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

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

Abstract. This article deduces geometric convergence rates for approximating matrix functions via inverse-free rational Krylov methods. In applications one frequently encounters matrix functions such as the matrix exponential or matrix logarithm; often the matrix under consideration is too large to compute the matrix function directly, and Krylov subspace methods are used to determine a reduced problem. If many evaluations of a matrix function of the form $f(A)\boldsymbol{v}$ with a large matrix A are required, then it may be advantageous to determine a reduced problem using *rational* Krylov subspaces. These methods may give more accurate approximations of $f(A)\boldsymbol{v}$ with subspaces of smaller dimension than standard Krylov subspace methods. Unfortunately, the system solves required to construct an orthogonal basis for a rational Krylov subspace may create numerical difficulties and/or require excessive computing time. This paper investigates a novel approach to determine an orthogonal basis of an approximation of a rational Krylov subspace of (small) dimension from a standard orthogonal Krylov subspace basis of larger dimension. The approximation error will depend on properties of the matrix A and on the dimension of the original standard Krylov subspace. We show that our inverse-free method for approximating the rational Krylov subspace converges geometrically (for increasing dimension of the standard Krylov subspace) to a rational Krylov subspace. The convergence rate may be used to predict the dimension of the standard Krylov subspace necessary to obtain a certain accuracy in the approximation. Computed examples illustrate the theory developed.

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 [18].

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 [7–9,13, 16,19,30], network analysis [10], as well as to inverse problems [5]. Other functions f

*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 Program, 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`.

of interest include $f(t) = \sqrt{t}$ with application to the solution of systems of stochastic differential equations [2], and $f(t) = \ln(t)$.

Higham [18] 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 [18].

Generally, the implementation of rational Krylov subspace methods requires the solution of linear systems of equations with matrices of the form $A - \zeta_j I$, where ζ_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 [25, 26] described an implementation of approximate rational Krylov methods that circumvents the solution of linear systems of equations by using a standard Krylov subspace of sufficiently high dimension and compressing it to approximate the desired rational space. This paper is concerned with 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[1] of the symmetric Lanczos method to A with initial vector $\boldsymbol{v} \neq \boldsymbol{0}$ 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

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

with $\boldsymbol{v}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|_2$. The vector $\boldsymbol{g}_\ell \in \mathbb{R}^n$ satisfies $V_\ell^T \boldsymbol{g}_\ell = \boldsymbol{0}$. Throughout this paper, $\boldsymbol{e}_j = [0, \ldots, 0, 1, 0, \ldots, 0]^T$ is the jth 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. Then the subdiagonal entries of T_ℓ are nonvanishing. This is the generic situation. We approximate the expression (1.1) by

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

see, e.g., [4, 8] and references therein for error bounds.

The relation (1.2) and the fact that T_ℓ is tridiagonal with nonvanishing subdiagonal entries 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) requires that a large number of Lanczos steps ℓ be carried out to determine an accurate approximant.

[1] We assume that no breakdown of the Lanczos method occurs.

The computation of many Lanczos steps ℓ is undesirable because it 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 an approximation of a rational Krylov subspace instead of the standard Krylov subspace (1.3). This replacement is important, e.g., when f depends on a 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 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., [3,4,7,8,15,21,22,24,28]. The main drawback of the rational Lanczos method is the already mentioned need to solve linear systems of equations.

Mach et al. [25] 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

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

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., [7]. The approximation method in [25] 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 a form that matches that of 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 [21,22]. 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.

Extensions of the Implicit Q Theorems proved in [25, 26] show that when the polynomial Krylov subspace (1.3) contains the rational Krylov subspace (1.5), the methods proposed by Mach et al. [25,26] determine rational approximants. Computed examples in [25,26] illustrate that when the rational Krylov subspace (1.5) is close to being contained in the polynomial Krylov subspace (1.3), the computed approximants are close to rational approximants. Error bounds are provided in [25]. 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 and of fairly low dimension are approximated by linear combinations of elements of a standard Krylov subspace of higher dimension. This situation is described in [26] 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. [25, 26]. 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.

We finally remark that other approaches to reduce the size of the matrix T_ℓ in (1.4) have been proposed. They use restarting and are discussed, e.g., by Afanasjew et al. [1], Frommer et al. [11], and Meerbergen and Spence [27]. These methods are quite different from the ones in [25, 26], and we therefore will not dwell on them further.

2. Inverse-free approximation of rational Krylov subspaces. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric nonsingular matrix for which we would like to compute an approximation of the rational Krylov subspace $\mathbb{K}_{p,q}(A,\boldsymbol{v})$ without explicit system solves. To achieve this, we first determine a standard Krylov subspace $\mathbb{K}_\ell(A,\boldsymbol{v})$. In applications, we will choose $\ell \gg p+q$, but for now we let $\ell = p+q-1$.

The symmetric Lanczos method applied to A with initial vector $\boldsymbol{v}$ yields, after $\ell-1$ steps, the decomposition (1.2). We assume that $\dim(\mathbb{K}_\ell(A,\boldsymbol{v})) = \ell$; otherwise the Lanczos process breaks down. Breakdown is rare but fortuitous as will be explained below.

It is shown in [22] that the rational Lanczos method for generating an orthonormal basis $\{\boldsymbol{w}_j\}_{j=1}^{p+q-1}$ for the rational Krylov subspace $\mathbb{K}_{p,q}(A,\boldsymbol{v})$ satisfies short recursion relations. Define the matrix $W_{p+q-1} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{p+q-1}] \in \mathbb{R}^{n\times(p+q-1)}$. Then the recursion relations for the rational Lanczos method can be expressed as

$$AW_{p+q-1} = W_{p+q-1}H_{p+q-1} + G_{p+q-1}, \tag{2.1}$$

where $H_{p+q-1} \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$ is a structured rank matrix, whose precise structure depends on the order in which positive and negative powers of A are introduced when forming an orthonormal basis for the rational Krylov subspace $\mathbb{K}_{p,q}(A,\boldsymbol{v})$. The remainder term $G_{p+q-1} \in \mathbb{R}^{n\times(p+q-1)}$ is a matrix of, typically, low rank such that $W_{p+q-1}^T G_{p+q-1} = O$. Moreover, $\boldsymbol{w}_1 = \boldsymbol{v}_1$. For instance, it is shown in [22] that if every negative power of A is followed by at least one positive power of A when sequentially constructing the orthonormal rational Krylov subspace basis $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots\}$, then the matrix $H_{p+q-1} \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$ is pentadiagonal. The structure of H_{p+q-1} obtained when two or more negative powers of A are used consecutively can be deduced from [28, Theorem 2.2].

Let T_ℓ be the symmetric tridiagonal matrix in (1.2). There is an orthonormal matrix $Q_\ell \in \mathbb{R}^{\ell\times\ell}$ such that

$$\widetilde{H}_\ell = Q_\ell^T T_\ell Q_\ell \tag{2.2}$$

has the same structure as the matrix H_{p+q-1} in (2.1), i.e., $\widetilde{H}_\ell$ and H_{p+q-1} have zero entries at the same positions, and the first row and column of $\widetilde{H}_\ell$ are the same as those

of T_ℓ, where we recall that $\ell = p+q-1$. We remark that the correction terms G_{p+q-1} in (2.1) and $\boldsymbol{g}_\ell \boldsymbol{e}_\ell^T Q_\ell$ obtained from (2.1) typically differ. An algorithm for carrying out the reverse similarity transformation (from $\widetilde{H}_\ell$ to T_ℓ) is described in [23, Section 4]. Substituting (2.2) into (1.2) yields

$$A\widehat{W}_\ell = \widehat{W}_\ell \widetilde{H}_\ell + \boldsymbol{g}_\ell \boldsymbol{e}_\ell^T Q_\ell, \tag{2.3}$$

where $\widehat{W} = [\widehat{\boldsymbol{w}}_1, \widehat{\boldsymbol{w}}_2, \ldots, \widehat{\boldsymbol{w}}_\ell] = V_\ell Q_\ell$.

We now examine the situation when the Lanczos method breaks down at step ℓ. Then the Krylov subspace $\mathbb{K}_\ell(A, \boldsymbol{v})$ is an A-invariant subspace.

THEOREM 2.1. *Assume that $\mathbb{K}_\ell(A, \boldsymbol{v})$ is an A-invariant subspace of dimension ℓ. Then the matrices $\widetilde{H}_\ell$ in (2.2) and H_{p+q-1} in (2.1) with $p+q-1 = \ell$ are essentially the same, and the columns of the matrices W_{p+q-1} in (2.1) and $\widehat{W}_\ell$ in (2.3) satisfy*

$$\operatorname{span}\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_j\} = \operatorname{span}\{\widehat{\boldsymbol{w}}_1, \ldots, \widehat{\boldsymbol{w}}_j\}, \qquad j = 1, 2, \ldots, \ell. \tag{2.4}$$

Proof. We have $A^{-1}\boldsymbol{v} = s(A)\boldsymbol{v}$ for some polynomial s of degree at most ℓ. The rational Lanczos method described in [22] requires the computation of $A^{-1}\boldsymbol{z}$ for vectors $\boldsymbol{z} \in \mathbb{K}_{p,q}(A, \boldsymbol{v})$. The fact that $A^{-1}\boldsymbol{z} = s(A)\boldsymbol{z}$ shows that the rational Lanczos basis determined by the rational Lanczos method is an orthonormal basis for a subspace of $\mathbb{K}_\ell(A, \boldsymbol{v})$. It is easy to see that the rational Lanczos method breaks down at step ℓ, because the standard Lanczos method does. Hence, $\ell - 1$ steps of the rational Lanczos method yield the decomposition

$$AW_\ell = W_\ell H_\ell,$$

which is analogous to (2.1), but where the columns the matrix $W_\ell = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_\ell]$ form an orthonormal basis for $\mathbb{K}_\ell(A, \boldsymbol{v})$ with $\boldsymbol{w}_1 = \boldsymbol{v}$.

Since $\mathbb{K}_\ell(A, \boldsymbol{v})$ is an A-invariant subspace, the relation (2.3) simplifies to

$$A\widehat{W}_\ell = \widehat{W}_\ell \widetilde{H}_\ell.$$

It follows from the implicit Q theorem [25, Theorem 3.1] that the matrices H_ℓ and $\widetilde{H}_\ell$ are essentially the same, i.e., $H_\ell = D\widetilde{H}_\ell D$, where $D \in \mathbb{R}^{\ell\times\ell}$ is a diagonal matrix with nontrivial entries ± 1. The property (2.4) follows from this relation between H_ℓ and $\widetilde{H}_\ell$. ∎

While it is unusual that a Krylov subspace $\mathbb{K}_\ell(A, \boldsymbol{v})$ is an A-invariant subspace of dimension ℓ, often only a small change of the matrix A is required to achieve this.

THEOREM 2.2. *The symmetric Lanczos method applied to the matrix*

$$\breve{A} = A - \boldsymbol{g}_\ell \boldsymbol{v}_\ell^T - \boldsymbol{v}_\ell \boldsymbol{g}_\ell^T$$

with initial vector $\boldsymbol{v}$ suffers from breakdown at step ℓ and gives the decomposition

$$\breve{A}V_\ell = V_\ell T_\ell,$$

where V_ℓ and T_ℓ are the same matrices as in (1.2) and $\boldsymbol{v}_\ell$ is the last column of V_ℓ.

Proof. The result follows by evaluating

$$\breve{A}V_\ell = (A - \boldsymbol{g}_\ell \boldsymbol{v}_\ell^T - \boldsymbol{v}_\ell \boldsymbol{g}_\ell^T)V_\ell = AV_\ell - \boldsymbol{g}_\ell \boldsymbol{v}_\ell^T V_\ell = \boldsymbol{0},$$

where the last equality follows from (1.2). Thus, computations with the symmetric Lanczos method cannot be continued. □

Assume that the vector $\boldsymbol{g}_\ell$ in (1.2) is nonvanishing. Then the number of Lanczos steps can be increased by one and this would result in the decomposition

$$AV_{\ell+1} = V_{\ell+1}T_{\ell+1} + \boldsymbol{g}_{\ell+1}\boldsymbol{e}_{\ell+1}^T.$$

Comparing this decomposition to (1.2) shows that

$$\boldsymbol{g}_\ell = (\boldsymbol{e}_\ell^T T_{\ell+1}\boldsymbol{e}_{\ell+1})\boldsymbol{v}_{\ell+1},$$

i.e., the norm of $\boldsymbol{g}_\ell$ is the magnitude of the last subdiagonal entry of $T_{\ell+1}$. The subdiagonal entries of T_ℓ typically decrease to zero in magnitude with increasing row number. Therefore, after sufficiently many Lanczos steps, only a small perturbation of A is required to obtain a matrix $\breve{A}$ for which Theorem 2.2 holds.

It is clear from the discussion preceding Theorem 2.1 that the norm of the remainder term $\|\boldsymbol{g}_\ell\boldsymbol{e}_\ell^T Q_\ell\|_2 = \|\boldsymbol{g}_\ell\|_2$ in (2.3) may be important for the performance of the method proposed in this paper. Assume that there is a matrix $E \in \mathbb{R}^{n\times n}$ of small norm such that

$$A^{-1}\boldsymbol{v} = s(A)\boldsymbol{v} + E\boldsymbol{v},$$

for some polynomial s of degree at most ℓ. We can compare the orthonormal basis for the rational Krylov subspace $\mathbb{K}_{p,q}(A,\boldsymbol{v})$ determined by the rational Lanczos method [22] to the orthonormal Krylov subspace basis for $\mathbb{K}_\ell(A,\boldsymbol{v})$ determined by applying $s(A)$ instead of A^{-1}, similarly as in the proof of Theorem 2.1. Such an analysis shows that when $A^{-1}\boldsymbol{v}$ can be approximated accurately by $s(A)\boldsymbol{v}$, where s is a polynomial of degree at most ℓ, the standard Krylov subspace $\mathbb{K}_\ell(A,\boldsymbol{v})$ is close to the rational Krylov subspace $\mathbb{K}_{p,q}(A,\boldsymbol{v})$. In this situation it is typically not necessary to evaluate $A^{-1}\boldsymbol{z}$ for vectors $\boldsymbol{z} \in \mathbb{K}_{p,q}(A,\boldsymbol{v})$, but the standard symmetric Lanczos method can be applied instead to determine an orthonormal basis for (1.3) for a suitable $\ell \geq p+q-1$. We will discuss how well A^{-k}, for some integer $k \geq 1$, can be approximated by a polynomial in A in Sections 3 and 4. This shows, in particular, when $A^{-k}\boldsymbol{v}$ can be approximated well by a polynomial in A times $\boldsymbol{v}$.

In applications of the method of this paper, one typically chooses $\ell \gg p+q$. An analogue of the expression (2.3) can be obtained by only keeping the first $p+q-1$ columns in each term. Let $\widetilde{H}_{p+q-1}$ denote the $(p+q-1)\times(p+q-1)$ leading principal submatrix of $\widetilde{H}_\ell$ and let $\widehat{W}_{p+q-1}$ be made up of the first $p+q-1$ columns of $\widehat{W}_\ell$. Assume for the moment that $\widetilde{H}_\ell$ is pentadiagonal. Then

$$\begin{aligned} A\widehat{W}_{p+q-1} &= \widehat{W}_{p+q-1}\widetilde{H}_{p+q-1} + [\widehat{\boldsymbol{w}}_{p+q}, \widehat{\boldsymbol{w}}_{p+q+1}]M[\boldsymbol{e}_{p+q-2}, \boldsymbol{e}_{p+q-1}]^T \\ &\quad + [\boldsymbol{g}_\ell\boldsymbol{e}_\ell^T Q_\ell]_{1:n,1:p+q-1}, \end{aligned} \tag{2.5}$$

where the first column of $M \in \mathbb{R}^{2\times 2}$ is made up of elements of column $p+q$ of $\widetilde{H}_\ell$ and the second column of M vanishes or is made up of entries of the column $p+q+1$ of $\widetilde{H}_\ell$. Thus, when the Lanczos method does not break down, the closeness of the recursion relations for the rational Lanczos method and the polynomial Lanczos method depends both on the size of the remainder term $\boldsymbol{g}_\ell\boldsymbol{e}_\ell^T Q_\ell$ as well as on certain entries in the matrix $\widetilde{H}_\ell$ that are not used in the recursions. Expressions analogous to (2.5) can easily be derived when $\widetilde{H}_\ell$ has bandwidth larger than five.

Our analysis in the following sections provides bounds for the rate of the convergence of the best polynomial approximants in A to A^{-k} for some integer $k \geq 1$, but does not give error bounds. The results are helpful to identify problems for which the approach proposed by Mach et al. [25, 26] is attractive to apply. These are problems with a fast rate of convergence. The appropriate dimension ℓ of the Krylov subspace (1.3) has to be determined during application of the methods [25, 26]. How this can be done is described in Example 6.2 below.

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 $1 \leq |w| < \infty$. 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 [12] or Walsh [31, 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). Introduce 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

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

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., [31, 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. We remark that by (3.4), we have

$$\limsup_{\ell\to\infty} \|A^{-k}\boldsymbol{v} - p_\ell(A)\boldsymbol{v}\|_2^{1/\ell} \leq 1/\rho_0. \tag{3.6}$$

Thus, the bound (3.4) applies to the approximation problem of interest in [25]. For certain special vectors $\boldsymbol{v}$, such as vectors that are void of components of certain eigenvectors of A, the left-hand side of (3.6) may be much smaller than the left-hand side of (3.4). However, for most vectors $\boldsymbol{v}$ the left-hand sides of (3.4) and (3.6) are of comparable size.

The situation when A is symmetric is considered in [25]. 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

$$M = \frac{(d+c)I - 2A}{d-c} \tag{3.7}$$

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

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

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

$$\phi^{-1}(z) = z + \sqrt{z^2 - 1}, \tag{3.8}$$

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 [17] for a detailed discussion on the properties of ϕ^{-1}.

The transformation (3.7) 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 3.2. 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

$$M = (cI - A)/r. \tag{3.9}$$

Its eigenvalues are in the unit disk. The relevant conformal mapping is $\phi(w) = w$ with inverse $\phi^{-1}(z) = z$. The transformation (3.9) 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. [26] describe the approximation of rational Krylov subspaces with several distinct poles $\zeta_1, \zeta_2, \ldots, \zeta_p$ by a standard Krylov subspace (1.3). Define the rational Krylov subspace

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

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 pole ζ_j is used p_j times. The selection of good poles is subject of recent and ongoing research, see, e.g., [14].

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 3.1 used Walsh [31, 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). ∎

Similarly as in Section 3, we have

$$\limsup_{\ell\to\infty} \|A^{-k}\boldsymbol{v} - p_\ell(A)\boldsymbol{v}\|_2^{1/\ell} \leq 1/\rho_0;$$

see (3.6). Thus, Theorem 4.1 provides an upper bound of the rate of convergence of interest when using the method by Mach et al. [26].

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 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 on the magnitude of the coefficients of the former basis in the approximation of (1.1).

5. Simultaneous and individual approximations of the inverse. It may be advantageous to approximate the function f in (1.1) by a rational function with suitably allocated poles, rather than by a polynomial, when f has one or several singularities close to the spectrum of A. Moreover, Druskin and Knizhnerman [7] consider matrix functions (1.1) with a Markov function f and a symmetric positive definite matrix A, and show that it may be beneficial to use extended Krylov subspaces (1.5) also when approximating this kind of matrix functions.

Rational approximations of expressions of the form (1.1) can be computed with the aid of extended or rational Krylov subspaces. When the matrix A is sparse and has a structure that allows the computation of its LU or Cholesky factorizations for a reasonable cost, then the implementation of extended Krylov subspace methods is quite straightforward; see [9, 21, 22]. For instance, rational Krylov subspace methods based on the subspaces (4.1) can be implemented with the aid of LU or Cholesky factorizations of the matrices $A-\zeta_j I$ for $j = 1, 2, \ldots, p$. However, when the structure of A does not allow the computation of an LU or Cholesky factorization of A or $A-\zeta_j I$ for a reasonable cost, then other techniques must be employed. One could, for example, solve each linear system of equations with A or $A-\zeta_j I$ by a Krylov subspace method using Krylov subspaces of the form $\mathbb{K}_\ell(A, \boldsymbol{w})$ for suitable vectors $\boldsymbol{w}$. The vectors $\boldsymbol{w}$ that arise in the algorithms in [9, 21, 22] are of the form $\boldsymbol{w} = s(A)\boldsymbol{v}$, where s is a polynomial. It follows that the approximations of (1.1) computed in this manner live in a standard Krylov subspace (1.3), similarly as the approximants determined by the methods [25, 26]. The latter methods typically require the evaluation of fewer matrix-vector products with the matrix A than when each linear system of equations with the matrices A and $A-\zeta_j I$ is solved by using separate standard Krylov subspaces $\mathbb{K}_\ell(A, \boldsymbol{w})$.

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 Sections 3 and 4 with actual approximation errors. All experiments were executed in MATLAB with about 15 significant decimal digits.

6.1. Savings in computing time. A simple system of ODEs was solved with an exponential integrator from EXPODE [20] and compared to the approximate rational Krylov approach. Consider the heat equation in one space-dimension,

$$\dot{x} = \Delta x, \qquad 0 < s < 1,$$

for $x = x(t, s)$, where $t \geq 0$ denotes time. The initial condition is a sinusoidal heat distribution and x vanishes at the end points $s = 0$ and $s = 1$. The solution is discretized at equidistant nodes with $n = 200$ interior nodes and the second derivative Δ is approximated by the standard 3-point finite difference. The semi-discretized solution at time $t > 0$ is given by

$$\boldsymbol{x}(t) = \exp(At)\boldsymbol{x}_0,$$

where the matrix $A \in \mathbb{R}^{n\times n}$ is symmetric, tridiagonal, and Toeplitz with the entries $-0.2(n+1)^2$ on the diagonal and the entries $0.1(n+1)^2$ on the subdiagonal.

We have used the following parameters for EXPODE: The initial step size is 0.05, the relative and absolute error tolerances are 10^{-5}, and the maximum dimension of Krylov subspaces is 300. Since our algorithm is not exploiting symmetry we did not tell EXPODE that the matrix A is symmetric. The solver exprb, an exponential Rosenbrock-type method, is applied.

We also solved the problem using a standard Krylov subspace of dimension 180 reduced to an approximation of an extended Krylov subspace of dimension 10 with five poles at infinity and five poles at zero. The matrix exponential function is frequently approximated by extended Krylov subspaces with poles at zero and infinity. However, other choices of poles may be preferable for certain problems. Our solution approach also can be used with poles different from zero and infinity.

The accuracy of both methods is 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 required 1.5s (where s stands for seconds). A smooth solution as computed by EXPODE and via 200 evaluations of the approximate rational Krylov space is shown in Figure 6.2. EXPODE required 22s, whereas the approximate approach only needed 0.65s to achieve the same result.

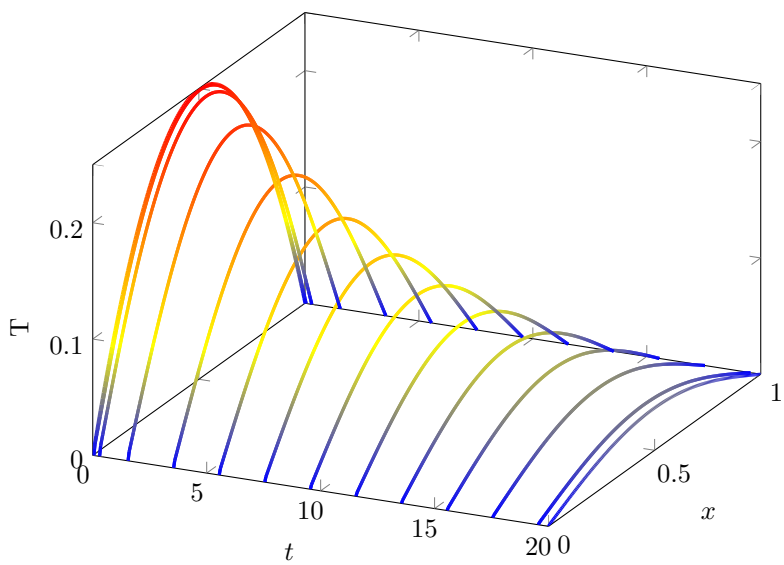

FIG. 6.1. *Solution via EXPODE, required time: 1.5s.*

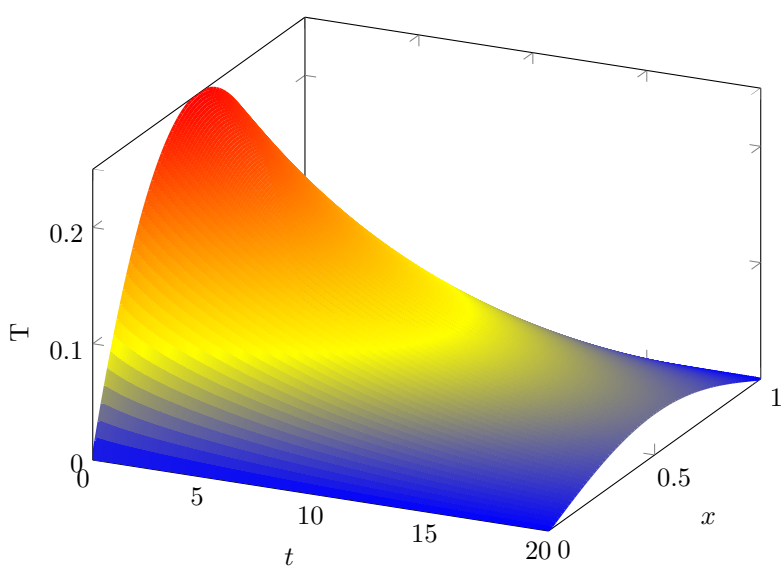

FIG. 6.2. *200 intermediate solutions via approximate rational Krylov, required time: 0.65s.*

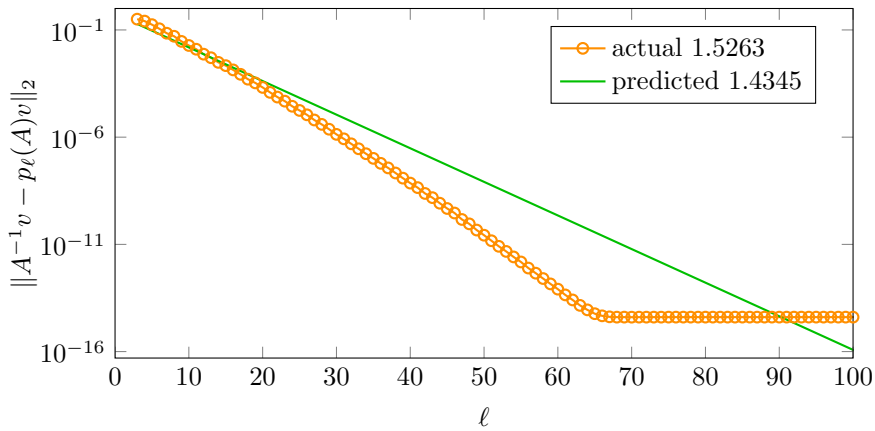

FIG. 6.3. *Example 6.1–Convergence behavior as a function of ℓ for matrix order $n = 1000$.*

6.2. Prediction of convergence rates.

Example 6.1. In this example, the matrix from [22, Example 5.1] was used. Consider the symmetric positive definite Toeplitz matrix $A \in \mathbb{R}^{1000\times 1000}$ with entries

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

As in [22, Example 5.1] we use a random vector $\boldsymbol{v}$.

The vector $A^{-1}\boldsymbol{v}$ was computed using the MATLAB command $A\backslash\boldsymbol{v}$ and serves as reference. A sequence of subspaces (1.3) approximating the extended Krylov subspace $\{\boldsymbol{v}, A^{-1}\boldsymbol{v}\}$ was computed for increasing dimension 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 obtain the same ρ_0-value by computing the level curves (3.2) by using the Schwarz–Christoffel toolbox [6] and 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 increasing matrix order n as shown in Figure 6.4.

We have so far not discussed how to determine the dimension ℓ of the standard Krylov subspace (1.3). Algorithm 1 describes a heuristic approach for symmetric positive definite matrices A. It is based on comparing the approximation error for a fairly small value of ℓ, and then apply the mapping (3.8) to estimate the actual dimension required to achieve the desired accuracy. In case the result with the ℓ-value determined by the algorithm is not sufficiently accurate, Algorithm 1 can be used to get a new estimate by using the determined ℓ-value as input. Related approaches can also be developed for nonsymmetric matrices.

In the example displayed in Figure 6.3, Algorithm 1 yields $\ell = 90$ for the target tolerance $1\,\mathrm{e}{-15}$. Since ρ_0 underestimates the convergence rate, it is not surprising that the heuristic overestimates the required dimension of the standard Krylov subspace.

The eigenvalues of $A \in \mathbb{R}^{1000\times 1000}$ lie between 0.3863 and 12.1259. We let Ω be a rectangle with height $2\,\mathrm{e}{-10}$ around the interval $[0.3863, 12.1259]$. The conformal

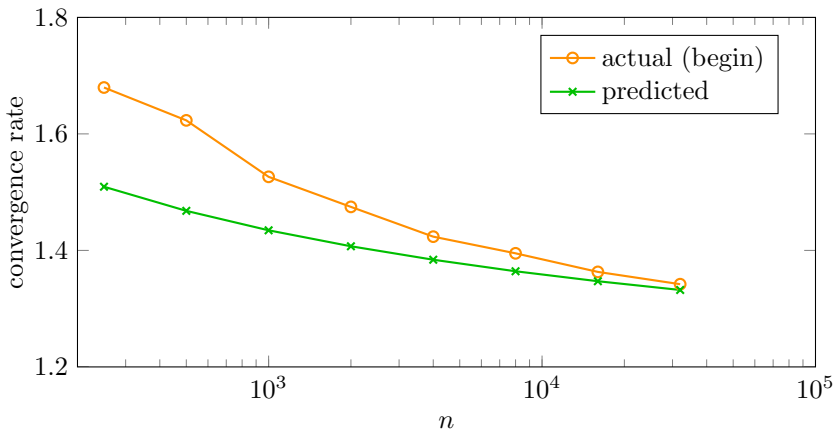

FIG. 6.4. *Example 6.1–Convergence rate over n.*

Algorithm 1 Heuristic for finding ℓ

Input: A symmetric positive definite, a target tolerance *tol*

1) Compute approximation of $\mathbb{K}_{p,q}(A, \boldsymbol{v})$ using $\mathbb{K}_{\ell_1}(A, \boldsymbol{v})$ with ℓ_1 small, say $10+p+q$.
2) Compute the approximation error $\delta = \left\|\boldsymbol{e}_\ell^T Q_\ell \mathbf{1}_{p+q}\right\|$, with $\mathbf{1}_{p+q} = [1, 1, \ldots, 1, 0, \ldots, 0]^T$ having the first $p+q$ entries equal to one.
3) Project A on $\mathbb{K}_{\ell_1}(A, \boldsymbol{v})$ to compute approximations c for the smallest eigenvalue, and d for the largest.
4) Compute $z_0 = \frac{d+c}{d-c}$ and $\rho_0 = z_0 + \sqrt{z_0^2 - 1}$.
5) Estimate the dimension ℓ by $\ell_1 + \frac{\log \delta - \log tol}{\log \rho_0}$.

mapping ϕ, computed with the Schwarz–Christoffel toolbox [6], that maps the concentric circles with radii $1, 1/0.9, 1/0.8, 1/0.7, 1/0.6, 1/0.5, 1/0.4$, is shown in Figure 6.5. The green cross in the right-hand side figure marks the origin; in the left-hand side figure the cross marks the image of the origin under ϕ^{-1}.

Schwarz–Christoffel conformal mappings can be applied to map the exterior of polygons to the exterior of the unit disk with infinity a fixed point; see Henrici [17]. To be able to use a Schwarz–Christoffel conformal mapping, we replace the interval $[0.3863, 12.1259]$ by a narrow polygon. This replacement will not affect the estimated rate of convergence significantly. We use a Schwarz–Christoffel conformal mapping because user-friendly public domain software has been made available by Driscoll and Trefethen [6], and because we would like to illustrate how rates of convergence can be computed easily. Schwarz–Christoffel conformal mappings can be applied to approximately map the exterior of fairly general sets Ω to the exterior of the unit circle with infinity a fixed point, by approximating the boundary Γ of Ω by a polygon. This is done in the following example.

Example 6.2. We use the Toeplitz matrix $A \in \mathbb{C}^{n\times n}$ with symbol

$$f(z) = 2iz^{-1} - 3i + z^2 + 0.7z^3, \qquad i = \sqrt{-1}. \tag{6.1}$$

This is a shifted version of the Toeplitz matrix "head of a bull" used in [29]. 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 $f(\mathbb{T})$,

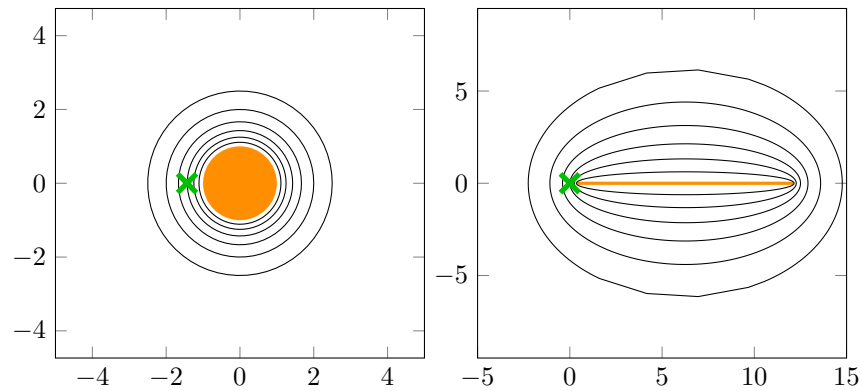

FIG. 6.5. *Example 6.1–Conformal mapping.*

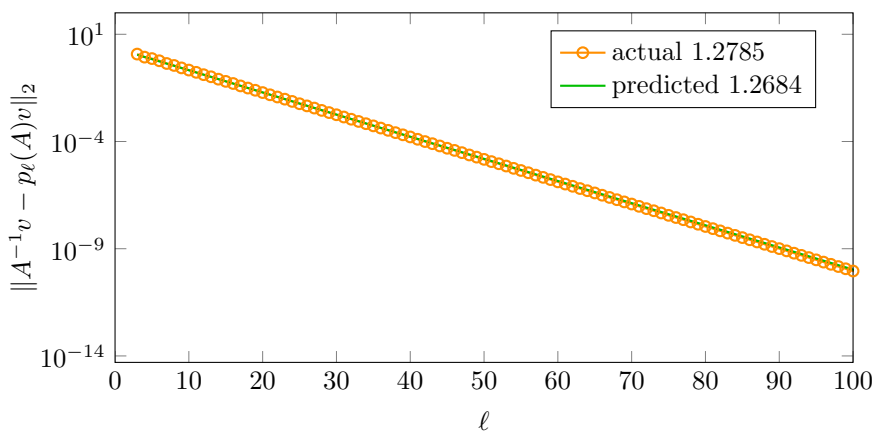

FIG. 6.6. *Example 6.2–Convergence behavior for* $n = 1000$.

where $\mathbb{T}$ denotes the unit circle in $\mathbb{C}$. Thus, we approximated $f(\mathbb{T})$ by a polygon with 210 corners, which we used as input to the Schwarz–Christoffel toolbox, see Figure 6.7, where the interior of $f(\mathbb{T})$ is displayed in orange.

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

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

and $\boldsymbol{v}$ a random vector, by using a rational Krylov subspace with poles at ∞, 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 absolute error $1\,\mathrm{e}{-9}$. The first predicted convergence rate is the value of the conformal mapping at 0. The second prediction is the value at 1. Figure 6.8 shows the actual convergence to be faster than the predicted one until the presence of round-off errors seriously affects the computed results.

We repeat the same for the nonsymmetric "bull's head" Toeplitz matrix. This time we shift by 3.5 instead of $-3i$ and divide by 6; c.f. (6.1). The idea is to gather the

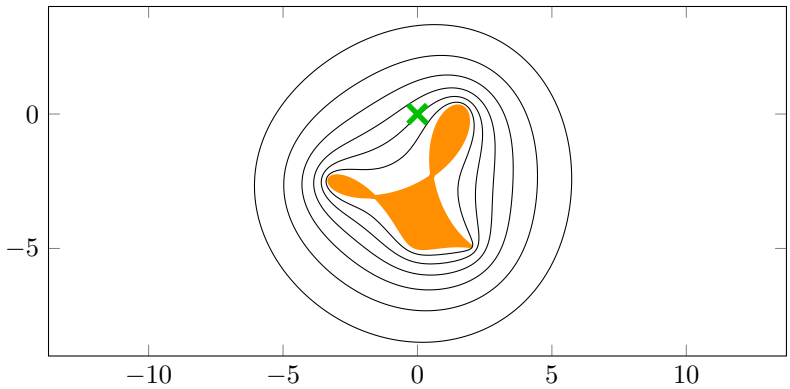

FIG. 6.7. *Example 6.2–Conformal mapping.*

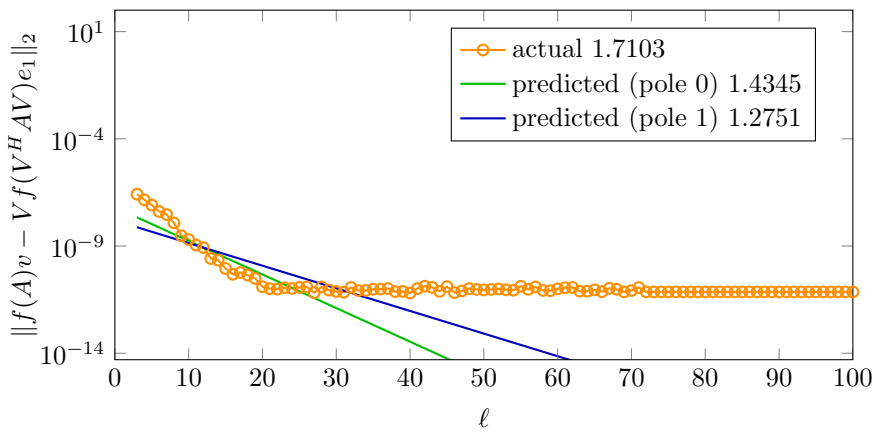

FIG. 6.8. *Example 6.3–symmetric matrix*

eigenvalues inside a band $[0,1] \times i\mathbb{R}$ near the real axis. 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.

The examples of this section illustrate that rates of convergence determined by the complex variable methods of Sections 3 and 4 provide fairly accurate predictions of actual convergence rates. Theorems 3.1 and 4.1 provide upper bounds for rates of convergence. This is verified by the computed examples that, generally, display slightly faster convergence than the predicted ones.

The convergence rates derived in this paper can be used to assess whether the methods proposed by Mach et al. [25, 26] are suitable for application to a problem at hand. This is the case when ρ_0 in Theorems 3.1 or 4.1 can be chosen significantly larger than unity, and $f(A)\boldsymbol{v}$ can be approximated well by a rational function of much smaller order than the degree required by an approximating polynomial.

We envision the analysis of the rate of convergence with conformal mappings be carried out once for a class of problems and that these problems have to be solved many times for different parameter values or initial vectors $\boldsymbol{v}$ for the Krylov subspace (1.3). Then the computational effort required for the analysis, such as the computation of the conformal mapping, is insignificant in comparison with the work required to

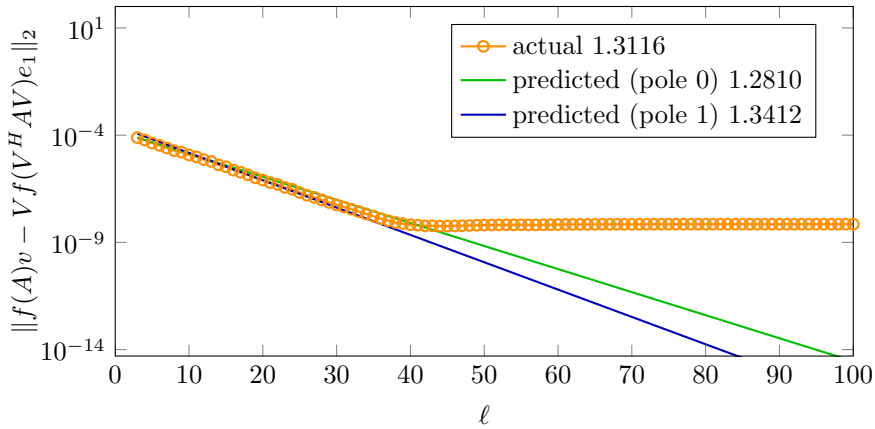

FIG. 6.9. *Example 6.3–"head of bull"*

compute approximations of many expressions $f(A)\boldsymbol{v}$. It is important that the analysis is simple to carry out for a user. We have used the Schwarz–Christoffel toolbox [6] in our computed examples to demonstrate that the computations can be carried out with public domain software that is easy to use.

7. Conclusion. In this article we analyzed the methods proposed by Mach et al. [25,26] to construct approximate rational Krylov subspaces. We deduced convergence rates based on complex analysis methods. These rates enable us to assess whether the methods proposed in [25,26] are suitable to apply. Computed examples illustrate the theory. In particular, timings show the approximate rational Krylov subspace method analyzed in this paper to be competitive with the use of an exponential integrator.

Acknowledgment. The authors would like to thank a referee for carefully reading the paper and comments that improved the presentation.

REFERENCES

[1] M. Afanasjew, M. Eiermann, O. G. Ernst, and S. Güttel, Implementation of a restarted Krylov subspace method for the evaluation of matrix functions, Linear Algebra Appl., 429 (2008), pp. 2293–2314.

[2] 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.

[3] 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.

[4] 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.

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

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

[7] 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.

[8] 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.

[9] V. Druskin, R. Remis, and M. Zaslavsky, An extended Krylov subspace model-order reduction technique to simulate wave propagation in unbounded domains, J. Comput. Phys., 272 (2014), pp. 608–618.

[10] E. Estrada and D. J. Higham, Network properties revealed through matrix functions, SIAM Rev., 52 (2010), pp. 696–714.
[11] A. Frommer, S. Güttel, and M. Schweitzer, Convergence of restarted Krylov subspace methods for Stieltjes functions of matrices, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1602–1624.
[12] D. Gaier, Lectures on Complex Approximation, Birkhäuser, Basel, 1987.
[13] E. Gallopoulos and Y. Saad, Efficient solution of parabolic equations by Krylov approximation methods, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 1236–1264.
[14] S. Güttel, Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection, GAMM-Mitteilungen, 36 (2013), pp. 8–31.
[15] S. Güttel and L. Knizhnerman, A black-box rational Arnoldi variant for Cauchy–Stieltjes matrix functions, BIT, 53 (2013), pp. 595–616.
[16] G. I. Hargreaves and N. J. Higham, Efficient algorithms for the matrix cosine and sine, Numer. Algorithms, 40 (2005), pp. 383–400.
[17] P. Henrici, Applied and Computational Complex Analysis, vol. 1, Wiley, New York, 1974.
[18] N. J. Higham, Functions of Matrices: Theory and Computation, SIAM, Philadelphia, 2008.
[19] M. Hochbruck and C. Lubich, On Krylov subspace approximations to the matrix exponential operator, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.
[20] M. Hochbruck and A. Ostermann, Exponential integrators, Acta Numer., 19 (2010), pp. 209–286.
[21] C. Jagels and L. Reichel, The extended Krylov subspace method and orthogonal Laurent polynomials, Linear Algebra Appl., 431 (2009), pp. 441–458.
[22] C. Jagels and L. Reichel, Recursion relations for the extended Krylov subspace method, Linear Algebra Appl., 434 (2011), pp. 1716–1732.
[23] C. Jagels and L. Reichel, On the computation of Gauss quadrature rules for measures with a monomial denominator, J. Comput. Appl. Math., 286 (2015), pp. 102–113.
[24] 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.
[25] 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.
[26] 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.
[27] K. Meerbergen and A. Spence, Implicitly restarted Arnoldi with purification for the shift-invert transformation, Math. Comp., 66 (1997), pp. 667–689.
[28] M. S. Pranić and L. Reichel, Rational Gauss quadrature, SIAM J. Numer. Anal., 52 (2014), pp. 832–851.
[29] L. Reichel and L. N. Trefethen, Eigenvalues and pseudo-eigenvalues of Toeplitz matrices, Linear Algebra Appl., 162–164 (1992), pp. 153–185.
[30] R. B. Sidje, Expokit: a software package for computing matrix exponentials, ACM Trans. Math. Software, 24 (1998), pp. 130–156.
[31] J. L. Walsh, Interpolation and Approximation by Rational Functions in the Complex Domain, 5th ed., Amer. Math. Society, Providence, 1969.