Electronic Transactions on Numerical Analysis.
Volume 58, pp. 538–567, 2023.

ISSN 1068–9613.
DOI: 10.1553/etna_vol58s538

ETNA
Kent State University **and**
Johann Radon Institute (RICAM)

INEXACT RATIONAL KRYLOV SUBSPACE METHODS FOR APPROXIMATING THE ACTION OF FUNCTIONS OF MATRICES[*]

SHENGJIE XU[†] AND FEI XUE[†]

Abstract. This paper concerns the theory and development of inexact rational Krylov subspace methods for approximating the action of a function of a matrix $f(A)$ to a column vector b. At each step of the rational Krylov subspace methods, a shifted linear system of equations needs to be solved to enlarge the subspace. For large-scale problems, such a linear system is usually solved approximately by an iterative method. The main question is how to relax the accuracy of these linear solves without negatively affecting the convergence of the approximation of $f(A)b$. Our insight into this issue is obtained by exploring the residual bounds for the rational Krylov subspace approximations of $f(A)b$, based on the decaying behavior of the entries in the first column of certain matrices of A restricted to the rational Krylov subspaces. The decay bounds for these entries for both analytic functions and Markov functions can be efficiently and accurately evaluated by appropriate quadrature rules. A heuristic based on these bounds is proposed to relax the tolerances of the linear solves arising in each step of the rational Krylov subspace methods. As the algorithm progresses toward convergence, the linear solves can be performed with increasingly lower accuracy and computational cost. Numerical experiments for large nonsymmetric matrices show the effectiveness of the tolerance relaxation strategy for the inexact linear solves of rational Krylov subspace methods.

Key words. matrix functions, rational Krylov subspace, inexact Arnoldi algorithm, decay bounds

AMS subject classifications. 65D15, 65F10, 65F50, 65F60

1. Introduction. Consider a matrix $A \in \mathbb{R}^{n\times n}$ and a function f that is analytic in a neighborhood of the numerical range of A. This paper studies efficient iterative methods for approximating a matrix function $f(A)$ multiplied by a vector $b \in \mathbb{R}^n$. For large-scale problems, we approximate $f(A)b$ by restricting A to a subspace of dimension m ($m \ll n$) and obtain

$$f(A)b \approx V_m f(A_m) V_m^* b,$$

where $V_m \in \mathbb{R}^{n\times m}$ contains orthonormal basis vectors of the subspace and $A_m = V_m^* A V_m$ is the restriction of A to this subspace. Matrix function problems arise in the numerical solution of differential equations [14, 37, 47], matrix functional integrators [44, 45], model order reduction [2, 31], optimization problems [9, 64], and others.

One of the classical methods of subspace projection for matrix function approximations is the standard Krylov subspace method that generates the subspaces

$$\mathcal{K}_m(A,b) = \operatorname{span}\left\{b, Ab, \ldots, A^{m-1}b\right\};$$

see, e.g., [42, 55]. A few restarted variants were proposed in [1, 26, 27, 32, 33]. Methods based on rational approximations have also been studied, such as the extended Krylov subspace method (EKSM) [22, 46] and the adaptive rational Krylov subspace method (RKSM) [23, 24, 38, 39, 50]. In this paper, we consider a generic RKSM that generates subspaces of the form

$$\mathcal{Q}_m(A,b) = q_{m-1}(A)^{-1}\mathcal{K}_m(A,b),$$

where $q_{m-1}(A)$ is a polynomial of degree not larger than $m-1$ with respect to A.

[*]Received July 5, 2022. Accepted May 6, 2023. Published online on September 12, 2023. Recommended by Stefan Güttel. This research was supported by the U.S. National Science Foundation under grants DMS-2111496 and DMS-1819097.

[†]School of Mathematical and Statistical Sciences, Clemson University, O-110 Martin Hall, Box 340975, Clemson, SC 29634 ({shengjx, fxue}@g.clemson.edu).

The error norm of the approximation, which is the quantity to determine convergence in an ideal stopping criterion, is defined as

$$\|R_m\| = \|f(A)b - V_m f(A_m) V_m^* b\| .$$

However, it is impossible to compute the residual norm directly because $f(A)b$ is unknown. A practical stopping criterion is to monitor the difference between the approximations obtained in two successive iterations. The RKSM can be terminated if such a difference becomes sufficiently small, but this criterion may lead to premature termination if the approximation stagnates without actual convergence to $f(A)b$. A more reliable alternative stopping criterion is to evaluate upper bounds for the residual norm [21, 43, 63], especially for exponential-type functions. There are also some results on a posteriori error bounds [35]. In addition, we may compute $|e_m^* f(A_m)e_1|$ to evaluate the accuracy of the approximation; see, e.g., [11, 21, 46, 59]. In this paper, our stopping criterion is based on the norm of $(AV_m - V_m A_m)f(A_m)V_m^* b$, which can be interpreted as the residual norm of an associated differential equation; see, e.g., [11, 21, 56].

Given a square band matrix B and a sufficiently regular function f, the magnitude of the entries of $f(B)$ below the main diagonal can be characterized by a decaying behavior that depends on the row index relative to the diagonal [4, 6, 7, 52]. A priori estimates of the decay rate have been discussed in [5, 15, 18, 34]. For RKSMs, the restricted matrix A_m is not banded, but upper bounds for the entries of $f(A_m)$ have been derived, which also exhibit a decaying behavior below the main diagonal [57]. In this paper, we further show a similar decaying behavior for the entries of $K_m^{-1} f(A_m)$ below the diagonal, where K_m is an upper Hessenberg matrix generated by the RKSM. The matrix $K_m^{-1} f(A_m)$ is directly related to the residual of the RKSM, and it can be used to determine an a priori tolerance relaxation for the linear solve at each RKSM step to enable an inexact RKSM for approximating $f(A)b$.

Specifically, at each step of the RKSM, we compute a shift-invert matrix-vector product of the form $(A - sI)^{-1}(A - \sigma I)u$, which is equivalent to the solution of the linear system $(A - sI)x = (A - \sigma I)u$. For large-scale problems, these linear systems are solved approximately by iterative methods. Earlier studies on the inexact Krylov methods based on inexact matrix vector products can be found in [13, 61]. Errors are introduced at each RKSM step, and they accumulate in the rational Krylov subspace. The motivation of this paper is to find a strategy to relax the accuracy of the linear solves without negatively impacting the convergence of the RKSM to $f(A)b$. This motivation is the same as that for the study of inexact standard [20, 56] and rational [8, 40] Krylov methods for approximating $f(A)b$ for Hermitian matrices. In particular, in [40] a strategy of relaxing the inner tolerance of the shift-invert Lanczos method or the EKSM is proposed that applies only one fixed pole repeatedly; in [8] an effective preconditioner construction for the iterative solution of linear systems with different shifts in the RKSM are considered, but a relaxation of the tolerance of the inner linear solves is not discussed. Inexact RKSMs has also been used in evolution equations [41], Lyapunov equations [48], eigenvalue problems [49, 66], and model reductions [65]. In this study, we consider RKSMs for matrices regardless of their symmetry and focus on how the tolerances of the inner linear systems with different shifts can be relaxed without delaying the convergence of the RKSM.

The inexact RKSM in our problem setting relies on the association between the upper bounds for the residual norm for approximating $f(A)b$ and the decay bounds for the entries below the main diagonal of $K_m^{-1} f(A_m)$. Such associations can be established for both analytic functions and Markov functions. These results are largely consistent with the relationships between the poles and the convergence of the RKSM for approximating the actions of the exponential function [53] and Markov functions [3]. A tolerance relaxation strategy for the iterative linear solve at each step of the inexact RKSM is derived based on the decay bounds for the entries of $K_m^{-1}f(A_m)$. Compared with the decay bounds for the entries of $f(A_m)$ in [57], our bounds for the entries of $K_m^{-1}f(A_m)$ are sharper as they keep the original integrals, which can be evaluated efficiently by appropriate quadrature rules; more importantly, they directly inform and enable an inexact RKSM in our problem setting.

The rest of the paper is organized as follows. In Section 2, we review the RKSM for approximating $f(A)b$ and derive a sparsity pattern of certain rational functions of matrices restricted to the RKSM subspaces. In Section 3, we introduce the theorems for the decay bounds for the entries of $K_m^{-1}f(A_m)$ below the diagonal. A tolerance relaxation strategy is derived for the inexact RKSM in Section 4 to ensure that the difference between the true and the derived residuals of the inexact method is bounded by a given tolerance. A heuristic tolerance relaxation strategy is proposed for the inexact linear solves arising at each RKSM step. In Section 5, we show numerical results to support the theorems of the decay bounds for the RKSM residual norms and also show the advantage of the inexact method with a heuristic tolerance relaxation strategy over the exact method. In Section 6, our main theorems are proved using the Faber-Dzhrbashyan rational approximations. Conclusions of this paper are given in Section 7.

2. Preliminaries.

In this section, we give some preliminary results to facilitate our later discussion of the inexact RKSM for approximating $f(A)b$.

2.1. A brief review of the RKSM.

The RKSM starts with a vector $b \in \mathbb{R}^n \setminus \{0\}$ to construct rational Krylov subspaces $\mathcal{Q}_m(A, v_1)$, where $A \in \mathbb{R}^{n\times n}$ and $v_1 = b/\|b\|_2$. At step k, the RKSM chooses a pole $s_k \neq 0$ and a zero $\sigma_k \neq s_k$, and applies the linear operator $(I - A/s_k)^{-1}(A - \sigma_k I)$ to the vector v_k, which is the last vector of the orthonormal basis vectors $\{v_1, v_2, \dots, v_k\}$ that span the current subspace $\mathcal{Q}_k(A, v_1)$. To build an orthonormal basis of the enlarged subspace $\mathcal{Q}_{k+1}(A, v_1)$, we adopt the modified Gram-Schmidt orthogonalization and obtain

$$(I - A/s_k)^{-1}(A - \sigma_k I)v_k = \sum_{i=1}^{k+1} h_{ik} v_i. \tag{2.1}$$

Repeat the above operation for each index value $k = 1, 2, \dots, m$, assuming that there is no breakdown. It is not difficult to get the rational Arnoldi relation:

$$AV_m(H_m D_m + I) + \frac{1}{s_m} h_{m+1,m} A v_{m+1} e_m^* = V_m(H_m + P_m) + h_{m+1,m} v_{m+1} e_m^*,$$

or equivalently,

$$AV_{m+1}\underline{K}_m = V_{m+1}\underline{G}_m, \tag{2.2}$$

where $H_m \in \mathbb{R}^{m\times m}$ is upper Hessenberg, $V_{m+1} = [v_1, v_2, \dots, v_{m+1}]$ contains the orthonormal basis vectors of the rational Krylov subspace

$$\mathcal{Q}_{m+1}(A, v_1) = \left(\prod_{k=1}^{m} (A - s_k I)^{-1}\right) \text{span}\left\{v_1, Av_1, A^2 v_1, \dots, A^m v_1\right\}, \tag{2.3}$$

$D_m = \text{diag}(1/s_1, \dots, 1/s_m)$ and $P_m = \text{diag}(\sigma_1, \dots, \sigma_m)$, and $\underline{K}_m, \underline{G}_m \in \mathbb{R}^{(m+1)\times m}$ are both upper Hessenberg matrices:

$$\begin{aligned} \underline{K}_m &= \begin{bmatrix} K_m \\ k_{(m+1)m} e_m^* \end{bmatrix} = \begin{bmatrix} H_m D_m + I \\ \frac{1}{s_m} h_{(m+1)m} e_m^* \end{bmatrix}, \\ \underline{G}_m &= \begin{bmatrix} G_m \\ g_{(m+1)m} e_m^* \end{bmatrix} = \begin{bmatrix} H_m + P_m \\ h_{(m+1)m} e_m^* \end{bmatrix}. \end{aligned} \tag{2.4}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

From (2.2), it follows that

$$A_m = V_m^* A V_m = \left(G_m - \frac{h_{m+1,m}}{s_m} V_m^* A v_{m+1} e_m^* \right) K_m^{-1}. \tag{2.5}$$

An alternative approach to enlarge the rational Krylov subspace is to apply the operator $(I - A/s_k)^{-1}$ to the vector v_k; see, e.g., [24, 57]. It generates exactly the same subspace as in (2.3) if all poles s_k $(1 \leq k \leq m)$ remain the same for both approaches. In this paper, we follow (2.1) to enlarge the subspace because our numerical experience suggests that applying the operator $(I - A/s_k)^{-1}(A - \sigma_k I)$ tends to achieve a smaller final residual norm. This can probably be attributed to an improvement in floating-point accuracy, though we have no additional insight here. The choice of $\sigma_k \neq s_k$ does not impact the generated rational Krylov subspace, and we follow [38] to set $\sigma_k = \sigma = 1$ for all $1 \leq k < m$.

2.2. Residual of the RKSM approximation for $f(A)b$. We use the RKSM to approximate $y = f(A)b$, where $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, and f is analytic in a neighborhood of the numerical range of A. Since the initial basis vector is $v_1 = b/\beta$, where $\beta = ||b||_2$, the RKSM approximation at step k is defined as

$$y_m = V_m f(A_m) V_m^* b = V_m f(A_m) \beta e_1,$$

where $A_m = V_m^* A V_m$ is the restriction of A to the subspace $\mathcal{Q}_m(A, v_1)$.

The residual norm of the approximation is $\|f(A)b - V_m f(A_m)\beta e_1\|$, but it is impossible to compute it directly since $f(A)b$ is unknown. For certain functions f, we may instead define an alternative residual, associated with an ordinary differential equation that depends on f; see, e.g., [11, 56]. For the exact solution $y = f(A)b$, this alternative residual is zero. By the argument of continuity, if an approximate solution y_m is sufficiently close to y, the alternative residual should be sufficiently small in norm. We give one example to demonstrate this point.

EXAMPLE 2.1. Assume that the matrix A has no negative real eigenvalues. Consider the elliptic Dirichlet problem:

$$\begin{cases} Ay - y''(t) = 0, & t > 0, \\ y(0) = b, & \\ y(+\infty) = 0. & \end{cases}$$

The exact solution is $y(t) = \exp(-t\sqrt{A})b$. If we denote the RKSM approximation as $y_m(t) = V_m \exp(-t\sqrt{A_m})\beta e_1$, then the residual of this approximation with respect to the differential equation is $R_m(t) = Ay_m(t) - y_m''(t)$. Since $y_m''(t) = V_m A_m \exp(-t\sqrt{A_m})\beta e_1$, it follows that

$$\begin{aligned} R_m(t) &= AV_m \exp(-t\sqrt{A_m})\beta e_1 - V_m A_m \exp(-t\sqrt{A_m})\beta e_1 \\ &= (AV_m - V_m A_m) \exp(-t\sqrt{A_m})\beta e_1. \end{aligned}$$

Since the residual of the exact solution is $R(t) = Ay(t) - y''(t) = 0$ for all $t \geq 0$, the residual $R_m(t) = Ay_m(t) - y_m''(t)$ should have a small norm if $y_m(t) \approx y(t)$. In particular, at $t = 1$, the residual of $y_m(t)$ is

$$R_m := R_m(1) = (AV_m - V_m A_m) f(A_m) \beta e_1, \tag{2.6}$$

where $f(z) = \exp(-\sqrt{z})$. If $y_m(1) = V_m \exp(-\sqrt{A_m})\beta e_1 \approx y(1) = \exp(-\sqrt{A})b$, then $\|R_m\|$ is expected to be small.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

There are more examples based on differential equations showing that the residual $\|R_m\| = \|(AV_m - V_m A_m) f(A_m)\beta e_1\|$ can be used to determine the accuracy of the RKSM approximation $y_m \approx f(A)b$ for other functions; see, e.g., [10, 11, 12, 21].

From the Arnoldi relation in (2.2) and (2.5), we get

$$R_m = h_{m+1,m} \left(I - A/s_m + V_m V_m^* A/s_m\right) v_{m+1} e_m^* K_m^{-1} f(A_m)\beta e_1. \tag{2.7}$$

Following the implementation by Güttel in [38], at step k ($k \le m$), we can first temporarily choose the infinite pole $s_k = +\infty$, so that the Arnoldi relation in (2.2) becomes

$$\begin{aligned} AV_k(H_k D_k + I) &= V_k(H_k + P_k) + h_{k+1,k} v_{k+1} e_k^*, \qquad \text{or} \\ AV_k K_k &= V_k G_k + h_{k+1,k} v_{k+1} e_k^*. \end{aligned} \tag{2.8}$$

We can easily obtain the restricted matrix $A_k = V_k^* A V_k = G_k K_k^{-1}$ with the temporary G_k and K_k and compute the residual of $y_k = V_k f(A_k)\beta e_1$. If the residual norm $\|(AV_k - V_k A_k) f(A_k)\beta e_1\|$ is not sufficiently small, then we choose a finalized pole $s_k \neq 0$ and form the finalized G_k, K_k by updating the last column of the temporary G_k and K_k and then proceed to the next RKSM step. The description of this method is shown in Algorithm 1.

Algorithm 1 RKSM for approximating $f(A)b$.

Input: $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n \setminus \{0\}$, function f, maximum step m, tolerance $tol > 0$.
Output: an approximate solution $y_m \approx f(A)b$.

```
Compute the initial vector v_1 = b/β, where β = ||b||_2.
for k = 1, 2, ..., m do
    Let w_{k+1} = (A − σ_k I) v_k, orthogonalize against v_1, v_2, ..., v_k, and normalize into
    (a temporary) v_{k+1}.
    Compute the restricted matrix A_k = V_k^* A V_k = G_k K_k^{-1}.
    Compute the approximate solution y_k = V_k f(A_k) β e_1.
    if ||R_k|| = ||(A V_k − V_k A_k) f(A_k) β e_1|| ≤ tol then
        Return y_k as the approximation to f(A)b.
    end if
    Determine the finalized pole s_k ≠ σ_k.
    Recompute w_{k+1} = (I − A/s_k)^{-1} (A − σ_k I) v_k, orthogonalize w_{k+1} against
    v_1, v_2, ..., v_k, and normalize into (the finalized) v_{k+1}.
    Update the last columns of G_k and K_k.
end for
```

Based on the Arnoldi relation in (2.8), for $k = m$ and $s_m = \infty$, we get

$$\begin{aligned} R_m &= h_{m+1,m} v_{m+1} e_m^* K_m^{-1} f(A_m)\beta e_1, \qquad \text{and} \\ \|R_m\|_2 &= |\beta h_{m+1,m}|\, |e_m^* K_m^{-1} f(A_m) e_1|. \end{aligned} \tag{2.9}$$

Note that, since we usually have $h_{m+1,m} = \mathcal{O}(1)$, the residual norm is directly associated with the $(m, 1)$-entry of the matrix $K_m^{-1} f(A_m)$.

2.3. A sparsity pattern of functions of restricted matrices for the RKSM. In this section, we show that the entries of certain rational functions of restricted matrices constructed by the RKSM have a sparsity pattern. This observation will be used to prove two main theorems in Section 3 on the decay bounds for the entries of $K_m^{-1} f(A_m)$ below the diagonal

ETNA
Kent State University and
Johann Radon Institute (RICAM)

for analytic functions and Markov functions. To this end, we first propose a lemma that states several properties of $A_m = V_m^* A V_m$, the restriction of A to the rational Krylov subspace (2.3).

LEMMA 2.2. *Let $V_m \in \mathbb{R}^{n\times m}$ contain an orthonormal basis of $\mathcal{Q}_m(A, v_1)$ in (2.3). Define $\widetilde{q} = q_{m-1}(A)^{-1}v_1$, where $q_{m-1}(z) = \prod_{j=1}^{m-1}(1 - z/s_j)$. Define $\mathcal{P}_m$ as the set of all polynomials of degree less than or equal to m. The following statements hold:*

(i) *For any matrix $X \in \mathbb{R}^{m\times m}$ and $0 \le j \le m$,*

$$V_m X V_m^* A^j \widetilde{q} = V_m X A_m^j V_m^* \widetilde{q}. \tag{2.10}$$

(ii) *For any matrix $X \in \mathbb{R}^{m\times m}$ and $r_m \in \mathcal{P}_m/q_{m-1}$,*

$$V_m X V_m^* r_m(A) v_1 = V_m X r_m(A_m) V_m^* v_1.$$

Proof. (i) For $j = 0$, the conclusion is trivial. For $1 \le j \le m$, we have

$$\begin{aligned} V_m X V_m^* A^j \widetilde{q} &= V_m X V_m^* A \left(A^{j-1}\widetilde{q}\right) = V_m X V_m^* A V_m V_m^* A^{j-1} \widetilde{q} \\ &= V_m X V_m^* A V_m A_m^{j-1} V_m^* \widetilde{q} = V_m X A_m^j V_m^* \widetilde{q}, \end{aligned}$$

where the second and the third equalities hold by [38, Lemma 3.1].

(ii) Let $X = I$ in (i). By left multiplying V_m^* on both sides of (2.10), we get

$$V_m^* A^j \widetilde{q} = A_m^j V_m^* \widetilde{q},$$

for $0 \le j \le m$. It follows that

$$V_m^* q_{m-1}(A)\widetilde{q} = q_{m-1}(A_m) V_m^* \widetilde{q} \implies V_m^* \widetilde{q} = q_{m-1}(A_m)^{-1} V_m^* v_1.$$

Substitute $\widetilde{q}$ and $V_m^*\widetilde{q}$ with $q_{m-1}(A)^{-1}v_1$ and $q_{m-1}(A_m)^{-1}V_m^* v_1$ in (2.10), respectively, and we eventually get

$$V_m X V_m^* A^j q_{m-1}(A)^{-1} v_1 = V_m X A_m^j q_{m-1}(A_m)^{-1} V_m^* v_1 \qquad (0 \le j \le m),$$

which is sufficient to complete the proof. ∎

LEMMA 2.3. *Suppose that $m-1$ steps of the RKSM are performed without breakdown as in (2.1), with $s_i \ne 0$, $s_i \ne \sigma_i$, for $1 \le i < m$, which leads to the Arnoldi relation in (2.2). Assume that K_m is nonsingular. Define two new vector spaces*

$$\begin{aligned} \widetilde{\mathcal{Q}}_m(A, v_1) &= q_{m-1}(A)^{-1}\operatorname{span}\{v_1, Av_1, \ldots, A^{m-2}v_1\}, \qquad \textit{and} \\ \widehat{\mathcal{Q}}_m(A, v_1) &= \operatorname{span}\left\{(I - A/s_1)^{-1}v_1, \ldots, (I - A/s_{m-1})^{-1}v_{m-1}\right\}. \end{aligned} \tag{2.11}$$

It holds that $\widetilde{\mathcal{Q}}_m(A, v_1) = \widehat{\mathcal{Q}}_m(A, v_1)$.

The quality of a candidate approximation to $f(A)b$ from the RK subspace

$$\widehat{\mathcal{Q}}_m(A, v_1) = \operatorname{span}\left\{(I - A/s_1)^{-1}v_1, \ldots, (I - A/s_{m-1})^{-1}v_1\right\}$$

follows the quality of a corresponding rational function $r(z) = \sum_{j=1}^{m-1} \frac{c_j}{z - s_j}$ of degree $(m-2, m-1)$ for approximating $f(z)$. The property $\widetilde{\mathcal{Q}}_m(A, v_1) = \widehat{\mathcal{Q}}_m(A, v_1)$ is used in most RKSM implementations, where the continuation vector is chosen as the last vector of the basis that has been generated.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

LEMMA 2.4. *With the definitions above, consider the rational functions*

$$r_j^{(t)}(z) = \frac{p_{j-1}(z)}{\prod_{i=t}^{t+j-1}(z-s_i)}, \tag{2.12}$$

where $t \geq 1$ and $p_{j-1}(z) \in \mathcal{P}_{j-1}$. For any indices k, ℓ ($1 \leq \ell < k \leq m$) such that $j+t \leq k \leq m$ and $\ell \leq t$, it holds that

$$e_k^* K_m^{-1} r_j^{(t)}(A_m) e_\ell = 0, \qquad 1 \leq j \leq k-t.$$

Proof. The ℓ-th orthonormal basis vector v_ℓ of $\mathcal{Q}_m(A, v_1)$ can be written as $r_{\ell-1}(A)v_1$, where $r_{\ell-1}(z) \in \mathcal{P}_{\ell-1}/q_{\ell-1}$, such that $r_j^{(t)}(A)v_\ell = r_j^{(t)}(A)r_{\ell-1}(A)v_1$. Since $\ell \leq t$, it holds that

$$r_j^{(t)}(z)r_{\ell-1}(z) \in \mathcal{P}_{t+j-2}/q_{t+j-1} \subseteq \mathcal{P}_m/q_{m-1},$$

which ensures that $r_j^{(t)}(A)v_\ell \in \widetilde{\mathcal{Q}}_{t+j}(A, v_1)$. We set $X = K_m^{-1}$ in (ii), and hence

$$\begin{aligned} V_m K_m^{-1} r_j^{(t)}(A_m) r_{\ell-1}(A_m) V_m^* v_1 &= V_m K_m^{-1} V_m^* r_j^{(t)}(A) r_{\ell-1}(A) v_1 \\ &= V_m K_m^{-1} V_m^* r_j^{(t)}(A) v_\ell. \end{aligned} \tag{2.13}$$

Left multiplying V_m^* on both sides of (ii) and letting $X = I$, we obtain the relation $V_m^* r_m(A)v_1 = r_m(A_m)V_m^* v_1$ for $r_m \in \mathcal{P}_m/q_{m-1}$. Note from (ii) that this equality also holds if r_m is replaced with $r_{\ell-1}$ because $r_{\ell-1} \in \mathcal{P}_m/q_{m-1}$. Therefore,

$$r_{\ell-1}(A_m)V_m^* v_1 = V_m^* r_{\ell-1}(A) v_1 = V_m^* v_\ell = e_\ell. \tag{2.14}$$

Combining (2.13) and (2.14), we get

$$V_m K_m^{-1} r_j^{(t)}(A_m) e_\ell = V_m K_m^{-1} r_j^{(t)}(A_m) \left(r_{\ell-1}(A_m) V_m^* v_1 \right) = V_m K_m^{-1} V_m^* r_j^{(t)}(A) v_\ell.$$

Left multiplying v_k^* on both sides, we have

$$e_k^* K_m^{-1} r_j^{(t)}(A_m) e_\ell = e_k^* K_m^{-1} V_m^* r_j^{(t)}(A) v_\ell. \tag{2.15}$$

By Lemma 2.3 and the fact that $r_j^{(t)}(A)v_\ell \in \widetilde{\mathcal{Q}}_{t+j}(A, v_1)$, there exist scalars α_i, with $1 \leq i \leq j+t-1$, such that $r_j^{(t)}(A)v_\ell = \sum_{i=1}^{j+t-1} \alpha_i (I - A/s_i)^{-1} v_i$. By (2.15), we obtain

$$e_k^* K_m^{-1} r_j^{(t)}(A_m) e_\ell = \sum_{i=1}^{j+t-1} \alpha_i e_k^* K_m^{-1} V_m^* (I - A/s_i)^{-1} v_i. \tag{2.16}$$

By (2.1), we have $(I - A/s_i)^{-1}(A - \sigma_i I)v_i = V_m H_m e_i$, for $1 \leq i < m$. Given the identity $(I - A/s_i)^{-1}(A - \sigma_i I) = s_i \left(\frac{s_i - \sigma_i}{s_i} (I - A/s_i)^{-1} - I \right)$, it follows that

$$\begin{aligned} & s_i \left(\frac{s_i - \sigma_i}{s_i} (I - A/s_i)^{-1} - I \right) v_i = V_m H_m e_i \\ \Longrightarrow\ & \frac{s_i - \sigma_i}{s_i} (I - A/s_i)^{-1} v_i = V_m (H_m D_m + I) e_i = V_m K_m e_i. \end{aligned} \tag{2.17}$$

Left multiplying $K_m^{-1}V_m^*$ on both sides of (2.17), we get

$$\frac{s_i - \sigma_i}{s_i} K_m^{-1} V_m^* (I - A/s_i)^{-1} v_i = e_i. \tag{2.18}$$

Combining (2.16) and (2.18), we have

$$e_k^* K_m^{-1} r_j^{(t)}(A_m) e_\ell = \sum_{i=1}^{j+t-1} \frac{s_i}{s_i - \sigma_i} \alpha_i e_k^* e_i = 0,$$

for all $k \geq j + t$. Note that since $\ell \leq t$ and $t \leq k - j$, we have $\ell \leq k - j$ with $j \geq 1$ and hence $\ell < k$. This means that the above sparsity pattern holds in the strictly lower triangular portion of $K_m^{-1} r_j^{(t)}(A_m)$. □

Lemma 2.4 shows that for the RKSM, there exists a sparsity pattern for the entries of $K_m^{-1} r_j^{(t)}(A_m)$ involving the class of rational functions (2.12) and the restricted matrices A_m obtained by the RKSM, and Figure 2.1 illustrates two examples of the sparsity patterns for certain t- and j-values. In [57], a corresponding result has been derived for the entries of $r_j^{(t)}(A_m)$, but our result involving K_m^{-1} is directly associated with the residual of the RKSM approximations for $f(A)b$ as given in (2.9). We will show that this sparsity property helps to establish the two main theorems in Section 3, derive the convergence of the RKSM, and develop the tolerance relaxation for the iterative linear solve at each step of the inexact RKSM.

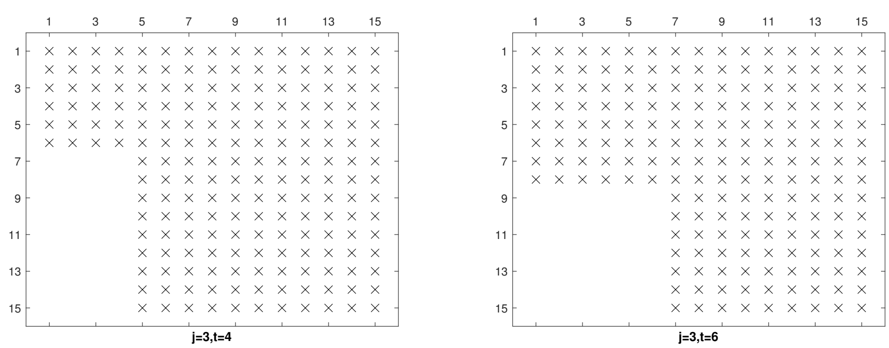

FIG. 2.1. *Sparsity pattern of $K_{15}^{-1} r_j^{(t)}(A_{15})$ in Lemma 2.4.*

3. Decay bounds for functions of matrices.

In this section, we investigate the upper bounds for the entries in the first column of $K_m^{-1} f(A_m)$, namely $\left|e_k^* K_m^{-1} f(A_m) \beta e_1\right|$ in (2.9), for $1 \leq k \leq m$. The core point is to show how quickly these entries decay with the row index k. In Section 4, we shall explore the connections between the decay bounds and the residual of the RKSM for approximating $f(A)b$.

3.1. Decay bounds for $\left|e_k^* K_m^{-1} f(A_m) \beta e_1\right|$ for analytic functions and Markov functions.

Several estimates for decay bounds for the entries of functions of matrices have been proposed; see, e.g., [4, Theorem 10], [6, Theorem 3.7], [52, Theorem 2.6], and [56, Theorem 2.3]. In this paper, we use the Faber-Dzhrbashyan (FD) rational functions [62, Ch. XIII, Section 3] and [51] to find upper bounds for $\left|e_k^* K_m^{-1} f(A_m) \beta e_1\right|$.

We begin with some definitions. For any matrix $A \in \mathbb{R}^{n \times n}$, we let

$$W(A) = \{x^* A x \mid x \in \mathbb{C}^n, \|x\|_2 = 1\}$$

be the numerical range of A. Let $E \subset \mathbb{C}$ be a connected compact metric space such that $W(A) \subset E$. Also denote the extended complex plane as $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, and the unit disk as $D = \{|w| \leq 1\}$. Define ϕ as the Riemann mapping that maps $\overline{\mathbb{C}} \setminus E$ conformally onto $\overline{\mathbb{C}} \setminus D$ such that $\phi(\infty) = \infty$ and $\lim_{z\to\infty} \frac{\phi(z)}{z} > 0$. Let $\psi = \phi^{-1}$ be the inverse mapping of ϕ.

The matrix function $f(A)$ can be defined by Cauchy's integral formula as follows [42]:

DEFINITION 3.1. *If f is analytic in a region $E \subseteq \mathbb{C}$ and $W(A) \subset E$, we have*

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma_E} f(z)(zI - A)^{-1} dz,$$

where Γ_E is a closed contour in E that encloses the spectrum of A.

In [57, Theorem 4.2], decay bounds for the entries of $f(A_m)$ are derived. Since K_m in (2.4) is an upper Hessenberg matrix, K_m^{-1} is the inverse function of a band matrix, and decay bounds for the entries of K_m^{-1} can be derived [19]. One can combine the results of the decay bounds for both $f(A_m)$ and K_m^{-1} to get those for $K_m^{-1} f(A_m)$. However, the decay bounds for K_m^{-1} require spectral information of K_m defined in (2.4), which has no straightforward connections to $W(A)$. In this paper, we follow the work in [57] to directly derive decay bounds for $K_m^{-1} f(A_m)$ by using the Faber-Dzhrbashyan (FD) rational functions. Our first main theorem reads as follows.

THEOREM 3.2. *Assume that K_m defined in (2.4) is nonsingular and $A_m = V_m^* A V_m$ is the restriction of A to the rational Krylov subspace $\mathcal{Q}_m(A, v_1)$ defined in (2.3), with orthonormal basis vectors $[v_1, \ldots, v_m]$. Suppose that all poles $s_1, \ldots, s_m$ of the RKSM are located in the exterior of the set E, where $E \subset \mathbb{C}$ is a connected compact metric space such that $W(A) \subset E$. For $k, \ell \in \mathbb{N}^+$ and $\ell < k \leq m$, define $\alpha_j = \left[\overline{\phi(s_{j+\ell-1})}\right]^{-1}$ for $1 \leq j \leq k - \ell$, where ϕ is the Riemann mapping. Let $\psi = \phi^{-1}$ be the inverse Riemann mapping. Let $\tau > 1$ be such that f is analytic in $E_\tau = E \cup \{z \in \mathbb{C} \setminus E \mid |\phi(z)| \leq \tau\}$. It holds that*

$$\left|e_k^* K_m^{-1} f(A_m) e_\ell\right| \leq 3 \|e_k^* K_m^{-1}\| \sum_{j=k-\ell}^{\infty} |c_j|, \tag{3.1}$$

where

$$c_j = \frac{1}{2\pi} \int_0^{2\pi} f\left(\psi(\tau e^{i\theta})\right) \left(-\frac{1}{\tau}\right)^{j-(k-\ell)} e^{-i\theta[j-(k-\ell)]} \prod_{i=1}^{k-\ell} \frac{\tau e^{i\theta}\overline{\alpha_i} - 1}{\tau e^{i\theta} - \alpha_i} \frac{|\alpha_i|}{\overline{\alpha_i}} d\theta \tag{3.2}$$

and c_j is independent of the value of τ. Moreover, a simplified bound holds in the form

$$\left|e_k^* K_m^{-1} f(A_m) e_\ell\right| \leq \frac{3}{2\pi} \|e_k^* K_m^{-1}\| \frac{\tau}{\tau - 1} \int_0^{2\pi} \left| f\left(\psi(\tau e^{i\theta})\right) \prod_{i=1}^{k-\ell} \frac{\tau e^{i\theta}\overline{\alpha_i} - 1}{\tau e^{i\theta} - \alpha_i} \right| d\theta. \tag{3.3}$$

The proof of Theorem 3.2 is given in Section 6.

Next, we study the bounds for the entries of $K_m^{-1} f(A_m)$ for an important class of non-analytic functions, namely, Markov (Cauchy-Stieltjes) functions, defined as [3]

$$f(z) = \int_{-\infty}^0 \frac{d\mu(\zeta)}{z - \zeta}, \qquad z \in \mathbb{C} \setminus (-\infty, 0], \tag{3.4}$$

where μ is a positive measure with $\operatorname{supp}(\mu) \subset (-\infty, 0]$. Markov functions are not analytic in the entire set E_τ for any $\tau > 1$ that is not sufficiently small. Below are two examples of Markov functions.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

EXAMPLE 3.3. For $f(z) = z^{-1/2}$, we can write

$$f(z) = z^{-1/2} = \int_{-\infty}^{0} \frac{d\mu(\zeta)}{z-\zeta}, \qquad \text{where } \mu'(\zeta) = \frac{1}{\pi\sqrt{-\zeta}}.$$

EXAMPLE 3.4. for $f(z) = e^{-\sqrt{z}}$, we can write

$$f(z) = e^{-\sqrt{z}} = \int_{-\infty}^{0} \frac{d\mu(\zeta)}{z-\zeta}, \qquad \text{where } \mu'(\zeta) = \frac{\sin(\sqrt{-\zeta})}{\pi}. \tag{3.5}$$

The two Markov functions above are not analytic in $(-\infty, 0]$. If we use Theorem 3.2 to determine an upper bound, we should choose τ such that $1 < \tau < |\phi(0)|$. Such a bound is usually a significant overestimate of the actual rate of decay. Instead, we present a similar theorem for Markov functions with decay bounds that are much sharper. Note that [7, 34] have established decay bounds for Markov functions of matrices with banded or Kronecker structure, whereas our results hold without assumptions on the matrix structure.

THEOREM 3.5. *With the same setting as Theorem 3.2, except that f is a Markov function defined in (3.4), and assuming that E lies strictly in the right half complex plane, it holds that*

$$\left|e_k^* K_m^{-1} f(A_m) e_\ell\right| \le 3\|e_k^* K_m^{-1}\| \sum_{j=k-\ell}^{\infty} |c_j|, \tag{3.6}$$

where

$$c_j = \int_{-\infty}^{\phi(0)} \left(-\frac{1}{w}\right)^{j+1-(k-\ell)} \prod_{i=1}^{k-\ell} \frac{w\overline{\alpha_i}-1}{w-\alpha_i} \frac{|\alpha_i|}{\overline{\alpha_i}} \frac{d\mu(\psi(w))}{\psi'(w)}.$$

Moreover, a simplified bound holds in the form

$$\left|e_k^* K_m^{-1} f(A_m) e_\ell\right| \le 3\|e_k^* K_m^{-1}\| \int_{-\infty}^{\phi(0)} \left|\frac{1}{\psi'(w)}\right| \frac{1}{|w+1|} \left|\prod_{i=1}^{k-\ell} \frac{w\overline{\alpha_i}-1}{w-\alpha_i}\right| |d\mu(\psi(w))|. \tag{3.7}$$

The proof of Theorem 3.5 is also given in Section 6.

Similar to the upper bounds for $|e_k^* f(A_m) e_\ell|$ and $|e_k^* A_m e_\ell|$ in [57], we may further relax the bounds for $\left|e_k^* K_m^{-1} f(A_m) e_\ell\right|$ in Theorem 3.2 and Theorem 3.5 by replacing the integrals in the formulas with rough bounds in terms of elementary functions. However, we point out that keeping the integrals and efficiently approximating them by customized quadrature rules can achieve significantly sharper bounds at a cost not much higher than that needed to evaluate the rough bounds without integrals. In fact, based on the numerical tests in [57], theoretical bounds roughly give overestimates that are four orders of magnitude larger than the actual values. We shall present customized numerical quadrature rules to efficiently approximate our bounds with integrals. This will be discussed in detail in Section 5.1.

3.2. Poles and the rate of convergence of the RKSM. Though the main goal of this paper is to study the mechanism to enable the inexact RKSM for approximating $f(A)b$, in this section we give a brief discussion about the implications of the bounds (3.1) and (3.6) to explore numerical or theoretical relationships between the poles and the asymptotic convergence factor of the RKSM in this problem setting. These relationships seem consistent with those established in [53] and [3] for the exponential function and Markov functions, respectively. For the exponential function $f(z) = e^{-hz}$, which is analytical in the entire complex plane, the Restricted Denominator (RD) rational approximation is a competitive method; see, e.g., [53].

ETNA
Kent State University and
Johann Radon Institute (RICAM)

RD rational approximations can be regarded as a special variant of the RKSM with a fixed repeated pole. It is shown in [53] that if $W(A)$ is a sector in the right half complex plane with vertex at the origin and is symmetric with respect to the real axis, then the optimal pole of RD is $s_0 = -m/h$, where m is the maximum number of RD steps. From the definition of the residual norm in (2.9) and the upper bound (3.3) for analytical functions we get

$$\begin{aligned}\|R_m\|_2 &= |\beta h_{m+1,m}| \left|e_m^* K_m^{-1} f(A_m) e_1\right| \\ &\le \frac{3}{2\pi} |\beta h_{m+1,m}| \|e_m^* K_m^{-1}\| \frac{\tau}{\tau-1} \int_0^{2\pi} \left| f\left(\psi(\tau e^{i\theta})\right) \prod_{i=1}^{m-1} \frac{\tau e^{i\theta}\overline{\alpha_i}-1}{\tau e^{i\theta}-\alpha_i} \right| d\theta.\end{aligned}$$

Assume that there exists a uniform upper bound for both $|h_{m+1,m}|$ and $\|e_m^* K_m^{-1}\|$ independent of m. Then for the RKSM with a fixed repeated pole $s \in \mathbb{R}$, $\alpha_i = \left[\overline{\phi(s_i)}\right]^{-1}$ $(1 \le i \le m-1)$, and therefore,

$$\|R_m\|_2 \le C_1 \frac{\tau}{\tau-1} \int_0^{2\pi} \left| f\left(\psi(\tau e^{i\theta})\right) \left(\frac{\tau e^{i\theta}-\phi(s)}{\tau e^{i\theta}\phi(s)-1}\right)^{m-1} \right| d\theta.$$

The optimal single repeated pole $s = s^* \le 0$ is defined as

$$s^* = \arg\min_{s\le 0} \min_{\tau>1} \frac{\tau}{\tau-1} \int_0^{2\pi} \left| f\left(\psi(\tau e^{i\theta})\right) \left(\frac{\tau e^{i\theta}-\phi(s)}{\tau e^{i\theta}\phi(s)-1}\right)^{m-1} \right| d\theta. \tag{3.8}$$

We use a composite trapezoidal rule to approximate the integral and use MATLAB's `fminbnd` (a function which aims to find the minimum of a continuous single-variable function on a finite interval) to approximate the optimal single repeated pole s^*.

Assume that A is a real nonsymmetric matrix such that $W(A)$ can be covered by an ellipse with semi-major axis of length a parallel to the real axis and semi-minor axis of length b parallel to the imaginary axis ($a > b \ge 0$), centered at $c \in \mathbb{C}$. The conformal map $\phi(z)$ is then defined by

$$\phi(z) = \begin{cases} \frac{z-c+\sqrt{(z-c)^2-\rho^2}}{\rho\kappa}, & \Re(z-c) > 0, \\ \frac{z-c-\sqrt{(z-c)^2-\rho^2}}{\rho\kappa}, & \Re(z-c) < 0, \end{cases}$$

and its inverse is defined by

$$\psi(w) = \frac{\rho}{2}\left(\kappa w + \frac{1}{\kappa w}\right) + c,$$

where $\rho = \sqrt{a^2-b^2}$ and $\kappa = (a+b)/\rho$; see, e.g., [57]. For $a = b$, $W(A)$ can be covered by a circle, so that

$$\phi(z) = \begin{cases} \frac{z-c+\sqrt{(z-c)^2}}{a+b}, & \Re(z-c) > 0, \\ \frac{z-c-\sqrt{(z-c)^2}}{a+b}, & \Re(z-c) < 0, \end{cases} \qquad \psi(w) = \frac{a+b}{2} w + c.$$

For $a < b$, $W(A)$ can be covered by an ellipse with semi-major axis parallel to the imaginary axis, and we can derive the similar expressions for both $\phi(z)$ and $\psi(w)$.

For example, assume that a matrix A is such that its numerical range $W(A)$ can be covered by an ellipse centered at $c = 101$, with semi-major axis of length $a = 100$ lying on the

ETNA
Kent State University and
Johann Radon Institute (RICAM)

real axis and semi-minor axis of length $b = 10$. Table 3.1 shows the comparison of the optimal single pole s^* computed numerically in (3.8) and $s_0 = -\frac{m}{h}$ in [53] for approximating $e^{-hA}b$. One can see from the table that the two poles are relatively close. The difference between these two poles might be attributed to the different shape of $W(A)$, which is assumed to be an infinite sector in [53] but is a finite ellipse in our experiments. The result in [53] is more suitable for some problems arising from discretizing PDEs with different mesh sizes, because all of them can be fitted into identical sector with infinite radius. In this paper, following the assumptions in the literature on the convergence of the RKSM based on Riemann mappings $\phi(z)$ and the inverses $\psi(w)$, we focus on matrices with a finite numerical range.

TABLE 3.1

Comparison of the optimal single pole s^* (3.8) and $s_0 = -\frac{m}{h}$ [53] for approximating $e^{-hA}b$, where $W(A)$ can be covered by an ellipse centered at $c = 101$, with semi-major axis of length $a = 100$ lying on the real axis and semi-minor axis of length $b = 10$.

	$h = 0.1$			$h = 1$			$h = 10$		
m	s^*	s_0	s^*/s_0	s^*	s_0	s^*/s_0	s^*	s_0	s^*/s_0
20	−174.35	−200	0.8717	−16.60	−20	0.8298	−0.92	−2	0.4578
40	−298.70	−400	0.7468	−38.08	−40	0.9521	−2.70	−4	0.6749
60	−437.64	−600	0.7294	−64.02	−60	1.0669	−4.45	−6	0.7418
80	−585.49	−800	0.7319	−93.91	−80	1.1739	−6.20	−8	0.7755
100	−738.70	−1000	0.7387	−126.80	−100	1.2680	−7.97	−10	0.7975

For Markov functions, from the definition of the residual norm (2.9) and (3.7) in Theorem 3.5, we get

$$\begin{aligned}\|R_m\|_2 &= |\beta h_{m+1,m}| \left|e_m^* K_m^{-1} f(A_m) e_1\right| \\ &\le 3|\beta h_{m+1,m}| \|e_m^* K_m^{-1}\| \int_{-\infty}^{\phi(0)} \frac{|\mu'(\psi(w))|}{|w+1|} \left|\prod_{i=1}^{m-1} \frac{w\overline{\alpha_i} - 1}{w - \alpha_i}\right| dw \\ &\le 3|\beta h_{m+1,m}| \|e_m^* K_m^{-1}\| \int_{-\infty}^{\phi(0)} \frac{|\mu'(\psi(w))|}{|w+1|} dw \max_{w\in(-\infty,\phi(0)]} \left|\prod_{i=1}^{m-1} \frac{w - \phi(s_i)}{\phi(s_i)w - 1}\right|.\end{aligned}$$

Assume that there exists a uniform upper bound for $|\beta h_{m+1,m}| \|e_m^* K_m^{-1}\| \int_{-\infty}^{\phi(0)} \frac{|\mu'(\psi(w))|}{|w+1|} dw$ independent of m. Then the optimal poles s_i $(1 \le i \le m)$ can be found by minimizing

$$\max_{w\in(-\infty,\phi(0)]} \left|\prod_{i=1}^{m-1} \frac{w - \phi(s_i)}{\phi(s_i)w - 1}\right|,$$

which is consistent with the findings in [3]. In particular, for a fixed repeated pole $s_i = s < 0$, there exists $C_2 \in \mathbb{R}^+$ such that

$$\|R_m\|_2 \le C_2 \max_{w\in(-\infty,\phi(0)]} \left|\frac{w - \phi(s)}{\phi(s)w - 1}\right|^{m-1} = C_2 \max\left\{\left|\frac{\phi(0) - \phi(s)}{\phi(s)\phi(0) - 1}\right|, \left|\frac{1}{\phi(s)}\right|\right\}^{m-1},$$

where the last equality holds since $\frac{w-\phi(s)}{\phi(s)w-1}$ is monotonic in w on $(-\infty, \phi(0)]$. Essentially the same result for the residual defined by $\|f(A)b - V_m f(A_m)\beta e_1\|$ can be found in [3, Corollary 6.4 (a)]. A similar residual bound obtained with two cyclic poles can also be derived, corresponding to the result in [3, Corollary 6.4 (b)]. The above discussion gives an alternative proof of the residual bounds for approximating the action of a Markov function $f(A)b$ by the RKSM with a few cyclic poles.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

4. An inexact RKSM for approximating $f(A)b$. Inexact Arnoldi algorithms have been widely used in solving numerical linear algebra problems, including approximating $f(A)b$. In general, these algorithms include inexact standard (polynomial) Krylov methods and inexact rational Krylov methods. They can be applied to symmetric and nonsymmetric matrices, while f can be an analytic function or a Markov function. Preliminary test results were given in [8] for the inexact standard Krylov method to approximate $f(A)b$, where A is symmetric and positive definite and f is analytic. Inexact standard Krylov subspace methods have also been studied in [20] and [56] for approximating $f(A)b$, where A is nonsymmetric and f is analytic. Several inexact rational Krylov methods, including the shift-and-invert Lanczos method and the EKSM (which rely on only one fixed pole), have been investigated in [40] for approximating the action of Markov functions of Hermitian matrices to vectors. To the best of our knowledge, no studies have been carried out to explore the inexact rational Krylov method with variable poles for approximating $f(A)b$ involving nonsymmetric matrices for either analytic functions or Markov functions. Our goal of study is to fill this research gap.

For large-scale problems, the approximate computation of the shift-invert matrix vector product $w_{k+1} = (I - A/s_k)^{-1}(A - \sigma I)v_k$ in (2.1) at step k of the RKSM is done by an iterative linear solver. Errors are introduced in the approximate solution and hence into the basis vectors of the rational Krylov subspaces. Let $\widehat{w}_{k+1}$ be an approximate solution such that the residual of this linear solve is $\xi_k = (A - \sigma I)v_k - (I - A/s_k)\widehat{w}_{k+1}$. Then (2.1) turns into

$$\widehat{w}_{k+1} = (I - A/s_k)^{-1}\left((A - \sigma I)v_k - \xi_k\right) = \sum_{i=1}^{k+1} h_{ik} v_i.$$

If we choose the pole $s_m = \infty$, then the inexact Arnoldi relation of RKSM after step m is

$$AV_m(H_m D_m + I) = V_m(H_m + P_m) + h_{m+1,m} v_{m+1} e_m^* + \Xi_m, \tag{4.1}$$

where $\Xi_m = [\xi_1, \xi_2, \ldots, \xi_m]$ contains the residual vectors of the approximate linear solves in the first m steps of the RKSM. The *true residual* of the inexact method for approximating $f(A)b$ is defined as

$$\begin{aligned} \widetilde{R}_m &= (AV_m - V_m A_m) f(A_m)\beta e_1 \\ &= \left[h_{m+1,m} v_{m+1} e_m^* + (I - V_m V_m^*)\Xi_m\right] K_m^{-1} f(A_m)\beta e_1. \end{aligned} \tag{4.2}$$

We are interested in exploring strategies to make the difference between the *derived residual* in (2.9) and the *true residual* in (4.2) sufficiently small, so that the error term Ξ_m has little impact on the convergence of the inexact RKSM. The difference between the two residuals is

$$\Delta_m = R_m - \widetilde{R}_m = (V_m V_m^* - I)\Xi_m K_m^{-1} f(A_m)\beta e_1.$$

From the definition of Δ_m, we have

$$\begin{aligned} \|\Delta_m\|_2 &= \left\|(V_m V_m^* - I)\,\Xi_m K_m^{-1} f(A_m)\beta e_1\right\|_2 \leq \|V_m V_m^* - I\|_2 \left\|\Xi_m K_m^{-1} f(A_m)\beta e_1\right\|_2 \\ &= \left\|\sum_{k=1}^{m} \xi_k e_k^* K_m^{-1} f(A_m)\beta e_1\right\|_2 \leq \sum_{k=1}^{m} \|\xi_k\|_2 \left|e_k^* K_m^{-1} f(A_m)\beta e_1\right|, \end{aligned} \tag{4.3}$$

where the second equality holds since $I - V_m V_m^*$ is an orthogonal projector.

In order to make $\|\Delta_m\|$ sufficiently small, either $\|\xi_k\|$ or $|e_k^* K_m^{-1} f(A_m)\beta e_1|$ should be sufficiently small for each index value $1 \leq k \leq m$. An a priori evaluation of the upper bound

ETNA
Kent State University and
Johann Radon Institute (RICAM)

for $|e_k^* K_m^{-1} f(A_m)\beta e_1|$ can be used to determine how large $\|\xi_k\|$ could be at each step. We have already derived the decay bounds for $|e_k^* K_m^{-1} f(A_m)\beta e_1|$ in Theorem 3.2 and Theorem 3.5 for analytic functions and Markov functions, respectively. Our next step is to investigate a relaxation strategy for the accuracy of the linear solve $(I - A/s_k)w_{k+1} = (A - \sigma_k I)v_k$ at each RKSM step.

The inexact Arnoldi relation for a matrix A in (4.1) is equivalent to an exact Arnoldi relation for the perturbed matrix $\widetilde{A} = A - \Xi_m K_m^{-1} V_m^*$. It follows that the restricted matrix $\widetilde{A}_m$ equals to $V_m^* \left(A - \Xi_m K_m^{-1} V_m^*\right) V_m$, and

$$
\begin{aligned}
W(\widetilde{A}_m) &\subset W\left(A - \Xi_m K_m^{-1} V_m^*\right) \\
&\subset \left\{ z \mid z = z_1 - z_2, z_1 \in W(A), z_2 \in W\left(\Xi_m K_m^{-1} V_m^*\right) \right\}.
\end{aligned}
$$

Suppose that c_m is an upper bound for $\left\|K_m^{-1}\right\|$. For any vector w with unit 2-norm, we have

$$
\left|w^* \Xi_m K_m^{-1} V_m^* w\right| \leq \|\Xi_m\| \left\|K_m^{-1}\right\| \leq c_m \|\Xi_m\| \leq c_m \sqrt{\sum_{k=1}^{m} \|\xi_k\|^2}. \tag{4.4}
$$

Define $\varepsilon = c_m \sqrt{\sum_{k=1}^m \|\xi_k\|^2}$ such that $\left|w^* \Xi_m K_m^{-1} V_m^* w\right| \leq \varepsilon$. Assume that $W(A)$ is a subset of an ellipse centered at $c = c_1 + c_2 i \in \mathbb{C}$, with semi-major axis of length a parallel to the real axis and semi-minor axis of length b ($a \geq b \geq 0$), where $c_1 > a$. For any point on the boundary of the ellipse that covers $W(A)$, denoted as $p = (c_1 + a\cos\theta, c_2 + b\sin\theta)$, consider a corresponding point $p^* = (c_1 + a\cos\theta + \varepsilon\cos\alpha, c_2 + b\sin\theta + \varepsilon\sin\alpha)$. Since the sum of the distances from p to the two foci of the ellipse is $2a$, it is easy to show that the sum of the distances from p^* to the two foci is less than or equal to $2a + 2\varepsilon$ by applying the triangle inequality involving the three triangles with vertices p, p^*, and the two foci. Therefore, $W(A - \Xi_m K_m^{-1} V_m^*)$ can be covered by a larger ellipse centered at $c = c_1 + c_2 i$ with semi-major axis of length $a_t = a + \varepsilon$ and semi-minor axis of length $b_t = \sqrt{(a+\varepsilon)^2 - a^2 + b^2} \geq 0$, which also has the same foci and focal distance $\sqrt{a^2 - b^2}$ as the original ellipse.

The following theorem provides a tolerance relaxation strategy for the inexact linear solve at each step of the RKSM.

THEOREM 4.1. *Let $\widetilde{R}_m$ and R_m be the true residual (4.2) and the derived residual (2.7) after m steps of the inexact RKSM for approximating $f(A)b$, where f is analytic in a connected compact metric space $E \subset \mathbb{C}$.*

Let $tol > 0$, and let $\varepsilon > 0$ be small and arbitrary. Define χ_k as the upper bound, either (3.1) and (3.3) for $\left|e_k^* K_m^{-1} f(A_m)\beta e_1\right|$ ($1 \leq k \leq m$) for analytic functions in Theorem 3.2, or (3.6) and (3.7) for Markov functions in Theorem 3.5. Let c_m be a uniform upper bound for $\left\|K_k^{-1}\right\|$ for $1 \leq k \leq m$.

Assume that $W(A)$ is a subset of an ellipse centered at $c = c_1 + c_2 i \in \mathbb{C}$, with semi-major axis of length a parallel to the real axis and semi-minor axis of length b ($a \geq b \geq 0$) and that E covers an elliptic boundary centered at $c = c_1 + c_2 i$, with semi-major axis of length $a_t = a + \varepsilon$ and semi-minor axis of length $b_t = \sqrt{(a+\varepsilon)^2 - a^2 + b^2}$, ($a_t \geq b_t$).

Suppose that for every $1 \leq k \leq m$, $\|\xi_k\| \leq \epsilon_k$, where

$$
\epsilon_k = \min\left\{ \frac{tol}{m\chi_k}, \frac{1}{m-k+1} \sqrt{\frac{\varepsilon^2}{c_m^2} - \sum_{i=1}^{k-1} \epsilon_i^2} \right\}. \tag{4.5}
$$

Then $\|\Delta_m\| = \left\|\widetilde{R}_m - R_m\right\| \leq tol$, and $\sum_{i=1}^m \epsilon_i^2 \leq \varepsilon^2 / c_m^2$.

Proof. From (4.5), we conclude that $\epsilon_k \le \frac{1}{m-k+1}\sqrt{\varepsilon^2/c_m^2 - \sum_{i=1}^{k-1}\epsilon_i^2}$ for all $1 \le k \le m$. Therefore, we have $\sum_{i=1}^{k}\epsilon_i^2 \le \sum_{i=1}^{k-1}\epsilon_i^2 + (m-k+1)^2\epsilon_k^2 \le \varepsilon^2/c_m^2$. Specifically, when $k = m$ we conclude that $\sum_{i=1}^{m}\epsilon_i^2 \le \varepsilon^2/c_m^2$.

From (4.4), we have $\left|w^*\Xi_k K_k^{-1}V_k^* w\right| \le c_m\sqrt{\sum_{i=1}^{k}\|\xi_i\|^2} \le c_m\sqrt{\sum_{i=1}^{m}\|\xi_i\|^2} \le \varepsilon$ for all $1 \le k \le m$. It follows that $W(A - \Xi_k K_k^{-1}V_k^*) \subset E$, so that f is analytic in the numerical range of all the perturbed matrices $\widetilde{A}_k = A - \Xi_k K_k^{-1}V_k^*$ $(1 \le k \le m)$.

From (4.5), we also conclude that $\epsilon_k \le \frac{tol}{m\chi_k}$ for all $1 \le k \le m$. By the expression for $\|\Delta_m\|$ in (4.3), it follows that

$$\|\Delta_m\| \le \sum_{k=1}^{m}\|\xi_k\|_2\left|e_k^*K_m^{-1}f(A_m)\beta e_1\right| \le \sum_{k=1}^{m}\epsilon_k\chi_k \le \sum_{k=1}^{m}\frac{tol}{m\chi_k}\chi_k = tol. \tag{4.6}$$ $\square$

In Theorems 3.2 and 3.5 we derived a decaying behavior of the entries in the first column of $K_m^{-1}f(A_m)$. Since these entries decrease in modulus with the row index, it follows from (4.6) that the tolerance of the inexact linear solves can be relaxed with the RKSM progress. In practice, the upper bounds for $\left|e_k^*K_m^{-1}f(A_m)\beta e_1\right|$ suggested by Theorem 3.2 or Theorem 3.5 involve integrals, which may take some time to be approximated to a reasonable accuracy. Also, these bounds could give significant overestimates of the actual entries at certain RKSM steps, which may lead to an excessively conservative relaxation estimate.

Instead, we consider a heuristic estimate of $\left|e_k^*K_m^{-1}f(A_m)\beta e_1\right|$ based on $\|R_k\|$, which usually gives a less conservative tolerance relaxation for the approximate linear solve at each RKSM step and works well in practice. To derive this heuristic, we define the actual entry of interest $\chi_k = \left|e_k^*K_m^{-1}f(A_m)\beta e_1\right|$, where $K_m, A_m \in \mathbb{R}^{m\times m}$ are obtained after applying the temporary pole $s_m = \infty$ at step m of the RKSM. From the expression of $\|R_m\|$ in (2.9), we have $\|R_k\| = |\overline{h}_{k+1,k}|\left|e_k^*\overline{K}_k^{-1}f(\overline{A}_k)\beta e_1\right|$, where the scalar $\overline{h}_{k+1,k}$ and the restricted matrices $\overline{K}_k, \overline{A}_k \in \mathbb{R}^{k\times k}$ are obtained after applying the finalized finite pole s_k at step k. From (3.1) in Theorem 3.2 or (3.6) in Theorem 3.5, we have

$$\chi_k = \left|e_k^*K_m^{-1}f(A_m)\beta e_1\right| \le 3\|e_k^*K_m^{-1}\|\sum_{j=k-1}^{\infty}|c_j|, \qquad \text{and} \tag{4.7}$$

$$\|R_k\|/|\overline{h}_{k+1,k}| = \left|e_k^*\overline{K}_k^{-1}f(\overline{A}_k)\beta e_1\right| \le 3\|e_k^*\overline{K}_k^{-1}\|\sum_{j=k-1}^{\infty}|c_j|. \tag{4.8}$$

Since $W(A_m), W(\overline{A}_k) \subseteq W(A) \subset E$ and the first $k-1$ poles remain the same for generating K_k, A_k, $h_{k+1,k}$, and $\overline{K}_k$, $\overline{A}_k$, $\overline{h}_{k+1,k}$, the definitions of c_j in (4.7) and (4.8) have identical expressions, for both analytic functions and Markov functions. Suppose that $\|e_k^*K_m^{-1}\|$ and $\|e_k^*\overline{K}_k^{-1}\|$ are close and that the bounds in (4.7) and (4.8) are comparably sharp. Then χ_k can be approximated by $\frac{\|R_k\|}{|\overline{h}_{k+1,k}|}$.

Based on Algorithm 1, it is possible to get both $\|R_k\|$ and $\left|\overline{h}_{k+1,k}\right|$ with the temporary infinite pole at step k, before we need to use $\chi_k \approx \frac{\|R_k\|}{|\overline{h}_{k+1,k}|}$ to set the tolerance for the iterative linear solve $(I - A/s_k)w_{k+1} = (A - \sigma_k I)v_k$ with the finalized finite pole s_k. Since this is a heuristic estimate of χ_k, we may have a lower risk of applying excessive relaxation by slightly increasing this estimate so that the tolerance of the inexact linear solve can be tightened moderately, and the inexact RKSM may follow the behavior of the exact algorithm more reliably. In practice, we set $\chi_k = \frac{10\|R_k\|}{|\overline{h}_{k+1,k}|}$ in our numerical tests.

5. Numerical experiments. In this section, we first provide numerical evidence to show the sharpness of our upper bounds for the entries of $K_m^{-1} f(A_m)$ for both analytic functions and Markov functions and discuss efficient quadrature rules to approximate these bounds. Then we numerically show the advantage of the inexact RKSM over the exact method for approximating $f(A)b$. All experiments were carried out in MATLAB R2021b on a laptop running in Windows 10 with 16GB DDR4 2400 MHz memory and a 2.81 GHz Intel Dual Core CPU.

5.1. Upper bounds for the entries of $K_m^{-1} f(A_m)$. To study the decaying pattern for the residual norm $\|R_m\|_2$ in (2.9), we are mostly interested in approximating the upper bound for $\left|e_k^* K_m^{-1} f(A_m) e_1\right|$, for $k \leq m$, accurately and efficiently.

For analytical functions, Theorem 3.2 provides upper bounds in both (3.1) and (3.3), referred to as the original bound and the simplified bound, respectively. Although $|c_j|$ defined in (3.2) is independent of the values of τ, numerical tests show that it is unstable to approximate $|c_j|$ in (3.2) for a wide range of values of τ. A better approach is to use `fminbnd` in MATLAB to find the optimal $\tau > 1$ that minimizes the partial sum of the infinite series of the $|c_j|$'s. To compute $|c_j|$ in (3.2) with a fixed value of τ, we use a composite trapezoid rule to approximate the integral in (3.2) and then employ the fast Fourier transform (FFT) to evaluate a partial sum of the infinite series in (3.1). In a neighborhood of the optimal value of τ, numerical experiments show the efficiency of the composite trapezoid rule evaluated by the FFT.

Specifically, we evaluate the expression for c_j in (3.2) by the composite trapezoid rule

$$c_j \approx \sum_{p=0}^{N-1} w_p Q_p \left(\frac{1}{\tau}\right)^{j-(k-\ell)} e^{-i\theta_p[j-(k-\ell)]},$$

where N is the number of quadrature points, $\theta_p = \frac{2\pi p}{N}$ $(0 \leq p \leq N-1)$ are the quadrature nodes, $w_p = \frac{2\pi}{N-1}$ are the quadrature weights for all $0 \leq p \leq N-1$, and

$$Q_p = \frac{1}{2\pi} f\left(\psi(\tau e^{i\theta_p})\right) \prod_{i=1}^{k-\ell} \frac{\tau e^{i\theta_p}\overline{\alpha_i} - 1}{\tau e^{i\theta_p} - \alpha_i}.$$

To approximate the infinite series for $|c_j|$, we compute the first $N = 2^{14} = 16384$ terms of c_j. It follows that

$$\begin{aligned}
\sum_{j=k-\ell}^{\infty} |c_j| &\approx \sum_{j=k-\ell}^{k-\ell+N-1} \left| \sum_{p=0}^{N-1} w_p Q_p \left(\frac{1}{\tau}\right)^{j-(k-\ell)} e^{-i\theta_p[j-(k-\ell)]} \right| \\
&= \sum_{j=k-\ell}^{k-\ell+N-1} \left(\frac{1}{\tau}\right)^{j-(k-\ell)} \left| \sum_{p=0}^{N-1} w_p Q_p e^{-i\frac{2\pi p}{N}[j-(k-\ell)]} \right| \\
&= \sum_{j=0}^{N-1} \left(\frac{1}{\tau}\right)^{j} \left| \sum_{p=0}^{N-1} w_p Q_p e^{-i\frac{2\pi p}{N} j} \right| = \sum_{j=0}^{N-1} \left(\frac{1}{\tau}\right)^{j} |P_j|,
\end{aligned}$$

where we use MATLAB's `fft` to compute all $P_j = \sum_{p=0}^{N-1} w_p Q_p e^{-i\frac{2\pi p}{N} j}$ $(0 \leq j \leq N-1)$. Similarly, for the simplified bound in (3.3), we also use the composite trapezoid rule to approximate the integral and then call `fminbnd` in MATLAB to find the optimal $\tau > 1$.

For Markov functions, Theorem 3.5 provides the original bound in (3.6) and also the simplified bound in (3.7). Since the original bound in (3.6) involves an infinite series, we approximate it by computing the first $N = 2^{12} = 4096$ terms. For the test function $f_2(z) = e^{-\sqrt{z}}$,

ETNA
Kent State University and
Johann Radon Institute (RICAM)

since $\mu'(\zeta)$ defined in (3.5) is oscillatory, we apply integration by substitution and divide the interval of integration into several subintervals. We apply Gauss-Legendre quadrature on the first few subintervals where the quadrature values are relatively large. The remaining subintervals are approximated by a trigonometric integral. For the test function $f_3(z) = z^{-1/2}$, we divide the interval of integration into several subintervals and apply integration by substitution and Gauss-Jacobi quadrature on appropriate subintervals. We omit these technical details but point out that the quadrature can be evaluated accurately with efficiency. Compared with MATLAB's `integral` with default setting, numerical experiments show that our quadrature for approximating c_j is more accurate and faster.

EXAMPLE 5.1. Consider a diagonal (symmetric) matrix $A \in \mathbb{R}^{100001\times100001}$ with diagonal entries $a_{kk} = -\cos\left(\frac{\pi k}{100000}\right)\frac{10^3-10^{-3}}{2} + \frac{10^3+10^{-3}}{2}$, for $0 \le k \le 100000$. We compute several iterations of the RKSM for approximating $f(A)b$ with 4 different functions.

We test an analytic hyperbolic sine function $f_0(z) = \sin(hz) = \frac{e^{hz}-e^{-hz}}{2}$ for $h = 0.01$, and we use one repeated single pole $s = -m/h = -4000$ for 40 steps of the RKSM. We also test the analytic exponential function $f_1(z) = e^{-z}$ and use the repeated single pole $s = m = -40$ suggested in [53]. Comparisons between the actual values of $\left|e_k^* K_m^{-1} f(A_m)e_1\right|$ and the two upper bounds in Theorem 3.2 are reported in the upper left and upper right plots in Figure 5.1 for $f_0(z)$ and $f_1(z)$, respectively. Both upper bounds give accurate estimates of the actual values. Compared to the bounds for $|e_k^* A_m e_\ell|$ and $|e_k^* f(A_m)e_\ell|$ investigated for analytic functions in [57], our approach with composite trapezoid quadrature and FFT evaluates the upper bounds efficiently with higher accuracy.

For the Markov functions $f_2(z) = e^{-\sqrt{z}}$ and $f_3(z) = z^{-1/2}$, we apply the upper bounds from Theorem 3.5, with different expressions for $\mu'(\psi(w))$, to Example 3.3 and Example 3.4, respectively. The lower left and lower right plots in Figure 5.1 display the results. We can see that for $f_3(z) = z^{-1/2}$, both upper bounds give accurate estimates of the actual values, while for $f_2(z) = e^{-\sqrt{z}}$, both upper bounds give overestimates, and the simplified bound is even further away from the actual value. This might be related to the fact that $\mu'(\psi(w)) = \frac{\sin\left(\sqrt{-\psi(w)}\right)}{\pi}$ for $f_2(z) = e^{-\sqrt{z}}$ does not decrease in absolute value but keeps oscillating infinitely many times on the interval of integration $(-\infty, \phi(0)]$.

EXAMPLE 5.2. Consider a block diagonal non-Hermitian matrix $A \in \mathbb{R}^{200002\times200002}$ with 2×2 blocks $B_k = \begin{bmatrix} c_k & d_k \\ -d_k & c_k \end{bmatrix}$ $(0 \le k \le 100000)$ along its diagonal, where

$$c_k = -\cos\left(\frac{\pi k}{100000}\right)\frac{10^3-10^{-3}}{2} + \frac{10^3+10^{-3}}{2} \qquad \text{and}$$

$$d_k = 10\sqrt{1 - \frac{(c_k - \frac{10^3+10^{-3}}{2})^2}{(\frac{10^3-10^{-3}}{2})^2}}.$$

The eigenvalues of A are located in an ellipse centered at $\frac{10^3+10^{-3}}{2} = 500.0005$, with semi-major axis of length $\frac{10^3-10^{-3}}{2} = 499.9995$ lying on the real axis and semi-minor axis of length 10. Similar to Example 5.1, we compute several iterations of the RKSM for approximating $f(A)b$ with 4 functions. From the results shown in Figure 5.2, we can see for this non-Hermitian matrix whose eigenvalues are located in an ellipse that the upper bounds we derived have similar behavior as those for the Hermitian matrix given in Example 5.1.

5.2. Comparison between the exact and inexact RKSM for approximating $f(A)b$. We first give an example showing that the inexact RKSM can track the behavior of the exact method if the error introduced at each RKSM step satisfies the bound given in Theorem 4.1.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

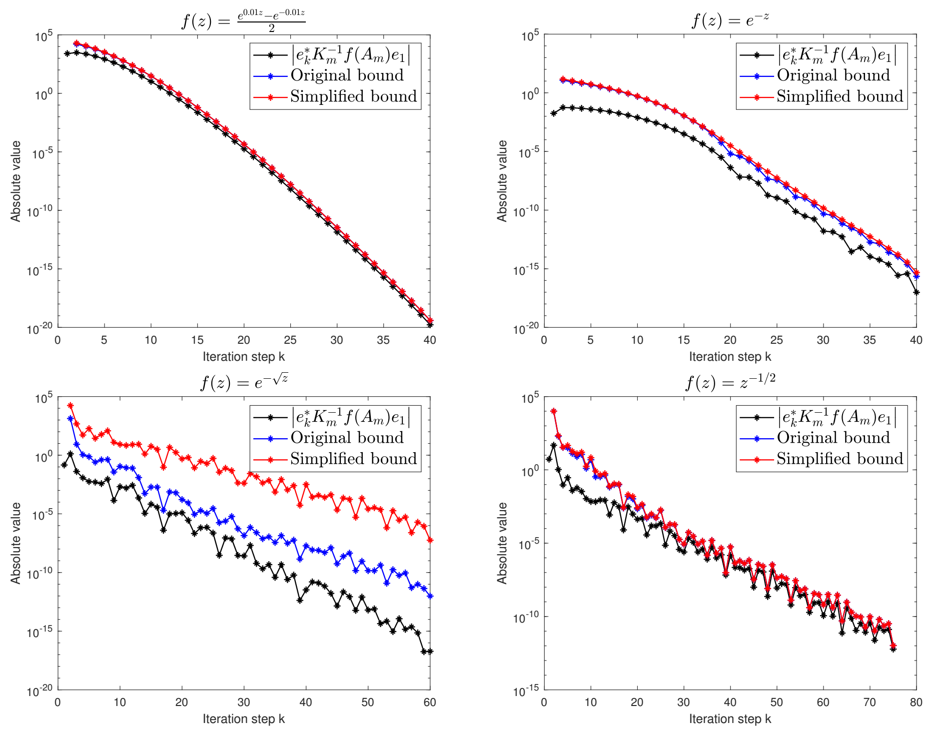

FIG. 5.1. *Comparison between the values of* $\left|e_k^* K_m^{-1} f(A_m) e_1\right|$ *and their upper bounds. Upper left:* $f_0(z) = \sinh(0.01z)$. *Upper right:* $f_1(z) = e^{-z}$. *Lower left:* $f_2(z) = e^{-\sqrt{z}}$. *Lower right:* $f_3(z) = z^{-1/2}$.

Then we consider a few practical problems for which the exact method is simulated by an inexact RKSM, where the linear solve at each RKSM step is performed to a fixed high level of accuracy, to compare with the inexact RKSM with the relaxation strategy discussed in Section 4.

EXAMPLE 5.3. We consider a non-Hermitian matrix $A \in \mathbb{R}^{23560 \times 23560}$ from the MartrixMarket problem *af23560*, whose eigenvalues are located on the right half complex plane. We test the analytic function $f_1(z) = e^{-z}$ and the Markov function $f_2(z) = e^{-\sqrt{z}}$ with the adaptive RKSM in [38, Section 4] to approximate $f(A)b$, where $b \in \mathbb{R}^{23560 \times 1}$ is a vector with standard normally distributed random entries. We apply both the exact RKSM and inexact RKSM to approximate $f(A)b$. For the exact method, the linear solves are performed by MATLAB's backslash operation, while for the inexact method, the linear systems are solved by a right-preconditioned GMRES(100) method with the relaxation strategy discussed in Theorem 4.1, where $tol = 10^{-9}$, $\xi_j = \frac{tol}{m\chi_j}$, and $\chi_k = \frac{10\|R_k\|}{|\bar{h}_{k+1,k}|}$. The preconditioner is the incomplete LU factorization preconditioner with threshold and pivoting (ILUTP) [60, Section 10.4.4, p. 327], using a drop tolerance 0.01. Figure 5.3 illustrates that if we properly set the relaxed accuracy of the approximate linear solve at each RKSM step, we can get the desired residual norm for the inexact method, and the norm of the difference between the residuals of the exact method and the inexact method remains small through the entire process of the RKSM iterations.

EXAMPLE 5.4. We test 11 nonsymmetric real matrices, all of which have the entire spectrum strictly in the right half complex plane. Two of these matrices are of the form $A = M^{-1}K$, where both M and K are sparse, but A is not formed explicitly. Specifically,

ETNA
Kent State University and
Johann Radon Institute (RICAM)

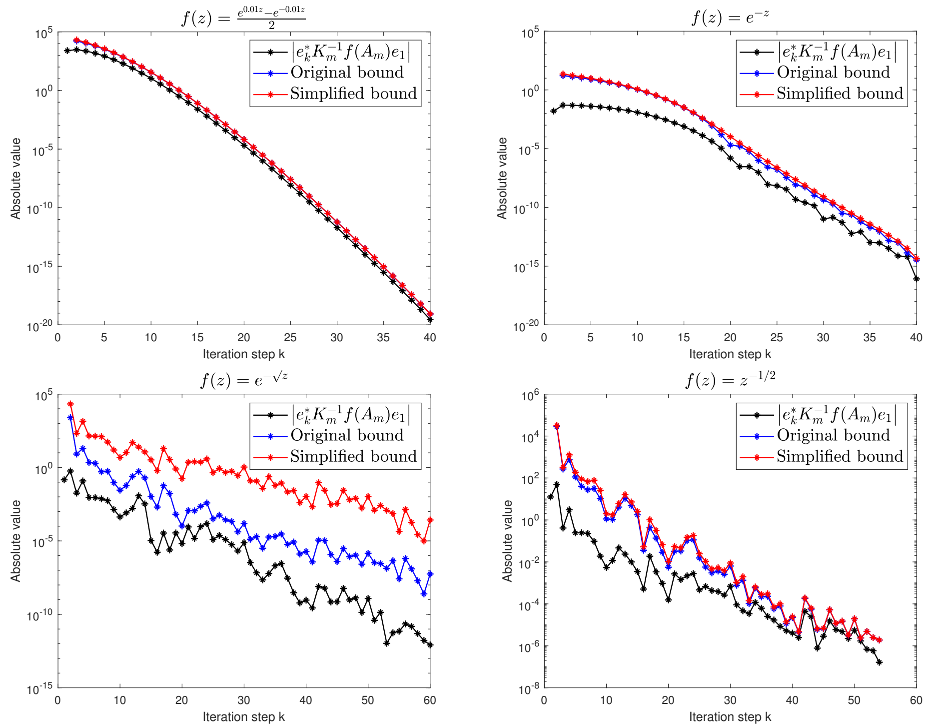

FIG. 5.2. *Comparison between the values of* $\left|e_k^* K_m^{-1} f(A_m) e_1\right|$ *and their upper bounds. Upper left:* $f_0(z) = \sinh(0.01z)$. *Upper right:* $f_1(z) = e^{-z}$. *Lower left:* $f_2(z) = e^{-\sqrt{z}}$. *Lower right:* $f_3(z) = z^{-1/2}$.

the problems *obstacle* and *plate*, which involve matrices of saddle-point structure arising from modeling incompressible fluid flows in 2D domains, are generated by the IFISS package version 3.6 [30]. The *obstacle* problem is generated with grid parameter 6, using the biquadratic-bilinear ($Q2$-$Q1$) element on a stretched rectangular grid, with viscosity parameter $\nu = \frac{1}{175}$ corresponding to a Reynolds number $Re = \frac{2}{\nu} = 350$. The *plate* problem is constructed with grid parameter 7, using the biquadratic-bilinear element on a non-stretched rectangular grid, with viscosity parameter $\nu = \frac{1}{500}$ that corresponds to a Reynolds number $Re = \frac{2}{\nu} = 1000$. Both problems give a matrix pair (K, M), where

$$K = \begin{bmatrix} F & B^T \\ B & 0 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} G & \eta B^T \\ \eta B & 0 \end{bmatrix},$$

with F being the discrete convection-diffusion operator, B^T the gradient operator for the pressure, B the divergence operator for the velocity, G the velocity mass matrix, and $\eta = 0.01$ so that the n_p (the degree of freedom of the pressure space) infinite eigenvalues of

$$\begin{bmatrix} F & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \lambda \begin{bmatrix} G & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix}$$

are mapped to $\frac{1}{\eta} = 100$ without changing the finite eigenvalues [16]. These mapped finite eigenvalues are in the deep interior of the spectrum and should have essentially no impact on the convergence of the RKSM for approximating $f(A)b$ with $A = M^{-1}K$. Such a matrix pair (K, M) has been used to study the linear stability of the steady-state solution of the

ETNA
Kent State University and
Johann Radon Institute (RICAM)

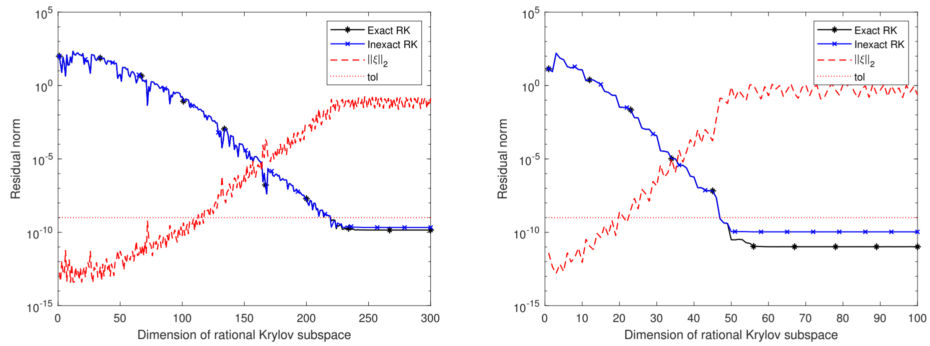

FIG. 5.3. *Comparison of the exact and inexact RKSM for approximating $f(A)b$, where A is the MatrixMarket af23560 matrix. Left: $f_1(z) = e^{-z}$. Right: $f_2(z) = e^{-\sqrt{z}}$.*

Navier-Stokes equation by matrix exponentials [58]. For these two problems, where M is not the identity, we in addition let $\xi_k = \frac{tol}{m\chi_k(1+\|M\|_2)}$ to accommodate the corresponding residual norm $\|(KV_m - MV_mA_m)f(A_m)\beta e_1\|$.

The two larger problems *LinDir2D* and *LinDir3D* are generated from finite difference discretizations of the second-order linear PDE

$$-\triangle u + \mathbf{v} \cdot \nabla u + wu = f$$

on a 2D domain $\Omega_{2D} = [0,1] \times [0,e]$ and a 3D domain $\Omega_{3D} = [0,1] \times [0,e] \times [0, \sqrt{2}\pi]$, respectively. The artificial "wind" $\mathbf{v}$ and w are defined as

$$\mathbf{v}_{2D} = \begin{bmatrix} e^{-2xy}(y^2 + 2\sin(x)) \\ \cos(4x+y)(x^3 + 3e^{-y}) \end{bmatrix}, \qquad w_{2D} = \frac{\operatorname{erf}(x-y^2)^2 + 2^{-8}}{\arctan(x^2\cos(y)) + \pi/2},$$

and

$$\mathbf{v}_{3D} = \begin{bmatrix} e^{-2xyz}(y^2 + 2z\sin(x)) \\ \cos(4x+y+2z)(x^3 + 3e^{-y} - z) \\ \ln(1+x+2y+3z)\left(x + 3\cos(z) + \frac{1}{z+\sqrt{2}\pi+0.01}\right) \end{bmatrix},$$
$$w_{3D} = \frac{\operatorname{erf}(x+z-y^2)^2 + 2^{-8}}{\arctan(x^2\cos(y)z) + \pi/2}.$$

We use a standard second-order centered finite difference to approximate the first and second derivatives based on a uniform $2^9 \times 2^{10}$ mesh grid of Ω_{2D} and a $2^6 \times 2^7 \times 2^8$ mesh grid of Ω_{3D}. Both problems are based on Dirichlet boundary condition but with the boundary nodes included in the matrix. This leads to matrices of order $(2^9+1) \times (2^{10}+1) = 525825$ for *LinDir2D* and $(2^6+1) \times (2^7+1) \times (2^8+1) = 2154945$ for *LinDir3D*, respectively. The original matrices corresponding to the linear differential operator of this PDE were scaled by $\max\{h_x, h_y\}^2$ and $\max\{h_x, h_y, h_z\}^2$, respectively, where h_x, h_y, and h_z are the mesh size in the x, y, and z directions, so that the scaled matrices have bounded norms independent of the mesh size (consistent with the assumption that $W(A) \subset E$ and $E \subset \mathbb{C}$ is compact). An application of MATLAB's backslash to solve a shifted linear system involving the scaled matrices of *LinDir2D* is still faster than our iterative method, whereas for *LinDir3D* it used up to 16 GB memory on our machine in a few minutes and led MATLAB to crash (in fact, 32 GB memory was still not sufficient to solve the linear systems involving *LinDir3D* by backslash). Other matrices are selected from the SuiteSparse Matrix Collection [17].

ETNA
Kent State University and
Johann Radon Institute (RICAM)

To compare the behavior of the inexact RKSM with and without the relaxation strategy for the inner linear solves, we use the RKSM to approximate $f(A)b$ for $f_1(z) = e^{-z}$, $f_2(z) = e^{-\sqrt{z}}$, $f_3(z) = z^{-1/2}$ and a random vector b whose entries follow a standard normal distribution. For the inexact method, we use the same strategy as in Example 5.3, and for the exact method, we let $\widetilde{\xi}_k = \min_{1\leq i\leq k} \xi_k$ to simulate the behavior of the ideal exact RKSM that performs an exact linear solve at each step. The adaptive poles of the RKSM are chosen following the strategy adopted in [38, Section 4]. We use the right-preconditioned GMRES(70) method as the inner linear solver for the RKSM. The maximum number of GMRES restart cycles is set to be $J = 20$. We use the ideal least-squares commutator (LSC) preconditioners [28, 29] for the problems *obstacle* and *plate* from IFISS. For *LinDir2D* and *LinDir3D*, the preconditioner is one W-cycle of the geometric multigrid (GMG) mtehod with two applications of the Gauss-Seidel method as pre- and post-smoothers. For all other matrices (from SuiteSparse), we use the incomplete LU preconditioner with threshold dropping and pivoting (ILUTP) preconditioners [60, Section 10.4.4]; also see MATLAB's documentation for `ilu` with the option for the `ilutp`.

The results of the performance of the exact and inexact RKSM for $f_1(z) = e^{-z}$, $f_2(z) = e^{-\sqrt{z}}$, and $f_3(z) = z^{-1/2}$ are summarized in Tables 5.1, 5.2, and 5.3, respectively. We show the size of the matrices n, the residual tolerance tol, the type of preconditioners (including the drop tolerance for ILUTP), the max number of RKSM steps m, the total number of GMRES iterations, the runtime for both the exact and the inexact methods, and the number of RKSM steps to converge. We chose the smallest tolerance tol (a negative integer power of 10) for each test matrix across all test functions such that the inexact RKSM can successfully converge to this tolerance for all functions of interest here. Such a problem-dependent tolerance is preferred to a uniform tolerance because an absolute tolerance for the residual (2.6) depends on the norm (and probably the condition number) of the matrix A. A uniform absolute tolerance similarly does not indicate the quality of approximation for different problems, nor does it show whether the computed approximation is close to the most accurate approximation achievable in double precision. Note that the total runtime includes the time for constructing the preconditioners, applying GMRES, the orthogonalization of the basis vectors of the RKSM, evaluating $f(A_m)$ for the small restricted matrices, and estimating the level of relaxation for the inner linear solves.

TABLE 5.1
Performance of the exact and inexact RKSM for $f_1(z) = e^{-z}$.

problem	size n	tol	precond-itioner	m	# GMRES iter. exact	# GMRES iter. inexact	time (secs.) exact	time (secs.) inexact	RKSM steps
af23560	23560	10^{-10}	ILUTP, 0.01	300	14353	3943	130.44	72.34	224
chipcool1	20082	10^{-11}	ILUTP, 0.01	100	4578	1891	24.48	12.62	58
obstacle	37168	10^{-11}	LSC	300	37666	20723	691.54	427.87	281
plate	37507	10^{-10}	LSC	400	35324	20556	702.77	460.10	264
venkat	62424	10^{-11}	ILUTP, 0.1	200	8754	4424	125.56	85.19	101
poli3	16955	10^{-13}	ILUTP, 0.1	100	1457	440	4.86	1.28	21
epb1	14734	10^{-12}	ILUTP, 0.01	100	3159	1135	10.63	4.11	42
goodwin030	10142	10^{-12}	ILUTP, 0.01	100	3598	1362	19.52	12.00	47
pesa	11738	10^{-10}	ILUTP, 0.01	100	5515	2643	18.89	9.21	66
LinDir2D	525825	10^{-10}	GMG	100	3672	897	988.45	205.40	61
LinDir3D	2154945	10^{-10}	GMG	100	1442	589	1080.83	385.53	49

Overall, from the results in Tables 5.1–Table 5.3, the inexact methods need fewer GMRES iterations to solve the inner linear systems, so that they need less time to converge than the "exact" RKSM. However, the level of advantage of the inexact RKSM over the exact method

TABLE 5.2
Performance of the exact and inexact RKSM for $f_2(z) = e^{-\sqrt{z}}$.

problem	size n	*tol*	precond-itioner	m	# GMRES iter. exact	# GMRES iter. inexact	time (secs.) exact	time (secs.) inexact	RKSM steps
af23560	23560	10^{-10}	ILUTP, 0.01	100	1999	794	23.63	15.26	50
chipcool1	20082	10^{-11}	ILUTP, 0.01	100	4109	1577	22.60	11.38	54
obstacle	37168	10^{-11}	LSC	100	7575	4090	140.23	84.13	62
plate	37507	10^{-10}	LSC	100	6344	3480	126.76	77.61	52
venkat	62424	10^{-11}	ILUTP, 0.1	100	5804	2880	83.11	55.73	69
poli3	16955	10^{-13}	ILUTP, 0.1	100	1231	304	3.97	0.81	18
epb1	14734	10^{-12}	ILUTP, 0.01	100	2902	1050	9.60	3.81	39
goodwin030	10142	10^{-12}	ILUTP, 0.01	100	3017	1098	16.26	9.83	40
pesa	11738	10^{-10}	ILUTP, 0.01	100	4329	2049	14.15	7.14	46
LinDir2D	525825	10^{-10}	GMG	100	2811	664	755.41	151.98	44
LinDir3D	2154945	10^{-10}	GMG	100	1898	476	1631.24	336.33	30

TABLE 5.3
Performance of the exact and inexact RKSM for $f_3(z) = z^{-1/2}$.

problem	size n	*tol*	precond-itioner	m	# GMRES iter. exact	# GMRES iter. inexact	time (secs.) exact	time (secs.) inexact	RKSM steps
af23560	23560	10^{-10}	ILUTP, 0.01	100	3874	1002	35.01	18.10	54
chipcool1	20082	10^{-11}	ILUTP, 0.01	100	4491	1777	24.19	12.62	57
obstacle	37168	10^{-11}	LSC	100	8835	4875	163.53	99.49	72
plate	37507	10^{-10}	LSC	100	7636	4110	154.99	92.60	63
venkat	62424	10^{-11}	ILUTP, 0.1	100	6461	3104	91.49	59.16	74
poli3	16955	10^{-13}	ILUTP, 0.1	100	1238	402	4.05	1.19	18
epb1	14734	10^{-12}	ILUTP, 0.01	100	2944	1151	9.71	4.06	39
goodwin030	10142	10^{-12}	ILUTP, 0.01	100	3199	1242	16.95	10.66	42
pesa	11738	10^{-10}	ILUTP, 0.01	100	4820	2722	15.84	9.36	49
LinDir2D	525825	10^{-10}	GMG	100	3314	875	894.54	211.29	49
LinDir3D	2154945	10^{-10}	GMG	100	2582	545	2233.17	388.70	38

varies for different matrices and preconditioners. In general, if the proportion of time used to construct preconditioners is small, then the relative advantage of the inexact RKSM is significant.

To demonstrate this point, we choose $f_2(z) = e^{-\sqrt{z}}$ as an example and show the computation time used for constructing the preconditioners and applying GMRES in Table 5.4. For the problem *af23560*, the exact RKSM needs 47% of the total runtime to construct the preconditioners, and the inexact RKSM requires 35% less runtime than the exact method; by comparison, for the problem *LinDir3D*, the exact RKSM takes only 1% of the total runtime to construct the preconditioners, and the inexact RKSM takes 79% less runtime than the exact method. In general, excluding the cost for constructing the preconditioners, the inexact RKSM can save about 35–81% of the runtime needed for the exact method.

6. Proofs. In this section, we prove Theorems 3.2 and 3.5 in Section 3. We use a rational approximation approach called the Faber-Dzhrbashyan (FD) rational functions introduced in [25]; see also [62, Ch. XIII, Section 3] and the references therein. Our proofs of Lemma A.1 and Theorem 3.2 largely follow the ideas of those in [57], especially the introduction to FD rational functions before the proofs of Theorems 3.2 and 3.5. To make our paper self-contained, the background details are presented in Appendix A. There are two minor differences between the materials in the appendix and in [57, Section 7]. First, a more complete description of the conditions for the expansions of the FD rational functions is presented in the appendix based on the original reference [62]. Second, a sharper upper bound is constructed in Lemma A.1 by using the least number of inequalities. In the proofs of Theorems 3.2 and 3.5, with Lemma 2.4,

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 5.4
Itemized runtime (sec.) of the exact and inexact RKSM for $f_2(z) = e^{-\sqrt{z}}$.

problem	exact RKSM constructing preconditioners	exact RKSM applying GMRES	exact RKSM total time	inexact RKSM constructing preconditioners	inexact RKSM applying GMRES	inexact RKSM total time
af23560	11.16	11.00	23.63	11.04	3.37	15.26
chipcool1	4.79	17.06	22.60	4.79	5.83	11.38
obstacle	19.79	117.61	140.23	19.49	61.91	84.13
plate	19.44	153.41	175.31	20.15	98.13	120.94
venkat	27.82	52.14	83.11	28.53	24.04	55.73
poli3	0.04	3.85	3.97	0.04	0.69	0.81
epb1	1.00	8.34	9.60	1.00	2.55	3.81
goodwin030	5.86	10.02	16.26	6.16	3.28	9.83
pesa	2.25	10.78	14.15	2.18	4.53	7.14
LinDir2D	6.33	737.34	755.41	6.29	134.75	151.98
LinDir3D	16.97	1591.79	1631.24	16.51	298.58	336.33

our derivation is developed for the entries of $K_m^{-1}f(A_m)$ instead of A_m or $f(A_m)$, and we keep the integrals of the upper bounds to provide sharper bounds and propose efficient quadrature rules to evaluate them accurately. In addition, to the best of our knowledge, Theorem 3.5 for Markov functions and its proof in this paper are new.

6.1. Proof of Theorem 3.2. In the description of Theorem 3.2, for $k, \ell \in \mathbb{N}^+$ and $\ell < k \leq m$, we define $\alpha_j = \left[\overline{\phi(s_{j+\ell-1})}\right]^{-1}$, for $1 \leq j \leq k-\ell$, and in addition we set $\alpha_j = 0$ for $j > k-\ell$. Since $s_{j+\ell-1} \in \mathbb{C} \setminus E$, we get $|\phi(s_{j+\ell-1})| > 1$, so that $|\alpha_j| < 1$. The FD rational function becomes

$$M_j^{(k,\ell)}(z) = \frac{p_j(z)}{\prod_{i=\ell}^{\ell+j}(z-s_i)}, \qquad 0 \leq j \leq k-\ell-1.$$

Define the boundary of $E_\tau = E \cup \{z \mid z \in \mathbb{C} \setminus E, |\phi(z)| \leq \tau\}$ as Γ_τ. If f is analytic in E_τ, since $W(A) \subset E \subset E_\tau$ by Definition 3.1 and (A.3), it holds that

$$\begin{aligned} f(A) &= \frac{1}{2\pi i}\int_{\Gamma_\tau} f(z)(zI-A)^{-1}dz = \frac{1}{2\pi i}\int_{|w|=\tau} f\left(\psi(w)\right)\left(\psi(w)I-A\right)^{-1}\psi'(w)dw \\ &= \frac{1}{2\pi i}\sum_{j=0}^{\infty} M_j(A)\int_{|w|=\tau}\frac{f\left(\psi(w)\right)}{w}\overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)}dw, \end{aligned}$$

yielding the following expansion:

$$f(A) = \sum_{j=0}^{\infty} c_j M_j(A), \quad \text{where} \quad c_j = \frac{1}{2\pi i}\int_{|w|=\tau}\frac{f\left(\psi(w)\right)}{w}\overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)}dw. \tag{6.1}$$

By Lemma 2.4, we have

$$e_k^* K_m^{-1} f(A_m)e_\ell = e_k^* K_m^{-1}\sum_{j=0}^{\infty} c_j M_j^{(k,\ell)}(A_m)e_\ell = e_k^* K_m^{-1}\sum_{j=k-\ell}^{\infty} c_j M_j^{(k,\ell)}(A_m)e_\ell. \tag{6.2}$$

Note that since we set $\alpha_j = 0$, for $j > k-\ell$, from (A.1) with $j \geq k-\ell$, we have

$$\begin{aligned} \overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)} &= \overline{\frac{\sqrt{1-|\alpha_{j+1}|^2}}{1-\overline{\alpha_{j+1}}\frac{1}{\overline{w}}}\prod_{i=1}^{j}\frac{\alpha_i-\frac{1}{\overline{w}}}{1-\overline{\alpha_i}\frac{1}{\overline{w}}}\frac{|\alpha_i|}{\alpha_i}} = \frac{w\sqrt{1-|\alpha_{j+1}|^2}}{w-\alpha_{j+1}}\prod_{i=1}^{j}\frac{w\overline{\alpha_i}-1}{w-\alpha_i}\frac{|\alpha_i|}{\overline{\alpha_i}} \\ &= \left(-\frac{1}{w}\right)^{j-(k-\ell)}\prod_{i=1}^{k-\ell}\frac{w\overline{\alpha_i}-1}{w-\alpha_i}\frac{|\alpha_i|}{\overline{\alpha_i}}. \end{aligned} \tag{6.3}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Combining c_j in (6.1) and (6.3), for $j \geq k-\ell$, it holds that

$$
\begin{aligned}
c_j &= \frac{1}{2\pi i}\int_{|w|=\tau} \frac{f\left(\psi(w)\right)}{w}\overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)}dw \\
&= \frac{1}{2\pi i}\int_{|w|=\tau} \frac{f\left(\psi(w)\right)}{w}\left(-\frac{1}{w}\right)^{j-(k-\ell)}\prod_{i=1}^{k-\ell}\frac{w\overline{\alpha_i}-1}{w-\alpha_i}\frac{|\alpha_i|}{\overline{\alpha_i}}dw \\
&= \frac{1}{2\pi}\int_0^{2\pi} f\left(\psi(\tau e^{i\theta})\right)\left(-\frac{1}{\tau}\right)^{j-(k-\ell)} e^{-i\theta[j-(k-\ell)]}\prod_{i=1}^{k-\ell}\frac{\tau e^{i\theta}\overline{\alpha_i}-1}{\tau e^{i\theta}-\alpha_i}\frac{|\alpha_i|}{\overline{\alpha_i}}d\theta. \qquad (6.4)
\end{aligned}
$$

Since we set $\alpha_j = 0$, for $j > k-\ell$, (A.11) implies

$$
\|M_j(A_m)\| \leq 2\frac{\sqrt{1+|\alpha_{j+1}|}}{\sqrt{1-|\alpha_{j+1}|}} + \sqrt{1-|\alpha_{j+1}|^2} = 3 \qquad (j \geq k-\ell). \tag{6.5}
$$

Note that similar to (A.11), the bound in (6.5) is valid for both analytic functions and Markov functions. Combining (6.2), (6.4), and (6.5), we get

$$
\begin{aligned}
\left|e_k^* K_m^{-1} f(A_m)e_\ell\right| &= \left|e_k^* K_m^{-1}\sum_{j=0}^{\infty} c_j M_j^{(k,\ell)}(A_m)e_\ell\right| \\
&\leq \|e_k^* K_m^{-1}\| \sum_{j=k-\ell}^{\infty} |c_j|\|M_j^{(k,\ell)}(A_m)\|\|e_\ell\| \leq 3\|e_k^* K_m^{-1}\| \sum_{j=k-\ell}^{\infty} |c_j|.
\end{aligned}
$$

The simplified bound in (3.3) can be derived as follows:

$$
\begin{aligned}
\sum_{j=k-\ell}^{\infty} |c_j| &\leq \frac{1}{2\pi}\sum_{j=k-\ell}^{\infty}\int_0^{2\pi}\left| f\left(\psi(\tau e^{i\theta})\right)\left(\frac{1}{\tau}\right)^{j-(k-\ell)} e^{-i\theta[j-(k-\ell)]}\prod_{i=1}^{k-\ell}\frac{\tau e^{i\theta}\overline{\alpha_i}-1}{\tau e^{i\theta}-\alpha_i}\right| d\theta \\
&= \frac{1}{2\pi}\int_0^{2\pi}\left| f\left(\psi(\tau e^{i\theta})\right)\prod_{i=1}^{k-\ell}\frac{\tau e^{i\theta}\overline{\alpha_i}-1}{\tau e^{i\theta}-\alpha_i}\right| d\theta \sum_{j=k-\ell}^{\infty}\left(\frac{1}{\tau}\right)^{j-(k-\ell)} \\
&= \frac{1}{2\pi}\frac{\tau}{\tau-1}\int_0^{2\pi}\left| f\left(\psi(\tau e^{i\theta})\right)\prod_{i=1}^{k-\ell}\frac{\tau e^{i\theta}\overline{\alpha_i}-1}{\tau e^{i\theta}-\alpha_i}\right| d\theta.
\end{aligned}
$$

We emphasize that c_j is independent of $\tau > 1$. In fact, from (6.4), if we define

$$
g(w) = \frac{f\left(\psi(w)\right)}{w}\left(-\frac{1}{w}\right)^{j-(k-\ell)}\prod_{i=1}^{k-\ell}\frac{w\overline{\alpha_i}-1}{w-\alpha_i}\frac{|\alpha_i|}{\overline{\alpha_i}},
$$

the residue theorem shows that

$$
c_j = \frac{1}{2\pi i}\int_{|w|=\tau} g(w)dw = \sum_{i=1}^{r}\mathrm{Res}(g, w_i),
$$

where w_i $(1 \leq i \leq r)$ denote all the poles of g in the set $\{w : |w| \leq \tau\}$. Since $\tau > 1$ and $|\alpha_i| < 1$, these poles are $0, \alpha_1, \ldots, \alpha_{k-\ell}$, which means that $\mathrm{Res}(g, w_i)$ is independent of the value of τ for $1 \leq i \leq r$. Therefore, the coefficients c_j are also independent of the value of τ. This independence of τ is important for us to develop reliable FFT-based composite trapezoid quadrature rule to evaluate these coefficients for analytic functions.

6.2. Proof of Theorem 3.5. From (3.4) and (A.3), we have

$$\begin{aligned} f(A) &= \int_{-\infty}^{0} (A-\zeta I)^{-1} d\mu(\zeta) = \int_{-\infty}^{\phi(0)} (A-\psi(w)I)^{-1} d\mu(\psi(w)) \\ &= -\int_{-\infty}^{\phi(0)} \frac{1}{w} \sum_{j=0}^{\infty} \overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)} M_j(A) \frac{1}{\psi'(w)} d\mu(\psi(w)) \\ &= -\sum_{j=0}^{\infty} M_j(A) \int_{-\infty}^{\phi(0)} \frac{1}{w} \overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)} \frac{d\mu(\psi(w))}{\psi'(w)}. \end{aligned}$$

It is possible to represent $f(A)$ in an expression similar to (6.1) with a slightly different definition of c_j:

$$f(A) = \sum_{j=0}^{\infty} c_j M_j(A), \quad \text{where} \quad c_j = -\int_{-\infty}^{\phi(0)} \overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)} \frac{1}{w} \frac{d\mu(\psi(w))}{\psi'(w)}. \tag{6.6}$$

From the definition of $\overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)}$ in (6.3) and c_j in (6.6), we have

$$c_j = \int_{-\infty}^{\phi(0)} \left(-\frac{1}{w}\right)^{j+1-(k-\ell)} \prod_{i=1}^{k-\ell} \frac{w\overline{\alpha_i}-1}{w-\alpha_i} \frac{|\alpha_i|}{\overline{\alpha_i}} \frac{d\mu(\psi(w))}{\psi'(w)}. \tag{6.7}$$

Combining the upper bounds for $\|M_j(A)\|$ in (6.5) and $|c_j|$ in (6.7) with (6.2), it holds that

$$\left|e_k^* K_m^{-1} f(A_m) e_\ell\right| \le \|e_k^* K_m^{-1}\| \sum_{j=k-\ell}^{\infty} |c_j| \|M_j^{(k,\ell)}(A)\| \|e_\ell\| \le 3 \|e_k^* K_m^{-1}\| \sum_{j=k-\ell}^{\infty} |c_j|.$$

To establish the simplified bound, we have

$$\begin{aligned} \left|e_k^* K_m^{-1} f(A_m) e_\ell\right| &\le 3\|e_k^* K_m^{-1}\| \sum_{j=k-\ell}^{\infty} \left| \int_{-\infty}^{\phi(0)} \left(\frac{1}{|w|}\right)^{j+1-(k-\ell)} \prod_{i=1}^{k-\ell} \frac{w\overline{\alpha_i}-1}{w-\alpha_i} \frac{d\mu(\psi(w))}{\psi'(w)} \right| \\ &\le 3\|e_k^* K_m^{-1}\| \sum_{j=k-\ell}^{\infty} \int_{-\infty}^{\phi(0)} \left|\frac{1}{\psi'(w)}\right| \frac{1}{|w|} \left(\frac{1}{|w|}\right)^{j-(k-\ell)} \left| \prod_{i=1}^{k-\ell} \frac{w\overline{\alpha_i}-1}{w-\alpha_i} \right| |d\mu(\psi(w))| \\ &\le 3\|e_k^* K_m^{-1}\| \int_{-\infty}^{\phi(0)} \left|\frac{1}{\psi'(w)}\right| \frac{1}{|w|} \frac{|w|}{|w|-1} \left| \prod_{i=1}^{k-\ell} \frac{w\overline{\alpha_i}-1}{w-\alpha_i} \right| |d\mu(\psi(w))| \\ &\le 3\|e_k^* K_m^{-1}\| \int_{-\infty}^{\phi(0)} \left|\frac{1}{\psi'(w)}\right| \frac{1}{|w+1|} \left| \prod_{i=1}^{k-\ell} \frac{w\overline{\alpha_i}-1}{w-\alpha_i} \right| |d\mu(\psi(w))|. \end{aligned}$$

7. Conclusion. In this paper, we studied the residual of the RKSM for approximating the action of a function of a matrix $f(A)$ to a vector b. We explored the decay bounds for the off-diagonal entries of a restricted matrix that arise in the RKSM approximation of $f(A)b$ for analytic functions and Markov functions. For the inexact RKSM, upper bounds for the allowable errors for the inner linear solves are derived, and a heuristic tolerance relaxation strategy is proposed to enable that the inexact RKSM keeps track of the convergence of the exact RKSM. Numerical experiments show that the inexact RKSM can exhibit a convergence behavior similar to that of the exact method, but it entails lower computational cost thanks to the relaxed accuracy for the inner linear systems.

Appendix A. This appendix provides an introduction to the Faber-Dzhrbashyan (FD) rational functions for Theorems 3.2 and 3.5. To this end, we first review the definition of the Takenaka-Malmquist (TM) system of rational functions:

$$\varphi_1(w) = \frac{\sqrt{1-|\alpha_1|^2}}{1-\overline{\alpha_1}w},$$

$$\varphi_j(w) = \frac{\sqrt{1-|\alpha_j|^2}}{1-\overline{\alpha_j}w}\prod_{k=1}^{j-1}\frac{\alpha_k - w}{1-\overline{\alpha_k}w}\frac{|\alpha_k|}{\alpha_k}, \qquad j \geq 2, \tag{A.1}$$

where $\alpha_k = \left[\overline{\phi(z_k)}\right]^{-1} \in D = \{w : |w| \leq 1\}$ $(k \geq 1)$ and $\{z_k\}_{k=1}^{\infty}$ is a sequence of points that all lie in the exterior of E. Here $\frac{|\alpha_k|}{\alpha_k}$ is defined as 1 if $\alpha_k = 0$. Since $\phi(z_k)$ is in the exterior of D, it implies that $\left|\left[\overline{\phi(z_k)}\right]^{-1}\right| < 1$. The Takenaka-Malmquist systems $\{\varphi_n(w)\}_{n=0}^{\infty}$ form an orthonormal basis on the subspace $\mathcal{T} = \{w \in \mathbb{C} : |w| = 1\}$, i.e.,

$$\langle \varphi_m(w), \varphi_n(w)\rangle := \frac{1}{2\pi}\int_0^{2\pi}\varphi_m(e^{it})\overline{\varphi_n(e^{it})}dt = \delta_{mn}, \qquad m,n \in \mathbb{N},$$

where δ_{mn} is the Kronecker delta; see, e.g., [54].

The FD rational function $M_j(z)$ is defined as the sum of the principal part and the constant in the Laurent decomposition of $\varphi_j(\phi(z))$ in the neighborhoods of the points $\{z_k\}_{k=1}^{j+1}$. Therefore, $M_j(z)$ can be represented in the form

$$M_j(z) = \frac{p_j(z)}{\prod_{k=1}^{j+1}(z-z_k)}, \qquad j \geq 0,$$

where $p_j(z)$ is a polynomial with degree no higher than j. If we define Γ_D as the preimage of the unit disk under the map $w = \phi(z)$ and G_D denotes the interior of the boundary Γ_D, then the FD rational functions can also be represented in the form (see [62, Section 13, equation (4)]):

$$M_j(z) = \frac{1}{2\pi i}\int_{\Gamma_D}\frac{\varphi_{j+1}\left[\phi(\zeta)\right]}{\zeta - z}d\zeta, \qquad z \in G_D. \tag{A.2}$$

If the conditions

$$\sum_{k=1}^{\infty}(1-|\alpha_k|) = +\infty, \qquad \text{and} \qquad \lim_{r\to 1^+}\int_0^{2\pi}\left|\psi'(re^{i\theta})\right|^2 d\theta < \infty$$

hold, then we have the following expansions:

$$\frac{\psi'(w)}{\psi(w)-z} = \frac{1}{w}\sum_{j=0}^{\infty}\overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)}M_j(z), \qquad z \in G,\ |w| > 1,$$

where G is the interior of $E \supset W(A)$; see, e.g., [62, p. 259]. If $W(A) \subset G$, it follows that

$$\psi'(w)\left(\psi(w)I - A\right)^{-1} = \frac{1}{w}\sum_{j=0}^{\infty}\overline{\varphi_{j+1}\left(\frac{1}{\overline{w}}\right)}M_j(A), \qquad |w| > 1. \tag{A.3}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Our next step is to derive the upper bounds for both $\left|\overline{\varphi_j\left(\frac{1}{\overline{w}}\right)}\right|$ and $\|M_j(A)\|$.

LEMMA A.1. *From the definition of the Takenaka-Malmquist system of functions in* (A.1)*, for* $\tau > 1$ *and* $|\alpha_j| < 1$*, for all* $j \geq 1$*, it holds that*

$$\max_{|w|=\tau}\left|\overline{\varphi_j\left(\frac{1}{\overline{w}}\right)}\right| \leq \frac{\tau\sqrt{1-|\alpha_j|^2}}{\tau-|\alpha_j|}\prod_{k=1}^{j-1}\frac{\tau|\alpha_k|+1}{\tau+|\alpha_k|}.$$

Proof. First, if $j = 1$, we have from (A.1) that

$$\left|\overline{\varphi_1\left(\frac{1}{\overline{w}}\right)}\right| = \left|\frac{\sqrt{1-|\alpha_1|^2}}{1-\overline{\alpha_1}\frac{1}{\overline{w}}}\right| \leq \frac{\tau\sqrt{1-|\alpha_1|^2}}{\tau-|\alpha_1|}.$$

For $j > 1$ and $|w| = \tau$, it holds that

$$\left|\overline{\varphi_j\left(\frac{1}{\overline{w}}\right)}\right| = \left|\frac{\sqrt{1-|\alpha_j|^2}}{1-\overline{\alpha_j}\frac{1}{\overline{w}}}\prod_{k=1}^{j-1}\frac{\alpha_k-\frac{1}{\overline{w}}}{1-\overline{\alpha_k}\frac{1}{\overline{w}}}\frac{|\alpha_k|}{\alpha_k}\right| = \frac{\tau\sqrt{1-|\alpha_j|^2}}{|\overline{w}-\overline{\alpha_j}|}\prod_{k=1}^{j-1}\left|\frac{\overline{w}\alpha_k-1}{\overline{w}-\overline{\alpha_k}}\right|. \tag{A.4}$$

Let $w = \tau e^{\theta_0 i}$ and $\alpha_j = \rho_j e^{\theta_j i}$ for some $\rho_j = |\alpha_j| \in [0, 1)$. We define

$$\begin{aligned} g_j(\theta_0) &:= |\overline{w}-\overline{\alpha_j}| = \left|\tau e^{-\theta_0 i}-\rho_j e^{-\theta_j i}\right| = \left|\tau e^{(\theta_j-\theta_0)i}-\rho_j\right| \\ &= \sqrt{\tau^2+\rho_j^2-2\tau\rho_j\cos(\theta_j-\theta_0)} \geq \tau-\rho_j. \end{aligned} \tag{A.5}$$

We also define

$$h_k(\theta_0) := \left|\frac{\overline{w}\alpha_k-1}{\overline{w}-\overline{\alpha_k}}\right|^2 = \left|\frac{\tau\rho_k e^{(\theta_k-\theta_0)i}-1}{\tau e^{-\theta_0 i}-\rho_k e^{-\theta_k i}}\right|^2 = \frac{\tau^2\rho_k^2+1-2\tau\rho_k\cos(\theta_k-\theta_0)}{\tau^2+\rho_k^2-2\tau\rho_k\cos(\theta_k-\theta_0)}.$$

Since

$$\begin{aligned} &(\tau^2-1)(\rho_k^2-1) = \tau^2\rho_k^2-\tau^2-\rho_k^2+1 < 0 \Longrightarrow \tau^2\rho_k^2+1 < \tau^2+\rho_k^2 \qquad \text{and} \\ &\tau^2+\rho_k^2-2\tau\rho_k\cos(\theta_k-\theta_0) \geq (\tau-\rho_k)^2 > 0, \end{aligned}$$

it is easy to show that $h_k(\theta_0)$ achieves its maximum when $\theta_k - \theta_0 = \pi$. Therefore,

$$\max h_k(\theta_0) = h_k(\theta_k-\pi) = \frac{(\tau\rho_k+1)^2}{(\tau+\rho_k)^2}. \tag{A.6}$$

Combining (A.5) and (A.6) into (A.4), we get

$$\max_{|w|=\tau}\left|\overline{\varphi_j\left(\frac{1}{\overline{w}}\right)}\right| \leq \frac{\tau\sqrt{1-|\alpha_j|^2}}{\tau-|\alpha_j|}\prod_{k=1}^{j-1}\frac{\tau|\alpha_k|+1}{\tau+|\alpha_k|}.$$

For $j = 1$, it is easy to verify that the above inequality also holds. $\square$

We can write the FD rational functions $M_j(z)$ with the Faber transformation of a Takenaka-Malmquist system. For every function f continuous on the boundary of D and analytic in the interior of D, the Faber transformation is defined as

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

see, e.g., [36]. By letting $\zeta = \psi(w)$ in (A.2) with w on the unit circle, we get

$$M_j(z) = \mathcal{F}(\varphi_{j+1})(z), \qquad z \in G,\ j \geq 0.$$

Define the modified Faber operator $\mathcal{F}_+(f)(z) := \mathcal{F}(f)(z) + f(0)$. It has been proved in [3] that for any matrix A such that $W(A) \subset E$,

$$\|\mathcal{F}_+(f)(A)\| = \|\mathcal{F}(f)(A) + f(0)I\| \leq 2 \sup_{w\in D} |f(w)|,$$

where f is analytic in the interior of D and continuous on D. Then,

$$\begin{aligned} \|M_j(A)\| &= \|\mathcal{F}(\varphi_{j+1})(A)\| = \|\mathcal{F}_+(\varphi_{j+1})(A) - \varphi_{j+1}(0)I\| \\ &\leq 2 \sup_{w\in D} |\varphi_{j+1}(w)| + |\varphi_{j+1}(0)| \,. \end{aligned} \tag{A.7}$$

As in Lemma A.1, if $|\alpha_j| < 1$, it can be concluded analogously that

$$\max_{|w|=1/\tau} |\varphi_{j+1}(w)| \leq \frac{\tau\sqrt{1-|\alpha_{j+1}|^2}}{\tau - |\alpha_{j+1}|} \prod_{k=1}^{j} \frac{\tau|\alpha_k|+1}{\tau+|\alpha_k|}. \tag{A.8}$$

It is easy to show that the right-hand side of (A.8) decreases when τ increases, so

$$\sup_{w\in D} |\varphi_{j+1}(w)| \leq \frac{\sqrt{1-|\alpha_{j+1}|^2}}{1-|\alpha_{j+1}|} \prod_{k=1}^{j} \frac{|\alpha_k|+1}{1+|\alpha_k|} = \sqrt{\frac{1+|\alpha_{j+1}|}{1-|\alpha_{j+1}|}}. \tag{A.9}$$

We also know from (A.1) that

$$|\varphi_{j+1}(0)| = \sqrt{1-|\alpha_{j+1}|^2} \prod_{k=1}^{j} |\alpha_k| \leq \sqrt{1-|\alpha_{j+1}|^2}. \tag{A.10}$$

Combining (A.9) and (A.10) into (A.7), we obtain

$$\|M_j(A)\| \leq 2\sqrt{\frac{1+|\alpha_{j+1}|}{1-|\alpha_{j+1}|}} + \sqrt{1-|\alpha_{j+1}|^2}. \tag{A.11}$$

Both upper bounds for $\max_{|w|=\tau} \left|\varphi_j\left(\frac{1}{w}\right)\right|$ in Lemma A.1 and $\|M_j(A)\|$ in (A.11) are fundamentals for the proofs of Theorems 3.2 and 3.5 in Section 3.

REFERENCES

[1] M. Afanasjew, M. Eiermann, O. G. Ernst, and S. Güttel, *Implementation of a restarted Krylov subspace method for the evaluation of matrix functions*, Linear Algebra Appl., 429 (2008), pp. 2293–2314.
[2] Z. Bai, *Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems*, Appl. Numer. Math., 43 (2002), pp. 9–44.
[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] M. Benzi and P. Boito, *Decay properties for functions of matrices over C^*-algebras*, Linear Algebra Appl., 456 (2014), pp. 174–198.
[5] M. Benzi and G. H. Golub, *Bounds for the entries of matrix functions with applications to preconditioning*, BIT Numer. Math., 39 (1999), pp. 417–438.
[6] M. Benzi and N. Razouk, *Decay bounds and $O(n)$ algorithms for approximating functions of sparse matrices*, Electron. Trans. Numer. Anal., 28 (2007/08), pp. 16–39.
https://etna.ricam.oeaw.ac.at/vol.28.2007-2008/pp16-39.dir/pp16-39.pdf

[7] M. BENZI AND V. SIMONCINI, *Decay bounds for functions of Hermitian matrices with banded or Kronecker structure*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1263–1282.
[8] D. BERTACCINI AND F. DURASTANTE, *Computing function of large matrices by a preconditioned rational Krylov method*, in Numerical Mathematics and Advanced Applications—ENUMATH 2019, F. J. Vermolen and C. Vuik, eds., vol. 139 of Lect. Notes Comput. Sci. Eng., Springer, Cham, 2021, pp. 343–351.
[9] G. BIROS AND O. GHATTAS, *Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization. I. The Krylov-Schur solver*, SIAM J. Sci. Comput., 27 (2005), pp. 687–713.
[10] M. BOTCHEV, L. KNIZHNERMAN, AND M. SCHWEITZER, *Krylov subspace residual and restarting for certain second order differential equations*, Preprint on arXiv, 2022. https://arxiv.org/abs/2206.06909
[11] M. A. BOTCHEV, V. GRIMM, AND M. HOCHBRUCK, *Residual, restarting, and Richardson iteration for the matrix exponential*, SIAM J. Sci. Comput., 35 (2013), pp. A1376–A1397.
[12] M. A. BOTCHEV, L. KNIZHNERMAN, AND E. E. TYRTYSHNIKOV, *Residual and restarting in Krylov subspace evaluation of the φ function*, SIAM J. Sci. Comput., 43 (2021), pp. A3733–A3759.
[13] A. BOURAS AND V. FRAYSSÉ, *Inexact matrix-vector products in Krylov methods for solving linear systems: a relaxation strategy*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 660–678.
[14] K. BURRAGE, N. HALE, AND D. KAY, *An efficient implicit FEM scheme for fractional-in-space reaction-diffusion equations*, SIAM J. Sci. Comput., 34 (2012), pp. A2145–A2172.
[15] C. CANUTO, V. SIMONCINI, AND M. VERANI, *On the decay of the inverse of matrices that are sum of Kronecker products*, Linear Algebra Appl., 452 (2014), pp. 21–39.
[16] K. A. CLIFFE, T. J. GARRATT, AND A. SPENCE, *Eigenvalues of block matrices arising from problems in fluid mechanics*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 1310–1318.
[17] T. A. DAVIS AND Y. HU, *The University of Florida sparse matrix collection*, ACM Trans. Math. Software, 38 (2011), Art. 1, 25 pages.
[18] N. DEL BUONO, L. LOPEZ, AND R. PELUSO, *Computation of the exponential of large sparse skew-symmetric matrices*, SIAM J. Sci. Comput., 27 (2005), pp. 278–293.
[19] S. DEMKO, W. F. MOSS, AND P. W. SMITH, *Decay rates for inverses of band matrices*, Math. Comp., 43 (1984), pp. 491–499.
[20] K. N. DINH AND R. B. SIDJE, *Analysis of inexact Krylov subspace methods for approximating the matrix exponential*, Math. Comput. Simulation, 138 (2017), pp. 1–13.
[21] V. DRUSKIN AND L. KNIZHNERMAN, *Krylov subspace approximation of eigenpairs and matrix functions in exact and computer arithmetic*, Numer. Linear Algebra Appl., 2 (1995), pp. 205–217.
[22] ———, *Extended Krylov subspaces: approximation of the matrix square root and related functions*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 755–771.
[23] V. DRUSKIN, L. KNIZHNERMAN, AND M. ZASLAVSKY, *Solution of large scale evolutionary problems using rational Krylov subspaces with optimized shifts*, SIAM J. Sci. Comput., 31 (2009), pp. 3760–3780.
[24] V. DRUSKIN AND V. SIMONCINI, *Adaptive rational Krylov subspaces for large-scale dynamical systems*, Systems Control Lett., 60 (2011), pp. 546–560.
[25] M. M. DŽRBAŠYAN, *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.
[26] M. EIERMANN AND O. G. ERNST, *A restarted Krylov subspace method for the evaluation of matrix functions*, SIAM J. Numer. Anal., 44 (2006), pp. 2481–2504.
[27] M. EIERMANN, O. G. ERNST, AND S. GÜTTEL, *Deflated restarting for matrix functions*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 621–641.
[28] H. ELMAN, V. E. HOWLE, J. SHADID, R. SHUTTLEWORTH, AND R. TUMINARO, *Block preconditioners based on approximate commutators*, SIAM J. Sci. Comput., 27 (2006), pp. 1651–1668.
[29] H. ELMAN, V. E. HOWLE, J. SHADID, D. SILVESTER, AND R. TUMINARO, *Least squares preconditioners for stabilized discretizations of the Navier-Stokes equations*, SIAM J. Sci. Comput., 30 (2007/08), pp. 290–311.
[30] H. C. ELMAN, A. RAMAGE, AND D. J. SILVESTER, *IFISS: a computational laboratory for investigating incompressible flow problems*, SIAM Rev., 56 (2014), pp. 261–273.
[31] R. W. FREUND, *Krylov-subspace methods for reduced-order modeling in circuit simulation*, J. Comput. Appl. Math., 123 (2000), pp. 395–421.
[32] A. FROMMER, S. GÜTTEL, AND M. SCHWEITZER, *Convergence of restarted Krylov subspace methods for Stieltjes functions of matrices*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1602–1624.
[33] ———, *Efficient and stable Arnoldi restarts for matrix functions based on quadrature*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 661–683.
[34] A. FROMMER, C. SCHIMMEL, AND M. SCHWEITZER, *Bounds for the decay of the entries in inverses and Cauchy-Stieltjes functions of certain sparse, normal matrices*, Numer. Linear Algebra Appl., 25 (2018), Art. e2131, 17 pages.
[35] A. FROMMER AND V. SIMONCINI, *Stopping criteria for rational matrix functions of Hermitian and symmetric matrices*, SIAM J. Sci. Comput., 30 (2008), pp. 1387–1412.

[36] D. GAIER, *The Faber operator and its boundedness*, J. Approx. Theory, 101 (1999), pp. 265–277.
[37] V. GRIMM AND M. HOCHBRUCK, *Rational approximation to trigonometric operators*, BIT Numer. Math., 48 (2008), pp. 215–229.
[38] S. GÜTTEL, *Rational Krylov approximation of matrix functions: numerical methods and optimal pole selection*, GAMM-Mitt., 36 (2013), pp. 8–31.
[39] S. GÜTTEL AND L. KNIZHNERMAN, *A black-box rational Arnoldi variant for Cauchy-Stieltjes matrix functions*, BIT Numer. Math., 53 (2013), pp. 595–616.
[40] S. GÜTTEL AND M. SCHWEITZER, *A comparison of limited-memory Krylov methods for Stieltjes functions of Hermitian matrices*, SIAM J. Matrix Anal. Appl., 42 (2021), pp. 83–107.
[41] Y. HASHIMOTO AND T. NODERA, *Inexact rational Krylov method for evolution equations*, BIT Numer. Math., 61 (2021), pp. 473–502.
[42] N. J. HIGHAM, *Functions of Matrices*, SIAM, Philadelphia, 2008.
[43] M. HOCHBRUCK AND C. LUBICH, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.
[44] ———, *Exponential integrators for quantum-classical molecular dynamics*, BIT Numer. Math., 39 (1999), pp. 620–645.
[45] M. HOCHBRUCK AND A. OSTERMANN, *Exponential Runge-Kutta methods for parabolic problems*, Appl. Numer. Math., 53 (2005), pp. 323–339.
[46] L. KNIZHNERMAN AND V. SIMONCINI, *A new investigation of the extended Krylov subspace method for matrix function evaluations*, Numer. Linear Algebra Appl., 17 (2010), pp. 615–638.
[47] D. KRESSNER AND C. TOBLER, *Krylov subspace methods for linear systems with tensor product structure*, SIAM J. Matrix Anal. Appl., 31 (2009/10), pp. 1688–1714.
[48] P. KÜRSCHNER AND M. A. FREITAG, *Inexact methods for the low rank solution to large scale Lyapunov equations*, BIT Numer. Math., 60 (2020), pp. 1221–1259.
[49] R. B. LEHOUCQ AND K. MEERBERGEN, *Using generalized Cayley transformations within an inexact rational Krylov sequence method*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 131–148.
[50] L. LOPEZ AND V. SIMONCINI, *Analysis of projection methods for rational function approximation to the matrix exponential*, SIAM J. Numer. Anal., 44 (2006), pp. 613–635.
[51] A. M. LUKATSKII, *On the system of rational functions of M. M. Dzhrbashyan for an arbitrary continuum*, Sib. Math. J., 15 (1974), pp. 147–152.
[52] N. MASTRONARDI, M. NG, AND E. E. TYRTYSHNIKOV, *Decay in functions of multiband matrices*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2721–2737.
[53] I. MORET AND P. NOVATI, *RD-rational approximations of the matrix exponential*, BIT Numer. Math., 44 (2004), pp. 595–615.
[54] M. PAP AND F. SCHIPP, *Malmquist-Takenaka systems and equilibrium conditions*, Math. Pannon., 12 (2001), pp. 185–194.
[55] M. POPOLIZIO AND V. SIMONCINI, *Acceleration techniques for approximating the matrix exponential operator*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 657–683.
[56] S. POZZA AND V. SIMONCINI, *Inexact Arnoldi residual estimates and decay properties for functions of non-Hermitian matrices*, BIT Numer. Math., 59 (2019), pp. 969–986.
[57] ———, *Functions of rational Krylov space matrices and their decay properties*, Numer. Math., 148 (2021), pp. 99–126.
[58] M. W. ROSTAMI AND F. XUE, *Robust linear stability analysis and a new method for computing the action of the matrix exponential*, SIAM J. Sci. Comput., 40 (2018), pp. A3344–A3370.
[59] Y. SAAD, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 29 (1992), pp. 209–228.
[60] ———, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, PA, 2003.
[61] 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.
[62] P. K. SUETIN, *Series of Faber Polynomials*, Gordon and Breach, Amsterdam, 1998.
[63] H. WANG AND Q. YE, *Error bounds for the Krylov subspace methods for computations of matrix exponentials*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 155–187.
[64] S. WANG, E. DE STURLER, AND G. H. PAULINO, *Large-scale topology optimization using preconditioned Krylov subspace methods with recycling*, Internat. J. Numer. Methods Engrg., 69 (2007), pp. 2441–2468.
[65] S. WYATT, *Issues in Interpolatory Model Reduction: Inexact Solves, Second-order Systems and DAEs*, PhD. Thesis, Virginia Polytechnic Institute and State University, Blacksburg, ProQuest LLC, Ann Arbor, 2012.
[66] S. XU AND F. XUE, *Inexact rational Krylov subspace method for eigenvalue problems*, Numer. Linear Algebra Appl., 29 (2022), Art. e2437, 25 pages.