Noname manuscript No.
(will be inserted by the editor)

Superlinear convergence of the rational Arnoldi method for the approximation of matrix functions

Bernhard Beckermann · Stefan Güttel

Received: date / Accepted: date

Abstract A superlinear convergence bound for rational Arnoldi approximations to functions of matrices is derived. This bound generalizes the well-known superlinear convergence bound for the CG method to more general functions with finite singularities and to rational Krylov spaces. A constrained equilibrium problem from potential theory is used to characterize a max-min quotient of a nodal rational function underlying the rational Arnoldi approximation, where an additional external field is required for taking into account the poles of the rational Krylov space. The resulting convergence bound is illustrated at several numerical examples, in particular, the convergence of the extended Krylov method for the matrix square root.

Keywords matrix functions · rational Arnoldi algorithm · superlinear convergence · logarithmic potential theory

1 Introduction

An important problem arising in science and engineering is the computation of matrix functions $f(A)\boldsymbol{b}$, where $A \in \mathbb{C}^{N\times N}$ is a Hermitian matrix, $\boldsymbol{b} \in \mathbb{C}^N$ is a vector of unit length, and f is a function such that $f(A)$ is defined. In

The work of the second author was partially supported by the Swiss National Science Foundation.

B. Beckermann
Université des Sciences et Technologies de Lille, Laboratoire Painlevé UMR 8524,
UFR Mathématiques, F-59655 Villeneuve d'Ascq CEDEX, France
Tel.: +33-320434562
Fax: +33-320434302
E-mail: bbecker@math.univ-lille1.fr

S. Güttel
Université de Genève, Section de mathématiques, CH-1211 Genève, Switzerland
E-mail: stefan@guettel.com

most applications the matrix A is large and sparse or structured, and first computing the generally dense unstructured matrix $f(A)$ and then forming the product with $\boldsymbol{b}$ is infeasible. The polynomial Arnoldi method (see, e.g., [12,17,20,21,38]) circumvents this problem by utilizing matrix-vector products to iteratively build an orthonormal basis $V_n = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n] \in \mathbb{C}^{N\times n}$ for a polynomial Krylov space of order n,

$$\mathcal{K}_n(A, \boldsymbol{b}) = \operatorname{span}\{\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^{n-1}\boldsymbol{b}\},$$

and to compute the associated *Arnoldi approximation* as

$$\boldsymbol{f}_n := V_n f(V_n^* A V_n) V_n^* \boldsymbol{b}. \tag{1.1}$$

The feasibility of this approach rests on the observation that $\boldsymbol{f}_n$ is often a good approximation to $f(A)\boldsymbol{b}$ for n being much smaller than N, and algorithms for dense matrices can be used to evaluate $f(V_n^* A V_n)$, for example by diagonalization of the small matrix $V_n^* A V_n$. We refer to [26] for a detailed account on algorithms for computing functions of dense matrices.

The Arnoldi approximations $\boldsymbol{f}_n$ of (1.1) often exhibit superlinear convergence towards $f(A)\boldsymbol{b}$, even if $f(z)$ has finite singularities, although the standard convergence analysis by polynomial approximation theory only predicts linear convergence in this case. A well-known example is obtained for the function $f(z) = z^{-1}$ and a positive definite matrix A, in which case the Arnoldi approximations $\boldsymbol{f}_n$ equal the iterates of the conjugate gradient (CG) method with zero initial guess (cf. [39, Section 6.7]). By standard estimates involving Chebyshev polynomials, it can be shown that these iterates converge at least at a linear rate determined by the condition number $\kappa := \lambda_{\max}/\lambda_{\min}$, namely

$$\|A^{-1}\boldsymbol{b} - \boldsymbol{f}_n\| \le C \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n \lesssim C \cdot \exp\left(-\frac{2n}{\sqrt{\kappa}}\right), \tag{1.2}$$

with some constant $C > 0$ independent of n. For small orders n this bound can be expected to be almost sharp because a low-degree residual polynomial that is small at all eigenvalues needs to be uniformly small on the spectral interval $S(0) := [\lambda_{\min}, \lambda_{\max}]$. However, for larger orders n the residual polynomials may have some of their zeros close to some of the eigenvalues of A and need to be uniformly small only on a remaining set $S(t) \subseteq S(0)$, $t = n/N$, which is shrinking as $0 \le t \le 1$ increases. This is the reason for the superlinear convergence behavior of the CG method, as was explained by Beckermann & Kuijlaars [6–8]. In their proofs these authors characterize the shrinking sets $S(t)$ as the support of a constrained equilibrium measure from logarithmic potential theory and show that the Ritz values $\Lambda(V_n^* A V_n)$ in the set $S(0)\setminus S(t)$ converge to eigenvalues with at least a geometric rate (see also [1, 3]). This allows to conclude that the CG method actually converges superlinearly like

$$\|A^{-1}\boldsymbol{b} - \boldsymbol{f}_n\| \lessapprox \exp\left(-N\int_0^t g_{S(\tau)}(0,\infty)\,\mathrm{d}\tau\right), \quad t = \frac{n}{N}, \tag{1.3}$$

where $g_S(x,y)$ denotes the Green function of (the unbounded connected component of) $\overline{\mathbb{C}} \setminus S$ with pole at y. For excellent expositions of the fruitful interaction between logarithmic potential theory and Krylov subspace methods, the interested reader may be referred to [16, 31].

The aim of the present paper is to explain such superlinear convergence effects more generally, with a generalization that is two-fold compared to the existing theory for the CG method. First, we allow that f be given as a Cauchy–Stieltjes (or Markov) representation

$$f(z) = \int_\Gamma \frac{\mathrm{d}\gamma(x)}{x - z}, \tag{1.4}$$

where γ is a complex measure supported on a closed set $\Gamma \subset \overline{\mathbb{C}} \setminus [\lambda_{\min}, \lambda_{\max}]$. As a consequence, the associated Newton potential given by

$$\widehat{f}(z) := \int_\Gamma \frac{\mathrm{d}|\gamma|(x)}{|x - z|} \tag{1.5}$$

is finite on the spectral interval $[\lambda_{\min}, \lambda_{\max}]$ of the matrix A. As a second generalization, we approximate $f(A)\boldsymbol{b}$ by the *rational Arnoldi method*, in which case the approximations (1.1) are computed with an orthonormal basis V_n of a *rational Krylov space* [36, 37]

$$\mathcal{Q}_n(A, \boldsymbol{b}) := q_{n-1}(A)^{-1}\mathcal{K}_n(A, \boldsymbol{b}) \quad \text{with} \quad q_{n-1}(z) := \prod_{\substack{j=1 \\ \xi_j \neq \infty}}^{n-1} (z - \xi_j), \tag{1.6}$$

where all the "poles" $\xi_j \in \overline{\mathbb{C}}$ are distinct from the eigenvalues $\Lambda(A)$. Note that our notation is such that with every rational Krylov space $\mathcal{Q}_n(A, \boldsymbol{b})$ there is implicitly associated a "denominator polynomial" q_{n-1}, and we obtain a polynomial Krylov space when $q_{n-1} \equiv 1$, i.e., when all poles ξ_j are infinite. Our analysis heavily builds on results from [5], in particular Theorem 3.1, which characterizes the convergence of rational Ritz values by the solution of a constrained equilibrium problem, where an additional external field is required for taking into account the poles of the rational Krylov space. We will derive a Buyarov–Rakhmanov-type asymptotic error bound for rational Arnoldi approximations. In case of cyclically repeated poles $\xi_{j+p} = \xi_j$ of periodicity p, it takes the form

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_n\| \lessapprox \exp\left(-N \min_{x \in \Gamma} \int_0^t \frac{\sum_{j=1}^p g_{S(\tau)}(x, \xi_j)}{p} \, \mathrm{d}\tau\right), \quad t = \frac{n}{N}. \tag{1.7}$$

Our analysis also allows for more general non-periodic pole sequences described by an increasing family of measures ν_t (cf. Theorem 4.1), but all of our examples have cyclically repeated poles. In particular, for the CG method, one recovers (1.3) from (1.7) by setting all $\xi_j = \infty$ and $\Gamma = \{0\}$.

Although our results are of an asymptotic nature, there is numerical evidence that the superlinear convergence phenomena analyzed here also occur

for finite N. To demonstrate this, we consider in Figure 1.1 the convergence of the so-called extended Krylov subspace method [18,28,29], which is equivalent to the rational Arnoldi method with cyclic pole sequence $\xi_{2j} = \infty$, $\xi_{2j+1} = 0$. We approximate $f(A)\boldsymbol{b}$ for the function

$$f(z) = z^{-1/2} = \int_\Gamma \frac{1}{x-z}\,\frac{-\,\mathrm{d}x}{\pi\sqrt{-x}}, \quad \Gamma = [-\infty, 0],$$

the matrix $A = \mathrm{tridiag}(-1, 2, -1)$ (which can be interpreted as a finite-difference approximation of the 1D-Laplacian), and a random vector $\boldsymbol{b}$ of size $N = 100$. In Figure 1.1 (left) we show the rational Ritz values of all orders $n = 1, \ldots, 100$, with a color indicating the distance to a closest eigenvalue of A. We will show in Section 5.2 that the eigenvalues of A outside the interval $S(t) = [4t^2(2-t)^{-2}, 4]$, $t = n/N$, are well approximated by Ritz values of order n. We visually confirm this by showing the left endpoint of $S(t)$ as the solid black curve. Our asymptotic error formula (1.7) is shown as the dashed red curve in Figure 1.1 (right). The dotted red curve is a closed expression for an upper bound of our integral formula, which we will also derive in Section 5.2. The dash-dotted blue curve is the linear convergence rate given by Knizhnerman & Simoncini [28, Theorem 3.4],

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_n\| \le C\left(\frac{\sqrt[4]{\kappa}-1}{\sqrt[4]{\kappa}+1}\right)^n \lesssim C\cdot\exp\left(-\frac{2n}{\sqrt[4]{\kappa}}\right), \tag{1.8}$$

and it coincides with the slope of our curves in the first few iterations, because none or only a few of the left-most Ritz values have converged. At later iterations, however, we clearly observe superlinear convergence of the extended Krylov subspace iteration (shown as the solid black curve), and this behavior is captured by our integral formula.

We will present some other examples of superlinear convergence in Section 5. Another example can be found in the thesis [23, Section 8.3], where such effects have been described and analyzed for the transfer function $f(z) = (z - i\omega)^{-1}$ approximated from a rational Krylov space with a single repeated pole different from $i\omega$. The remainder of this paper is organized as follows: In Section 2 we will recall some relevant facts about rational Arnoldi approximations. In particular, we will bound the error of these approximations in terms of a max-min quotient of a nodal rational function $s_n(z)$ in Theorem 2.1. Since there exists an order $n \le N$ for which the rational Arnoldi approximation $\boldsymbol{f}_n$ is exact, the notion of *convergence* of these approximations is meaningless for a single problem $f(A)\boldsymbol{b}$. In Section 3 we therefore translate this problem into a sequence of problems $f(A_N)\boldsymbol{b}_N$ having increasing dimension N and a well-defined limit. After this asymptotic reformulation, we are able to apply results from [5] for characterizing these potentials and the set $S(t)$ in terms of a minimal energy problem with an external field. Our main result is given in Section 4, where we present an integral representation for the difference of potentials, which completes the derivation of an asymptotic error bound for rational Arnoldi approximations. To make this paper also interesting for

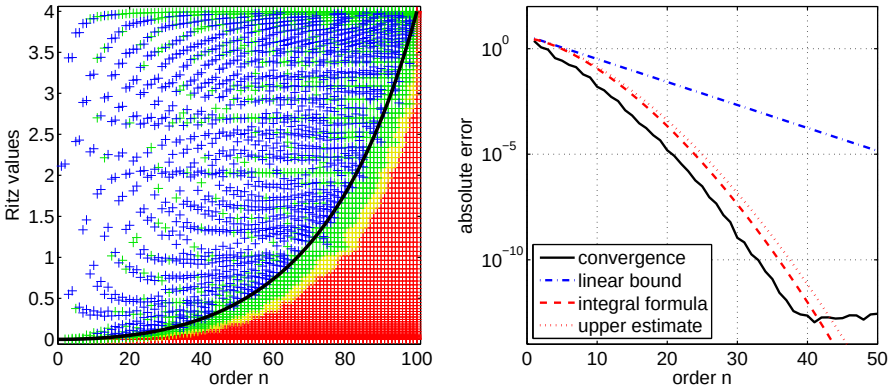

Fig. 1.1 On the left we plot the Ritz values of order $n = 1, \ldots, N$, associated with the matrix $A = \text{tridiag}(-1, 2, -1)$ and a random vector $\boldsymbol{b}$ of size $N = 100$. The colors indicate the distance of each Ritz value to a closest eigenvalue of A, this distance being decreasing from blue, yellow, green, to red (cf. Table 5.1 on page 16). The solid black curve is our prediction for the region of converged Ritz values. On the right we show the convergence of the extended Krylov subspace method towards $f(A)\boldsymbol{b}$ with $f(z) = z^{-1/2}$. Note how our asymptotic convergence bound mimics the superlinear convergence behavior of this method.

readers which are not very familiar with potential theory and to emphasize our results, we have decided to present the proof of our integral formula in a separate Section 6.

If not otherwise stated, $\langle \cdot, \cdot \rangle$ refers to the standard scalar product in $\mathbb{C}^N$ and $\| \cdot \|$ is the induced norm. We remark that the Arnoldi algorithm for Hermitian matrices is often implemented as a one-sided Lanczos algorithm, making use of three-term recurrences for orthogonal polynomials [32] or orthogonal rational functions [10, 14]. With this implementation, the numerical orthogonality of the Krylov basis vectors cannot be guaranteed. Since we assume exact arithmetic throughout this paper, we found it more appropriate to consider Arnoldi approximations instead of "Lanczos approximations."

2 Rational Arnoldi approximations and interpolation

We recall some relevant facts about rational Arnoldi approximations. For algorithmic details on the iterative construction of an orthonormal Krylov basis by Ruhe's rational Krylov sequence algorithm we refer to [36, 37]. Some implementation-related issues for computing rational Arnoldi approximations are discussed in [9, 23]. By definition of the rational Arnoldi approximation (cf. (1.1) and (1.6)) it is clear that we have a rational function representation

$$\boldsymbol{f}_n = r_n(A)\boldsymbol{b} = \frac{p_{n-1}}{q_{n-1}}(A)\boldsymbol{b},$$

where $p_{n-1} \in \mathcal{P}_{n-1}$ is a polynomial of degree at most $n-1$. Moreover, it is known that r_n actually is a rational interpolant of f with fixed denominator

q_{n-1} and interpolation nodes given by the *rational Ritz values* $\Lambda(V_n^* A V_n) = \{\theta_1, \ldots, \theta_n\}$ (see, e.g., [9], [23, Theorem 4.8]), which can be expressed as

$$f(z) - r_n(z) = s_n(z)h(z), \quad \text{where} \quad s_n(z) := \frac{(z-\theta_1)\cdots(z-\theta_n)}{q_{n-1}(z)} \tag{2.1}$$

with some function h analytic in $\overline{\mathbb{C}} \setminus \Gamma$. We refer to s_n as a nodal rational function, as its zeros are nodes for rational interpolation with prescribed denominator q_{n-1}. The following theorem is now easily derived.

Theorem 2.1 *Let $f(z)$ be given by* (1.4) *analytic in $\overline{\mathbb{C}} \setminus \Gamma$ containing the spectral interval $[\lambda_{\min}, \lambda_{\max}]$ of A, and let $\widehat{f}(z)$ be the associated Newton potential* (1.5)*. Then the rational Arnoldi approximation $\boldsymbol{f}_n$ satisfies*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_n\| \le \|\widehat{f}(A)\boldsymbol{b}\| \, \frac{\max_{z\in\Lambda(A)} |s_n(z)|}{\min_{x\in\Gamma} |s_n(x)|}.$$

Proof As a consequence of the interpolation property (2.1), the interpolation error can be represented as (see, e.g., [41, Theorem VIII.2], [9, p. 24])

$$f(z) - r_n(z) = s_n(z) \int_\Gamma \frac{\mathrm{d}\gamma(x)}{s_n(x)(x-z)}.$$

Since the Euclidean norm is invariant under multiplication with a unitary factor and A is supposed to be Hermitian, we may suppose without loss of generality that $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ is diagonal. Then

$$\begin{aligned}
\|f(A)\boldsymbol{b} - \boldsymbol{f}_n\| &= \|f(A)\boldsymbol{b} - r_n(A)\boldsymbol{b}\| \\
&\le \|s_n(A)\| \left\| \int_\Gamma \frac{1}{s_n(x)} (xI - A)^{-1} \boldsymbol{b} \, \mathrm{d}\gamma(x) \right\| \\
&\le \frac{\|s_n(A)\|}{\min_{x\in\Gamma} |s_n(x)|} \left\| \left(\int_\Gamma \frac{|b_j|}{|x-\lambda_j|} \, \mathrm{d}|\gamma|(x) \right)_j \right\|.
\end{aligned}$$

The minimum in the last term is nonzero because the zeros of the nodal rational function s_n (the rational Ritz values) are contained in $[\lambda_{\min}, \lambda_{\max}]$ and therefore bounded away from the closed set Γ. Hence the assertion follows from

$$\|s_n(A)\| = \max_{z\in\Lambda(A)} |s_n(z)|.$$

The upper bound of Theorem 2.1 requires some further knowledge on the nth rational Ritz values of A, which will be available in the setting of Section 3. For completeness we suggest an alternate error estimate inspired by [9, Proposition 3.1 and Theorem 3.2] roughly saying that we may replace s_n in Theorem 2.1 by any other nodal-type function.

Theorem 2.2 *With the setting of Theorem 2.1, we have for any* $\widetilde{\theta}_1, \ldots, \widetilde{\theta}_n \in \mathbb{C}$ *that*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_n\| \leq 2 \max_{z \in [\lambda_{\min}, \lambda_{\max}]} |\widehat{f}(z)| \, \frac{\|\widetilde{s}_n(A)\boldsymbol{b}\|}{\min_{x \in \Gamma} |\widetilde{s}_n(x)|},$$

where

$$\widetilde{s}_n(z) = \frac{(z - \widetilde{\theta}_1) \cdots (z - \widetilde{\theta}_n)}{q_{n-1}(z)}.$$

Proof It is sufficient to consider the case $\widetilde{\theta}_j \notin \Gamma$ since otherwise the upper bound equals $+\infty$. Denote by $\widetilde{r}_n$ the rational interpolant of f with fixed denominator q_{n-1} and numerator of degree $< n$ interpolating f at $\widetilde{\theta}_1, \ldots, \widetilde{\theta}_n$. Then the exactness property [23, Lemma 4.6] tells us that $\widetilde{r}_n(A)\boldsymbol{b} = V_n \widetilde{r}_n(V_n^* A V_n) V_n^* \boldsymbol{b}$, and thus, using similar arguments as in the proof of Theorem 2.1,

$$\begin{aligned} \|f(A)\boldsymbol{b} - \boldsymbol{f}_n\| &\leq \|(f - \widetilde{r}_n)(A)\boldsymbol{b}\| + \|(f - \widetilde{r}_n)(V_n^* A V_n) V_n^* \boldsymbol{b}\| \\ &\leq \max_{z \in \Lambda(A)} |\widehat{f}(z)| \, \frac{\|\widetilde{s}_n(A)\boldsymbol{b}\|}{\min_{x \in \Gamma} |\widetilde{s}_n(x)|} + \max_{z \in \Lambda(V_n^* A V_n)} |\widehat{f}(z)| \, \frac{\|\widetilde{s}_n(V_n^* A V_n) V_n^* \boldsymbol{b}\|}{\min_{x \in \Gamma} |\widetilde{s}_n(x)|}. \end{aligned}$$

We also obtain from [23, Lemma 4.6] that

$$\|\widetilde{s}_n(V_n^* A V_n) V_n^* \boldsymbol{b}\| = \|V_n V_n^* \widetilde{s}_n(A)\boldsymbol{b}\| \leq \|\widetilde{s}_n(A)\boldsymbol{b}\|,$$

and the observation $\Lambda(A) \cup \Lambda(V_n^* A V_n) \subset [\lambda_{\min}, \lambda_{\max}]$ allows to conclude.

For the special case $\Gamma \subset [-\infty, \lambda_{\min})$, by slightly changing the above arguments and arguing in terms of energy norms, it is possible to drop in Theorem 2.2 the factor 2.

By choosing a suitable function $\widetilde{s}_n$ in Theorem 2.2 we may recover several known error bounds. For instance, the linear error bound (1.2) for the iterates of the CG method (which is a polynomial Krylov method, i.e., $q_{n-1} \equiv 1$) is obtained by representing $f(z) = z^{-1}$ as a Markov function (1.4) with generating Dirac measure $\gamma = \delta_0$ supported at 0 and thus $\Gamma = \{0\}$. Finally, the term on the right-hand side can be bounded from above by replacing $\Lambda(A)$ by the spectral interval $[\lambda_{\min}, \lambda_{\max}]$ and taking a shifted Chebyshev polynomial for $\widetilde{s}_n$.

Remark 2.1 Besides $f(z) = z^{-1}$ and $f(z) = z^{-1/2}$ discussed in Section 1, examples of other Markov functions are (see, e.g., [9, 18])

$$f(z) = \frac{\log(1+z)}{z} = \int_{\Gamma} \frac{(1/x)\,\mathrm{d}x}{x - z}, \quad \text{where} \quad \Gamma = [-\infty, -1],$$

$$f(z) = \frac{\exp(\theta\sqrt{z}) - 1}{z} = \int_{\Gamma} \frac{1}{x - z} \frac{\sin(\theta\sqrt{-x})\,\mathrm{d}x}{\pi x}, \quad \text{where} \quad \Gamma = [-\infty, 0].$$

Notice that for all these functions the measure γ is positive, and $\Gamma \subset [-\infty, \lambda_{\min})$, implying that $\widehat{f}(z) = -f(z)$ for $z \in \Lambda(A)$. As a consequence, $\|\widehat{f}(A)\boldsymbol{b}\| = \|f(A)\boldsymbol{b}\|$ in Theorem 2.1, that is, we give an upper bound for the relative error.

This last observation is no longer true for general functions f being analytic in some neighborhood of the spectral interval. In this case, by choosing a suitable contour Γ encircling the spectral interval, we still obtain from the Cauchy integral formula a representation as in (1.4) (for z in the interior of Γ, as required), but now the function $\widehat{f}$ of (1.5) might be of much larger modulus than f, depending of course on the choice of Γ.

3 Asymptotic setting and potential theoretic tools

To study the N-th root of the error bound of Theorem 2.1, we investigate the behavior of the nodal rational function $s_n(z)$ of (2.1), or more precisely the quotient

$$Z_{n,N}(\Lambda(A_N),\Gamma) := \left(\frac{\max_{z\in\Lambda(A_N)}|s_n(z)|}{\min_{x\in\Gamma}|s_n(x)|}\right)^{1/N},$$

where we have added an index N for the matrix $A = A_N$ to indicate its size. For simplicity we will assume in Sections 3 and 4 that all poles are finite, and refer the reader to Remark 3.2(b) for the general case. We recall the definition of the logarithmic potential of a measure μ

$$U^{\mu}(z) = \int \log\frac{1}{|x-z|}\,\mathrm{d}\mu(x),$$

and consider the function

$$\chi_N \;:\; \Sigma \mapsto \frac{1}{N}\sum_{\substack{x\in\Sigma\\ x\neq\infty}}\delta_x,$$

which associates a counting measure with a multiset $\Sigma \subset \overline{\mathbb{C}}$ (δ_x denoting the Dirac unit measure at the point x). A simple calculation verifies that with the measures $\mu_n := \chi_N(\{\theta_1,\ldots,\theta_n\})$ and $\nu_n := \chi_N(\{\xi_1,\ldots,\xi_{n-1}\})$ we have

$$U^{\mu_n-\nu_n}(z) = \frac{1}{N}\sum_{j=1}^{n}\log\frac{1}{|z-\theta_j|} - \frac{1}{N}\sum_{\substack{j=1\\ \xi_j\neq\infty}}^{n-1}\log\frac{1}{|z-\xi_j|} = -\log|s_n(z)|^{1/N},$$

which shows that

$$\log\big(Z_{n,N}(\Lambda(A_N),\Gamma)\big) = \max_{x\in\Gamma}U^{\mu_n-\nu_n}(x) - \min_{z\in\Lambda(A_N)}U^{\mu_n-\nu_n}(z). \tag{3.1}$$

However, this expression is still of limited use, mainly because discrete measures do not have a finite logarithmic energy

$$I(\mu) = I(\mu,\mu), \quad I(\mu_1,\mu_2) = \iint\log\frac{1}{|x-z|}\,\mathrm{d}\mu_1(x)\,\mathrm{d}\mu_2(z)$$

and our asymptotic description of the nodal rational function $s_n(z)$ will be in terms of a (continuous) measure with minimal logarithmic energy. Following

Kuijlaars and his successors [1,5–8,25,30], we therefore consider a sequence of matrices $A_N \in \mathbb{C}^{N\times N}$ whose eigenvalues $\lambda_{1,N} < \cdots < \lambda_{N,N}$ have an asymptotic distribution given by a probability measure σ,

$$\chi_N(\{\lambda_{1,N}, \ldots, \lambda_{N,N}\}) \to \sigma$$

in the weak-star sense. We recall that, for a sequence of measures σ_n, the relation $\sigma_n \to \sigma$ means that $\int h \,\mathrm{d}\sigma_n \to \int h \,\mathrm{d}\sigma$ for all continuous functions h. Accordingly, we consider a sequence of vectors $\boldsymbol{b}_N \in \mathbb{C}^N$ of unit length and define the eigencoordinates

$$w_N(\lambda_{j,N}) := |\langle \boldsymbol{u}_{j,N}, \boldsymbol{b}_N\rangle|, \quad j = 1, \ldots, N, \tag{3.2}$$

where $\boldsymbol{u}_{j,N}$ is the normalized eigenvector of A_N associated with $\lambda_{j,N}$. Moreover, we consider a sequence of rational Krylov spaces $\mathcal{Q}_{n,N}(A_N, \boldsymbol{b}_N)$ of orders $n = n(N)$ such that $n/N \to t$ for some $t \in (0,1)$ as $N \to \infty$ and such that the poles $\xi_{1,N}, \ldots, \xi_{n-1,N}$ are asymptotically distributed according to the measure ν_t,

$$\chi_N(\{\xi_{1,N}, \ldots, \xi_{n-1,N}\}) \to \nu_t,$$

and thus $\|\nu_t\| = t$.

A few other technical assumptions are required: We impose in particular a separation between the eigenvalues of A_N and the poles of the rational Krylov spaces. This assumption is natural as we are approximating a function f that is analytic on the spectral interval and the poles of the approximant should stay away from this set.

Assumption 1: There exist disjoint compact sets $\Lambda, \Xi \subset \mathbb{C}$ such that $\lambda_{j,N} \in \Lambda$ $(j = 1, \ldots, N)$ and $\xi_{j,N} \in \Xi$ $(j = 1, \ldots, n(N) - 1)$ for all N.

We need to exclude the possibility that eigenvalues of A_N cluster exponentially, because this situation could not be resolved by N-th root asymptotics. The following assumption prevents exponential clustering, but still allows for equidistant eigenvalues, Chebyshev eigenvalues (the eigenvalues of the 1D-Laplacian), and more general sets of points [15]. It also guarantees that U^σ is continuous [5, Lemma A.4].

Assumption 2: For any sequence $\lambda_{k(N),N} \to \lambda$ for $N \to \infty$,

$$\limsup_{\delta\to 0+} \limsup_{N\to\infty} \frac{1}{N} \sum_{0<|\lambda_{j,N}-\lambda_{k(N),N}|\le\delta} \log \frac{1}{|\lambda_{j,N} - \lambda_{k(N),N}|} = 0.$$

The last technical assumption ensures that the vectors $\boldsymbol{b}_N$ have sufficiently large coordinates in all eigenvectors of A_N.

Assumption 3: The eigencomponents $w_N(\lambda_{j,N}) \in [0,1]$ defined in (3.2) satisfy

$$\liminf_{N\to\infty} \min_k w_N(\lambda_{k,N})^{1/N} = 1.$$

Let $\mathcal{M}_t^\sigma := \{\mu : \mu \text{ Borel measure with } \|\mu\| = t, \mu \le \sigma\}$ denote a set of σ-constrained measures. The following lemma summarizes results from [5, Lemma A.1] and [5, Theorem 3.1].

Lemma 3.1 *Under the Assumptions 1–3, the extremal problem*

$$\inf\{I(\mu) - 2I(\mu, \nu_t) : \mu \in \mathcal{M}_t^\sigma\} \tag{3.3}$$

has a unique minimizer μ_t*. This minimizer satisfies* $\operatorname{supp}(\mu_t) = \operatorname{supp}(\sigma)$ *and there exists a constant* $F_t > 0$ *such that*

$$G(t,z) := U^{\mu_t - \nu_t}(x) - F_t \begin{cases} = 0 \text{ for } x \in S(t) := \operatorname{supp}(\sigma - \mu_t), \\ \le 0 \text{ for } x \in \mathbb{C}. \end{cases} \tag{3.4}$$

Furthermore, the rational Ritz values $\Theta_N := \{\theta_{1,N}, \ldots, \theta_{n,N}\}$ *of order* $n = n(N)$ *with* $n/N \to t$ *have an asymptotic distribution* $\chi_N(\Theta_N) \to \mu_t$ *and converge geometrically to eigenvalues located outside the set* $S(t)$*.*

We remark that the constraint $\mu_t \le \sigma$ arises from the so-called *interlacing property* of Ritz values (cf. [34, Theorem 10.1.1]), which states that in every interval the number of Ritz values does not exceed the number of eigenvalues by more than one. For more general extremal problems following these lines we refer to [15, 35] and [2, Theorem 1.1].

The upper bound derived in the following theorem together with the integral expression for $G(t,z)$ of the next section concludes our derivation of the error bound (1.7) claimed in the introduction.

Theorem 3.1 *Under the Assumptions 1–3, denote by* $\boldsymbol{f}_{n,N}$ *the* n*th rational Arnoldi approximation of* $f(A_N)\boldsymbol{b}_N$*. Provided that* $\Gamma \cap \Lambda$ *is empty, there holds*

$$\limsup_{\substack{n,N\to\infty \\ n/N\to t}} \log\left(\|f(A_N)\boldsymbol{b}_N - \boldsymbol{f}_{n,N}\|\right)^{1/N} \le \max_{x\in\Gamma} G(t,x).$$

Proof According to Theorem 2.1, it is sufficient to show that

$$\limsup_{\substack{n,N\to\infty \\ n/N\to t}} \log\left(Z_{n,N}(\Lambda(A_N), \Gamma)\right) \le \max_{x\in\Gamma} G(t,x). \tag{3.5}$$

Let us analyze separately each term on the right-hand side of (3.1), where we write more explicitly

$$\begin{aligned} \mu_{n,N} &:= \chi_N(\{\theta_{1,N}, \ldots, \theta_{n,N}\}) \to \mu_t, \\ \nu_{n,N} &:= \chi_N(\{\xi_{1,N}, \ldots, \xi_{n-1,N}\}) \to \nu_t. \end{aligned}$$

From [2, Theorem 1.3] (see also [15, Theorem 3.3] and [5, Theorem 5.4]) we obtain

$$\lim_{\substack{n,N\to\infty \\ n/N\to t}} \log\left(\min_{z\in\Lambda(A_N)} U^{\mu_{n,N}-\nu_{n,N}}(z)\right) = F_t.$$

In order to discuss the other term, let $x_N \in \Gamma$ such that

$$\min_{x\in\Gamma} U^{\nu_{n,N}-\mu_{n,N}}(x) =: U^{\nu_{n,N}-\mu_{n,N}}(x_N).$$

By passing to subsequences if necessary, we may suppose that $x_N \to \overline{x} \in \overline{\mathbb{C}}$. Since $\operatorname{supp}(\mu_{n,N}) \subset \Lambda$ and $\Lambda \cap \Gamma$ is empty, we conclude that $U^{\mu_{n,N}} - U^{\mu_t} \to 0$ uniformly on Γ. By applying the principle of descent (see, e.g., [40, Theorem I.6.8]) for $\nu_{n,N}$ we hence arrive at

$$\liminf_{\substack{n,N\to\infty \\ n/N\to t}} U^{\nu_{n,N}-\mu_{n,N}}(x_N) \geq U^{\nu_t-\mu_t}(\overline{x}) \geq -\max_{x\in\Gamma} U^{\mu_t-\nu_t}(x),$$

as required for the assertion of (3.5).

Remark 3.1 With the assumptions and the notation of the preceding proof, if Γ is sufficiently dense at each of its elements (as is typically the case for our contours), we may apply [33, Theorem 5.4.3] saying that

$$\liminf_{\substack{n,N\to\infty \\ n/N\to t}} \min_{x\in\Gamma} U^{\nu_{n,N}-\mu_{n,N}}(x) = \min_{x\in\Gamma} U^{\nu_t-\mu_t}(x),$$

and thus (3.5) holds with equality. Moreover, the discrete Bernstein–Walsh inequality of [2, Theorem 1.2(c)] allows to show that we do not obtain any better rate of convergence by using Theorem 2.2 instead of Theorem 2.1.

We conclude this section by a discussion of the assumptions of Theorem 3.1.

Remark 3.2 **(a)** It is possible to obtain similar upper bounds without Assumption 3 and the technical separation condition of Assumption 2, by just assuming that U^σ is continuous. In this case, Lemma 3.1 fails to hold, and the asymptotic behavior of the $n(N)$-th Ritz values is no longer known. Here we use the bounds of Theorem 2.2 instead of Theorem 2.1, and replace the nodal rational functions s_n by $\widetilde{s}_n$ obtained by discretization of μ_t. Since such a discretization is technically involved, we refer the reader to [7, Proof of Theorem 2.1] and omit further details.
(b) In order to include unbounded poles or eigenvalues or even poles being equal to ∞, it suffices to assume in Assumption 1 only that the sets Λ and Ξ containing respectively the eigenvalues and the poles are closed and disjoint subsets of $\overline{\mathbb{C}}$. In this case, a rational transformation of the plane as in [5, Corollary 3.3] allows to go back to compacts: with some fixed $\rho \in \mathbb{R} \setminus (\Lambda \cup \Xi)$, we consider the new variable

$$\underline{z} = T(z) = 1/(z-\rho),$$

where we notice that the nth rational Krylov space for A_N, $\boldsymbol{b}_N$ and poles $\xi_{j,N}$ is the same as the nth rational Krylov space for $T(A_N)$, $\boldsymbol{b}_N$ and poles $T(\xi_{j,N})$, both being spanned by the columns of, say, $V_{n,N}$. However, in our analysis we have to be a bit careful since the Rayleigh–Ritz matrix $V_{n,N}^* T(A_N) V_{n,N}$ can be shown to be $T(V_{n,N}^* A_N V_{n,N})$ plus a perturbation of rank 1. In particular, the transformed Ritz values $T(\theta_{j,N})$ are no longer Ritz values of the transformed matrix $T(A_N)$. One could go back to Ritz values of $T(A_N)$ by applying Theorem 2.2 instead of Theorem 2.1, with $\widetilde{s}_n$ having as roots the

pre-images under T of the rational Ritz values of $T(A_N)$. This again leads to an asymptotic bound as in Theorem 3.1, but now in terms of an extremal problem in the transformed variables $T(\lambda_{j,N})$, $T(\xi_{j,N})$ and their asymptotic counterparts.

(c) As in some of our numerical examples, we may also allow for poles lying in $\Lambda \backslash \Lambda(A_N)$, compare with [5, Theorem 3.2]. We again omit the (quite technical) details.

(d) In most of our examples, the superlinear convergence behavior is particularly pronounced if Γ and Λ are not disjoint but have one point in common, say, the point $0 \notin S(t)$, which is no longer covered by Theorem 3.1. The situation is similar in [7, Proof of Theorem 2.1] for superlinear CG convergence where an additional separation condition between $\Gamma = \{0\}$ and the eigenvalues close to 0 is imposed. In our setting we have to revisit the proof of Theorem 3.1 and impose other conditions in order to insure that $U^{\mu_{n,N}}(x_N) \to U^{\mu_t}(\overline{x})$ in the case $x_N \to \overline{x} = 0$. For instance, if $\Gamma = \Gamma_1 \cup \Gamma_2$ with $\Gamma_1 \subset \{Re(z) \leq 0\}$, $\Lambda \subset [0, +\infty)$ and Γ_2 some compact set having empty intersection with Λ, we can argue as in [7, Proof of Theorem 2.1] and [5] by imposing the additional condition (similar to Assumption 2) that

$$\limsup_{\delta \to 0+} \limsup_{N \to \infty} \frac{1}{N} \sum_{|\lambda_{j,N}| \leq \delta} \log \frac{1}{|\lambda_{j,N}|} = 0,$$

which again is true for all our examples.

4 The integral formula

Beside Theorem 3.1, there is another important ingredient in our derivation of the error bound (1.7) claimed in the introduction. Namely, given the solution μ_t of the extremal problem (3.3) with external field $-U^{\nu_t}$ and a constraint σ, we require a representation of $G(t,z) = U^{\mu_t - \nu_t}(z) - F_t$ in terms of a negative integral mean of Green functions of a family of monotone sets $S(\tau)$ for $0 < \tau < t$.

A first result in this direction for extremal problems without constraints is given by the so-called Buyarov–Rakhmanov formula [11] with $S(t) = \mathrm{supp}(\mu_t)$ being increasing in t. For constrained problems without external fields (or $\nu_t = 0$), a similar formula was established in [7, Theorem 2.1] with $S(t) = \mathrm{supp}(\sigma - \mu_t)$ being decreasing in t.

For a general external field (possibly depending on t) and constraint, the existence of a Buyarov–Rakhmanov formula is an open problem, and only partial results have been given by Coussement & Van Assche in [13, Theorem A.10], but under technical assumptions which are not easy to verify. The aim of this section is to show that for the particular external field $-U^{\nu_t}$ these difficulties can be eliminated. We have the following result.

Theorem 4.1 *Let $G(t,z)$ and $S(t)$ be defined as in Lemma 3.1, with a probability measure σ having compact support* $\mathrm{supp}(\sigma) \subset \mathbb{R}$*, and with an increasing*

family of measures $(0,1) \ni t \mapsto \nu_t$ *being (weak-star) differentiable with respect to* t *for almost all* t*, with derivative* $\widetilde{\nu}_t$*, and* $\|\nu_t\| = t$*,* $\operatorname{supp}(\nu_t)$ *compact having empty intersection with* $\operatorname{supp}(\sigma)$*. Then for all* $t \in (0,1)$ *and for all* $z \in \mathbb{C}$ *we have*

$$G(t,z) = -\int_0^t \int g_{S(\tau)}(z,y)\,\mathrm{d}\widetilde{\nu}_\tau(y)\,\mathrm{d}\tau, \tag{4.1}$$

where the sets $S(t) \subseteq \operatorname{supp}(\sigma)$ *are decreasing in* t*.*

Remark 4.1 The extremal quantity $G(t,z)$ is also well defined if $\|\nu_t\| < t$ in Theorem 4.1, in which case the derivative $\widetilde{\nu}_t$ has mass ≤ 1. In this case, we have to replace $\widetilde{\nu}_\tau$ in (4.1) by $\widetilde{\nu}_\tau + (1 - \|\widetilde{\nu}_\tau\|)\delta_\infty$, the additional δ_∞ term occurring due to a rational variable transformation, compare with Remark 3.2(b). Such a situation occurs if a portion of the poles of the rational Krylov space are equal to ∞: by translating the asymptotic error bound of Remark 3.2(b) in terms of the new variable $\underline{z} = T(z)$ back to the original variable $z = T^{-1}(\underline{z})$, we obtain ν_t describing the asymptotic distribution of the *finite* poles.

Remark 4.2 In most applications one has a fixed asymptotic pole distribution $\nu_t = t\nu$ for some probability measure ν, here the assumptions of Theorem 4.1 hold with constant derivative $\widetilde{\nu}_t = \nu$, leading to some simplification in (4.1). In particular, in case of cyclically repeated poles $\xi_1, \ldots, \xi_p$ for all N, one obtains $\nu_t = t\nu$ for $\nu = \frac{1}{p}\sum_{j=1}^p \delta_{\xi_j}$, and hence

$$G(t,z) = -\int_0^t \frac{\sum_{j=1}^p g_{S(\tau)}(z,\xi_j)}{p}\,\mathrm{d}\tau,$$

in accordance with (1.7). By Remark 4.1, this formula remains true if one of the repeated poles is equal to ∞, for instance $p = 1$ and $\xi_1 = \infty$ for polynomial Krylov spaces (all poles are at ∞), or $p = 2$ and $(\xi_1, \xi_2) = (0, \infty)$ for the extended Krylov subspace method.

Remark 4.3 Let us have a closer look at the case $\nu_t = t\nu$. We will see in the proof of Theorem 4.1 that $S(t) = \operatorname{supp}(\sigma - \mu_t)$. In particular, it may happen for small t that $S(t) = S(0) = \operatorname{supp}(\sigma)$, or, roughly speaking, none of the eigenvalues of A_N is well approximated by Ritz values. In this case, (4.1) reduces to $G(t,z) = -t\int g_{S(0)}(z,y)\,\mathrm{d}\nu(y)$, a linear convergence rate in Theorem 3.1. Otherwise, we use Lemma 6.5 established below in order to show that, for $0 < t_1, t_2 < 1$,

$$G(t_2,z) \leq G(t_1,z) - (t_2 - t_1)\int g_{S(t_1)}(z,y)\,\mathrm{d}\nu(y),$$

from which it follows that $t \mapsto G(t,z)$ is concave for fixed z. We will see in Section 5, Figure 5.2, that $t \mapsto \max_{z\in\Gamma} G(t,z)$ is no longer necessarily convex, and thus the convergence rate in Theorem 3.1 is more complicated.

We will require families of pole measures ν_t not necessarily depending linearly on t in case we want to optimize poles uniformly for all iterations. The weak-star differentiability of $t \mapsto \nu_t$ at t means that the family of (in general signed) measures $\frac{\nu_t - \nu_\tau}{t-\tau}$ has a weak-star limit $\widetilde{\nu}_t$ for $\tau \to t$, or, in other words, the following limit exists

$$\forall h \in \mathcal{C}: \qquad \lim_{\tau \to t} \int h \, \mathrm{d} \frac{\nu_t - \nu_\tau}{t-\tau} = \int h \, \mathrm{d}\widetilde{\nu}_t. \tag{4.2}$$

In the setting of Theorem 4.1, putting $\nu_0 = 0$, one deduces that $\frac{\nu_t - \nu_\tau}{t-\tau}$ is a probability measure for all $0 \le \tau \le t < 1$, and thus also $\widetilde{\nu}_t$ is a probability measure.

The above Theorem 4.1 allows to deduce a result on an optimal pole distribution with respect to a given closed Γ of positive capacity in terms of condensers $(\Gamma, S(t))$, with capacity $\mathrm{cap}(\Gamma, S(t))$ [40, Section VIII]. The main message of Corollary 4.1 (and its proof) is that an increasing family of pole measures being optimal for all t should take into account that the sets $S(t)$ are shrinking (and hence the derivative $\widetilde{\nu}_t$ should not be constant). As in [4, 22] we will suppose that Γ has connected complement and empty interior, though we believe that the following statement remains also true without these two technical assumptions.

Corollary 4.1 *Under the assumptions of Theorem 4.1, let $\Gamma \subset \overline{\mathbb{C}}$ be closed, of positive capacity, with connected complement and empty interior. If* $\mathrm{cap}(\Gamma \cap \mathrm{supp}(\sigma)) = 0$ *then there exists a family of pole distributions $t \mapsto \underline{\nu}_t$ minimizing for all t the quantity*

$$\max_{z \in \Gamma} G(t, z).$$

Denoting by $t \mapsto \underline{\mu}_t$ the corresponding family of extremal measures, this family of extremal pole measures is characterized by the fact that, for almost all t, the weak-star derivatives $\widetilde{\underline{\mu}}_t$ and $\widetilde{\underline{\nu}}_t$ exist, and $\widetilde{\underline{\mu}}_t - \widetilde{\underline{\nu}}_t$ is the charge of a condenser with plate $\mathrm{supp}(\sigma - \underline{\mu}_t)$ *carrying a positive unit charge and the plate Γ carrying a negative unit charge. In addition, the maximum of the corresponding function $G(t, z)$ for $z \in \Gamma$ is given by*

$$-\int_0^t \frac{1}{\mathrm{cap}(\mathrm{supp}(\sigma - \underline{\mu}_t), \Gamma)} \, \mathrm{d}\tau.$$

5 Examples

Before proving Theorem 4.1 and Corollary 4.1 we would like to illustrate these results at some examples. In order to evaluate the superlinear error bound, we need to know a family of Green functions $g_{S(t)}(x, y)$. Hence we need to determine the family of sets $S(t) = \mathrm{supp}(\sigma - \mu_t)$ beforehand, which can be done analytically for the following model examples.

5.1 Kac–Murdock–Szegő example

The following Lemma is easily deduced from results in [7, 27] and [5, Lemma A.3].

Lemma 5.1 *Let $0 < q < 1$ be given. Then the eigenvalues of*

$$A_N = \begin{pmatrix} q^0 & q^1 & q^2 & \\ q^1 & q^0 & q^1 & \ddots \\ q^2 & q^1 & q^0 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix} \in \mathbb{R}^{N\times N}$$

are asymptotically (for $N \to \infty$) distributed like

$$\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x} = \frac{1}{\pi x\sqrt{(x-\alpha)(\beta-x)}}, \quad \operatorname{supp}(\sigma) = [\alpha,\beta] = \left[\frac{1-q}{1+q}, \frac{1+q}{1-q}\right].$$

For $\xi > \beta$ and $\nu_t = t\delta_\xi$ (single repeated pole) we have

$$S(t) = [\alpha, b(t)], \quad b(t) = \begin{cases} \beta, & t < \sqrt{\frac{\alpha(\alpha\xi-1)}{\xi-\alpha}}; \\ \alpha\xi/(\alpha + t^2\xi - t^2\alpha), & t \geq \sqrt{\frac{\alpha(\alpha\xi-1)}{\xi-\alpha}}. \end{cases}$$

Proof Under the assumption that $\operatorname{supp}(\nu_t) \subset (\beta, +\infty)$, it can be shown (see [7, 27] and [5, Lemma A.3]) that $S(t) = \operatorname{supp}(\sigma - \mu_t) = [\alpha, b(t)]$ if

$$\sqrt{\frac{\alpha}{b(t)}} = \int \sqrt{\frac{y-\alpha}{y-b(t)}}\,\mathrm{d}\nu_t(y) \tag{5.1}$$

has a solution $b(t) \in [\alpha, \beta]$ (which is unique in this case), and otherwise $b(t) = \beta$. The assertion follows by solving this equation with $\nu_t = t\delta_\xi$.

Example 5.1 We apply Lemma 5.1 with $q = 1/2$ and $\xi = 4$, yielding

$$S(t) = [1/3, b(t)], \quad b(t) = \begin{cases} 3, & t < 1/\sqrt{33} \approx 0.174; \\ 4/(11t^2+1), & t \geq 1/\sqrt{33}. \end{cases}$$

We expect that if A is of order, say, $N = 100$, convergence of Ritz values sets in at iteration $n = \lceil N/\sqrt{33}\rceil = 18$, which is confirmed numerically in Figure 5.1 (left). To indicate the distance of each Ritz value to a closest eigenvalue (relative to the width of the spectral interval) we have used the color code given in Table 5.1. The solid black curve indicates the right endpoint of the interval $S(t)$.

We now approximate $f(A)\boldsymbol{b}$ for the function $f(z) = \log(3-z)$, which can be represented as a polynomial modification of a Markov function whose generating measure has support $\Gamma = [3, +\infty]$ (the branch cut of f, cf. Remark 2.1).

Table 5.1 Color code for the following figures.

Color	Relative distance of a Ritz value θ to the spectrum
Red	$\operatorname{dist}(\theta, \Lambda(A)) < 10^{-13}$
Yellow	$10^{-13} \le \operatorname{dist}(\theta, \Lambda(A)) < 10^{-8}$
Green	$10^{-8} \le \operatorname{dist}(\theta, \Lambda(A)) < 10^{-3}$
Blue	$10^{-3} \le \operatorname{dist}(\theta, \Lambda(A))$

The evaluation of the integral formula (1.7) involves the determination of the minimum

$$\min_{x\in\Gamma} \int_0^t \int g_{S(\tau)}(x,y)\,\mathrm{d}\widetilde{\nu}_t(y)\,\mathrm{d}\tau = \min_{x\in\Gamma} \int_0^t g_{S(\tau)}(x,\xi)\,\mathrm{d}\tau.$$

Since the sets $S(t)$ are intervals, the associated Green functions are explicitly known. To be precise, if $S = [a,b]$ then the Green function $g_S(x,y)$ of $\overline{\mathbb{C}} \setminus S$ with pole at y is

$$g_S(x,y) = \log \left| \frac{\phi_S(x) - 1/\overline{\phi_S(y)}}{1 - \phi_S(x)/\phi_S(y)} \right|, \tag{5.2}$$

where $\phi_S(x)$ is the external conformal map for $\overline{\mathbb{C}} \setminus S$,

$$\phi_S(x) = \frac{2x-a-b}{b-a} + \sqrt{\left(\frac{2x-a-b}{b-a}\right)^2 - 1}. \tag{5.3}$$

To locate the minimum in (1.7), we make use of the fact that $S(t)$ lies to the left of the pole $\xi = 4$ and hence $g_{S(t)}(x,\xi)$ is strictly monotonically increasing for $x \in [\beta = 3, \xi)$ and strictly monotonically decreasing for $x \in (\xi, +\infty]$. Therefore the minimum in (1.7) can either be located at $x = \beta$ or $x = +\infty$. We have used a numerical quadrature formula to approximate the involved integral for both values of $x \in \{3, +\infty\}$ and to take the minimal value for each t. The resulting superlinear error estimate (1.7) is shown as the dashed red curve in Figure 5.1 (right). The predicted convergence rate is in good agreement with the actual convergence of rational Arnoldi. Note that in the first 17 iterations we have $S(t) = [\alpha, \beta] = [1/3, 3]$ (i.e., none of the Ritz values has converged), which is not away from $\Gamma = [3, +\infty]$, and therefore the predicted convergence rate must be 1.

5.2 The 1D-Laplacian

The following Lemma characterizes $S(t) = \operatorname{supp}(\sigma - \mu_t)$ when A is a finite-difference discretization of the 1D-Laplacian.

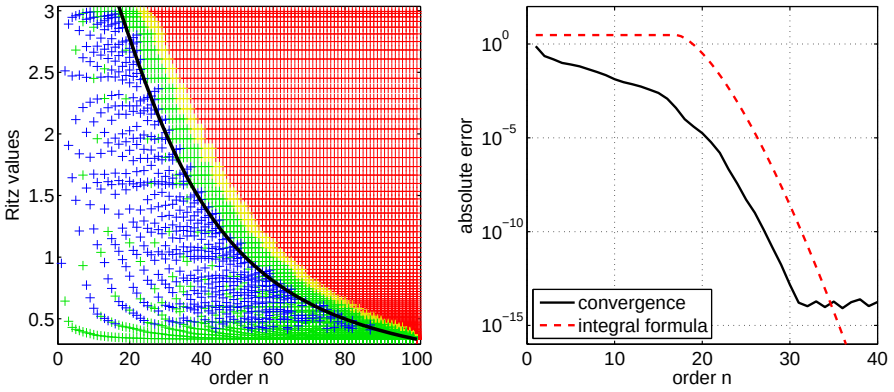

Fig. 5.1 On the left we plot the Ritz values of order $n = 1, \ldots, N$, associated with a Toeplitz matrix A and a random vector $\boldsymbol{b}$ of size $N = 100$. All poles are at $\xi = 4$. The colors indicate the distance of each Ritz value to a closest eigenvalue of A (cf. Table 5.1). The solid black curve is the right endpoint of the interval $S(t)$. On the right we plot the convergence of the rational Arnoldi method towards $f(A)\boldsymbol{b}$ with $f(z) = \log(3 - z)$ (solid black curve), together with the predicted convergence rate from our integral formula (dashed red curve).

Lemma 5.2 *The eigenvalues of*

$$A_N = \begin{pmatrix} 2 & -1 & \\ -1 & 2 & \ddots \\ & \ddots & \ddots \end{pmatrix} \in \mathbb{R}^{N\times N}$$

are asymptotically distributed like

$$\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x} = \frac{1}{\pi\sqrt{(x-\alpha)(\beta-x)}}, \quad \operatorname{supp}(\sigma) = [0,4].$$

For $\xi \le 0$ and $\nu_t = t\delta_\xi$ (single repeated pole) we have

$$S(t) = [a(t), 4], \quad a(t) = \begin{cases} 0, & t < \xi/(\xi-4); \\ \xi + t^2(4-\xi), & t < \xi/(\xi-4). \end{cases}$$

For $\nu_t = t(\delta_0 + \delta_\infty)/2$ (extended Krylov) we have

$$S(t) = [a(t), 4], \quad a(t) = 4t^2/(2-t)^2.$$

Proof We start with Lemma 5.1, in particular formula (5.1) of its proof, setting $\nu_t = t_1\delta_{\xi_1} + t_2\delta_\infty$, $t_1 + t_2 = t$ and $\xi_1 \ge \beta$. Using the linear transformation

$$\ell_q : [\alpha, \beta] \to [0, 4], \quad \ell_q(x) := x \cdot (q^2-1)/q + (1+q)^2/q,$$

we find that we may as well look for the solution $a(t) = \ell_q(b(t)) \in [0, 4]$ of

$$\sqrt{\frac{(1-q)^2}{(1+q)^2 - q\cdot a(t)}} = t_1\sqrt{\frac{4-\xi}{a(t)-\xi}} + t_2, \quad \xi = \ell_q(\xi_1). \tag{5.4}$$

It is interesting to see what happens when $q \to 0$. Because $a(t) = \ell_q(b(t))$ varies continuously in q, the solution of the limiting problem is

$$a(t) = \xi + \frac{t_1^2(4-\xi)}{(1-t_2)^2}.$$

A simple calculation shows that

$$\ell_q(\mathrm{toep}(\ldots, q^2, q^1, q^0, q^1, q^2, \ldots)) \to \mathrm{tridiag}(-1, 2, -1)$$

as $q \to 0$, and hence $a(t)$ is the left boundary of $S(t)$.

Example 5.2 Assume we run the rational Arnoldi method for the 1D-Laplacian $A = \mathrm{tridiag}(-1, 2, -1) \in \mathbb{R}^{100\times 100}$ and the function $f(z) = z^{-1/2}$, hence $\Gamma = [-\infty, 0]$. Let us also assume that there is only one single pole $\xi = -\sqrt{\lambda_{\min}\lambda_{\max}} \le 0$, the pole for which the linear convergence rate becomes minimal[1] with value

$$R = \frac{\sqrt[4]{\kappa} - 1}{\sqrt[4]{\kappa} + 1}, \quad \kappa = \frac{\lambda_{\max}}{\lambda_{\min}},$$

and in fact coincides with the linear convergence rate of the extended Krylov subspace method given in [28] (cf. also (1.8)).

As one can see in Figure 5.2, superlinear convergence sets in at iteration $\approx$ 14. It is interesting to note that our integral formula captures the nonconcavity of the convergence curve. This is explained by the fact that the minimum

$$\min_{x\in\Gamma} \int_0^t g_{S(\tau)}(x, \xi)\, \mathrm{d}\tau$$

is attained at $x = 0$ for early iterations, and at $x = -\infty$ for later iterations (i.e., after the cusp in the convergence curve occurred).

Example 5.3 We return to the example from the introduction, the convergence of the extended Krylov subspace method (i.e., $\nu_t = t(\delta_0 + \delta_\infty)/2$) for the 1D-Laplacian. By Lemma 5.2 we have $S(t) = \mathrm{supp}(\sigma - \mu_t) = [4t^2/(2-t)^2, 4]$. (The left endpoint of $S(t)$ is indicated as the solid black curve in the left plot of Figure 1.1.) To estimate the right-hand side of (1.7) we use

$$\min_{x\le 0} \frac{1}{2}\int_0^t g_{S(t)}(x,0) + g_{S(t)}(x,\infty)\, \mathrm{d}\tau \ge \frac{1}{2}\int_0^t \min_{x\le 0} \left(g_{S(t)}(x,0) + g_{S(t)}(x,\infty)\right)\, \mathrm{d}\tau. \tag{5.5}$$

It is shown in [28, Proposition 3.6] that for a positive interval $S = [a, b]$, the minimum on the right-hand side of (5.5) is attained at the point $x = -\sqrt{ab}$, leading to the linear convergence rate

$$\frac{\sqrt[4]{\kappa} - 1}{\sqrt[4]{\kappa} + 1} \le \exp\left(-\frac{2}{\sqrt[4]{\kappa}}\right), \quad \kappa = \frac{b}{a},$$

[1] This can be easily verified by the Bernstein–Walsh theory of polynomial approximation. In fact, every rational Krylov space with only one repeated pole ξ can be interpreted as a polynomial Krylov space with the matrix $(A - \xi I)^{-1}$.

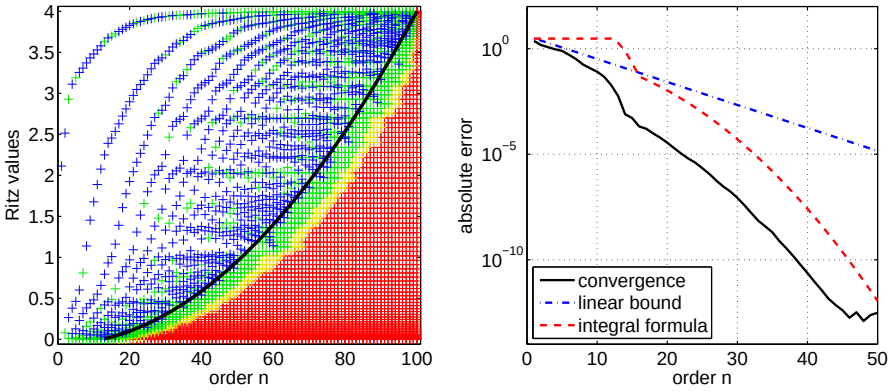

Fig. 5.2 On the left we plot the Ritz values of order $n = 1, \dots, N$, associated with the matrix $A = \text{tridiag}(-1, 2, -1)$ and a random vector $\boldsymbol{b}$ of size $N = 100$. The colors indicate the distance of each Ritz value to a closest eigenvalue of A (cf. Table 5.1). The solid black curve is our prediction for the region of converged Ritz values. On the right we show the convergence of the rational Arnoldi method towards $f(A)\boldsymbol{b}$ with $f(z) = z^{-1/2}$. The single pole $\xi < 0$ is asymptotically optimal and the linear convergence rate (dash-dotted blue line) equals that of the extended Krylov subspace method.

of the extended Krylov subspace method (cf. (1.8)). Therefore our integral formula (1.7) can be interpreted as a continuous geometric mean of linear convergence rates, the latter of which depend on a decreasing "effective condition number" $\kappa_t = b/a(t) = (2-t)^2/t^2$. Hence by some simple calculation,

$$\begin{aligned}
\limsup_{\substack{n,N\to\infty \\ n/N\to t}} \|f(A_N)\boldsymbol{b}_N - \boldsymbol{f}_{n,N}\|^{1/N} &\le \exp\left(\int_0^t \log\left(\frac{\sqrt[4]{\kappa_\tau}-1}{\sqrt[4]{\kappa_\tau}+1}\right) \mathrm{d}\tau\right) \\
&\le \exp\left(\int_0^t -\frac{2}{\sqrt[4]{\kappa_\tau}} \,\mathrm{d}\tau\right) \\
&= \exp\left(-4\arcsin\left(\sqrt{t/2}\right) + 2\sqrt{2t-t^2}\right) \\
&\le \exp\left(-\frac{\sqrt{8}\,t^{3/2}}{3}\right),
\end{aligned}$$

where the last bound has been obtained by a power series expansion about 0. The last expression is shown on the right-hand side of Figure 1.1 (the dotted red curve). To summarize, we can expect that the extended Krylov subspace method for the 1D-Laplacian (with N large enough) converges superlinearly as

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_n\| \lessapprox \exp\left(-\frac{\sqrt{8}\,n^{3/2}}{3\sqrt{N}}\right).$$

Note that at least this convergence is obtained by the extended Krylov subspace method with any Hermitian matrix which, when being scaled to spectral

interval $[0,4]$, has an eigenvalue density such that $S(t) \subseteq [4t^2(2-t)^2, 4]$, i.e., the spectrum should at least not be denser close the origin than the eigenvalues of the 1D-Laplacian.

6 Proofs

For a proof of Theorem 4.1 we will show several auxiliary results, always assuming without explicit mention that the assumptions of Theorem 4.1 hold. Our first Lemma 6.1 establishes that μ_t increases in t. Subsequently, we show in Lemma 6.2 that we may express the potential of the pole measure ν_t as an integral of the potentials of its derivative $\widetilde{\nu}_t$. Our integral formula for $G(t,z)$ is derived by integrating its derivative with respect to t. Inequalities for the difference quotient in terms of Green functions are established in Lemma 6.3, and some elementary facts about these Green functions established in Lemma 6.5 will allow to show in Lemma 6.6 that the t-derivative exists almost everywhere. After examining the limit for $t \to 0+$ in Lemma 6.4, we will be prepared to conclude with the proofs of Theorem 4.1 and Corollary 4.1.

Lemma 6.1 *For all $0 < \tau < t < 1$, the quantity $\mu_t - \mu_\tau$ is a positive measure, with $\operatorname{supp}(\mu_t - \mu_\tau) \subset S(\tau)$. In particular, $S(t)$ is decreasing in t.*

Proof The statement of Lemma 6.1 has already been shown in a more general setting in [13, Theorem A.10(a)], we give a (simplified) proof for completeness. Let $0 < \tau < t < 1$, then, by (3.4),

$$U^{\mu_t - \mu_\tau}(z) \le F_t - F_\tau + U^{\nu_t - \nu_\tau}(z) \tag{6.1}$$

for $z \in S(\tau)$. Consider the Hahn decomposition $\mu_+ - \mu_- = \mu_t - \mu_\tau = \sigma - \mu_\tau - (\sigma - \mu_t)$ with (positive) measures μ_+, μ_-. Since $\sigma - \mu_t \ge 0$, it follows that $\operatorname{supp}(\mu_+) \subset \operatorname{supp}(\sigma - \mu_\tau) = S(\tau)$. Taking into account that $\nu_t - \nu_\tau \ge 0$, we may apply the Principle of Domination [40, Theorem II.3.2], telling us that (6.1) holds for all $z \in \mathbb{C}$, but, again according to (3.4), there must be $\ge$ and thus equality for $z \in S(t)$. Since in addition all involved measures have finite potential in some open neighborhood of $S(t)$, we may apply de la Vallée–Poussin's Theorem [40, Theorem IV.4.5] to conclude that

$$(\mu_t - \mu_\tau - (\nu_t - \nu_\tau))|_{S(t)} \ge 0.$$

Since ν_t, ν_τ are supported in a set having an empty intersection with $S(t)$, it follows that $(\sigma - \mu_\tau)|_{S(t)} - (\sigma - \mu_t) \ge 0$ or $\mu_\tau \le \mu_t$, as claimed in Lemma 6.1. The other two claims of Lemma 6.1 are an immediate consequence.

In a next step, let us show that our assumptions on ν_t allow for the derivation of an integral formula similar to the one of Theorem 4.1.

Lemma 6.2 *For all $0 < t < 1$ and $z \in \mathbb{C}$ we have*

$$U^{\nu_t}(z) = \int_0^t U^{\widetilde{\nu}_\tau}(z)\,\mathrm{d}\tau.$$

Proof Let $0 < T < 1$. For some fixed function f continuous in $\mathrm{supp}(\nu_T)$, we consider the function $(0,T] \ni t \mapsto h(t) = \int f(y)\,\mathrm{d}\nu_t(y)$. We claim that $h(0+) = 0$, that h is Lipschitz continuous on $(0,T]$, and that h is differentiable almost everywhere on $(0,T]$, with $h'(t) = \int f(y)\,\mathrm{d}\widetilde{\nu}_t(y)$. In this case, we get for each $0 < t \le T$

$$h(t) = \int_0^t h'(\tau)\,\mathrm{d}\tau,$$

which can be rewritten as

$$\int f(y)\,d\nu_t(y) = \int_0^t \int f(y)\,\mathrm{d}\widetilde{\nu}_\tau(y)\,d\tau. \tag{6.2}$$

For a proof of our claims we set $C := \max_{y\in\mathrm{supp}(\nu_T)} |f(y)| < \infty$, and notice that, by monotonicity of the ν_t, we have $|f(y)| \le C$ for all $y \in \mathrm{supp}(\nu_t)$ and for all $0 < t \le T$. First recall that $\nu_t \ge 0$ of mass t implies that

$$|h(t)| \le \int |f(y)|\,d\nu_t(y) \le \int C\,\mathrm{d}\nu_t(y) = tC \to 0, \quad t \to 0+,$$

and thus $h(0+) = 0$. By a similar argument, we obtain for $0 < \tau < t \le T$ the condition

$$\Big|\frac{h(t)-h(\tau)}{t-\tau}\Big| \le \int C\,\mathrm{d}\frac{\nu_t - \nu_\tau}{t-\tau}(y) = C$$

and thus Lipschitz continuity, since again by assumption on ν_t the (signed) measure occurring on the right is a probability measure. Finally, the differentiability of h at t and the above formula $h'(t) = \int f(y)\,\mathrm{d}\widetilde{\nu}_t(y)$ is an immediate consequence of our assumption that $t \mapsto \nu_t$ has a weak-star derivative at t, see (4.2). Thus formula (6.2) is true for any function f continuous on $\mathrm{supp}(\nu_T)$.

In particular, for fixed $z \notin \mathrm{supp}(\nu_T)$, we may take $f(y) = \log(1/|z-y|)$ in (6.2), which gives the claim of Lemma 6.2. The remaining case $z \in \mathrm{supp}(\nu_T)$ is slightly more involved. For $M \ge 1$, we consider the regularized logarithmic kernel $f_M(y) = \min(M, \log(1/|y-z|))$ being clearly continuous in y and bounded below in $\mathrm{supp}(\nu_T)$ uniformly in M in terms of the diameter of $\mathrm{supp}(\nu_T)$. Also, for fixed y, the quantity $f_M(y)$ is increasing in M. We observe in addition that $\int f_M(y)\,\mathrm{d}\widetilde{\nu}_\tau(y)$ is bounded below uniformly in τ and M (since $\widetilde{\nu}_\tau$ is a probability measure), and increasing in M for fixed τ. Hence, applying the monotone convergence theorem (for several times), we obtain

$$\begin{aligned} U^{\nu_t}(z) &= \lim_{M\to\infty} \int f_M(y)\,\mathrm{d}\nu_t(y) = \lim_{M\to\infty} \int_0^t \int f_M(y)\,\mathrm{d}\widetilde{\nu}_\tau(y)\,\mathrm{d}\tau \\ &= \int_0^t \lim_{M\to\infty} \int f_M(y)\,\mathrm{d}\widetilde{\nu}_\tau(y)\,\mathrm{d}\tau = \int_0^t U^{\widetilde{\nu}_\tau}(z)\,\mathrm{d}\tau, \end{aligned}$$

as claimed in Lemma 6.2.

We notice that $z \mapsto G(t,z)$ is continuous also at infinity, with value $G(t,\infty) = -F_t$. We now establish a basic inequality refining (6.1), a similar one may be found in [13, Eqns. (A.22),(A.25)].

Lemma 6.3 *For all* $0 < \tau < t < 1$ *and* $z \in \overline{\mathbb{C}}$ *we have*

$$-\int g_{S(t)}(z,y)\,\mathrm{d}\frac{\nu_t-\nu_\tau}{t-\tau}(y) \le \frac{G(t,z)-G(\tau,z)}{t-\tau} \le -\int g_{S(\tau)}(z,y)\,\mathrm{d}\frac{\nu_t-\nu_\tau}{t-\tau}(y).$$

Proof First recall from [40, Theorem II.5.1(iv) and Eqn. (5.7)] that

$$\int g_S(z,y)\,\mathrm{d}\rho(y)$$

is $= 0$ quasi-everywhere on S for any compact $S \subset \mathbb{R}$ of positive capacity, where we use the abbreviation $\rho = \frac{\nu_t-\nu_\tau}{t-\tau}$ being a probability measure by the assumptions of Theorem 4.1. We may rewrite this expression as

$$\int g_S(z,y)\,\mathrm{d}\rho(y) = U^\rho(z) - U^{\widetilde{\rho}}(z) + \int g_S(\infty,y)\,\mathrm{d}\rho(y)$$

with $\widetilde{\rho}$ the balayage measure of ρ onto the given compact S (namely the unique measure supported in S with the same mass as ρ such that $U^{\rho-\widetilde{\rho}}$ is constant quasi-everywhere on S).

We start by showing the right-hand inequality and put $S = S(\tau)$. Recall from (3.4) that $G(t,z) \le 0$ and $G(\tau,z) = 0$ on S. Hence

$$\begin{aligned}&\frac{G(t,z)-G(\tau,z)}{t-\tau} + \int g_S(z,y)\,\mathrm{d}(\nu_t-\nu_\tau)(y) \\ &= U^\kappa(z) - U^{\widetilde{\rho}}(z) - \frac{F_t-F_\tau}{t-\tau} + \int g_S(\infty,y)\,\mathrm{d}\rho(y)\end{aligned} \tag{6.3}$$

is ≤ 0 quasi-everywhere on S, where we have set $\kappa = \frac{\mu_t-\mu_\tau}{t-\tau}$, which again by Lemma 6.1 is a probability measure. Since $\mathrm{supp}(\kappa) \subset S$, the Principle of Domination tells us that the expression (6.3) is ≤ 0 everywhere in $\overline{\mathbb{C}}$, as claimed above.

In order to show the left-hand inequality, we put $S = S(t)$, and thus $G(t,z) = 0$ and $G(\tau,z) \le 0$ on S. Hence the expression (6.3) is ≥ 0 quasi-everywhere on S which contains $\mathrm{supp}(\widetilde{\rho})$. Applying again the Principle of Domination gives the required inequality.

Lemma 6.4 *For all* $z \in \mathbb{C} \setminus \mathrm{supp}(\sigma)$*, we have*

$$\lim_{t\to 0+} U^{\mu_t}(z) = 0, \quad \lim_{t\to 0+} F_t = 0.$$

Proof The first claim is a consequence of the facts that, by Lemma 6.1, $\mu_t \to 0$ for $t \to 0+$, and $y \mapsto \log(1/|z-y|)$ is continuous on $\mathrm{supp}(\sigma) = \mathrm{supp}(\mu_t)$ by assumption on z. Also, by (3.4),

$$(1-t)F_t = I(\mu_t-\nu_t, \sigma-\mu_t) = -I(\nu_t, \sigma-\mu_t) + I(\sigma,\mu_t) - I(\mu_t,\mu_t).$$

We now recall that $\nu_t \to 0$ and $\sigma - \mu_t \to \sigma$ for $t \to 0$, the first (and the second) measure being supported on subsets of $\mathrm{supp}(\nu_T)$ (and $\mathrm{supp}(\sigma)$), where by assumption these two sets have empty intersection. Hence $I(\nu_t, \sigma-\mu_t) \to$

$I(0,\sigma)=0$. Also, since U^σ is assumed to be continuous, we have by definition of weak-star convergence that $I(\mu_t,\sigma)=\int U^\sigma \,\mathrm{d}\mu_t \to 0$. For the final term we obtain by the definition of logarithmic capacity the lower bound $I(\mu_t,\mu_t) \geq t^2 \log(1/\operatorname{cap}(\operatorname{supp}(\sigma)))$, and hence

$$\limsup_{t\to 0+} F_t \leq \limsup_{t\to 0+} \frac{-t^2}{1} \log \frac{1}{\operatorname{cap}(\operatorname{supp}(\sigma))} = 0.$$

On the other hand, letting $x\to\infty$ in (3.4) yields $F_t \geq 0$, as required for the second claim of Lemma 6.4.

In view of Lemma 6.4 we may define the quantities F_t, μ_t for $t=0$ by continuity, by setting $F_0=0$ and $\mu_0=\nu_0=0$, $S(0)=\operatorname{supp}(\sigma-\mu_0)=\operatorname{supp}(\sigma)$. It is not difficult to check that in this case Lemma 6.3 remains true also for $\tau=0$.

In order to be able to consider the limit $\tau\to t$ in Lemma 6.3, we require some auxiliary properties on Green functions stated below.

Lemma 6.5 *For fixed $z,y\in\overline{\mathbb{C}}\setminus\operatorname{supp}(\sigma)$ and $0\leq\tau<t<1$ we have*

$$0 \leq g_{S(\tau)}(z,y) \leq g_{S(t)}(z,y). \tag{6.4}$$

Moreover, if the map $t\mapsto \operatorname{cap}(S(t))$ is continuous in t then

$$\lim_{\tau\to t} \max_{y\in\Delta} |g_{S(t)}(z,y) - g_{S(\tau)}(z,y)| = 0 \tag{6.5}$$

for any closed $\Delta\subset\overline{\mathbb{C}}\setminus\operatorname{supp}(\sigma)$.

Proof From [40, Eqn. (II.4.3) and Theorem II.4.9] we know that $g_S(z,y)=g_S(y,z)\geq 0$ for all $x,y\in\mathbb{C}$ and all compact $S\subset\mathbb{R}$. We will first show the above claims for $z=\infty$, where we recall from [40, Corollary I.4.5 and Eqn. (I.4.8)] that

$$g_S(y,\infty) = \log\frac{1}{\operatorname{cap}(S)} - U^{\omega_S}(y),$$

with ω_S the Robin equilibrium measure of the compact set S, $\operatorname{supp}(\omega_S)\subset S$, and in particular $g_S(y,\infty)=0$ for quasi all $y\in S$. Letting $\tau<t$ and thus $S(t)\subset S(\tau)$, we conclude that the function $u_{\tau,t}(y)=g_{S(t)}(y,\infty)-g_{S(\tau)}(y,\infty)$ is $=0$ quasi-everywhere on $S(t)$, and hence ≥ 0 on $\mathbb{C}$ by the Principle of Domination, implying (6.4) for $z=\infty$.

Suppose now that $\operatorname{cap}(S(\tau))\to\operatorname{cap}(S(t))$ for $\tau\to t$, or, in other words, $u_{\tau,t}(\infty)\to 0$ for $\tau\to t$. Since $(u_{\tau,t})$ is harmonic in $\overline{\mathbb{C}}\setminus\operatorname{supp}(\sigma)$ and decreasing in τ by (6.4), and $u_{t,t}(y)=0$, we may conclude from the Harnack Principle [40, Theorem 0.4.10] that $u_{\tau,t}\to 0$ for $\tau\to t$ uniformly in Δ, a closed subset of $\overline{\mathbb{C}}\setminus\operatorname{supp}(\sigma)$, as claimed in (6.5) for $z=\infty$.

Finally, if $z\in\mathbb{C}\setminus\operatorname{supp}(\sigma)$, we denote by φ a Moebius transformation sending z to $\varphi(z)=\infty$, and hence $\varphi(S(t))$ is some compact subset of $\mathbb{C}$ with connected complement and empty interior. From [40, Eqn. (II.4.4)] we know

that $g_{S(\tau)}(z,y) = g_{S(\tau)}(y,z) = g_{\varphi(S(\tau))}(\varphi(y),\infty)$. Hence we may conclude as above provided that we are able to show that

$$\lim_{\tau\to t} \operatorname{cap}(S(\tau)) = \operatorname{cap}(S(t)) \qquad \text{implies} \qquad \lim_{\tau\to t} \operatorname{cap}(\varphi(S(\tau))) = \operatorname{cap}(\varphi(S(t))).$$

Indeed, for any decreasing family of compact sets E_t, we have $\operatorname{cap}(E_\tau) \to \operatorname{cap}(E_t)$ for $\tau \to t$ if and only if

$$\operatorname{cap}\Big((\bigcap_{\tau<t} E_\tau) \setminus (\bigcup_{\tau>t} E_\tau) \Big) = 0,$$

and this last property is invariant under the Moebius transformation φ.

Since the sets $S(t)$ are decreasing, also $t \mapsto \operatorname{cap}(S(t))$ is decreasing, meaning that this map only has a countable number of discontinuities. Let us denote by Θ the set of $\tau \in (0,1)$ such that $t \mapsto \operatorname{cap}(S(t))$ is continuous in τ, and $t \mapsto \nu_t$ is differentiable in τ.

Lemma 6.6 *Let $z \in \mathbb{C} \setminus \operatorname{supp}(\sigma)$, $0 < T < 1$.*

(a) The function $[0,T] \ni t \mapsto F_t$ is Lipschitz continuous and, for $t \in \Theta$, the following derivative exists

$$\frac{\partial}{\partial t} F_t = \int g_{S(t)}(\infty, y)\, \mathrm{d}\widetilde{\nu}_t(y).$$

(b) The function $[0,T] \ni t \mapsto U^{\mu_t}(z) - F_t$ is Lipschitz continuous and, for $t \in \Theta$, the following derivative exists

$$\frac{\partial}{\partial t}\Big(U^{\mu_t}(z) - F_t \Big) = \int \Big(\log \frac{1}{|z-y|} - g_{S(t)}(z,y) \Big)\, \mathrm{d}\widetilde{\nu}_t(y).$$

Proof From Lemma 6.3 with $z = \infty$ we obtain for $0 \le \tau < t \le T$ the inequalities

$$\int g_{S(\tau)}(y,\infty)\, \mathrm{d}\frac{\nu_t - \nu_\tau}{t-\tau}(y) \le \frac{F_t - F_\tau}{t-\tau} \le \int g_{S(t)}(y,\infty)\, \mathrm{d}\frac{\nu_t - \nu_\tau}{t-\tau}(y). \qquad (6.6)$$

Taking into account (6.4), we arrive at

$$\Big| \frac{F_t - F_\tau}{t-\tau} \Big| \le C := \max_{y \in \operatorname{supp}(\nu_T)} \max\{ g_{S(T)}(y,\infty), -g_{\operatorname{supp}(\sigma)}(y,\infty) \} < \infty,$$

and thus $t \mapsto F_t$ is indeed Lipschitz continuous. From the weak-star differentiability of $t \mapsto \nu_t$ and the continuity of $y \mapsto g_{S(t)}(y,\infty)$ on $\operatorname{supp}(\nu_T)$ we obtain from (6.6) the relations

$$\limsup_{\tau\to t-} \frac{F_t - F_\tau}{t-\tau} \le \int g_{S(t)}(y,\infty)\, \mathrm{d}\widetilde{\nu}_t(y),$$
$$\liminf_{\tau\to t+} \frac{F_t - F_\tau}{t-\tau} \ge \int g_{S(t)}(y,\infty)\, \mathrm{d}\widetilde{\nu}_t(y),$$

where the second inequality is obtained by exchanging the role of t and τ in (6.6).

Let $t \in \Theta$ be a point of continuity of the map $t \mapsto \operatorname{cap}(S(t))$. Then using again the weak-star differentiability of $t \mapsto \nu_t$ we obtain from (6.6)

$$\begin{aligned}&\liminf_{\tau\to t-} \frac{F_t - F_\tau}{t-\tau} - \int g_{S(t)}(y,\infty)\,\mathrm{d}\widetilde{\nu}_t(y)\\ &\geq -\limsup_{\tau\to t-} \max_{y\in\operatorname{supp}(\nu_T)} \Big|g_{S(t)}(y,\infty) - g_{S(\tau)}(y,\infty)\Big| \int \mathrm{d}\frac{\nu_t-\nu_\tau}{t-\tau}(x) = 0,\end{aligned}$$

the last equality following from (6.5). Similarly, by exchanging t and τ in (6.6), one obtains the missing expression for the lim sup for $\tau \to t+$, which allows us to insure that Lemma 6.6(a) holds.

A proof of part (b) is very similar, we just outline the necessary adjustments. First, from Lemma 6.3 for fixed $z \in \mathbb{C} \setminus \operatorname{supp}(\sigma)$ we obtain for $0 \leq \tau < t \leq T$ the inequality

$$\begin{aligned}&\int \Big(\log\frac{1}{|z-y|} - g_{S(t)}(z,y)\Big)\,\mathrm{d}\frac{\nu_t-\nu_\tau}{t-\tau}(y) \leq \frac{(U^{\mu_t}(z)-F_t)-(U^{\mu_\tau}(z)-F_\tau)}{t-\tau}\\ &\leq \int \Big(\log\frac{1}{|z-y|} - g_{S(\tau)}(z,y)\Big)\,\mathrm{d}\frac{\nu_t-\nu_\tau}{t-\tau}(y).\end{aligned}$$

Then we observe that the expression

$$\log\frac{1}{|z-y|} - g_{S(t)}(z,y)$$

is continuous in $y \in \operatorname{supp}(\nu_T)$ for fixed t (even if z itself is an element of $\operatorname{supp}(\nu_T)$, since two singularities cancel in this case), and decreasing in t for fixed y by (6.4). Hence, as before, we obtain Lipschitz continuity with constant

$$C = \max_{y\in\operatorname{supp}(\nu_T)} \max\left\{g_{S(T)}(y,z) - \log\frac{1}{|z-y|}, \log\frac{1}{|z-y|} - g_{\operatorname{supp}(\sigma)}(y,z)\right\} < \infty.$$

Also we may conclude as before for the derivative since (6.5) is true not only for $z = \infty$.

We are now prepared to present a proof of Theorem 4.1.

Proof of Theorem 4.1. We use slightly different arguments depending on whether z is finite or not. If $z = \infty$, then $G(t,\infty) = -F_t$. We have shown in Lemma 6.6(a) that this function of t is Lipschitz continuous and thus absolutely continuous, and have computed the (almost everywhere existing) derivative. The value at zero has been computed in Lemma 6.4, and hence for all $0 < t < 1$ we have

$$G(t,\infty) = \int_0^t \frac{\partial}{\partial t} G(\tau,\infty)\,\mathrm{d}\tau = -\int_0^t \int g_{S(\tau)}(\infty,y)\,\mathrm{d}\widetilde{\nu}_\tau(y)\,\mathrm{d}\tau,$$

as claimed in Theorem 4.1.

For $z \in \mathbb{C} \setminus \operatorname{supp}(\sigma)$ we write $G(t,z) = (U^{\mu_t}(z) - F_t) - U^{\nu_t}(z)$, argue as before for $U^{\mu_t}(z) - F_t$ by using Lemma 6.6(b) and Lemma 6.4, and use for $U^{\nu_t}(z)$ the representation of Lemma 6.2. Writing the two integrals under the same sign of integration, we find that

$$G(t,z) = -\int_0^t \left(\int \Big(g_{S(\tau)}(z,y) - \log \frac{1}{|z-y|} \Big) \, \mathrm{d}\widetilde{\nu}_\tau(y) + U^{\widetilde{\nu}_\tau}(z) \right) \mathrm{d}\tau,$$

and thus obtain the expression claimed in Theorem 4.1.

For the remaining case $z \in \operatorname{supp}(\sigma)$ we have to be a bit more careful. Let us first show that $t \mapsto \mu_t$ is differentiable almost everywhere, with derivative given by $\widetilde{\mu}_t$, the balayage measure of $\widetilde{\nu}_t$ onto $S(t)$. Indeed, subtracting the expressions for the derivatives obtained in Lemma 6.6(a),(b) we conclude that, for all $t \in \Theta$ and for all $z \notin \operatorname{supp}(\sigma)$ the following derivative exists

$$\frac{\partial}{\partial t} U^{\mu_t}(z) = U^{\widetilde{\nu}}(z) - \int g_{S(t)}(z,y) \, \mathrm{d}\widetilde{\nu}(y) + \int g_{S(t)}(\infty,y) \, \mathrm{d}\widetilde{\nu}_t(y) = U^{\widetilde{\mu}_t}(z).$$

Denote by ω any weak-star limit of the probability measures $\frac{\mu_t - \mu_\tau}{t-\tau}$ for $\tau \to t$, and thus $\operatorname{supp}(\omega) \subset \operatorname{supp}(\sigma)$. Since for $z \notin \operatorname{supp}(\sigma)$ the function $y \mapsto \log(1/|y-z|)$ is continuous on $\operatorname{supp}(\sigma)$, we conclude that $U^{\omega}(z) = U^{\widetilde{\mu}_t}(z)$ for all $z \notin \operatorname{supp}(\sigma)$, and hence $\omega = \widetilde{\mu}_t$ by the Unicity Theorem [40, Theorem II.2.1], that is, $\tau \mapsto \mu_\tau$ is weak-star differentiable at $t \in \Theta$. We thus may apply Lemma 6.2 for both families of measures $t \mapsto \mu_t$ and $t \mapsto \nu_t$, and with the above result of the first part of the proof we obtain, for all $0 < t < 1$ and all $z \in \mathbb{C}$,

$$\begin{aligned} G(t,z) &= \int_0^t \Big(U^{\widetilde{\mu}_\tau - \widetilde{\nu}_\tau}(z) - \int g_{S(\tau)}(\infty,y) \, \mathrm{d}\widetilde{\nu}_\tau(y) \Big) \, \mathrm{d}\tau \\ &= -\int_0^t \int g_{S(\tau)}(z,y) \, \mathrm{d}\widetilde{\nu}_\tau(y) \, \mathrm{d}\tau, \end{aligned}$$

as claimed in Theorem 4.1. □

Remark 6.1 Notice that, for all continuous functions f, formula (6.2) can be written equivalently in a more abstract setting as $\nu_t = \int_0^t \widetilde{\nu}_\tau \, \mathrm{d}\tau$. We therefore have shown in the preceding proof that

$$\mu_t = \int_0^t \widetilde{\mu}_\tau \, \mathrm{d}\tau, \quad \text{where } \widetilde{\mu}_t \text{ is the balayage of } \widetilde{\nu}_t \text{ onto } S(t).$$

This integral formula for the extremal measure (and for the constraint σ for $t \to 1$) is part of what is usually called the Buyarov–Rakhmanov formula [11]. This formula has been established in a more general setting in [13, Theorem A.10], but under technical assumptions which are not easy to verify.

Remark 6.2 In general it is quite difficult to determine for a given family of pole measures ν_t and a constraint σ the corresponding sets $S(t) = \mathrm{supp}(\sigma - \mu_t)$ which are required in our integral formula of Theorem 4.1. Even worse, given a family of decreasing compact sets $\widehat{S}(t)$ (say, a finite union of compact intervals with endpoints depending continuously on t), and a family of pole measures ν_t as described in Theorem 4.1, $\mathrm{supp}(\nu_t) \cap \widehat{S}(0)$ being empty, and $\widetilde{\mu}_t$ being the balayage of $\widetilde{\nu}_t$ onto $\widehat{S}(t)$, then it is possible to show using Remark 6.1 that $S(t) = \widehat{S}(t)$ holds for the constraint

$$\sigma = \int_0^1 \widetilde{\mu}_t \, \mathrm{d}t.$$

Proof of Corollary 4.1. Writing more explicitly

$$\mathcal{G}(\nu_t) := \sup_{z \in \Gamma} G(t, z),$$

for an increasing family of pole measures ν_t, we first claim the optimality property of Corollary 4.1 that

$$\mathcal{G}(\nu_t) \geq \mathcal{G}(\underline{\nu}_t) \tag{6.7}$$

for all $t \in (0, 1)$ for some particular $\underline{\nu}_t$ to be constructed as follows: Following [4], we consider the problem of minimizing $I(\mu - \nu)$ over all measures μ, ν of mass t satisfying the constraints $\mu \leq \sigma$ and $\nu \leq \widehat{\sigma}$, for some fixed $\widehat{\sigma}$ supported in Γ with continuous potential to be specified later. From [4, Theorem 2.2] we know that there is a unique such couple of extremal measures $\underline{\mu}_t, \underline{\nu}_t$, being characterized by the equilibrium conditions

$$U^{\underline{\mu}_t - \underline{\nu}_t}(z) \begin{cases} = C_{1,t}, & z \in \mathrm{supp}(\sigma - \underline{\mu}_t), \\ = C_{2,t}, & z \in \mathrm{supp}(\widehat{\sigma} - \underline{\nu}_t), \\ \in [C_{2,t}, C_{1,t}] & \text{otherwise}, \end{cases} \tag{6.8}$$

for some constants $C_{1,t} > C_{2,t}$. Here it is possible to choose $\widehat{\sigma}$ sufficiently large so that $\mathrm{supp}(\widehat{\sigma} - \nu) = \Gamma$, in other words, this constraint is not active. Then, comparing the first and last condition of (6.8) with [5, Lemma A.1(d)] we conclude that $\underline{\mu}_t$ describes the Ritz value distribution for the pole distribution $\underline{\nu}_t$, and hence

$$\mathcal{G}(\underline{\nu}_t) = \max_{z \in \Gamma} U^{\underline{\mu}_t - \underline{\nu}_t}(z) - C_{1,t} = C_{2,t} - C_{1,t}. \tag{6.9}$$

Let ν_t be now some arbitrary pole distribution, with the corresponding Ritz value distribution μ_t. The Domination Lemma of Charges [40, Lemma VIII.2.4] applied to the right-hand side of

$$U^{\mu_t - \nu_t}(z) - U^{\underline{\mu}_t - \underline{\nu}_t}(z) = U^{(\sigma - \underline{\mu}_t) - \nu_t}(z) - U^{(\sigma - \mu_t) - \underline{\nu}_t}(z),$$

tells us that

$$\text{"}\inf\text{"}_{z \in \mathrm{supp}(\sigma - \mu_t)} U^{\mu_t - \nu_t}(z) - U^{\underline{\mu}_t - \underline{\nu}_t}(z) \leq \text{"}\sup\text{"}_{z \in \mathrm{supp}(\underline{\nu}_t)} U^{\mu_t - \nu_t}(z) - U^{\underline{\mu}_t - \underline{\nu}_t}(z),$$

where " inf " (or " sup ") means infimum (or supremum) neglecting sets of zero capacity. Keeping in mind that $S(t) = \mathrm{supp}(\sigma - \mu_t)$ and $\mathrm{supp}(\underline{\nu}_t) \subset \Gamma$, we conclude that

$$\inf_{z\in S(t)} U^{\mu_t-\nu_t}(z) - U^{\underline{\mu}_t-\underline{\nu}_t}(z) \le \sup_{z\in\Gamma} U^{\mu_t-\nu_t}(z) - U^{\underline{\mu}_t-\underline{\nu}_t}(z). \tag{6.10}$$

Using (3.4) and the last relation of (6.8), we also find that

$$\inf_{z\in S(t)} U^{\mu_t-\nu_t}(z) - U^{\underline{\mu}_t-\underline{\nu}_t}(z) \ge F_t - C_{1,t},$$

and, by the second relation of (6.8),

$$\sup_{z\in\Gamma} U^{\mu_t-\nu_t}(z) - U^{\underline{\mu}_t-\underline{\nu}_t}(z) = \sup_{z\in\Gamma} U^{\mu_t-\nu_t}(z) - C_{2,t} = \mathcal{G}(\nu_t) + F_t - C_{2,t}.$$

Combining with (6.9) and (6.10), this allows to conclude that $\mathcal{G}(\nu_t) \ge C_{2,t} - C_{1,t} = \mathcal{G}(\underline{\nu}_t)$, as claimed in (6.7).

Finally, from [22, Lemme 6.3.1] we know that $\underline{\nu}_t$ is increasing in t. The integral formula for $G(\underline{\nu}_t)$ has been given in [22, Théorème 6.3.4] (see also [4, Theorem 5.1]), and, combining the idea of the proof of [4, Theorem 5.1] with the techniques of the present paper we obtain the characterization of the derivatives $\widetilde{\underline{\mu}}_t$, and $\widetilde{\underline{\nu}}_t$ as stated in Corollary 4.1. □

7 Conclusion

We have analyzed the error of the rational Arnoldi method for approximating the expression $f(A)\boldsymbol{b}$. Several authors exhibited a superlinear convergence rate for entire functions like the exponential function [9, 23, 38], but only a linear convergence rate for functions f with finite singularities [9,17,18,23,28]. We presented several numerical examples for symmetric A where superlinear convergence also takes place for functions with finite singularities. For the most prominent example $f(z) = z^{-1}$ and polynomial Krylov spaces, where the Arnoldi method reduces to CG, such a superlinear convergence behavior has been quantified in [6–8] for sequences of matrices having a joint eigenvalue distribution. Based on previous work [5] on the asymptotic behavior of rational Ritz values, we have extended in this paper the findings of [6–8] to general f and general distributions of poles of the underlying rational Krylov spaces.

One drawback of the rational Arnoldi method is that the user has to device parameters, namely the poles of the rational Krylov spaces. In Corollary 4.1 we have shown that there are asymptotically optimal poles, but for the moment we have no numerical method to efficiently generate such poles from A, $\boldsymbol{b}$ and the set Γ of singularities of f. We suspect that the adaptive pole selection of [24] (see also [19]) is asymptotically optimal in the sense of Corollary 4.1, but for the moment we do not have a rigorous proof of this statement.

References

1. B. Beckermann, A note on the convergence of Ritz values for sequences of matrices, Technical Report ANO 408, Labo Paul Painlevé, Université de Lille I, France (2000).
2. B. Beckermann, On a conjecture of E.A. Rakhmanov, Constr. Approx., 16, 427–448 (2000).
3. B. Beckermann, Discrete orthogonal polynomials and superlinear convergence of Krylov subspace methods in numerical linear algebra, in Orthogonal Polynomials and Special Functions, F. Marcellan and W. Van Assche, eds., vol. 1883 of Lecture Notes in Mathematics, Springer, 119–185 (2006).
4. B. Beckermann and A. Gryson, Extremal rational functions on symmetric discrete sets and superlinear convergence of the ADI method, Constr. Approx., 32, 393–428 (2010).
5. B. Beckermann, S. Güttel and R. Vandebril, On the convergence of rational Ritz values, SIAM J. Matrix Anal. Appl., 31, 1740–1774 (2010).
6. B. Beckermann and A. B. J. Kuijlaars, On the sharpness of an asymptotic error estimate for conjugate gradients, BIT, 41, 856–867 (2001).
7. B. Beckermann and A. B. J. Kuijlaars, Superlinear convergence of conjugate gradients, SIAM J. Numer. Anal., 39, 300–329 (2001).
8. B. Beckermann and A. B. J. Kuijlaars, Superlinear CG convergence for special right-hand sides, Electronic Transactions on Numerical Analysis, 14, 1–19 (2002).
9. B. Beckermann and L. Reichel, Error estimation and evaluation of matrix functions via the Faber transform, SIAM J. Numer. Anal., 47, 3849–3883 (2009).
10. A. Bultheel, P. González-Vera, E. Hendriksen and O. Njåstad, Orthogonal Rational Functions, vol. 5 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, United Kingdom (1999).
11. V. Buyarov and E. A. Rakhmanov, Families of equilibrium measures with external field on the real axis, Sb. Math., 190, 791–802 (1999).
12. D. Calvetti, L. Reichel and Q. Zhang, Iterative exponential filtering for large discrete ill-posed problems, Numer. Math., 83, 535–556 (1999).
13. J. Coussement and W. Van Assche, A continuum limit of relativistic Toda lattice: asymptotic theory of discrete Laurent orthogonal polynomials with varying recurrence coefficients, J. Phys. A, 38, 3337–3366 (2005).
14. K. Deckers and A. Bultheel, Rational Krylov sequences and orthogonal rational functions, Tech. Rep. TW499, Katholieke Universiteit Leuven, Departement Computerwetenschappen (2008).
15. P. D. Dragnev and E. B. Saff, Constrained energy problems with applications to orthogonal polynomials of a discrete variable, Journal d'Analyse Mathématique, 72, 223–259 (1997).
16. T. A. Driscoll, K.-C. Toh and L. N. Trefethen, From potential theory to matrix iterations in six steps, SIAM Review, 40, 547–578 (1998).
17. V. Druskin and L. Knizhnerman, Two polynomial methods of calculating functions of symmetric matrices, USSR Comput. Maths. Math. Phys., 29, 112–121 (1989).
18. V. Druskin and L. Knizhnerman, Extended Krylov subspaces: Approximation of the matrix square root and related functions, SIAM J. Matrix Anal. Appl., 19, 775–771 (1998).
19. V. Druskin, C. Lieberman and M. Zaslavsky, On adaptive choice of shifts in rational Krylov subspace reduction of evolutionary problems, SIAM J. Sci. Comput. 32, 2485–2496 (2010).
20. T. Ericsson, Computing functions of matrices using Krylov subspace methods, Technical Report, Department of Computer Science, Chalmers University of Technology, Sweden (1990).
21. E. Gallopoulos and Y. Saad, Efficient solution of parabolic equations by Krylov approximation methods, Numer. Linear Algebra Appl., 13, 1236–1264 (1992).
22. A. Gryson, Minimisation d'énergie sous contraintes, Applications en algèbre linéaire et en contrôle linéaire, PhD Thesis, University of Lille (2009).
23. S. Güttel, Rational Krylov Methods for Operator Functions, PhD Thesis, Technische Universität Bergakademie Freiberg (2010).

24. S. Güttel and L. Knizhnerman, Automated parameter selection for rational Arnoldi approximation of Markov functions, Proc. Appl. Math. Mech., submitted (2011).
25. S. Helsen, A. B. J. Kuijlaars and M. Van Barel, Convergence of the isometric Arnoldi process, SIAM J. Matrix Anal. Appl., 26, 782–809 (2005).
26. N. J. Higham, Functions of Matrices: Theory and Computation, SIAM, Philadelphia, PA, USA (2008).
27. M. Kac, W. L. Murdock and G. Szegő, On the eigenvalues of certain Hermitian forms, Indiana Univ. Math. J. (formerly known as Journal of Rational Mechanics and Analysis), 2, 767–800 (1953).
28. L. Knizhnerman and V. Simoncini, A new investigation of the extended Krylov subspace method for matrix function evaluations, Numer. Linear Algebra Appl., 17, 615–638 (2010).
29. L. Knizhnerman and V. Simoncini, Convergence analyis of the extended Krylov subspace method for the Lyapunov equation, to appear in Numer. Math. (2011).
30. A. B. J. Kuijlaars, Which eigenvalues are found by the Lanczos method?, SIAM J. Matrix Anal. Appl., 22, 306–321 (2000).
31. A. B. J. Kuijlaars, Convergence analysis of Krylov subspace iterations with methods from potential theory, SIAM Review, 48, 3–40, (2006).
32. C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Nat. Bur. Standards, 45, 225–280 (1950).
33. E. M. Nikishin and V. N. Sorokin, Rational Approximations and Orthogonality, Transl. Amer. Math. Soc., 92, Providence, R.I. (1991).
34. B. N. Parlett, The Symmetric Eigenvalue Problem, Prentice-Hall, Englewood Cliffs, NJ, USA (1980).
35. E. A. Rakhmanov, Equilibrium measure and the distribution of zeros on the extremal polynomials of a discrete variable, Sb. Math., 187, 1213–1228 (1996).
36. A. Ruhe, Rational Krylov sequence methods for eigenvalue computation, Lin. Alg. Appl., 58, 391–405 (1984).
37. A. Ruhe, Rational Krylov algorithms for nonsymmetric eigenvalue problems, in Recent Advances in Iterative Methods, G. H. Golub, A. Greenbaum, and M. Luskin, eds., IMA Volumes in Mathematics and its Applications, Springer-Verlag, New York, 149–164 (1994).
38. Y. Saad, Analysis of some Krylov subspace approximations to the exponential operator, SIAM J. Numer. Anal., 29, 209–228 (1992).
39. Y. Saad, Iterative Methods for Sparse Linear Systems, SIAM, PA, USA, second ed. (2003).
40. E. B. Saff and V. Totik, Logarithmic Potentials with External Fields, Springer-Verlag, Berlin (1997).
41. J. L. Walsh, Interpolation and Approximation by Rational Functions in the Complex Domain, 5th ed., Amer. Math. Soc., Providence (1969).