BIT

A black-box rational Arnoldi variant for Cauchy–Stieltjes matrix functions

Stefan Güttel · Leonid Knizhnerman

Abstract Rational Arnoldi is a powerful method for approximating functions of large sparse matrices times a vector. The selection of asymptotically optimal parameters for this method is crucial for its fast convergence. We present and investigate a novel strategy for the automated parameter selection when the function to be approximated is of Cauchy–Stieltjes (or Markov) type, such as the matrix square root or the logarithm. The performance of this approach is demonstrated by numerical examples involving symmetric and nonsymmetric matrices. These examples suggest that our black-box method performs at least as well, and typically better, as the standard rational Arnoldi method with parameters being manually optimized for a given matrix.

1 Introduction

An important problem arising in science and engineering is the computation of the matrix-vector product $f(A)\boldsymbol{v}$, where $A \in \mathbb{C}^{N \times N}$, $\boldsymbol{v} \in \mathbb{C}^{N}$, and f is a function such

S. Güttel (✉)
School of Mathematics, The University of Manchester, Manchester M13 9PL, UK
e-mail: stefan.guettel@manchester.ac.uk

L. Knizhnerman
Central Geophysical Expedition, 38/3 Narodnogo opolcheniya St., 123298 Moscow, Russia
e-mail: lknizhnerman@gmail.com

that $f(A)$ is defined. The term $f(A)$ is called a *matrix function*, and a sufficient condition for $f(A)$ to be defined is that $f(z)$ be analytic in a neighborhood of $\Lambda(A)$, the set of eigenvalues of A. For more detailed information on matrix functions and their possible definitions we refer to the monograph by Higham [28].

In most applications, A is very large and sparse (e.g., a finite-difference or finite-element discretization of a differential operator), so that explicitly computing and storing the generally dense matrix $f(A)$ is infeasible. In the recent years, polynomial and rational Krylov methods have proven to be the methods of choice for computing approximations to $f(A)v$ efficiently, without forming $f(A)$ explicitly. Rational Krylov methods require the solution of shifted linear systems with A, and the approximations they deliver are rational matrix functions of the form $r_n(A)v$, with r_n being a rational function of type $(n-1, n-1)$ and $n \ll N$. Polynomial Krylov methods are a special case obtained when r_n reduces to a polynomial. Although each iteration of a rational Krylov method may be considerably more expensive than a polynomial Krylov iteration, rational functions often have superior approximation properties than polynomials, which may lead to a reduction of the overall Krylov iteration number.

An important pitfall, which possibly prevents rational Krylov methods from being used more widely in practice, is the selection of optimal poles of the rational functions r_n. These poles are parameters that should be chosen based on the function f, the spectral properties of the matrix A, and the vector v. While the function f is usually known a priori, spectral properties of A may be difficult to access when A is large. Recently, interesting strategies for the automated selection of the poles have been proposed for the exponential function and the transfer function of symmetric and nonsymmetric matrices, see [18] and [19], respectively. The algorithms proposed in these two papers gather spectral information from quantities computed during the rational Krylov iteration, and they only require some initial rough estimate of the "spectral region" of the matrix A (see [19]). Because the exponential function can be represented as a Cauchy (or Fourier) integral along the imaginary axis, one can motivate that it is natural to choose the poles of the rational Krylov space as mirrored images of rational Ritz values for A with respect to the imaginary axis, and a similar reasoning applies for the resolvent function, see [18, 19]. In the case of functions of Cauchy–Stieltjes (or Markov) type the imaginary axis does not play a prominent role, such that a direct application of the idea of mirrored Ritz values is not helpful. The aim of this paper is to give analysis and numerical evidence for a heuristic pole selection strategy for such functions of non-necessarily symmetric matrices proposed in [26]. Cauchy–Stieltjes functions are of the form

$$f(z) = \int_\Gamma \frac{\mathrm{d}\gamma(x)}{z - x} \tag{1.1}$$

with a (complex) measure γ supported on a prescribed closed set $\Gamma \subset \mathbb{C}$. Particularly important examples of such functions are

$$f_1(z) = z^{-1/2} = \int_{-\infty}^0 \frac{1}{z - x} \frac{\mathrm{d}x}{\pi\sqrt{-x}},$$

$$f_2(z) = \frac{e^{-t\sqrt{z}} - 1}{z} = \int_{-\infty}^0 \frac{1}{x - z} \frac{\sin(t\sqrt{-x})\mathrm{d}x}{\pi x}.$$

For instance, certain solutions of the equation

$$Au(t) - \frac{d^2u}{dt^2}(t) = g(t)v$$

can be represented as $u(t) = f(A)v$ with f being a rational function of f_1 and f_2 (cf. [15, 17]). Functions of this type also arise in the context of computation of Neumann-to-Dirichlet and Dirichlet-to-Neumann maps [3, 16], the solution of systems of stochastic differential equations [2], and in quantum chromodynamics [22]. Another relevant Cauchy–Stieltjes function is

$$f_3(z) = \frac{\log(1 + z)}{z} = \int_{-\infty}^{-1} \frac{-1/x}{z - x} \, dx;$$

see [28] for applications of this function. The variant of the rational Arnoldi method presented here is parameter-free and seems to enjoy remarkable convergence properties and robustness. We believe that our method outperforms (in terms of required iteration numbers) any other available rational Krylov method for the approximation of $f(A)v$, and we will demonstrate this by a number of representative numerical tests.

This paper is structured as follows. In Sect. 2 we review the rational Arnoldi method and some of its important properties. In Sect. 3 we present our automated version of the rational Arnoldi method for functions of Cauchy–Stieltjes type (1.1). The problem of estimating the error of Arnoldi approximations is dealt with in Sect. 4. In Sect. 5 we study the asymptotic convergence of our method and compare it to other available methods for the approximation of $f(A)v$. Finally, in Sect. 6 we demonstrate the performance of our parameter-free algorithm for a large-scale numerical example. Throughout this paper, $\| \cdot \|$ denotes the Euclidian norm, I is the identity matrix of size $N \times N$, and $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ is the extended complex plane. Vectors are printed in bold face.

2 Rational Arnoldi method

A popular rational Krylov method for the approximation of $f(A)v$ is known as the *rational Arnoldi method*. It is based on the extraction of an approximation $f_n = r_n(A)v$ from a rational Krylov space [35, 36]

$$\mathcal{Q}_n(A, v) := \text{span}\left\{ \frac{p_{n-1}}{q_{n-1}}(A)v \; : \; p_{n-1} \text{ polynomial of degree } \leq n - 1 \right\},$$

$$q_{n-1}(z) := \prod_{\substack{j=1 \\ \xi_j \neq \infty}}^{n-1} (z - \xi_j), \tag{2.1}$$

where the parameters $\xi_j \in \overline{\mathbb{C}}$ (the *poles*) are different from the eigenvalues $\Lambda(A)$. Note that fractions in (2.1) range over the linear space of rational functions of type $(n - 1, n - 1)$ with a prescribed denominator q_{n-1}, and that $\mathcal{Q}_n(A, v)$ reduces to

a polynomial Krylov space if we set all poles $\xi_j = \infty$. If $\mathscr{Q}_n(A, \boldsymbol{v})$ is of full dimension n, as we assume in the following, we can compute an orthonormal basis $V_n = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n] \in \mathbb{C}^{N \times n}$ of this space. The rational Arnoldi approximation is then defined as

$$\boldsymbol{f}_n := V_n f(A_n) V_n^* \boldsymbol{v}, \quad A_n := V_n^* A V_n. \tag{2.2}$$

If n is relatively small, then $f(A_n)$ can be evaluated easily using algorithms for dense matrix functions (see [28]). A stable iterative procedure for computing the orthonormal basis V_n is the *rational Arnoldi algorithm* by Ruhe [36], which we briefly review in the following. Let σ be a finite number different from all ξ_j, and consider the translated operator $\tilde{A} := A - \sigma I$ and the translated poles $\tilde{\xi}_j := \xi_j - \sigma$. Note that the rational Krylov space $\tilde{\mathscr{Q}}_n(\tilde{A}, \boldsymbol{v})$ built with the poles $\tilde{\xi}_j$ coincides with $\mathscr{Q}_n(A, \boldsymbol{v})$. We may therefore equally well construct an orthonormal basis for $\tilde{\mathscr{Q}}_n(\tilde{A}, \boldsymbol{v})$ as follows:

Starting with $\boldsymbol{v}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|$, in each iteration $j = 1, \ldots, n$ one utilizes a modified Gram–Schmidt procedure to orthogonalize the vector

$$\boldsymbol{w}_{j+1} = (I - \tilde{A}/\tilde{\xi}_j)^{-1} \tilde{A} \boldsymbol{v}_j \tag{2.3}$$

against $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j\}$, yielding the vector $\boldsymbol{v}_{j+1}$, $\|\boldsymbol{v}_{j+1}\| = 1$ which satisfies

$$\boldsymbol{v}_{j+1} h_{j+1,j} = \boldsymbol{w}_{j+1} - \sum_{i=1}^{j} \boldsymbol{v}_i h_{i,j}, \quad h_{i,j} = \boldsymbol{v}_i^* \boldsymbol{w}_{j+1}. \tag{2.4}$$

Equating (2.3) and (2.4), and collecting the orthogonalization coefficients in $H_n = [h_{i,j}] \in \mathbb{C}^{n \times n}$, we obtain in the n-th iteration of the rational Arnoldi algorithm a decomposition

$$\tilde{A} V_n \big(I_n + H_n \operatorname{diag}(\tilde{\xi}_1^{-1}, \ldots, \tilde{\xi}_n^{-1})\big) + \tilde{A} \boldsymbol{v}_{n+1} h_{n+1,n} \tilde{\xi}_n^{-1} \boldsymbol{e}_n^T = V_n H_n + \boldsymbol{v}_{n+1} h_{n+1,n} \boldsymbol{e}_n^T,$$

or in more compact form after defining $K_n := I_n + H_n \operatorname{diag}(\tilde{\xi}_1^{-1}, \ldots, \tilde{\xi}_n^{-1})$,

$$\tilde{A} V_n K_n + \tilde{A} \boldsymbol{v}_{n+1} h_{n+1,n} \tilde{\xi}_n^{-1} \boldsymbol{e}_n^T = V_n H_n + \boldsymbol{v}_{n+1} h_{n+1,n} \boldsymbol{e}_n^T, \tag{2.5}$$

where I_n denotes the $n \times n$ identity matrix and $\boldsymbol{e}_n$ its last column. Using the convention that $\tilde{\xi}_n = \infty$ (i.e., $\xi_n = \infty$, which corresponds to a polynomial Krylov step, cf. [8, 24]), Eq. (2.5) reduces to

$$\tilde{A} V_n K_n = V_n H_n + \boldsymbol{v}_{n+1} h_{n+1,n} \boldsymbol{e}_n^T. \tag{2.6}$$

If the matrix H_n appended with the row $h_{n+1,n} \boldsymbol{e}_n^T$ is an unreduced upper Hessenberg matrix (that is, all the coefficients $h_{j+1,j}$ of (2.4) are nonzero), then the right-hand side of (2.6) is of full rank n and therefore K_n is invertible. Otherwise, if $h_{n+1,n} = 0$, then $\mathscr{Q}_n(A, \boldsymbol{v})$ is an A-invariant subspace and we have a lucky early termination. The matrix A_n required for computing the rational Arnoldi approximation (2.2) can be calculated from (2.6) without explicit projection as

$$A_n = V_n^* A V_n = V_n^* \tilde{A} V_n + \sigma I_n = H_n K_n^{-1} + \sigma I_n. \tag{2.7}$$

Note that A_n corresponds to the projection of the non-translated operator A onto the rational Krylov space $\mathcal{Q}_n(A, \boldsymbol{v})$.

Remark 2.1 In exact arithmetic the rational Arnoldi approximation (2.2) is independent of the choice of the translation σ. However, for numerical stability of the rational Arnoldi algorithm, σ should have a large enough distance to the poles ξ_j relative to $\|A\|$, because otherwise the pole and the zero in the fraction of (2.3) may "almost cancel", causing accuracy loss in the rational Krylov basis [32].

It is well known that the rational Arnoldi approximation $\boldsymbol{f}_n$ defined in (2.2) is (in some sense) a near-optimal approximation for $f(A)\boldsymbol{v}$ from the space $\mathcal{Q}_n(A, \boldsymbol{v})$ (see, e.g., [8, 15, 24]), that is, $\boldsymbol{f}_n$ is very close to the orthogonal projection $V_n V_n^* f(A)\boldsymbol{v}$. Therefore the poles ξ_j need to be chosen such that $\mathcal{Q}_n(A, \boldsymbol{v})$ contains a good approximation to $f(A)\boldsymbol{v}$, and of course, such a choice depends both on the spectral properties of A and the function f. This necessity for choosing optimal parameters is a serious problem that prevents rational Arnoldi from being used in practice more widely. Before discussing our automated pole selection strategy, we list some well-known properties of the rational Arnoldi approximation (2.2). The interested reader is referred to [8, 24, 25] for further details.

1. By the definition of a rational Krylov space $\mathcal{Q}_n(A, \boldsymbol{v})$ (cf. (2.1)), there exists a rational function r_n of type $(n-1, n-1)$ such that

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

2. This function r_n is a rational interpolant for f with prescribed denominator q_{n-1} and interpolation nodes $\Lambda(A_n) = \{\theta_1, \ldots, \theta_n\}$, the so-called *rational Ritz values*. Defining the *rational nodal function* s_n of type $(n, n-1)$,

$$s_n(z) := \frac{\prod_{k=1}^n (z - \theta_k)}{q_{n-1}(z)}, \tag{2.8}$$

by the Hermite–Walsh formula for rational interpolants (see, e.g., [39, Theorem VIII.2] or [6]) we have

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

and therefore

$$\left\| f(A)\boldsymbol{v} - r_n(A)\boldsymbol{v} \right\| \le \left\| s_n(A)\boldsymbol{v} \right\| \cdot \left\| \int_\Gamma \frac{(xI - A)^{-1}}{s_n(x)} \, \mathrm{d}\gamma(x) \right\|. \tag{2.9}$$

3. The term $\|s_n(A)\boldsymbol{v}\|$ in (2.9) is minimal among all rational functions of the form $\tilde{s}_n(z) = (z^n + \alpha_{n-1} z^{n-1} + \cdots + \alpha_0)/q_{n-1}(z)$ (see, e.g., [24, Lemma 4.5]).

3 Automated pole selection

Note that the rational nodal function s_n of (2.8) is explicitly known in the n-th iteration of the rational Arnoldi method: it has poles $\xi_1, \ldots, \xi_{n-1}$, and its zeros are the rational Ritz values $\Lambda(A_n)$. The aim of an automated pole selection strategy is, of course, to achieve a smallest possible (bound for the) approximation error $\|f(A)\boldsymbol{v} - \boldsymbol{f}_n\|$ at every iteration of the rational Arnoldi method. In view of (2.9) we will therefore try to make $|s_n(x)|$ uniformly large on Γ (the support of the measure γ in (1.1)) by choosing the next pole $\xi_n \in \Gamma$ such that

$$\left|s_n(\xi_n)\right| = \min_{x \in \Gamma}\left|s_n(x)\right|.$$

This choice is inspired by the pole selection strategy proposed in [18, 19], where the nodal function has to be large on a negative real interval Γ and small on $-\Gamma$ (the spectral interval of a symmetric matrix). In our case we do not necessarily have such symmetry, but still we can achieve that our nodal rational function s_n is large on Γ. Recall from above that the term $\|s_n(A)\boldsymbol{v}\|$ in (2.9) is guaranteed to be minimal among all rational functions with the prescribed poles. This justifies our strategy to minimize explicitly only the second factor on the right-hand side of (2.9). In Algorithm 1 we summarize our rational Arnoldi method with automated pole selection, see also [26].

Algorithm 1 Rational Arnoldi method for $f(A)\boldsymbol{v}$ with automated pole selection

Input: Function f, associated set Γ (see (1.1)), $A \in \mathbb{C}^{N \times N}$, $\boldsymbol{v} \in \mathbb{C}^N \setminus \{\boldsymbol{0}\}$,
and translation parameter σ if necessary (i.e., when $0 \in \Gamma$, see Remark 2.1).
Output: Rational Arnoldi approximations $\boldsymbol{f}_j$ and pole sequence $\xi_1, \xi_2, \ldots$

1. Set $\tilde{A} := A - \sigma I$, $q_0(z) := 1$, and $\boldsymbol{v}_1 := \boldsymbol{v}/\|\boldsymbol{v}\|$.
2. For $j = 1, 2, \ldots$
3.   Compute $\tilde{\boldsymbol{w}}_{j+1} := \tilde{A}\boldsymbol{v}_j$.
4.   Compute $h_{i,j} := \boldsymbol{v}_i^* \tilde{\boldsymbol{w}}_{j+1}$ $(i = 1, \ldots, j)$.
5.   Set $H_j := [h_{i,j}] \in \mathbb{C}^{j \times j}$ and $K_j := I_j + H_j \mathrm{diag}(\tilde{\xi}_1^{-1}, \ldots, \tilde{\xi}_{j-1}^{-1}, 0) \in \mathbb{C}^{j \times j}$.
6.   Compute $A_j := H_j K_j^{-1} + \sigma I_j$.
7.   Compute j-th order Arnoldi approximation $\boldsymbol{f}_j := V_j f(A_j) V_j^* \boldsymbol{v} \in \mathcal{Q}_j(A, \boldsymbol{v})$.
8.   Compute an error estimate for $\boldsymbol{f}_j$ (see Sect. 4) and halt if a stopping criterium is satisfied.
9.   Compute $\Lambda(A_j) = \{\theta_1, \ldots, \theta_j\}$ and set

$$\xi_j := \arg\min_{z \in \Gamma}\left|s_j(z)\right| = \arg\min_{z \in \Gamma}\left|\prod_{k=1}^{j}(z - \theta_k)/q_{j-1}(z)\right|.$$

10.   Set $q_j(z) := (z - \xi_j)q_{j-1}(z)$.
11.   Set $\tilde{\xi}_j := \xi_j - \sigma$ and compute $\boldsymbol{w}_{j+1} := (I - \tilde{A}/\tilde{\xi}_j)^{-1}\tilde{\boldsymbol{w}}_{j+1}$.
12.   Redefine $h_{i,j} := \boldsymbol{v}_i^* \boldsymbol{w}_{j+1}$ $(i = 1, \ldots, j)$ and compute $\hat{\boldsymbol{v}}_{j+1} := \boldsymbol{w}_{j+1} - \sum_{i=1}^{j} \boldsymbol{v}_i h_{i,j}$.
13.   Compute $h_{j+1,j} := \|\hat{\boldsymbol{v}}_{j+1}\|$ and $\boldsymbol{v}_{j+1} := \hat{\boldsymbol{v}}_{j+1}/h_{j+1,j}$.

Remark 3.1 Note that, in contrast to the algorithms presented in [18, 19], Algorithm 1 does not require any estimation for the spectral region of A (except for a very rough estimate of $\|A\|$ to choose the parameter σ, see Remark 2.1). In fact, we will demonstrate in Sects. 5 and 6 that our algorithm also performs well for highly nonsymmetric and nonnormal matrices.

Remark 3.2 In a practical implementation of Algorithm 1 one can use an a-posteriori error estimator in Step 8 to halt the iteration if $\|f(A)\boldsymbol{v} - \boldsymbol{f}_j\|$ is below some tolerance. We will discuss such error estimators in Sect. 4.

Remark 3.3 Typically, the cost for orthogonalization of the Krylov basis vectors, equations (2.3) and (2.4), is negligible compared with the cost of solving the shifted linear systems with A. Step 3 of Algorithm 1 corresponds to a polynomial Krylov step (i.e., $\tilde{\xi}_j = \infty$) and allows for the cheap computation of the projection matrix $A_j = V_j^* A V_j$ in Step 6 using the relation (2.7). Overall two orthogonalizations are required in one iteration of the algorithm. However, techniques for reducing the number of inner products in the rational Arnoldi algorithm using so-called *auxiliary vectors* could be employed [24, Chap. 6]. Reorthogonalization techniques may be employed as well, however, we have not found this necessary in the numerical experiments presented in this paper.

4 Error criteria

In this section we derive some practical estimates for the approximation error $\|f(A)\boldsymbol{v} - \boldsymbol{f}_n\|$. Some of these techniques were adopted from [14, 24, 31], where they were used for other problems: computation of some non-Markov matrix functions and solution of matrix equations. This section also contains a comparison of these error estimators in Fig. 1 for a simple test matrix.

4.1 The difference of iterates

Given some delay integer d, by the triangle inequality $\|f(A)\boldsymbol{v} - \boldsymbol{f}_n\| \leq \|f(A)\boldsymbol{v} - \boldsymbol{f}_{n+d}\| + \|\boldsymbol{f}_{n+d} - \boldsymbol{f}_n\|$. Under the assumption that the rational Arnoldi method converges sufficiently fast so that $\|f(A)\boldsymbol{v} - \boldsymbol{f}_{n+d}\|$ is relatively small in comparison to $\|f(A)\boldsymbol{v} - \boldsymbol{f}_n\|$, a primitive estimator for the approximation error is

$$\left\| f(A)\boldsymbol{v} - \boldsymbol{f}_n \right\| \approx \|\boldsymbol{f}_{n+d} - \boldsymbol{f}_n\|. \tag{4.1}$$

Note that the Euclidian norm of the difference of two iterates can be computed cheaply using only their coordinates in the orthonormal basis V_{n+d} and V_n, without forming the long Arnoldi approximation vectors $\boldsymbol{f}_{n+d}$ and $\boldsymbol{f}_n$. We warn that this estimator may be too optimistic, in particular, when the approximations $\boldsymbol{f}_n$ (almost) stagnate for d iterations or more. This underestimation of the error can be seen in Fig. 1.
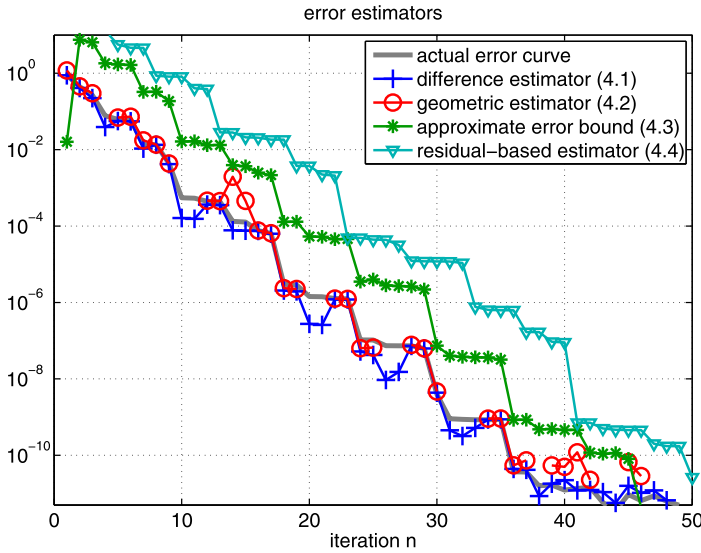
Fig. 1 Comparison of the error estimators from Sect. 4, given in (4.1), (4.2), (4.3), and (4.4) for a diagonal matrix whose eigenvalues are 10^4 Chebyshev points in $[10^{-3}, 10^3]$. The delay integer for the first two error estimators (4.1) and (4.2) is chosen as $d = 2$. Note that the corresponding curves (*blue with crosses* and *red with circles*) partly overlay the actual error curve (*solid gray*). The integrals involved in the other two estimators have been approximated by adaptive Gauss–Kronrod quadrature (Color figure online)

4.2 Exploiting geometric convergence

Another estimate can be derived by assuming that the error is decaying approximately geometrically (as is typically the case, see Sect. 5). A similar estimator has been successfully applied in [14, 31]. One starts by assuming the ideal equalities

$$\|f_n - f_{n+d}\| = cR^{-n} \quad \text{and} \quad \|f_{n+d} - f_{n+2d}\| = cR^{-(n+d)},$$

and defining $\chi_j = \log \|f_{j+d} - f_j\|$. Moreover, one defines

$$R = \exp\left(\frac{\chi_n - \chi_{n+d}}{d}\right) \quad \text{and} \quad c = \exp\left(\frac{(n+d)\chi_n - n\chi_{n+d}}{d}\right).$$

From

$$\|f(A)v - f_n\| \leq \|f_n - f_{n+d}\| + \|f_{n+d} - f_{n+2d}\| + \cdots$$
$$\approx cR^{-n}\left(1 + R^{-d} + R^{-2d} + \cdots\right)$$

we obtain the estimator

$$\|f(A)v - f_n\| \approx \frac{cR^{-n}}{1 - R^{-d}}. \tag{4.2}$$

The last equality is only valid if $R > 1$. In practice it may happen that $R \leq 1$, in which case this estimator becomes negative or infinite. This typically indicates an

error increase or stagnation in the iterates, and one should iterate further to reobtain a reliable estimator. This effect can be seen in Fig. 1, where the curve corresponding to this error indicator shows several "gaps".

4.3 Approximate error bound

An approximate bound for the approximation error can be derived by replacing A with A_n in the integrand of (2.9), yielding

$$\left\| f(A)\boldsymbol{v} - r_n(A)\boldsymbol{v} \right\| \lesssim \left\| s_n(A)\boldsymbol{v} \right\| \cdot \left\| \int_\Gamma \frac{(xI_n - A_n)^{-1}}{s_n(x)} \, d\gamma(x) \right\|. \qquad (4.3)$$

If A_n is diagonalizable, then this new integral can be approximated by scalar quadrature. We still need to compute $\|s_n(A)\boldsymbol{v}\|$. Note that $s_n(A)\boldsymbol{v}$ is just a scalar multiple of $\boldsymbol{v}_{n+1}$, say $s_n(A)\boldsymbol{v} = \delta_n \boldsymbol{v}_{n+1}$. A simple trick to get a hand on this scalar δ_n is to run the rational Arnoldi algorithm with a modified matrix and starting vector

$$\hat{A} = \begin{bmatrix} A & \\ & \tau \end{bmatrix}, \qquad \hat{\boldsymbol{v}} = \begin{bmatrix} \boldsymbol{v} \\ 1 \end{bmatrix},$$

where $\tau \in \mathbb{C}$ is away from Γ and all Ritz values, still performing all inner products only on the first N components (and hence not changing the orthogonality of V_n). We therefore have

$$s_n(\hat{A})\hat{\boldsymbol{v}} = \begin{bmatrix} s_n(A)\boldsymbol{v} \\ s_n(\tau) \end{bmatrix} = \delta_n \hat{\boldsymbol{v}}_{n+1}.$$

Comparing the last (that is, $(N + 1)$-st) element of $\hat{\boldsymbol{v}}_{n+1}$ with $s_n(\tau)$, we obtain the desired scaling factor $|\delta_n| = \|s_n(A)\boldsymbol{v}\|$ as

$$\delta_n = s_n(\tau)/[\hat{\boldsymbol{v}}_{n+1}]_{N+1}.$$

The behavior of the resulting approximate upper bound is shown in Fig. 1. Typically, this approximate bound cannot be trusted for small iteration numbers n, but it becomes more reliable in later iterations when more spectral information about A has been captured in A_n.

4.4 Residual-based estimator

As explained in Sect. 2, the matrix $A_n = V_n^* A V_n$ required for the Arnoldi approximation (2.2) can be computed via $A_n = H_n K_n^{-1} + \sigma I_n$ without explicit projection. This allows us to use a shifted version of the decomposition (2.6) in the form

$$(\tilde{x}I - A)V_n = V_n(\tilde{x}I - A_n) - \boldsymbol{v}_{n+1} h_{n+1,n} \boldsymbol{e}_n^T K_n^{-1},$$

where $\tilde{x} = x + \sigma$ for an arbitrary $x \in \mathbb{C}$. Let us consider a shifted linear system $(\tilde{x}I - A)\boldsymbol{x}(\tilde{x}) = \boldsymbol{v}$ and the corresponding rational Arnoldi approximation $\boldsymbol{x}_n(\tilde{x}) = V_n(\tilde{x}I_n - A_n)^{-1}V_n^*\boldsymbol{v}$. The residual of this approximation satisfies

$$
\begin{aligned}
\boldsymbol{v} - (\tilde{x}I - A)\boldsymbol{x}_n(\tilde{x}) &= \boldsymbol{v} - (\tilde{x}I - A)V_n(\tilde{x}I_n - A_n)^{-1}V_n^*\boldsymbol{v} \\
&= \boldsymbol{v} - \left[V_n(\tilde{x}I - A_n) - \boldsymbol{v}_{n+1}h_{n+1,n}\boldsymbol{e}_n^T K_n^{-1}\right](\tilde{x}I_n - A_n)^{-1}V_n^*\boldsymbol{v} \\
&= \boldsymbol{v}_{n+1}h_{n+1,n}\boldsymbol{e}_n^T K_n^{-1}(\tilde{x}I_n - A_n)^{-1}V_n^*\boldsymbol{v}.
\end{aligned}
$$

Using the fact that $V_n^*\boldsymbol{v} = \|\boldsymbol{v}\|\boldsymbol{e}_1$ by construction of the rational Arnoldi algorithm, and $K_n^{-1}(\tilde{x}I_n - A_n)^{-1} = (xK_n - H_n)^{-1}$, we obtain

$$
\left\|\boldsymbol{v} - (\tilde{x}I - A)\boldsymbol{x}_n(\tilde{x})\right\| = h_{n+1,n}\|\boldsymbol{v}\| \cdot \left|\boldsymbol{e}_n^T(xK_n - H_n)^{-1}\boldsymbol{e}_1\right|.
$$

This allows for the definition of a "residual" of a Cauchy–Stieltjes matrix function

$$
\mathrm{residual}(f, n) := \boldsymbol{v}_{n+1}h_{n+1,n}\int_{\Gamma-\sigma}\boldsymbol{e}_n^T(xK_n - H_n)^{-1}\boldsymbol{e}_1\,\mathrm{d}\gamma(x),
$$

whose norm is given by

$$
\left\|\mathrm{residual}(f, n)\right\| = h_{n+1,n}\|\boldsymbol{v}\| \cdot \left\|\int_{\Gamma-\sigma}\boldsymbol{e}_n^T(xK_n - H_n)^{-1}\boldsymbol{e}_1\,\mathrm{d}\gamma(x)\right\|. \tag{4.4}
$$

See also [10, 13, 29, 37] for related constructions. In our numerical experiments this appeared to be a good indicator, being almost proportional to the actual error; see again Fig. 1.

5 Convergence studies

A thorough convergence analysis of our algorithm appears to be complicated by the interaction between the Ritz values $\Lambda(A_n)$ (which vary in each iteration) and the selected poles $\{\xi_j\}$. Although there is hope of characterizing these two sets asymptotically as equilibrium charges on a condenser (at least in the case of a symmetric matrix A; see our Remark 5.1), we decided to present here a numerical comparison of our method with competing approaches for computing approximations for $f(A)\boldsymbol{v}$. Our comparison is two-fold. In Sect. 5.1 we compare our method with two other methods, both of which use asymptotically optimal poles computed by assuming knowledge of the spectral properties of A. In Sect. 5.2 we then compare our method to well-established Krylov methods with prescribed pole sequences independent of A, namely the polynomial and extended Krylov subspace methods.

5.1 Comparison with asymptotically optimal pole sequences

Our algorithm can be seen as a strategy for constructing the nodal function s_n of (2.8) such that this function is large on Γ and small on the numerical range $\mathbb{W}(A)$. The numerical range $\mathbb{W}(A) := \{\boldsymbol{x}^*A\boldsymbol{x} : \|\boldsymbol{x}\| = 1\}$ is a convenient set for bounding the norm $\|s_n(A)\|$: by a theorem of Crouzeix [12] we have

$$
\left\|s_n(A)\right\| \leq 11.08 \max_{z\in\mathbb{W}(A)}\left|s_n(z)\right|. \tag{5.1}
$$

Unfortunately, bounds based on the numerical range may be crude, in which case it is not clear on which set $\Sigma \subset \mathbb{C}$ the function s_n actually needs to be small so that $\|s_n(A)\|$ is guaranteed to be small. Although the convergence bounds below are given in terms of $\Sigma = \mathbb{W}(A)$, the reader should keep in mind that possibly a smaller set Σ may be relevant for the actual convergence of the Krylov methods under consideration. For example, adaptation of polynomial Krylov methods to the operator spectrum is treated in [30], or in [6] for the rational Krylov case, and improved (theoretical) bounds for $\|s_n(A)\|$ may be obtainable by considering sets $\Sigma = \Sigma_n$ that shrink as the iteration progresses (see, e.g., [6]), or by considering pseudospectra of A instead of the numerical range (see, e.g., [38]).

Assuming that $\Sigma = \mathbb{W}(A)$ and Σ is disjoint from Γ, we may compare the performance of our automated pole selection strategy with that of explicit selection of (asymptotically) optimal poles ξ_j. One choice of such poles is so-called *generalized Leja points* (or *Leja–Bagby points*, see [5]), which are constructed as follows: Starting with a point $\sigma_1 \in \Sigma$ such that $\max_{z \in \Sigma} |z - \sigma_1|$ is minimal, the points $\sigma_{j+1} \in \Sigma$ and $\xi_j \in \Gamma$ are determined recursively such that with the nodal function

$$s_j(z) = \frac{\prod_{i=1}^{j} (z - \sigma_i)}{\prod_{i=1}^{j-1} (z - \xi_i)}$$

the conditions

$$\max_{z \in \Sigma} |s_j(z)| = |s_j(\sigma_{j+1})| \quad \text{and} \quad \min_{z \in \Gamma} |s_j(z)| = |s_j(\xi_{j+1})|$$

are satisfied. Note that the function s_j defined here would agree with the nodal function defined in (2.8) at iteration j of the rational Arnoldi method if all the σ_i were to coincide with Ritz values θ_i, and all the poles ξ_i were the same. Results from logarithmic potential theory [23, 33] assert that there exists a positive number $\mathrm{cap}(\Sigma, \Gamma)$, called the *condenser capacity*, such that

$$\limsup_{n \to \infty} \left(\frac{\max_{z \in \Sigma} |s_n(z)|}{\min_{z \in \Gamma} |s_n(z)|} \right)^{1/n} = e^{-1/\mathrm{cap}(\Sigma, \Gamma)}.$$

Determining the capacity of an arbitrary condenser (Σ, Γ) is a nontrivial problem. The situation simplifies if both Σ and Γ are simply connected sets (and not single points): then by the *Riemann mapping theorem* (cf. [27, Theorem 5.10h]) there exists a bijective function Φ that conformally maps the complement of $\Sigma \cup \Gamma$ onto a circular annulus $\mathbb{A}_R := \{w : 1 < |w| < R\}$. The number R is called the *Riemann modulus of* $\mathbb{A}_R$ and it satisfies

$$R^{-1} = e^{-1/\mathrm{cap}(\Sigma, \Gamma)}.$$

To relate the asymptotic behavior of s_n to that of the error $\|f(A)\boldsymbol{v} - r_n(A)\boldsymbol{v}\|$ we use (2.9) and (5.1), and obtain

$$\limsup_{n \to \infty} \|f(A)\boldsymbol{v} - r_n(A)\boldsymbol{v}\|^{1/n} \leq R^{-1}.$$

In the following examples we demonstrate that our adaptive rational Arnoldi method, Algorithm 1, (asymptotically) converges at least with rate R^{-1}, i.e., not slower than a rational Krylov method with asymptotically optimal poles would converge if the set Σ were known a priori. To this end, we numerically compare our method with two reference methods, both of which are known to converge asymptotically at least with rate R^{-1}.

The first reference method is the so-called *PAIN (poles and interpolation nodes) method*, which is a two-term recurrence described in [24]

$$\boldsymbol{v}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|,$$

$$\beta_j \boldsymbol{v}_{j+1} = (I - A/\xi_j)^{-1}(A - \sigma_j I)\boldsymbol{v}_j, \quad j = 1, \ldots, n,$$

where the numbers β_j are chosen to normalize the vectors $\boldsymbol{v}_{j+1}$, and σ_j and ξ_j are the generalized Leja points for the condenser (Σ, Γ). The corresponding PAIN approximation is defined as

$$\boldsymbol{f}_n^{(P)} := [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n] f\left(R_n L_n^{-1}\right) \|\boldsymbol{v}\| \boldsymbol{e}_1,$$

where $\boldsymbol{e}_1 \in \mathbb{R}^n$ denotes the first unit coordinate vector,

$$L_n = \begin{bmatrix} 1 & & & \\ \beta_1/\xi_1 & 1 & & \\ & \ddots & \ddots & \\ & & \beta_{n-1}/\xi_{n-1} & 1 \end{bmatrix} \quad \text{and} \quad R_n = \begin{bmatrix} \sigma_1 & & & \\ \beta_1 & \sigma_2 & & \\ & \ddots & \ddots & \\ & & \beta_{n-1} & \sigma_n \end{bmatrix}.$$

It can be shown that $\boldsymbol{f}_n^{(P)} = r_n^{(P)}(A)\boldsymbol{v}$, where $r_n^{(P)}$ is the rational interpolant for f with prescribed poles $\xi_1, \ldots, \xi_{n-1}$ and interpolation nodes $\sigma_1, \ldots, \sigma_n$ [24]. Note that the PAIN method is not spectrally adaptive: both the poles and the interpolation nodes are chosen a priori and no discrete spectral information about A is taken into account.

The second reference method is the rational Arnoldi method where the poles ξ_j are chosen a priori as generalized Leja points, and we will refer to this as the *standard rational Arnoldi method* in the following. We denote the corresponding approximations as $\boldsymbol{f}_n^{(S)}$. Note that the standard rational Arnoldi method chooses the interpolation nodes spectrally adaptive as Ritz values associated with the rational Krylov space. It is therefore an adaptive method for the interpolation nodes, but still the poles are chosen a priori. The methods under consideration are summarized in Table 1.

Table 1 Overview of the methods to be compared, all of which compute approximations to $f(A)\boldsymbol{v}$ of the form $r_n(A)\boldsymbol{v}$, where r_n is a rational interpolating function for f

Method	Interpolation nodes for r_n	Poles of r_n
PAIN $\boldsymbol{f}_n^{(P)}$	generalized Leja	generalized Leja
standard rational Arnoldi $\boldsymbol{f}_n^{(S)}$	adaptive (Ritz values)	generalized Leja
adaptive rational Arnoldi $\boldsymbol{f}_n$	adaptive (Ritz values)	adaptive

The interval case   Let A be a symmetric matrix with $\Sigma = \mathbb{W}(A) = [a, b]$ being a positive spectral interval. Moreover, let f be a Markov function (1.1) whose generating measure γ is supported on $\Gamma = (-\infty, 0]$. Then the conformal map Φ that carries the complement of $\Gamma \cup \Sigma = (-\infty, 0] \cup [a, b]$ onto the annulus $\mathbb{A}_R$ can be given explicitly in terms of elliptic functions. In particular, the Riemann modulus R is given as (see [23, § 3])

$$R = \exp\left(\frac{\pi}{2} \frac{K(\sqrt{1 - \kappa^2})}{K(\kappa)}\right), \quad \text{where } \kappa = \frac{\sqrt{b/a} - 1}{\sqrt{b/a} + 1} \tag{5.2}$$

and

$$K(\kappa) = \int_0^1 \frac{1}{\sqrt{(1 - t^2)(1 - \kappa^2 t^2)}} \, dt \tag{5.3}$$

is the complete elliptic integral of the first kind (see [4]).[1]

Let $f(z) = z^{-1/2}$ without further mention in this Sect. 5.1. We first consider a diagonal matrix A_1 with $N = 10^4$ eigenvalues being scaled and shifted Chebyshev points of the second kind,

$$\lambda_j = a + \frac{\cos(\pi j/(N-1)) + 1}{2}(b - a), \quad j = 0, 1, \ldots, N-1,$$

in the interval $[a, b] = [10^{-3}, 10^3]$, and a vector $\boldsymbol{v}$ whose entries are normally distributed pseudo-random numbers. In Fig. 2 (top left) we show the convergence of our adaptive rational Arnoldi method, Algorithm 1, in comparison with the two reference methods (the PAIN method and standard rational Arnoldi, cf. Table 1). The theoretical convergence rate R^{-1} from (5.2) is indicated by the slope of the dashed line. Note that all the three methods converge almost linearly with the predicted rate. The reason for the rational Arnoldi methods behaving like this is that the Chebyshev eigenvalues are denser at the endpoints of the spectral interval, and almost no spectral adaption takes place during the first 50 iterations shown here. In some sense, the rational Krylov methods behave initially as if the spectrum were not discrete; see [6] for a potential theoretic explanation. We expect our adaptive method to choose roughly the same poles as were chosen in the generalized Leja case, and the plot below confirms this expectation by depicting the (smoothed) empirical distribution functions of the first 50 adaptive poles and generalized Leja poles; the two distributions are visually hard to distinguish.

We next consider a diagonal matrix A_2 with $N = 10^4$ equispaced eigenvalues

$$\lambda_j = a + j(b - a)/(N - 1), \quad j = 0, 1, \ldots, N-1,$$

in the interval $[a, b] = [10^{-3}, 10^3]$. In Fig. 2 (top right) we again show the convergence of the three methods. While the PAIN method still converges linearly with rate R given by (5.2), the standard rational Arnoldi method is somewhat faster because

---

[1] The definition of $K(\kappa)$ is not consistent in the literature. We stick to the definition used in [34, Chap. VI]. In MATLAB one would type `ellipke(kappa^2)` to obtain the value $K(\kappa)$.
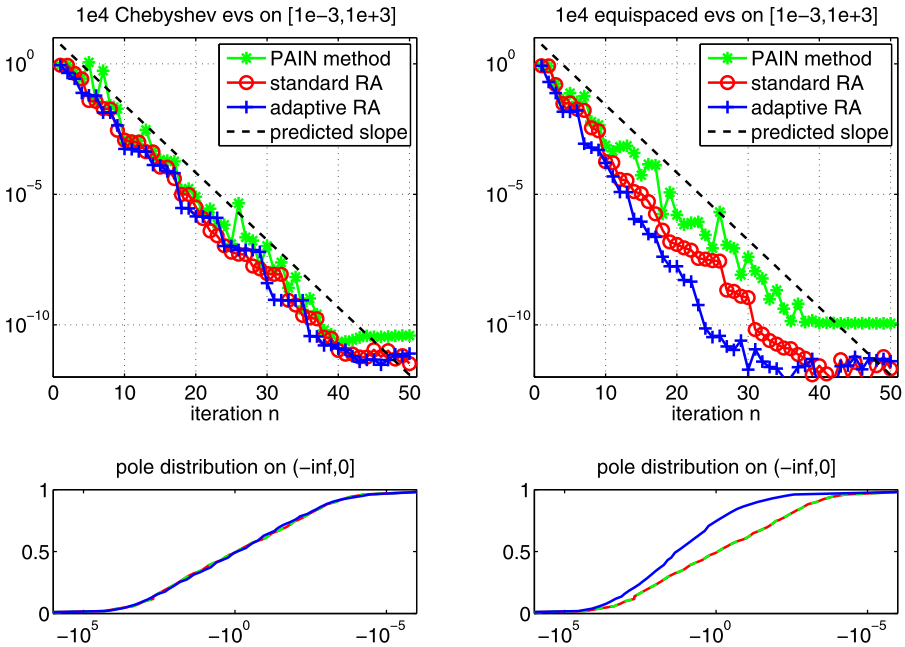
Fig. 2 *Left*: Convergence curves (*top*) and distribution of poles (*below*) when approximating $A_1^{-1/2} v$ with different rational Krylov methods. *Right*: Similar plots for the matrix A_2

the interpolation nodes (Ritz values) "deflate" some of the left-most eigenvalues of A in early iterations, causing a superlinear convergence speedup (see [6] for an analysis of this effect). The adaptive rational Arnoldi method converges even faster than standard rational Arnoldi, because the poles of the rational Krylov space are selected by taking into account the deflation of left-most eigenvalues. The plot of the pole distribution functions below reveals that the adaptive method has the tendency to place the poles ξ_j somewhat farther away from the origin.

Union of intervals In Fig. 3 (left) we consider a diagonal matrix A_3 whose spectrum is the union of 10 Chebyshev points on the interval $[10^{-3}, 10^{-1}]$ and 9990 Chebyshev points on $[10^1, 10^3]$. Note that the PAIN method with poles optimized for the spectral interval $[10^{-3}, 10^3]$ converges linearly. However, our adaptive method changes its slope after a few iterations to converge linearly as if the spectral interval were $[10^1, 10^3]$. Both slopes are depicted in this figure. The spectral adaption becomes also visible in the pole distribution function (Fig. 3, bottom left).

Remark 5.1 In view of the behavior of our adaptive rational Arnoldi method for the above symmetric matrices, we believe that the convergence can be asymptotically (that is, for a sequence of symmetric matrices growing larger in size and having a joint limit eigenvalue distribution) compared to min-max rational functions with poles on Γ and zeros on Σ being constrained Leja points in the sense of [11]. The
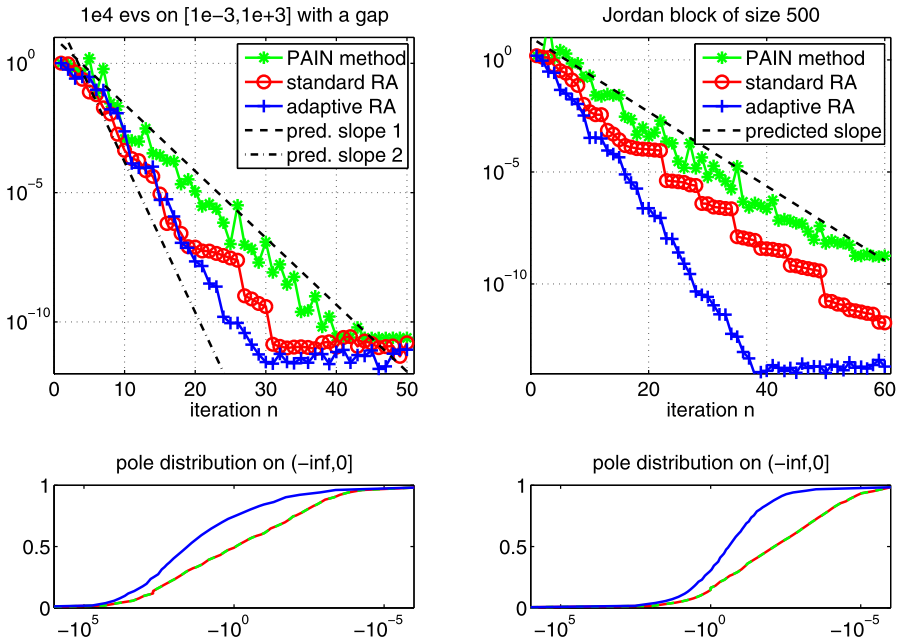
Fig. 3 *Left*: Convergence curves (*top*) and distribution of poles (*below*) when approximating $A_3^{-1/2}v$ with different rational Krylov methods. *Right*: Similar plots for the matrix A_4

constraint for the zeros is given by the interlacing property of Ritz values associated with symmetric matrices (see, e.g., [7]).

A Jordan block   The matrix A_4 is a single Jordan block

$$A_4 = \begin{bmatrix} 1 & 1 & & \\ & 1 & \ddots & \\ & & \ddots & 1 \\ & & & 1 \end{bmatrix} \in \mathbb{C}^{N \times N},$$

and its numerical range is a disk $\mathbb{W}(A_4) = \{z : |z - 1| \le r\}$ with radius $r = \cos(\pi/(N+1))$. The conformal mapping of the complement of $\mathbb{W}(A_4)$ in the slit complex plane onto an annulus is known in terms of elliptic functions (see [34, p. 293–294]), and the Riemann modulus of this domain is

$$R = \exp\left(\frac{\pi}{4} \frac{K(\sqrt{1 - \kappa^2})}{K(\kappa)}\right), \quad \text{where } \kappa = \left(\frac{c}{r} - \sqrt{\frac{c^2}{r^2} - 1}\right)^2,$$

with $K(\kappa)$ defined in (5.3). The resulting convergence of the three methods is shown in Fig. 3 (right), together with the theoretical rate R^{-1}. Note that the PAIN method converges as predicted from the bound involving the numerical range, whereas the

two rational Arnoldi methods converge faster due to spectral adaption. In particular, our adaptive method converges significantly faster.

5.2 Comparison with fixed pole sequences

In this section we will briefly discuss polynomial and rational Krylov methods with poles prescribed independently of Σ, and therefore not leading to the optimal convergence rate associated with the condenser capacity $\mathrm{cap}(\Sigma, \Gamma)$.

The simplest of these methods is the *polynomial Arnoldi method*, which is the special case of rational Arnoldi in which all poles ξ_j are set to infinity. This method has the obvious advantage that no linear system solves are required. If A is Hermitian and we consider the approximation of functions with generating measure supported on $\Gamma = (-\infty, 0]$, such as $f(z) = z^{-1/2}$, then the convergence rate of the polynomial Arnoldi method equals that of the CG method, i.e.,

$$\|f(A)\boldsymbol{v} - \boldsymbol{f}_n\| \le C\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n \lesssim C \cdot \exp\left(-\frac{2n}{\sqrt{\kappa}}\right), \quad \kappa = \frac{\lambda_{\max}}{\lambda_{\min}},$$

where the approximate inequality is valid for large condition numbers κ. Obviously, convergence can be slow if the condition number κ gets large, and therefore many Krylov iterations will be required to approximate $f(A)\boldsymbol{v}$ to a prescribed accuracy. Note that Arnoldi (and also Lanczos) methods for matrix functions require the Krylov basis V_n to be stored for the final computation of the Arnoldi approximation $\boldsymbol{f}_n$ of (2.2), which renders this method impractical if n is large. Although restarted variants of the polynomial Arnoldi method for $f(A)\boldsymbol{v}$ have been proposed, which prevent the dimension of the Krylov space to grow above memory limit (see [1, 20, 21]), the use of finite poles ξ_j typically is a worthwhile alternative if linear systems with shifted versions of A can be solved efficiently.

If the poles alternate between $\xi_{2j} = \infty$ and $\xi_{2j+1} = 0$, we obtain the so-called *extended Krylov subspace method* with convergence (see [15] and [31, Theorem 3.4])

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

A computational advantage of the extended Krylov subspace method is that only the actions of A and A^{-1} on vectors are required. In particular, if a direct solver is applicable, only one factorization of A needs to be computed. The convergence of the polynomial and extended Krylov subspace methods is illustrated in Fig. 4, and compared with that of our adaptive rational Arnoldi method. In this figure, $f(z) = z^{-1/2}$ and A is the finite-difference discretization of the negative 2D Laplacian with 100 discretization points in each coordinate direction (i.e., $N = 100^2$). Note that the predicted convergence rate for the extended Krylov subspace method is only observable in the first few iterations because superlinear convergence effects take place when some rational Ritz values start converging to the left-most eigenvalues of A (which are close to the poles at 0, see [6] for an explanation).

Our last test is more challenging: We consider the computation of the logarithm $\log(A)\boldsymbol{v}$ of a random diagonalizable matrix $A \in \mathbb{C}^{200 \times 200}$ having eigenvalues in the
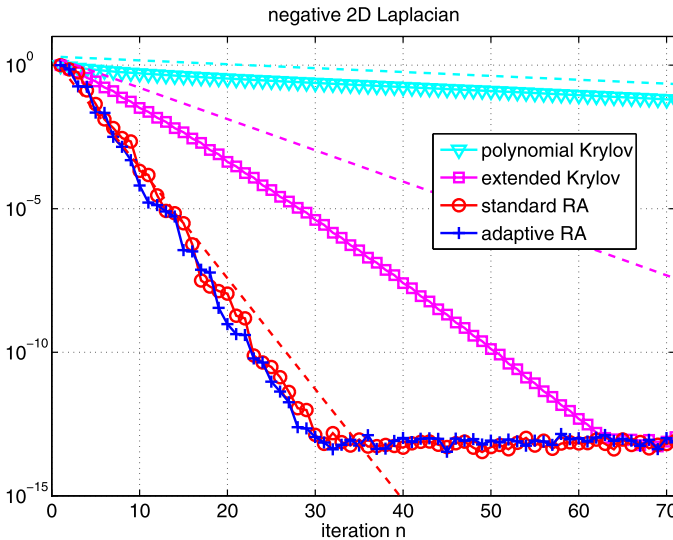
Fig. 4 Comparison of polynomial and rational Arnoldi methods for the approximation of $f(A)\boldsymbol{v} = A^{-1/2}\boldsymbol{v}$ for the negative 2D Laplacian. The *dashed lines* indicate the expected linear convergence slopes
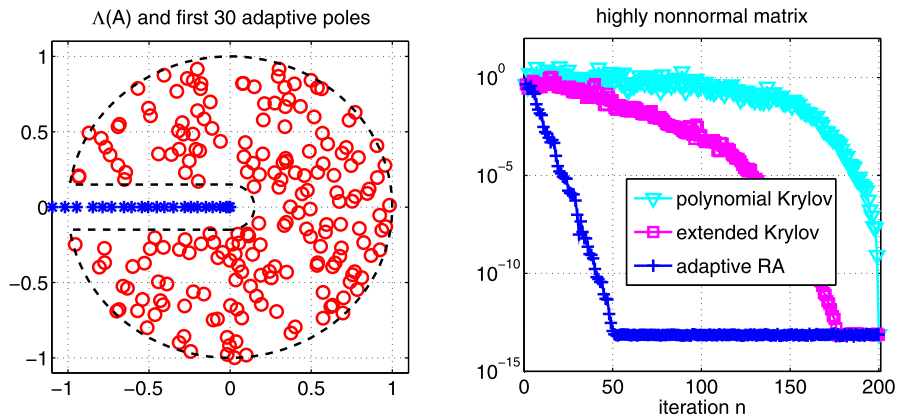

Fig. 5 *Left*: Eigenvalues of a highly nonnormal random matrix $A \in \mathbb{C}^{200 \times 200}$. *Right*: Convergence of rational Arnoldi for $f(A)\boldsymbol{v} = \log(A)\boldsymbol{v}$, random vector $\boldsymbol{v}$, with adaptive poles on $\Gamma = (-\infty, 0]$

unit disk under the constraint that the distance of each eigenvalue to $\Gamma = (-\infty, 0]$ is at least 0.1. The eigenvalues of this matrix are shown in Fig. 5 (left). We remark that A is highly nonnormal: although the moduli of its eigenvalues are nicely bounded above and below, it has a condition number of $\approx 2.5 \times 10^4$. To our best knowledge, no existing convergence theory is able to explain why Algorithm 1 converges so robustly even for this matrix (see Fig. 5, right). Note that the usual arguments involving the numerical range $\mathbb{W}(A)$ fail here, as this set is not even disjoint from Γ. We have failed to implement the extended Krylov subspace method in such a way that it reproduces

the exact solution at least at iteration $n = N = 200$, as theory predicts. This instability is probably caused by the $(n/2)$-fold pole 0 being surrounded by eigenvalues of A and lying inside the numerical range.

6 A large-scale numerical example with inexact solves

The following tests are run on a desktop computer with 3.7 GB of RAM, running an AMD Phenom II X3 705e processor at 2.5 GHz. The software environment is Matlab 7.12.0 (R2011a) under Ubuntu Release 10.04.

We shall consider the problem of computing the impedance function $f(z) = z^{-1/2}$ of a discretization A of the convection–diffusion operator

$$\mathscr{A}u = -\operatorname{div}(a \operatorname{grad} u) + b^T \operatorname{grad} u$$

on $\Omega = [0, 1]^3$. We assume that a is a uniformly positive and bounded function defined on Ω, and $b = (b_1, b_2, b_3)^T$ is a vector function whose components posses the same properties. We have discretized this operator by the standard second-order finite difference scheme with 100 regular interior grid points.

In Fig. 6 (left) we show the convergence behavior of our adaptive method and the extended Krylov subspace method for smooth low-contrast conductivity a_1 and smooth convective field b, namely

$$a_1(x, y, z) = 1 + \exp(x - 2y), \qquad b(x, y, z) = \begin{bmatrix} \sin(x + y) \\ \cos(x + y) \\ \sin(y + z) \end{bmatrix}. \qquad (6.1)$$

In Fig. 6 (right) we show the results for piecewise constant high-contrast conductivity a_2 and the same b as in (6.1),

$$a_2(x, y, z) = \begin{cases} 100, & \text{if } |x| \leq 0.5 \text{ and } |y| \leq 0.6, \\ 1 & \text{otherwise.} \end{cases}$$

The resulting discretization matrices are of size $10^6 \times 10^6$ and are referred to as A_1 and A_2, respectively. All components of the vector $\boldsymbol{v}$ are set to 1.

All shifted linear systems are solved with a relative error tolerance of 10^{-5}, which is sufficiently smaller than our target relative error of 10^{-4} for the approximation $f(A_\ell)\boldsymbol{v}$ ($\ell = 1, 2$). The linear system solver is BICGSTAB preconditioned by ILU(0). This combination works quite well for the shifted linear systems under consideration, as is indicated in Table 2. The errors of the linear system solves are estimated by exploiting the almost geometric convergence of BICGSTAB with the estimator presented in Sect. 4.2 (we chose the delay integer $d = 2$). As indicated in the last column of Table 2, the measured errors are typically below 10^{-5}, or at least of that order. We have also tried other combinations of iterative solvers and preconditioners, such as BICGSTABL, restarted GMRES, GMR, TFQMR, and IDRS(s)[2] in combination with drop-tolerance ILU or Gauss–Seidel preconditioners. The results of these
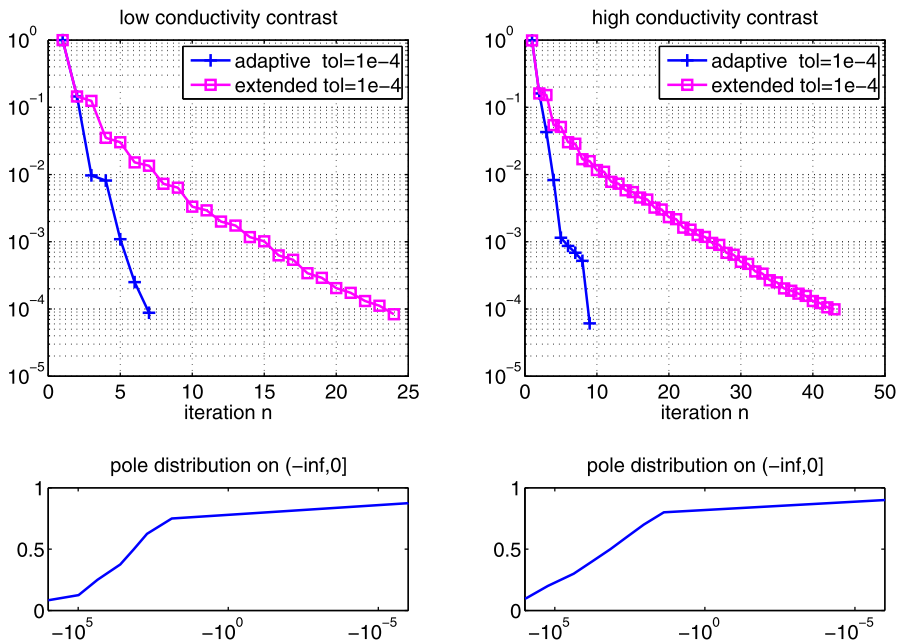
---

[2]As available from http://ta.twi.tudelft.nl/nw/users/gijzen/IDR.html.

Fig. 6 *Left*: Convergence curves (*top*) and distribution of poles (*below*) when approximating $A_1^{-1/2}\boldsymbol{v}$. The iteration is stopped when the relative error is below 10^{-4}. The linear systems involved are solved with a relative error tolerance of 10^{-5}. *Right*: The same plots for the matrix A_2

comparisons are not reported here, but the combination ILU(0) and BICGSTAB consistently outperformed the others. Moreover, BICGSTAB and ILU(0) are parameter-free methods, which is important in our case where we try to develop a black-box method. The initial guess for all linear systems was the vector of all zeroes.

Our adaptive shifts ξ_j are chosen by a greedy search on a discretization of the interval $[-10^6, -10^{-6}]$ with 10^5 logarithmically equispaced points. We have found experimentally that this is a sufficiently fine approximation to the continuous set $\Gamma = (-\infty, 0]$: Taking more discretization points or increasing the width of the search interval did not give any visible improvement in the convergence of our adaptive method. As can be seen in Table 2, the computation time of BICGSTAB clearly dominates that of the ILU(0) factorization, the latter being more or less shift-independent. Note how linear systems with a large shift (in modulus) are typically solved faster than systems with a shift of moderate size. The reason for this observation may be the stronger diagonal dominance of systems with larger shifts, which renders the ILU(0) preconditioner to be more effective, a well-known effect [9].

Our adaptive method clearly outperforms the extended Krylov subspace method in terms of required iterations and computation time. For example, in the case of low-contrast conductivity, the adaptive method requires $n = 7$ iterations, whereas the extended Krylov subspace method requires $n = 24$ iterations (see Fig. 6, left). With the conservative assumption that each shifted linear system solve requires about 30 seconds (see Table 2), our adaptive method requires at least $6 \times 30 = 180$ seconds

Table 2 Solving linear systems with the matrices $A_1 - \xi I$ and $A_2 - \xi I$ using BICGSTAB preconditioned by ILU(0). The BICGSTAB iteration is terminated when our estimator of Sect. 4.2 indicated a relative error below 10^{-5}

Matrix	Shift ξ	ILU(0) time in s	BICGSTAB #iterations	Time in s	Relative error
low-contrast conductivity	$-0e+00$	0.39	85	28.5	$8.72e{-}07$
	$-1e{-}06$	0.42	82	27.5	$1.31e{-}06$
	$-1e{-}04$	0.42	82	27.4	$1.14e{-}06$
	$-1e{-}02$	0.43	83	27.7	$1.84e{-}06$
	$-1e+00$	0.42	78	26.2	$1.43e{-}05$
	$-1e+02$	0.42	55	20.1	$1.12e{-}06$
	$-1e+04$	0.43	11	3.4	$1.26e{-}08$
	$-1e+06$	0.42	6	1.7	$1.33e{-}12$
high-contrast conductivity	$-0e+00$	0.39	71	23.7	$2.90e{-}06$
	$-1e{-}06$	0.42	71	27.3	$2.70e{-}06$
	$-1e{-}04$	0.42	71	26.4	$3.49e{-}06$
	$-1e{-}02$	0.74	71	30.9	$3.43e{-}06$
	$-1e+00$	0.43	70	26.4	$2.86e{-}06$
	$-1e+02$	0.43	65	25.2	$1.19e{-}06$
	$-1e+04$	0.42	35	14.1	$2.44e{-}06$
	$-1e+06$	0.43	9	3.7	$1.74e{-}11$

computation time (the first iteration only utilizes the vector $\boldsymbol{v}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|$ and does not require a linear system solve). The extended Krylov subspace method, on the other hand, requires at least $11 \times 30 = 330$ seconds (only every second iteration of this method requires a linear system solve). Note that we have still neglected the computational costs for orthogonalization and memory management of the long Krylov basis vectors. These costs are larger for the extended Krylov subspace method, because the associated Krylov basis is of higher dimension, but in comparison to the time spent in the BICGSTAB routine these computations are negligible. The gap in iteration numbers between our adaptive method and the extended Krylov subspace method becomes even larger for the example with high conductivity contrast: in this case the methods required 9 versus 43 iterations, respectively (see Fig. 6, right).

We finally remark that the extended Krylov subspace method does not perform well in these examples due to the use of an iterative solver which cannot exploit the fact that only one finite shift $\xi = 0$ appears. The use of direct methods is typically prohibitive for 3D problems. For 2D problems, however, the situation is different and the extended Krylov subspace method in combination with direct solvers may still be competitive with our adaptive method in terms of computation time. In any case, our method tends to require lower-dimensional Krylov subspaces, so that our advantage of lower memory consumption and fewer orthogonalizations still persists.

7 Summary and future work

We have presented a parameter-free rational Arnoldi method for the efficient computation of certain matrix functions $f(A)$ acting on a vector v. We provided numerical evidence that this method converges at least as well as rational Krylov methods using optimal pole sequences constructed with the knowledge of the spectrum. In fact, our new method typically even outperforms such methods due to the spectral adaption of the poles during the iteration. A rigorous convergence analysis, perhaps involving tools from potential theory as in [6, 7], for explaining spectral adaption of this rational Arnoldi variant applied with a symmetric matrix, may be an interesting research problem. The first author also has experienced cases where it seems profitable to take Γ different from the "canonical choice" $(-\infty, 0]$ when approximating the inverse matrix square root, an observation that will be subject of future investigation.

References

1. Afanasjew, M., Eiermann, M., Ernst, O.G., Güttel, S.: Implementation of a restarted Krylov subspace method for the evaluation of matrix functions. Linear Algebra Appl. **429**, 2293–2314 (2008)
2. Allen, E.J., Baglama, J., Boyd, S.K.: Numerical approximation of the product of the square root of a matrix with a vector. Linear Algebra Appl. **310**, 167–181 (2000)
3. Arioli, M., Loghin, D.: Matrix square-root preconditioners for the Steklov–Poincaré operator. Technical Report RAL-TR-2008-003, Rutherford Appleton Laboratory (2008)
4. Abramowitz, M., Stegun, I.A.: Pocketbook of Mathematical Functions. Verlag Harri Deutsch, Frankfurt am Main (1984)
5. Bagby, T.: On interpolation by rational functions. Duke Math. J. **36**, 95–104 (1969)
6. Beckermann, B., Güttel, S.: Superlinear convergence of the rational Arnoldi method for the approximation of matrix functions. Numer. Math. **121**, 205–236 (2012)
7. Beckermann, B., Güttel, S., Vandebril, R.: On the convergence of rational Ritz values. SIAM J. Matrix Anal. Appl. **31**, 1740–1774 (2010)
8. Beckermann, B., Reichel, L.: Error estimation and evaluation of matrix functions via the Faber transform. SIAM J. Numer. Anal. **47**, 3849–3883 (2009)
9. Benzi, M.: Preconditioning techniques for large linear systems: a survey. J. Comput. Phys. **182**, 418–477 (2002)
10. Botchev, M.A.: Residual, restarting and Richardson iteration for the matrix exponential. Memorandum 1928, Department of Applied Mathematics, University of Twente, Enschede (2010)
11. Coroian, D.I., Dragnev, P.: Constrained Leja points and the numerical solution of the constrained energy problem. J. Comput. Appl. Math. **131**, 427–444 (2001)
12. Crouzeix, M.: Numerical range and functional calculus in Hilbert space. J. Funct. Anal. **244**, 668–690 (2007)
13. Druskin, V., Greenbaum, A., Knizhnerman, L.: Using nonorthogonal Lanczos vectors in the computation of matrix functions. SIAM J. Sci. Comput. **19**, 38–54 (1998)
14. Druskin, V., Knizhnerman, L.: Spectral approach to solving three-dimensional Maxwell's diffusion equations in the time and frequency domains. Radio Sci. **29**, 937–953 (1994)
15. Druskin, V., Knizhnerman, L.: Extended Krylov subspaces: approximation of the matrix square root and related functions. SIAM J. Matrix Anal. Appl. **19**, 755–771 (1998)
16. Druskin, V., Knizhnerman, L.: Gaussian spectral rules for the three-point second differences, I: a two-point positive definite problem in a semi-infinite domain. SIAM J. Numer. Anal. **37**, 403–422 (1999)

17. Druskin, V., Knizhnerman, L., Zaslavsky, M.: Solution of large scale evolutionary problems using rational Krylov subspaces with optimized shifts. SIAM J. Sci. Comput. **31**, 3760–3780 (2009)
18. Druskin, V., Lieberman, C., Zaslavsky, M.: On adaptive choice of shifts in rational Krylov subspace reduction of evolutionary problems. SIAM J. Sci. Comput. **32**, 2485–2496 (2010)
19. Druskin, V., Simoncini, V.: Adaptive rational Krylov subspaces for large-scale dynamical systems. Syst. Control Lett. **60**, 546–560 (2011)
20. Eiermann, M., Ernst, O.G.: A restarted Krylov subspace method for the evaluation of matrix functions. SIAM J. Numer. Anal. **44**, 2481–2504 (2006)
21. Eiermann, M., Ernst, O.G., Güttel, S.: Deflated restarting for matrix functions. SIAM J. Matrix Anal. Appl. **32**, 621–641 (2011)
22. van den Eshof, J., Frommer, A., Lippert, T., Schilling, K., van der Vorst, H.A.: Numerical methods for the QCD overlap operator, I: sign-function and error bounds. Comput. Phys. Commun. **146**, 203–224 (2002)
23. Gonchar, A.A.: Zolotarev problems connected with rational functions. Math. USSR Sb. **7**, 623–635 (1969)
24. Güttel, S.: Rational Krylov methods for operator functions. PhD thesis, TU Bergakademie Freiberg (2010)
25. Güttel, S.: Rational Krylov approximation of matrix functions: numerical methods and optimal pole selection. In: GAMM Mitteilungen (2013, accepted for publication)
26. Güttel, S., Knizhnerman, L.: Automated parameter selection for Markov functions. Proc. Appl. Math. Mech. **11**, 15–18 (2011)
27. Henrici, P.: Applied and Computational Complex Analysis, vol. I. Wiley, New York (1988)
28. Higham, N.J.: Functions of Matrices. Theory and Computation. SIAM, Philadelphia (2008)
29. Hochbruck, M., Lubich, C., Selhofer, H.: Exponential integrators for large systems of differential equations. SIAM J. Sci. Comput. **19**, 1552–1574 (1998)
30. Knizhnerman, L.: Adaptation of the Lanczos and Arnoldi methods to the spectrum, or why the two Krylov subspace methods are powerful. Chebyshev Dig. **3**(2), 141–164 (2002)
31. Knizhnerman, L., Simoncini, V.: A new investigation of the extended Krylov subspace method for matrix function evaluations. Numer. Linear Algebra Appl. **17**, 615–638 (2010)
32. Lehoucq, R.B., Meerbergen, K.: Using generalized Cayley transformations within an inexact rational Krylov sequence method. SIAM J. Matrix Anal. Appl. **20**, 131–148 (1998)
33. Levin, A.L., Saff, E.B.: Optimal ray sequences of rational functions connected with the Zolotarev problem. Constr. Approx. **10**, 235–273 (1994)
34. Nehari, Z.: Conformal Mapping. Dover, New York (1975)
35. Ruhe, A.: Rational Krylov sequence methods for eigenvalue computation. Linear Algebra Appl. **58**, 391–405 (1984)
36. Ruhe, A.: Rational Krylov algorithms for nonsymmetric eigenvalue problems. IMA Vol. Math. Appl. **60**, 149–164 (1994)
37. Saad, Y.: Analysis of some Krylov subspace approximations to the matrix exponential operator. SIAM J. Numer. Anal. **29**, 209–228 (1992)
38. Trefethen, L.N.: Pseudospectra of matrices. In: Griffiths, D.F., Watson, G.A. (eds.) Numerical Analysis 1991, pp. 234–266. Longman Scientific and Technical, Harlow (1992)
39. Walsh, J.L.: Interpolation and Approximation by Rational Functions in the Complex Domain. AMS, Providence (1969)