Functions of rational Krylov space matrices and their decay properties

Noname manuscript No.
(will be inserted by the editor)

Functions of Rational Krylov Space Matrices and Their Decay Properties

Stefano Pozza · Valeria Simoncini

Received: date / Accepted: date

Abstract Rational Krylov subspaces have become a fundamental ingredient in numerical linear algebra methods associated with reduction strategies. Nonetheless, many structural properties of the reduced matrices in these subspaces are not fully understood. We advance in this analysis by deriving bounds on the entries of rational Krylov reduced matrices and of their functions, that ensure an a-priori decay of their entries as we move away from the main diagonal. As opposed to other decay pattern results in the literature, these properties hold in spite of the lack of any banded structure in the considered matrices. Numerical experiments illustrate the quality of our results.

Keywords Rational krylov spaces · Decay pattern · Matrix functions · Lyapunov matrix equations · Faber approximation

1 Introduction

Order reduction of discretized dynamical systems is currently one of the crucial steps in scientific computing and engineering modeling. In geometric terms, the reduction procedure consists in computing a linear vector subspace of small dimension where the original system is projected, in a way so as to preserve the major properties of the original problem. If the reduced problem has significantly smaller dimensions, memory savings and low computational efforts

S. Pozza
Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic. Associated member of ISTI-CNR, Pisa, Italy, and member of INdAM-GNCS group, Italy.
E-mail: pozza@karlin.mff.cuni.cz

V. Simoncini
Dipartimento di Matematica and AM^2, Alma Mater Studiorum - Università di Bologna, Italia, and IMATI-CNR, Pavia, Italia.
E-mail: valeria.simoncini@unibo.it

for its numerical solution can be achieved. In this context, rational Krylov subspaces (RKSs) play a crucial role. Originally introduced by Axel Ruhe in [38] for eigenvalue problems, it has become a standard ingredient for reduction of linear and nonlinear dynamical systems [1,4,5,35], for the approximation of matrix function evaluations ([21]) and in the numerical solution of linear matrix equations [42]. The success of RKSs is mainly due to their ability of capturing accurate information associated with the dynamical system coefficient matrix. As a result, a space dimension dramatically lower than that of the *polynomial* Krylov subspace is sufficient to satisfactorily reduce the original problem. This important property has been experimentally observed and theoretically analyzed over the past few decades, making rational Krylov subspace based strategies the methods of choice.

From the matrix analysis perspective, standard Krylov subspaces enjoy favourable structural properties through the generated reduced matrices, that lead to various computational and analytical advantages. In particular, the reduced matrix is upper Hessenberg, and if the original coefficient matrix is symmetric, so is the reduced one, thus becoming tridiagonal. Thanks to this structure, functions of this (banded) reduced matrix typically show an exponential decay behavior of their components' magnitude, away from the main diagonal. Insightful theoretical results provide solid ground for this phenomenon; we refer for instance to [6,7] and their references for a thorough bibliographic account. Reduced matrices typically have small size and can be explicitly computed. Nevertheless, their decay behavior and that of related quantities can provide the theoretical tools for devising practical computational enhancements, such as truncation, flexible and inner-outer iterations in the solution of linear systems and eigenvalue problems; see, e.g, [8,39,43] for early contributions. Decay bounds for small reduced matrices have also been used to motivate and devise new relaxed approaches and stopping criteria for iterative solvers in matrix function evaluations and matrix equation solving, see, e.g., [23],[31],[36].

Reduced matrices stemming from RKSs loose the explicit banded structure, hence a decay behavior of the entries of functions of the reduced matrix cannot be granted by the known theory. However, numerical experiments illustrate that this decay is still present. Hence, there must be some hidden structure in the constructed reduced matrices that allow the entry decay to be preserved. We aim to uncover this structure by deriving a-priori bounds on the entries of the RKS reduced matrices and of their functions; this will ensure a decay of the entries of these matrix functions as we move away from the main diagonal. as we move away from the main diagonal. The decay pattern of the solution to linear matrix equations will also be considered.

This is a synopsis of the paper. After recalling the necessary notation and main definitions, section 2 introduces new structural properties of RKS reduced matrices. A-priori decay bounds for RKS reduced matrices are presented in section 3, while bounds for functions of the reduced matrices are given in section 4. Analogous results are derived for the solution matrix of reduced Lyapunov matrix equations in section 5. Section 6 reports numerical

tests illustrating the quality of the newly introduced bounds. All main results are proved in section 7 by Faber-Dzhrbashyan[1] approximation. Section 8 concludes the paper.

Notation. For a matrix A, we denote with A^* (A^T) its conjugate (real) transpose, and with $\|A\|$ the matrix norm induced by the Euclidean vector norm. For a vector d, $\mathrm{diag}(d)$ is the diagonal matrix having the components of d along its diagonal.

2 Rational Krylov subspaces and related matrices

We start by introducing standard and rational Krylov subspaces, emphasizing their structural differences.

Given a matrix $A \in \mathbb{R}^{N\times N}$ with spectrum $\lambda(A)$ and a vector $\boldsymbol{v} \neq 0$, the mth step of the Arnoldi algorithm produces the $N \times m$ matrix $U_m = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]$ whose orthonormal columns are a basis of the (polynomial) Krylov subspace

$$\mathcal{P}_m(A, \boldsymbol{v}) := \mathrm{span}\big\{\boldsymbol{v}, A\,\boldsymbol{v}, \ldots, A^{m-1}\,\boldsymbol{v}\big\}\,.$$

Starting with $\boldsymbol{u}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|$, the Arnoldi method uses the Gram-Schmidt orthogonalization process to define the recurrence

$$t_{j+1,j}\boldsymbol{u}_{j+1} = A\boldsymbol{u}_j - \sum_{i=1}^{j} t_{i,j}\boldsymbol{u}_i, \quad j = 1, \ldots, m,$$

and it can be represented in matrix form as

$$AU_m = U_m T_m + t_{m+1,m}\boldsymbol{u}_{m+1}\boldsymbol{e}_m^T, \tag{2.1}$$

with T_m the $m \times m$ upper Hessenberg matrix with nonzero entries $t_{i,j}$, and $\boldsymbol{e}_m$ the mth element of the canonical basis. By orthogonality, (2.1) yields

$$T_m = U_m^* A\, U_m. \tag{2.2}$$

The matrix T_m plays two roles in the algorithm: it represents both the orthogonalization process and the projection and restriction of A in the Krylov subspace $\mathcal{P}_m(A, \boldsymbol{v})$.

In rational Krylov methods, these two roles are decoupled and represented by different matrices. Setting $\boldsymbol{\sigma}_{m-1} = [\sigma_1, \ldots, \sigma_{m-1}]$ with $\sigma_j \notin \lambda(A)$, a rational Krylov subspace is defined as

$$\mathcal{K}_m(A, \boldsymbol{v}, \boldsymbol{\sigma}_{m-1}) := \mathrm{span}\left\{\boldsymbol{v}, (A-\sigma_1 I)^{-1}\,\boldsymbol{v}, \ldots, \prod_{j=1}^{m-1}(A-\sigma_j I)^{-1}\,\boldsymbol{v}\right\}. \tag{2.3}$$

[1] Note that Dzhrbashyan is often also transliterated as Djrbashian. We chose to use Dzhrbashyan in accordance with [44].

Note that

$$\mathcal{K}_m(A, \boldsymbol{v}, \boldsymbol{\sigma}_{m-1}) = \pi_{m-1}(A)^{-1}\mathcal{P}_m(A, \boldsymbol{v}), \quad \pi_{m-1}(A) = \prod_{j=1}^{m-1}(A - \sigma_j I).$$

Analogously to the Arnoldi algorithm, the mth iteration of the rational Krylov subspace method (RKSM) ([37,38]) gives the matrix $V_m = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_m]$ whose orthonormal columns are a basis of $\mathcal{K}_m(A, \boldsymbol{v}, \boldsymbol{\sigma}_{m-1})$. Moreover the RKSM is a Gram-Schmidt orthogonalization process for this space, and it can be expressed by the following recurrence relation

$$h_{j+1,j}\boldsymbol{v}_{j+1} = (A - \sigma_j I)^{-1}\boldsymbol{v}_j - \sum_{i=1}^{j} h_{i,j}\boldsymbol{v}_i, \quad j = 1, \dots, m, \tag{2.4}$$

with $\boldsymbol{v}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|$ and

$$h_{i,j} = \boldsymbol{v}_i^*(A - \sigma_j I)^{-1}\boldsymbol{v}_j, \quad h_{j+1,j} = \|\boldsymbol{v}_{j+1}\|, \quad i, j = 1, \dots, m. \tag{2.5}$$

The recurrence (2.4) can be represented in matrix form as

$$A\,V_m H_m = V_m K_m - h_{m+1,m}(A - \sigma_m I)\boldsymbol{v}_{m+1}\boldsymbol{e}_m^T, \tag{2.6}$$

with $H_m := (h_{i,j})_{i,j=1,\dots,m}$ upper Hessenberg, and $K_m = I + H_m\,\mathrm{diag}(\boldsymbol{\sigma}_m)$; see, e.g., [21,22,38]. Hence, in the rational case, the information about the orthogonalization and its recurrence are carried by the matrix H_m. At the same time, we can define the *reduced-order matrix*

$$J_m := V_m^* A V_m = K_m\,H_m^{-1} - h_{m+1,m}V_m^*(A - \sigma_m I)\,\boldsymbol{v}_{m+1}\boldsymbol{e}_m^T\,H_m^{-1}, \tag{2.7}$$

which is the projection and restriction of A onto $\mathcal{K}_m(A, \boldsymbol{v}, \boldsymbol{\sigma}_{m-1})$. The matrix J_m is also known as *compression* of A or as *matrix Rayleigh quotient*; e.g., [21]. The matrix J_m is not generally a Hessenberg matrix, except for some special choices of the shifts; see, e.g., [11,26,27,40].

From (2.7) a relation between the entries of H_m and those of a function of J_m can be determined. Let $\boldsymbol{w}_m = h_{m+1,m}V_m^*(A - \sigma_m I)\,\boldsymbol{v}_{m+1}$. Then $J_m H_m - K_m = -\boldsymbol{w}_m\boldsymbol{e}_m^T$, and substituting K_m,

$$J_m H_m - H_m\mathrm{diag}(\boldsymbol{\sigma}_m) = I - \boldsymbol{w}_m\boldsymbol{e}_m^T. \tag{2.8}$$

For the jth column $(H_m)_{:,j}$ with $j < m$, it holds that $(J_m - \sigma_j I)(H_m)_{:,j} = \boldsymbol{e}_j$, so that

$$h_{i,j} = \boldsymbol{e}_i^T(J_m - \sigma_j I)^{-1}\boldsymbol{e}_j, \quad i = 1, \dots, m,\ j = 1, \dots, m-1. \tag{2.9}$$

This provides a relation between the pattern of J_m and H_m; note in particular that the elements with $i > j + 1$ will be zero. In terms of the Arnoldi-type recurrence, the relation $(H_m)_{:,j} = (J_m - \sigma_j I)^{-1}\boldsymbol{e}_j$ substituted in (2.4) yields

$$(A - \sigma_j I)^{-1}\boldsymbol{v}_j = V_m(J_m - \sigma_j I)^{-1}\boldsymbol{e}_j, \quad j < m,$$

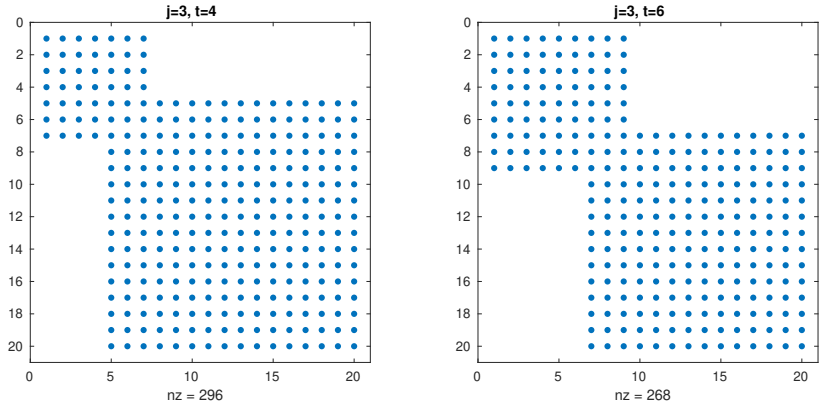

Fig. 2.1 Sparsity pattern of $s_j^{(t)}(J_m)$ in Proposition 2.1 for J_{20} and Hermitian matrix A.

explicitating the role of J_m in the projection operation. For $m = N$, $\boldsymbol{w}_m = 0$ so that $J_N = K_N\, H_N^{-1}$, and the result also holds for $j = m$. The consequences of this fact are described in the next proposition, which will also be used in later proofs.

Proposition 2.1 *Let J_m be as defined in* (2.7) *and consider the rational function*

$$s_j^{(t)}(x) := \frac{q_j(x)}{(x-\sigma_t)\cdots(x-\sigma_{t+j-1})},$$

with $t \geq 1$ and $q_j(x)$ a polynomial of degree at most j. If the indexes k, ℓ are such that $k \geq t+2$ and $\ell \leq t$, then

$$\left(s_j^{(t)}(J_m)\right)_{k,\ell} = 0, \quad j = 1, \ldots, k-t-1.$$

In the Arnoldi algorithm, the connection between the orthogonalization process and the sparsity pattern of the matrix T_m is self-evident. Proposition 2.1 shows that J_m has a hidden sparsity structure, determined by the orthogonality property of V_m; see the proof in section 7.1. Figure 2.1 illustrates the revealed structure in the Hermitian case. The left plot shows $|s_3^{(4)}(J_{20})|_{k,\ell} = 0$ for $k \geq j+t+1 = 8$, and $\ell \leq 4$, while the right plot displays $|s_3^{(6)}(J_{20})|_{k,\ell} = 0$ for $k \geq j+t+1 = 10$, and $\ell \leq 6$.

A particular case of Proposition 2.1 is that for $t \in \{1, \ldots, m-1\}$, it holds that

$$(J_m - \sigma_t I)^{-1} = \left[\begin{array}{c|c} \mathcal{J}_{11} & \mathcal{J}_{12} \\ \hline \mathcal{O} & \mathcal{J}_{22} \end{array}\right],$$

with $\mathcal{J}_{11} \in \mathbb{C}^{(t+1)\times t}, \mathcal{J}_{12} \in \mathbb{C}^{(t+1)\times(m-t)}, \mathcal{J}_{22} \in \mathbb{C}^{(m-t-1)\times(m-t)}$. That is, all elements with indexes (k, ℓ) with $k \geq t+2$ and $\ell \leq t$ are zero. In the case of J_m Hermitian a corresponding zero block will appear in the right-top corner.

Another striking structural property is that the inverse of $J_m - \widehat{S}_m$ is upper Hessenberg, where $\widehat{S}_m$ is a diagonal matrix related to the σ_t's.

Proposition 2.2 *Let* $\widehat{S}_m = \mathrm{diag}(\star, \sigma_1, \ldots, \sigma_{m-1})$ *where* $\star$ *stands for any complex scalar. If* $J_m - \widehat{S}_m$ *is invertible, then* $(J_m - \widehat{S}_m)^{-1}$ *is upper Hessenberg, from which it follows that* $(J_m - \widehat{S}_m)^{-1} J_m$ *is also upper Hessenberg.*

Proof Let $S_m = \mathrm{diag}(\boldsymbol{\sigma}_m)$. From (2.8) we can write

$$J_m - \widehat{S}_m = (I + H_m S_m - \boldsymbol{w}_m \boldsymbol{e}_m^T) H_m^{-1} - \widehat{S}_m$$

that is $J_m - \widehat{S}_m = (I + H_m S_m - \boldsymbol{w}_m \boldsymbol{e}_m^T - \widehat{S}_m H_m) H_m^{-1}$, from which we can write

$$(J_m - \widehat{S}_m)^{-1} = H_m (I + H_m S_m - \boldsymbol{w}_m \boldsymbol{e}_m^T - \widehat{S}_m H_m)^{-1}. \tag{2.10}$$

To proceed, we notice that for any jth column with $j < m$, it holds that

$$(I + H_m S_m - \boldsymbol{w}_m \boldsymbol{e}_m^T - \widehat{S}_m H_m)\boldsymbol{e}_j = \boldsymbol{e}_j + H_{:,j}\sigma_j - 0 - \begin{bmatrix} \star h_{1,j} \\ \sigma_1 h_{2,j} \\ \vdots \\ \sigma_j h_{j+1,j} \\ 0 \\ \vdots \end{bmatrix},$$

that is,

$$(I + H_m S_m - \boldsymbol{w}_m \boldsymbol{e}_m^T - \widehat{S}_m H_m)\boldsymbol{e}_j = \begin{bmatrix} (\sigma_j - \star) h_{1,j} \\ (\sigma_j - \sigma_1) h_{2,j} \\ \vdots \\ 1 + (\sigma_j - \sigma_{j-1}) h_{j,j} \\ 0 \\ \vdots \end{bmatrix},$$

so that all the elements below the jth component are zero. As a consequence, the matrix $(I + H_m S_m - \boldsymbol{w}_m \boldsymbol{e}_m^T - \widehat{S}_m H_m)$ and its inverse are upper triangular. Left multiplication by the upper Hessenberg matrix H_m as in (2.10) yields again an upper Hessenberg matrix. We thus have in (2.10) that $(J_m - \widehat{S}_m)^{-1}$ is upper Hessenberg.

To prove the second property, we write

$$(J_m - \widehat{S}_m)^{-1} J_m = (J_m - \widehat{S}_m)^{-1}(J_m - \widehat{S}_m + \widehat{S}_m) = I + (J_m - \widehat{S}_m)^{-1}\widehat{S}_m.$$

Since $(J_m - \widehat{S}_m)^{-1}$ is upper Hessenberg and $\widehat{S}_m$ is diagonal, $(J_m - \widehat{S}_m)^{-1} J_m$ is also upper Hessenberg. □

When A is Hermitian, $(J_m - \widehat{S}_m)^{-1}$ and $(J_m - \widehat{S}_m)^{-1} J_m$ are Hermitian, and must thus be tridiagonal. In this case, the structural property of Proposition 2.2 is also associated with the construction of a sequence of orthogonal

rational functions generating a related rational Krylov space; see further discussion in section 7.

In the next section we will show that, despite having lost the Hessenberg structure, the reduced-order matrix J_m shows an exponential decay in the magnitude of the lower part of J_m as we move away from the first lower diagonal. As we will show in section 7, this decay is connected with the structural properties of J_m described in Proposition 2.1.

While the upper Hessenberg structure of $(J_m - \widehat{S}_m)^{-1}$ ensures a decay property of $J_m - \widehat{S}_m$ [10,34] with a decay rate depending on the spectral properties of $(J_m - \widehat{S}_m)^{-1}$, the matrix J_m^{-1} is no longer upper Hessenberg, hence the decay in the lower part of J_m away from the main diagonal needs to be analyzed explicitly. To give a first insight into these decay properties, consider the expression derived from (2.8)

$$J_m^{-1} = H_m(I + H_m S_m - \boldsymbol{w}_m \boldsymbol{e}_m^T)^{-1}.$$

From the proof of Proposition 2.2, $(I+H_m S_m-\boldsymbol{w}_m \boldsymbol{e}_m^T)$ is an upper Hessenberg matrix. As a consequence, $(I+H_m S_m-\boldsymbol{w}_m \boldsymbol{e}_m^T)^{-1}$ has a decay in its lower part. Since H_m is also upper Hessenberg, we can expect the elements in the lower part of J_m^{-1} to decay away from the main diagonal. Rigorous decay bounds for general functions of J_m are obtained in the next sections.

3 Decay bounds for the RKSM reduced-order matrix

We present an a-priori bound for the absolute value of the elements composing the reduced-order matrix J_m. The proof is postponed to section 7. The bound uses as spectral information the *field of values* (or *numerical range*) of the matrix A, i.e., the (convex) set ([25])

$$W(A) = \{\boldsymbol{v}^* A \boldsymbol{v} \,|\, \boldsymbol{v} \in \mathbb{C}^n, ||\boldsymbol{v}|| = 1\}.$$

If A is Hermitian, then $W(A)$ is the smallest interval containing the spectrum of A.

We also recall that for every convex continuum E there exists a conformal map ϕ (with inverse ψ) from the exterior of E onto the exterior of the closed unit disk $D := \{z \in \mathbb{C} : |z| \leq 1\}$ satisfying the following conditions

$$\phi(\infty) = \infty, \quad \lim_{z\to\infty} \frac{\phi(w)}{w} = d > 0. \tag{3.1}$$

Moreover, we choose $E \subset \mathbb{C}$ such that $W(A) \subset E$, and we let $E_\tau = E \cup \{w \in \mathbb{C} \setminus E : |\phi(w)| \leq \tau\}$ with $\tau > 1$. Figure 3.1 illustrates the role of all these quantities.

We are ready to state the main result of this section.

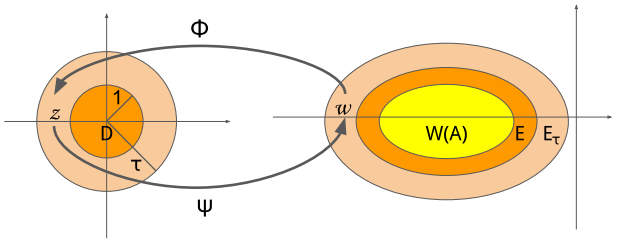

Fig. 3.1 Conformal mappings associated with $W(A)$.

Theorem 3.1 *Let A be a matrix with $W(A) \subset E$, with E, ϕ and ψ as defined above. Let $J_m = V_m^* A V_m$, with $V_m = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_m]$ having orthonormal columns and such that range$(V_m) = \mathcal{K}_m(A, \boldsymbol{v}, \boldsymbol{\sigma}_{m-1})$, with $\boldsymbol{\sigma}_{m-1} = [\sigma_1, \dots, \sigma_{m-1}]$, $\sigma_j \notin E$. For $k, \ell = 1, \dots, m$ with $k - \ell > 1$ it holds*

$$|(J_m)_{k,\ell}| \le 3\frac{\tau}{\tau - 1} \max_{|z| \le \tau} |\psi(z)| \prod_{t=\ell}^{k-2} \frac{\tau + |\phi(\sigma_t)|}{|\phi(\sigma_t)|\tau + 1}, \tag{3.2}$$

for every $\tau > 1$.

We can specialize the previous result as follows.

Corollary 3.2 *With assumptions of Theorem 3.1, for every $\tau > 1$ the bound*

$$|(J_m)_{k,\ell}| \le 3(a\tau + |c|)\frac{\tau}{\tau - 1} \prod_{t=\ell}^{k-2} \frac{\tau + |\phi(\sigma_t)|}{|\phi(\sigma_t)|\tau + 1}, \tag{3.3}$$

holds for the following cases:

1. *$W(A)$ is contained in an ellipse with center c, major axis a, and minor axis r, and the conformal map is defined as*

$$\phi(w) = \frac{w - c + \sqrt{(w-c)^2 - \rho^2}}{\rho R},$$

   with $\rho = \sqrt{a^2 - r^2}$, $R = (a + r)/\rho$.
2. *A is Hermitian with $\lambda(A) \subset (c - a, c + a)$ (with $a > 0$) and the conformal map defined as*

$$\phi(w) = \frac{w - c + \sqrt{(w-c)^2 - a^2}}{a}. \tag{3.4}$$

The previous bounds can be improved by carefully choosing the value for the parameter $\tau > 1$. The optimal (or near optimal) value of τ depends on the set E, on the parameters σ_j, and on k, ℓ. We will describe our strategy for τ together with numerical examples in section 6.

In the Hermitian case, the matrix $G_m = (J_m - \widehat{S}_m)^{-1}$ is tridiagonal by Proposition 2.2. Since J_m can be expressed as $J_m = G_m^{-1} + \widehat{S}_m$, one can use bounds for the inverses of tridiagonal matrices in [10] to derive decay bounds for J_m. However, we did not follow such a strategy for several reasons. First, the bounds in [10] require spectral information of G_m, which can only be obtained in some quite rough form, without running RKSM. Secondly, the bound slopes from [10] strongly depend on the condition number of G_m. Finally, note that while the spectral information on G_m is not readily related to that of A, the bounds in Theorem 3.1 are based on A's field of values, and the shifts σ_j can carry additional information on the properties of A.

4 Decay bounds for functions of the reduced-order matrix J_m

A matrix function can be defined in several ways (see [24, Section 1]); here we will use the definition based on the Cauchy integral formula.

Definition 4.1 *Let A be a complex matrix and f be an analytic function in some open set $\Omega \subset \mathbb{C}$ such that $\lambda(A) \subset \Omega$. Then*

$$f(A) = \int_\Gamma f(z)\,(zI - A)^{-1}\,dz,$$

with $\Gamma \subset \Omega$ a system of Jordan curves encircling each eigenvalue of A exactly once, with mathematical positive orientation.

In the classical Arnoldi iteration, the matrix function $f(T_m)$, with T_m upper Hessenberg determined in (2.2), displays a decay behavior in its lower part, whose slope depends also on f; see, e.g., [6,36] and their references. In this case, the decay phenomenon is derived by the banded structure of T_m. We next show that in the rational case we still have a decay phenomenon for $f(J_m)$ despite J_m not being banded. This is possible thanks to the orthogonality of the columns of V_m, see section 7 for a rigorous argumentation.

Theorem 4.2 *Consider the setting of Theorem 3.1, assume $k - \ell > 1$, and let f be an analytic function in E_τ, for $\tau > 1$. Then*

$$|f(J_m)_{k,\ell}| \le 3\frac{\tau}{\tau - 1} \max_{|z|=\tau} |f(\psi(z))| \prod_{t=\ell}^{k-2} \frac{\tau + |\phi(\sigma_t)|}{|\phi(\sigma_t)|\tau + 1} =: B(k,\ell). \tag{4.1}$$

Setting the coefficients

$$\alpha_j = \frac{1}{2\pi i} \int_{|z|=\tau} \frac{f(\psi(z))}{z} \prod_{t=\ell}^{k-2} \frac{z - \phi(\sigma_t)}{\overline{\phi(\sigma_t)}z - 1} \frac{\phi(\sigma_t)}{|\phi(\sigma_t)|} \left(-\frac{1}{z}\right)^{j-k+\ell+2} dz. \tag{4.2}$$

and a positive integer s, we have the following more refined bound

$$|f(J_m)_{k,\ell}| \le 3 \sum_{j=0}^{s-1} |\alpha_{j+k-\ell-1}| + \frac{B(k,\ell)}{\tau^s}. \tag{4.3}$$

Theorem 4.2 describes families of bounds depending on E_τ. These families can be specialized by choosing the shape of E, i.e., the map ϕ and its inverse ψ. The bounds can be further improved by choosing an optimal (or near optimal) value of τ, which depends on E, f, $\boldsymbol{\sigma}$, k, ℓ. The second bound (4.3) comes at the cost of computing the coefficients α_j by approximating s integrals. As we will see in the numerical examples, a value for the index s of the order of m appears to be enough to get an effective bound.

Using (2.9) and the fact that J_m is Hermitian whenever A is Hermitian, Theorem 4.2 immediately gives the following result for the upper entries of the Hessenberg matrix H_m whose elements are defined in (2.5).

Theorem 4.3 *Consider the setting of Theorem 3.1 with A Hermitian. Let $dist(E_\tau, \sigma_\ell)$ be the distance between the set E_τ and the coefficient σ_ℓ. For $k, \ell = 1, \ldots, m$ such that $\ell - k > 1$, and for every $\tau > 1$ so that $dist(E_\tau, \sigma_\ell) > 0$, it holds*

$$|h_{k,\ell}| \le 3\frac{\tau}{\tau - 1}\,\frac{1}{dist(E_\tau, \sigma_\ell)}\prod_{t=k}^{\ell-2}\frac{\tau + |\phi(\sigma_t)|}{|\phi(\sigma_t)|\tau + 1}.$$

When A is Hermitian, the matrix T_m in (2.2) stemming from the Arnoldi method is Hermitian, i.e., it is tridiagonal. In the rational case, H_m from (2.4) is in general not Hermitian. Nonetheless, for A Hermitian Theorem 4.3 shows that the upper elements in H_m, though nonzero in general, are characterized by a decay behavior.

5 Decay bound for the reduced-order solution of a Lyapunov equation

Krylov subspaces, and in particular their rational version, have been used successfully in the past few years for solving linear matrix problems such as the Lyapunov and Sylvester equations [42]. These equations have a prominent role in control, eigenvalue computations and in reduction strategies for discretized partial differential equations.

Let us assume that A is stable, i.e., all its eigenvalues lay on the complex half-plane with negative real part $\mathbb{C}_-$, and consider the Lyapunov equation

$$AX + XA^* + \boldsymbol{v}\boldsymbol{v}^* = 0\,. \tag{5.1}$$

For simplicity of exposition and without loss of generality we assume $\|\boldsymbol{v}\| = 1$. Under certain assumptions on A, it can be shown that the solution X can be well approximated by a low rank matrix. This property motivated the approximation $X \approx X_m = V_m Y_m V_m^T$, where the columns of V_m span a small dimensional approximation space. In our setting, this space is just the rational Krylov subspace. The reduced matrix Y_m is determined by imposing a so-called Galerkin orthogonality condition on the residual matrix $R_m = AX_m + X_m A^* + \boldsymbol{v}\boldsymbol{v}^*$, namely

$$V_m^T R_m V_m = 0.$$

Substituting the residual matrix into this equation and taking into account that $V_m^T V_m = I$, we obtain the reduced-order Lyapunov equation

$$J_m Y_m + Y_m J_m^* + \boldsymbol{e}_1 \boldsymbol{e}_1^T = 0\,, \tag{5.2}$$

with $\boldsymbol{e}_1$ the first vector of the canonical basis. The solution to this equation yields the sought after reduced approximate solution Y_m; note that Y_m is Hermitian. For more information on the general methodology and its properties we refer, e.g, to [42] and references therein. It is important to also recall that the bottom entries of Y_m are involved in $\|R_m\|$, thus their magnitude closely follows the convergence history of the method. In other words, The matrix Y_m is also characterized by a decay behavior as we move away from the top left corner entries. This property was analyzed for the standard Krylov subspace in [41], where estimates for the entry absolute values were also given.

The following theorem discusses such behavior for the rational Krylov subspace when the field of values of A is contained in a disk; its proof can be found in section 7. First attempts for proving decay bounds for rational Krylov subspaces can be found in [31].

Theorem 5.1 *Let A be a stable matrix with $W(A) \subset E$, where $E \subset \mathbb{C}_-$ is a disk of center $c \in \mathbb{C}_-$ and radius a, and $\phi(w) = (w-c)/a$ is the related conformal map. Let Y_m be the solution to the Lyapunov equation* (5.2) *with V_m spanning the subspace $\mathcal{K}_m(A, \boldsymbol{v}, \boldsymbol{\sigma}_{m-1})$, with real positive parameters σ_j. Given $1 < \tau < -c/a$, then*

$$|(Y_m)_{k,\ell}| \le \min\left\{C(k,\ell), C(\ell,k)\right\}, \tag{5.3}$$

where

$$C(k,\ell) := \frac{9}{|a\tau + 2c| - a} \frac{\tau}{\tau - 1} \prod_{j=1}^{k-2} \frac{\tau + |\phi(\sigma_j)|}{|\phi(\sigma_j)|\tau + 1} \prod_{j=1}^{\ell-2} B_j(\tau) \tag{5.4}$$

and for $j = 1, \ldots, \ell - 2$,

$$B_j(\tau) := \max\left\{ \left| \frac{a\tau + \phi(\sigma_j)a + 2c}{\phi(\sigma_j)(a\tau + 2c) + a} \right|, \left| \frac{-a\tau + \phi(\sigma_j)a + 2c}{\phi(\sigma_j)(-a\tau + 2c) + a} \right| \right\}.$$

Similarly to Theorems 3.1 and 4.2, it is possible to extend the results in Theorem 5.1 for a set E with a different shape. However, this would lead to a more complicated expression for the factors $B_j(\tau)$.

To give a more intuitive expression for the bound, we assume that $\phi(\sigma_j) \ge -3c/a$, for $j = 1, \ldots, \max\{k, \ell\}$, so that $B_j(\tau)$ always takes the second value. Letting $\tau \to -\frac{c}{a}$, we get the bound

$$\frac{-c/a + |\phi(\sigma_j)|}{|\phi(\sigma_j)|(-c/a) + 1} < -\frac{4ac}{3c^2 + a^2}, \quad \text{for} \quad \phi(\sigma_j) = -3\frac{c}{a},$$

obtained by noticing that for larger values of $\phi(\sigma_j)$ the left-hand function is decreasing. Moreover,

$$B_j\left(-\frac{c}{a}\right) = \left| \frac{c + \phi(\sigma_j)a}{\phi(\sigma_j)c + a} \right| < -\frac{a}{c}, \quad \text{for} \quad \phi(\sigma_j) \to +\infty,$$

where B_j is instead increasing with $\phi(\sigma_j)$. Hence we get the simplified bound

$$C(k,\ell) \le \frac{9c}{(a+c)^2}\left(-\frac{4ac}{3c^2+a^2}\right)^{k-1}\left(-\frac{a}{c}\right)^{\ell-1}.$$

Note that since $a < |c|$

$$0 < -\frac{a}{c} < -\frac{4ac}{3c^2+a^2} < 1.$$

Therefore the decay slopes along the rows and the columns are different.

The results of Theorem 5.1 can be generalized to Sylvester linear equations, $A_1X + XA_2 + \boldsymbol{v}\boldsymbol{v}^* = 0$, where A_1, A_2 do not necessarily have the same dimensions, under the assumption that the fields of values of the matrices A_1, and $-A_2$ are disjoint sets; see, e.g., [2].

6 Numerical examples

In this section we present a number of numerical examples to illustrate our findings. To this end we describe how we chose the parameters of the bounds introduced in the previous sections. For the bounds in Corollary 3.2, taking the natural logarithm of the right-hand side of (3.3) and omitting $\log(3)$ yields the function $L : (1,+\infty) \to (0,+\infty)$,

$$L(\tau) := \log(a\tau + |c|) + \log\left(\frac{\tau}{\tau-1}\right) + \sum_{t=\ell}^{k-2} \log\left(\frac{\tau + |\phi(\sigma_t)|}{|\phi(\sigma_t)|\tau + 1}\right),$$

with limits $\lim_{\tau\to+\infty} L(\tau) = +\infty$, $\lim_{\tau\to 1+} L(\tau) = +\infty$ and with global minimum at some τ_*. Moreover, for τ large enough $L(\tau)$ behaves like $\log(a\tau + |c|)$. Therefore any τ not (too) much larger than τ_* is a good choice for the bound (3.3). Fixing $\rho_{k,\ell} = \max\{|\phi(\sigma_\ell)|, \dots, |\phi(\sigma_{k-2})|\}$, in our experiments we used the value of $\tau > 1$ that minimizes

$$\log(a\tau + |c|) + \log\left(\frac{\tau}{\tau-1}\right) + (k-1-\ell)\log\left(\frac{\tau+\rho_{k,\ell}}{\rho_{k,\ell}\,\tau+1}\right);$$

the Matlab function `fminbnd` was used for this purpose.

We applied a similar strategy for the bounds (4.1), (4.3) in Theorem 4.2. In both cases, we considered a set E with $W(A) \subset E$ whose boundary is the ellipse $\Gamma = \{\psi(z) \,|\, |z| = 1\}$, with $\psi(z) = \frac{1}{2}\rho(Rz + (Rz)^{-1}) + c$. The quasi-optimal τ was determined as

$$\tau^* = \mathrm{argmin}_{\tau>1} \log\left(\frac{\tau}{\tau-1}\left|f\left(\frac{\rho}{2}\left(R\tau + \frac{1}{R\tau}\right) + |c|\right)\right| \prod_{t=\ell}^{k-2} \frac{\tau + |\phi(\sigma_t)|}{|\phi(\sigma_t)|\tau + 1}\right).$$

In addition, for the bound (4.3), we set the value of s as the first index such that

$$\sum_{j=0}^{s-1} |\alpha_{j+k-\ell-1}| \ge \frac{B(k,\ell)}{\tau^s},$$

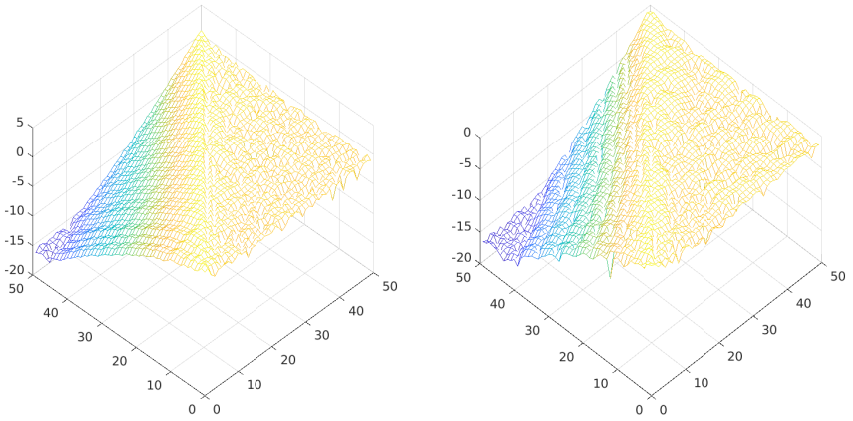

Fig. 6.1 Example 6.1. Left: magnitude of the elements of J_{50}. Right: magnitude of the elements of $\exp(J_{50})$. Logarithmic scale is used.

while the coefficients in (4.2) were approximated by the Matlab function `integral` with default parameters.

In our implementation, the bound (4.3) is computationally much more expensive than the bound (4.1) due to the approximation of the coefficients α_j in (4.2). We included this bound to illustrate that it is possible to obtain a qualitative sharp bound with the given spectral information. Clearly, if computational costs are a concern, a looser approximation to the α_js can still provide a valuable bound. Finally, when using the bound in Theorem 5.1, we set $\tau = -c/a$.

In all experiments, the elements of the vector $\boldsymbol{v}$ were taken from a normally distributed random sequence (Matlab function `randn`).

Example 6.1 We consider the matrix A stemming from the scaled discretization of the 2D Laplacian, in the open unit square,

$$A = L \otimes I + I \otimes L,$$

where $\otimes$ is the Kronecker product and L is a 40×40 tridiagonal matrix whose main diagonal elements are equal to -2, and first upper and lower diagonals have elements equal to 1. The matrix A is symmetric with size 1600×1600, and its spectrum is contained in $E = [-7.9883, -0.0117]$. Following the shift selection in [14] we computed 50 iterations of RKSM on $A, \boldsymbol{v}$ obtaining the reduced-order matrix J_{50}.

Figure 6.1 displays the elements magnitude in J_{50} (left) and that in $\exp(J_{50})$ (right); logarithmic scale is used. Both plots show a decay phenomenon in the lower triangular part of the matrix. An illustration of our bounds is reported in Figure 6.2. In the left plot, the symbol "+" (in red) corresponds to the entries of $|(J_{50})_{:,2}|$, while the blue circles represent the bound in Corollary 3.2 for the

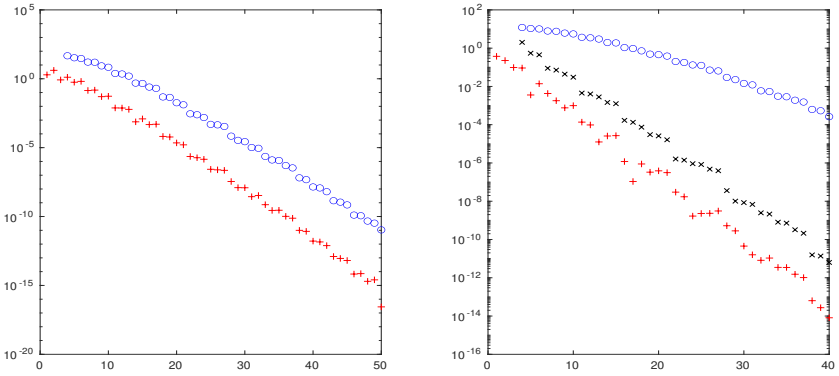

Fig. 6.2 Example 6.1. Left: values of $|(J_{50})_{:,2}|$ (red "+"), Hermitian case bound in Corollary 3.2 (blue circles). Right: values of $|(\exp(J_{50}))_{:,2}|$ (red "+"), bound in (4.1) (blue circles), bound in (4.3) (black "×"). Logarithmic scale is used.

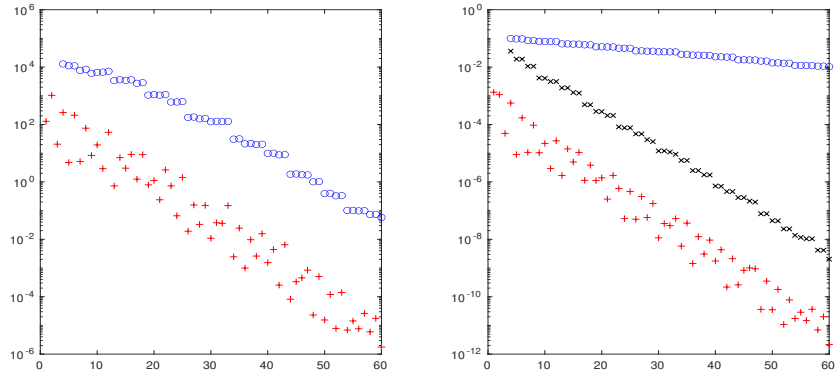

Fig. 6.3 Example 6.2. Left: values of $|(J_{60})_{:,2}|$ (red "+"), Hermitian case bound in Corollary 3.2 (blue circles). Right: values of $|((J_{60} - 100i\, I)^{-1})_{:,2}|$ (red "+"), bound in (4.1) (blue circles), bound in (4.3) (black "×"). Logarithmic scale is used.

Hermitian case. The right plot shows the entries of $|(\exp(J_{50}))_{:,2}|$ (red symbol "+"), the blue circles reproduce the bound in (4.1) and the black symbol "×" reports the bound in (4.3). The latter bound provides a significantly sharper estimate of the actual decay than that based on (4.1). The maximum reached value for s in bound (4.3) is 27.

Example 6.2 We consider the matrix `flowmeter0` from the Oberwolfach Model Reduction Benchmark Collection [29]. The matrix is symmetric, it has size 9669×9669, and its spectrum is contained in the interval $E = [-2.0885\,10^3, -1.3140\,10^{-4}]$. After 60 iterations of RKSM on $A, \boldsymbol{v}$ the reduced-order matrix J_{60} is obtained. In the analysis of dynamical systems an impor-

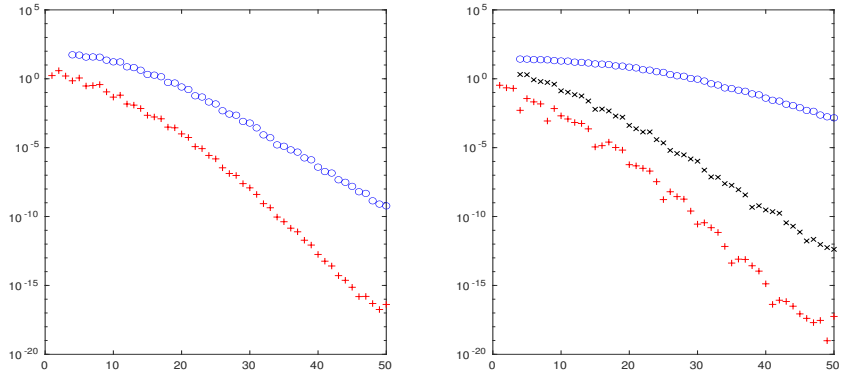

Fig. 6.4 Example 6.3. See description of Figure 6.2. In the left plot the bound for the non-Hermitian case in Corollary 3.2 is employed.

tant role is played by the resolvent. In a reduced model context, and the same input and output control locations, this corresponds to analyzing $(J_{60}-wi\,I)^{-1}$ with $w \in \mathbb{R}$. The plots in Figure 6.3 display information on the reduced resolvent matrix for $w = 100$. More precisely, the left plot reports $|(J_{60})_{:,2}|$ (red "+") and the Hermitian bound in Corollary 3.2 (blue circles). The right plot concerns $|(J_{60} - wi\,I)^{-1}|_{:,2}$ (red "+"), the bound in (4.1) (blue circles) and the bound in (4.3) (black ×). Once again the bound in (4.3) correctly captures the actual decay. The maximum reached value for s in bound (4.3) is 53. The strategy for optimizing the parameter τ in both bounds (4.1) and (4.3) was analogous to the one used for the exponential function in Example 6.1.

Example 6.3 Let A stem from the centered finite difference discretization of the operator $L(u) = -\Delta u+35u_x+35u_y$, on the unit square, with homogeneous Dirichlet boundary conditions. The matrix is non-symmetric and has size 784×784. The left plot in Figure 6.5 reports relevant spectral information associated with A. We computed 50 iterations of RKSM on $A, \boldsymbol{v}$ obtaining the reduced-order matrix J_{50}.

In both plots in Figure 6.4 the color and symbol coding is the same as before, except that the blue circles now correspond to the bound for the non-Hermitian case in Corollary 3.2. On the right-hand side plot, for $f = \exp$ the bound in (4.1) (blue circles) does not perform well, whereas the bound in (4.3) (black ×) is in good agreement with the true slope. For the bound in (4.3) (black × in the right plot), the maximum reached value for s is 20. In all these bounds the set E is chosen as the ellipse in Figure 6.5 delimited by the red thick line.

Example 6.4 With the same data as in Example 6.3, we illustrate the decay behavior for the entries of the solution Y_{50} to the reduced Lyapunov equation (5.2). The right plot of Figure 6.5 displays the values $|(Y_{50})_{k,\ell}|$. The plot shows

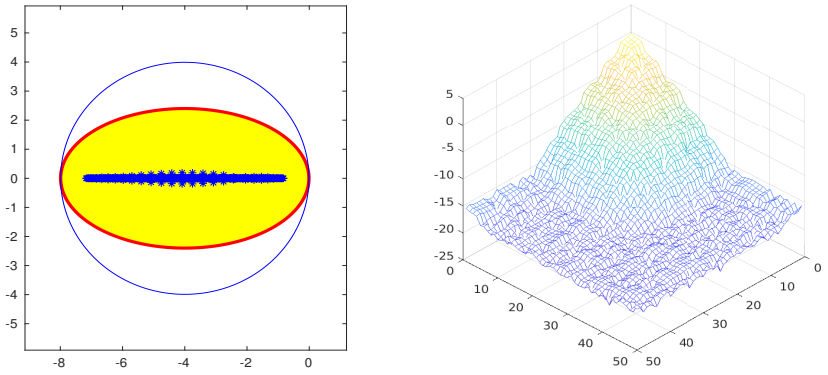

Fig. 6.5 Example 6.4. Left: field of values of A (yellow area), eigenvalues of A (blue stars), ellipse (red thick line) and circle (blue thin line) used in the bounds. Right (logarithmic scale): $|Y_{50}|$, the solution of the reduced-order Lyapunov equation (5.2).

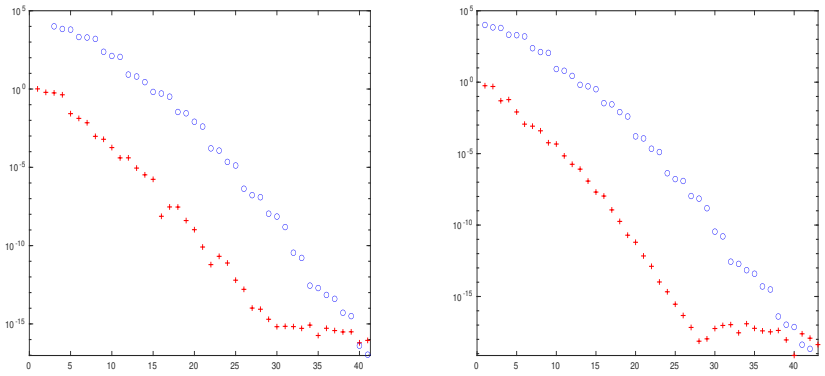

Fig. 6.6 Example 6.4. True values (red "+") and related bound in (5.3) (blue circles). Left: column $|(Y_{50})_{:,3}|$. Right: diag($|Y_{50}|$). Logarithmic scale is used.

that the magnitude of the elements in Y_{50} exponentially decays as we move away from the $(1,1)$ element. In the left plot of Figure 6.6, the values of $|(Y_{50})_{:,3}|$ (red "+") are reported, together with the bound (5.3) (blue circles), where the set E is delimited by the blue thin circumference in Figure 6.5. The right plot of Figure 6.6 reports the diagonal elements of Y_{50} and the corresponding bound, with similar behavior. After their 30th component, the values in the diagonal and in the 3rd column have magnitude below machine precision. The corresponding bound values reach machine precision around their 40th component.

7 Proofs

In this section we will prove Proposition 2.1 from which we will derive Theorem 4.2, Corollary 3.2, and Theorem 5.1 by using a rational approximation approach similar to the one in [12,28]. Note that Theorem 3.1 is derived as a special case of Theorem 4.2. In our analysis we use the Faber-Dzhrbashyan (FD) rational functions introduced in [15]; see also [44, Ch. XIII, section 3] and the references therein. For FD rational approximation, it will be more natural to consider the rational Krylov subspaces defined as

$$\widetilde{\mathcal{K}}_m(A,\widetilde{\boldsymbol{v}},\widetilde{\boldsymbol{\sigma}}_{m-1}) := \mathrm{span}\left\{(A-\sigma_0 I)^{-1}\,\widetilde{\boldsymbol{v}}, (A-\sigma_0 I)^{-1}(A-\sigma_1 I)^{-1}\,\widetilde{\boldsymbol{v}}, \ldots, \prod_{j=0}^{m-1}(A-\sigma_j I)^{-1}\,\widetilde{\boldsymbol{v}}\right\},$$

with $\widetilde{\boldsymbol{\sigma}}_{m-1} = [\sigma_0, \ldots, \sigma_{m-1}]$. Note that

$$\widetilde{\mathcal{K}}_m(A,\widetilde{\boldsymbol{v}},\widetilde{\boldsymbol{\sigma}}_{m-1}) = (A-\sigma_0)^{-1}\mathcal{K}_m(A,\widetilde{\boldsymbol{v}},\boldsymbol{\sigma}_{m-1}).$$

Both rational Krylov subspaces $\mathcal{K}_m$ and $\widetilde{\mathcal{K}}_m$ are commonly used in the literature; see, e.g., [38,21] for the former, and [13,12] for the latter. See [20] for a discussion on both approaches. It is interesting to observe that the Arnoldi-type recurrence associated with $\widetilde{\mathcal{K}}_m(A,\widetilde{\boldsymbol{v}},\widetilde{\boldsymbol{\sigma}}_{m-1})$ generates a set of orthogonal rational functions [9].

Let $\{\widetilde{\boldsymbol{v}}_1,\ldots,\widetilde{\boldsymbol{v}}_m\}$ be the orthonormal basis for $\widetilde{\mathcal{K}}_m(A,\widetilde{\boldsymbol{v}},\widetilde{\boldsymbol{\sigma}}_{m-1})$ obtained by an Arnoldi type process. Each basis vector can be expressed by the matrix rational function

$$\widetilde{\boldsymbol{v}}_j = r_{j-1}(A)\,\widetilde{\boldsymbol{v}}, \quad j = 1,\ldots,m, \tag{7.1}$$

where

$$r_{j-1}(x) = \frac{p_{j-1}(x)}{(x-\sigma_0)\cdots(x-\sigma_{j-1})}, \quad j = 1,\ldots,m, \tag{7.2}$$

with $p_{j-1}(x)$ a polynomial of degree at most $j-1$. Moreover, for $\widetilde{V}_m = [\widetilde{\boldsymbol{v}}_1,\ldots,\widetilde{\boldsymbol{v}}_m]$, the orthonormal projection and restriction of A onto the subspace $\widetilde{\mathcal{K}}_m(A,\widetilde{\boldsymbol{v}},\widetilde{\boldsymbol{\sigma}}_{m-1})$ is defined as

$$\widetilde{J}_m = \widetilde{V}_m^* A \widetilde{V}_m. \tag{7.3}$$

Remark 7.1 Let the columns of V_m form the Arnoldi-based orthonormal basis of the rational Krylov subspace $\mathcal{K}_m(A,\boldsymbol{v},\boldsymbol{\sigma}_{m-1})$ defined in (2.3), with $\boldsymbol{\sigma}_{m-1} = [\sigma_1,\ldots,\sigma_{m-1}]$, and let $J_m = V_m^* A V_m$. If $\boldsymbol{v} = (A-\sigma_0 I)^{-1}\widetilde{\boldsymbol{v}}$, then $\mathcal{K}_m(A,\boldsymbol{v},\boldsymbol{\sigma}_{m-1}) = \widetilde{\mathcal{K}}_m(A,\widetilde{\boldsymbol{v}},\widetilde{\boldsymbol{\sigma}}_{m-1})$. Therefore $\widetilde{V}_m = V_m$ and $\widetilde{J}_m = J_m$.

We also remark that structural properties have already been identified for $\widetilde{J}_m$. Indeed, for A real symmetric, and under certain conditions on the shifts, the $N\times N$ matrix $M = \widetilde{J}_N(I - D_N\widetilde{J}_N)^{-1}$ is symmetric tridiagonal with positive subdiagonals, where D_N is the diagonal matrix containing the shift reciprocals [9]. In fact, this property also holds for $m < N$, that is $\widetilde{J}_m(I - D_m\widetilde{J}_m)^{-1}$ is

tridiagonal. This follows from Proposition 2.2, noticing that $\widetilde{J}_m = J_m$ and $D_m = \widehat{S}_m^{-1}$ having chosen the first diagonal element of $\widehat{S}_m$ to be equal to σ_0. Interestingly, the property discussed in [9] stems from a short-term recurrence determined in [9] to generate an orthonormal basis for the rational Krylov subspace $\widetilde{\mathcal{K}}_m(A, \widetilde{\boldsymbol{v}}, \widetilde{\boldsymbol{\sigma}}_{m-1})$, which was later named Rational-Lanczos iteration in [20, Alg.2].

Let E be a convex continuum, ϕ be the related conformal map as in (3.1), and ψ be its inverse. Let D be the closed unit disk. For every function f continuous on ∂D and analytic in the interior of D the *Faber transformation* is defined as

$$\mathcal{F}(f)(w) = \frac{1}{2\pi i} \int_{|z|=1} f(z) \frac{\psi'(z)}{\psi(z) - w} \mathrm{d}z, \quad w \in G,$$

with G the interior of E; see, e.g., [17, Ch.I, sec. 6], [18], and [19]. Let us set the points $\theta_0, \theta_1, \ldots$, with $|\theta_j| > 1$ in the complex plane, and let us define the Takenaka-Malmquist system of functions

$$\begin{aligned}
\varphi_0(z) &= \frac{\sqrt{|\theta_0|^2 - 1}}{\theta_0 - z}, \\
\varphi_j(z) &= \frac{\sqrt{|\theta_j|^2 - 1}}{\theta_j - z} \prod_{k=0}^{j-1} \frac{1 - \overline{\theta}_k z}{\theta_k - z} \frac{\overline{\theta}_k}{|\theta_k|}, \quad j = 1, 2, \ldots ;
\end{aligned}$$

see [45,33], [46, sec.9.1], and [44, Ch.13,sec.3]. As noticed, e.g., in [12,28], the Faber-Dzhrbashyan rational functions $M_0, M_1, \ldots$ can be defined as the Faber transformation of a Takenaka-Malmquist system, i.e.,

$$M_j(w) = \mathcal{F}(\varphi_j)(w), \quad w \in G, \quad j = 0, 1, 2, \ldots .$$

The Faber transformation maps every rational function r to a rational function with numerator and denominator of the same degrees of the original function r and whose poles are the image of the original poles by ψ (see, e.g., [17, pp. 49–50]). Therefore

$$M_j(w) = \frac{q_j(w)}{(w - \sigma_0) \cdots (w - \sigma_j)}, \tag{7.4}$$

with $\sigma_t = \psi(\theta_t)$, and $q_j(w)$ a polynomial of degree at most j.

Consider the modified Faber operator $\mathcal{F}_+$ defined as

$$\mathcal{F}_+(g)(w) := \mathcal{F}(g)(w) + g(0).$$

If the matrix A satisfies $W(A) \subset E$, for every g analytic in the interior of D and continuous on ∂D we get

$$\|\mathcal{F}_+(g)(A)\| \leq 2 \sup_{z \in D} |g(z)|,$$

as explicitly proved by Beckermann and Reichel in [3, Theorem 2.1][2]. Since $M_j(A) = \mathcal{F}_+(\varphi_j)(A) - \varphi_j(0)$, we can now bound the matrix FD rational functions by

$$\|M_j(A)\| \leq 2\left(\sup_{z\in D}|\varphi_j(z)|\right) + |\varphi_j(0)| \leq 3, \quad j = 0, 1, \ldots . \tag{7.5}$$

Under the assumption that

$$\sum_{j=0}^{\infty} 1 - |\theta_j|^{-1} = \infty, \tag{7.6}$$

FD rational functions satisfy the relation

$$\frac{\psi'(z)}{\psi(z) - w} = \frac{1}{z}\sum_{j=0}^{\infty} \overline{\varphi_j\left(\frac{1}{\bar{z}}\right)} M_j(w), \quad w \in G, |z| > 1; \tag{7.7}$$

see [16] and [44, p. 259].

Lemma 7.2 *Given the points $\theta_0, \ldots, \theta_j$, with $|\theta_*| = \min\{|\theta_0|, \ldots, |\theta_j|\} > 1$, then for $\tau \geq 1$ it holds*

$$\max_{|z|=\tau}\left|\varphi_j\left(\frac{1}{\bar{z}}\right)\right| \leq \sqrt{\frac{|\theta_j|+1}{|\theta_j|-1}} \prod_{t=0}^{j-1} \frac{\tau + |\theta_t|}{|\theta_t|\tau + 1}, \tag{7.8}$$

$$\leq \sqrt{\frac{|\theta_*|+1}{|\theta_*|-1}} \left(\frac{\tau + |\theta_*|}{|\theta_*|\tau + 1}\right)^j. \tag{7.9}$$

Proof Let us assume that for $t \in \{0, \ldots, j-1\}$ and $\tau \geq 1$ it holds

$$\max_{|z|=\tau}\left|\frac{z - \theta_t}{\overline{\theta}_t z - 1}\right| = \frac{\tau + |\theta_t|}{|\theta_t|\tau + 1}. \tag{7.10}$$

In case (7.10) holds, then it also follows that

$$\max_{|z|=\tau}\left|\varphi_j\left(\frac{1}{\bar{z}}\right)\right| \leq \frac{\tau\sqrt{|\theta_j|^2-1}}{|\theta_j|\tau - 1} \prod_{t=0}^{j-1} \frac{\tau + |\theta_t|}{|\theta_t|\tau + 1}.$$

The bound in (7.8) follows observing that

$$\frac{\tau\sqrt{|\theta_j|^2-1}}{|\theta_j|\tau - 1} = \frac{\sqrt{|\theta_j|^2-1}}{|\theta_j| - 1/\tau} \leq \frac{\sqrt{|\theta_j|^2-1}}{|\theta_j| - 1} = \sqrt{\frac{|\theta_j|+1}{|\theta_j|-1}}.$$

Since $\tau > 1$, $(\tau + |\theta|)/(|\theta|\tau + 1)$ is decreasing for $|\theta| > 1$. Hence, we get (7.9).

[2] As remarked in [3], the bound $\|\mathcal{F}_+\| \leq 2$ was implicitly given in [30]; moreover, the same bound stands for the operator $\mathcal{F}_-(g)(w) := \mathcal{F}(g)(w) - g(0)$ and can be obtained, e.g., modifying the proof of Theorem 2 in [17, p. 49].

To conclude the proof, we show that (7.10) indeed holds. Consider the polar coordinates $z = \tau\exp(i\alpha)$, $\theta_t = \rho\exp(i\beta)$, with $\tau, \rho > 1$. Then

$$\left|\frac{z-\theta_t}{\bar\theta_t z - 1}\right|^2 = \left|\exp(i\beta)\frac{\tau\exp(i(\alpha-\beta)) - \rho}{\tau\rho\exp(i(\alpha-\beta)) - 1}\right|^2 = \frac{\tau^2+\rho^2-2\tau\rho\cos(\alpha-\beta)}{\tau^2\rho^2+1-2\tau\rho\cos(\alpha-\beta)} =: R(\alpha).$$

The derivative of $R(\alpha)$ is

$$R'(\alpha) = \frac{2\tau\rho\sin(\alpha-\beta)(\rho^2-1)(\tau^2-1)}{(\tau^2\rho^2+1-2\tau\rho\cos(\alpha-\beta))^2}.$$

Therefore $R'(\alpha)$ is nonnegative for $\alpha \in [\beta, \beta+\pi]$ and negative for $\alpha \in (\beta+\pi, \beta+2\pi)$. Hence its maximum is reached at $\alpha = \beta+\pi$, giving (7.10). □

Note that $(\tau+|\theta|)/(|\theta|\tau+1) < 1$ for every $\tau, |\theta| > 1$. Moreover, the bound (7.5) implies $|M_j(w)| \le 3$ for $w \in G$. Hence by (7.9) the series in (7.7) is uniformly convergent. In our framework, the assumptions (7.6) and $|\theta_*| > 1$ are not restrictive. Indeed, our analysis considers only finite sequences of poles $\theta_0, \dots, \theta_{m-1}$ which can be completed by suitable nodes $\theta_m, \theta_{m+1}, \dots$.

Let E be a convex compact set with interior G. By the Definition 4.1 and by (7.7) for every matrix A with $W(A) \subset G$ it holds

$$\psi'(z)(\psi(z)I - A)^{-1} = \frac{1}{z}\sum_{j=0}^{\infty}\overline{\varphi_j\left(\frac{1}{\bar z}\right)}M_j(A), \quad |z| > 1, \tag{7.11}$$

which is also uniformly convergent. For $\tau > 1$, define the set $E_\tau = E \cup \{w \in \mathbb{C}\setminus E : |\phi(w)| \le \tau\}$ with boundary Γ_τ and assume f analytic in E_τ, then

$$\begin{aligned} f(A) &= \frac{1}{2\pi i}\int_{\Gamma_\tau} f(\eta)(\eta I - A)^{-1}\,\mathrm{d}\eta = \frac{1}{2\pi i}\int_{|z|=\tau} f(\psi(z))\psi'(z)(\psi(z)I-A)^{-1}\,\mathrm{d}z \\ &= \frac{1}{2\pi i}\sum_{j=0}^{\infty} M_j(A)\int_{|z|=\tau}\frac{f(\psi(z))}{z}\overline{\varphi_j\left(\frac{1}{\bar z}\right)}\,\mathrm{d}z, \end{aligned}$$

yielding the matrix function expansion

$$f(A) = \sum_{j=0}^{\infty}\alpha_j M_j(A), \quad \alpha_j = \frac{1}{2\pi i}\int_{|z|=\tau}\frac{f(\psi(z))}{z}\overline{\varphi_j\left(\frac{1}{\bar z}\right)}\,\mathrm{d}z. \tag{7.12}$$

7.1 Proof of Proposition 2.1

Let $\widetilde{\boldsymbol{v}}$ be such that $\boldsymbol{v} = (A-\sigma_0 I)^{-1}\widetilde{\boldsymbol{v}}$. We first prove Proposition 2.1 for the matrix $\widetilde{J}_m$ in (7.3). By (7.1) we get

$$s_j^{(t)}(A)\,\widetilde{\boldsymbol{v}}_\ell = s_j^{(t)}(A)\,r_{\ell-1}(A)\,\widetilde{\boldsymbol{v}}. \tag{7.13}$$

Since $\ell \le t$, by Lemma 3.1 in [13] it holds

$$s_j^{(t)}(A)\widetilde{\boldsymbol{v}}_\ell = \widetilde{V}_m\, s_j^{(t)}(\widetilde{J}_m)\, r_{\ell-1}(\widetilde{J}_m)\widetilde{V}_m^*\, \widetilde{\boldsymbol{v}}, \quad j \le m-t.$$

Combining Lemma 3.1 in [13] and (7.1) gives

$$r_{\ell-1}(\widetilde{J}_m)\widetilde{V}_m^*\, \widetilde{\boldsymbol{v}} = \widetilde{V}_m^* r_{\ell-1}(A)\, \widetilde{\boldsymbol{v}} = \widetilde{V}_m^*\, \widetilde{\boldsymbol{v}}_\ell = \boldsymbol{e}_\ell.$$

Therefore we obtain

$$\widetilde{\boldsymbol{v}}_k^*\, s_j^{(t)}(A)\, \widetilde{\boldsymbol{v}}_\ell = \boldsymbol{e}_k^T\, s_j^{(t)}(\widetilde{J}_m)\, \boldsymbol{e}_\ell.$$

By (7.13), $s_j^{(t)}(A)\widetilde{\boldsymbol{v}}_\ell$ is in $\widetilde{\mathcal{K}}_{j+t}(A, \widetilde{\boldsymbol{v}})$, while $\widetilde{\boldsymbol{v}}_k \perp \widetilde{\mathcal{K}}_{j+t}(A, \widetilde{\boldsymbol{v}})$ for $k \ge j+t+1$. This orthogonality allows us to conclude, as

$$0 = \widetilde{\boldsymbol{v}}_k^*\, s_j^{(t)}(A)\, \widetilde{\boldsymbol{v}}_\ell = \boldsymbol{e}_k^T\, s_j^{(t)}(\widetilde{J}_m)\, \boldsymbol{e}_\ell, \quad j \le k-t-1. \tag{7.14}$$

Finally, for our choice of $\widetilde{\boldsymbol{v}}$, Remark 7.1 ensures that $\widetilde{J}_m = J_m$. □

7.2 Proof of Theorem 4.2

For fixed values of the indexes k, ℓ with $k-\ell-2 \ge 0$, let $\{M_j^{(k,\ell)}(w)\}_{j=0,1,\dots}$ be the sequence of FD rational functions defined by ϕ and by the sequence of parameters

$$\theta_j = \phi(\sigma_{j+\ell}), \quad \text{for } j = 0, \dots, k-\ell-2, \quad \text{and} \quad \theta_j = +\infty, \quad \text{for } j \ge k-\ell-1.$$

Then

$$M_j^{(k,\ell)}(w) = \frac{q_j(w)}{(w-\sigma_\ell)\cdots(w-\sigma_{\ell+j})}, \quad j = 0, \dots, k-\ell-2, \tag{7.15}$$

with $q_j(w)$ a polynomial of degree at most j. Let $\widetilde{\boldsymbol{v}}$ be such that $\boldsymbol{v} = (A-\sigma_0 I)^{-1}\widetilde{\boldsymbol{v}}$. Once again, with this choice of $\widetilde{\boldsymbol{v}}$ and using Remark 7.1, it holds that $J_m = \widetilde{J}_m$ so that we can prove the result for $\widetilde{J}_m$.

Similarly to (7.12), we have the expansion $f(\widetilde{J}_m) = \sum_{j=0}^{\infty} \alpha_j M_j^{(k,\ell)}(\widetilde{J}_m)$, which combined with Proposition 2.1 gives[3]

$$f(\widetilde{J}_m)_{k,\ell} = \sum_{j=0}^{+\infty} \alpha_j\, \boldsymbol{e}_k^T\, M_j^{(k,\ell)}(\widetilde{J}_m)\, \boldsymbol{e}_\ell = \sum_{j=k-\ell-1}^{+\infty} \alpha_j\, \boldsymbol{e}_k^T\, M_j^{(k,\ell)}(\widetilde{J}_m)\, \boldsymbol{e}_\ell.$$

Since $W(\widetilde{J}_m) \subseteq W(A) \subset E$, the bound (7.5) gives

$$\left| f(\widetilde{J}_m)_{k,\ell} \right| \le \|\boldsymbol{e}_k\|\, \|\boldsymbol{e}_\ell\| \sum_{j=k-\ell-1}^{+\infty} |\alpha_j| \|M_j^{(k,\ell)}(\widetilde{J}_m)\| \le 3 \sum_{j=k-\ell-1}^{+\infty} |\alpha_j|.$$

[3] Note that in section 7.1 we proved that Proposition 2.1 holds also for $\widetilde{J}_m$.

Since f is analytic in E_τ for $\tau > 1$, the coefficients α_j are given as in (7.12) and are bounded by

$$|\alpha_j| \le \frac{1}{2\pi} \left| \int_{|z|=\tau} \frac{f(\psi(z))}{z} \overline{\varphi_j\left(\frac{1}{\overline{z}}\right)} \, \mathrm{d}z \right| \le \max_{|z|=\tau} |f(\psi(z))| \max_{|z|=\tau} \left| \varphi_j\left(\frac{1}{\overline{z}}\right) \right|.$$

For $t \ge k-\ell-1$ we have $\theta_t = +\infty$ and hence

$$\sqrt{\frac{|\theta_t|+1}{|\theta_t|-1}} = 1, \quad \text{and} \quad \frac{\tau+|\theta_t|}{|\theta_t|\tau+1} = \frac{1}{\tau}.$$

Hence by (7.8) we get

$$\begin{aligned}
\left| f(\widetilde{J}_m)_{k,\ell} \right| &\le 3 \sum_{j=k-\ell-1}^{+\infty} |\alpha_j| \le 3 \max_{|z|=\tau} |f(\psi(z))| \sum_{j=k-\ell-1}^{+\infty} \sqrt{\frac{|\theta_j|+1}{|\theta_j|-1}} \prod_{t=0}^{j-1} \frac{\tau+|\theta_t|}{|\theta_t|\tau+1} \\
&\le 3 \max_{|z|=\tau} |f(\psi(z))| \prod_{t=0}^{k-\ell-2} \frac{\tau+|\theta_t|}{|\theta_t|\tau+1} \sum_{j=0}^{+\infty} \left(\frac{1}{\tau}\right)^j \\
&\le 3 \frac{\tau}{\tau-1} \max_{|z|=\tau} |f(\psi(z))| \prod_{t=0}^{k-\ell-2} \frac{\tau+|\theta_t|}{|\theta_t|\tau+1}. \qquad (7.16)
\end{aligned}$$

We recall that $\theta_t = \phi(\sigma_{t+\ell})$ for $t = 0, \dots, k-\ell-2$. Hence we can redefine the index t in the product to go from ℓ to $k-2$, thus establishing the first bound.

The refined bound in (4.3) is obtained as follows.

$$\begin{aligned}
\left| f(\widetilde{J}_m)_{k,\ell} \right| &\le 3 \sum_{j=k-\ell-1}^{k-\ell+s-2} |\alpha_j| + 3 \sum_{j=k-\ell+s-1}^{+\infty} |\alpha_j| \\
&\le 3 \sum_{j=k-\ell-1}^{k-\ell+s-2} |\alpha_j| + 3 \max_{|z|=\tau} |f(\psi(z))| \prod_{t=0}^{k-\ell-2} \frac{\tau+|\theta_t|}{|\theta_t|\tau+1} \sum_{j=0}^{+\infty} \left(\frac{1}{\tau}\right)^{j+s} \\
&\le 3 \sum_{j=k-\ell-1}^{k-\ell+s-2} |\alpha_j| + \frac{B(k,\ell)}{\tau^s}.
\end{aligned}$$

Redefining j in the summation to start from zero, the refined bound follows. □

7.3 Proof of Corollary 3.2

We consider a compact set E containing $(c-a, c+a)$ whose boundary is the ellipse with center c, semi-minor axis r, and distance between the foci and the center $\rho = \sqrt{a^2 - r^2}$. Setting $R = (a+r)/\rho$, we can associate to E the conformal map

$$\phi(w) = \frac{w - c + \sqrt{(w-c)^2 - \rho^2}}{\rho R}, \qquad (7.17)$$

with inverse

$$\psi(z) = \frac{\rho}{2}\left(Rz + \frac{1}{Rz}\right) + c. \tag{7.18}$$

see, e.g., [44, Ch.II, Ex.3]. Notice that for every $\tau > 1$ and $r = 0$ we get

$$\max_{|z|\le\tau} |(\psi(z))| \le \frac{a}{2}(\tau + 1) + |c| \le a\tau + |c|.$$

The proof is concluded using Theorem 3.1, and letting $r \to 0$ in the Hermitian case. □

7.4 Proof of Theorem 5.1

Let $\phi(w) = (w - c)/a$ be the conformal map for E, and ψ be its inverse. The solution $\widetilde{Y}_m$ of the Lyapunov equation

$$\widetilde{J}_m\widetilde{Y}_m + \widetilde{Y}_m\widetilde{J}_m^* + \boldsymbol{e}_1\boldsymbol{e}_1^T = 0, \tag{7.19}$$

can be represented as

$$\widetilde{Y}_m = \frac{1}{2\pi}\int_{\Gamma_\tau} (wI - \widetilde{J}_m)^{-1}\boldsymbol{e}_1\,\boldsymbol{e}_1^T(wI + \widetilde{J}_m^*)^{-1}\,\mathrm{d}w, \tag{7.20}$$

where $\Gamma_\tau = \{\psi(z) : |z| = \tau\} \subset \mathbb{C}_-$; see, e.g., [32, Eq. (26)], [28]. Consider a family of sequences of FD rational functions $M_t^{(j)}(w) = \mathcal{F}(\varphi_t^{(j)})(w)$, $s \ge 0$, $w \in G$ (the interior of E), defined by ϕ and by the sequence of parameters[4]

$$\begin{aligned} &\theta_s^{(j)} = \phi(\sigma_{s+1}), \quad \text{for } s = 0, \dots, j-3, \\ &\theta_s^{(j)} = +\infty, \qquad \text{for } s \ge j-2. \end{aligned} \tag{7.21}$$

From (7.11) we obtain

$$(wI - \widetilde{J}_m)^{-1} = \frac{1}{\psi'(\phi(w))}\sum_{s=0}^{\infty}\frac{1}{\phi(w)}\overline{\varphi_s^{(k)}\left(\frac{1}{\overline{\phi(w)}}\right)}M_s^{(k)}(\widetilde{J}_m), \quad w \notin E,$$

$$(wI + \widetilde{J}_m^*)^{-1} = \frac{-1}{\psi'(\phi(-w))}\sum_{t=0}^{\infty}\frac{1}{\phi(-w)}\overline{\varphi_t^{(\ell)}\left(\frac{1}{\overline{\phi(-w)}}\right)}M_t^{(\ell)}(\widetilde{J}_m^*), \quad w \notin -E;$$

see also (3.7) and (3.8) in [28]. Using the expressions above in (7.20) gives

$$\left(\widetilde{Y}_m\right)_{k,\ell} = \sum_{s=0}^{\infty}\sum_{t=0}^{\infty}\alpha_{s,t}\left(M_s^{(k)}(\widetilde{J}_m)\right)_{k,1}\left(M_t^{(\ell)}(\widetilde{J}_m^*)\right)_{1,\ell},$$

[4] The index $j = 3, 4, \dots$ parametrizes the family of sequences of FD rational functions $\{M_t^{(j)}\}_{t=0,\dots}$. All sequences in such family share the initial (finite) shifts $\phi(\sigma_{s+1})$ up to $j-3$. This means that given the indexes $i < j$, it holds that $M_t^{(i)} = M_t^{(j)}$ for $t = 0, \dots, i-3$.

with

$$\alpha_{s,t}=\frac{1}{2\pi}\int_{\Gamma_\tau}\frac{1}{\phi(w)\psi'(\phi(w))}\overline{\varphi_s^{(k)}\left(\frac{1}{\overline{\phi(w)}}\right)}\frac{1}{\phi(-w)\psi'(\phi(-w))}\overline{\varphi_t^{(\ell)}\left(\frac{1}{\overline{\phi(-w)}}\right)}\,\mathrm{d}w.$$

By the change of variable $z=\phi(w)$, we get

$$\alpha_{s,t}=\frac{1}{2\pi}\int_{|z|=\tau}\overline{\varphi_s^{(k)}\left(\frac{1}{\overline{z}}\right)}\frac{1}{z\,\phi(-\psi(z))\psi'(\phi(-\psi(z)))}\overline{\varphi_t^{(\ell)}\left(\frac{1}{\overline{\phi(-\psi(z))}}\right)}\,\mathrm{d}z;$$

cf. (3.11) and (3.12) in [28]. Then Proposition 2.1 and (7.5) give

$$\left|\left(\widetilde{Y}_m\right)_{k,\ell}\right|\le 9\sum_{s=k-2}^{\infty}\sum_{t=\ell-2}^{\infty}|\alpha_{s,t}|. \tag{7.22}$$

We notice that for our choice of ϕ it holds that $\frac{1}{\overline{\phi(-\psi(z))}}=-a/(\overline{z}a+2c)$. For $t\ge\ell-2$ and $|z|=\tau$, we substitute this quantity into the definition of the φ_t's and we recall (7.21), yielding

$$\left|\varphi_t^{(\ell)}\left(\frac{1}{\overline{\phi(-\psi(z))}}\right)\right|\le\left|\frac{a}{az+2c}\right|^{t-\ell+2}\prod_{j=0}^{\ell-3}\left|\frac{az+\theta_j^{(\ell)}a+2c}{\theta_j^{(\ell)}(az+2c)+a}\right|.$$

Since $\theta_j^{(\ell)}\in\mathbb{R}$, it holds that for $z=\tau\exp(i\alpha)$, $\alpha\in[0,2\pi]$,

$$\left|\frac{az+\theta_j^{(\ell)}a+2c}{\theta_j^{(\ell)}(az+2c)+a}\right|\le\max_{z\in\{\tau,-\tau\}}\left|\frac{az+\theta_j^{(\ell)}a+2c}{\theta_j^{(\ell)}(az+2c)+a}\right|=:B_j^{(\ell)}(\tau). \tag{7.23}$$

Indeed, for $z=\tau\exp(i\alpha)$, $\alpha\in[0,2\pi]$, consider the general rational function

$$\chi(\alpha):=\left|\frac{a_1\tau\exp(i\alpha)+a_2}{b_1\tau\exp(i\alpha)+b_2}\right|^2=\frac{a_1^2\tau^2+a_2^2+2a_1a_2\tau\cos(\alpha)}{b_1^2\tau^2+b_2^2+2b_1b_2\tau\cos(\alpha)},$$

with $a_1,a_2,b_1,b_2\in\mathbb{R}$ collecting all other real terms. Then its derivative is

$$\chi'(\alpha)=\frac{2\tau\sin(\alpha)(a_1^2b_1b_2\tau^2+a_2^2b_1b_2-a_1a_2b_1^2\tau^2-a_1a_2b_2^2)}{(b_1^2\tau^2+b_2^2+2b_1b_2\tau\cos(\alpha))^2}.$$

Hence the critical points of $\chi(\alpha)$ are $\alpha=0,\pi$, and $\chi(\alpha)\le\max\{\chi(0),\chi(\pi)\}$ from which the bound (7.23) follows.

Using (7.8) and assuming $s\ge k-2$, $t\ge\ell-2$, we get

$$|\alpha_{s,t}|\le\left|\frac{1}{a\tau+2c}\right|\prod_{j=0}^{k-3}\frac{\tau+|\theta_j^{(k)}|}{|\theta_j^{(k)}|\tau+1}\left(\frac{1}{\tau}\right)^{s-k+2}\prod_{j=0}^{\ell-3}B_j^{(\ell)}(\tau)\left|\frac{a}{a\tau+2c}\right|^{t-\ell+2}.$$

Inequality (7.22) then gives

$$\left|\left(\widetilde{Y}_m\right)_{k,\ell}\right| \le \left|\frac{9}{a\tau+2c}\right| \prod_{j=0}^{k-3} \frac{\tau+|\theta_j^{(k)}|}{|\theta_j^{(k)}|\tau+1} \prod_{j=0}^{\ell-3} B_j^{(\ell)}(\tau) \sum_{s=0}^{\infty} \left(\frac{1}{\tau}\right)^s \sum_{t=0}^{\infty} \left|\frac{a}{a\tau+2c}\right|^t$$
$$\le \left|\frac{9}{a\tau+2c}\right| \frac{\tau}{\tau-1} \frac{|a\tau+2c|}{|a\tau+2c|-a} \prod_{j=0}^{k-3} \frac{\tau+|\theta_j^{(k)}|}{|\theta_j^{(k)}|\tau+1} \prod_{j=0}^{\ell-3} B_j^{(\ell)}(\tau).$$

To conclude, we need to show that the same result holds for Y_m. Let $\widetilde{\boldsymbol{v}}$ be such that $\boldsymbol{v} = (A-\sigma_0 I)^{-1}\widetilde{\boldsymbol{v}}$. From Remark 7.1, it holds that $\widetilde{J}_m = J_m$, so that $\widetilde{Y}_m = Y_m$ for the solution uniqueness of the Lyapunov equation. □

8 Conclusion

The (classical) Arnoldi method produces an orthonormal basis U_n for the Krylov subspace $\mathcal{P}_m(A,\boldsymbol{v}) = \operatorname{span}\{\boldsymbol{v}, A\boldsymbol{v}, \ldots, A^{m-1}\boldsymbol{v}\}$. In this framework, the matrix $T_m = U_m^* A U_m$ is an upper-Hessenberg matrix (tridiagonal for A Hermitian). In the rational case, V_m is the orthonormal basis for the rational Krylov subspace $\mathcal{K}_m(A,\boldsymbol{v},\sigma)$ obtained by RKSM. By Theorem 3.1, the matrix $J_m = V_m^* A V_m$ has a decay in the values of its elements away from the first sub-diagonal, i.e., exactly where the elements of T_m would be zero. Moreover, the proof in section 7 shows that the decay property of J_m is related to the fact that rational functions of J_m in the rational Krylov subspace can be projected back to subspaces of $\mathcal{K}_m(A,\boldsymbol{v},\boldsymbol{\sigma}_{m-1})$, thus ensuring orthogonality with respect to later basis vectors. This property is a generalization of the orthogonality property behind the sparsity pattern of T_m. This hidden structure has allowed us to prove decay properties for $f(J_m)$ which are typical of banded matrices, without J_m being banded.

Acknowledgments

The authors would like to thank the referees for the helpful comments and improvements suggested. The authors are members of Indam-GNCS, which support is gratefully acknowledged. This work has also been supported by Charles University Research program No. UNCE/SCI/023.

References

1. A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*, vol. 6 of Advances in Design and Control, SIAM, Philadelphia, PA, 2005. With a foreword by Jan C. Willems.
2. B. Beckermann, *An Error Analysis for Rational Galerkin Projection applied to the Sylvester Equation*, SIAM J. Numerical Analysis, 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. P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, eds., *Model reduction and approximation: theory and Algorithms*, Computational Science & Engineering, SIAM, PA, 2017.
5. P. Benner, V. Mehrmann, and D. Sorensen, eds., *Dimension Reduction of Large-Scale Systems*, Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin/Heidelberg, 2005.
6. M. Benzi and N. Razouk, *Decay bounds and $O(n)$ algorithms for approximating functions of sparse matrices*, ETNA, 28 (2007), pp. 16–39.
7. M. Benzi and V. Simoncini, *Decay bounds for functions of hermitian matrices with banded or Kronecker structure*, SIAM J. Matrix Analysis and Applications, 36 (2015), pp. 1263–1282.
8. A. Bouras and V. Frayssé, *Inexact matrix-vector products in Krylov methods for solving linear systems: A relaxation strategy*, SIAM Journal on Matrix Analysis and Applications, 26 (2005), pp. 660 – 678.
9. K. Deckers and A. Bultheel, *Rational Krylov sequences and orthogonal rational functions*, tech. rep., Department of Computer Science, K.U.Leuven, 2007.
10. S. Demko, W. F. Moss, and P. W. Smith, *Decay rates for inverses of band matrices*, Math. Comp., 43 (1984), pp. 491–499.
11. V. Druskin and L. Knizhnerman, *Extended Krylov subspaces: approximation of the matrix square root and related functions*, SIAM Journal on Matrix Analysis and Applications, 19 (1998), pp. 755–771.
12. V. Druskin, L. Knizhnerman, and V. Simoncini, *Analysis of the rational Krylov subspace and ADI methods for solving the Lyapunov equation*, SIAM J. Numer. Anal., 49 (2011), pp. 1875–1898.
13. V. Druskin, L. Knizhnerman, and M. Zaslavsky, *Solution of Large Scale Evolutionary Problems Using Rational Krylov Subspaces with Optimized Shifts*, SIAM Journal on Scientific Computing, 31 (2009), pp. 3760–3780.
14. V. Druskin and V. Simoncini, *Adaptive rational Krylov subspaces for large-scale dynamical systems*, Systems Control Lett., 60 (2011), pp. 546–560.
15. M. M. Dzhrbashyan, *On expansion of analytic functions in rational functions with preassigned poles*, Izv. Akad. Nauk Armyan. SSR. Ser. Fiz.-Mat. Nauk, 10 (1957), pp. 21–29.
16. ———, *Expansions in rational functions with fixed poles*, Dokl. Akad. Nauk SSSR, 143 (1962), pp. 17–20.
17. D. Gaier, *Lectures on complex approximation*, Birkhäuser Boston, Inc., Boston, MA, 1987. Translated from the German by Renate McLaughlin.
18. ———, *The Faber operator and its boundedness*, J. Approx. Theory, 101 (1999), pp. 265–277.
19. T. Ganelius, *Degree of rational approximation*, in Lectures on approximation and value distribution, vol. 79 of Sém. Math. Sup., Presses Univ. Montréal, Montreal, Que., 1982, pp. 9–78.
20. S. Güttel, *Rational Krylov Methods for Operator Functions*, PhD thesis, TU Bergakademie Freiberg, Germany, 2010.
21. S. Güttel, *Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection*, GAMM-Mitteilungen, 36 (2013), pp. 8–31.
22. S. Güttel and L. Knizhnerman, *A black-box rational Arnoldi variant for Cauchy–Stieltjes matrix functions*, BIT Numerical Mathematics, 53 (2013), pp. 595–616.
23. S. Güttel and M. Schweitzer, *A comparison of limited-memory Krylov methods for Stieltjes functions of Hermitian matrices*, arXiv:2006.05922, (2020).
24. N. J. Higham, *Functions of Matrices. Theory and Computation*, SIAM, Philadelphia, PA, 2008.
25. R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
26. C. Jagels and L. Reichel, *The extended Krylov subspace method and orthogonal Laurent polynomials*, Linear Algebra Appl., 431 (2009), pp. 441–458.
27. ———, *The structure of matrices in rational Gauss quadrature*, Math. Comp., 82 (2013), pp. 2035–2060.
28. L. Knizhnerman and V. Simoncini, *Convergence analysis of the extended Krylov subspace method for the Lyapunov equation*, Numer. Math., 118 (2011), pp. 567–586.

29. J. G. Korvink and E. B. Rudnyi, *Oberwolfach benchmark collection*, in Dimension Reduction of Large-Scale Systems, P. Benner, D. C. Sorensen, and V. Mehrmann, eds., Berlin, Heidelberg, 2005, Springer Berlin Heidelberg, pp. 311–315.
30. T. Kövari and C. Pommerenke, *On Faber polynomials and Faber expansions*, Math. Z., 99 (1967), pp. 193–206.
31. P. Kürschner and M. Freitag, *Inexact methods for the low rank solution to large scale lyapunov equations*, BIT Numer Math, 60 (2020), pp. 1221–1259.
32. P. Lancaster, *Explicit Solutions of Linear Matrix Equations*, SIAM Review, 12 (1970), pp. 544–566.
33. F. Malmquist, *Sur la détermination d'une classe de fonctions analytiques par leurs valeurs dans un ensemble donné de points*, Comptes Rendus du Sixièmme Congrèss (1925) des mathématiciens scandinaves. Kopenhagen, (1926), pp. 253–259.
34. R. Nabben, *Decay rates of the inverse of nonsymmetric tridiagonal and band matrices*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 820–837.
35. K. H. A. Olsson and A. Ruhe, *Rational Krylov for eigenvalue computation and model order reduction*, BIT Numerical Mathematics, 46 (2006), pp. 99–111.
36. S. Pozza and V. Simoncini, *Inexact Arnoldi residual estimates and decay properties for functions of non-Hermitian matrices*, BIT Numerical Mathematics, 59 (2019), pp. 969–986.
37. A. Ruhe, *Rational Krylov sequence methods for eigenvalue computation*, Lin. Alg. Appl., 58 (1984), pp. 391–405.
38. A. Ruhe, *The rational Krylov algorithm for nonsymmetric eigenvalue problems. III: Complex shifts for real matrices*, BIT Numerical Mathematics, 34 (1994), pp. 165–176.
39. V. Simoncini, *Variable accuracy of matrix-vector products in projection methods for eigencomputation*, SIAM J. Numerical Analysis, 43 (2005), pp. 1155–1174.
40. ———, *The extended Krylov subspace for parameter dependent systems*, Applied Num. Math., 60 (2010), pp. 550–560.
41. ———, *The Lyapunov matrix equation. Matrix analysis from a computational perspective*, in Quaderno UMI - Topics in Mathematics, vol. 55, UMI, 2015, pp. 157–174.
42. ———, *Computational Methods for Linear Matrix Equations*, SIAM Review, 58 (2016), pp. 377–441.
43. V. Simoncini and D. B. Szyld, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477.
44. P. K. Suetin, *Series of Faber polynomials*, Gordon and Breach Science Publishers, 1998. Translated from the 1984 Russian original by E. V. Pankratiev [E. V. Pankrat′ev].
45. S. Takenaka, *On the orthogonal functions and a new formula of interpolation*, Jpn. J. Math., 2 (1925), pp. 129–145.
46. J. L. Walsh, *Interpolation and approximation by rational functions in the complex domain*, Fourth edition. American Mathematical Society Colloquium Publications, Vol. XX, American Mathematical Society, Providence, R.I., 1965.