}\right) \left(R_B^G(\bar{z}I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{v}\right)^H dz. \end{aligned}$$

Applying Lemma 6.6, we have

$$\begin{aligned} R_{h,k} = &\left(R_A^G(\infty \cdot I_n) \circ^{-1} R_A^G(A) \circ \mathbf{u}\right) \left(R_B^G(\infty \cdot I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{v}\right)^H \\ &+ \frac{1}{2\pi i}\int_{\gamma_B} \left(R_A^G(zI_n) \circ^{-1} R_A^G(A) \circ \mathbf{u}\right) \tilde{\mathbf{y}}^H V^H dz \\ &+ \frac{1}{2\pi i}\int_{\gamma_A} U\tilde{\mathbf{x}} \left(R_B^G(\bar{z}I_m) \circ^{-1} R_B^G(B^H) \circ \mathbf{v}\right)^H dz, \end{aligned}$$

and thanks to Lemma 3.8 the above term can be rewritten as

$$\begin{aligned} R_{h,k} = &\left(R_A^G(A) \circ \mathbf{u} \left(R_A^G(\infty)\right)^{-1}\right) \left(R_B^G(B^H) \circ \mathbf{v} \left(R_B^G(\infty)\right)^{-1}\right)^H \\ &+ \frac{1}{2\pi i}\int_{\gamma_B} \left(R_A^G(A) \circ \mathbf{u}\right) \left({R_A^G}^H(\bar{z}I_{bk}) \circ^{-1} \tilde{\mathbf{y}}\right)^H V^H dz \\ &+ \frac{1}{2\pi i}\int_{\gamma_A} U \left({R_B^G}^H(zI_{bh}) \circ^{-1} \tilde{\mathbf{x}}\right) \left(R_B^G(B^H) \circ \mathbf{v}\right)^H dz. \end{aligned}$$

Finally, by Theorem 3.9 we have

$$\begin{aligned} R_{h,k} = &\left(R_A^G(A) \circ \mathbf{u} \left(R_A^G(\infty)\right)^{-1}\right) \left(R_B^G(B^H) \circ \mathbf{v} \left(R_B^G(\infty)\right)^{-1}\right)^H \\ &+ \left(R_A^G(A) \circ \mathbf{u}\right) \left({R_A^G}^H(B_k) \circ^{-1} \mathbf{v}^{(k)}\right)^H V^H \\ &+ U \left({R_B^G}^H(A_h) \circ^{-1} \mathbf{u}^{(h)}\right) \left(R_B^G(B^H) \circ \mathbf{v}\right)^H . \end{aligned}$$

To prove (6.1) consider the orthogonal projectors $\Pi_A = UU^H$ and $\Pi_B = VV^H$. Applying Corollary 6.4 we obtain the sought identities

$$\begin{aligned} \Pi_A R_{h,k}(I_{bk} - \Pi_B) = \rho_{1,2}, \quad (I_{bh} - \Pi_A)R_{h,k}\Pi_B = \rho_{2,1}, \\ \text{and} \quad (I_{bh} - \Pi_A)R_{h,k}(I_{bk} - \Pi_B) = \rho_{2,2}. \end{aligned}$$ ∎

6.2. Pole selection. The results of Theorem 6.1 can be used to adaptively find good poles for the block rational Arnoldi Algorithm 4.1 for the resolution of Sylvester equations.

During this discussion we assume that one of the poles in $\boldsymbol{\xi}_h^{(A)}$ or $\boldsymbol{\xi}_k^{(B)}$ is chosen equal to infinity; hence for Remark 6.2 the term $\rho_{2,2}$ in the formulation of the residual is equal to zero. With this assumption, the norm of the residual is monitored by the norms of

$$\rho_{1,2} = U({R_B^G}^H(A_h) \circ^{-1} \mathbf{u}^{(h)})(R_B^G(B^H) \circ \mathbf{v})^H$$

and

$$\rho_{2,1} = \left(R_A^G(A) \circ \mathbf{u}\right) ({R_A^G}^H(B_k) \circ^{-1} \mathbf{v}^{(k)})^H V^H.$$

Let us start by considering the norm of $\rho_{1,2}$. We have that

$$\|\rho_{1,2}\|_F \le \|{R_B^G}^H(A_h) \circ^{-1} \mathbf{u}^{(h)}\|_F \cdot \|R_B^G(B^H) \circ \mathbf{v}\|_F.$$

By Corollary 6.4, the vector $R_B^G(B^H) \circ \mathbf{v}$ minimizes $\|R(B^H) \circ \mathbf{v}\|_F$ over all $R(z) \in \mathbb{P}_h(\mathbb{C}^{b\times b})/Q_B(z)$ with monic numerator; for this reason, we choose the new pole by minimizing the norm of ${R_B^G}^H(A_h) \circ^{-1} \mathbf{u}^{(h)}$.

Let $\chi_B(z) = \sum_{i=0}^k \Gamma_i z^i$ be a monic block characteristic polynomial of B_k. By the definition of the operator $\circ^{-1}$, we have

$$\left\|{R_B^G}^H(A_h) \circ^{-1} \mathbf{u}^{(h)}\right\|_F = \left\|(I_b \otimes \bar{Q}_B(A_h)) \left(\sum_{i=0}^k \bar{\Gamma}_i \otimes A_h^i\right)^{-1} \operatorname{vec}(\mathbf{v})\right\|_2 ,$$

where $\bar{Q}_B(z)$ is the conjugate of $Q_B(z)$ and $\bar{\Gamma}_i$ denotes the conjugate of the matrix Γ_i.

Assuming for simplicity that A_h is diagonalizable, i.e., $A_h = Z_h D_h {Z_h}^{-1}$ with D_h diagonal matrix, we have the following bound:

$$\begin{aligned} \left\|(I_b \otimes \bar{Q}_B(A_h)) \left(\sum_{i=0}^k \bar{\Gamma}_i \otimes A_h^i\right)^{-1} \operatorname{vec}(\mathbf{v})\|_2 \le \kappa(Z_h)\|\mathbf{v}\right\|_F \\ \times \left\|(I_b \otimes \bar{Q}_B(D_h)) \left(\sum_{i=0}^k \bar{\Gamma}_i \otimes D_h^i\right)^{-1}\right\|_2 , \end{aligned}$$

where $\kappa(Z_h)$ denotes the condition number of Z_h. The two norm of $(I_b \otimes \bar{Q}_B(D_h))$ $(\sum_{i=0}^{k} \bar{\Gamma}_i \otimes D_h^i)^{-1}$ is equal to the two norm of the matrix

$$(\bar{Q}_B(D_h) \otimes I_b) \left(\sum_{i=0}^{k} D_h^i \otimes \bar{\Gamma}_i\right)^{-1} = \begin{bmatrix} \bar{R}_B^{-1}(\lambda_1) & & \\ & \ddots & \\ & & \bar{R}_B^{-1}(\lambda_h) \end{bmatrix},$$

where $\bar{R}_B(z) = \bar{\chi}_B(z)/\bar{Q}_B(z) = (\sum_{i=0}^{h} z^i \bar{\Gamma}_i)/\bar{Q}_B(z)$ and $\lambda_1, \ldots, \lambda_h$ are the eigenvalues of A_h. In particular,

$$\left\| (I_b \otimes \bar{Q}_B(D_h)) \left(\sum_{i=0}^{k} \bar{\Gamma}_i \otimes D_h^i\right)^{-1} \right\|_2 = \max_{i=1,\ldots,h} \|\bar{R}_B^{-1}(\lambda_i)\|_2.$$

This shows that keeping the function $\|\bar{R}_B^{-1}(z)\|_2$ small over the eigenvalues of A_h guarantees a small norm for $\rho_{1,2}$. In order to obtain a condition independent of h, we can ask for $\|\bar{R}_B^{-1}(z)\|_2$ to be small on the field of values of A, which encloses the spectra of all A_h.

In the following, we describe practical methods to adaptively choose poles for $\boldsymbol{\xi}_k^{(B)}$. The same techniques can be used to provide poles for $\boldsymbol{\xi}_h^{(A)}$.

Let us assume we know the matrix B_{k-1} obtained after $k-1$ steps of the block rational Arnoldi Algorithm 4.1 with poles $\boldsymbol{\xi}_{k-1}$ and that we want to choose a new pole to perform the next step of the algorithm. As we saw before, the norm of $\rho_{1,2}$ after the kth step can be monitored by

$$\|\bar{R}_B^{-1}(\lambda)\|_2 = |\lambda - \bar{\xi}_k| \cdot \|\bar{\chi}_k(\lambda)^{-1}\bar{Q}_{k-1}(\lambda)\|_2 \tag{6.17}$$

for $\lambda \in \mathbb{W}(A)$, where $Q_{k-1}(z) = \prod_{\xi\in\boldsymbol{\xi}_{k-1},\xi\neq\infty} z-\xi$ and $\chi_k(z)$ is the block characteristic polynomial of B_k. In practice we assume that the block characteristic polynomial of B_{k-1}, say $\chi_{k-1}(z)$, well approximates $\chi_k(z)$ over $\mathbb{W}(A)$; hence we approximate (6.17) by

$$|\lambda - \bar{\xi}_k| \cdot \|\bar{\chi}_{k-1}(\lambda)^{-1}\bar{Q}_{k-1}(\lambda)\|_2. \tag{6.18}$$

To keep (6.18) small over $\mathbb{W}(A)$ we can choose ξ_k as the conjugate of

$$\arg\max_{\lambda\in\mathbb{W}(A)} \|\bar{\chi}_k(\lambda)^{-1}\bar{Q}_{k-1}(\lambda)\|_2.$$

Remark 6.7. If $\mathbb{W}(A)$ has a nonempty interior, for the maximum modulus principle it is sufficient to maximize the function over its boundary.

Remark 6.8. In the case of classical rational block Krylov, i.e., $b = 1$ for the resolution of Lyapunov equations, that is $B = -A$, this result reduces to the choice of poles developed in [9] for the case of A symmetric and in [10] for generic A.

The numerical computation of ξ_k, using the definition of the block characteristic polynomial given by Theorem 3.14, is often inaccurate, because the condition number of the matrices W_i is often large. This problem can be overcome by developing an alternative way to compute the norm of the evaluation of block characteristic polynomials. We leave this for future research since the result is beyond the purpose of this work.

We now provide two methods to monitor the Euclidean norm of $\bar{\chi}_{k-1}(\lambda)^{-1}\bar{Q}_{k-1}(\lambda)$ avoiding an explicit computation, noting that it equals $1/\sigma_{\min}(\lambda)$, where $\sigma_{\min}(\lambda)$ is the minimum singular value of $\bar{\chi}_{k-1}(\lambda)/\bar{Q}_{k-1}(\lambda)$.

The first method is to approximate the maximizer of $1/\sigma_{min}(\lambda)$, for $\lambda \in \mathbb{W}(A)$ with the maximizer of the inverse of $|\det(\bar{\chi}_{k-1}(\lambda)/\bar{Q}_{k-1}(\lambda))|$ since the absolute value of the determinant is the product of all the singular values. From Theorem 3.14 it can be noticed that

$$\det(\bar{\chi}_{k-1}(\lambda)) = \prod_{\mu\in\Lambda(B_{k-1})} (\lambda - \bar{\mu});$$

hence the choice of the new pole reduces to the conjugate of

$$\arg\max_{\lambda\in\mathbb{W}(A)} \frac{\prod_{\xi\in\boldsymbol{\xi}_{k-1},\xi\neq\infty} |\lambda-\bar{\xi}|^b}{\prod_{\mu\in\Lambda(B_{k-1})}|\lambda-\bar{\mu}|}. \tag{6.19}$$

We refer to this pole selection strategy as *adaptive determinant minimization* (ADM).

Remark 6.9. In the case of the solution of Lyapunov equations, this choice of poles has already been suggested in [10] as a possible generalization of the technique developed for nonblock rational Krylov methods. This result produces a theoretical justification of such a generalization and an extension to the resolution of Sylvester equations.

To introduce the second method assume B_{k-1} diagonalizable. In such a case, if we let

$$\chi_{k-1}(z) = \prod_{i=1}^{k-1}(zI_b - S_i),$$

as described in Theorem 3.14, the matrices S_i are also diagonalizable; hence

$$\begin{aligned} \|\bar{\chi}_{k-1}(\lambda)^{-1}\bar{Q}_{k-1}(\lambda)\|_2 &\le |\bar{Q}_{k-1}(\lambda)| \prod_{i=1}^{k-1} \|(\lambda I_b - \bar{S}_i)^{-1}\|_2 \\ &\le |\bar{Q}_{k-1}(\lambda)| \prod_{i=1}^{k-1} \frac{\kappa(X_i)}{|\Lambda_{\min}(\lambda-\bar{S}_i)|}, \end{aligned} \tag{6.20}$$

where X_i is the matrix of eigenvectors of $\bar{S}_i$ and $\Lambda_{\min}(\lambda - \bar{S}_i)$ denotes the smallest modulus eigenvalue of $\lambda-\bar{S}_i$ for each i. From Theorem 3.14 we see that the matrices S_i can be recovered by an arbitrary eigendecomposition of the matrix B_{k-1}; in particular, for a fixed λ we can construct S_i using an ordered eigendecomposition of B_{k-1}, where the eigenvalues $\{\mu_i\}$ of B_{k-1} are ordered such that $|\bar{\lambda}-\mu_1| \le |\bar{\lambda}-\mu_2| \le \cdots \le |\bar{\lambda}-\mu_{k-1}|$. With this construction the eigenvalues of S_i are $\mu_{(i-1)b+1}, \mu_{(i-1)b+2}, \ldots, \mu_{ib}$ and (6.20) can be rewritten as

$$\|\bar{\chi}_B(\lambda)^{-1}\bar{Q}_{k-1}(\lambda)\|_2 \le \left(\prod_{i=1}^{k-1}\kappa(X_i)\right)|\bar{Q}_B(\lambda)| \left(\prod_{i=1}^{k-1}(|\lambda-\bar{\mu}_{(i-1)b+1}|)^{-1}\right).$$

This suggests a new method to choose the next shift: ξ_k can be taken as the conjugate of

$$\arg\max_{\lambda\in\mathbb{W}(A)} \left(\prod_{\xi\in\boldsymbol{\xi}_B,\xi\neq\infty} |\lambda-\bar{\xi}| \prod_{i=1}^{k-1}(|\lambda-\bar{\mu}_{(i-1)b+1}|)^{-1}\right), \tag{6.21}$$

where μ_i are the eigenvalues of B_{k-1} ordered as described before.

We refer to this pole selection strategy as *subsampled ADM* (`sADM`).

Remark 6.10. The main advantage of this choice of poles with respect to the previous one is that we have to maximize a rational function with a much smaller degree.

7. Numerical experiments.

In this section we provide some numerical experiment to show the convergence of the block rational Arnoldi Algorithm 4.1 using poles determined in section 6.2: throughout the section, the algorithms that choose poles accordingly to (6.19) and (6.21) are denoted by `ADM` and `sADM`, respectively. The pole ξ_0 is always chosen equal to infinity, and the techniques developed in section 4.1 are employed to guarantee the last pole equal to infinity at each step. This allows computing the residual as described in section 5 avoiding extra computational costs. The implementation of block rational Arnoldi algorithms is based on the `rktoolbox` for MATLAB, developed in [6].

The numerical simulations have been run on a Intel(R) Core(TM) i5-8250U CPU processor running Ubuntu and MATLAB R2022b.

The experiments only involve real matrices; hence, if a nonreal pole is employed, the subsequent is chosen as its conjugate; this allows us to avoid complex matrices. We refer the reader to [22] for a more complete discussion.

In the first experiment, we compute the approximate solution of the Poisson equation

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ u \equiv 0 & \text{on } \partial\Omega, \end{cases} \qquad \Omega = [0,1]^2.$$

We discretize the domain with a uniformly spaced grid with $n = 4096$ points in each direction, and the operator Δ by finite differences, which yields the Lyapunov equation

$$AX + XA = F \quad \text{with} \quad A = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & & \\ 1 & -2 & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & -2 \end{bmatrix},$$

where $h = \frac{1}{n-1}$ is the distance between the grid points and F is the matrix obtained evaluating f on the grid points. If the function f is a smooth bivariate function, the matrix F is numerically low rank, that is, it can be approximated by a low-rank matrix UV^H, where $U, V \in \mathbb{C}^{n\times b}$ for an appropriate $b \ll n$; see, e.g., [14, section 2.7].

Figure 1 shows the behavior of the normalized residual $R_k/\|UV^H\|_F$ for the solution of the Poisson equation with $f(x,y) = 1/(1+x+y)$ with the two proposed choices of poles. In this case, the matrix F has numerical rank 8. We also compared the results with the extended Krylov proposed in [24], which is a block rational Krylov method that alternates a pole equal to zero and a pole equal to infinity. We remark that the iterations of the extended Krylov method are usually faster than a generic block rational Krylov method since in the iterations associated with poles equal to infinity the linear systems are replaced by matrix products and the iterations associated with poles equal to zero are improved using a factorization of the matrix. Table 1 contains times and number of iterations required to reach a relative norm of the residual less than 10^{-8} for the solution of a discretized Poisson equation with block rational Krylov methods with different choices of poles.

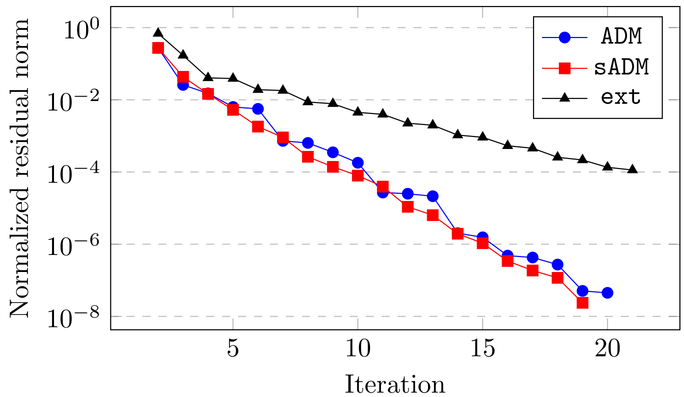

FIG. 1. *Behavior of the residual produced by solving the Poisson equation with block rational Krylov methods, with different choices of poles. Note: color appears only in the online article.*

TABLE 1
Iterations and time needed to reach a relative norm of the residual less than 10^{-8} *for the solution of a discretized Poisson equation with block rational Krylov methods with different choices of poles.*

Poles	Iter	Residual	Time (s)
ADM	21	$8.82e-09$	0.92
sADM	20	$9.19e-09$	1.10
Ext	53	$9.30e-09$	5.91

The second experiment is the computation of an approximate solution for the convection-diffusion partial differential equation

$$\begin{cases} -\epsilon\Delta u + \mathbf{w}\cdot\nabla u = f & \text{in } \Omega, \\ u \equiv 0 & \text{on } \partial\Omega, \end{cases} \qquad \Omega = [0,1]^2,$$

where $\epsilon \in \mathbb{R}_+$ is the viscosity parameter and $\mathbf{w}$ is the convection vector. Assuming $\mathbf{w} = (\Phi(x), \Psi(y))$, and discretizing the domain with a uniformly spaced grid as before, we obtain the Sylvester equation

$$(\epsilon A + \mathbf{\Phi} B)X + X(\epsilon A + B^H \mathbf{\Psi}) = F,$$

where A and F are defined as in the first experiment;

$$\mathbf{\Phi} = \begin{bmatrix} \Phi(h) & & & \\ & \Phi(2h) & & \\ & & \ddots & \\ & & & \Phi((n-2)h) \end{bmatrix}, \quad \mathbf{\Psi} = \begin{bmatrix} \Psi(h) & & & \\ & \Psi(2h) & & \\ & & \ddots & \\ & & & \Psi((n-2)h) \end{bmatrix},$$

and

$$B = \frac{1}{2h}\begin{bmatrix} 0 & 1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & -1 & 0 \end{bmatrix}$$

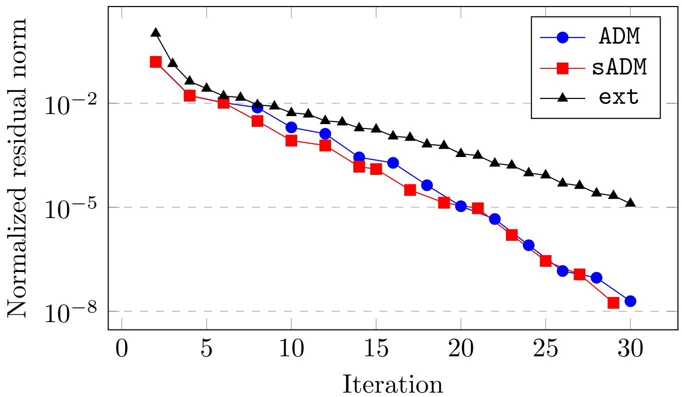

FIG. 2. *Behavior of the residual produced by solving the convection-diffusion equation with block rational Krylov methods with different choices of poles. Note: color appears only in the online article.*

TABLE 2
Iterations and time needed to reach a relative norm of the residual less than 10^{-8} *for the solution of a discretized convection-diffusion equation with block rational Krylov methods with different choices of poles.*

Poles	Iter	Residual	Time (s)
ADM	32	$2.18e-09$	2.12
sADM	31	$9.38e-09$	2.05
Ext	54	$7.55e-09$	7.42

is the discretization by centered finite differences of the first order derivative in each direction.

Figure 2 shows the behavior of the normalized residual for the solution of the convection-diffusion equation with $\epsilon = 0.0083$, $f(x, y) = 1/(1 + x + y)$, $\mathbf{w} = (1 + \frac{(x+1)^2}{4}, \frac{1}{2}y)$ with the two proposed choices of poles and the extended Krylov method. Table 2 contains times and number of iterations required to reach a relative norm of the residual less than 10^{-8} for the solution of a discretized confection-diffusion equation with block rational Krylov methods with different choices of poles.

8. Conclusions. In this work we have proposed a method for solving low-rank Sylvester equations by means of projection onto block rational Krylov subspaces. The key advantage of the method with respect to state-of-the-art techniques is the possibility of exploiting the reordering of poles to maintain the "last" pole of the space equal to ∞. This choice makes the residual of the large-scale equation easily computable in the projected one, without the need of artificially increasing the size of the subspace by introducing unnecessary poles at infinity.

We have also reconsidered the convergence analysis for Krylov solvers for Sylvester equations of [2], extending it to block rational Krylov subspaces by means of the theoretical tools used in [19] for the polynomial case. The analysis allows us to design new strategies for adaptive pole selection, obtained by minimizing the norm of a small $b \times b$ rational matrix, where b is the block size. The minimization problem can be made simpler by replacing the norm with a surrogate function that is easier to evaluate. In [10] the authors propose a heuristic for the pole selection in a block rational Krylov method, based on their analysis of the nonblock case. Choosing the determinant as a surrogate function yields exactly this heuristic, and it gives a solid theoretical justification to this approach. Other choices, instead, yield completely novel strategies. One of these, called `sADM` in the paper, has comparable or better performances than the state of the art on the considered examples.

We expect that the results in this work will help to devise other pole selection strategies and convergence analyses in rational block Krylov methods. This will be subject to future research.

The resulting algorithm is a robust solver for Sylvester equations, and the code has been made freely available at https://github.com/numpi/rk_adaptive_sylvester.

Reproducibility of computational results. This paper has been awarded the "SIAM Reproducibility Badge: Code and data available" as a recognition that the authors have followed reproducibility principles valued by SISC and the scientific computing community. Code and data that allow readers to reproduce the results in this paper are available at https://github.com/numpi/rk_adaptive_sylvester and in the supplementary materials (125720_2_supp_537920_object_rzygf9.zip [local/web 7.20KB]).

REFERENCES

[1] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*, SIAM, Philadelphia, 2005.

[2] B. Beckermann, *An error analysis for rational Galerkin projection applied to the Sylvester equation*, SIAM J. Numer. Anal., 49 (2011), pp. 2430–2450.

[3] B. Beckermann and L. Reichel, *Error estimates and evaluation of matrix functions via the Faber transform*, SIAM J. Numer. Anal., 47 (2009), pp. 3849–3883.

[4] B. Beckermann and A. Townsend, *On the singular values of matrices with displacement structure*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 1227–1248.

[5] P. Benner, S. Gugercin, and K. Willcox, *A survey of projection-based model reduction methods for parametric dynamical systems*, SIAM Rev., 57 (2015), pp. 483–531.

[6] M. Berljafa, S. Elsworth, and S. Güttel, *A Rational Krylov Toolbox for MATLAB*, https://eprints.maths.manchester.ac.uk/2773/1/guide.pdf (2014).

[7] D. A. Bini, G. Latouche, and B. Meini, *Numerical Methods for Structured Markov Chains*, Oxford University Press, Oxford, 2005.

[8] A. Casulli and I. Simunec, *Computation of Generalized Matrix Functions with Rational Krylov Methods*, Math. Comp., 92 (2023), pp. 749–777.

[9] V. Druskin, C. Lieberman, and M. Zaslavsky, *On adaptive choice of shifts in rational Krylov subspace reduction of evolutionary problems*, SIAM J. Sci. Comput., 32 (2010), pp. 2485–2496.

[10] V. Druskin and V. Simoncini, *Adaptive rational Krylov subspaces for large-scale dynamical systems*, Systems Control Lett., 60 (2011), pp. 546–560.

[11] S. Elsworth and S. Güttel, *The block rational Arnoldi method*, SIAM J. Matrix Anal. Appl., 41 (2020), pp. 365–388.

[12] I. Gohberg, P. Lancaster, and L. Rodman, *Matrix Polynomials*, Academic, New York, 1982.

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 2013.

[14] L. Grasedyck, D. Kressner, and C. Tobler, *A literature survey of low-rank tensor approximation techniques*, GAMM-Mitt, 36 (2013), pp. 53–78.

[15] S. Güttel, *Rational Krylov Methods for Operator Functions*, Ph.D. thesis, Technische Universitat Bergakademie Freiberg, Freiberg, Germany, 2010.

[16] M. D. Kent, *Chebyshev, Krylov, Lanczos: Matrix Relationships and Computations*, Ph.D. thesis, Stanford University, Stanford, CA, 1989.

[17] P. Lancaster, *Explicit solutions of linear matrix equations*, SIAM Rev., 12 (1970), pp. 544–566.

[18] P. Lancaster and F. Tisseur, *Hermitian quadratic matrix polynomials: Solvents and inverse problems*, Linear Algebra Appl., 436 (2012), pp. 4017–4026.

[19] K. Lund, *A New Block Krylov Subspace Framework with Applications to Functions of Matrices Acting on Multiple Vectors*, Ph.D. thesis, Temple University, Philadelphia, PA, 2018.

[20] D. Palitta, S. Pozza, and V. Simoncini, *The short-term rational Lanczos method and applications*, SIAM J. Sci. Comput., 44 (2022), pp. A2843–A2870.
[21] D. Palitta and V. Simoncini, *Matrix-equation-based strategies for convection–diffusion equations*, BIT, 56 (2016), pp. 751–776.
[22] A. Ruhe, *The rational Krylov algorithm for nonsymmetric eigenvalue problems.* III*: Complex shifts for real matrices*, BIT, 34 (1994), pp. 165–176.
[23] V. Simoncini, *Ritz and pseudo-Ritz values using matrix polynomials*, Linear Algebra Appl., 241 (1996), pp. 787–801.
[24] V. Simoncini, *A new iterative method for solving large-scale Lyapunov matrix equations*, SIAM J. Sci. Comput., 29 (2007), pp. 1268–1288.
[25] V. Simoncini, *Computational methods for linear matrix equations*, SIAM Rev., 58 (2016), pp. 377–441.
[26] V. Simoncini and E. Gallopoulos, *Convergence properties of block GMRES and matrix polynomials*, Linear Algebra Appl., 247 (1996), pp. 97–119.
[27] A. Townsend and S. Olver, *The automatic solution of partial differential equations using a global spectral method*, J. Comput. Phys., 299 (2015), pp. 106–123.