Numer. Math. (2011) 118:567–586
DOI 10.1007/s00211-011-0366-3

Numerische Mathematik

Convergence analysis of the extended Krylov subspace method for the Lyapunov equation

L. Knizhnerman · V. Simoncini

Received: 31 December 2009 / Revised: 2 August 2010 / Published online: 26 February 2011

Abstract The extended Krylov subspace method has recently arisen as a competitive method for solving large-scale Lyapunov equations. Using the theoretical framework of orthogonal rational functions, in this paper we provide a general a priori error estimate when the known term has rank-one. Special cases, such as symmetric coefficient matrix, are also treated. Numerical experiments confirm the proved theoretical assertions.

Mathematics subject classification 65F10 · 65F30 · 15A06 · 93B40

1 Introduction

The numerical solution of the large-scale matrix Lyapunov equation

$$AX + XA^* + BB^* = 0 \tag{1.1}$$

is of great relevance in a large variety of scientific and engineering applications; see, e.g., [1,4,29]. Here and later on x^* denotes conjugate transposition of the vector x.

Version of 20 December, 2010.

L. Knizhnerman
Central Geophysical Expedition, house 38, building 3, Narodnogo opolcheniya St., 123298 Moscow, Russia
e-mail: mmd@cge.ru

V. Simoncini (✉)
Dipartimento di Matematica, CIRSA, Università di Bologna,
Piazza di Porta S. Donato 5, 40127 Bologna, Ravenna, Italy
e-mail: valeria@dm.unibo.it

In the following we shall assume that $B = b$ is a vector, but our results can be generalized to the case when B is a tall and slim matrix.

Projection type approaches are particularly appealing approximation methods to reduce the order of the problem, so that the projected matrix equation is solved in a much smaller approximation space. Usually some condition is imposed on the residual to produce the smaller equation so as to uniquely derive an approximate solution within the approximation space. A very common choice, the Galerkin condition, is to require that the residual matrix be orthogonal to the approximation space. Various alternative approximation spaces have been proposed in the past, starting with the now classical Krylov subspace $K_m(A, b) = \text{span}\{b, Ab, \ldots, A^{m-1}b\}$ [28]; see also [18]. Although the Galerkin procedure associated with the standard Krylov subspace has been used for a long time since it was first proposed in [28], its convergence analysis was only recently developed in [30]. Other approximation methods based on projection include the well-established ADI method, in its factorized form [25,27], the global Krylov solver [19], and the Kronecker formulation approach [16]. Except for the approach based on the Kronecker formulation, projection approaches have the convenient feature of determining an approximate solution of low rank, which can be written as $X_m = Z_m Z_m^*$, where Z_m has few columns, so that only Z_m needs to be stored.

In this paper we analyze a recently developed Galerkin-type method, which projects the original problem onto an enriched space that *extends* the Krylov subspace recurrence to inverse powers of A. This *extended Krylov subspace* is given by

$$\mathbf{EK}_m(A, b) := K_m(A, b) + K_m(A^{-1}, A^{-1}b).$$

This approximation space was first introduced in [8] to approximate the action of matrix functions to a vector. It was later used in [31] for solving the Lyapunov equation. The spaces $\mathbf{EK}_m(A, b)$ in the sequence are nested, that is

$$\mathbf{EK}_m(A, b) \subseteq \mathbf{EK}_{m+1}(A, b),$$

and it was shown in [31] that the spaces can be iteratively generated by means of a recurrence similar to that used for computing the standard Krylov subspace $K_m(A, b)$. Results in [31] demonstrate the effectiveness of the enriched space with respect to the state-of-the-art ADI method, which suffers from a somewhat difficult parameter selection. However, a complete theoretical understanding of the method's performance was not presented in [31]. A first attempt in this direction was recently proposed in [22] for A real symmetric. In this paper we solve this open problem: we provide new general estimates for the convergence rate of the extended Krylov subspace method for A real nonsymmetric with field of values in the right half-plane. Numerical results confirm the quality of our estimates.[1] We also show how these bounds can be simplified under additional hypotheses on A, e.g., A symmetric, recovering for instance the leading convergence rate in [22]. Finally, we propose a conjecture leading to an even sharper rate.

[1] Numerical experiments with data stemming from control theory problems can be found in [31, Sect. 5].

In the following we assume that A is positive definite, although possibly nonsymmetric, namely, $\Re(x^*Ax) > 0$ for any $x \in \mathbb{C}^n$, $\|x\| = 1$, where $\|\cdot\|$ is the Euclidean norm. This condition ensures that (1.1) has a unique solution. In fact, a unique solution is obtained by requiring the less strict condition that the eigenvalues of A all have positive real part. However, this latter hypothesis is not necessarily preserved by projection type methods such as those discussed in this paper. Instead, positive definiteness is preserved under projection. Throughout the paper we also assume that $\|b\| = 1$.

2 Galerkin approximate solution in the extended Krylov subspace

Let the orthonormal columns of $\mathcal{Q}_m \in \mathbb{R}^{n\times 2m}$ form a basis for $\mathbf{EK}_m(A, b)$, and let $\mathcal{H}_m := \mathcal{Q}_m^* A \mathcal{Q}_m$. We shall assume that $\mathcal{Q}_m^* b = e_1$; In particular, the basis built according to the algorithm from [31] satisfies this condition. An approximate solution $\mathcal{Q}_m Y \mathcal{Q}_m^* \approx X$ is sought by imposing that the residual $R_m = AX_m + X_m A^* + bb^*$ be orthogonal to the space, which in this case corresponds to setting $\mathcal{Q}_m^* R_m \mathcal{Q}_m = 0$, yielding

$$\mathcal{H}_m Y + Y\mathcal{H}_m^* + (\mathcal{Q}_m^* b)(b^* \mathcal{Q}_m) = 0.$$

Let Y_m be the symmetric positive semidefinite solution to the reduced Lyapunov equation above. Then an approximation to X is obtained as $X_m = \mathcal{Q}_m Y_m \mathcal{Q}_m^*$.

We are thus interested in evaluating the accuracy of the approximation by estimating the error norm $\|X - X_m\|$, where $\|\cdot\|$ is the matrix norm induced by the Euclidean vector norm. To this end, we define the field of values $W := W(A) = \{x^*Ax,\ x \in \mathbb{C}^n,\ \|x\| = 1\}$, and we assume that W is symmetric with respect to the real axis $\mathbb{R}$ and (strictly) lies in the right half-plane. The solution to Eq. (1.1) may be written as (cf., e.g., [24], [13, exercise 4.4]):

$$\begin{aligned} X &= \int_0^{+\infty} e^{-tA} b (e^{-tA} b)^* dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (i\omega I - A)^{-1} bb^* (i\omega I - A)^{-*} d\omega \\ &= -\frac{i}{2\pi} \int_{-i\infty}^{+i\infty} (zI - A)^{-1} bb^* (-zI - A^*)^{-1} dz \\ &= -\frac{i}{2\pi} \int_{\Gamma} (zI - A)^{-1} bb^* (-zI - A^*)^{-1} dz, \end{aligned} \tag{2.1}$$

where Γ is a closed contour in $\mathbb{C}\backslash[W \cup (-W)]$ homeotopic in $\overline{\mathbb{C}}\backslash[W \cup (-W)]$ to the imaginary axis $i\mathbb{R}$ being passed from $-i\infty$ to $+i\infty$.

A similar expression can be used for X_m (see also [28]), namely

$$X_m = \int_0^{+\infty} \mathcal{Q}_m e^{-t\mathcal{H}_m} e_1 (\mathcal{Q}_m e^{-t\mathcal{H}_m} e_1)^* dt$$
$$= -\frac{i}{2\pi} \int_\Gamma \mathcal{Q}_m (zI - \mathcal{H}_m)^{-1} e_1 e_1^* (-zI - \mathcal{H}_m^*)^{-1} \mathcal{Q}_m^* dz. \tag{2.2}$$

Note that these relations are common to any projection method with approximation space spanned by $\mathcal{Q}_m$, and they are not restricted to the extended Krylov space.

We can thus write the error of the approximate solution as

$$X - X_m = -\frac{i}{2\pi} \int_\Gamma \Big((zI - A)^{-1} bb^* (-zI - A^*)^{-1} - \mathcal{Q}_m (zI - \mathcal{H}_m)^{-1}$$
$$\times e_1 e_1^* (-zI - \mathcal{H}_m^*)^{-1} \mathcal{Q}_m^* \Big) dz.$$

3 Error estimates for the approximate solution

Let D denote the closed unit circle, and let $\Psi : \overline{\mathbb{C}} \backslash D \to \overline{\mathbb{C}} \backslash W$, $\Phi = \Psi^{-1}$ be the direct and inverse Riemann mappings for W (preserving infinity and having positive derivatives there). Introduce canonical contours (isolines of Green's function) and canonical domains for W

$$\Gamma_r = \{z \in \mathbb{C} \backslash W : |\Phi(z)| = r\} \quad \text{and} \quad G_r = W \cup \{z \in \mathbb{C} \backslash W : |\Phi(z)| < r\}, \quad r > 1.$$

Consider the Takenaka–Malmquist system of rational functions (see [34,26], [35, Sect. 9.1], [32, Chap. 13, Sect. 3]) for the cyclic sequence of poles $0, \infty, 0, \infty, 0, \ldots$:

$$\phi_{2l}(w) = B(w)^l, \quad \phi_{2l+1}(w) = -\frac{\sqrt{1 - |\Phi(0)|^{-2}}}{1 - \Phi(0)^{-1} w} w B(w)^l, \qquad l \in \mathbb{N}, \tag{3.1}$$

where

$$B(w) = w \left[\frac{1 - \Phi(0) w}{\Phi(0) - w} \right]. \tag{3.2}$$

Function (3.2) is a finite Blaschke product (see, e.g., [5, Sect. 1.3]).[2] As such, B possesses the properties $|B(w)| = 1$ for $|w| = 1$ and, thus,

$$|B(w)| < 1 \qquad \text{locally uniformly in the open unit disk.} \tag{3.3}$$

[2] Note that Blaschke products are also used in [3] for other computational mathematical purposes.

The functions defined in (3.1) are orthogonal and uniformly bounded on the unit circumference (see, e.g., [32]):

$$\frac{1}{2\pi}\int\limits_{|w|=1}\phi_n(w)\overline{\phi_m(w)}|dw| = \delta_{n,m}, \quad n,m\in\mathbb{N}, \tag{3.4}$$

$$\max_{n\in\mathbb{N}}\max_{|w|=1}|\phi_n(w)| < +\infty. \tag{3.5}$$

In [10] Dzhrbashyan introduced the following rational functions that are now called the Faber–Dzhrbashyan rational functions[3]:

$$M_n(z) = \frac{1}{2\pi i}\int\limits_{\Gamma_r}\frac{\phi_n(\Phi(\zeta))}{\zeta - z}d\zeta, \qquad r > 1, \quad z \notin \overline{G_r}. \tag{3.6}$$

Taking into account the definition and the properties of the Faber transformation $\mathcal{F}$ (see [33,11], [32, Sect. 7.4, Sect. 9.3], [12]), one concludes from (3.5) and (3.6) that $M_n = \mathcal{F}[\phi_n]$, therefore M_n is a rational function of type $[2l/l]$ if $n = 2l$, of type $2l+1/l+1$ if $n = 2l+1$ ($l \in \mathbb{N}$). Moreover, M_n has only a zero finite pole (of multiplicity l and, due to the convexity of W, $l+1$, respectively) and is uniformly bounded in n on W.

In [21] Kocharyan established for M_n the following generating relation

$$\frac{\Psi'(w)}{\Psi(w)-u} = \frac{1}{w}\sum_{k=0}^{\infty}\overline{\phi_k\left(\frac{1}{\overline{w}}\right)}M_k(u), \qquad |w| > 1,$$

which in our case gives

$$(zI - A)^{-1} = \frac{1}{\Phi(z)\Psi'(\Phi(z))}\sum_{k=0}^{\infty}\phi_k(\Phi(z)^{-1})M_k(A), \qquad z \notin W, \tag{3.7}$$

and

$$(-zI - A^*)^{-1} = \frac{1}{\Phi(-z)\Psi'(\Phi(-z))}\sum_{l=0}^{\infty}\phi_l(\Phi(-z)^{-1})M_l(A^*), \qquad z \notin -W. \tag{3.8}$$

[3] In the symmetric case, a similar system of rational functions (Laurent polynomials) was used in [17].

Owing to [6, Theorem 2 and formula (1)] and the uniform boundedness of M_k on W, we have[4]

$$\max_{k\in\mathbb{N}} \max\{\|M_k(A)\|, \|M_k(A^*)\|, \|M_k(\mathcal{H}_m)\|, \|M_k(\mathcal{H}_m^*)\|\} < +\infty. \tag{3.9}$$

Substituting (3.7)–(3.8) and the analogous decompositions for $\mathcal{H}_m$ into (2.1) and (2.2), respectively, we obtain [5]

$$X = -\frac{i}{2\pi} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} a_{k,l} M_k(A) b b^* M_l(A^*) \tag{3.10}$$

and

$$X_m = -\frac{i}{2\pi} \mathcal{Q}_m \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} a_{k,l} M_k(\mathcal{H}_m) e_1 e_1^* M_l(\mathcal{H}_m^*) \mathcal{Q}_m^*, \tag{3.11}$$

where

$$a_{k,l} = \int\limits_{\Gamma} \frac{\phi_k\left(\Phi(z)^{-1}\right) \phi_l\left(\Phi(-z)^{-1}\right) dz}{\Phi(z)\Psi'\left(\Phi(z)\right) \Phi(-z)\Psi'\left(\Phi(-z)\right)}. \tag{3.12}$$

Combining (3.9) with the exactness of the extended Krylov subspace for M_k, $k \le 2m-1$ (i.e., that $M_k(A)b = \mathcal{Q}M_k(H)e_1$; see [9, Lemma 3.1]) and subtracting (3.11) from (3.10), we derive the bound

$$\|X - X_m\| \lesssim \sum_{k,l\in\mathbb{N},\ \max\{k,l\}\ge 2m} |a_{k,l}| \le \sum_{k=2m}^{\infty} \sum_{l=0}^{\infty} |a_{k,l}| + \sum_{k=0}^{\infty} \sum_{l=2m}^{\infty} |a_{k,l}|, \tag{3.13}$$

the symbol $\lesssim$ denoting the same as O.[6]

It remains to estimate the coefficients $a_{k,l}$.

Lemma 3.1 *There exists a number μ, $0 < \mu < 1$, such that the coefficients in* (3.12) *satisfy the inequality*

$$a_{k,l} \lesssim \mu^{k+l}. \tag{3.14}$$

4 See related results in [2] (for Faber polynomials) and in [3, Theorem 2.1].

5 Due to (3.3) and (3.9), series (3.7) and (3.8) uniformly converge for $z \in \Gamma$, so we may exchange summation and integration.

6 The multiplicative constants hidden under $\lesssim$ symbols shall depend on W and be trackable, so our results are constructive.

Proof It follows from (3.12) with $\Gamma = i\mathbb{R}$, (3.1), (3.3) and properties of Riemann mappings that

$$a_{k,l} \lesssim \int_{i\mathbb{R}} \frac{\left|\phi_k\left(\Phi(z)^{-1}\right)\phi_l\left(\Phi(-z)^{-1}\right)\right| \cdot |dz|}{\left|\Phi(z)\Psi'\left(\Phi(z)\right)\Phi(-z)\Psi'\left(\Phi(-z)\right)\right|} \lesssim \int_{i\mathbb{R}} \frac{\left|B\left(\Phi(z)^{-1}\right)\right|^{\frac{k+l}{2}} \cdot |dz|}{\max\{1, |z|^2\}} \lesssim \mu^{k+l},$$

where $\mu = \sqrt{\max_{z\in i\mathbb{R}} \left|B\left(\Phi(z)^{-1}\right)\right|} < 1.$ □

Introduce the function

$$h(w) = B(\Phi(-\Psi(w^{-1}))^{-1}) \tag{3.15}$$

and the number

$$\rho := \max_{|w|=1} |h(w)| < 1 \tag{3.16}$$

(the inequality follows from (3.3)).

Theorem 3.2 *The following error estimate holds:*

$$\|X - X_m\| \lesssim m\rho^m. \tag{3.17}$$

Proof Let $\Gamma = \partial W$. Making in (3.12) the change of variable $z = \Psi(w^{-1}), dz = -\Psi'(w^{-1})w^{-2}dw$, we get

$$\begin{aligned} a_{k,l} &= -\int_{\partial W} \frac{\phi_k\left(\Phi(z)^{-1}\right)\phi_l\left(\Phi(-z)^{-1}\right)dz}{\Phi(z)\Psi'\left(\Phi(z)\right)\Phi(-z)\Psi'\left(\Phi(-z)\right)} \\ &= \int_{|w|=1} \phi_k(w) \frac{\phi_l(\Phi\left(-\Psi(w^{-1})\right)^{-1})dw}{w\Phi\left(-\Psi(w^{-1})\right)\Psi'\left(\Phi\left(-\Psi(w^{-1})\right)\right)} \\ &= \int_{|w|=1} \phi_k(w)h(w)^{[l/2]}g_l(w)dw = i\int_{|w|=1} \phi_k(w)h(w)^{[l/2]}g_l(w)\exp(i\,\mathrm{Arg}\,w)|dw|, \end{aligned}$$

where

$$g_l(w) = \begin{cases} \left.\frac{1}{wu\Psi'(u)}\right|_{u=\Phi\left(-\Psi(w^{-1})\right)}, & \text{if } l \text{ is even,} \\ \frac{\sqrt{1-\Phi(0)^{-2}}}{1-\Phi(0)^{-1}u^{-1}} \cdot \left.\frac{1}{w\Psi'(u)}\right|_{u=\Phi\left(-\Psi(w^{-1})\right)}, & \text{if } l \text{ is odd.} \end{cases}$$

All the functions h, g_l and $\exp(i\,\mathrm{Arg}\,w)$ are smooth on the unit circumference. Accounting for the orthogonality (3.4), the definition (3.16) and Parceval's theorem, we derive the inequalities

$$\sum_{k=0}^{\infty} a_{k,l}^2 \lesssim \int\limits_{|w|=1} \left| h(w)^{[l/2]} g_l(w) \right|^2 d|w| \lesssim \rho^l$$

and

$$\sum_{k=0}^{\infty} \sum_{l=2m}^{\infty} a_{k,l}^2 \lesssim \rho^{2m}. \tag{3.18}$$

Analogously, exploiting in (3.12) the contour $-\partial W$, we "symmetrically" obtain the inequality

$$\sum_{k=2m}^{\infty} \sum_{l=0}^{\infty} a_{k,l}^2 \lesssim \rho^{2m}. \tag{3.19}$$

Combining (3.13), (3.18) and (3.19) yields

$$\sum_{k,l \in \mathbb{N},\ \max\{k,l\} \geq 2m} a_{k,l}^2 \lesssim \rho^{2m}. \tag{3.20}$$

Setting $j_0 = \left[m \frac{\log \rho}{\log \mu} \right]$ and exploiting (3.20) and (3.14), we obtain

$$\sum_{\max\{k,l\} \geq 2m} |a_{k,l}| = \sum_{\max\{k,l\} \geq 2m,\ k+l < j_0} |a_{k,l}| + \sum_{\max\{k,l\} \geq 2m,\ k+l \geq j_0} |a_{k,l}|$$
$$\lesssim j_0 \sqrt{\sum_{\max\{k,l\} \geq 2m} a_{k,l}^2} + j_0 \mu^{j_0} \lesssim m \rho^m.$$

Substitution of this into (3.13) gives (3.17). □

The scheme of the proof is quite general for projection-type methods, as emphasized by the following remark.

Remark 3.3 With a similar approach as in the proof of Theorem 3.2, and using plain Faber polynomials instead of Faber–Dzhrbashyan rational functions, one can obtain the following error estimate for solving the same Lyapunov Eq. (1.1) with $B = b$ by means of m steps of Arnoldi's method, that is, within the standard Krylov subspace ([28]):

$$\|X - X_m\| \lesssim m\rho^m, \quad \text{where} \quad \rho = |\Phi(-\alpha_{\min})|^{-1} \quad \text{and} \quad \alpha_{\min} = \min \mathbb{R} \cap W. \tag{3.21}$$

The closed form expression for ρ is obtained due to the convexity of canonical domains for the convex compact W.[7] This result generalizes the ones in [30].

Our experience with bound (3.17), and specialized results for A symmetric ([22]) seem to show that the factor m is an artifact of our proof methodology, and that $\|X - X_m\| = O(\rho^m)$ may be a realistic estimate. For this reason, our further discussion focuses on the leading term ρ^m. In fact, relying on the results of our numerical experiments partially reported below (cf. Example 5.4–4.4), we would like to formulate the following conjecture which, if true, would considerably sharpen the result of Theorem 3.2 and of available bounds especially for ρ close to one.

Conjecture 3.4 *Under the conditions of Theorem 3.2, the following estimate is valid:*

$$\|X - X_m\| \lesssim \frac{\rho^m}{\sqrt{m}}. \tag{3.22}$$

We think that one way to improve (3.17) is to account for the non-constancy of $|h(w)|$ on the unit circumference in the integral formula for $a_{k,l}$. However, it can be realized that the estimate (3.22) cannot be derived from (3.13); thus, to achieve (3.22), a different estimation strategy should be adopted. Conjecture 3.4 appears to be particularly appropriate as $\rho \to 1$, that is for difficult problems, where ρ^m becomes looser than that in (3.22).

In the following sections we report on numerical experiments that confirm the quality of the bound in Theorem 3.2, and the appropriateness of our setting. In some specific case, such as A symmetric, the estimates are simplified thanks to the explicit knowledge of the Riemann's mapping. Unless explicitly stated, our experiments use the vector of all ones as b, normalized so as to have unit norm.

4 Numerical evidence: symmetric A

We first give an explicit form for the convergence factor ρ in terms of the extreme eigenvalues of A.

Proposition 4.1 *Let A be symmetric and positive definite with spectral interval $[\lambda_{\min}, \lambda_{\max}]$, and define $\kappa = \lambda_{\max}/\lambda_{\min}$. Then*

$$\rho = \left(\frac{\kappa^{1/4} - 1}{\kappa^{1/4} + 1}\right)^2. \tag{4.1}$$

Proof We show that the function $B(u)$ defined in (3.2) has a single local maximum for $u \in \mathbb{R}$, attained at some u_+; since such u_+ is contained in the image of $(\Phi(-\Psi(1/w)))^{-1}$ with $|w| = 1$, we will conclude that $\rho = B(u_+) = ((\kappa^{1/4} - 1)/(\kappa^{1/4} + 1))^2$.

[7] The authors thank V. Dubinin and S. Suetin for indicating the fact that this convexity follows from [14, Chap. 4, Sect. 5].

Let $c = (\lambda_{\max} + \lambda_{\min})/2$, $r = (\lambda_{\max} - \lambda_{\min})/2$. We first recall that for real segments, $\Phi(z) = (z-c)/r + \sqrt{(z-c)^2/r^2 - 1}$, or its reciprocal, so that $|\Phi(z)| \geq 1$. Moreover, $\Psi(w) = c + r(w + 1/w)/2$.

We have

$$\max_{|w|=1} \left| B\left(\frac{1}{\Phi(-\Psi(w))}\right)\right| = \max_{\lambda\in[\lambda_{\min},\lambda_{\max}]} \left| B\left(\frac{1}{\Phi(-\lambda)}\right)\right| = \max_{u\in[\ell_1,\ell_2]} |B(u)|,$$

where $\ell_1 = 1/\Phi(-\lambda_{\min})$ and $\ell_2 = 1/\Phi(-\lambda_{\max})$. The function $B = B(u)$ is positive for $\Phi(0)^{-1} < u < 0$, and because of the monotonicity of Φ, $\Phi(0)^{-1} < 1/\Phi(-\lambda_{\min}) = \ell_1$, therefore B is positive in $[\ell_1, \ell_2]$. Hence,

$$\max_{|w|=1} \left| B\left(\frac{1}{\Phi(-\Psi(w))}\right)\right| = \max_{u\in[\ell_1,\ell_2]} B(u).$$

For $u \in \mathbb{R}$, $B'(u) = \Phi(0)(u^2 - 2\Phi(0)u + 1)/(\Phi(0) - u)^2$. Therefore, B has the following two critical points: $u_\pm = \Phi(0) \pm \sqrt{\Phi(0)^2 - 1}$, where

$$\Phi(0) = \left(-\frac{c}{r} + \sqrt{\left(\frac{c}{r}\right)^2 - 1}\right)^{-1} = -\frac{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}. \tag{4.2}$$

A sign analysis of B' shows that B increases between u_- and u_+ and decreases on the right of u_+, which implies that u_+ is a local maximum point. After some little algebra we can write that $B(u_+) = B(\Phi(0) + \sqrt{\Phi(0)^2 - 1}) = u_+^2$. Explicitly writing u_+ in terms of $\lambda_{\min}, \lambda_{\max}$ we obtain

$$\begin{aligned}
|u_+| &= \left| -\frac{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}} + \sqrt{\left(\frac{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}\right)^2 - 1}\right| \\
&= \left| -\frac{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}} + \frac{2(\lambda_{\min}\lambda_{\max})^{1/4}}{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}\right| \\
&= \frac{\lambda_{\max}^{1/4} - \lambda_{\min}^{1/4}}{\lambda_{\max}^{1/4} + \lambda_{\min}^{1/4}} = \frac{\kappa^{1/4} - 1}{\kappa^{1/4} + 1}.
\end{aligned}$$

Using (4.2) it can be verified that $(-\lambda_{\max} - c)/r \leq \Phi(0) \leq (-\lambda_{\min} - c)/r$, therefore again for the monotonicity on $\mathbb{R}$ of Riemann maps for real-symmetric domains on $\mathbb{R}$, we have that $u_+ \in [\ell_1, \ell_2]$. In addition, it can also be verified that $u_- < -1 < \ell_1$. Therefore, $B(u_+)$ is the maximum in $[\ell_1, \ell_2]$, completing the proof. □

Example 4.2 We consider a 5,000 × 5,000 diagonal matrix A with eigenvalues $\lambda = c + r\cos(\theta)$, $c = (10+1/10)/2$, $r = (10-1/10)/2$, and with θ uniformly distributed in $[0, 2\pi]$, so that the eigenvalues belong to $[0.1, 10]$. These values give $\rho = 0.26987$ in (4.1). The true convergence history and the estimate are reported in the left plot of

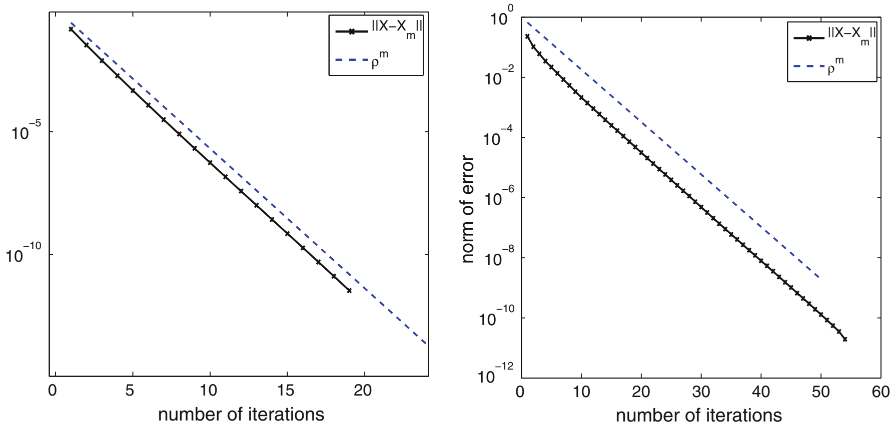

Fig. 1 True error norm and asymptotic estimate for A symmetric. Left: Spectrum in [0.1, 10]. Right: Spectrum in [0.01, 100]

Fig. 1, showing the good agreement of the a-priori bound with respect to the actual error norm. Here and in the following the Frobenius norm is used.

Example 4.3 In this example the diagonal matrix A of size $n = 10{,}000$ has higher condition number, having spectrum uniformly distributed in [0.01, 100], generated as in Example 4.2. This setting gives $\rho \approx 0.66942$. Here b is a random vector with normally distributed values and unit norm. The true convergence history and the estimate are reported in the right plot of Fig. 1, showing once again the good agreement of the leading a-priori bound ρ^m with respect to the actual error norm also for a higher condition number.

Example 4.4 We wish to numerically justify our Conjecture 3.4. To clean out the influence of adaptation to a discrete spectrum from the results of our numerical experiments, we used the rational Krylov quadrature [7, Sect. 3] initially designed for computing scalar products in the space $L_2(]-1, 1[; (1-x^2)^{-\frac{1}{2}})$. Let the Hilbert space $\mathcal{H}$ be the result of linear translation of the independent variable in $L_2([-1, 1]; (1-x^2)^{-\frac{1}{2}})$ from $[-1, 1]$ to the spectral interval of a symmetric matrix A. Let $\mathcal{A}$ be the operator of multiplication by the independent variable in $\mathcal{H}$. If A is a diagonal matrix, whose eigenvalues are the nodes of the n degree quadrature with suitably chosen shifts, and if the vector $b \in \mathbb{R}^n$ consists of the square roots of the quadrature's weights, then a rational Krylov process (and thus also the extended Krylov method) with A and b up to step $n-1$ gives the same error as the rational Krylov process in $\mathcal{H}$ with the operator $\mathcal{A}$ and the initial "vector" $\mathbf{1}$ (the constant function "unit" from $\mathcal{H}$). Since the spectral measure of the pair $(\mathcal{A}, \mathbf{1})$ is regular and its support is a segment, such an example shall not show adaptation. The true convergence curve and the conjectured rate $\rho^m/\sqrt{m}$ for these data are reported in Fig. 2, for $n = 5{,}000$ (the value of n is shown for completeness) and spectral interval [0.002, 500]. The true rate is optimally captured throughout the convergence history.

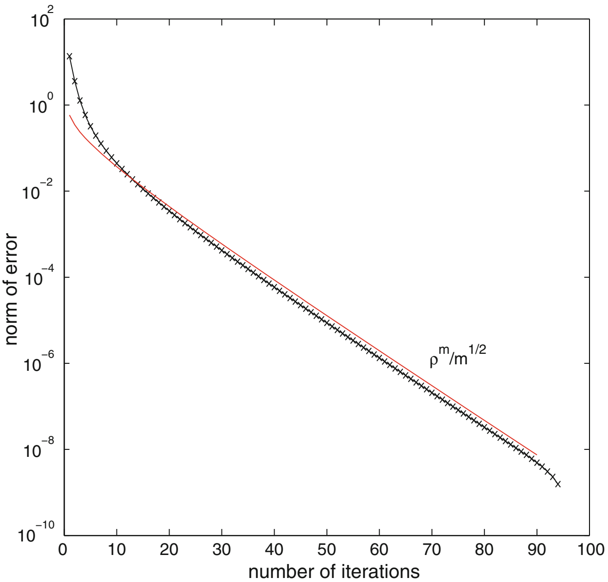

Fig. 2 Example 4.4. True error norm and asymptotic estimate (solid line) of Conjecture 3.4 for A symmetric. Spectrum in [0.002, 500]

4.1 Further considerations for A symmetric

In [22] an estimate of the error norm $\|X - X_m\|$ for A real symmetric was proposed. In our notation, the bound in [22] can be written as

$$\|X - X_m\| \leq \left(1 + \frac{\lambda_{\max}}{\lambda_{\min}}\right) \frac{1}{\widetilde{\lambda}_{\min}} \frac{\sqrt{\widetilde{\kappa}}+1}{\sqrt{\widetilde{\kappa}}} \left(\frac{\sqrt{\widetilde{\kappa}}-1}{\sqrt{\widetilde{\kappa}}+1}\right)^m, \tag{4.3}$$

where $\widetilde{\lambda}_{\min} = 2\sqrt{\lambda_{\max}\lambda_{\min}}$ and $\widetilde{\kappa} = \frac{1}{4}(\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}})^2/\sqrt{\lambda_{\max}\lambda_{\min}}$; see also the revised published version [23]. Here we propose a simple derivation of the estimate in [22] by rewriting the extended Krylov subspace approximation in terms of the Standard Krylov subspace with a suitably chosen coefficient matrix, and then by applying the bounds in [30].

We consider the two equivalent equations

$$AX + XA^* + bb^* = 0, \quad A^{-1}X + XA^{-*} + A^{-1}bb^*A^{-*} = 0$$

(A^{-*} is a short-hand notation for $(A^{-1})^*$). We multiply the second equation by γ^2, where $\gamma \in \mathbb{R}$, $\gamma > 0$, and we sum the two equations and obtain the new equivalent one:

$$(A + \gamma^2 A^{-1})X + X(A + \gamma^2 A^{-1})^* + [b, c][b, c]^* = 0, \quad c = \gamma A^{-1} b.$$

Analogously, Y solves the following two equivalent equations:

$$\mathcal{H}_m Y + Y\mathcal{H}_m^* + e_1 e_1^* = 0, \quad \mathcal{H}_m^{-1} Y + Y\mathcal{H}_m^{-*} + \mathcal{H}_m^{-1} e_1 e_1^* \mathcal{H}_m^{-*} = 0,$$

so that we can write once again

$$(\mathcal{H}_m + \gamma^2 \mathcal{H}_m^{-1})Y + Y(\mathcal{H}_m + \gamma^2 \mathcal{H}_m^{-1})^* + [e_1, \gamma \mathcal{H}_m^{-1} e_1][e_1, \gamma \mathcal{H}_m^{-1} e_1]^* = 0.$$

Therefore, we can write the analytic solutions as

$$X = \int_0^\infty \exp(-t(A + \gamma^2 A^{-1}))[b, \gamma A^{-1} b][b, \gamma A^{-1} b]^* \exp(-t(A + \gamma^2 A^{-1}))^* dt,$$

and

$$X_m = \mathcal{Q}_m Y \mathcal{Q}_m^* = \int_0^\infty x_m(t) x_m(t)^* dt,$$
$$x_m(t) = \mathcal{Q}_m \exp(-t(\mathcal{H}_m + \gamma^2 \mathcal{H}_m^{-1}))[e_1, \gamma \mathcal{H}_m^{-1} e_1].$$

We next recall from [20] that polynomials of degree up to $m-1$ in $A + \gamma^2 A^{-1}$ with vector b or $A^{-1}b$ can be represented exactly in the extended Krylov subspace, that is,

$$p_k(A + \gamma^2 A^{-1})[b, \gamma A^{-1} b] = \mathcal{Q}_m p_k(\mathcal{H}_m + \gamma^2 \mathcal{H}_m^{-1})[e_1, \gamma \mathcal{H}_m^{-1} e_1],$$

where p_k is a polynomial of degree $k \le m-1$. Therefore, the obtained approximation X_m belongs to $\mathbf{EK}_m(A, b)$. It is interesting to remark that the derivation up to this point is very general, and it also applies to the nonsymmetric case. However, it is not completely clear how the optimization performed below could be generalized to the nonsymmetric case, and whether this approach would lead to a sharp estimate of the convergence rate.

If Θ_k denotes the Chebyshev polynomial of degree k, then the exponential function of the matrix $-t(A+\gamma^2 A^{-1})$ can be expanded in terms of Chebyshev series as follows (see, e.g., [30]):

$$\exp(-t(A + \gamma^2 A^{-1}))[b, \gamma A^{-1} b] = \sum_{k=0}^\infty \alpha_k \Theta_k (A + \gamma^2 A^{-1})[b, \gamma A^{-1} b]$$
$$= \mathcal{Q}_m \exp(-t(\mathcal{H}_m + \gamma^2 \mathcal{H}_m^{-1}))[e_1, \gamma \mathcal{H}_m^{-1} e_1] + \sum_{k=m}^\infty \alpha_k \Theta_k (A + \gamma^2 A^{-1})[b, \gamma A^{-1} b].$$

This setting allows us to use the estimate in [30] for $\mathcal{A} = A + \gamma^2 A^{-1}$ symmetric. We only need to chose γ so as to minimize $\kappa(A + \gamma^2 A^{-1})$. This is determined next. But first, we would like to remark that the fact that polynomials in $A + \gamma^2 A^{-1}$ are exactly represented in the extended Krylov subspace does not necessarily imply that the obtained bound is sharp. In other words, the extended Krylov subspace solution could be the result of a possibly better rational function in A, A^{-1}. Our argument below seems to support the claim that using $A + \gamma^2 A^{-1}$ does yield the desired formulation, at least in the symmetric case.

Lemma 4.5 *Let A be symmetric with spectral interval $[\alpha, \beta]$. We have*

$$\min_{\gamma^2 \in \mathbb{R}^+} \kappa(A + \gamma^2 A^{-1}) = \kappa(A + \alpha\beta A^{-1}) = \frac{\alpha + \beta}{2} \frac{1}{\sqrt{\alpha\beta}},$$

attained at $\gamma = \sqrt{\alpha\beta}$, and $\lambda_{\min}(A + \alpha\beta A^{-1}) = 2\sqrt{\alpha\beta}$, $\lambda_{\max}(A + \alpha\beta A^{-1}) = \alpha + \beta$.

Proof Let $f_\gamma(\lambda) = \lambda + \gamma^2/\lambda$. It holds that $\min_{\lambda \in [\alpha,\beta]} f_\gamma(\lambda) = 2\gamma$, while $\max_{\lambda \in [\alpha,\beta]} f_\gamma(\lambda) = \max\left\{\alpha + \frac{\gamma^2}{\alpha}, \beta + \frac{\gamma^2}{\beta}\right\}$. By collecting the two extremes we obtain $\kappa(A + \gamma^2 A) = \frac{1}{2}\max\left\{\frac{\alpha}{\gamma} + \frac{\gamma}{\alpha}, \frac{\beta}{\gamma} + \frac{\gamma}{\beta}\right\}$. The minimum over all possible γ is obtained when the two quantities are equal, that is for $\gamma = \sqrt{\alpha\beta}$. □

Using the bound in [30, Proposition 3.1] for the symmetric matrix $\mathcal{A} = A + \alpha\beta A^{-1}$ we obtain

$$\|X - X_m\| \le 4\frac{\sqrt{\widehat{\kappa}} + 1}{2\lambda_{\min}(\mathcal{A})\sqrt{\widehat{\kappa}}}\left(\frac{\sqrt{\widehat{\kappa}} - 1}{\sqrt{\widehat{\kappa}} + 1}\right)^m \|[b, \sqrt{\alpha\beta}A^{-1}b]\|^2$$

where $\widehat{\kappa} = (\lambda_{\max}(\mathcal{A}) + \lambda_{\min}(\mathcal{A}))/(2\lambda_{\min}(\mathcal{A}))$. Using the formulas for the extreme eigenvalues in Lemma 4.5 we find $\widehat{\kappa} = (\sqrt{\alpha} + \sqrt{\beta})^2/(4\sqrt{\alpha\beta})$, which is precisely $\widetilde{\kappa}$ in (4.3). We conclude this discussion with a comparison between the rate $\frac{\sqrt{\widetilde{\kappa}}-1}{\sqrt{\widetilde{\kappa}}+1}$ and ρ given in (4.1). We have

$$\frac{\sqrt{\widetilde{\kappa}} - 1}{\sqrt{\widetilde{\kappa}} + 1} = \frac{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}} - 2\sqrt[4]{\lambda_{\min}\lambda_{\max}}}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}} + 2\sqrt[4]{\lambda_{\min}\lambda_{\max}}} = \frac{(\lambda_{\max}^{1/4} - \lambda_{\min}^{1/4})^2}{(\lambda_{\max}^{1/4} + \lambda_{\min}^{1/4})^2} = \rho.$$

Therefore, our convergence rate ρ in (4.1) coincides with that in [22].

Finally, we would like to compare the convergence rate for A symmetric, obtained by the extended Krylov subspace method, with that obtained either by the Standard Krylov method ([30]) or by the ADI method with a single pole. In the latter case, the rate is known to be ([25, (2.12)])

$$\varepsilon_{adi} \approx \left(\frac{1 - \sqrt{\kappa^{-1}}}{1 + \sqrt{\kappa^{-1}}}\right)^2, \quad \kappa = \frac{\lambda_{\max}}{\lambda_{\min}}.$$

In Fig. 3 we show the different rates of convergence (considering $\rho^{1/2}$ for the extended Krylov subspace) as $\lambda_{\max}$ varies in $[10^2, 10^5]$ and $\lambda_{\min} = 1$. The curves clearly show that the extended Krylov subspace provides the best (smallest) rate among the considered methods, with a significant gap for large condition numbers.

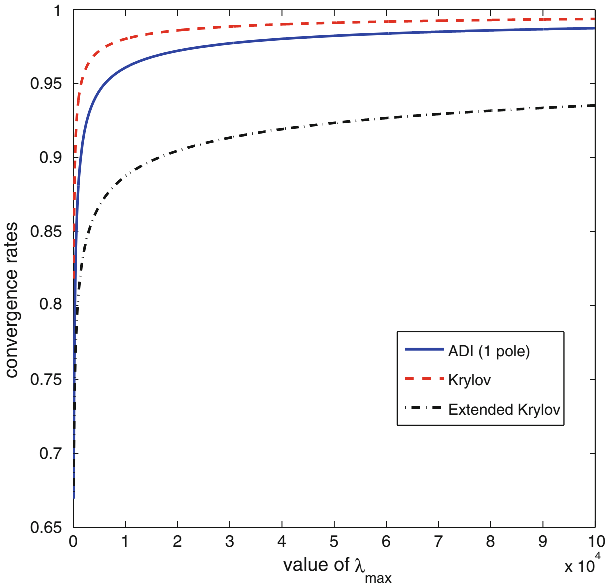

Fig. 3 Convergence rate of competing methods as the condition number of the symmetric and positive definite matrix A increases

5 Numerical evidence: nonsymmetric A

We first derive an explicit formula for ρ in the case of a disk.

Proposition 5.1 *Assume that $W(A) \subset \mathbb{C}^+$ is a disk with center $c > 0$ and radius $r < c$. Then*

$$\rho = \left| B\left(\frac{1}{\Phi(-\Psi(w_\star^{-1}))}\right) \right| = \frac{r^2}{4c^2 - 3r^2}, \quad \textit{with } \Re(w_\star) = -\frac{r}{c},\ |w_\star| = 1.$$

Proof For a disk $D(c, r)$, $\Phi(z) = (z - c)/r$, for z in the exterior of $D(c, r)$ and on $\partial D(c, r)$; moreover, $\Psi(w) = c + rw$, $|w| = 1$. Note that $\Psi(1/w) = \Psi(\bar{w})$, therefore we can directly work with $\Psi(w)$, $|w| = 1$. Let $\alpha = r/c$. After some little algebra we have

$$\begin{aligned}
\left| B\left(\frac{1}{\Phi(-\Psi(w^{-1}))}\right) \right| &= \left| \frac{rw}{2cw + r} \cdot \frac{rcw + r^2}{(2c^2 - r^2)w + cr} \right| \\
&= r^2 \left| \frac{cw + r}{(2cw + r)((2c^2 - r^2)w + cr)} \right| \\
&= \alpha^2 \left| \frac{w + \alpha}{(2w + \alpha)((2 - \alpha^2)w + \alpha)} \right| \\
&= \frac{\alpha^2}{2(2 - \alpha^2)} \left| \frac{w + \alpha}{(w + \frac{\alpha}{2})(w + \frac{\alpha}{2-\alpha^2})} \right|, \qquad (5.1)
\end{aligned}$$

where we used the fact that $|w| = 1$ and $0 < \alpha < 1$. Using once again $|w| = 1$, we have

$$\frac{|w+\alpha|^2}{|(w+\frac{\alpha}{2})(w+\frac{\alpha}{2-\alpha^2})|^2} = \frac{1+\alpha^2+2\alpha\Re(w)}{(1+\frac{\alpha^2}{4}+\alpha\Re(w))(1+\alpha^2/(2-\alpha^2)^2+2\alpha/(2-\alpha^2)\Re(w))}$$
$$= \beta_0 \frac{\frac{1+\alpha^2}{2\alpha}+\Re(w)}{\left(\frac{1+\frac{\alpha^2}{4}}{\alpha}+\Re(w)\right)\left(\frac{(1+\alpha^2/(2-\alpha^2)^2)(2-\alpha^2)}{2\alpha}+\Re(w)\right)}$$
$$= \beta_0 \frac{\beta_1+\Re(w)}{(\beta_2+\Re(w))(\beta_3+\Re(w))} =: f(\Re(w)),$$

where $\beta_j(\alpha)$ are positive constants only depending on α.

Critical points of f are obtained for $f'(\Re(w)) = 0$, whose numerator, except for the positive constant β_0, equated to zero gives

$$-\Re(w)^2 - 2\beta_1\Re(w) + \beta_2\beta_3 - \beta_1\beta_2 - \beta_1\beta_3 = 0,$$

which yields $\Re(w) = -\beta_1 \mp \sqrt{\beta_1^2 + \beta_2\beta_3 - \beta_1\beta_2 - \beta_1\beta_3}$. A sign analysis shows that the local maximum of f is obtained for the root $w_\star$ with positive square root sign. Some tedious algebra shows that the square root argument can be written as

$$\beta_1^2 + \beta_2\beta_3 - \beta_1\beta_2 - \beta_1\beta_3 = \frac{-\alpha^6+4\alpha^4-5\alpha^2+2}{4\alpha^2(2-\alpha^2)} = \frac{-(\alpha^2-1)^2(\alpha^2-2)}{4\alpha^2(2-\alpha^2)},$$

so that

$$\Re(w_\star) = -\frac{1+\alpha^2}{2\alpha} + \frac{1-\alpha^2}{2\alpha} = -\frac{r}{c}.$$

Substituting in (5.1) we find $\rho = \alpha^2/(4-3\alpha^2) = r^2/(4c^2-3r^2)$. We conclude by observing that substituting $w = \pm 1$ into (5.1) we obtain that $|B(1/(\Phi(-\Psi(\pm 1))))| = \alpha^2/(4-\alpha^2) < \rho$. □

Example 5.2 Our first experiment for A nonsymmetric aims to confirm the result of Proposition 5.1. We consider the $5,000 \times 5,000$ upper bidiagonal matrix with diagonal c and off-diagonal r, with $c = (g+1/g)/2$ and $r = (g-1/g)/2$ and $g = 6$, resulting in $W(A)$ being tightly enclosed in the disk of center c and radius r, yielding $\rho = 0.68018$. Figure 4 displays the true convergence rate and the leading asymptotic estimate ρ^m as predicted by Theorem 3.2 with the help of Proposition 5.1. The slope is well captured by the a-priori estimate.

Example 5.3 We consider an external mapping from the family used for instance in [15], [30],

$$\psi(w) = 2 + 2\left(1+\frac{1}{w}\right)^{1.5} w, \qquad |w| \geq 1. \tag{5.2}$$

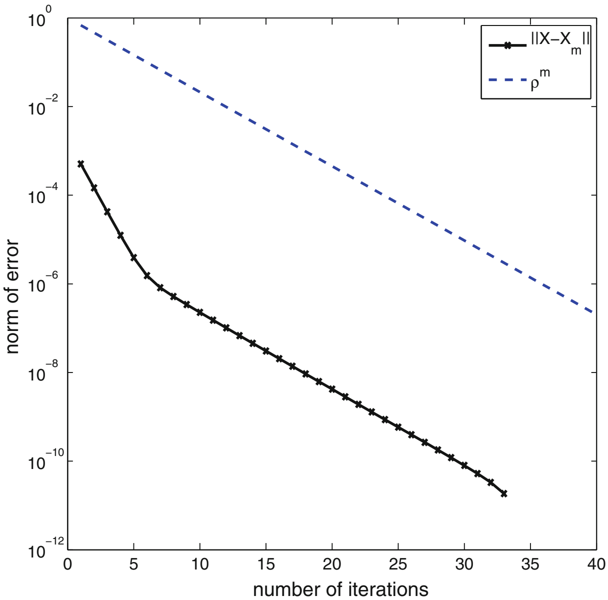

Fig. 4 Example 5.2. True error norm and asymptotic estimate ρ^m for $W(A)$ being a disk

The function ψ maps the exterior and boundary of the unit circle onto the exterior of a wedge-shaped convex set Ω in $\mathbb{C}^+$ (cf. left plot of Fig. 5). We consider the 500×500 (normal) diagonal matrix A whose eigenvalues are the image through ψ of uniformly distributed points on the unit circle. The optimal rate is computed numerically as the maximum of the function h in (3.15), yielding $\rho \approx 0.051971$. The true convergence history and the estimated convergence curve with leading term ρ^m are reported in the right plot of Fig. 5, confirming once again the good agreement of the bound.

Example 5.4 We generalize Example 4.4 to the nonsymmetric case to numerically verify our Conjecture 3.4. Since we are not aware of a rational quadrature like the one in [7, Sect. 3] for the non-Hermitian case, we content ourselves with the construction of an approximately adaptation-free test for a disk, by orthogonally lifting the eigenvalues (rational Chebyshev nodes) from the corresponding segment to points on the given disk's boundary (circumference). The data used to generate the results of Fig. 6 were obtained for $n = 100,000$ and a disk with $c = (g + 1/g)/2$, $r = (g - 1/g)/2$ and $g = 50$. Because of the problem size, the residual norm is plotted rather than the error norm. The plot shows how well the rate $\rho^m/\sqrt{m}$ captures the true (slow) asymptotic regime, which is missed by ρ^m.

6 Conclusions

In this paper we have analyzed the convergence of the extended Krylov subspace method for numerically solving large scale Lyapunov equations. Our main

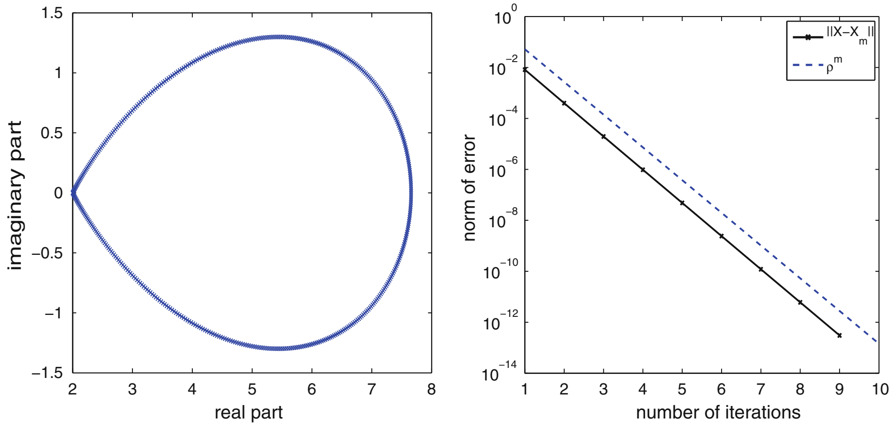

Fig. 5 Example 5.3. Left: spectrum of A. Right: True error norm and asymptotic estimate for $W(A)$ enclosed by a wedge-shaped curve

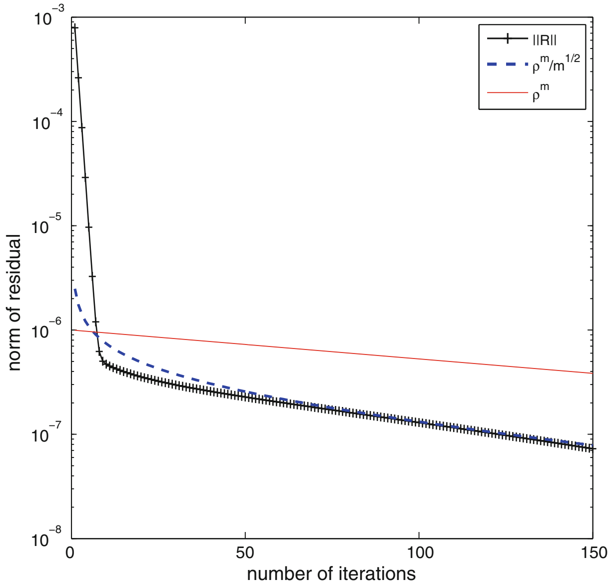

Fig. 6 Example 5.4. Residual norm and estimate of Conjecture 3.4 for the disk W:$W \cap \mathbb{R} = [0.02, 50]$. A normal with spectrum on ∂W

Theorem 3.2 shows that after m iterations the error norm decreases as least as $m\rho^m$, where ρ is related to information on the field of values of the coefficient matrix. We also conjecture that the faster rate $\rho^m/\sqrt{m}$ should hold, which appears to be appropriate in some specifically chosen tough numerical examples.

Acknowledgments The authors thank L. Reichel, B. Beckermann, C.-C. Kjelgaard Mikkelsen and the referees for their comments and suggestions.

References

1. Antoulas, A.C.: Approximation of large-scale dynamical systems. SIAM, Philadelphia (2005)
2. Beckermann, B.: Image numérique, GMRES et polynômes de Faber. C. R Acad. Sci. Paris Ser. I **340**, 855–860 (2005)
3. Beckermann, B., Reichel, L.: Error estimation and evaluation of matrix functions via the Faber transform. SIAM J. Numer. Anal. **47**(5), 3849–3883 (2009)
4. Benner, P., Mehrmann, V., Sorensen, D. (eds.): Dimension reduction of large-scale systems. Lecture notes in computational science and engineering. Springer-Verlag, Berlin (2005)
5. Bultheel, A., González-Vera, P., Hendriksen, E., Njåstad, O.: Orthogonal rational functions. Cambridge University Press, Cambridge (1999)
6. Crouzeix, M.: Numerical range and numerical calculus in Hilbert space. J. Funct. Anal. **244**, 668–690 (2007)
7. Van Deun, J., Deckers, K., Bultheel, A., Weideman, J.A.C.: Algorithm 882: Near best fixed pole rational interpolation with applications in spectral methods. ACM Trans. Math. Softw. **35**(2), 14:1–14:21 (2008)
8. Druskin, V., Knizhnerman, L.: Extended Krylov subspaces: approximation of the matrix square root and related functions. SIAM J. Matrix Anal. Appl. **19**(3), 755–771 (1998)
9. Druskin, V., Knizhnerman, L., Zaslavsky, M.: Solution of large scale evolutionary problems using rational Krylov subspaces with optimized shifts. SIAM J. Sci. Comput. **31**(5), 3760–3780 (2009)
10. Dzhrbashyan, M.M.: On decomposition of analytic functions in a series in rational functions with a given set of poles. Izv. AN Arm. SSR, Ser. Fiz.-Matem. N. **10**(1), 21–29 (1957)
11. Ellacott, S.W.: On the Faber transformation and efficient numerical rational approximation. SIAM J. Numer. Anal. **20**(5), 989–1000 (1983)
12. Gaier, D.: The Faber operator and its boundedness. J. Approx. Theory **101**(2), 265–277 (1999)
13. Godunov, S.K.: Lectures on modern aspects of linear algebra. Nauchnaya Kniga (IDMI), Novosibirsk, (In Russian) (2002)
14. Goluzin, G.M.: Geometric theory of functions of a complex variable. American Mathematical Society, Providence, RI (1969)
15. Hochbruck, M., Lubich, C.: On Krylov subspace approximations to the matrix exponential operator. SIAM J. Numer. Anal. **34**(5), 1911–1925 (1997)
16. Hochbruck, M., Starke, G.: Preconditioned Krylov subspace methods for Lyapunov matrix equations. SIAM Matrix Anal. and Appl. **16**(1), 156–171 (1995)
17. Jagels, C., Reichel, L.: The extended Krylov subspace method and orthogonal Laurent polynomials. Lin. Alg. Appl. **431**, 441–458 (2009)
18. Jaimoukha, I.M., Kasenally, E.M.: Krylov subspace methods for solving large Lyapunov equations. SIAM J. Numer. Anal. **31**(1), 227–251 (1994)
19. Jbilou, K., Riquet, A.J.: Projection methods for large Lyapunov matrix equations. Lin. Algebra Appl. **415**(2–3), 344–358 (2006)
20. Knizhnerman, L., Simoncini, V.: A new investigation of the extended Krylov subspace method for matrix function evaluations. Num. Lin. Algebra Appl. **17**(4), 615–638 (2010)
21. Kocharyan, G.S.: On approximation by rational functions in the complex domain. Izv. AN Arm. SSR, Ser. Fiz.-Matem. N. **11**(4), 53–77 (1958)
22. Kressner, D., Tobler., C.: Krylov subspace methods for linear systems with tensor product structure. Technical Report SAM-2009-16, ETH Zurich (2009)
23. Kressner, D., Tobler, C.: Krylov subspace methods for linear systems with tensor product structure. SIAM J. Matrix Anal. Appl. **31**(4), 1688–1714 (2010)
24. Lancaster, P.: Explicit solutions of linear matrix equations. SIAM Review **12**(4), 544–566 (1970)
25. Lu, A., Wachspress, E.: Solution of Lyapunov equations by Alternating Direction Implicit iteration. Comput. Math. Appl. **21**(9), 43–58 (1991)
26. Malmquist, F.: Sur la détermination d'une class de fonctions analytiques par leur valeurs dans un ensemble donné de points. In Comptes Rendus du Sixième Congrès (1925) des mathématiciens scandinaves, pp. 253–259. Kopenhagen (1926)
27. Penzl, T.: A cyclic low-rank Smith method for large sparse Lyapunov equations. SIAM J. Sci. Comput. **21**(4), 1401–1418 (2000)

28. Saad Y.: Numerical solution of large Lyapunov equations. In: Kaashoek, M. A., van Schuppen, J. H., and Ran, A. C., (eds,) Signal Processing, Scattering, Operator Theory, and Numerical Methods. Proceedings of the international symposium MTNS-89, vol III, pp. 503–511, Boston, Birkhauser (1990)
29. Schilders, W.H.A., van der Vorst, H.A., Rommes, J.: Model order reduction: theory, research aspects and applications. Springer-Verlag, Berlin (2008)
30. Simoncini, V., Druskin, V.: Convergence analysis of projection methods for the numerical solution of large Lyapunov equations. SIAM J. Numer. Anal. **47**(2), 828–843 (2009)
31. Simoncini, V.: A new iterative method for solving large-scale Lyapunov matrix equations. SIAM J. Sci. Comput. **29**(3), 1268–1288 (2007)
32. Suetin P.K.: Series of Faber Polynomials (Analytical Methods and Special Functions). Gordon and Breach Science Publishers, Amsterdam, Translated by E.V. Pankratiev (1998)
33. Suetin, S.P.: On the Montessus de Ballore theorem for nonlinear Padé approximations of orthogonal expansions and Faber series. Dokl. AN SSSR **253**(6), 1322–1325 (1980)
34. Takenaka, S.: On the orthogonal functions and a new formula for interpolation. Jpn. J. Math. **2**, 129–145 (1925)
35. Walsh, J.L.: Interpolation and Approximation by Rational Functions in the Complex Plane. Amer. Math. Soc., Providence, RI (1965)