CONVERGENCE RATES FOR INVERSE-FREE RATIONAL APPROXIMATION OF MATRIX FUNCTIONS*

C. JAGELS[†], T. MACH[‡], L. REICHEL[§], AND R. VANDEBRIL[‡]

Abstract. In this article we will deduce geometrical convergence rates for approximating matrix functions via inverse-free rational Krylov methods. In practical applications one frequently encounters matrix functions such as, e.g., the matrix exponential or matrix logarithm; often the matrix under consideration is too large to compute the matrix function directly, and an approximation by using Krylov subspaces is used instead. If many matrix-vector products of the form $f(A)\boldsymbol{v}$ are required, then it pays off to use smaller *rational* Krylov subspaces leading to good approximations which can be evaluated with little computational effort. Unfortunately the required system solves in constructing the rational Krylov space often create numerical problems or require quite some computing time.

A novel approach to obtain compact rational Krylov subspaces is to first build a large Krylov subspace and then transform this space in a numerically reliable manner to an approximate rational Krylov space. The approximation error depends of course on the size of the original Krylov subspace. In this article we will prove that the approximation converges geometrically (for increasing size of the Krylov subspace) to the actual rational Krylov space. These convergence rates can then be used to predict which size of Krylov space is required to obtain a certain accuracy in the approximation.

Key words. rational Krylov, approximation, convergence rate, matrix function, iterative method

AMS subject classifications. 65F60, 65F10, 47J25, 15A16

1. Introduction.

Many applications in science and engineering require the evaluation of expressions of the form

$$f(A)\boldsymbol{v}, \tag{1.1}$$

where $A \in \mathbb{R}^{n\times n}$ is a large, possibly sparse or structured, matrix, f is a suitable function, and $\boldsymbol{v} \in \mathbb{R}^n$ is a vector. The function $f(A)$ can be defined in terms of the spectral factorization or Jordan canonical form of A; see, e.g., Higham [14].

The evaluation of (1.1) is of interest for entire functions such as

$$f(t) = \exp(t), \quad f(t) = (1-\exp(t))/t, \quad f(t) = \cos(t), \quad f(t) = \sin(t),$$

with applications to the solution of ordinary and partial differential equations [6, 7, 10, 12, 16, 24], network analysis [8], as well as to inverse problems [4]. Other functions f of interest include $f(t) = \sqrt{t}$ with application to the solution of systems of stochastic differential equations [1], and $f(t) = \ln(t)$.

*The research was partially supported by the Research Council KU Leuven, projects CREA-13-012 Can Unconventional Eigenvalue Algorithms Supersede the State of the Art, OT/11/055 Spectral Properties of Perturbed Normal Matrices and their Applications, and CoE EF/05/006 Optimization in Engineering (OPTEC); by the Fund for Scientific Research–Flanders (Belgium) project G034212N Reestablishing Smoothness for Matrix Manifold Optimization via Resolution of Singularities; and by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office, Belgian Network DYSCO (Dynamical Systems, Control, and Optimization). The research also is supported in part by NSF grant DMS-1115385.

[†]Department of Mathematics and Computer Science, Hanover College, Hanover, IN 47243, USA. E-mail: `jagels@hanover.edu`.

[‡]Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001 Leuven (Heverlee), Belgium. E-mail: {`thomas.mach,raf.vandebril`}`@cs.kuleuven.be`.

[§]Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: `reichel@math.kent.edu`.

Higham [14] discusses and analyzes many methods for the evaluation of $f(A)$ that can be used when the matrix A is small enough to allow factorization. We are interested in the approximation of expressions (1.1) with matrices A that are too large to factor, and discuss the reduction of such matrices to small ones with standard and rational Krylov subspace methods. The evaluation of f applied to a small matrix so obtained can be carried out with methods described in [14]. Generally, the implementation of rational Krylov subspace methods requires the solution of linear systems of equations with matrices of the form $A-\psi_j I$, where ψ_j is a prescribed pole. Typically, an LU factorization has to be computed for every distinct pole. These factorizations can be very demanding computationally when the matrix A is large and does not posses structure that can be exploited. Recently, Mach, Pranić, and Vandebril [20, 21] described an implementation of approximate rational Krylov methods that circumvents the solution of linear systems of equations by using standard Krylov subspaces of sufficiently high dimension and compressing it to the desired rational space. This paper is concerned with the investigation of the convergence properties of these methods, i.e., how quickly the approximate rational Krylov subspaces generated converge to rational Krylov subspaces when the dimension of the standard Krylov subspace is increased.

Assume for the moment that the matrix A is large and symmetric. Application of ℓ steps of the symmetric Lanczos method to A with initial vector $\boldsymbol{v}$ yields a decomposition of the form

$$AV_\ell = V_\ell T_\ell + \boldsymbol{g}_\ell \boldsymbol{e}_\ell^T, \tag{1.2}$$

where the columns of $V_\ell = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\ell] \in \mathbb{R}^{n\times\ell}$ form an orthonormal basis for the Krylov subspace

$$\mathbb{K}_\ell(A, \boldsymbol{v}) = \text{span}\{\boldsymbol{v}, A\boldsymbol{v}, \ldots, A^{\ell-1}\boldsymbol{v}\}, \tag{1.3}$$

with $\boldsymbol{v}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|_2$ and $\boldsymbol{g}_\ell \in \mathbb{R}^n$ satisfies $V_\ell^T \boldsymbol{g} = \boldsymbol{0}$. Throughout this paper, $\boldsymbol{e}_\ell = [0, \ldots, 0, 1]^T$ is the ℓth axis vector, the superscript ^{T} stands for transposition, and $\|\cdot\|_2$ denotes the Euclidean vector norm or spectral matrix norm. The matrix $T_\ell \in \mathbb{R}^{\ell\times\ell}$ is symmetric and tridiagonal. We assume to be in the generic situation, that means that ℓ is small enough so that no breakdown of the Lanczos method occurs. As a result the subdiagonal elements of T are nonvanishing. We then can approximate the expression (1.1) by

$$V_\ell f(T_\ell)\boldsymbol{e}_1 \|\boldsymbol{v}\|_2; \tag{1.4}$$

see, e.g., [3, 7] and references therein for error bounds.

The relation (1.2) and the properties of T_ℓ show that the column $\boldsymbol{v}_j$ of V_ℓ can be expressed as a polynomial in A of exact degree $j-1$ times the vector $\boldsymbol{v}$. It follows that the expression (1.4) is an approximation of (1.1) in which f is replaced by a polynomial in A of degree at most $\ell-1$. In particular, when the function f cannot be approximated well by a polynomial of fairly low degree on the spectrum of A, accurate approximation of $f(A)\boldsymbol{v}$ by an expression of the form (1.4) generally requires that a large number of Lanczos steps ℓ be carried out to determine an accurate approximant. The computation of many Lanczos steps ℓ is undesirable because this yields a large matrix T_ℓ in (1.4) and this makes the evaluation of $f(T_\ell)$ computationally demanding. It is the aim of the present paper to discuss how T_ℓ can be replaced by a matrix of smaller size by using a rational Krylov subspace instead of the standard subspace (1.3).

This replacement is important, in particular, when f depends on some parameter and $f(T_\ell)$ has to be evaluated for many values of this parameter. This situation is illustrated in Section 6 for an exponential integrator.

The rational symmetric Lanczos method can be applied to determine rational approximants of f with poles at or near singularities of the function f in the complex plane. These approximants may converge to f much faster than polynomial approximants. Therefore, the rational symmetric Lanczos method may require significantly fewer steps than than the standard symmetric Lanczos method to deliver an approximation of (1.1) of desired quality. Because of this, the development and application of rational Lanczos methods has received considerable attention in the literature; see, e.g., [2, 3, 6, 7, 11, 17–19, 22]. The main drawback of the rational Lanczos method is the already mentioned need to solve linear systems of equations.

Mach et al. [20] proposed that rational Krylov subspaces determined by a symmetric nonsingular matrix A and by poles at 0 and ∞ be approximated by a standard Krylov subspace (1.3). Assume that a rational Krylov subspace

$$\mathbb{K}_{p,q}(A,\boldsymbol{v}) = \mathrm{span}\{A^{-p+1}\boldsymbol{v}, A^{-p+2}\boldsymbol{v}, \ldots, A^{-1}\boldsymbol{v}, \boldsymbol{v}, A\boldsymbol{v}, \ldots, A^{q-1}\boldsymbol{v}\} \tag{1.5}$$

of dimension $p+q-1$ is desired. Sometimes rational Krylov subspaces of this form, with poles at zero and infinity only, are referred to as *extended* Krylov subspaces; see, e.g., [6]. The approximation method in [20] first generates a standard Krylov subspace (1.3) of dimension $\ell \geq p+q-1$ with the symmetric Lanczos method. Then the symmetric tridiagonal matrix T_ℓ in (1.2) is transformed by orthogonal similarity transformations to a symmetric block diagonal matrix with overlapping blocks of the form matching the desired rational Krylov subspace. The block structure of the latter matrix is chosen to correspond to the structure of the recursion relations for orthogonal rational functions with poles at 0 and ∞ described in [17, 18]. No linear systems of equations with the matrix A have to be solved. The subspace determined by the transformation can accurately approximate the desired rational Krylov subspace (1.5). This approach can be thought of as a scheme for approximating the expression (1.1) in three steps: i) compute an orthonormal basis for a standard Krylov subspace (1.3), ii) apply this basis to determine an orthogonal basis that approximately spans the rational Krylov subspace (1.5), and iii) consider the computed basis a basis for the rational Krylov subspace (1.5) and use it to compute a rational approximation with poles at 0 and ∞ of f. This scheme for approximating $f(A)\boldsymbol{v}$ is attractive when an accurate approximation of $f(A)\boldsymbol{v}$ can be determined in a rational Krylov subspace (1.5) of low dimension, while the required dimension of the standard Krylov subspace (1.3) is large.

The present paper aims to shed light on how the dimension ℓ of the standard Krylov subspace (1.3) should be chosen so that elements in the rational Krylov subspace (1.5) can be approximated sufficiently accurately by elements in the standard Krylov subspace. We will investigate this question with the aid of complex variable methods. Our analysis is applicable also when the matrix A is nonsymmetric and rational Krylov subspaces with several finite poles are approximated by a standard Krylov subspace. This situation is described in [21] and will be analyzed in Section 4.

This paper is organized as follows. Section 2 revisits the basic principles of the algorithm by Mach et al. [20, 21]. Section 3 considers the choice of the dimension ℓ of the standard Krylov subspace (1.3) when the rational Krylov subspace only has one finite pole such as (1.5). The approximation of rational Krylov subspaces with several finite poles is discussed in Section 4. Different ways of approximating $A^{-1}\boldsymbol{v}$ are described and compared in Section 5. Computed examples are presented in Section 6, and concluding remarks can be found in Section 7.

2. The inverse-free approximation of rational Krylov spaces. Let A be a nonsingular matrix for which we would like to compute an approximate rational Krylov subspace $\mathbb{K}_{p,q}(A, \boldsymbol{v})$ without explicit system solves. To do so we need a large oversampled standard Krylov subspace $\mathbb{K}_\ell(A, \boldsymbol{v})$, i.e., $\ell \gg p + q$. Let columns of the matrix V_ℓ form an orthonormal basis for this Krylov subspace, cf. (1.2). Let us denote the orthonormal basis vectors for the rational Krylov subspace (1.5) by $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{p+q-1}$, and set $W_{p+q-1} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{p+q-1}]$. To keep our discussion simple, we consider extended Krylov subspaces (1.5) in this section. However, suitably modified, the approach outlined applies to more general rational Krylov subspaces.

The QR factorization of the symmetric tridiagonal matrix $T_\ell = V_\ell^T A V_\ell$ in (1.2) can be computed easily with rotators. It is of the form

$$T_\ell = C_1 \cdots C_{\ell-1} R, \tag{2.1}$$

where each C_i is the identity except for the part $(i : i+1, i : i+1)$ which equals a 2×2 rotation. The matrix $R \in \mathbb{R}^{\ell \times \ell}$ is upper triangular with small bandwidth. It was shown in [20] that the matrix $W_{p+q-1}^T A W_{p+q-1}$ admits a similar factorization namely

$$W_{p+q-1}^T A W_{p+q-1} = \tilde{C}_{\sigma_1} \cdots \tilde{C}_{\sigma_{p+q-2}} \tilde{S},$$

where the $\tilde{C}_i$'s are rotators analogous to the rotators C_i, $\tilde{S} \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$ is upper triangular, and $(\sigma_1, \ldots, \sigma_{p+q-2})$ is a permutation of $(1, \ldots, p+q-2)$. The extended Krylov subspace (1.5) is constructed by starting with $p = q = 1$ and increasing either p or q by one at a time until the desired extended Krylov subspace is obtained. The permutation vector $(\sigma_1, \ldots, \sigma_{p+q-2})$ is determined by the order in which p or q are increased. For instance, a standard Krylov space (1.3) of length ℓ corresponds to the ordering $(1, 2, \ldots, \ell-1)$; a rational Krylov space of length ℓ obtained by replacing A by A^{-1} in (1.3) is associated with the ordering $(\ell-1, \ell-2, \ldots, 1)$; a CMV-like ordering $(1, 3, 5, \ldots, 2, 4, 6, \ldots)$ corresponds to an extended Krylov space (1.5) that is built up by alternatingly increasing p and q.

We outline the algorithm described in [20]. The algorithm first constructs a standard Krylov subspace $\mathbb{K}_\ell(A, \boldsymbol{v})$ of generally fairly high dimension ℓ. Substituting the factorization (2.1) into (1.2) yields

$$AV_\ell = V_\ell T_\ell + \boldsymbol{g}_\ell \boldsymbol{e}_\ell^T = V_\ell(C_1 \ldots C_{\ell-1} R) + \boldsymbol{g}_\ell \boldsymbol{e}_\ell^T.$$

Next, we apply at most $p+q-3$ unitary similarity transformations, say Q_i, to reorder the rotations C_i. More precisely, the purpose of the transformation Q_i is to position the rotator that acts on rows and columns i and $i+1$ on the proper side of the rotator that acts on the rows and columns $i-1$ and i. These similarity transformations preserve the upper triangular form of the matrix R. At the end we get, for $Q = Q_1 \ldots Q_{p+q-3}$ and $\tilde{V}_\ell = V_\ell Q$,

$$\begin{aligned} A\tilde{V}_\ell = AV_\ell Q &= V_\ell Q Q^H (C_1 \ldots C_{\ell-1} R) Q + \boldsymbol{g}_\ell \boldsymbol{e}_\ell^T Q \\ &= \tilde{V}_\ell (\tilde{C}_{\sigma_1} \cdots \tilde{C}_{\sigma_{p+q-2}} \tilde{C}_{p+q-1} \cdots \tilde{C}_{\ell-1} \tilde{R}) + \boldsymbol{g}_\ell \boldsymbol{e}_\ell^T Q, \end{aligned}$$

where $\tilde{R}$ is upper triangular. Note, that only at most the first $p+q-2$ rotators are reordered[1]. Truncating the right-hand side and left-hand side of the above expression

[1] We remark that the positioning of $\tilde{C}_{p+q-1}$ plays a minor role. Details can be found in [20].

so that only the first $p+q-1$ vectors on each side of the equality are retained yields

$$A\tilde{W}_{p+q-1} = \tilde{W}_{p+q-1}(\tilde{C}_{\sigma_1}\cdots\tilde{C}_{\sigma_{p+q-2}}\tilde{S}) + \tilde{\boldsymbol{g}}_\ell \boldsymbol{q}^T,$$

where $\boldsymbol{q} = \boldsymbol{e}_\ell^T Q \in \mathbb{R}^{p+q-1}$. In the absence of the residual $\tilde{\boldsymbol{g}}_\ell \boldsymbol{q}^T$ it would follow from the implicit Q-theorem shown in [20] that the matrix $\tilde{W}_{p+q-1}$ agrees with the desired matrix W_{p+q-1}. The size of $\|\tilde{\boldsymbol{g}}_\ell \boldsymbol{q}^T\|_2$ therefore is important for gaining insight into how well the columns of $\tilde{W}_{p+q-1}$ approximate those of W_{p+q-1}.

3. Rational Krylov subspaces with one finite pole. Let A be a nonsingular matrix and consider the problem of approximating A^{-k} for some positive integer k by a polynomial in A. This problem can be studied with the aid of conformal mappings. Let $\lambda(A)$ denote the spectrum of A and let Ω be a simply connected compact set in the complex plane $\mathbb{C}$ such that

$$\lambda(A) \in \Omega, \qquad 0 \notin \Omega. \tag{3.1}$$

Assume that the boundary Γ of Ω is a Jordan curve and introduce the analytic function ϕ that maps the set $\{w \in \mathbb{C} : |w| > 1\}$ conformally onto $\Omega_c := \bar{\mathbb{C}}\backslash\Omega$ so that $\phi(\infty) = \infty$ and $\phi'(\infty) > 0$. Here $\bar{\mathbb{C}}$ denotes the extended complex plane $\mathbb{C} \cup \{\infty\}$. We assume that ϕ is defined as a continuous and univalent function in $1 \le |w| < \infty$. Then ϕ has a Laurent expansion

$$\phi(w) = cw + d_0 + d_1 w^{-1} + d_2 w^{-2} + \dots\ ,$$

for $|w|$ sufficiently large. The coefficient c is known as the capacity of Ω; it depends on the scaling of Ω; see Gaier [9] or Walsh [25, Chapter 4] for details on the mapping ϕ.

Introduce the level curves

$$\Gamma_\rho := \{\phi(w) : |w| = \rho\}, \qquad \rho > 1. \tag{3.2}$$

Since $0 \in \Omega_c$, there is a constant $\rho_0 > 1$ such that

$$0 \in \Gamma_{\rho_0}. \tag{3.3}$$

Define the uniform norm

$$\|h\|_\Omega := \max_{z\in\Omega} |h(z)|,$$

for functions h that are analytic in the interior of Ω and continuous on Ω, and introduce the set $\mathbb{P}_\ell$ of all polynomials of degree at most ℓ.

THEOREM 3.1. *Let the set $\Omega \in \mathbb{C}$ satisfy (3.1) and let $\rho_0 > 1$ be defined by (3.3). Define the best polynomial approximant $p_\ell \in \mathbb{P}_\ell$ of z^{-k} on Ω, i.e., p_ℓ is the solution of*

$$\|z^{-k} - p_\ell(z)\|_\Omega = \min_{p\in\mathbb{P}_\ell} \|z^{-k} - p(z)\|_\Omega.$$

Then

$$\limsup_{\ell\to\infty} \|A^{-k} - p_\ell(A)\|_2^{1/\ell} \le 1/\rho_0. \tag{3.4}$$

Proof. It can be shown that

$$\limsup_{\ell\to\infty} \|z^{-k} - p_\ell(z)\|_\Omega^{1/\ell} = 1/\rho_0; \tag{3.5}$$

see, e.g., [25, Chapter 4, Theorem 5]. When A is diagonalizable, the theorem follows by substituting the spectral factorization of A into (3.4) and then applying (3.5). When A is defective, its Jordan decomposition can be used. We use the spectral norm in (3.4), but the bound holds for other matrix norms as well. □

The above theorem establishes geometric convergence. The rate of convergence increases with ρ_0. The size of ρ_0 depends on the choice of Ω and on the location of the origin in relation to Ω. We would like $\rho_0 > 1$ to be large. This implies that we would like the set Ω to be far away from the origin and of small size. Note that while ρ_0 is independent of k, the norm $\|A^{-k} - p_\ell(A)\|_2$ may depend on k.

The situation when A is symmetric is considered in [20]. Assume that, in addition, A is positive definite. Then Ω may be chosen as the smallest interval that contains $\lambda(A)$. This is illustrated in the following example.

Example 3.1. Let all eigenvalues of the symmetric matrix A live in the interval $[c, d]$ with $0 < c < d$. Then all eigenvalues of the matrix

$$M := \frac{(d + c)I - 2A}{d - c} \tag{3.6}$$

are in the interval $[-1, 1]$. The conformal mapping

$$\phi(w) := \frac{1}{2}\left(w + w^{-1}\right) \tag{3.7}$$

maps the exterior of the unit circle to the exterior of the interval $[-1, 1]$ and is known as the Joukowski map. Its inverse is given by

$$\phi^{-1}(z) := z + \sqrt{z^2 - 1},$$

where the branch of the square root is chosen so that $|z+\sqrt{z^2 - 1}| > 1$ for $z \notin [-1, 1]$; see, e.g., Henrici [13] for a detailed discussion on the properties of ϕ^{-1}.

The transformation (3.6) maps zero to the point

$$z_0 := \frac{d + c}{d - c}.$$

The solution of $\phi(w) = z_0$ yields

$$\rho_0 := \phi^{-1}(z_0) = z_0 + \sqrt{z_0^2 - 1}.$$

For instance, $c = 1$ and $d = 2$ give $z_0 = 3$ and $\rho_0 = 3 + 2\sqrt{2}$. □

Example 2.2. Let all eigenvalues of the (possibly nonsymmetric) matrix $A \in \mathbb{R}^{n\times n}$ lie in the disk with center $c > 0$ and radius $0 < r < c$. Consider the matrix

$$M := (cI - A)/r. \tag{3.8}$$

Its eigenvalues are in the unit disk. The relevant conformal mappings are $\phi(w) := w$ and $\phi^{-1}(z) := z$. The transformation (3.8) maps zero to $z_0 := c/r$. It follows that $\rho_0 = c/r$. □

4. Rational Krylov subspaces with several finite poles. Mach et al. [21] describe the approximation of rational Krylov subspaces with several poles $\zeta_1, \zeta_2, \ldots, \zeta_p$ by a standard Krylov subspace (1.3). Define the rational Krylov subspace

$$\begin{aligned}\mathbb{K}_{p_1,\ldots,p_s,\zeta_1,\ldots,\zeta_s}(A,\boldsymbol{v}) = \mathrm{span}\{&\boldsymbol{v}, (A-\zeta_1 I)^{-1}\boldsymbol{v}, \ldots, (A-\zeta_1 I)^{-p_1}\boldsymbol{v},\\ &(A-\zeta_2 I)^{-1}\boldsymbol{v}, \ldots, (A-\zeta_2 I)^{-p_2}\boldsymbol{v}, \ldots,\\ &(A-\zeta_s I)^{-1}\boldsymbol{v}, \ldots, (A-\zeta_s I)^{-p_s}\boldsymbol{v}\}.\end{aligned} \tag{4.1}$$

When $\zeta_j = \infty$, the negative power $(A - \zeta_j I)^{-s}$ should be replaced by the positive power A^s for $s = 1, 2, \ldots, p_j$. The poles ζ_j do not have to be distinct.

Let Ω be a compact simply connected set in $\mathbb{C}$ whose boundary is a Jordan curve and assume that (3.1) holds. Similarly as in Section 3, the rate of convergence is determined by the level curves (3.2). There is a largest constant $\rho_0 > 1$ such that all poles ζ_j are on or exterior to the level curve Γ_{ρ_0}. The following result is analogous to Theorem 3.1 and can be shown in the same manner.

THEOREM 4.1. *Let the conditions of Theorem 3.1 hold and define ρ_0 as described above. Then (3.4) holds.*

Proof. Let $p_\ell \in \mathbb{P}_\ell$ be the best polynomial approximant of z^{-k} on Ω. The proof of Theorem 4.1 used Walsh [25, Chapter 4, Theorem 5]. The latter result is also valid when there are several distinct poles. It follows that

$$\limsup_{\ell\to\infty} \|z^{-k} - p_\ell(z)\|_\Omega^{1/\ell} = 1/\rho_0.$$

This shows geometric convergence. The rate of convergence depends on the distance between Ω and the closest pole ζ_j. Here "distance" is measured using the level curves (3.2). □

The results of this and the previous section provide insight into how large the dimension ℓ of the standard Krylov subspace (1.3) has to be chosen. For instance, when the poles of the rational Krylov subspaces (1.5) or (4.1) are close to the spectrum of A, then it may be necessary to choose ℓ fairly large. The fact that elements in the rational Krylov subspaces (1.5) or (4.1) are approximated by elements in the standard standard Krylov subspace (1.3) leads to an approximation error in (1.1). This error depends both on how well the basis elements of the rational Krylov subspaces (1.5) or (4.1) can be approximated by elements in the standard Krylov subspace (1.3) and the magnitude of the coefficients of the former basis in the approximation of (1.1).

5. Simultaneous and individual approximations of the inverse. It is clear that when the function f in (1.1) has a singularity close to the spectrum of A, it may be advantageous to approximate $f(A)$ by a rational function with suitably allocated poles when compared with polynomial approximation. Rational approximations are obtained with the aid of extended or rational Krylov subspaces. Druskin and Knizhnerman [6] considered matrix functions (1.1) with an entire function f and a symmetric positive definite matrix, and showed that it may be beneficial to use extended Krylov subspaces (1.5) when approximating these kind of matrix functions. The numerical method described in [6] fixes p and first determines an orthonormal basis for the rational Krylov subspace (1.5) with $p > 1$ and $q = 1$. This basis is computed with the symmetric Lanczos process applied to the matrix A^{-1} and with initial vector $\boldsymbol{v}$. The matrix A^{-1} is not formed; instead $p - 1$ linear systems of equations with the matrix A are solved by using the conjugate gradient method. Having computed an orthonormal basis for the space $\mathbb{K}_{p,1}(A, \boldsymbol{v})$, Druskin and Knizhnerman proceed to compute an orthonormal basis for the space $\mathbb{K}_{p,q}(A, \boldsymbol{v})$ for a desired $q > 1$.

The approach from Mach, et al. [20, 21] differs from the one in [6] in that we approximate all negative powers in $A^{-1}\boldsymbol{v}, \ldots, A^{-p+1}\boldsymbol{v}$, required in the extended Krylov subspace (1.5), at once. However, we will show in this section that our approach, under some minor constraints, is closely related to one used in [6]. We restrict ourselves to the analysis of a single inverse $A^{-1}\boldsymbol{v}$, more inverses are treated identically.

Let the matrix $V_\ell = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\ell] \in \mathbb{R}^{n\times\ell}$, whose columns form an orthonormal basis for the standard Krylov subspace (1.3), be available. We described in Section 2 a transformation of the form $V_\ell Q$ such that the orthogonal columns of $\tilde{V}_\ell = V_\ell Q$ approximately span an extended Krylov subspace. Assume that the columns approximately span the space

$$\mathbb{K}_{2,\ell-1}(A, \boldsymbol{v}) = \operatorname{span}\{A^{-1}\boldsymbol{v}, \boldsymbol{v}, A\boldsymbol{v}, \ldots, A^{\ell-2}\boldsymbol{v}\}.$$

Since $\operatorname{range}(\tilde{V}_\ell) = \operatorname{range}(V_\ell)$, we have

$$\boldsymbol{y} := \underset{\boldsymbol{y}\in\operatorname{range}\{V_\ell\}}{\arg\min} \|A^{-1}\boldsymbol{v} - \boldsymbol{y}\|_2 = \underset{\boldsymbol{y}\in\operatorname{range}\{\tilde{V}_\ell\}}{\arg\min} \|A^{-1}\boldsymbol{v} - \boldsymbol{y}\|_2.$$

Thus, the approximation of $A^{-1}\boldsymbol{v}$ in $\mathbb{K}_\ell(A, \boldsymbol{v})$ determined by a minimal residual method is as accurate as the approximation of $A^{-1}\boldsymbol{v}$ determined in the approximation of the rational Krylov space $\mathbb{K}_{2,\ell-1}(A, \boldsymbol{v})$. It follows that solving a sequence of linear systems of equations with the matrix A, as advocated in [6], is in exact arithmetic equivalent to our approach if the Krylov subspaces for the MINRES method are chosen of sufficiently large dimension.

6. Computed examples. This section contains two types of experiments. In Section 6.1 we illustrate that our technique for approximating matrix functions can lead to savings in computing time. Section 6.2 compares the convergence bounds of Section 4 and 6 with actual approximation errors. All experiments were executed in MATLAB.

6.1. Savings in computing time. A simple ODE was solved with an exponential integrator from EXPODE [15] and compared to the approximate rational Krylov approach. The problem $\dot{x} = \Delta x$ is a 1D-heat-equation with a sinusoidal heat distribution as starting condition, the solution at time stamp t equals $x(t) = \exp(At)x_0$. The matrix $A \in \mathbb{R}^{n\times n}$, for $n = 200$, is a symmetric tridiagonal matrix with $-0.2(n+1)^2$ on the diagonal and $0.1(n+1)^2$ on the subdiagonals.

The accuracy of both methods was roughly 7 significant digits. The timings are as follows. A single matrix exponential in EXPODE $\exp(At)$ took 0.13s and several more evaluation as in Figure 6.1 resulted in 1.5s. A smooth solution as computed by EXPODE and via 200 evaluations of the approximate rational Krylov space is shown in Figure 6.1. EXPODE required 22s, whereas the approximate approach only needed 0.65s to achieve the same result.

6.2. Prediction of the convergence. Example 6.1. In this example, the matrix from [18, Example 5.1] was used. Consider a 1000×1000 symmetric positive definite Toeplitz matrix A, having entries

$$a_{i,j} = \frac{1}{1 + |i-j|}.$$

The vector $A^{-1}\boldsymbol{v}$ was computed with MATLAB and serves as a reference. A sequence of subspaces approximating the extended Krylov subspace $\{\boldsymbol{v}, A^{-1}\boldsymbol{v}\}$ was

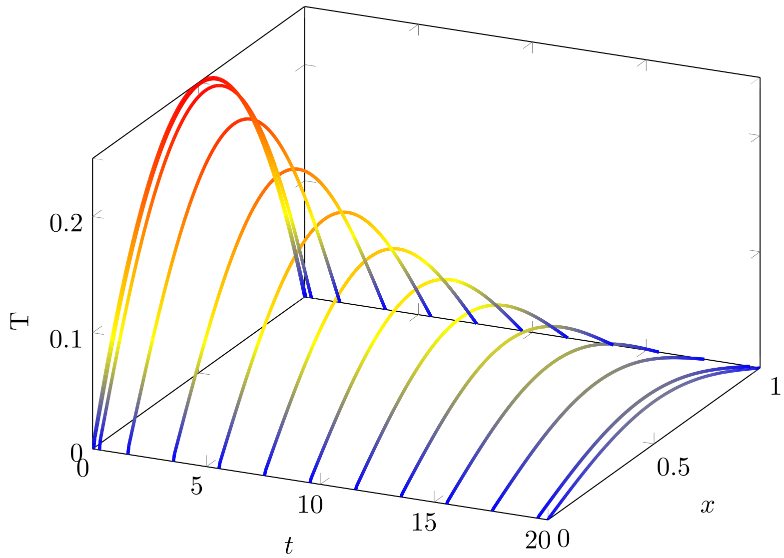

FIG. 6.1. *Solution via EXPODE, required time: 1.5s.*

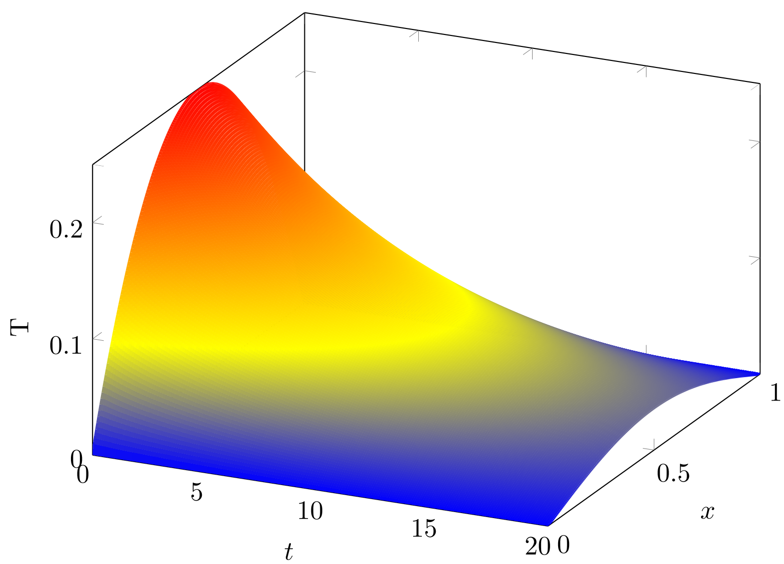

FIG. 6.2. *200 intermediate solutions via approximate rational Krylov, required time: 0.65.*

computed for varying ℓ. In Figure 6.3 the orange line reveals how well $A^{-1}\boldsymbol{v}$ is approximated in the computed subspace. The green line displays the expected convergence based on (3.4), with $\rho_0 \approx 1.4345$. We yield the same ρ_0 by computing the level curves (3.2) using the Schwarz-Christoffel toolbox [5] and by using the formulas in Example 3.1. However, the orange line seems to follow $\rho \approx 1.5263$ instead, until it reaches machine precision. But, the difference between the actual convergence and the predicted convergence is shrinking for growing n as shown in Figure 6.4.

The eigenvalues for $n = 1000$ lie between 0.3863 and 12.1259. We chose Ω to be

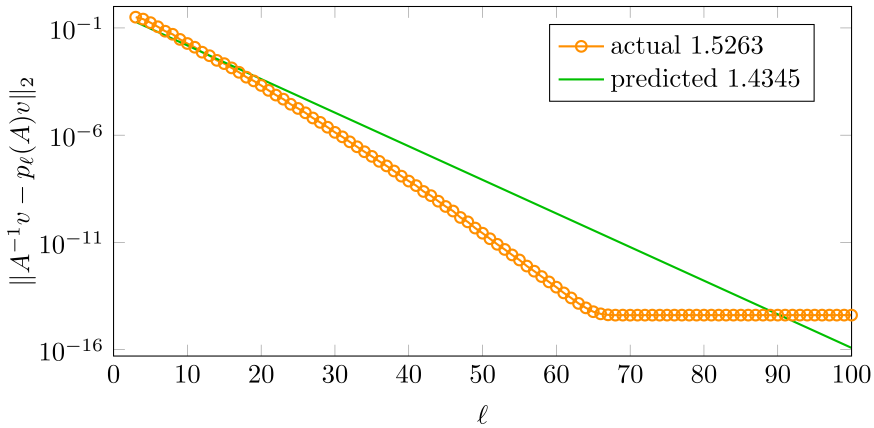

FIG. 6.3. *Example 6.1–Convergence behavior for $n = 1000$.*

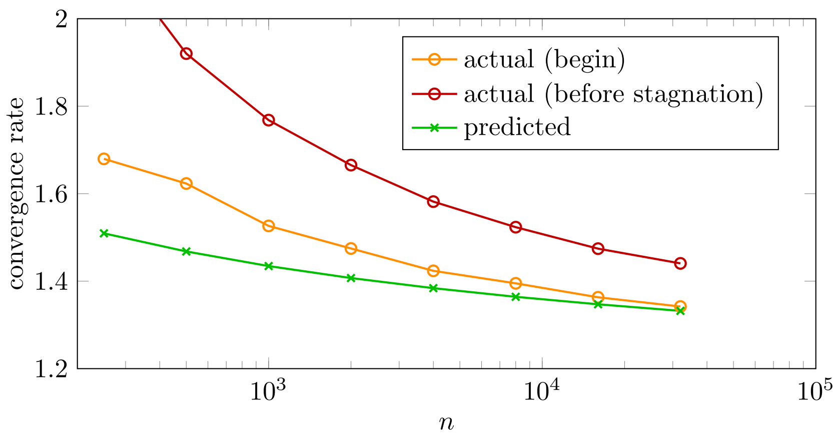

FIG. 6.4. *Example 6.1–Convergence rate over n.*

a rectangle with height $2\,\mathrm{e}{-10}$ around the interval $[0.3863, 12.1259]$. The conformal mapping computed with the Schwarz-Christoffel toolbox [5] mapping the concentric circles with radius $1, 1/0.9, 1/0.8, 1/0.7, 1/0.6, 1/0.5, 1/0.4$ is shown in Figure 6.5.

Example 6.2. We use the Toeplitz matrix with symbol

$$f(z) = 2iz^{-1} - 3i + z^2 + 0.7z^3.$$

This is a shifted version of the Toeplitz matrix "head of a bull" used by Reichel and Trefethen [23]. The results for $n = 1000$ are shown Figure 6.6. The actual convergence rate is predicted very well by the conformal mapping. We know that all eigenvalues lie within $f(\mathbb{T})$, with $\mathbb{T}$ the unit circle. Thus we used $f(\mathbb{T})$ as input for the Schwarz-Christoffel toolbox, see Figure 6.7, where $f(\mathbb{T})$ is surrounded by the orange line.

Example 6.3. Let us consider now the effect of different poles. We take the Toeplitz matrix from the first example ($n = 1000$) and divide it by 12.5 so that all eigenvalues lie in $[0, 1]$. We approximate the matrix function $f(A)v$ with

$$f(z) = \log(z) + \log(1 - z).$$

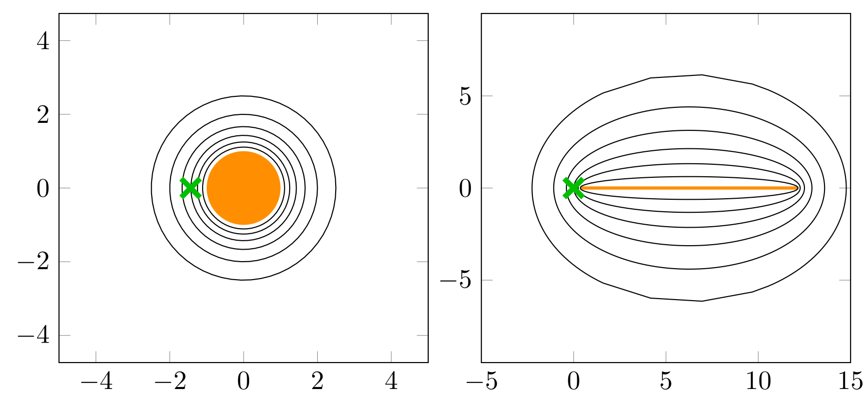

Fig. 6.5. *Example 6.1–Conformal mapping.*

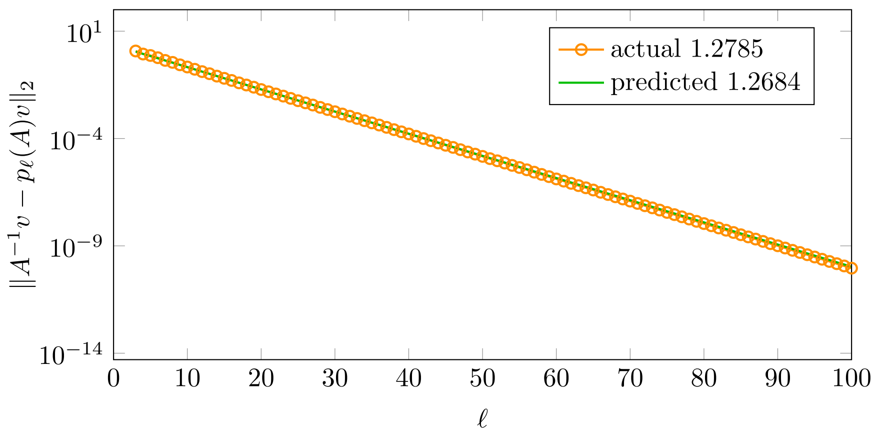

Fig. 6.6. *Example 6.2–Convergence behavior for* $n = 1000$.

For a successful approximation we require a rational Krylov subspace with poles at infinity, 0, and 1. A rational Krylov subspace where the poles $\infty, 0, \infty, 1$ are repeated 10 times, is sufficient to approximate the matrix function to $1\,\mathrm{e}{-9}$. The first predicted convergence rate is the value of the conformal mapping at 0. The second prediction is the value at 1. Unfortunately the results in Figure 6.8 are not as good as hoped for.

We repeat the same for the nonsymmetric "bull's head" Toeplitz matrix. This time we shift with 3.5 instead of $-3i$ and divide by 6. The idea is again to gather the eigenvalues inside a band $[0, 1] \times i\mathbb{R}$ near the real line. With the conformal mapping we can now predict a convergence rate for the pole 0 and for the pole 1. The results are much better now and shown in Figure 6.9.

7. Conclusion. In this article we analyzed the methods proposed by Mach et al. [20, 21] to construct approximate rational Krylov subspaces. We were able to deduce convergence rates based on complex analysis methods, predicting how large the original Krylov subspace prior to contraction had to be. We also showed the equivalence with Krylov subspace methods for approximating the matrix inverses. All statements were validated in the numerical experiments section, also illustrat-

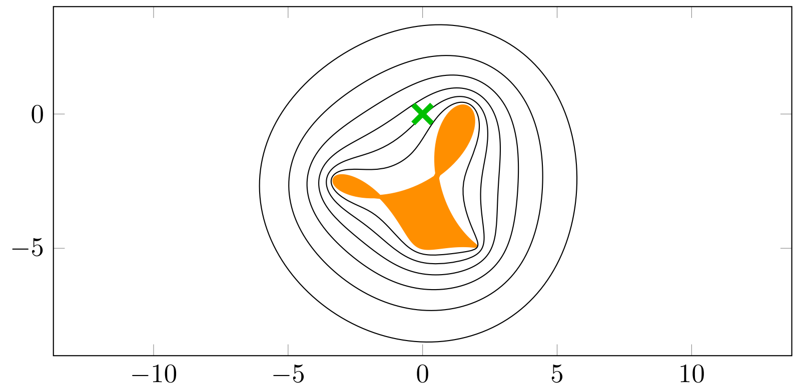

Fig. 6.7. *Example 6.2–Conformal mapping.*

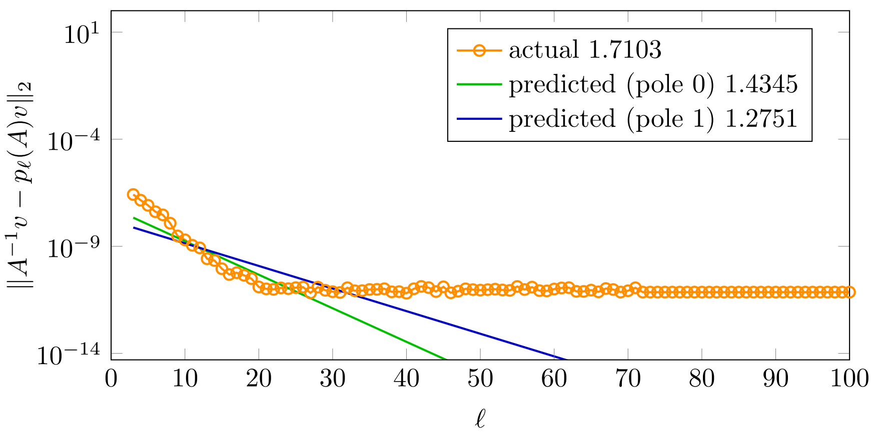

Fig. 6.8. *Example 6.3– symmetric matrix*

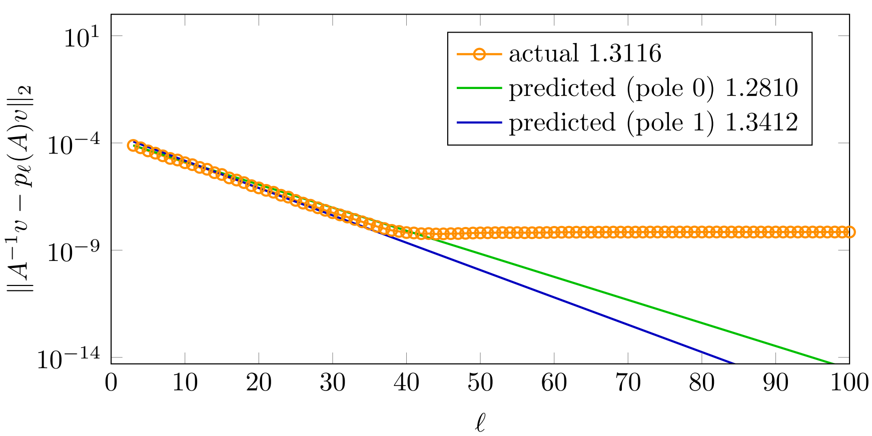

Fig. 6.9. *Example 6.3–"head of bull"*

ing the possible computational savings when using the approximate Krylov subspace techniques for an exponential integrator.

REFERENCES

[1] E. J. Allen, J. Baglama, and S. K. Boyd, Numerical approximation of the product of the square root of a matrix with a vector, Linear Algebra Appl., 310 (2000), pp. 167–181.

[2] B. Beckermann and S. Güttel, Superlinear convergence of the rational Arnoldi method for the approximation of matrix functions, Numer. Math., 121 (2012), pp. 205–236.

[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] D. Calvetti and L. Reichel, Lanczos-based exponential filtering for discrete ill-posed problems, Numer. Algorithms, 29 (2002), pp. 45–65.

[5] T. A. Driscoll and L. N. Trefethen, The Schwarz-Christoffel toolbox for MATLAB v2.3, http://www.math.udel.edu/~driscoll/SC/.

[6] V. Druskin and L. Knizhnerman, Extended Krylov subspaces: approximation of the matrix square root and related functions, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 755–771.

[7] 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.

[8] E. Estrada and D. J. Higham, Network properties revealed through matrix functions, SIAM Rev., 52 (2010), pp. 696–714.

[9] D. Gaier, Lectures on Complex Approximation, Birkhäuser, Basel, 1987.

[10] E. Gallopoulos and Y. Saad, Efficient solution of parabolic equations by Krylov approximation methods, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 1236–1264.

[11] S. Güttel and L. Knizhnerman, A black-box rational Arnoldi variant for Cauchy–Stieltjes matrix functions, BIT, 53 (2013), pp. 595–616.

[12] G. I. Hargreaves and N. J. Higham, Efficient algorithms for the matrix cosine and sine, Numer. Algorithms, 40 (2005), pp. 383–400.

[13] P. Henrici, Applied and Computational Complex Analysis, vol. 1, Wiley, New York, 1974.

[14] N. J. Higham, Functions of Matrices: Theory and Computation, SIAM, Philadelphia, 2008.

[15] M. Hochbruck and A. Ostermann, Exponential integrators, Acta Numer., 19 (2010), pp. 209–286.

[16] M. Hochbruck and C. Lubich, On Krylov subspace approximations to the matrix exponential operator, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.

[17] C. Jagels and L. Reichel, The extended Krylov subspace method and orthogonal Laurent polynomials, Linear Algebra Appl., 431 (2009), pp. 441–458.

[18] C. Jagels and L. Reichel, Recursion relations for the extended Krylov subspace method, Linear Algebra Appl., 434 (2011), pp. 1716–1732.

[19] 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.

[20] T. Mach, M. S. Pranić, and R. Vandebril, Computing approximate extended Krylov subspaces without explicit inversion, Electron. Trans. Numer. Anal., 40 (2013), pp 414–435.

[21] T. Mach, M. S. Pranić, and R. Vandebril, Computing approximate (block) rational Krylov subspaces without explicit inversion with extensions to symmetric matrices, Electron. Trans. Numer. Anal., 43 (2014), pp 100–124.

[22] M. S. Pranić and L. Reichel, Rational Gauss quadrature, SIAM J. Numer. Anal., 52 (2014), pp. 832–851.

[23] L. Reichel and L. N. Trefethen, Eigenvalues and pseudo-eigenvalues of Toeplitz matrices, Linear Algebra Appl., 162–164 (1992), pp. 153–185.

[24] R. B. Sidje, Expokit: a software package for computing matrix exponentials, ACM Trans. Math. Software, 24 (1998), pp. 130–156.

[25] J. L. Walsh, Interpolation and Approximation by Rational Functions in the Complex Domain, 5th ed., Amer. Math. Society, Providence, 1969.