SIAM J. SCI. COMPUT.
Vol. 44, No. 4, pp. A2843–A2870

THE SHORT-TERM RATIONAL LANCZOS METHOD AND APPLICATIONS*

DAVIDE PALITTA†, STEFANO POZZA‡, AND VALERIA SIMONCINI§

Abstract. Rational Krylov subspaces have become a reference tool in dimension reduction procedures for several application problems. When data matrices are symmetric, a short-term recurrence can be used to generate an associated orthonormal basis. In the past this procedure was abandoned because it requires twice the number of linear system solves per iteration compared with the classical long-term method. We propose an implementation that allows one to obtain the rational subspace reduced matrices at lower overall computational costs than proposed in the literature by also conveniently combining the two system solves. Several applications are discussed where the short-term recurrence feature can be exploited to avoid storing the whole orthonormal basis. We illustrate the advantages of the proposed procedure with several examples.

Key words. rational Krylov, rational Lanczos, short-term recurrence

MSC codes. 65F30, 15B57, 41A20

DOI. 10.1137/21M1403254

1. Introduction.

Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$ and a unit norm vector $v \in \mathbb{R}^n$, we are interested in analyzing the algebraic recurrence that generates the rational Krylov subspace

$$\mathcal{K}_m(A, v, \boldsymbol{\xi}_m) = \operatorname{span}\left\{ v, (I - \xi_1^{-1}A)^{-1}v, \ldots, \prod_{j=1}^{m-1} (I - \xi_j^{-1}A)^{-1}v \right\}, \tag{1.1}$$

and in the applicability of the computed quantities; here $\boldsymbol{\xi}_m = [\xi_1, \ldots, \xi_{m-1}]$ with $\xi_i \neq 0$ are such that $I - \xi_i^{-1}A$ is nonsingular for $i = 1, \ldots, m-1$. In the considered applications A is definite (either positive or negative), for which strong theoretical arguments for the selection of the ξ_i have been discussed; see [23, 34] and references therein. In our context, the corresponding theoretical setting suggests that the eigenvalues of A and ξ_i have opposite signs, a hypothesis that we will assume throughout. By relying on efficient sparse solvers for linear systems, rational Krylov subspaces have become a major tool in a variety of application problems, including eigenvalue approximation, dynamical system reduction, matrix equation solution, and matrix

*Submitted to the journal's Methods and Algorithms for Scientific Computing section March 8, 2021; accepted for publication (in revised form) March 31, 2022; published electronically August 31, 2022.

https://doi.org/10.1137/21M1403254

Funding: The first and third authors are members of Indam-GNCS, whose support is gratefully acknowledged. This work was also supported by Charles University Research programs PRIMUS/21/SCI/009 and UNCE/SCI/023.

†Research Group Computational Methods in Systems and Control Theory (CSC), Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany (palitta@mpi-magdeburg.mpg.de). Current address: Junior Assistant Professor at Dipartimento di Matematica, Alma Mater Studiorum - Università di Bologna, 40126, Bologna, Italy (davide.palitta@unibo.it).

‡Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic (pozza@karlin.mff.cuni.cz)

§Dipartimento di Matematica and AM^2, Alma Mater Studiorum - Università di Bologna, Piazza di Porta S. Donato, 5, I-40127 Bologna, Italy, and IMATI-CNR, 27100 Pavia, Italy (valeria.simoncini@unibo.it).

function and bilinear form evaluations [6, 21, 22, 28, 32, 33, 51]. An orthonormal basis $\{q_1, \dots, q_m\}$ for $\mathcal{K}_m(A, v, \boldsymbol{\xi}_m)$ can be determined by using the Gram–Schmidt procedure as $q_1 = v$ and

$$\hat{q}_{j+1} := \left(I - \xi_j^{-1} A\right)^{-1} q_j, \qquad h_{j+1,j} q_{j+1} = \hat{q}_{j+1} - Q_j h_j,$$

where $h_j = Q_j^T \hat{q}_{j+1}$, $Q_j = [q_1, \dots, q_j]$, and $h_{j+1,j}$ is the normalization factor for q_{j+1} for $j = 1, \dots, m$ [51, eq. (4.15)]. This can be viewed as a rational variant of the Arnoldi iteration. The parameters (or shifts) $\boldsymbol{\xi}_m = [\xi_1, \dots, \xi_{m-1}]$ can be computed a priori or determined adaptively as the space grows. The vector h_j contains the orthogonalization coefficients so that Q_j has orthonormal columns for $j = 1, \dots, m$.

For A symmetric, in [18] a short-term recurrence was introduced to generate an orthonormal basis of rational functions associated with $\mathcal{K}_m(A, v, \boldsymbol{\xi}_m)$, from which a short-term recurrence can be derived for the basis $\{q_1, \dots, q_m\}$ [33].

The short-term recurrence yielding an orthonormal basis of the *polynomial* Krylov subspace generated by a symmetric A was introduced by Lanczos [40], and it is thus referred to as the Lanczos iteration. This recurrence allows one to store a few n-dimensional vectors, leading to major savings in various approximation problems where the whole basis is otherwise not needed. Thanks to the work in [18, 33], similar advantages can be envisioned in the rational case. Güttel in [33, p. 46] says:

> "Note that in each iteration of Algorithm 2 two linear systems with $I - A/\xi_j$ need to be solved [...]. Hence, this algorithm is in general not competitive with the rational Arnoldi algorithm if the poles ξ_j vary often. Moreover, we will make explicit use of the orthogonality of the rational Krylov basis V_{m+1} when computing Rayleigh–Ritz approximations for $f(A)b$ (see Chapter 6). In this case full orthogonalization of V_{m+1} is required anyway and one cannot take advantage of the short recurrence."

In spite of the elegant derivation, these considerations led Güttel to discard the short-term iteration in his application setting.

We claim that in spite of the extra cost per iteration, the rational short-term recurrence does provide an appealing framework for a variety of approximation problems. Our contribution is two-fold. First, we propose a new implementation of the short-term recurrence that (i) alleviates the computational costs associated with the two solves by combining them into a single linear system solve with multiple right-hand sides; (ii) derives the entries of the reduced matrix $J_m := Q_m^T A Q_m$ as the iterations proceed, without using Q_m or explicitly solving $m \times m$ linear systems. Second, we illustrate the advantages of the obtained implementation in the numerical treatment of several application problems that do not require the whole basis matrix $Q_m = [q_1, \dots, q_m]$; these include the approximation of quadratic and bilinear forms in general, and quantities of interest in control, such as the estimation of the $\mathcal{H}_2$-norm of (parametric) linear time-invariant systems and of the optimal feedback control function.We will refer to this implementation as the Q_m-less computation.

We also start a discussion on the behavior of the obtained recurrence in finite precision arithmetic. Our matrix relations and experimental evidence seem to suggest that the short-term iteration is affected by round-off error accumulations similar to those of the classical Lanczos method. A deeper analysis of this crucial aspect deserves dedicated research, which will be postponed until future work.

The synopsis of this paper is as follows. In section 2 we revisit the rational Lanczos iteration. An efficient, basis-free procedure to compute the matrix $J_m = Q_m^T A Q_m$

is derived in section 2.1, while in section 2.2 the novel implementation of the rational Lanczos method is illustrated. A panel of applications where the rational Lanczos method can be successfully employed is presented in section 3. These include numerical approximations of quadratic and bilinear forms (section 3.1), matrix function trace estimation (section 3.2), $\mathcal{H}_2$-norm computation for linear time-invariant (LTI) systems (section 3.3), and linear-quadratic regulator (LQR) feedback control approximations (section 3.4). In section 4 some preliminary remarks on the behavior of the rational Lanczos method in finite precision arithmetic are reported. Our conclusions are given in section 5, while in the appendix the *block* rational Lanczos algorithm is given.

2. The short-term rational Krylov iteration. Based on the thorough analysis of orthogonal rational functions in [13], the authors of [18, Thm. 4.2] developed a three-term recurrence relation to generate a sequence of orthogonal rational functions associated with the rational Krylov subspace (1.1); see also [18]. This elegant construction was further developed in [33, section 5.2], leading to the following vector recurrence for $j \geq 1$:

$$\beta_j \left(I - \xi_j^{-1} A\right) q_{j+1} = Aq_j - \alpha_j \left(I - \xi_{j-1}^{-1} A\right) q_j - \beta_{j-1} \left(I - \xi_{j-2}^{-1} A\right) q_{j-1}, \tag{2.1}$$

where, with the usual convention that $1/\infty = 0$, $\xi_{-1} := \infty$, $\xi_0 := \infty$, $\beta_0 = 0$, $q_0 = 0$, and $q_1 = v/\|v\|$. By setting

$$r := (I - \xi_j^{-1} A)^{-1}(Aq_j + \beta_{j-1}\xi_{j-2}^{-1} Aq_{j-1}) - \beta_{j-1} q_{j-1}, \quad s := (I - \xi_j^{-1} A)^{-1}(I - \xi_{j-1}^{-1} A) q_j,$$

the coefficients α_j and β_j are computed as $\alpha_j = (r^T q_j)/(s^T q_j)$ and $\beta_j = \|r - \alpha_j s\|$. A simple rearrangement of the terms leads to the following more compact notation:

$$A[q_{j-1}, q_j, q_{j+1}] \begin{bmatrix} \frac{\beta_{j-1}}{\xi_{j-2}} \\ \frac{\alpha_j}{\xi_{j-1}} + 1 \\ \frac{\beta_j}{\xi_j} \end{bmatrix} = [q_{j-1}, q_j, q_{j+1}] \begin{bmatrix} \beta_{j-1} \\ \alpha_j \\ \beta_j \end{bmatrix},$$

which, after m iterations, gives the Arnoldi-like relation

$$AQ_{m+1}\underline{K}_m = Q_{m+1}\underline{H}_m, \quad \text{with} \quad \underline{K}_m = \underline{I}_m + D_m^{-1}\underline{H}_m \in \mathbb{R}^{(m+1)\times m}, \tag{2.2}$$

where $\underline{I}_m$ is the $(m+1) \times m$ identity matrix and $D_m = \operatorname{diag}(\xi_0, \ldots, \xi_m)$. Hence

$$\underline{K}_m = \begin{bmatrix} 1 & 0 & & & \\ \frac{\beta_1}{\xi_1} & 1+\frac{\alpha_2}{\xi_1} & \frac{\beta_2}{\xi_1} & & \\ & \frac{\beta_2}{\xi_2} & 1+\frac{\alpha_3}{\xi_2} & \ddots & \\ & & \ddots & \ddots & \frac{\beta_{m-1}}{\xi_{m-2}} \\ & & & \frac{\beta_{m-1}}{\xi_{m-1}} & 1+\frac{\alpha_m}{\xi_{m-1}} \\ & & & & \frac{\beta_m}{\xi_m} \end{bmatrix}, \quad \underline{H}_m = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \beta_2 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_{m-1} \\ & & & \beta_{m-1} & \alpha_m \\ & & & & \beta_m \end{bmatrix}.$$

The relation in (2.2) can also be written as

$$AQ_m K_m = Q_m H_m + \beta_m \left(I - \xi_m^{-1} A\right) q_{m+1} e_m^T, \tag{2.3}$$

with K_m and H_m the leading $m \times m$ upper parts of $\underline{K}_m$ and $\underline{H}_m$, respectively. Here and in the following, e_j denotes the jth column of the identity matrix, whose dimension is clear from the context. Hence, we resume the standard form for the rational

Krylov iteration [51]. This means that in exact arithmetic the rational Lanczos and rational Arnoldi recurrences compute the same basis, provided the same set of shifts is employed in the basis construction.

Thanks to the irreducibility of $\underline{H}_m$, it follows from (2.3) that if the matrix K_m is singular, then $[Q_m, (I - \xi_m^{-1}A)q_{m+1}]$ is not full rank, which occurs if and only if the subspace $\mathcal{K}_m(A, v, \boldsymbol{\xi}_m)$ is A-invariant, i.e., $A\mathcal{K}_m(A, v, \boldsymbol{\xi}_m) = \mathcal{K}_m(A, v, \boldsymbol{\xi}_m)$. In the A-invariant case we have a lucky termination of the algorithm. Hereafter, we always assume K_m to be invertible. As for the standard Lanczos procedure, the matrix $J_m = Q_m^T A Q_m$ represents the projection and restriction of A onto the range space of Q_m. However, while in the (standard) Arnoldi process this matrix is tridiagonal, the matrix J_m is generally full. A detailed analysis of the structure and decay properties of the entries of J_m can be found in [49]; see also [50]. The matrix J_m appears in projection methods for solving problems such as linear and quadratic matrix equations and matrix functions evaluations. Classical implementations of the rational Lanczos recurrence rely on the whole matrix Q_m. In the following we show that this can be avoided, leading to memory savings in a variety of application problems.

2.1. On the computation of $J_j = Q_j^T A Q_j$. In this section we derive a recurrence for computing the small dimensional matrix $J_j = Q_j^T A Q_j$ without using Q_j or explicitly solving small size linear systems with K_j at each iteration j.

Setting $J_{j+1} = Q_{j+1}^T A Q_{j+1}$ and multiplying (2.3) by Q_{j+1}^T from the left, we obtain

$$J_{j+1}\underline{K}_j = \underline{H}_j. \tag{2.4}$$

Considering the first j rows we can write $J_j K_j + w_j e_j^T = H_j$ with $w_j = Q_j^T A q_{j+1} \frac{\beta_j}{\xi_j}$, that is, except for the last column, the matrix $J_j K_j$ is tridiagonal, and

$$J_j = H_j K_j^{-1} - w_j e_j^T K_j^{-1}. \tag{2.5}$$

To get a Q_j-less computation of J_j we need to obtain a different expression for w_j. Setting $u = Q_j^T A q_{j+1}$ and exploiting symmetry, from (2.4) we have

$$\begin{bmatrix} J_j & u \\ u^T & \eta \end{bmatrix} \begin{bmatrix} I \\ \beta_j/\xi_j e_j^T K_j^{-1} \end{bmatrix} = \underline{H}_j K_j^{-1}, \tag{2.6}$$

with $w_j = u\beta_j/\xi_j$. For the last row it holds that $u^T = e_{j+1}^T \underline{H}_j K_j^{-1} - \eta\beta_j/\xi_j e_j^T K_j^{-1}$, which gives the sought-after expression for w_j. Summarizing, at step $j+1$ we can completely define J_j without storing the whole Q_j. In particular, its last column is given by

$$\begin{aligned} J_j e_j =& H_j K_j^{-1} e_j - w_j e_j^T K_j^{-1} e_j \\ =& H_j K_j^{-1} e_j - \left(K_j^{-T} \underline{H}_j^T e_{j+1} - K_j^{-T} e_j \eta \beta_j/\xi_j\right) \beta_j/\xi_j e_j^T K_j^{-1} e_j \\ =& H_j K_j^{-1} e_j - K_j^{-T} e_j \left(\xi_j - \eta\right) \frac{\beta_j^2}{\xi_j^2} \left(e_j^T K_j^{-1} e_j\right). \end{aligned} \tag{2.7}$$

The most expensive steps in (2.7) are the solution of the linear systems with K_j and K_j^T. The tridiagonal structure of these matrices allows us to derive a recurrence for the two solution vectors as the iterations proceed, making the overall computation cheaper than explicitly solving the linear systems from scratch at each iteration j.

LEMMA 2.1. *With the previous notation, for $j = 1, \dots, m$ the solutions (y_j, t_j) to the systems $K_j y = e_j$ and $K_j^T t = e_j$ can be obtained via the following recurrences:*

$$y_j = \frac{1}{\omega_j}\left(e_j - \frac{\beta_{j-1}}{\xi_{j-2}}\begin{bmatrix} y_{j-1} \\ 0 \end{bmatrix}\right), \quad t_j = \frac{1}{\omega_j}\left(e_j - \frac{\beta_{j-1}}{\xi_{j-1}}\begin{bmatrix} t_{j-1} \\ 0 \end{bmatrix}\right), \quad j \geq 2, \tag{2.8}$$

with $y_1 = 1$ and $t_1 = 1$, where $\omega_j \in \mathbb{R}$ is given by the following recursive formula:

$$\omega_1 = 1, \quad \omega_2 = \frac{\alpha_2}{\xi_1} + 1, \quad \omega_j = \frac{\alpha_j}{\xi_{j-1}} + 1 - \frac{\beta_{j-1}^2}{\xi_{j-1}\xi_{j-2}\omega_{j-1}}, \ j \geq 3. \tag{2.9}$$

Proof. We focus on the computation of $y_j = K_j^{-1} e_j$. The computation of $t_j = K_j^{-T} e_j$ is analogous. Under the assumption that $\mathcal{K}_j(A, v, \boldsymbol{\xi}_j)$ is not A-invariant, K_i is a nonsingular matrix for $i = 1, \dots, j$. Hence the LU factorization $K_j = L_j U_j$ with no pivoting exists, $y_j = U_j^{-1} L_j^{-1} e_j$, and the factors are given by

$$L_j = \begin{bmatrix} 1 & & & & \\ \ell_2 & 1 & & & \\ & \ell_3 & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ell_j & 1 \end{bmatrix} \in \mathbb{R}^{j\times j}, \quad U_j = \begin{bmatrix} \omega_1 & 0 & & & \\ & \omega_2 & \beta_2/\xi_1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \beta_{j-1}/\xi_{j-2} \\ & & & & \omega_j \end{bmatrix} \in \mathbb{R}^{j\times j}.$$

Thanks to the structure of L_j, $x = L_j^{-1} e_j = e_j$. To solve $U_j y = e_j$ we first determine the diagonal elements of U_j. Direct computation gives the recursion in (2.9), that is, the computation of ω_j only requires information available in the current subspace. Moreover, if $y_{j-1} \in \mathbb{R}^{j-1}$ is such that $K_{j-1} y_{j-1} = e_{j-1}$, the solution y_j to $U_j y = e_j$ can be derived as in (2.8). Analogously, the solution t_j to $K_j^T t = e_j$ is obtained as in (2.8), where $t_{j-1} \in \mathbb{R}^{j-1}$ is such that $K_{j-1}^T t_{j-1} = e_{j-1}$. ∎

THEOREM 2.2. *With the notation and results of Lemma 2.1, at the jth iteration the last column (or row) of $J_j = Q_j^T A Q_j$ is given by*

$$J_j e_j = \widehat{y}_j - \frac{\beta_j^2}{\xi_j^2}\frac{\xi_j - \eta}{\omega_j} t_j,$$

where t_j is defined as in Lemma 2.1 while $\widehat{y}_j := H_j y_j$ satisfies

$$\widehat{y}_1 = \alpha_1, \quad \widehat{y}_j = \begin{bmatrix} -\widehat{y}_{j-1}\frac{\beta_{j-1}}{\xi_{j-2}\omega_j} \\ \beta_{j-1} e_{j-1}^T y_j + \frac{\alpha_j}{\omega_j} \end{bmatrix} + \frac{\beta_{j-1}}{\omega_j} e_{j-1} \quad \text{for } j > 1.$$

Proof. By plugging the expressions of $y_j = K_j^{-1} e_j$ and $t_j = K_j^{-T} e_j$ given in Lemma 2.1 into (2.7), we get

$$J_j e_j = H_j K_j^{-1} e_j - K_j^{-T} e_j (\xi_j - \eta)\frac{\beta_j^2}{\xi_j^2}\left(e_j^T K_j^{-1} e_j\right) = H_j y_j - \frac{\beta_j^2}{\xi_j^2}\frac{\xi_j - \eta}{\omega_j} t_j.$$

Using Lemma 2.1 and the tridiagonal structure of H_j we can write $\widehat{y}_j = H_j y_j$ as

$$\widehat{y}_j = \begin{bmatrix} -H_{j-1} y_{j-1}\frac{\beta_{j-1}}{\xi_{j-2}\omega_j} \\ \beta_{j-1} e_{j-1}^T y_j + \frac{\alpha_j}{\omega_j} \end{bmatrix} + \frac{\beta_{j-1}}{\omega_j} e_{j-1} = \begin{bmatrix} -\widehat{y}_{j-1}\frac{\beta_{j-1}}{\xi_{j-2}\omega_j} \\ \beta_{j-1} e_{j-1}^T y_j + \frac{\alpha_j}{\omega_j} \end{bmatrix} + \frac{\beta_{j-1}}{\omega_j} e_{j-1}. \qquad ∎$$

Algorithm 2.1. Q_m-less rational Lanczos.

input : $A \in \mathbb{R}^{n\times n}$, $v \in \mathbb{R}^n$, $m \in \mathbb{N}$, $m > 0$, $\boldsymbol{\xi}_m = [\xi_1, \dots, \xi_m]$.
output: $J_m \in \mathbb{R}^{m\times m}$ s.t. $J_m = Q_m^T A Q_m$, where $\mathrm{Range}(Q_m) = \mathcal{K}_m(A, v, \boldsymbol{\xi}_m)$.

1 Set $\widehat{q} = v/\|v\|$, $\xi_{-1} := \infty, \xi_0 := \infty, \beta_0 = 0, \bar{q} = 0$
2 **for** $j = 1, \dots, m$ **do**
3 Set $\widetilde{r} = A\widehat{q} - \beta_{j-1}(I - A/\xi_{j-2})\bar{q}$ and $\widetilde{s} = (I - A/\xi_{j-1})\widehat{q}$
4 Solve $(I - \frac{1}{\xi_j}A)[r, s] = [\widetilde{r}, \widetilde{s}]$
5 Compute $\alpha_j = \frac{r^T\widehat{q}}{s^T\widehat{q}}$
6 Set $q = r - \alpha_j s$
7 Set $\beta_j = \|q\|$, $\bar{q} = \widehat{q}$, $\widehat{q} = q/\beta_j$
8 **if** j=1 **then**
9 Set $\omega_1 = y_1 = t_1 = 1$ and $\widehat{y}_1 = \alpha_1$
10 **else**
11 Set $\omega_j = \alpha_j/\xi_{j-1} + 1 - \beta_{j-1}^2/(\xi_{j-1}\xi_{j-2}\omega_{j-1})$
12 Set $y_j = \begin{bmatrix} -y_{j-1}\frac{\beta_{j-1}}{\xi_{j-2}\omega_j} \\ \frac{1}{\omega_j} \end{bmatrix}$, $t_j = \begin{bmatrix} -t_{j-1}\frac{\beta_{j-1}}{\xi_{j-1}\omega_j} \\ \frac{1}{\omega_j} \end{bmatrix}$, and $\widehat{y}_j = \begin{bmatrix} -\widehat{y}_{j-1}\frac{\beta_{j-1}}{\xi_{j-2}\omega_j} \\ \beta_{j-1}e_{j-1}^T y_j + \frac{\alpha_j}{\omega_j} \end{bmatrix} + \frac{\beta_{j-1}}{\omega_j}e_{j-1}$
13 Compute $\eta = \widehat{q}^T A\widehat{q}$
14 Set $J_{1:j,j} = \widehat{y}_j - \frac{\beta_j^2}{\xi^2}\frac{\xi_j - \eta}{\omega_j}t_j$ and $J_{j,1:j} = J_{1:j,j}^T$

Theorem 2.2 also shows that by storing the low dimensional vectors y_{j-1}, $\widehat{y}_{j-1}$, and t_{j-1}, along with some additional scalar quantities, the allocation of the matrices H_j and K_j can be avoided.

The following proposition shows that the computation of $K_j^{-1}e_j$ and $K_j^{-T}e_j$ in Lemma 2.1 by means of the LU factorization is backward stable. This result ensures that using Gaussian elimination does not introduce any instability in the update of J_j in Theorem 2.2.

PROPOSITION 2.3. *Assume that the elements α_j and β_j of the matrix K_j are computed exactly. Moreover, let the matrix A be symmetric positive (negative) definite, and the shifts ξ_j be negative (positive). If the unit roundoff is small enough, then the solutions of the systems $K_j y = e_j$ and $K_j^T t = e_j$ computed by the recurrences* (2.8) *and* (2.9) *are backward stable.*

The proof is a direct consequence of the stability analysis in Proposition 4.6.

2.2. The Q_m-less procedure. The implementation of the proposed memory saving method is summarized in Algorithm 2.1. Step 4 relies on the fact that solving a single linear system with p right-hand sides is more efficient than sequentially solving p systems with the same coefficient matrix. Indeed, assuming for instance, that a sparse direct solver is used, the symbolic analysis phase and the factorization step can be performed once for all the considered right-hand sides. The same gains might be obtained in the sequential solution of the two systems if a very fine tuning of the adopted linear solver is possible. Nevertheless, also in the latter scenario, the block solution strategy is still advantageous thanks to a better computer handling of the dense kernels involved in the solution process. Moreover, the coefficient factors need to be accessed only once avoiding an increment in the storage requirements.

In Algorithm 2.1 we suppose that the shifts are given. Alternatively, dynamic shift computation strategies can be easily incorporated in the algorithm; see, e.g., [23, 24, 34] for different shift selection strategies. Since A is symmetric, all shifts can be taken to be real.

Algorithm 2.1 should be equipped with a stopping criterion that must not involve the whole basis Q_m. Such a stopping criterion depends on the application of interest, and different instances are discussed in section 3.

Remark 2.4. Algorithm 2.1 can be generalized by replacing the starting vector v with a full column rank matrix $V \in \mathbb{R}^{n\times p}$, $p > 1$. This generates the *block* rational Krylov subspace $\mathcal{K}_m(A, V, \boldsymbol{\xi}_m) = \text{range}([V, (I-\xi_1^{-1}A)^{-1}V, \ldots, \prod_{j=1}^{m-1}(I-\xi_j^{-1}A)^{-1}V])$. In this construction, many of the scalar quantities involved in Algorithm 2.1 are replaced by $p \times p$ matrices. Also in this case, the matrix $J_m = Q_m^T A Q_m \in \mathbb{R}^{mp\times mp}$ can still be computed Q_m-less at low computational cost. We include the corresponding implementation as Algorithm A.1 in the appendix. Analogously to the matrix form (2.3), the recurrences in Algorithm A.1 can be written as

$$AQ_mK_m = Q_mH_m + \left(I - \xi_m^{-1}A\right)\widehat{Q}_{m+1}\beta_m E_m^T, \tag{2.10}$$

where $\widehat{Q}_{m+1} \in \mathbb{R}^{n\times p}$, $Q_m = [\widehat{Q}_1, \ldots, \widehat{Q}_m] \in \mathbb{R}^{n\times mp}$, $K_m, H_m \in \mathbb{R}^{mp\times mp}$ are block tridiagonal matrices with $(p\times p)$-size blocks, $\beta_m \in \mathbb{R}^{p\times p}$, and $E_m = e_m \otimes I_p \in \mathbb{R}^{mp\times p}$. We also have the block counterpart of (2.5), that is,

$$J_m = H_mK_m^{-1} - W_mE_m^TK_m^{-1}, \quad W_m = (\xi_m)^{-1}Q_m^TA\widehat{Q}_{m+1}\beta_m. \tag{2.11}$$

3. Applications. In this section we illustrate the applicability of the Q_m-less rational Krylov algorithm to a variety of problems. All numerical results have been obtained by running MATLAB R2017b [42] on a standard node[1] of the Linux cluster `mechthild` hosted at the Max Planck Institute for Dynamics of Complex Technical Systems in Magdeburg, Germany.

3.1. Quadratic and bilinear forms. Consider a function f defined on the spectrum of the symmetric matrix A, and vectors u, v. The approximation of the bilinear form $u^T f(A)\, v$ (or quadratic form for $u = v$) arises in many applications including network analysis [25], regularization problems [26], electronic structure calculations [52], solution of PDEs [39], Gaussian processes [48], and many others [4, 9, 29]. Bilinear forms can also be used to estimate the trace of $f(A)$; see section 3.2.

Given a large and sparse A, the use of (polynomial) Krylov subspace methods for the approximation of $u^T f(A)\, v$ is well established and grounded in a theoretical framework comprising orthogonal polynomials and Gauss quadrature [29]. Under the assumptions that $u = v$ and $\|v\| = 1$, the mth Lanczos iteration produces an $m \times m$ tridiagonal matrix T_m (known as the *Jacobi* matrix) giving the approximation $v^T f(A)\, v \approx e_1^T f(T_m)e_1$. Such an approximation relies on the so-called *moment matching property*, that is, $v^T A^j\, v = e_1^T (T_m)^j e_1$, $j = 0, \ldots, 2m-1$, or, equivalently, $v^T p(A)\, v = e_1^T p(T_m)e_1$ for every polynomial $p(x)$ of degree at most $2m-1$. Such a property is connected with the Gauss quadrature approximation for a Riemann–Stieltjes integral determined by A and v; see, e.g., [29, 41].

An analogous result can be given for rational Krylov subspaces. Let Q_m and J_m be the matrices associated with the rational Krylov subspace $\mathcal{K}_m(A, v, \boldsymbol{\xi}_m)$. Given

[1] CPU: 2x Intel Xeon Skylake Silver 4110 @ 2.1 GHz, 8 cores per CPU. RAM: 192 GB DDR4 ECC. See also https://www.mpi-magdeburg.mpg.de/cluster/mechthild.

$q(x) = \prod_{j=1}^{m-1}(1 - x/\xi_j)$ and a polynomial $p = p(x)$ of degree at most $m-1$, it holds that

$$p(A)q(A)^{-1}\, v = Q_m p(J_m) q(J_m)^{-1} e_1 \in \mathcal{K}_m(A, v, \boldsymbol{\xi}_m); \tag{3.1}$$

see, e.g., [22, Lemma 3.1], [33, Lemma 4.6]. Left multiplication by $Q_m Q_m^T A$ yields $Q_m Q_m^T\, A\, p(A)q(A)^{-1}\, v = Q_m\, J_m\, p(J_m) q(J_m)^{-1} e_1$, and, hence, by linearity

$$Q_m Q_m^T p_m(A) q(A)^{-1}\, v = Q_m p_m(J_m) q(J_m)^{-1} e_1 \tag{3.2}$$

for every polynomial p_m of degree at most m. The following proposition extends the moment matching property to the rational case by using ideas borrowed from Vorobjev's moment problem (see, e.g., [41, section 3.7.1]). The result can also be obtained as an application of the results in [28, Thm. 2]. However, our approach provides a short alternative proof that, to the best of our knowledge, has not yet appeared in the literature.

PROPOSITION 3.1. *With the previous notation for $\mathcal{K}_m(A, v, \boldsymbol{\xi}_m)$, Q_m, and J_m, let $q(x) = \prod_{j=1}^{m-1}(1 - x/\xi_j)$. Then, for every polynomial $p(x)$ of degree at most $2m-1$,*

$$v^T p(A) q(A)^{-2}\, v = e_1^T p(J_m) q(J_m)^{-2} e_1.$$

Proof. For every $p_{m-1}(x)$ of degree at most $m-1$, (3.1) gives

$$v^T p_{m-1}(A) q(A)^{-1}\, v = e_1^T p_{m-1}(J_m) q(J_m)^{-1} e_1.$$

Consider a polynomial p_m of degree at most m. Given that $(Q_m Q_m^T)^2 = Q_m Q_m^T$, (3.2) becomes $Q_m Q_m^T \left(Q_m p_m(J_m) q(J_m)^{-1} e_1 - p_m(A) q(A)^{-1}\, v\right) = 0$, implying that the vector $Q_m p_m(J_m) q(J_m)^{-1} e_1 - p_m(A) q(A)^{-1}\, v$ is orthogonal to $\mathcal{K}_m(A, v, \boldsymbol{\xi}_m)$. Therefore, for any polynomial $p_{m-1}(x)$ of degree at most $m-1$ we get

$$v^T p_{m-1}(A) q(A)^{-1} \left(Q_m p_m(J_m) q(J_m)^{-1} e_1 - p_m(A) q(A)^{-1}\, v\right) = 0$$

(where we have used the symmetry of A), that is,

$$v^T p_{m-1}(A) q(A)^{-1} Q_m p_m(J_m) q(J_m)^{-1} e_1 = v^T p_{m-1}(A) p_m(A) (q(A)^{-1})^2\, v.$$

This equality concludes the proof since by (3.1) it holds that $v^T p_{m-1}(A) q(A)^{-1} = e_1^T p_{m-1}(J_m) q(J_m)^{-1} Q_m^T$. ∎

Proposition 3.1 holds for every orthogonalization process of a rational Krylov subspace, i.e., for every orthogonal basis. As a consequence, it is not related to short recurrences. We also remark that similar properties have been derived in [50, Thm. 3.1] for a different kind of rational Krylov subspaces. Extensions to the nonsymmetric case have also been studied; see, e.g., [19, 28] among many others.

Thanks to Algorithm 2.1, we can compute the matrix J_m by means of the Q_m-less short-term recurrence rational method, that is, we can compute the approximation

$$e_1^T f(J_m) e_1 \approx v^T f(A)\, v. \tag{3.3}$$

The approximation error can be characterized by adapting the results in [17] to our case. Indeed, since A is symmetric we can interpret the bilinear form as a Riemann–Stieltjes integral and the approximant (3.3) as a rational Gauss quadrature, that is,

$$v^T f(A) v = \int f(\lambda) \mathrm{d}\mu(\lambda) \approx \sum_{j=1}^{m} f(\lambda_j)\theta_j = e_1^T f(J_m) e_1,$$

where $\mu(\lambda)$ is a measure depending on the spectrum and the eigenvectors of A, and the λ_j's, θ_j's are the eigenvalues and eigenvectors of J_m, respectively (see the results in [29, Chapter 7], which can be easily adapted to this rational case). In this framework, the approximation (3.3) is a rational quadrature rule. Therefore, by [17, eq. (4)],

$$|e_1^T f(J_m)e_1 - v^T f(A)\, v| \le 2\|v\|^2 \min_{\deg(p)\le 2m-1} \left|v^T \left(f(A) - p(A)q(A)^{-2}\right) v\right|, \tag{3.4}$$

where we used the notation of Proposition 3.1 (cf. [33, Thm. 4.10] related to a similar but different problem).

The block case can be treated analogously. Consider the $n \times p$ matrix V and the $n \times n$ symmetric matrix A. Then using Algorithm A.1 we get the block $mp \times mp$ matrix J_m and, hence, by setting $E_1^T = [I_p, 0, \ldots, 0] \in \mathbb{R}^{p \times mp}$ we obtain the approximation

$$E_1^T f(J_m)E_1 \approx V^T f(A)V. \tag{3.5}$$

Algorithm 2.1 can also approximate a bilinear form $u^T f(A)\, v$ with $u \neq v$. We describe various alternative strategies. The first one is to rewrite the problem as

$$u^T f(A)\, v = \frac{1}{4}\left((u+v)^T f(A)\,(u+v) - (u-v)^T f(A)(u-v)\right),$$

and run Algorithm 2.1 twice [29, section 7.3]. Such a strategy maintains the same exactness of Proposition 3.1 at twice the cost. The second one considers the vector $u_m = Q_m^T\, u$ (computed on the fly) and the approximation

$$u_m^T f(J_m)e_1 \approx u^T f(A)v. \tag{3.6}$$

This approximant is exact for rational functions whose numerator has a degree up to $m-1$ and denominator q from (3.1). The third possibility uses (3.5) applied to the 2×2 block bilinear form

$$[\ u\ \ v\ ]^T\, f(A)\, [\ u\ \ v\ ] = \begin{bmatrix} u^T f(A)u & u^T f(A)v \\ v^T f(A)u & v^T f(A)v \end{bmatrix},$$

whose (1,2) position yields the sought-after quantity; see, e.g., [39, eqs. (6)–(7)]. Finally, another possibility is to consider the rational variant of the nonsymmetric Lanczos algorithm in [60].

Stopping criteria. For a general function f a cheap stopping criterion at the mth iteration is given by the difference between two iterates

$$|u_m^T f(J_m)e_1 - u_{m-s}^T f(J_{m-s})e_1|$$

for some fixed index s satisfying $1 \le s < m$. This criterion relies on the idea that the approximation error decreases as the iterations proceed. For the special case of the extended Krylov subspace and Laplace–Stieltjes functions the convergence to $f(A)v$ is indeed monotonic [53], hence the criterion is reliable. This simple criterion can be further developed following the results in [14].

A "residual-based" criterion can be obtained if the function f is such that $y(\tau) = f(\tau A)v$ is the solution to the differential equation $y^{(d)} = Ay$, with $y^{(d)}$ the dth derivative of y, $d \in \mathbb{N}$, and specified initial conditions for $\tau = 0$. Indeed, let $y_m(\tau) = Q_m f(\tau J_m)e_1$ be the approximant derived by (3.1) and define the differential equation residual

$$r_m(\tau) = Ay_m(\tau) - y_m^{(d)}(\tau). \tag{3.7}$$

Then the norm of $r_m(1)$ is commonly used as stopping criterion for Krylov subspace approximations to $y(1)$; see, e.g., [11, 20, 38]. Computing $\|r_m(1)\|$ would require storing Q_m; however, it is possible to use an upper bound with quantities available at the current step. Indeed, using (2.3), (2.5), and $J_m f(\tau J_m)e_1 = (f(\tau J_m))^{(d)}e_1$, we get

$$\begin{aligned} r_m(\tau) =& Ay_m(\tau) - y_m^{(d)}(\tau) = AQ_m f(\tau J_m)e_1 - Q_m(f(\tau J_m))^{(d)}e_1 \\ =& \left(Q_m H_m + \beta_m(I - \xi_m^{-1}A)q_{m+1}e_m^T\right) K_m^{-1} f(\tau J_m)e_1 - Q_m(f(\tau J_m))^{(d)}e_1 \\ =& Q_m J_m f(\tau J_m)e_1 + Q_m w_m e_m^T K_m^{-1} f(\tau J_m)e_1 \\ &+ \beta_m(I - \xi_m^{-1}A)q_{m+1}e_m^T K_m^{-1} f(\tau J_m)e_1 - Q_m(f(\tau J_m))^{(d)}e_1 \\ =& (I - Q_m Q_m^T)(I - \xi_m^{-1}A)q_{m+1}\beta_m e_m^T K_m^{-1} f(\tau J_m)e_1. \end{aligned}$$

Therefore,

$$\|r_m(\tau)\| \le |\beta_m|\,(1 + |\xi_m^{-1}|\,\|A\|)\,|e_m^T K_m^{-1} f(\tau J_m)e_1|, \tag{3.8}$$

and this holds, in particular, for $\tau = 1$. We recall that $t_m^T = e_m^T K_m^{-1}$ is computed iteratively during the recurrence (see Lemma 2.1), hence the only extra computational cost is given by the norm and the inner product $|t_m^T f(\tau J_m)e_1|$. As already mentioned, the approximant y_m is exact for rational functions with a numerator of degree at most $m-1$ and denominator q from (3.1), while (3.3) is exact on a much larger set of rational functions; cf. (3.1) with Proposition 3.1. Therefore, the previous stopping criterion may overestimate the error of (3.3), while it is more appropriate in (3.6) where $v \neq u$.

The previous stopping criteria can be extended to the block case. For the latter one, we can derive the bound $\|R_m(\tau)\| \le \|\beta_m\|\,(1 + |\xi_m^{-1}|\,\|A\|)\,\|E_m^T K_m^{-1} f(\tau J_m)E_1\|$. To prove the inequality above, consider the differential equation $Y^{(d)} = AY$, with $Y(\tau) = f(\tau A)V$, and the approximant $Y_m(\tau) = Q_m f(\tau J_m)E_1$. The bound follows by using the formulas (2.10) and (2.11) and adapting the scalar case arguments seen above to the block case.

Example 3.2. Let A be the adjacency matrix of a network. For any $i \in \{1, \dots, n\}$ the quantity $(\exp(A))_{ii}$ measures the importance of the ith node with respect to the network edge structure, the so-called exp-*centrality index* [25]. We consider the symmetric normalized adjacency matrices *Oregon*-1 and *ca-GrQc* of size 11492 and 5242, respectively, from the SuiteSparse Matrix Collection [16] and the node i with the largest exp-centrality. Both matrices are very sparse, with, respectively, 4 and 5 nonzero elements per row on average. We are interested in approximating the bilinear form $e_i^T \exp(A - 2I)e_i$ (note that $\exp(A) = \exp(A - 2I)\exp(2)$, with $A - 2I$ negative definite). As a quality measure, we use the error between the approximation obtained by the rational Lanczos/Arnoldi methods and the quantity obtained by the MATLAB function `expm`.

Figure 3.1 reports the absolute errors for both approximation methods as the iterations proceed, until the final accuracy is attained after which, not surprisingly, the full orthogonalization approach shows higher accuracy; see, e.g., Figure 3.1 (right). See also section 4. The figure also shows the norm of the rational Lanczos residual (3.7) for $\tau = 1$ and its estimate (3.8). In Table 3.1, we report the number of iterations and corresponding CPU times, confirming the similar behavior of the two approaches, in terms of computational costs. The very limited number of iterations balances the cost of the extra solves in the Lanczos process with that of the full orthogonalization in the Arnoldi iteration.

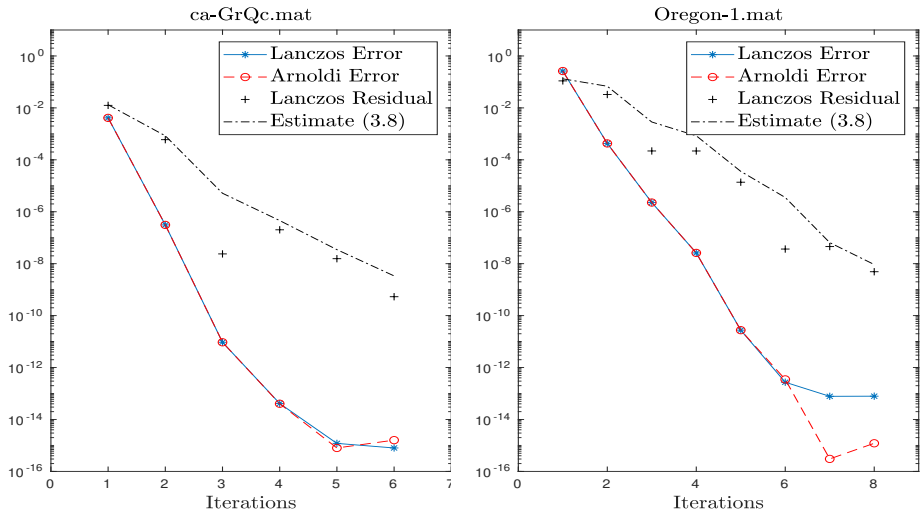

FIG. 3.1. *Example* 3.2. *Convergence history of the Arnoldi error, together with the rational Lanczos error, residual norm* (3.7) *for* $\tau = 1$*, and residual estimate* (3.8).

TABLE 3.1

Example 3.2 *with* $s = 1$*. Number of iterations of the* Q_m*-less rational Lanczos and Arnoldi methods and corresponding CPU times. Also reported are the dimension and the number of nonzeros* (nnz) *of each matrix.*

Matrix	Size	nnz	Method	It.	Time (secs)
ca-GrQc	5242	28980	Q_m-less Lanczos	6	0.079
			Rat. Arnoldi	6	0.067
Oregon-1	11492	46818	Q_m-less Lanczos	8	0.129
			Rat. Arnoldi	8	0.130

3.2. The trace of a matrix function. A problem strictly related to that of approximating a quadratic form is given by the approximation of the trace of a matrix function, $\mathrm{tr}(f(A)) = \sum_{i=1}^{n}(f(A))_{ii} = \sum_{i=1}^{n} f(\lambda_i)$, where we assume that A is symmetric with eigenvalues $\lambda_1, \ldots, \lambda_n$, and $f(\lambda_i)$ is well defined for $i = 1, \ldots, n$. The approximation of the trace by means of its definition is overly expensive for large matrices, since it requires estimates for $e_i^T f(A) e_i$ for all $i = 1, \ldots, n$.

A popular strategy is the use of a Monte-Carlo approximation. Let Z be a discrete random variable with values $\{1, -1\}$ with probability 0.5, and let z be a vector of n independent samples of Z. Then $z^T f(A) z$ is an unbiased estimator for $\mathrm{tr}(f(A))$; see, e.g., [37]. By exploiting this result, one can generate ℓ sample vectors $z^{(k)}$, $k = 1, \ldots, \ell$, estimate $(z^{(k)})^T f(A) z^{(k)} \approx \tau_k$ by means of the procedure from section 3.1, and obtain

$$\mathrm{tr}(f(A)) \approx \frac{1}{\ell} \sum_{k=1}^{\ell} \tau_k. \tag{3.9}$$

Paper [4] analyzes the case $f(\lambda) = \lambda^{-1}$, with a polynomial Lanczos approach; see also [44]. The effectiveness of the overall approach for general functions of symmetric matrices has been established in [59]; see also [45] for an improved method for stochastic trace approximation, and [15] and its references for general randomized approaches.

In the past few years probing methods have also emerged as an important alternative, especially in network analysis. These differ from Monte-Carlo approximations for the selection of the "probing" vectors $z^{(k)}$, which are then used to estimate $(z^{(k)})^T f(A) z^{(k)}$ by means, for instance, of Krylov methods; see, e.g., [8, 27, 45].

In all these strategies, a key step is the use of the Lanczos procedure to obtain an approximation to $(z^{(k)})^T f(A) z^{(k)}$; a block method suggests itself. Rational Lanczos can effectively be used to speed up convergence, in terms of number of iterations, with respect to polynomial approaches. A disadvantage of the rational approach lies in the solution of linear systems with the possibly large matrix $I - \xi_j^{-1} A$, whose cost depends on the sparsity structure of A.

Applications related to Gaussian processes often require estimating a parameter ϕ by maximizing the so-called *log-likelihood function*:

$$\log(p \,|\, \phi) = \frac{1}{2} \log\det(A(\phi)) - \frac{1}{2} x^T A(\phi) x - \frac{n}{2} \log(2\pi), \tag{3.10}$$

where the positive definite $n \times n$ matrix $A(\phi)$ is the inverse of the covariance matrix parametrized by ϕ, and x is a given vector; see, e.g., [59]. Estimating $\log\det(A(\phi))$ constitutes the main computational cost in (3.10). The relation $\log\det(A(\phi)) = \operatorname{tr}\log(A(\phi))$ allows one to use the stochastic trace estimator in (3.9) to reduce the overall computational cost. The values τ_k can be obtained by using the block rational Lanczos algorithm to approximate the bilinear form $B^T \log(A(\phi)) B$, with $B = [z^{(1)}, \dots, z^{(p)}]$.

Example 3.3. We consider the model in [48]; see also [27]. We generate uniformly distributed random pairs $s_i \in [0,1]^2$, $i = 1, \dots, n$, representing points on the real plane. A random Gaussian variable is associated with each point s_i. The model describes the association between random variables observed at fixed sites in the Euclidean space, thus imposing a neighborhood structure to the points. More precisely, two points s_i, s_j are associated if and only if the Euclidean distance $d_{ij} = \|s_i - s_j\|$ is smaller than a given parameter $\delta > 0$. Such a structure defines a planar graph. We considered $\phi = 20$, while the entries of $A(\phi)$ are defined as follows:

$$[A(\phi)]_{ij} = \begin{cases} 1 + \phi \sum_{k=1,\, k\neq i}^{n} \gamma_{ik}, & i = j, \\ -\phi \gamma_{ij}, & i \neq j, \end{cases}$$

with the so-called *reciprocal choice* [48]

$$\gamma_{ij} = \begin{cases} 1 - d_{ij}/\delta, & 0 < d_{ij} < \delta, \\ 0 & \text{otherwise.} \end{cases}$$

Table 3.2 reports the results obtained with the Q_m-less Lanczos and rational Arnoldi algorithms for different values of n, p, and δ. The sparsity pattern of $A(\phi)$ depends on δ; the larger δ, the denser $A(\phi)$. In spite of the possible cost increase in system solves, the rational Lanczos method turns out to be faster than rational Arnoldi for all the parameters we tested; see Table 3.2. This may be related to an increased cost of the full orthogonalization step in rational Arnoldi, which seems to suffer the large rank of the matrix B. Once again, we used as accuracy measure the error between the computed quantity and the value obtained by means of the MATLAB `logm` function.

For each iteration of the rational block Lanczos/Arnoldi algorithm, Figure 3.2 displays the error in the trace approximation for the rational block Lanczos/Arnoldi algorithms as the iterations proceed. The accuracy reached in the last iterations of the examples agrees with the estimated achievable accuracy of the stochastic strategy we used; see, e.g., [45]. The algorithms behave almost identically in terms of the error.

TABLE 3.2

Example 3.3. *Number of iterations of the rational Lanczos/Arnoldi method needed to reach the maximal achievable accuracy along with the CPU time in seconds. The approximation space dimension is at most* $p\cdot(\#\text{ It.}+1)$. nnz *refers to the number of nonzero elements in the considered matrices.*

Size	p	δ	nnz	Method	It.	Time (secs)
1 000	20	0.02	2 260	Q_m-less Lanczos	6	0.026
				Rat. Arnoldi	6	0.063
		0.06	11 440	Q_m-less Lanczos	7	0.047
				Rat. Arnoldi	7	0.094
10 000	200	0.002	11 278	Q_m-less Lanczos	5	1.82
				Rat. Arnoldi	5	1.99
		0.006	21 238	Q_m-less Lanczos	10	3.76
				Rat. Arnoldi	10	5.96

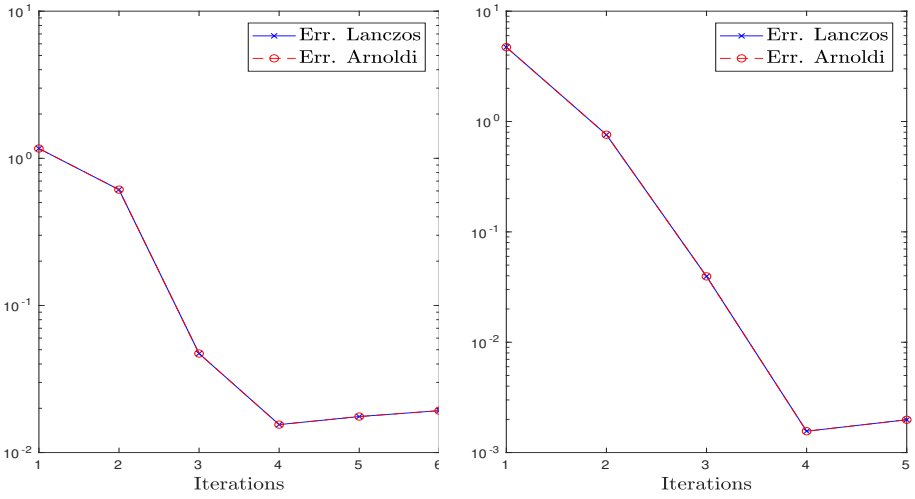

FIG. 3.2. *Example* 3.3. *Convergence error history of the trace estimation* (3.9) *obtained at each step of the rational block Lanczos/Arnoldi methods.*

3.3. $\mathcal{H}_2$-norm computation.

We consider LTI systems of the form

$$\Sigma : \begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0, \\ y(t) = Cx(t), \end{cases} \tag{3.11}$$

where $A \in \mathbb{R}^{n\times n}$ is stable, that is, its spectrum is contained in the left-half open complex plane $\mathbb{C}_-$, and $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{q\times n}$ are low rank, i.e., $p + q \ll n$. The $\mathcal{H}_2$-norm of Σ is defined as follows:

$$\|\Sigma\|_{\mathcal{H}_2}^2 = \text{trace}(C\mathbf{Q}C^T) = \text{trace}(B^T\mathbf{P}B),$$

where $\mathbf{Q}$, $\mathbf{P} \in \mathbb{R}^{n\times n}$ denote the controllability and the observability Gramian, respectively, i.e., $\mathbf{Q}$ and $\mathbf{P}$ are the solutions of the following Lyapunov equations:

$$A\mathbf{Q} + \mathbf{Q}A^T + BB^T = 0, \quad A^T\mathbf{P} + \mathbf{P}A + C^TC = 0;$$

see, e.g., [3, section 5.5.1]. The $\mathcal{H}_2$-norm gives the maximum amplitude of the system output resulting from input signals of the LTI system (3.11) with finite energy (see, e.g., [3, 32]), and thus its estimation is of interest.

For A of large dimension n, model order reduction (MOR) is used to make the dynamical system numerically tractable [3], so that the $\mathcal{H}_2$-norm can also be more cheaply estimated. Given a matrix Q_m whose columns span an appropriately chosen reduction space of dimension much lower than n, in MOR the following smaller system is introduced:

$$\Sigma_m : \begin{cases} \dot{\widehat{x}}(t) = J_m \widehat{x}(t) + B_m u(t), & \widehat{x}(0) = Q_m^T x_0,\ J_m = Q_m^T A Q_m, \\ \widehat{y}(t) = C_m \widehat{x}(t), & B_m = Q_m^T B,\ C_m = C Q_m. \end{cases} \tag{3.12}$$

This reduced system is hopefully able to reproduce the main features of the original large scale setting.[2] Rational Krylov subspaces have proven to be particularly effective for this task [3, 7, 28, 31]. For A symmetric, Algorithm 2.1 produces the small dimensional matrix J_m, while $B_m = Q_m^T B$ and $C_m = CQ_m$ can be computed incrementally during the iteration, without explicitly storing Q_m. After these computations, the Q_m-less rational Lanczos method can be employed to cheaply compute $\|\Sigma_m\|_{\mathcal{H}_2}$ as an approximation to $\|\Sigma\|_{\mathcal{H}_2}$. In the following we approximate $\|\Sigma\|_{\mathcal{H}_2} = \sqrt{\text{trace}(B^T \mathbf{P} B)}$ and we thus focus on the approximation of the observability Gramian $\mathbf{P}$. The same procedure can be adopted to compute $\|\Sigma\|_{\mathcal{H}_2} = \sqrt{\text{trace}(C \mathbf{Q} C^T)}$ if the latter formulation is preferred.

The rational Krylov subspace method can also effectively be used for solving the associated Lyapunov equation; see, e.g., [56]. Given the iteratively generated matrix Q_m for $\mathcal{K}_m(A, C^T, \boldsymbol{\xi})$, an approximation to $\mathbf{P}$ is sought in the form $\mathbf{P}_m = Q_m Y_m Q_m^T$, where the reduced matrix Y_m is obtained by imposing an orthogonality condition on the residual matrix $R_m = A\mathbf{P}_m + \mathbf{P}_m A + C^T C$. In terms of the Euclidean matrix inner product, this condition can be written as $Q_m^T R_m Q_m = 0$. Substituting $\mathbf{P}_m$ into the residual and using the orthogonality of the columns in Q_m yields the following reduced Lyapunov equation:

$$J_m Y_m + Y_m J_m + E_1 \boldsymbol{\gamma} \boldsymbol{\gamma}^T E_1^T = 0,$$

where $J_m = Q_m^T A Q_m$ and $C^T = Q_1 \boldsymbol{\gamma}$ for a nonsigular $\boldsymbol{\gamma} \in \mathbb{R}^{q \times q}$. Hence, the small size solution Y_m can be computed by means of a Schur decomposition based strategy [56]. Using the computed quantities we can write

$$\|\Sigma\|_{\mathcal{H}_2}^2 = \text{trace}(B^T \mathbf{P} B) \approx \text{trace}(B^T \mathbf{P}_m B) = \text{trace}\left(\left(B^T Q_m\right) Y_m \left(Q_m^T B\right)\right) =: \|\Sigma_m\|_{\mathcal{H}_2}^2.$$

All the required quantities can be computed without ever storing the whole matrix Q_m.

Stopping criterion. For the $\mathcal{H}_2$-norm computation we propose checking the distance between two subsequent norm approximations, that is, for $1 \le s \le m-1$,

$$\frac{|\|\Sigma_m\|_{\mathcal{H}_2} - \|\Sigma_{m-s}\|_{\mathcal{H}_2}|}{\|\Sigma_m\|_{\mathcal{H}_2}} = \frac{\left|\sqrt{\text{trace}\left(B_m^T Y_m B_m\right)} - \sqrt{\text{trace}\left(B_{m-s}^T Y_{m-s} B_{m-s}\right)}\right|}{\sqrt{\text{trace}\left(B_m^T Y_m B_m\right)}}. \tag{3.13}$$

The scheme presented in this section can be easily adapted to deal with certain parametric LTI systems like those studied in, e.g., [5], where only the matrices $B = B(\mu)$ and $C = C(\mu)$ *affinely* depend on a parameter μ belonging to a given parameter set $\mathcal{D}$.

[2] Usually two spaces, range(Q_m), range(W_m), can also be considered, so that $J_m = W_m^T A Q_m$, $B_m = W_m^T B$, and $C_m = CQ_m$. For our problem one reduction space suffices [3].

TABLE 3.3

Example 3.4. *Results for Q_m-less rational Lanczos and rational Arnoldi methods to achieve the prescribed accuracy. Shown are number of iterations, space dimension, CPU time, and relative error.*

	It.	$\dim\left(\mathcal{K}_{m+1}(\widetilde{A},\widetilde{C}^T,\boldsymbol{\xi}_{m+1})\right)$	Time (secs)	Rel. err.
Q_m-less Lanczos	12	65	0.219	7.13e-9
Rat. Arnoldi	12	65	0.199	7.13e-9

Example 3.4. We consider the two dimensional (2D) *Optical Tunable Filter* dataset available in the MOR WIKI repository [58] (see also [36]), giving the following LTI system:

$$\Sigma : \begin{cases} E\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{cases} \tag{3.14}$$

with $n = 1668$, $p = 1$, and $q = 5$. The mass matrix E is diagonal and positive definite so that we can consider the transformed system

$$\widetilde{\Sigma} : \begin{cases} \dot{\widetilde{x}}(t) = \widetilde{A}\widetilde{x}(t) + \widetilde{B}u(t), & \widetilde{A} = E^{-\frac{1}{2}}AE^{-\frac{1}{2}},\ \widetilde{B} = E^{-\frac{1}{2}}B \\ y(t) = \widetilde{C}\widetilde{x}(t), & \widetilde{C} = CE^{-\frac{1}{2}},\ \widetilde{x} = E^{\frac{1}{2}}x. \end{cases} \tag{3.15}$$

This transformation does not affect the $\mathcal{H}_2$-norm of the system, since $\|\widetilde{\Sigma}\|_{\mathcal{H}_2} = \|\Sigma\|_{\mathcal{H}_2}$. We construct $\mathcal{K}_m(\widetilde{A},\widetilde{C}^T,\boldsymbol{\xi}_m)$ for the approximation of $\|\widetilde{\Sigma}_m\|_{\mathcal{H}_2} \approx \|\widetilde{\Sigma}\|_{\mathcal{H}_2}$. The iterations are stopped as soon as the relative quantity in (3.13) for $s = 1$ is smaller than 10^{-8}. Table 3.3 collects the results for both the Q_m-less Lanczos and Arnoldi methods, using the same shifts. Thanks to the moderate dimension of the dataset, we are also able to explicitly compute the $\mathcal{H}_2$-norm of the full system[3] $\widetilde{\Sigma}$. Therefore, Table 3.3 also reports the relative error $|\|\widetilde{\Sigma}\|_{\mathcal{H}_2} - \|\widetilde{\Sigma}_m\|_{\mathcal{H}_2}|/\|\widetilde{\Sigma}\|_{\mathcal{H}_2}$.

The results in Table 3.3 show a very similar behavior for the Q_m-less rational Lanczos and rational Arnoldi methods. The Lanczos method allows us to store only three basis blocks, namely 15 vectors of length n, instead of the whole basis as in the rational Arnoldi method. In terms of CPU time, solving $2q$ linear systems per iteration in the Lanczos approach, instead of only q systems in rational Arnoldi, does not lead to a remarkable increment in the computational efforts.

Example 3.5. We consider yet another dataset from [58], the three dimensional (3D) *Gas Sensor* example. The LTI system has the form (3.14) with a diagonal positive definite mass matrix E, hence the transformed LTI system in (3.15) is employed. In this example we have $n = 66917$, $p = 1$, and $q = 28$. The large problem dimension n does not allow for the computation of the $\mathcal{H}_2$-norm of the full system $\widetilde{\Sigma}$, hence only the approximation $\|\widetilde{\Sigma}_m\|_{\mathcal{H}_2}$ is computed. Since p is significantly smaller than q (the number of columns of B and C^T, resp.), we proceed by constructing $\mathcal{K}_m(\widetilde{A},\widetilde{B},\boldsymbol{\xi}_m)$ and then we compute $\|\widetilde{\Sigma}_m\|^2_{\mathcal{H}_2} = \operatorname{trace}(\widetilde{C}\widetilde{\mathbf{Q}}_m\widetilde{C}^T)$. Both methods are stopped as soon as (3.13) for $s = 1$ becomes smaller than 10^{-8}. Table 3.4 collects the results. Also, for this example the rational Lanczos and Arnoldi methods perform similarly. This means that the cost per iteration of the two schemes is rather similar. On the other hand, for problems where a larger number of iterations is required to converge, the computational cost of the Arnoldi algorithm may significantly increase due to the explicit full orthogonalization.

[3] $\|\widetilde{\Sigma}\|_{\mathcal{H}_2}$ is computed by `norm(sys,2)` where `sys=ss`$(\widetilde{A},\widetilde{B},\widetilde{C},0)$.

TABLE 3.4

Example 3.5. This table presents results for Q_m-less rational Lanczos and rational Arnoldi methods to achieve the prescribed accuracy. Shown are the number of iterations, the space dimension, and the CPU time.

	It.	$\dim\left(\mathcal{K}_{m+1}(\widetilde{A},\widetilde{B},\boldsymbol{\xi}_{m+1})\right)$	Time (secs)
Q_m-less Lanczos	19	20	30.5
Rat. Arnoldi	19	20	28.8

3.4. LQR feedback control. We consider once again LTI systems Σ of the form (3.11), and we investigate the efficient computation of a different quantity related to the so-called LQR problem. Given the LTI system (3.11) with a stable A, this can be stated as

$$u_* = \arg\min_u \mathcal{J}(u), \text{ with } \mathcal{J}(u) = \int_0^\infty y(t)^T y(t) + u(t)^T R^{-1} u(t) dt,$$

where $\mathcal{J}$ is a quadratic cost functional and R is a $p \times p$ symmetric and positive definite matrix. Since A is stable, this u_* exists and is given by $u_*(t) = -Kx(t) = -R^{-1}B^T\mathbf{X}x(t)$, where $\mathbf{X} \in \mathbb{R}^{n\times n}$ is the unique positive semidefinite and stabilizing[4] solution of the following Riccati equation [43]:

$$A^T\mathbf{X} + \mathbf{X}A - \mathbf{X}BR^{-1}B^T\mathbf{X} + C^TC = 0. \tag{3.16}$$

Using u_* the first equation in (3.11) can be written in terms of a *closed-loop* dynamic

$$\dot{x}(t) = (A - BR^{-1}B^T\mathbf{X})x(t), \qquad x(0) = x_0,$$

whose solution is given by $x(t) = \exp((A - BR^{-1}B^T\mathbf{X})t)x_0$ for $t \geq 0$. Therefore,

$$u_*(t) = -R^{-1}B^T\mathbf{X} \cdot \exp((A - BR^{-1}B^T\mathbf{X})t)x_0. \tag{3.17}$$

In the following we show that for A symmetric an approximation to u_* can be cheaply obtained by the Q_m-less rational Lanczos method; see also [1] for related results.

Rational Krylov subspaces have appeared to lead to competitive methods for solving large-scale Riccati equations; see, e.g., [55] and references therein. In case a projection approach is employed, the overall scheme is very similar to the one reported in section 3.3 for Lyapunov equations. Once again, an approximate solution is sought in the form $\mathbf{X}_m = Q_m Y_m Q_m^T$, where $\text{Range}(Q_m) = \mathcal{K}_m(A, C^T, \boldsymbol{\xi})$, while Y_m is computed, for instance, by imposing an orthogonality (Galerkin) condition on the residual matrix. Explicitly imposing this condition and exploiting the property $Q_m^TQ_m = I$ determines a *reduced* Riccati equation to be solved in Y_m (see [10]), that is,

$$J_mY_m + Y_mJ_m - Y_mB_mR^{-1}B_m^TY_m + E_1\boldsymbol{\gamma}\boldsymbol{\gamma}^TE_1 = 0, \tag{3.18}$$

where $B_m = Q_m^TB_m$ and $\boldsymbol{\gamma} \in \mathbb{R}^{q\times q}$ is such that $C^T = Q_1\boldsymbol{\gamma}$. Since A is stable and symmetric, the matrix J_m is also stable so that Y_m exists and it is the unique positive semidefinite stabilizing solution to (3.18).

Algorithm 2.1 can be employed to construct (3.18) for a growing m, where the rows of B_m are computed iteratively during the recurrence. At the mth iteration an approximation u_m to u_* is obtained as

$$u_m(t) = R^{-1}B_m^TY_m \exp((J_m - B_mR^{-1}B_m^TY_m)t)(Q_m^Tx_0) \approx u_*(t), \tag{3.19}$$

[4] That is, $A - BR^{-1}B^T\mathbf{X}$ is a stable matrix.

which does not require storing the whole Q_m, since $Q_m^T x_0$ can also be constructed iteratively as m grows. Thanks to the stability of J_m, it can be shown that the function $u_m(t)$ defined in (3.19) is indeed the optimal control of the reduced model (3.12), namely (see [55, Corollary 3.2])

$$u_m = \arg\min_u \widehat{\mathcal{J}}(u), \quad \widehat{\mathcal{J}}(u) = \int_0^\infty \widehat{y}(t)^T \widehat{y}(t) + u(t)^T R^{-1} u(t) dt.$$

Stopping criterion. The L^2-distance between two iterates can be employed as a measure to assess the quality of the computed approximation $u_m(t)$,

$$\frac{\|u_m - u_{m-s}\|_{L^2}^2}{\|u_m\|_{L^2}^2} = \frac{\int_0^\infty \|u_m(\tau) - u_{m-s}(\tau)\|^2 d\tau}{\int_0^\infty \|u_m(\tau)\|^2 d\tau}, \quad s \in \mathbb{N},\ 0 < s < m. \tag{3.20}$$

This quantity can be cheaply approximated because it only involves small dimensional quantities. Furthermore, since $J_m - B_m R^{-1} B_m Y_m^T$ is a stable matrix, it holds that $\exp((J_m - B_m R^{-1} B_m Y_m^T)t) \to 0$ as $t \to \infty$ which may lead to an exponential convergence of the quadrature formula adopted to approximate (3.20); see, e.g., [12].

Example 3.6. We consider data in [2, Test 1]. The matrix $A \in \mathbb{R}^{n \times n}$ amounts to the 5-point finite differences discretization of the 2D Laplacian operator in the unit square $[0,1]^2$ with zero Dirichlet boundary conditions, namely $A = 1/(\bar{n}-1)^2 \cdot (T \otimes I_{\bar{n}} + I_{\bar{n}} \otimes T)$, $T = \text{tridiag}(1, -2, 1) \in \mathbb{R}^{\bar{n} \times \bar{n}}$, $n = \bar{n}^2$. The vector $B \in \mathbb{R}^n$ is such that the matrix BB^T corresponds to the discrete indicating function related to the square $[0.2, 0.8]^2$. Similarly, $C^T C$ with $C \in \mathbb{R}^{1 \times n}$ amounts to the discrete indicating function of $[0.1, 0.9]^2$. We select $R = 1$ and $x_0 = 1/(\bar{n}-1) \cdot \mathbf{1}_n$, where $\mathbf{1}_n \in \mathbb{R}^n$ is the vector of all ones.

We compare the Q_m-less rational Lanczos and Arnoldi methods for the computation of the approximate optimal control $u_m(t)$ in (3.19). Both schemes are stopped as soon as the value in (3.20) for $s = 4$ is smaller than 10^{-8}. The results in Table 3.5 show that the two methods perform similarly in terms of convergence trend and computational efficiency. Figure 3.3 also reports the convergence history of the two schemes for different values of n, illustrating that the lack of a full orthogonalization procedure does not affect the convergence of the rational Lanczos method for this example. See section 4 for a broader discussion on this topic.

TABLE 3.5

Example 3.6. *Number of iterations of the rational Lanczos/Arnoldi method needed to achieve the prescribed accuracy along with the dimension of the computed subspace and the running time in seconds.*

n	It.	$\dim\left(\mathcal{K}_{m+1}(A, C^T, \boldsymbol{\xi}_{m+1})\right)$	Method	Time (secs)
40 000	25	26	Q_m-less Lanczos	2.7
			Rat. Arnoldi	2.6
160 000	29	30	Q_m-less Lanczos	11.7
			Rat. Arnoldi	11.8
360 000	29	30	Q_m-less Lanczos	28.3
			Rat. Arnoldi	27.3

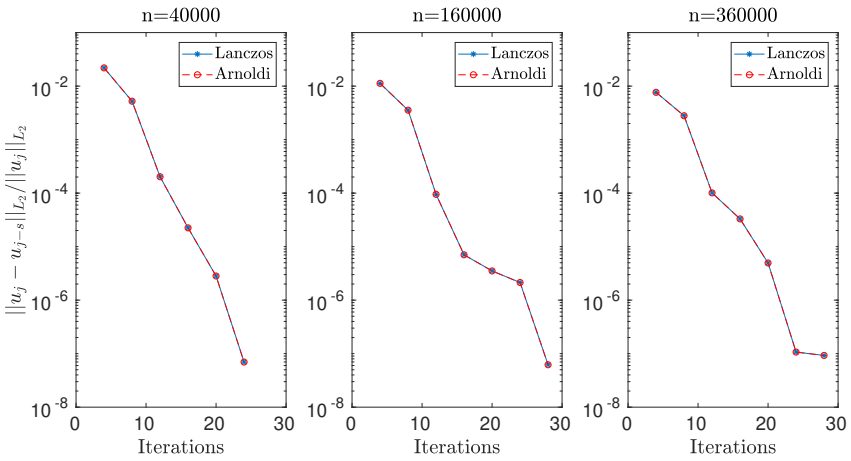

FIG. 3.3. *Example* 3.6. *Convergence history of the rational Lanczos and Arnoldi methods for different values of* n *and* $s = 4$.

4. Considerations on finite precision arithmetic computations. The polynomial Lanczos iteration is known to be prone to numerical instabilities, which cause loss of orthogonality in the computed basis. This fact has been deeply investigated by Paige in his seminal Ph.D. thesis [46] and successive works; see, e.g., [47]. Quoting [54, p. 108], loss of orthogonality can be viewed as the result of an amplification of each local error after its introduction into the computation, and its growth is determined by the eigenvalue distribution of A and by the starting vector v.

The rational Lanczos sequence is tightly related to its polynomial counterpart, with the added difficulty of the system solves, whose finite precision arithmetic computations may significantly increase the perturbation induced by round-off, even assuming a stable direct solver is employed. Notably, to the best of our efforts, we could not find in the literature a round-off error analysis for rational Krylov subpace computations within the considered applications. In this section we introduce very preliminary considerations on the round-off perturbation that may occur in the computed quantities during the iteration. We also report on our numerical experience, showing that the type of loss of orthogonality in the basis seems to be similar to that analyzed in the polynomial Lanczos method over several decades. We are conscious that performing a satisfactory quantitative analysis requires sophisticated tools that go beyond our current presentation. Hopefully, these preliminary results may be useful in a deeper analysis.

At iteration j, the generation of the rational Krylov orthonormal basis requires the following computations:

$$\begin{aligned}
&\widetilde{r} = Aq_j - \beta_{j-1}(I - \xi_{j-2}^{-1}A)q_{j-1}, \qquad \widetilde{s} = (I - \xi_{j-1}^{-1}A)q_j \\
&r = (I - \xi_j^{-1}A)^{-1}\widetilde{r}, \quad s = (I - \xi_j^{-1}A)^{-1}\widetilde{s} \\
&q = r - \alpha_j s, \qquad \text{with} \quad \alpha_j = \frac{r^T q_j}{s^T q_j} \\
&q_{j+1} = q/\beta_j, \quad \beta_j = \|q\|.
\end{aligned}$$

Using standard assumptions on finite precision arithmetic computations, and assuming a backward stable method is used for solving with the symmetric and positive

definite matrix $I - \xi_j^{-1}A$, we conjecture that an Arnoldi-type relation similar to that in the results of [47] holds, that is,

$$AQ_jK_j = Q_jH_j + (I - \xi_j^{-1}A)q_{j+1}e_j^T + F_j,$$

where the matrix F_j collects all round-off terms during the iteration. Here we envision that the columns of F_j will have an increasing norm as j grows, in accordance with the error accumulation argument known for the polynomial Lanczos. We stress that K_j and H_j are not the same as those computed in exact arithmetic, and that the columns of Q_j are no longer exactly orthonormal; however, we can assume that $v = Q_je_1$. Although our conjecture seems to be confirmed by numerical experiments, a rigorous analysis leading to upper bounds for the elements in F_j would be desirable, though it goes beyond the aim of this work. A different approach to the understanding of the finite precision arithmetic behavior could also follow the backward error analysis introduced by Greenbaum for the Lanczos method [30]. In particular, this approach may enlighten the interplay between the distribution of the eigenvalues of A and the shifts ξ_j, leading to a better understanding of the rational Lanczos convergence behavior in finite precision arithmetic.

Remark 4.1. In finite precision it no longer holds that $J_j = Q_j^TAQ_j$, and one could even question the symmetry of the computed J_j. However, since we are mainly concerned with the loss of accuracy in the computation of the length-n vectors, we can assume that K_j^{-1} and $J_j := H_jK_j^{-1} - we^TK_j^{-1}$ are computed accurately, with w as in the discussion after (2.6), so that $w^T = u^T\beta_j/\xi_j = \beta_j/\xi_j(\alpha_{j+1} - \eta\beta_j/\xi_j)e_j^TK_j^{-1} =: \upsilon_je_j^TK_j^{-1}$. We thus have

$$J_j = (H_j - \upsilon_jK_j^{-T}e_je_j^T)K_j^{-1} = H_jK_j^{-1} - \upsilon_jK_j^{-T}e_je_j^TK_j^{-1},$$

with $H_jK_j^{-1}$ symmetric. Hence, J_j remains symmetric also in finite precision arithmetic, as long as all quantities are determined using the computed coefficients.

According to Remark 4.1, we thus assume that $J_j = H_jK_j^{-1} - \upsilon_jK_j^{-T}e_je_j^TK_j^{-1}$ and K_j^{-1} are computed exactly. We can then write the perturbed relation as

$$\begin{aligned} AQ_j &= Q_jH_jK_j^{-1} + (I - \xi_j^{-1}A)q_{j+1}e_j^TK_j^{-1} + F_jK_j^{-1} \\ &= Q_jJ_j + [Q_j\upsilon_jK_j^{-T}e_j + (I - \xi_j^{-1}A)q_{j+1}]e_j^TK_j^{-1} + F_jK_j^{-1} \\ &=: Q_jJ_j + z_je_j^TK_j^{-1} + F_jK_j^{-1}. \end{aligned}$$

Subtracting ζQ_j for $\zeta \in \mathbb{C}$ such that $A - \zeta I$ and $J_j - \zeta I$ are nonsingular, we obtain

$$(A - \zeta I)Q_j = Q_j(J_j - \zeta I) + z_je_j^TK_j^{-1} + F_jK_j^{-1}.$$

Multiplying by $(A - \zeta I)^{-1}$ and $(J_j - \zeta I)^{-1}$, and rearranging the relation above we obtain

$$(A - \zeta I)^{-1}Q_j = Q_j(J_j - \zeta I)^{-1} - (A - \zeta I)^{-1}(z_je_j^TK_j^{-1} + F_jK_j^{-1})(J_j - \zeta I)^{-1}.$$

We analyze the effect of this computation in the approximation of the quadratic form $v^Tf(A)v$, for which the short-term recurrence seems to work particularly well (see section 3.1); see, e.g., [20] for a related analysis of the polynomial Lanczos method.

Using $f(A) = \int_\Gamma f(\zeta)(A-\zeta I)^{-1}d\zeta$ we can write

$$\begin{aligned} v^T f(A)v &= \int_\Gamma f(\zeta) v^T((A-\zeta I)^{-1}Q_j)e_1 d\zeta \\ &= \int_\Gamma f(\zeta) v^T Q_j (J_j-\zeta)^{-1}e_1 d\zeta \\ &\quad -v^T \int_\Gamma f(\zeta)(A-\zeta I)^{-1}(z_j e_j^T + F_j)K_j^{-1}(J_j-\zeta I)^{-1}e_1 d\zeta. \end{aligned}$$

Therefore,

$$v^T f(A)v = e_1^T Q_j^T Q_j f(J_j)e_1 - v^T \int_\Gamma f(\zeta)(A-\zeta I)^{-1}(z_j e_j^T + F_j)K_j^{-1}(J_j-\zeta I)^{-1}e_1 d\zeta.$$

In exact arithmetic it would hold that $Q_j^T Q_j = I_j$ so that the quantity $e_1^T f(J_j)e_1$ would correspond to the classical approximation in the given subspace. In finite precision arithmetic the distance from the ideal quantity can be estimated as follows:

$$\begin{aligned} |v^T f(A)v - e_1^T f(J_j)e_1| &\le |e_1^T f(J_j)e_1 - q_1^T Q_j f(J_j)e_1| \\ &\quad + \left| \int_\Gamma f(\zeta) v^T (A-\zeta I)^{-1} z_j e_j^T K_j^{-1}(J_j-\zeta I)^{-1}e_1 d\zeta \right| \\ &\quad + \left| \int_\Gamma f(\zeta) v^T (A-\zeta I)^{-1} F_j K_j^{-1}(J_j-\zeta I)^{-1}e_1 d\zeta \right| \\ &=: \mathcal{I} + \mathcal{II} + \mathcal{III}. \end{aligned}$$

We next analyze the right-hand side terms. Let $q_1^T Q_j = e_1 + \boldsymbol{\epsilon}_j$, where we can assume that $e_1^T \boldsymbol{\epsilon}_j = 0$ (exact normalization) while the quantity $|e_k^T \boldsymbol{\epsilon}_j|$ may grow with k. Then

$$\begin{aligned} |e_1^T f(J_j)e_1 - q_1^T Q_j f(J_j)e_1| &= |\boldsymbol{\epsilon}_j^T f(J_j)e_1| \\ &= \left| \sum_{\ell=1}^{j} (\boldsymbol{\epsilon}_j)_\ell (f(J_j)e_1)_\ell \right| \le \sum_{\ell=1}^{j} |(\boldsymbol{\epsilon}_j)_\ell|\,|(f(J_j)e_1)_\ell|. \end{aligned} \tag{4.1}$$

In exact arithmetic, it has been proved that $|e_\ell^T f(J_j)e_1|$ shows a decaying behavior—which is possibly exponential—as ℓ grows, and the slope depends on the spectral properties of the coefficient matrix [49]. If this property is maintained in finite precision arithmetic, then each term $|(\boldsymbol{\epsilon}_j)_\ell|$ is allowed to grow as long as the product $|(\boldsymbol{\epsilon}_j)_\ell|\,|(f(J_j)e_1)_\ell|$ remains small.

Assuming that K_j is computed exactly, the term $\mathcal{II}$ does not involve perturbation matrices; therefore its magnitude is related to the quality of the rational Krylov space approximation in exact precision arithmetic. Last, the term $\mathcal{III}$ shows that the columns of the round-off error matrix are *weighted* by the components of the vector $K_j^{-1}(J_j-\zeta I)^{-1}e_1$. By using the definition of J_j and K_j, it follows that $(J_j-\zeta I)K_j = H_j - \zeta K_j - w_j e_j^T$, which is a tridiagonal plus a rank-one matrix acting on the last column. Setting $G_j = H_j - \zeta K_j$ with $G_j = G_j(\zeta)$ and using the Sherman–Morrison formula, we get $K_j^{-1}(J_j-\zeta I)^{-1}e_1 = G_j^{-1}e_1 + G_j^{-1}w_j(e_j^T G_j^{-1}e_1)/(1-e_j^T G_j^{-1}w)$. If G_j has convenient spectral properties, then the components of the vector $G_j^{-1}e_1$ have a decaying magnitude, while the magnitude of the second term in the formula depends on $e_j^T G_j^{-1}e_1$, the last component of the first vector. Hence, in this case the propagated errors contained in the rightmost columns of F_j are weighted by small values.

As a consequence, the round-off effect appears to be mitigated. Although we have experimental evidence that round-off errors seem to only slightly affect computations, a thorough analysis is needed to make more definitive statements.

Remark 4.2. For $f(z) = z^{-1}$ the polynomial Lanczos analysis is closely related to that of the conjugate gradient (CG) method for solving linear systems. We refer the reader to the monograph [41] and its references for a thorough discussion.

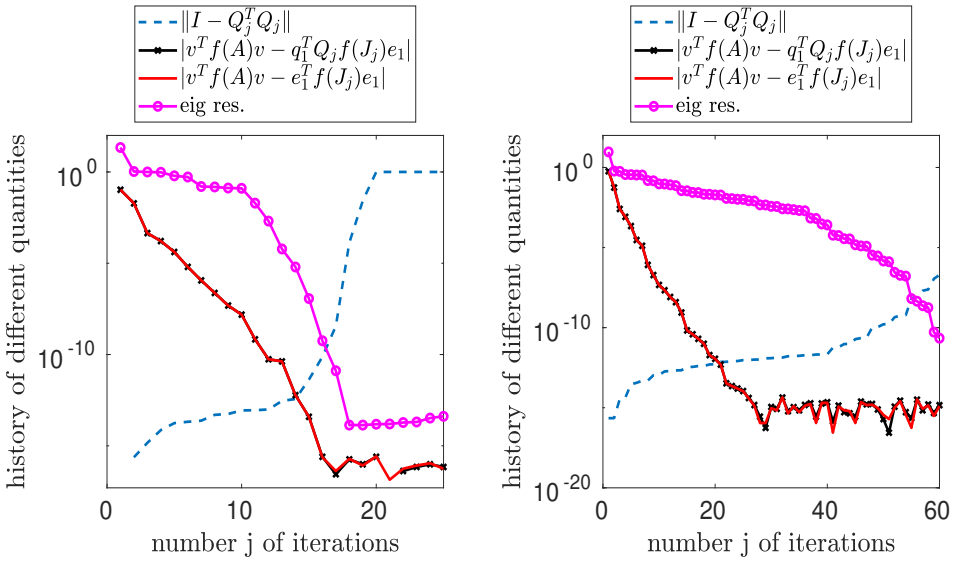

FIG. 4.1. *Example* 4.3. *Convergence history of approximation to* $v^T A^{1/2} v$. *Left:* $\rho = 0.45$. *Right:* $\rho = 0.85$.

Example 4.3. We consider an example first introduced in [57]. The matrix is diagonal with eigenvalues equal to $\lambda_i = \lambda_1 + (i-1)/(n-1)(\lambda_1 - \lambda_n)\rho^{n-i}$, $i = 1, \ldots, n$; the parameter ρ is used to control the eigenvalue distribution in the spectral interval, so that a value of ρ close to one distributes the eigenvalues almost uniformly in the interval. We considered $n = 900$, $\lambda_1 = 0.01$, and $\lambda_n = 100$, together with $f(\lambda) = \lambda^{1/2}$. Moreover, we analyzed two values of ρ, that is, $\rho = 0.45$ and $\rho = 0.85$. The plots in Figure 4.1 show the error $|v^T f(A)v - e_1^T f(J_j)e_1|$ together with the loss of orthogonality $\|I - Q_j^T Q_j\|$ and the true approximation error $|v^T f(A)v - q_1^T Q_j f(J_j)e_1|$ as the number j of iterations increases. Plots are reported for $\rho = 0.45$ (left) and $\rho = 0.85$ (right). The eigenvalue residual norm $\|Ax^{(j)} - x^{(j)}\lambda^{(j)}\|/|\lambda^{(j)}|$ is also shown, where $(\lambda^{(j)}, x^{(j)})$ is the Ritz eigenpair with $\lambda^{(j)}$ closest to λ_n. Similarly to the polynomial Lanczos method, for both values of ρ loss of orthogonality is related to the convergence of the Ritz eigenpair to the corresponding eigenpair of A. Concerning the bilinear form, we first remark that the convergence to $v^T f(A)v$ is not consistently related to that of the Ritz eigenpair, and in addition, convergence seems to be insensitive to the fact that $q_1^T Q_j \neq e_1$, that is, $\boldsymbol{\epsilon}_j \neq 0$. Though convergence is slower for $\rho = 0.85$ than for $\rho = 0.45$, the last property is maintained in both cases.

In the previous example we illustrated that the accuracy obtained by $q_1^T Q_j f(J_j)e_1$ is similar to that of $e_1^T f(J_j)e_1$, and this is related to the role of $|\sum_{\ell=1}^{j}(\boldsymbol{\epsilon}_j)_\ell (f(J_j)e_1)_\ell|$ in the discussion above. The next example investigates this issue further.

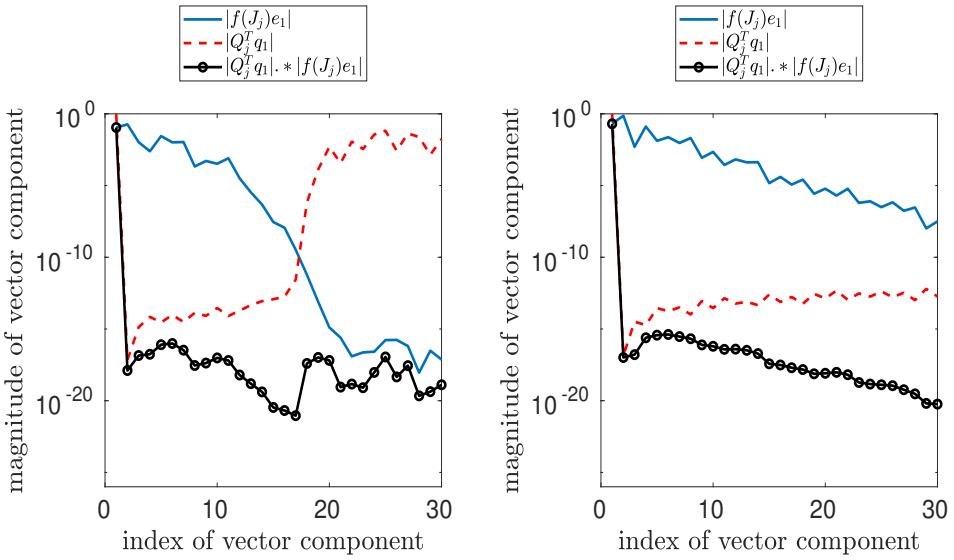

FIG. 4.2. *Example* 4.4. *Magnitude of components at iteration* $j = 30$ *in the approximation to* $v^T A^{1/2} v$. *Left:* $\rho = 0.45$. *Right:* $\rho = 0.85$.

Example 4.4. With the same data as in Example 4.3, we focus on iteration $j = 30$ and inspect the magnitude of the components of the vectors $f(J_j)e_1$ and $q_1^T Q_j$, which give the factors in the sum $|\sum_{\ell=1}^{j}(\boldsymbol{\epsilon}_j)_\ell (f(J_j)e_1)_\ell|$. The two plots in Figure 4.2 ($\rho = 0.45$ on the left and $\rho = 0.85$ on the right) report the quantities $|f(J_j)e_1|_\ell$, $|q_1^T Q_j|_\ell$, and $|q_1^T Q_j|_\ell |f(J_j)e_1|_\ell$ for $\ell = 1, \ldots, 30$. The two figures consistently report that the increasing pattern of $|q_1^T Q_j|_\ell$ inversely matches the decreasing one of $|f(J_j)e_1|_\ell$, so that the product of each component remains at the level of 10^{-15}. If J_j were exact, the decay pattern of $|f(J_j)e_1|_\ell$ would be expected, thanks to the theoretical results reported in [49]. The fact that the round-off error does not seem to significantly alter the interplay between $|q_1^T Q_j|_\ell$ and $|f(J_j)e_1|_\ell$ could be related to the theory developed by Greenbaum [30]. For the polynomial Lanczos method, Greenbaum showed that the computed projected matrix can be expressed as the output of the Lanczos iteration applied *in exact arithmetic* not to A but to a matrix $\tilde{A}$ of larger dimensions. In particular, the eigenvalues of $\tilde{A}$ cluster around those of A. If an analogous result could be derived for the projected matrix J_j computed by the rational Lanczos method, then the result in [49] would ensure that the decay pattern of $|f(J_j)e_1|_\ell$ also holds in finite precision. This would justify the persistency of the interplay between $|q_1^T Q_j|_\ell$ and $|f(J_j)e_1|_\ell$ in finite precision arithmetic. We will address this intriguing issue in future research.

4.1. Stability of the LU factorization of K_j. In section 2.1 it was stated that the LU factorization $K_j = L_j U_j$ with no pivoting exists. In this section we analyze its stability properties. Assuming that K_j is computed exactly, the stability of the LU factorization will ensure that the recurrences (2.8) and (2.9) are backward stable. To this end, we write

$$K_j = \begin{bmatrix} 1 & 0 \\ \frac{\beta_1}{\xi_1} & \widetilde{K}_j \end{bmatrix} \quad \text{with } \widetilde{K}_j = \widetilde{D}_{j-1}^{-1}\widetilde{T}_j, \quad \widetilde{T}_j := \begin{bmatrix} \xi_1+\alpha_2 & \beta_2 & & \\ \beta_2 & \xi_2+\alpha_3 & \ddots & \\ & \ddots & \ddots & \beta_{j-1} \\ & & \beta_{j-1} & \xi_{j-1}+\alpha_j \end{bmatrix},$$

where $\widetilde{T}_j \in \mathbb{R}^{(j-1)\times(j-1)}$ and $\widetilde{D}_{j-1} := \operatorname{diag}(\xi_1, \ldots, \xi_{j-1})$.

LEMMA 4.5. *Let $j_* \le n$ be the first index such that $\beta_{j_*} = 0$, giving subspace invariance in* (2.3)*. Using the notation of section* 2.1*, let A be symmetric and positive definite, and let the shifts $\xi_1, \ldots, \xi_{j_*-1}$ all be negative. Then the symmetric matrices H_j and $\widetilde{T}_j$ are both positive definite for $j \le j_*$. Conversely, if A is symmetric and negative definite with positive shifts, then H_j is negative definite while $\widetilde{T}_j$ is positive definite.*

Proof. We prove the result for A negative definite; the positive definite case follows similarly. We have $J_{j_*} = H_{j_*} K_{j_*}^{-1}$ so that $J_{j_*} K_{j_*} = J_{j_*} + J_{j_*} D_{j_*-1}^{-1} H_{j_*} = H_{j_*}$, from which

$$H_{j_*} = (I_{j_*} - J_{j_*} D_{j_*-1}^{-1})^{-1} J_{j_*} = (J_{j_*}^{-1} - D_{j_*-1}^{-1})^{-1}.$$

Since $J_{j_*} = Q_{j_*}^T A Q_{j_*}$ is symmetric and negative definite, and the shifts are all positive, it follows that H_{j_*} is a negative definite matrix. Moreover, the eigenvalues of K_{j_*} are positive since $K_{j_*} = H_{j_*} J_{j_*}^{-1}$ is the product of two symmetric negative definite matrices.

For any $j \le j_*$, the eigenvalues of H_j are contained in the spectral interval of H_{j_*}, so that H_j is positive definite. To derive the positive definiteness of $\widetilde{T}_j$, we observe that for any $j \le j_*$, the spectrum of K_j is composed of 1 and all the eigenvalues of the submatrix $\widetilde{K}_j = \widetilde{D}_{j-1}^{-1}\widetilde{T}_j$, with $j \le j_*$. For $j = j_*$, the positivity of the eigenvalues of K_{j_*} ensures that of the eigenvalues of $\widetilde{K}_{j_*}$. In particular, since $\widetilde{D}_{j_*-1} > 0$, this implies that $\widetilde{T}_{j_*}$ is positive definite. Hence, all principal $j \times j$ matrices of $\widetilde{T}_{j_*}$ are also positive definite, with $j \le j_*$. ∎

We can prove the backward stability of the Gaussian elimination procedure associated with K_j, thus proving Proposition 2.3. Here $|M|$ is the matrix obtained by taking the elementwise absolute values of the matrix M.

PROPOSITION 4.6. *Under the assumptions of Lemma* 4.5*, if the unit roundoff* $\mathtt{u}$ *is small enough, then the Gaussian elimination for the system $K_j y = e_j$ succeeds, and the computed solution $\hat{y}$ satisfies*

$$(K_j + \Delta K_j)\hat{y} = e_j, \quad |\Delta K_j| < h(\mathtt{u})|K_j|, \quad h(\mathtt{u}) = \frac{4\mathtt{u} + 3\mathtt{u}^2 + \mathtt{u}^3}{1 - \mathtt{u}}.$$

The same holds for the system $K_j^T t = e_j$.

Proof. Following the argument in [35, section 9.6], it is sufficient to prove that the LU factorization $K_j = L_j U_j$ satisfies $|L_j||U_j| = |L_j U_j|$ with U_j having positive diagonal elements, and the result will follow from [35, Theorem 9.14] and its proof. We restrict our attention to the matrix $\widetilde{K}_j$, as the first row and column of K_j are already in the desired form.

For A negative definite, from Lemma 4.5 it follows that the matrix $\widetilde{T}_j = \widetilde{D}_{j-1}\widetilde{K}_j$ is positive definite. In particular, Theorem 9.12 in [35] ensures that the LU factorization $\widetilde{T}_j = \widetilde{L}_j\widetilde{U}_j$ satisfies the condition $|\widetilde{L}_j||\widetilde{U}_j| = |\widetilde{L}_j\widetilde{U}_j|$. Therefore, the matrix $\widetilde{K}_j$ can be factorized as $\widetilde{K}_j = \widetilde{D}_{j-1}^{-1}\widetilde{L}_j\widetilde{U}_j$. Note that the matrix $\widehat{L}_j := \widetilde{D}_{j-1}^{-1}\widetilde{L}_j\widetilde{D}_{j-1}$ is lower bidiagonal with all the diagonal entries equal to 1. Then

$$\widetilde{K}_j = \widehat{L}_j\widetilde{D}_{j-1}^{-1}\widetilde{U}_j = \widehat{L}_j\widehat{U}_j$$

is the unique LU factorization of $\widetilde{K}_j$, with $\widehat{U}_j := \widetilde{D}_{j-1}^{-1}\widetilde{U}_j$. Note that the diagonal elements of $\widehat{U}_j$ are positive. Since $\widetilde{D}_{j-1}$ has positive diagonal entries, we get the following equalities:

$$\begin{aligned}|\widehat{L}_j||\widehat{U}_j| &= |\widehat{L}_j||\widetilde{D}_{j-1}^{-1}\widetilde{U}_j| = |\widehat{L}_j\widetilde{D}_{j-1}^{-1}||\widetilde{U}_j| = |\widetilde{D}_{j-1}^{-1}\widetilde{L}_j||\widetilde{U}_j| = |\widetilde{D}_{j-1}^{-1}||\widetilde{L}_j||\widetilde{U}_j| \\ &= |\widetilde{D}_{j-1}^{-1}||\widetilde{L}_j\widetilde{U}_j| = |\widetilde{D}_{j-1}^{-1}\widetilde{L}_j\widetilde{U}_j| = |\widehat{L}_j\widehat{U}_j|.\end{aligned}$$

Returning to K_j and using the notation of the proof of Lemma 2.1, we observe that the first two computed coefficients in the factorization $K_j = L_j U_j$ are $\omega_1 = 1$ and $\ell_2 = \beta_1/\xi_1 > 0$. Hence, it holds that $|L_j||U_j| = |L_j U_j|$ with U_j having positive diagonal elements, concluding the proof.

The case in which A is positive definite can be proved analogously. ∎

5. Conclusions. We have described a computationally and memory efficient implementation of the symmetric rational Lanczos method. The algorithm does not require storing the whole orthonormal basis Q_m to proceed with the iterations. We have illustrated a number of application problems where the proposed Q_m-less algorithm can effectively be employed. Very preliminary considerations of finite precision arithmetic computations seem to indicate that the behavior of the short-term recurrence rational method in this context is similar to that of its polynomial counterpart, although a comprehensive analysis is required to make more definitive statements.

Appendix. In this section we present the block variant of Algorithm 2.1.

Algorithm A.1. Block rational Lanczos.

input : $A \in \mathbb{R}^{n\times n}$, $V \in \mathbb{R}^{n\times p}$, $\boldsymbol{\xi}$, number of iterations $m > 0$.
output: $J_m \in \mathbb{R}^{pm\times pm}$, $J_m = Q_m^T A Q_m$, $\text{Range}(Q_m) = \mathcal{K}_m(A, V, \boldsymbol{\xi}_m)$.

Compute a skinny QR factorization of V, $\widehat{Q}\widehat{R} = V$
while $j \le m$ **do**
  if $j = 1$ **then**
    Set $\widetilde{R} = A\widehat{Q}$ and $\widetilde{S} = \widehat{Q}$
  else if $j = 2$ **then**
    Set $\widetilde{R} = A\widehat{Q} - \bar{Q}\beta_{j-1}^T$ and $\widetilde{S} = (I - A/\xi_{j-1})\widehat{Q}$
  else
    Set $\widetilde{R} = A\widehat{Q} - (I - A/\xi_{j-2})\bar{Q}\beta_{j-1}^T$ and $\widetilde{S} = (I - A/\xi_{j-1})\widehat{Q}$
  Solve $(I - A/\xi_j)[R, S] = [\widetilde{R}, \widetilde{S}]$
  Compute $\alpha_j = \left(\widehat{Q}^T S\right)^{-1}\left(\widehat{Q}^T R\right)$
  Set $Q = R - S\alpha_j$
  Set $\bar{Q} = \widehat{Q}$
  Compute a skinny QR factorization of Q, $\widehat{Q}\beta_j = Q$
  if j=1 **then**
    Set $u_j = y_1 = t_1 = I_p$ and $\widehat{y}_1 = \alpha_1$
  else
    Set $u_j = \alpha_j/\xi_{j-1} + I_p - \beta_{j-1}\omega_{j-1}^{-1}\beta_{j-1}^T/(\xi_{j-1}\xi_{j-2})$
    Set $y_j = \begin{bmatrix} -y_{j-1}\beta_{j-1}^T\omega_j^{-1}/\xi_{j-2} \\ \omega_j^{-1} \end{bmatrix}$, $t_j = \begin{bmatrix} -t_{j-1}\beta_{j-1}^T\omega_j^{-1}/\xi_{j-1} \\ \omega_j^{-1} \end{bmatrix}$, and $\widehat{y}_j = \begin{bmatrix} -\widehat{y}_{j-1}\beta_{j-1}^T\omega_j^{-1}/\xi_{j-2} \\ \beta_{j-1}^T E_{j-1}^T y_j + \alpha_j\omega_j^{-1} \end{bmatrix}$
      $+ E_{j-1}\beta_{j-1}u_j^{-1}$
  Compute $\eta = \widehat{Q}^T A\widehat{Q}$
  Set $J_j E_j = \widehat{y}_j - t_j\beta_j^T(I_p - \eta/\xi_j)\beta_j/\xi_j u_j^{-1}$ and $E_j^T J_j = (J_j E_j)^T$
  Set $j = j + 1$

Acknowledgments. The authors would like to thank Miroslav S. Pranić for an insightful discussion on [50], and Niel Van Buggenhout for the helpful comments about the rational Krylov moment matching property. The authors are also grateful to the two anonymous reviewers for their careful reading and for comments that led us to include section 4.1. The datasets and algorithms generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

[1] A. Alla, D. Kalise, and V. Simoncini, *State-dependent Riccati Equation Feedback Stabilization for Nonlinear PDEs*, preprint, https://arxiv.org/abs/2106.07163, 2021.
[2] A. Alla and V. Simoncini, *Order reduction approaches for the algebraic Riccati equation and the LQR problem*, in Numerical Methods for Optimal Control Problems, Springer INdAM Ser. 29, Springer, Cham, 2018, pp. 89–109, https://doi.org/10.1007/978-3-030-01959-4_5.
[3] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*, SIAM, Philadelphia, 2005, https://doi.org/10.1137/1.9780898718713.
[4] Z. Bai, G. Fahey, and G. Golub, *Some large-scale matrix computation problems*, J. Comput. Appl. Math., 74 (1996), pp. 71–89, https://doi.org/10.1016/0377-0427(96)00018-0.
[5] U. Baur, C. Beattie, P. Benner, and S. Gugercin, *Interpolatory projection methods for parameterized model reduction*, SIAM J. Sci. Comput., 33 (2011), pp. 2489–2518, https://doi.org/10.1137/090776925.
[6] 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, https://doi.org/10.1137/080741744.
[7] P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, *Model Reduction and Approximation: Theory and Algorithms*, Comput. Sci. Eng. 15, SIAM, Philadelphia, 2017, https://doi.org/10.1137/1.9781611974829.
[8] A. H. Bentbib, M. El Ghomari, K. Jbilou, and L. Reichel, *Shifted extended global Lanczos processes for trace estimation with application to network analysis*, Calcolo, 58 (2021), 4, https://doi.org/10.1007/s10092-020-00395-1.

[9] M. Benzi and G. H. Golub, *Bounds for the entries of matrix functions with applications to preconditioning*, BIT, 39 (1999), pp. 417–438, https://doi.org/10.1023/a:1022362401426.
[10] D. A. Bini, B. Iannazzo, and B. Meini, *Numerical Solution of Algebraic Riccati Equations*, SIAM, Philadelphia, 2011, https://doi.org/10.1137/1.9781611972092.
[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, https://doi.org/10.1137/110820191.
[12] J. P. Boyd, *Exponentially convergent Fourier-Chebyshev quadrature schemes on bounded and infinite intervals*, J. Sci. Comput., 2 (1987), pp. 99–109, https://doi.org/10.1007/BF01061480.
[13] A. Bultheel, P. Gonzalez-Vera, E. Hendriksen, and O. Njastad, *Orthogonal rational functions and tridiagonal matrices*, J. Comput. Appl. Math., 153 (2003), pp. 89–97.
[14] J. Chen and Y. Saad, *A posteriori error estimate for computing* $\mathrm{tr}(f(A))$ *by using the Lanczos method*, Numer. Linear Algebra Appl., 25 (2018), e2170.
[15] A. Cortinovis and D. Kressner, *On randomized trace estimates for indefinite matrices with an application to determinants*, Found. Comput. Math., 22 (2022), pp. 875–903, https://doi.org/10.1007/s10208-021-09525-9.
[16] T. A. Davis and Y. Hu, *The University of Florida Sparse Matrix Collection*, ACM Trans. Math. Software, 38 (2011), 1, https://doi.org/10.1145/2049662.2049663.
[17] B. de la Calle Ysern, *Error bounds for rational quadrature formulae of analytic functions*, Numer. Math., 101 (2005), pp. 251–271, https://doi.org/10.1007/s00211-005-0575-8.
[18] K. Deckers and A. Bultheel, *Rational Krylov Sequences and Orthogonal Rational Functions*, Technical report, Department of Computer Science, K.U. Leuven, Leuven, Belgium, 2007.
[19] K. Deckers and A. Bultheel, *The existence and construction of rational Gauss-type quadrature rules*, Appl. Math. Comput., 218 (2012), pp. 10299–10320, https://doi.org/10.1016/j.amc.2012.04.008.
[20] V. Druskin, A. Greenbaum, and L. Knizhnerman, *Using nonorthogonal Lanczos vectors in the computation of matrix functions*, SIAM J. Sci. Comput., 19 (1998), pp. 38–54, https://doi.org/10.1137/S1064827596303661.
[21] V. Druskin, L. Knizhnerman, and V. Simoncini, *Analysis of the rational Krylov subspace and ADI methods for solving the Lyapunov equation*, SIAM J. Numer. Anal., 49 (2011), pp. 1875–1898, https://doi.org/10.1137/100813257.
[22] 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, https://doi.org/10.1137/080742403.
[23] V. Druskin, C. Lieberman, and M. Zaslavsky, *On adaptive choice of shifts in rational Krylov subspace reduction of evolutionary problems*, SIAM J. Sci. Comput., 32 (2010), pp. 2485–2496, https://doi.org/10.1137/090774082.
[24] V. Druskin and V. Simoncini, *Adaptive rational Krylov subspaces for large-scale dynamical systems*, Systems Control Lett., 60 (2011), pp. 546–560, https://doi.org/10.1016/j.sysconle.2011.04.013.
[25] E. Estrada and J. A. Rodríguez-Velázquez, *Subgraph centrality in complex networks*, Phys. Rev. E, 71 (2005), 056103, https://doi.org/10.1103/PhysRevE.71.056103.
[26] C. Fenu, L. Reichel, and G. Rodriguez, *GCV for Tikhonov regularization via global Golub–Kahan decomposition*, Numer. Linear Algebra Appl., 23 (2016), pp. 467–484, https://doi.org/10.1002/nla.2034.
[27] A. Frommer, C. Schimmel, and M. Schweitzer, *Analysis of probing techniques for sparse approximation and trace estimation of decaying matrix functions*, SIAM J. Matrix Anal. Appl., 42 (2021), pp. 1290–1318, https://doi.org/10.1137/20M1364461.
[28] K. Gallivan, E. Grimme, and P. Van Dooren, *A rational Lanczos algorithm for model reduction*, Numer. Algorithms, 12 (1996), pp. 33–63.
[29] G. H. Golub and G. Meurant, *Matrices, Moments and Quadrature with Applications*, Princeton University Press, Princeton, NJ, 2010.
[30] A. Greenbaum, *Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences*, Linear Algebra Appl., 113 (1989), pp. 7–63.
[31] E. Grimme, *Krylov Projection Methods for Model Reduction*, Ph.D. thesis, The University of Illinois at Urbana-Champaign, Champaign, IL, 1997.
[32] S. Gugercin, A. C. Antoulas, and C. Beattie, $\mathcal{H}_2$ *model reduction for large-scale linear dynamical systems*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 609–638, https://doi.org/10.1137/060666123.
[33] S. Güttel, *Rational Krylov Methods for Operator Functions*, Ph.D. thesis, TU Bergakademie Freiberg, Freiberg, Germany, 2010.

[34] S. Güttel, *Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection*, GAMM-Mitt., 36 (2013), pp. 8–31, https://doi.org/10.1002/gamm.201310002.
[35] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, PA, 2002, https://doi.org/10.1137/1.9780898718027.
[36] D. Hohlfeld and H. Zappe, *An all-dielectric tunable optical filter based on the thermo-optic effect*, J. Opt. A: Pure Appl. Opt., 6 (2004), pp. 504–511, https://doi.org/10.1088/1464-4258/6/6/002.
[37] M. Hutchinson, *A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines*, Comm. Statist. Simulation Comput., 18 (1989), pp. 1059–1076.
[38] 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.
[39] J. V. Lambers, *Solution of time-dependent PDE through component-wise approximation of matrix functions*, IAENG Int. J. Appl. Math., 41 (2011), pp. 1–10.
[40] C. Lanczos, *Solution of systems of linear equations by minimized iterations*, J. Research Nat. Bur. Standards, 49 (1952), pp. 33–53.
[41] J. Liesen and Z. Strakoš, *Krylov Subspace Methods: Principles and Analysis*, Oxford University Press, Oxford, 2013.
[42] The MathWorks, Inc., *MATLAB*, http://www.matlab.com.
[43] V. L. Mehrmann, *The Autonomous Linear Quadratic Control Problem: Theory and Numerical Solution*, Lect. Notes Control Inf. Sci. 163, Springer-Verlag, Berlin, 1991, https://doi.org/10.1007/BFb0039443.
[44] G. Meurant, *Estimates of the trace of the inverse of a symmetric matrix using the modified Chebyshev algorithm*, Numer. Algorithms, 51 (2009), pp. 309–318, https://doi.org/10.1007/s11075-008-9246-z.
[45] R. A. Meyer, C. Musco, C. Musco, and D. P. Woodruff, *Hutch++: Optimal stochastic trace estimation*, in Symposium on Simplicity in Algorithms (SOSA), SIAM, Philadelphia, 2021, pp. 142–155, https://doi.org/10.1137/1.9781611976496.16.
[46] C. C. Paige, *The Computation of Eigenvalues and Eigenvectors of Very Large and Sparse Matrices*, Ph.D. thesis, London University, London, England, 1971.
[47] C. C. Paige, *Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix*, J. Inst. Math. Appl., 18 (1976), pp. 341–349.
[48] A. N. Pettitt, I. S. Weir, and A. G. Hart, *A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data*, Stat. Comput., 12 (2002), pp. 353–367, https://doi.org/10.1023/a:1020792130229.
[49] S. Pozza and V. Simoncini, *Functions of rational Krylov space matrices and their decay properties*, Numer. Math., 148 (2021), pp. 99–126, https://doi.org/10.1007/s00211-021-01198-4.
[50] M. S. Pranić and L. Reichel, *Rational Gauss quadrature*, SIAM J. Numer. Anal., 52 (2014), pp. 832–851, https://doi.org/10.1137/120902161.
[51] A. Ruhe, *Rational Krylov sequence methods for eigenvalue computation*, Linear Algebra Appl., 58 (1984), pp. 391–405.
[52] Y. Saad, J. R. Chelikowsky, and S. M. Shontz, *Numerical methods for electronic structure calculations of materials*, SIAM Rev., 52 (2010), pp. 3–54, https://doi.org/10.1137/060651653.
[53] M. Schweitzer, *Monotone convergence of the extended Krylov subspace method for Laplace–Stieltjes functions of Hermitian positive definite matrices*, Linear Algebra Appl., 507 (2016), pp. 486–498.
[54] H. D. Simon, *Analysis of the symmetric Lanczos algorithm with reorthogonalization methods*, Linear Algebra Appl., 61 (1984), pp. 101–131.
[55] V. Simoncini, *Analysis of the rational Krylov subspace projection method for large-scale algebraic Riccati equations*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 1655–1674, https://doi.org/10.1137/16M1059382.
[56] V. Simoncini, *Computational methods for linear matrix equations*, SIAM Rev., 58 (2016), pp. 377–441, https://doi.org/10.1137/130912839.
[57] Z. Strakoš, *On the real convergence rate of the conjugate gradient method*, Linear Algebra Appl., 154/156 (1991), pp. 535–549.
[58] The MORwiki Community, *MOR Wiki - Model Order Reduction Wiki*, http://modelreduction.org.

[59] S. Ubaru, J. Chen, and Y. Saad, *Fast Estimation of* $\mathrm{tr}(f(A))$ *via stochastic Lanczos quadrature*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 1075–1099, https://doi.org/10.1137/16M1104974.

[60] N. Van Buggenhout, M. Van Barel, and R. Vandebril, *Biorthogonal rational Krylov subspace methods*, Electron. Trans. Numer. Anal., 51 (2019), pp. 451–468, https://doi.org/10.1553/etna_vol51s451.