# Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection[*]

**Stefan Güttel**[1,**]

[1] University of Oxford, Mathematical Institute, 24–29 St Giles', Oxford OX1 3LB, UK

Matrix functions are a central topic of linear algebra, and problems of their numerical approximation appear increasingly often in scientific computing. We review various rational Krylov methods for the computation of large-scale matrix functions. Emphasis is put on the rational Arnoldi method and variants thereof, namely, the extended Krylov subspace method and the shift-and-invert Arnoldi method, but we also discuss the nonorthogonal generalized Leja point (or PAIN) method. The issue of optimal pole selection for rational Krylov methods applied for approximating the resolvent and exponential function, and functions of Markov type, is treated in some detail.

## 1 Introduction

An important problem of science and engineering is the efficient computation of the matrix-vector product $f(A)\boldsymbol{b}$, where $A \in \mathbb{C}^{N \times N}$ is a matrix, $\boldsymbol{b} \in \mathbb{C}^N$ is a vector, and $f$ is a function such that the *matrix function* $f(A)$ is defined. A most general definition of $f(A)$ is given by $f(A) := p_{f,A}(A)$, where $p_{f,A}$ is a Birkhoff interpolation polynomial for $f$ at the eigenvalues $\lambda_i \in \Lambda(A)$, counted by their multiplicity in the minimal polynomial of $A$. Hence, $f(A)$ is defined if and only if for every $\lambda_i \in \Lambda(A)$ with multiplicity $\nu_i$ the derivatives $f(\lambda_i), f'(\lambda_i), \ldots, f^{(\nu_i-1)}(\lambda_i)$ exist. If $f$ is analytic in a neighborhood of $\Lambda(A)$, an elegant definition of $f(A)$ is the Cauchy integral formula

$$f(A) = \frac{1}{2\pi i} \int_\Gamma f(\zeta)(\zeta I - A)^{-1}\, \mathrm{d}\zeta,$$

with a contour $\Gamma$ winding around every eigenvalue $\lambda_i$ exactly once. There are other equivalent definitions of $f(A)$, and we refer to the monographs by Golub & Van Loan [1, Chapter 11], Horn & Johnson [2, Chapter 6], and Higham [3], and the review by Frommer & Simoncini [4] for detailed expositions of these. (The last two of these references also include discussions of polynomial Krylov methods for approximating matrix functions.)

Matrix functions are interesting for scientific computing because they arise in explicit solution formulas for relevant algebraic and differential equations. In many applications, the function $f = f^\tau$ to be approximated also depends on a parameter $\tau$, which often corresponds to frequency or time. Here are some examples:

- The solution of a (shifted) linear system $(A - \tau I)\boldsymbol{x} = \boldsymbol{b}$ is given as $\boldsymbol{x}(\tau) = (A - \tau I)^{-1}\boldsymbol{b}$, or equivalently, $\boldsymbol{x}(\tau) = f^\tau(A)\boldsymbol{b}$ with $f^\tau(z) = (z - \tau)^{-1}$. In model order reduction problems in the frequency domain, the parameter $\tau$ typically is a purely imaginary number associated with frequency [5–7][1].

- The solution of the fundamental dynamical system $\boldsymbol{u}'(\tau) + A\boldsymbol{u}(\tau) = \boldsymbol{0}$ can be written in terms of the matrix exponential function $\boldsymbol{u}(\tau) = \exp(-\tau A)\boldsymbol{u}(0)$. This function and variants thereof arise when solving space-discretized evolution problems via exponential integrators [8–14], in network analysis applications [15–19], nuclear magnetic resonance spectroscopy [20], quantum physics [21], and geophysics [22–24].

- The solution of $\boldsymbol{u}''(\tau) = A\boldsymbol{u}(\tau)$ is $\boldsymbol{u}(\tau) = \exp(\tau\sqrt{A})\boldsymbol{c}_1 + \exp(-\tau\sqrt{A})\boldsymbol{c}_2$, with vectors $\boldsymbol{c}_1, \boldsymbol{c}_2$ chosen to match the initial conditions. Such problems arise from the space discretization of time-dependent hyperbolic problems [25].

- Fractional powers $f(z) = z^\alpha$, in particular with $\alpha = \pm 1/2$, arise in the context of (stochastic) differential equations in population dynamics [26], reaction–diffusion problems [27], neutron transport [28, 29], and domain decomposition methods [30].

- The sign function $f(z) = \mathrm{sgn}(z)$, often expressed in terms of the inverse square root as $\mathrm{sgn}(z) = z/\sqrt{z^2}$, arises in quantum chromodynamics [31, 32].

In many applications, the matrix $A$ is large and typically sparse or structured. In this case it is prohibitive to first compute the generally dense matrix $f(A)$ and then form the product with $\boldsymbol{b}$. The rational Krylov methods reviewed here avoid this problem by using a *projection* or *interpolation* approach for computing approximations to the vector $f(A)\boldsymbol{b}$ from a low-dimensional search space without forming $f(A)$ explicitly. Note that this is different from the *direct approximation* approach where $f$ is replaced by an explicitly computed rational function $r$ such that $r(A) \approx f(A)$ [13, 33–37]. However, all these methods have in common the fact that linear system solves with (shifted versions of) $A$ are required, and in rational Krylov methods one typically solves one linear system per iteration. Therefore a rational Krylov iteration may be considerably more expensive (in terms of computation time) than a polynomial Krylov iteration, which involves only a matrix-vector product with $A$. The applicability of rational Krylov methods hinges on the efficiency by which these linear systems can be solved. Since rational functions may exhibit approximation properties superior to polynomials, the number of overall iterations required by rational Krylov methods is hopefully smaller than that required by polynomial methods, provided that the poles of the rational functions involved have been chosen in a suitable way.

This review partly follows the exposition in my thesis [38] of 2010, but I have updated and extended the discussion of optimal pole selection, which has been an active research

---

[1] Whenever we cite several references in a row, such as [5–7], this is a possibly non-exhaustive list and should be read as "see, e.g., [5–7] and the references therein."

topic in recent years. The outline is as follows. Section 2 is an introduction to rational Krylov spaces and the rational Arnoldi algorithm. Various rational Krylov methods for approximating $f(A)\boldsymbol{b}$ are described in Section 3, with an emphasis on the rational Arnoldi method and variants thereof, namely, the extended Krylov subspace method and the shift-and-invert Arnoldi method, but also the nonorthogonal generalized Leja point (or PAIN) method is included. For obtaining fast convergent rational Krylov methods the selection of optimal or near-optimal poles is crucial, and Section 4 is devoted to this issue.

## 2 Rational Krylov spaces and the rational Arnoldi algorithm

Rational Krylov methods for computing $f(A)\boldsymbol{b}$ all have in common the fact that an approximation at iteration $m$ is of the form $r_m(A)\boldsymbol{b}$, where $r_m = p_{m-1}/q_{m-1}$ is a rational function with a prescribed denominator polynomial $q_{m-1} \in \mathcal{P}_{m-1}$. We will assume in the following that $q_{m-1}$ is factored as

$$q_{m-1}(z) = \prod_{j=1}^{m-1}(1 - z/\xi_j), \tag{1}$$

where the *poles* $\xi_1, \xi_2, \ldots$ are numbers in the extended complex plane $\overline{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ different from all eigenvalues $\lambda \in \Lambda(A)$ and 0. The exclusion of the pole 0 is a nonessential restriction, as one could exclude any other finite number $\sigma$ by shifting $z = \hat{z} + \sigma$ and all $\xi_j = \hat{\xi}_j + \sigma$. The *rational Krylov space of order $m$ associated with* $(A, \boldsymbol{b})$ is defined as (see [39, 40])

$$\mathcal{Q}_m(A, \boldsymbol{b}) := q_{m-1}(A)^{-1}\mathrm{span}\{\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^{m-1}\boldsymbol{b}\}.$$

The name "$\mathcal{Q}_m(A, \boldsymbol{b})$" is intended to remind the reader that there is always a denominator $q_{m-1}$ associated with it, even if this polynomial does not appear explicitly in our notation.

By assumption (1) on the denominators $q_{m-1}$, rational Krylov spaces are nested and of strictly increasing dimension $m$ until some invariance index $M \leq N$ is reached:

$$\mathcal{Q}_1(A, \boldsymbol{b}) \subset \cdots \subset \mathcal{Q}_m(A, \boldsymbol{b}) \subset \cdots \subset \mathcal{Q}_M(A, \boldsymbol{b}) = \mathcal{Q}_{M+1}(A, \boldsymbol{b}) = \cdots$$

If all the poles $\xi_1, \xi_2, \ldots, \xi_{m-1}$ are set to infinity, then $q_{m-1} \equiv 1$ and the rational Krylov space $\mathcal{Q}_m(A, \boldsymbol{b})$ reduces to a polynomial Krylov space

$$\mathcal{K}_m(A, \boldsymbol{b}) = \mathrm{span}\{\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^{m-1}\boldsymbol{b}\}.$$

As long as $m \leq M$, which we assume in the following, one can compute an orthonormal basis $V_m = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m] \in \mathbb{C}^{N \times m}$ of $\mathcal{Q}_m(A, \boldsymbol{b})$. This is typically done by Ruhe's *rational Arnoldi algorithm* [40], which we now briefly describe. In the first iteration one sets $\boldsymbol{v}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|$, which is a basis vector of $\mathcal{Q}_1(A, \boldsymbol{b})$. In subsequent iterations, a vector $\boldsymbol{v}_{j+1}$ is obtained by orthonormalizing

$$\boldsymbol{x}_j = (I - A/\xi_j)^{-1}A\boldsymbol{v}_j \tag{2}$$

against the previously computed orthonormal vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j$. The orthogonalization yields

$$\boldsymbol{x}_j = \sum_{i=1}^{j+1} \boldsymbol{v}_i h_{i,j}, \tag{3}$$

with a nonzero normalization coefficient $h_{j+1,j}$. At this stage we have computed an orthonormal basis $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{j+1}\}$ of $\mathcal{Q}_{j+1}(A,\boldsymbol{b})$. Equating (2) and (3) and left-multiplying both sides by $I - A/\xi_j$, and separation of the terms containing $A$, gives

$$A\left(\boldsymbol{v}_j + \sum_{i=1}^{j+1}\boldsymbol{v}_i h_{i,j}\xi_j^{-1}\right) = \sum_{i=1}^{j+1}\boldsymbol{v}_i h_{i,j}.$$

For $j = 1,\ldots,m < M$ we can collect these equations in a *rational Arnoldi decomposition*

$$AV_m(I_m + H_m D_m) + A\boldsymbol{v}_{m+1}h_{m+1,m}\xi_m^{-1}\boldsymbol{e}_m^T = V_m H_m + \boldsymbol{v}_{m+1}h_{m+1,m}\boldsymbol{e}_m^T, \tag{4}$$

where $D_m = \mathrm{diag}(\xi_1^{-1},\ldots,\xi_m^{-1})$ and $\boldsymbol{e}_m$ denotes the $m$-th unit coordinate vector in $\mathbb{R}^m$. Setting

$$\underline{H_m} := \begin{bmatrix} H_m \\ h_{m+1,m}\boldsymbol{e}_m^T \end{bmatrix} \quad \text{and} \quad \underline{K_m} := \begin{bmatrix} I_m + H_m D_m \\ h_{m+1,m}\xi_m^{-1}\boldsymbol{e}_m^T \end{bmatrix},$$

we may formulate (4) more succinctly as

$$AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}, \tag{5}$$

where $\underline{H_m}$ and $\underline{K_m}$ are unreduced upper Hessenberg matrices of size $(m+1)\times m$ (in our notation the underline always symbolizes an additional last row), and $V_{m+1} = [V_m, \boldsymbol{v}_{m+1}]$. Under the assumption that the last pole $\xi_m$ is infinite, the second summand on the left-hand side of (4) vanishes, and (5) reduces to

$$AV_m K_m = V_{m+1}\underline{H_m}, \tag{6}$$

where $K_m$ denotes the upper $m \times m$ part of $\underline{K_m}$. If all poles $\xi_j$ are infinite, then (5) reduces to the standard *(polynomial) Arnoldi decomposition* $AV_m = V_{m+1}\underline{H_m}$ (or *Lanczos decomposition* when $A$ is Hermitian). The following discussions therefore include polynomial Krylov methods as a special case.

## 3  Rational Krylov methods

In this section we review various rational Krylov methods for approximating $f(A)\boldsymbol{b}$ that have been proposed in the literature, and relate them to each other.

### 3.1  The rational Arnoldi method

Let $V_m \in \mathbb{C}^{N\times m}$ be an orthonormal basis of $\mathcal{Q}_m(A,\boldsymbol{b})$, which may for example be computed by the rational Arnoldi algorithm discussed in the previous section. The *rational Arnoldi approximation for $f(A)\boldsymbol{b}$ from $\mathcal{Q}_m(A,\boldsymbol{b})$* is defined as

$$\boldsymbol{f}_m^{\mathrm{RA}} := V_m f(A_m)V_m^*\boldsymbol{b}, \quad \text{where} \quad A_m := V_m^* A V_m. \tag{7}$$

The matrix $A_m \in \mathbb{C}^{m\times m}$ is often referred to as a *compression of $A$* or a *matrix Rayleigh quotient*. The applicability of the rational Arnoldi method rests on the fact that $\boldsymbol{f}_m^{\mathrm{RA}}$ potentially is a very good approximation of $f(A)\boldsymbol{b}$ even for small order $m$. In this case only the

computation of a matrix function $f(A_m)$ of size $m \times m$ is required, which is small compared to the original $f(A)$ problem of size $N \times N$. As was pointed out in [41] (and also in [42] for a variant of the rational Arnoldi algorithm), it is not necessary to compute $A_m = V_m^* A V_m$ by explicit projection provided that the last pole used in the rational Arnoldi algorithm was $\xi_m = \infty$: from the reduced rational Arnoldi decomposition (6) we find $A_m = H_m K_m^{-1}$. The matrix $K_m$ is indeed invertible because $\underline{H_m}$ is an unreduced upper Hessenberg matrix and hence both sides of (6) are of full rank $m$.

The rational Arnoldi approximation $\boldsymbol{f}_m^{\mathrm{RA}}$ enjoys several remarkable properties. First of all, it is exact if $f$ is a rational function represented in the rational Krylov space $\mathcal{Q}_m(A, \boldsymbol{b})$. This exactness property is well known for polynomial Arnoldi approximations [43–45], and generalizes to the rational Krylov case (see [46] for the special case of extended Krylov subspaces, and more generally in [38, 41]).

**Lemma 3.1** (Exactness) *Let $V_m \in \mathbb{C}^{N \times m}$ be an orthonormal basis of $\mathcal{Q}_m(A, \boldsymbol{b})$, and let $A_m = V_m^* A V_m$. Then for any rational function $\widetilde{r}_m \in \mathcal{P}_m / q_{m-1}$ we have*

$$(V_m V_m^*) \widetilde{r}_m(A) \boldsymbol{b} = V_m \widetilde{r}_m(A_m) V_m^* \boldsymbol{b},$$

*provided that $\widetilde{r}_m(A_m)$ is defined. In particular, if $r_m \in \mathcal{P}_{m-1} / q_{m-1}$, then*

$$r_m(A) \boldsymbol{b} = V_m r_m(A_m) V_m^* \boldsymbol{b},$$

*i.e., the rational Arnoldi approximation for $r_m(A) \boldsymbol{b}$ is exact.*

P r o o f. Define $\boldsymbol{q} = q_{m-1}(A)^{-1} \boldsymbol{b}$. We first show by induction that

$$(V_m V_m^*) A^j \boldsymbol{q} = V_m A_m^j V_m^* \boldsymbol{q} \quad \text{for all } j \in \{0, 1, \ldots, m\}. \tag{8}$$

Assertion (8) is obviously true for $j = 0$. Assume that it is true for some $j < m$. Then by the definition of a rational Krylov space we have $V_m V_m^* A^j \boldsymbol{q} = A^j \boldsymbol{q}$, and therefore

$$(V_m V_m^*) A^{j+1} \boldsymbol{q} = (V_m V_m^*) A V_m V_m^* A^j \boldsymbol{q} = (V_m V_m^*) A V_m A_m^j V_m^* \boldsymbol{q} = V_m A_m^{j+1} V_m^* \boldsymbol{q},$$

which establishes (8). Again from (8) we obtain by linearity

$$\boldsymbol{b} = q_{m-1}(A) \boldsymbol{q} = V_m q_{m-1}(A_m) V_m^* \boldsymbol{q},$$

or equivalently, $V_m^* \boldsymbol{q} = q_{m-1}(A_m)^{-1} V_m^* \boldsymbol{b}$. Replacing $V_m^* \boldsymbol{q}$ in (8) completes the proof. $\square$

**Remark 3.2** From Lemma 3.1 it follows that the rational Arnoldi method is closely related to rational Gauss quadrature when approximating the scalar expression $\boldsymbol{b}^* f(A) \boldsymbol{b}$, with $A$ being Hermitian. Let $f = \widetilde{r}_m \cdot r_m$ be the product of two rational functions $\widetilde{r}_m \in \mathcal{P}_m / q_{m-1}$ and $r_m \in \mathcal{P}_{m-1} / q_{m-1}$. By the definition of a rational Krylov space we have $r_m(A) \boldsymbol{b} = V_m V_m^* r_m(A) \boldsymbol{b}$, and therefore

$$
\begin{aligned}
\boldsymbol{b}^* f(A) \boldsymbol{b} &= \boldsymbol{b}^* \widetilde{r}_m(A) V_m V_m^* r_m(A) \boldsymbol{b} \\
&= (V_m V_m^* \overline{\widetilde{r}_m}(A) \boldsymbol{b})^* (r_m(A) \boldsymbol{b}) \\
&= (V_m \overline{\widetilde{r}_m}(A_m) V_m^* \boldsymbol{b})^* (V_m r_m(A_m) V_m^* \boldsymbol{b}) \\
&= (V_m^* \boldsymbol{b})^* f(A_m) (V_m^* \boldsymbol{b}),
\end{aligned}
$$

where we have used Lemma 3.1 for the third equality. The last term is a Gauss-type approximation for $\boldsymbol{b}^* f(A) \boldsymbol{b}$, being exact for all functions $f \in \mathcal{P}_{2m-1}/q_{m-1}^2$. This connection to Gauss quadrature is well known for polynomial Krylov methods [47, 48] and has been explored in [49–51] for the rational Krylov case.

The next theorem states an interpolation property which, together with its Corollary 3.4, is the basis for much of the available convergence analysis of the rational Arnoldi method. Again, this result was long known for the polynomial Arnoldi method [44, 45], and later generalized to the rational Krylov case [38, 41].

**Theorem 3.3** (Interpolation) *Assume that $f(A_m)$ and $q_{m-1}(A_m)^{-1}$ are defined. Then the rational function $r_m^{\mathrm{RA}}$ underlying the rational Arnoldi approximation $\boldsymbol{f}_m^{\mathrm{RA}} = r_m^{\mathrm{RA}}(A)\boldsymbol{b}$ defined by* (7) *interpolates $f$ at the* rational Ritz values *$\Lambda(A_m)$.*

P r o o f.  Define the function $\widetilde{f} := f q_{m-1}$. By the definition of a matrix function there exists a polynomial $\widetilde{p}_{m-1} \in \mathcal{P}_{m-1}$ such that $\widetilde{f}(A_m) = \widetilde{p}_{m-1}(A_m)$ and $\widetilde{p}_{m-1}$ interpolates $\widetilde{f}$ at $\Lambda(A_m)$. Equivalently, the rational function $\widetilde{p}_{m-1}/q_{m-1} = r_m^{\mathrm{RA}}$ interpolates $f$ at $\Lambda(A_m)$. By Lemma 3.1 the rational Arnoldi approximation $\boldsymbol{f}_m^{\mathrm{RA}}$ is exact for such functions, which concludes the proof.                                                                                                              □

The assumption that $q_{m-1}(A_m)^{-1}$ is defined is equivalent to the requirement that none of the rational Ritz values in $\Lambda(A_m)$ coincides with any of the poles $\xi_j$ $(j = 1, \ldots, m-1)$. Since all rational Ritz values are contained in the *numerical range*

$$\mathbb{W}(A) := \{\boldsymbol{v}^* A \boldsymbol{v} : \boldsymbol{v} \in \mathbb{C}^N, \ \|\boldsymbol{v}\| = 1\},$$

this requirement is necessarily satisfied if all poles stay away from this set. We will see in Section 4 that this is naturally true for reasonable choices of poles.

The numerical range is also a convenient set for bounding the norm of $f(A)$ and thereby the norm of the error of a rational Arnoldi approximation. By a theorem of Crouzeix [52] there exists a universal constant $C \le 11.08$ such that

$$\|f(A)\| \le C\|f\|_\Sigma, \tag{9}$$

where the norm on the right is the maximum norm on a compact set $\Sigma \supseteq \mathbb{W}(A)$. With the help of this inequality, the following near-optimality property of $\boldsymbol{f}_m^{\mathrm{RA}}$ can be established.

**Corollary 3.4** (Near-optimality) *Let $f$ be analytic in a neighborhood of a compact set $\Sigma \supseteq \mathbb{W}(A)$. Then the rational Arnoldi approximation $\boldsymbol{f}_m^{\mathrm{RA}}$ defined by* (7) *satisfies*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{\mathrm{RA}}\| \le 2C\|\boldsymbol{b}\| \min_{r_m \in \mathcal{P}_{m-1}/q_{m-1}} \|f - r_m\|_\Sigma,$$

*with a constant $C \le 11.08$. If $A$ is Hermitian, the result holds even with $C = 1$ and $\Sigma \supseteq \Lambda(A) \cup \Lambda(A_m)$.*

P r o o f.  By Lemma 3.1 we know that $r_m(A)\boldsymbol{b} = V_m r_m(A_m) V_m^* \boldsymbol{b}$ for every rational function $r_m \in \mathcal{P}_{m-1}/q_{m-1}$. Thus,

$$
\begin{aligned}
\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{\mathrm{RA}}\| &= \|f(A)\boldsymbol{b} - V_m f(A_m) V_m^* \boldsymbol{b} - r_m(A)\boldsymbol{b} + V_m r_m(A_m) V_m^* \boldsymbol{b}\| \\
&\le \|\boldsymbol{b}\| \left(\|f(A) - r_m(A)\| + \|f(A_m) - r_m(A_m)\|\right) \\
&\le 2C\|\boldsymbol{b}\| \cdot \|f - r_m\|_\Sigma,
\end{aligned}
$$

where we have used (9) for the last inequality. If $A$ is Hermitian then so is $A_m$, and (9) holds with $C = 1$ and $\Sigma \supseteq \Lambda(A)$. The proof is completed by taking the infimum over all $r_m \in \mathcal{P}_{m-1}/q_{m-1}$, and noting that this infimum is attained on the compact set $\Sigma$. $\qquad\square$

The near-optimality property of rational Arnoldi approximants is remarkable and deserves some further discussion. As the error bound in Corollary 3.4 is based on the numerical range, we will certainly not expect it to be sharp for highly nonnormal matrices. In practice one often observes that the approximation $\boldsymbol{f}_m^{\mathrm{RA}}$ is indeed much better than predicted by this bound, namely extremely close to the orthogonal projection of $f(A)\boldsymbol{b}$ onto the search space $\mathcal{Q}_m(A, \boldsymbol{b})$. The gap in the bound of Corollary 3.4 can be quite large even for symmetric matrices, as we illustrate by a simple example.

**Example 3.5** We consider the tridiagonal matrix

$$
T_n = \begin{bmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{n\times n},
$$

and the matrices $\widetilde{A}_1 = T_{900}$ and $\widetilde{A}_2 = T_{30} \oplus T_{30}$, both of which are of dimension $N = 900$. We now apply linear transformations to $\widetilde{A}_1$ and $\widetilde{A}_2$ such that the spectral interval of the transformed matrices becomes $\Sigma = [1, 1000]$, and denote these matrices as $A_1$ and $A_2$, respectively. Note that $A_1$ and $A_2$ correspond to shifted and scaled finite-difference discretization matrices of the 1D- and 2D-Laplacian, respectively. The matrix $A_3$ is a diagonal matrix with 900 equispaced eigenvalues in $\Sigma$. We run the rational Arnoldi method for approximating $\exp(-A_\ell)\boldsymbol{b}$ with these symmetric matrices ($\ell = 1, 2, 3$), where all poles are chosen (rather arbitrarily) as $\xi_j = -1$, and $\boldsymbol{b}$ is a random vector of norm 1. The convergence curves are shown in Figure 1, together with the error of the orthogonal projection of $\exp(-A_\ell)\boldsymbol{b}$ onto $\mathcal{Q}_m(A_\ell, \boldsymbol{b})$, and the error bound of Corollary 3.4. Although this bound is the same for all $A_\ell$, the observed convergence differs for these matrices. Only for $A_1$, the matrix with eigenvalues corresponding to the 1D-Laplacian, does the convergence closely follow the error bound. The error bounds would be sharper if, instead of working with the spectral interval $\Sigma = \mathbb{W}(A_\ell)$, we used the discrete sets $\Sigma_\ell = \Lambda(A_\ell) \cup \Lambda_{\ell,m}$, with $\Lambda_{\ell,m}$ denoting the set of $m$-th order rational Ritz values associated with $A_\ell$. However, these would not be very practical error bounds as we do not want to compute eigenvalues of large matrices. On the other hand, in some cases one has knowledge of the eigenvalue distribution of a matrix in terms of a density function (this is the case for the above three matrices), and there exists convergence theory that gives asymptotic bounds incorporating the fine structure of the spectrum and explaining the fast convergence of the rational Arnoldi method seen in Figure 1. It is beyond the scope of this paper to go into details, and we refer to [53, 54], as well as to related[2] work on the superlinear convergence behavior of the CG method [55, 56].

---

[2] The iterates of the Lanczos method for solving $A\boldsymbol{x} = \boldsymbol{b}$ with a Hermitian matrix $A$ and zero initial guess coincide with the polynomial Arnoldi approximations for $f(A)\boldsymbol{b} = A^{-1}\boldsymbol{b}$.
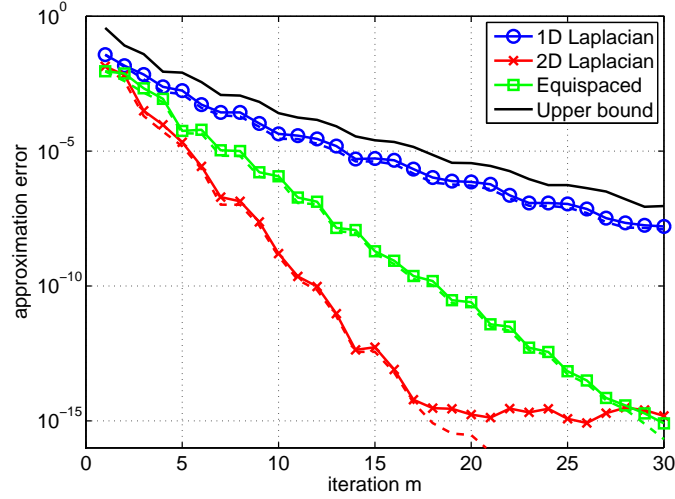
**Fig. 1** Convergence of the rational Arnoldi method for approximating $\exp(-A_\ell)\boldsymbol{b}$ with three different matrices $A_\ell$ described in Example 3.5, together with the error bound of Corollary 3.4. The dashed lines indicate the error of the orthogonal projection of $\exp(-A_\ell)\boldsymbol{b}$ onto $\mathcal{Q}_m(A_\ell, \boldsymbol{b})$, and these lines are partly overlaid by the error curves $\|\exp(-A_\ell)\boldsymbol{b} - \boldsymbol{f}_m^{\mathrm{RA}}\|$.

### 3.2 The extended Krylov subspace method

A popular special case of the rational Arnoldi method is known as the *extended Krylov subspace method*, obtained by choosing the poles alternately as $\xi_{2j} = \infty$ and $\xi_{2j-1} = 0$. This method was proposed by Druskin & Knizhnerman [46], and further investigated and improved in [57–59]. It is particularly suited for Markov functions

$$f(z) = \int_{-\infty}^{0} \frac{\mathrm{d}\gamma(x)}{z - x}, \tag{10}$$

where $\gamma$ is a (possibly signed) measure such that the integral converges absolutely for $z \in \mathbb{C} \setminus (-\infty, 0]$. The computational advantage of the extended Krylov subspace method is three-fold. First, the poles are chosen a priori and in this respect it is a black-box method. Second, if a direct solver is applicable for solving the linear systems with $A$, then only one LU factorization needs to be computed and it can be reused for all the solves in the rational Arnoldi algorithm. Third, for certain functions the alternating poles lead to the same convergence rate as can be achieved by using an asymptotically optimal single repeated pole. For example, it is known from [46] and [59, Theorem 3.4] that for a Hermitian matrix $A$ with spectral interval $[\lambda_{\min}, \lambda_{\max}] > 0$ one has

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{\mathrm{EK}}\| \leq C \left( \frac{\sqrt[4]{\kappa} - 1}{\sqrt[4]{\kappa} + 1} \right)^m \lesssim C \cdot \exp\left( -\frac{2m}{\sqrt[4]{\kappa}} \right), \quad \kappa = \frac{\lambda_{\max}}{\lambda_{\min}},$$

where $\boldsymbol{f}_m^{\mathrm{EK}}$ denotes the $m$-th extended Krylov subspace approximation, $C > 0$ is a constant independent of $m$, and the asymptotic upper bound denoted by $\lesssim$ is sharp for large condition numbers $\kappa$. The same linear convergence rate can be achieved with the rational Arnoldi

method by using the single repeated pole $\xi_j = \xi_{j+1} = -\sqrt{\lambda_{\min}\lambda_{\max}}$ (see [54]), but this choice requires knowledge of the extremal eigenvalues.

The 0-poles present in the extended Krylov subspace typically lead to "deflation" of the eigenvalues of $A$ closest to the origin, resulting in superlinear convergence. For example, it was shown in [54] that the extended Krylov subspace method applied for an $N \times N$ discretization of the 1D Laplacian actually converges superlinearly:

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{\mathrm{EK}}\| \lesssim C \cdot \exp\left(-\frac{\sqrt{8}m^{3/2}}{3\sqrt{N}}\right).$$

It should also be noted that an extended Krylov subspace associated with a Hermitian matrix $A$ can be constructed by a short-term recurrence on "blocks" of two vectors, each orthogonalization involving at most the last two blocks [59], although this may add computational complications due to loss of orthogonality of the basis vectors (see [60] for a discussion of this effect in the polynomial Krylov case). Such a short recurrence reduces the number of inner products in the rational Arnoldi algorithm, but still all Krylov basis vectors need to be stored in $V_m$ for forming $\boldsymbol{f}_m^{\mathrm{EK}} = V_m f(A_m) V_m^* \boldsymbol{b}$.

### 3.3 The shift-and-invert Arnoldi method

The *shift-and-invert Arnoldi method* for the approximation of matrix functions was introduced independently by Moret & Novati [61] (under the name of the *Arnoldi restricted denominator method*) and van den Eshof & Hochbruck [62]. The idea is to run the polynomial Arnoldi algorithm with the spectrally transformed matrix $(A - \xi I)^{-1}$, $\xi \notin \Lambda(A)$. This yields a polynomial Arnoldi decomposition

$$(A - \xi I)^{-1} V_m = V_{m+1} \underline{H_m}, \tag{11}$$

where $\underline{H_m}$ is an $(m+1) \times m$ unreduced upper Hessenberg matrix, and $V_{m+1} = [V_m, \boldsymbol{v}_{m+1}]$ is an orthonormal basis of $\mathcal{Q}_{m+1}(A, \boldsymbol{b}) = \mathcal{K}_{m+1}((A - \xi I)^{-1}, \boldsymbol{b})$, a rational Krylov space with all poles concentrated at $\xi$. The *shift-and-invert Arnoldi approximation* for $f(A)\boldsymbol{b}$ is defined as

$$\boldsymbol{f}_m^{\mathrm{SI}} := V_m f(S_m) V_m^* \boldsymbol{b}, \quad \text{where} \quad S_m := H_m^{-1} + \xi I_m. \tag{12}$$

This approximation can be easily related to the rational Arnoldi approximation $\boldsymbol{f}_m^{\mathrm{RA}}$: left-multiplication of (11) by $V_m^*(A - \xi I)$, and separation of the term $V_m^* A V_m$ yields

$$\begin{aligned} V_m^* A V_m &= H_m^{-1} + \xi I_m - V_m^* A \boldsymbol{v}_{m+1} h_{m+1,m} \boldsymbol{e}_m^T H_m^{-1} \\ &= S_m - V_m^* A \boldsymbol{v}_{m+1} h_{m+1,m} \boldsymbol{e}_m^T H_m^{-1} \\ &=: S_m - M_m, \end{aligned}$$

which shows that $S_m$ is a rank-1 modification of $A_m = V_m^* A V_m$. Therefore $\boldsymbol{f}_m^{\mathrm{SI}}$ is *not* a rational Arnoldi approximation for $f(A)\boldsymbol{b}$ from $\mathcal{Q}_m(A, \boldsymbol{b})$. However, with the function $\hat{f}(\hat{z}) := f(\hat{z}^{-1} + \xi)$ we have $\boldsymbol{f}_m^{\mathrm{SI}} = V_m \hat{f}(H_m) V_m^* \boldsymbol{b}$, hence $\boldsymbol{f}_m^{\mathrm{SI}}$ is a *polynomial* Arnoldi approximation for $\hat{f}((A - \xi I)^{-1})\boldsymbol{b} = f(A)\boldsymbol{b}$ associated with the Arnoldi decomposition (11). This connection allows us to conclude that there exists an interpolation characterization of the shift-and-invert approximation similar to Theorem 3.3 (see [62]).

**Theorem 3.6** *The rational function $r_m^{\mathrm{SI}}(z) = p_{m-1}^{\mathrm{SI}}(z)/(z-\xi)^{m-1}$ underlying the shift-and-invert Arnoldi approximation $\boldsymbol{f}_m^{\mathrm{SI}} = r_m^{\mathrm{SI}}(A)\boldsymbol{b}$ defined by* (12) *interpolates $f$ at the points in $\Lambda(S_m)$.*

A near-optimality result similar to Corollary 3.4 can also be given.

**Corollary 3.7** *Let $f$ be analytic in a neighborhood of a compact set $\Sigma \supseteq \mathbb{W}(A)$. Then the shift-and-invert Arnoldi approximation $\boldsymbol{f}_m^{\mathrm{SI}}$ defined by* (12) *satisfies*

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m^{\mathrm{SI}}\| \le 2C\|\boldsymbol{b}\| \min_{p_{m-1}\in\mathcal{P}_{m-1}} \|f(z) - p_{m-1}(z)/(z-\xi)^{m-1}\|_\Sigma,$$

*with a constant $C \le 11.08$.*

A computational advantage of the shift-and-invert Arnoldi method appears when $A - \xi I$ is Hermitian, because in this case $H_m$ is a tridiagonal symmetric matrix and the basis $V_m$ can be computed via the Lanczos three-term recurrence. However, the full basis $V_m$ still needs to be stored for forming the approximation $\boldsymbol{f}_m^{\mathrm{SI}}$ in (12).

### 3.4   The generalized Leja point method

A nonorthogonal rational Krylov method, proposed in [38] (under the name *PAIN*, which stands for *poles and interpolation nodes*), is given by the iteration

$$\boldsymbol{v}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|,$$
$$\alpha_j \boldsymbol{v}_{j+1} = (I - A/\xi_j)^{-1}(A - \sigma_j I)\boldsymbol{v}_j,$$

the numbers $\alpha_j, \sigma_j \in \mathbb{C}$ being arbitrary for the moment, except that we require $\alpha_j \ne 0$ and $\sigma_j \ne \xi_j$ for all $j = 1, \ldots, m$. It is easily seen that this iteration generates a decomposition $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$, where $V_{m+1} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m+1}]$ and

$$\underline{K_m} = \begin{bmatrix} 1 & & & \\ \alpha_1/\xi_1 & 1 & & \\ & \alpha_2/\xi_2 & \ddots & \\ & & \ddots & 1 \\ \hline & & & \alpha_m/\xi_m \end{bmatrix} \quad \text{and} \quad \underline{H_m} = \begin{bmatrix} \sigma_1 & & & \\ \alpha_1 & \sigma_2 & & \\ & \alpha_2 & \ddots & \\ & & \ddots & \sigma_m \\ \hline & & & \alpha_m \end{bmatrix}.$$

It can be shown (see [38, Thm. 5.8]) that the associated *generalized Leja point approximation*

$$\boldsymbol{f}_m^{\mathrm{GL}} := V_m f(H_m K_m^{-1})\|\boldsymbol{b}\|\boldsymbol{e}_1 \tag{14}$$

satisfies $\boldsymbol{f}_m^{\mathrm{GL}} = r_m^{\mathrm{GL}}(A)\boldsymbol{b}$ with a rational function $r_m^{\mathrm{GL}}$ having poles $\xi_1, \ldots, \xi_{m-1}$ and interpolating $f$ at the points $\sigma_1, \ldots, \sigma_m$. Therefore this approximation method corresponds to rational interpolation of $f(A)\boldsymbol{b}$ with preassigned poles and interpolation nodes. Compared with methods based on orthogonal Krylov bases, it has the advantage that the full matrix $V_m$ does not need to be stored for computing $\boldsymbol{f}_m^{\mathrm{GL}}$; instead the approximations can be updated using only the last Krylov basis vector, $\boldsymbol{f}_m^{\mathrm{GL}} = \boldsymbol{f}_{m-1}^{\mathrm{GL}} + c_m \boldsymbol{v}_m$, where $c_m$ denotes the last entry of the vector $f(H_m K_m^{-1})\|\boldsymbol{b}\|\boldsymbol{e}_1$.

In the polynomial case when all $\xi_j = \infty$, the generalized Leja point method reduces to the so-called *real Leja point method* [63–65]. In the rational case, the points $\{(\sigma_j, \xi_j)\}$ should

be chosen as *generalized Leja points*, hence the name of the method. We will discuss these points in the next section. It was also shown in [38] that the generalized Leja point method can be seen as a restarted rational Krylov method for matrix functions (and thereby generalizes certain polynomial restarted Krylov methods, see [66–69]).

## 4 Pole selection

Every rational Krylov approximation is of the form $\boldsymbol{f}_m = r_m(A)\boldsymbol{b}$, with a rational function $r_m \in \mathcal{P}_{m-1}/q_{m-1}$. It is therefore not surprising that the pole selection for rational Krylov methods is closely connected to rational approximation of the function $f$. Indeed, by Crouzeix's inequality (9) we know that

$$\|f(A)\boldsymbol{b} - \boldsymbol{f}_m\| \leq C\|\boldsymbol{b}\|\,\|f - r_m\|_\Sigma,$$

where $C \leq 11.08$ and $\Sigma \supseteq \mathbb{W}(A)$. In Example 3.5 we have demonstrated that such a bound based on the numerical range may be crude, but for convenience and simplicity we will rely on this universal bound. The aim for a small approximation error $\|f(A)\boldsymbol{b} - \boldsymbol{f}_m\|$ then leads to the problem of uniform rational approximation of $f$ on $\Sigma$. For the rational Arnoldi approximation $\boldsymbol{f}_m^{\mathrm{RA}}$, which by Corollary 3.4 is a near-optimal extraction from $\mathcal{Q}_m(A, \boldsymbol{b})$, the matter further reduces to the problem of finding an optimal denominator $q_{m-1}$, or equivalently, an optimal search space $\mathcal{Q}_m(A, \boldsymbol{b})$.

In many applications, the function $f = f^\tau$ also depends on a parameter $\tau$ from a *parameter set* $T$, and consequently the same is true for the approximations $\boldsymbol{f}_m^\tau \approx f^\tau(A)\boldsymbol{b}$. This needs to be taken into account when optimizing the poles of a rational Krylov space, which should depend on $T$ but not on each single $\tau \in T$. Moreover, it is sometimes necessary to restrict the poles $\xi_j$ to a *pole set* $\Xi$. For example, if complex arithmetic needs to be avoided in the rational Arnoldi algorithm, $\Xi = \mathbb{R} \cup \{\infty\}$ is an appropriate restriction.

Asymptotically optimal rational approximants can be constructed by interpolation, using tools from logarithmic potential theory. We will keep the amount of theory to a minimum here, and refer the interested reader to [70–72]. Let $\Sigma$ and $\Xi$ be closed subsets of $\overline{\mathbb{C}}$, both of nonzero logarithmic capacity (a property which holds for all sets we consider in the following) and of positive distance. The pair $(\Sigma, \Xi)$ is called a *condenser*, and associated with it is a positive number $\operatorname{cap}(\Sigma, \Xi)$ called the *condenser capacity* [73, 74]. Let us consider a sequence of rational nodal functions

$$s_m(z) = \frac{(z - \sigma_1)\cdots(z - \sigma_m)}{(1 - z/\xi_1)\cdots(1 - z/\xi_m)}, \quad m = 1, 2, \ldots \tag{15}$$

with zeros $\sigma_j \in \Sigma$ and poles $\xi_j \in \Xi$. Our aim is to make these rational functions asymptotically as large as possible on the set $\Xi$, and as small as possible on $\Sigma$. It can be proven (see [74, 75]) that for any sequence of rational functions $s_m$ of the form (15) one has

$$\limsup_{m \to \infty} \left( \frac{\sup_{z \in \Sigma} |s_m(z)|}{\inf_{z \in \Xi} |s_m(z)|} \right)^{1/m} \geq e^{-1/\operatorname{cap}(\Sigma, \Xi)}. \tag{16}$$

The problem of finding a sequence $\{s_m\}$ such that equality holds in (16) is called the *generalized Zolotarev problem for the condenser* $(\Sigma, \Xi)$, because it reduces to the third of Zolotarev's classical problems if $\Sigma$ and $\Xi$ are real intervals [74, 76, 77].

The problem of determining the capacity of an arbitrary condenser $(\Sigma, \Xi)$ is nontrivial. The situation simplifies if both $\Sigma$ and $\Xi$ are simply connected sets. Then by the Riemann mapping theorem (see [78, Thm. 5.10h]) there exists a bijective function $\Phi$ that conformally maps the complement of $\Sigma \cup \Xi$ 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 $R^{-1} = e^{-1/\operatorname{cap}(\Sigma, \Xi)}$.

A practical method for obtaining asymptotically optimal rational functions for the Zolotarev problem is the following greedy algorithm: starting with points $\sigma_1 \in \Sigma$ and $\xi_1 \in \Xi$ of minimal distance, the points $\sigma_{j+1} \in \Sigma$ and $\xi_{j+1} \in \Xi$ are determined recursively such that with $s_j$ defined in (15) the conditions

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

are satisfied. The points $\{(\sigma_j, \xi_j)\}$ are called *generalized Leja points* [79]. The connection of the sequence $\{s_m\}$ with rational interpolation is now easily explained. Assume that the function $f$ can be represented by a Cauchy-type integral over a contour $\Gamma$ winding around $\Sigma$. Then by the Walsh–Hermite formula, the error of a rational function $r_m$ with poles $\xi_1, \ldots, \xi_m$ that interpolates $f$ at the nodes $\sigma_1, \ldots, \sigma_m$ is given by

$$f(z) - r_m(z) = \frac{1}{2\pi i} \int_\Gamma \frac{s_m(z)}{s_m(\zeta)} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta, \quad z \in \Sigma.$$

The uniform approximation error can be bounded as

$$\|f - r_m\|_\Sigma \leq D \frac{\max_{z \in \Sigma} |s_m(z)|}{\min_{\zeta \in \Gamma} |s_m(\zeta)|},$$

where $D = D(f, \Gamma)$ is a constant independent of $m$. In conjunction with (16) we obtain

$$\limsup_{m \to \infty} \|f - r_m\|_\Sigma^{1/m} \leq e^{-1/\operatorname{cap}(\Sigma, \Gamma)} = R^{-1}, \tag{17}$$

provided the nodes $\sigma_j$ and poles $\xi_j$ are asymptotically distributed like generalized Leja points (or more generally, distributed according to the so-called *equilibrium measure* $\mu^*$) on the condenser $(\Sigma, \Gamma)$. This suggests that such points are reasonable nodes and poles for rational interpolation.

**Remark 4.1** It is interesting to note that a best uniform rational approximation $r_m^*$ of degree $m$ to $f$ on $\Sigma$ converges at most twice as fast as $r_m$ in (17). More precisely,

$$\liminf_{m \to \infty} \|f - r_m^*\|_\Sigma^{1/m} \geq e^{-2/\operatorname{cap}(\Sigma, \Gamma)} \tag{18}$$

(see [80–82]). This result is sharp in the sense that equality holds in (18) for certain classes of analytic functions whose region of analyticity contains $\Sigma$ and is bounded by $\Gamma$, for example Markov functions [83] or functions with a finite number of algebraic branch points [84]. The poles of best uniform rational approximants are typically free, i.e., they all alter when $m$ is increased to $m + 1$, which does not allow for the construction of nested rational Krylov spaces which achieve convergence like (18).

In the following subsections we will relate the above results to various approaches for computing asymptotically optimal poles of a rational Krylov space, depending on the function $f$ (or $f^\tau$ with a parameter $\tau \in T$) and the configuration of $\Sigma$ and $\Xi$. The reader will find that we only consider real intervals for $\Sigma$, which means that $A$ should be Hermitian. An analysis for more general $\Sigma$ is often possible (see [41]), but the corresponding conformal maps become more complicated.

### 4.1 Resolvent function

We first consider the uniform approximation of $f^\tau(z) = (z - \tau)^{-1}$ on a closed set $\Sigma \subset \overline{\mathbb{C}}$ for (frequency) parameters $\tau \in T$, by rational functions having their poles in a set $\Xi$. Closely related optimization problems arise in the parameter selection for the ADI method [85–88], and in model order reduction problems in the frequency domain, where the selected poles are typically called *expansion* or *interpolation points* (which is different from the nomenclature used here; see [5, 89, 90]).

**Unbounded $\Sigma = [0, +\infty]$, bounded $T = \pm i[\omega_{\mathbf{min}}, \omega_{\mathbf{max}}]$ symmetric to $\mathbb{R}$, $\Xi = T$:** This problem was considered in [7] and therein related to the classical Zolotarev problem for the weighted best rational approximation of the inverse square root on a positive interval. From this relation the authors deduced the capacity of $(\Sigma, \Xi = T)$ and the convergence rate [7, formulas (3.10) and (3.15)]

$$R = \exp\left(\frac{\pi}{2} \frac{K(\delta)}{K'(\delta)}\right) \simeq \exp\left(\frac{\pi^2}{4 \log(4/\delta)}\right), \quad \delta = \sqrt{\frac{\omega_{\min}}{\omega_{\max}}},$$

with the complete elliptic integral of the first kind and its complement defined as

$$K(\kappa) = \int_0^1 \frac{\mathrm{d}t}{\sqrt{(1 - t^2)(1 - \kappa^2 t^2)}}, \quad K'(\kappa) = K(\sqrt{1 - \kappa^2}).$$

The asymptotic estimate denoted by $\simeq$ is precise for small $\delta$. Asymptotically optimal interpolation nodes and poles can be obtained by discretization of the signed equilibrium measure for $(\Sigma, \Xi)$, or as explained above, by using generalized Leja points. The resulting poles $\xi_j \in \Xi$ lie on the imaginary axis, asymptotically distributed with a density symmetric to the real axis. Note that if both $A$ and $\boldsymbol{b}$ are real, and poles occur in complex conjugate pairs $\xi_{2j} = \overline{\xi_{2j-1}}$ $(j = 1, 2, \ldots)$, then this can be exploited such that the number of complex-shifted linear systems to be solved in the rational Arnoldi algorithm is only half the number of iterations (see also [91, 92] for computational aspects of solving complex-shifted linear systems).

**Unbounded $\Sigma = [0, +\infty]$, bounded $T = i[\omega_{\mathbf{min}}, \omega_{\mathbf{max}}]$, and poles on $\Xi = T$:** The complement of the condenser $(\Sigma, T) = ([0, +\infty], i[\omega_{\min}, \omega_{\max}])$ is doubly connected and therefore possesses a conformal map onto an annulus. Unfortunately, its nonsymmetric geometry makes this condenser more difficult to analyze. We have numerical evidence that the associated Riemann modulus may be

$$R = \exp\left(\frac{\pi^2}{4} \frac{K(\delta)}{K'(\delta)}\right) \simeq \exp\left(\frac{\pi^3}{8 \log(4/\delta)}\right), \quad \delta = \sqrt{\frac{\omega_{\min}}{\omega_{\max}}}.$$

**Unbounded $\Sigma = [0, +\infty]$, bounded $T = i[\omega_{\mathbf{min}}, \omega_{\mathbf{max}}]$, and real poles in $\Xi = (-\infty, 0)$:**
For real data $A$ and $\boldsymbol{b}$ it is favorable to have *real* poles $\xi_j$ in order to avoid complex
arithmetic in the linear system solves. For bounded $T$ and unbounded $\Sigma$ the following
construction was proposed in [38, Section 7.5]. Denote by $R_1(\tau, \xi) > 1$ the convergence
rate for the approximation of $f^\tau(z) = (z - \tau)^{-1}$ from a rational Krylov space with all
poles being concentrated at a finite pole $\xi$. The quantity $R_1(\tau, \xi)$ can be calculated via
a polynomial approximation problem in the transformed variable $\hat{z} = (z - \xi)^{-1}$. For
example, if $\Sigma = [0, +\infty]$, $T = i[\omega_{\min}, \omega_{\max}]$ and $\xi < 0$, we have

$$R_1(\tau, \xi) = \left( 1 + \frac{\sqrt{8}c^{3/4} + 4c^{1/2} + \sqrt{8}c^{1/4}}{1 + c} \right)^{1/2}, \quad c = |\tau|^2/\xi^2.$$

By considering $p$ cyclically repeated poles $\xi_1, \ldots, \xi_p$, one can then show that the geo-
metric mean $R(\tau) = (R_1(\tau, \xi_1) \cdots R_1(\tau, \xi_p))^{1/p}$ is the convergence rate for the approx-
imation of $f^\tau$ from a rational Krylov space built with these cyclic poles. The poles $\xi_j$
can be optimized by maximizing the worst-case convergence rate $\min_{\tau \in T} R(\tau)$.

**Bounded $\Sigma = [\lambda_{\min}, \lambda_{\max}] > 0$, unbounded $T = i\mathbb{R}$, real poles on $\Xi = -\Sigma$:** It would
be natural to consider generalized Leja points for the condenser $(\Sigma, T)$, that is, to place
the poles of asymptotically optimal rational functions on the imaginary axis. Via con-
formal mapping it was shown in [23, formula (4.8)] that the associated convergence rate
given by the Riemann modulus would be

$$R = \exp\left( \frac{\pi}{4} \frac{K'(\mu)}{K(\mu)} \right) \simeq \exp\left( \frac{\pi^2}{4 \log(2/\delta)} \right), \quad \mu = \left( \frac{1 - \delta}{1 + \delta} \right)^2, \quad \delta = \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}. \quad (19)$$

It was discovered in [93] that rational functions constructed for the symmetric condenser
$(\Sigma, -\Sigma)$ actually achieve the same convergence rate on $(\Sigma, T)$. An intuitive explanation
is the following: if a rational function $s_m$ has real zeros $\sigma_1, \ldots, \sigma_m \in \Sigma$ and "mirrored"
poles $\xi_j = -\sigma_j$, then the quotient $s_m(z)/s_m(-z)$ attains modulus 1 on the imaginary
axis. Indeed, the condenser $(\Sigma, -\Sigma)$ is of half the capacity of $(\Sigma, T)$. One can therefore
construct generalized Leja points for the former condenser, with poles in $\Xi = -\Sigma$, and
use these poles for the construction of an asymptotically optimal rational Krylov space.

It should be noted that poles and interpolation nodes symmetric to the imaginary axis
arise in the context of optimal $\mathcal{H}_2$ model order reduction, and the corresponding optimal
points can be constructed iteratively by repeated runs of the rational Arnoldi algorithm,
with the poles taken as "mirrored" rational Ritz values of the previous run (*iterative
rational Krylov algorithm*, see [94]).

**Bounded $\Sigma = [\lambda_{\min}, \lambda_{\max}] > 0$, unbounded $T = i\mathbb{R}$, and adaptive poles on $\Xi = -\Sigma$:**
An adaptive approach for choosing the poles of a rational Arnoldi approximation to
the resolvent function has been proposed by Druskin & Simoncini [95]. It is based
on the interpolation property of the rational Arnoldi method, stated in Theorem 3.3.
As at iteration $m$ of this method the poles $\xi_1, \ldots, \xi_m$ have already been chosen, and
the interpolation nodes $\sigma_1, \ldots, \sigma_m$ are explicitly known to be the rational Ritz values
$\Lambda(A_m)$, the corresponding nodal function $s_m$ of (15) is also known. The next pole $\xi_{m+1}$

is then selected as $\xi_{m+1} = \arg\min_{x \in \Xi} |s_m(x)|$, aiming to make $|s_m(x)|$ uniformly large on $\Xi = -\Sigma$. Numerical experiments indicate that this results in a rational Arnoldi method with convergence rate $R$ at least as great as (19).

### 4.2 Exponential function

We will now focus on the choice of optimal poles when approximating $\exp(-\tau A)\boldsymbol{b}$ for a (time) parameter interval $[\tau_{\min}, \tau_{\max}] > 0$. We assume that the numerical range $\mathbb{W}(A) \subseteq \Sigma$ is contained in the right half-plane (and we denote this as $\Sigma \geq 0$).

**Unbounded $\Sigma \geq 0$, a single parameter $T = \{\tau\}$, and a single repeated pole $\xi < 0$:**
Due to their computational advantages, rational approximations to $e^{-z}$ on $\Sigma = [0, +\infty]$ with real poles have been studied extensively in the literature, see [96–99]. In particular, it is known that the best uniform rational approximation of type $[m/m]$ has a single pole $\xi < 0$ of order $m$ [100], and $\xi \sim -m/\sqrt{2}$ as $m \to \infty$ [101]. The case of choosing a single optimal pole $\xi$ is relevant in the context of the shift-and-invert Arnoldi method, and this problem has been studied extensively for Hermitian and non-Hermitian $A$ (see [62, 102] and [41, 61, 103, 104], respectively).

Because $\Sigma$ can be assumed to be unbounded as long as it is contained in the right half-plane and only contains a finite subinterval of the imaginary axis (where $e^z$ is oscillatory), rational Krylov methods for the matrix exponential are often said to have *mesh-independent* convergence. If $A$ is the discretization of a spatial differential operator, then the number of iterations required by a rational Krylov method to achieve a prescribed error tolerance will be independent of the mesh width. On the other hand, in a practical situation one typically desires a more accurate result when refining the mesh, so that the number of required iterations actually increases.

**Bounded $\Sigma = [\lambda_{\min}, \lambda_{\max}] > 0$, parameters $T \subseteq [0, +\infty)$, and poles on $\Xi = -\Sigma$:**
This case was considered by Druskin, Knizhnerman & Zaslavsky [23]. The key idea is to exploit the inverse Fourier representation of the exponential function, namely

$$e^{-\tau z} = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{e^{-\tau z}}{\zeta - z} \, d\zeta, \quad z > 0, \ \tau \geq 0.$$

Using this formula in the rational Arnoldi approximation $\boldsymbol{f}_m^\tau = V_m e^{-\tau A_m} V_m^* \boldsymbol{b}$ one obtains

$$
\begin{aligned}
e^{-\tau A}\boldsymbol{b} - \boldsymbol{f}_m^\tau &= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{-\tau\zeta} \left( (\zeta I - A)^{-1}\boldsymbol{b} - V_m(\zeta I_m - A_m)^{-1} V_m^* \boldsymbol{b} \right) \, d\zeta \\
&= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{-\tau z} \left( r^\zeta(A)\boldsymbol{b} - r_m^\zeta(A)\boldsymbol{b} \right) \, d\zeta,
\end{aligned}
$$

where $r_m^\zeta(z) = p_{m-1}^\zeta(z)/q_{m-1}(z)$ interpolates the resolvent function $r^\zeta(z) := (\zeta - z)^{-1}$ at the rational Ritz values $\Lambda(A_m)$. Apparently, the aim is to make the error $r^\zeta - r_m^\zeta$ as small as possible on $\Sigma = [\lambda_{\min}, \lambda_{\max}]$ for all "parameters" $\zeta \in i\mathbb{R}$. As described in Subsection 4.1, this can be done by solving a Zolotarev problem on the condenser

$(\Sigma, -\Sigma)$. Using the near-optimality of the rational Arnoldi method and the Plancherel identity one can indeed show that

$$\limsup_{m \to \infty} \|f^\tau(A)\boldsymbol{b} - \boldsymbol{f}_m^\tau\|^{1/m} \le R^{-1} \quad \text{for all } \tau \in T, \tag{20}$$

with $R$ given in (19), provided that the poles of the rational Krylov space are distributed on $\Xi = -\Sigma$ according to the equilibrium measure of the condenser $(\Sigma, -\Sigma)$, see [23]. Note that the rate $R$ deteriorates with a growing condition number $\lambda_{\max}/\lambda_{\min}$, but is independent of the length of the time interval $T$. Hence this approach should be preferred to approaches for unbounded $\Sigma$ if the condition number of $A$ is moderate and $T$ is a large interval.

**Bounded $\Sigma = [\lambda_{\min}, \lambda_{\max}] > 0$, parameters $T \subseteq [0, +\infty)$, adaptive poles on $\Xi = -\Sigma$:**
As is the case for the resolvent function, an adaptive choice of poles is also possible for the matrix exponential function, see [24]. Also here one aims to make the nodal function $s_m$ of (15) uniformly large on $\Xi = -\Sigma$ by choosing $\xi_{m+1} = \arg\min_{x \in \Xi} |s_m(x)|$ at each iteration of the rational Arnoldi method.

**Example 4.2** We consider the computation of $\exp(-\tau A_1)\boldsymbol{b}$ with the matrix $A_1$ and vector $\boldsymbol{b}$ from Example 3.5. Recall that $\Sigma = [1, 1000]$. The time interval is $T = [10^{-4}, 1]$ and as poles for the rational Krylov space we use generalized Leja points on the plate $-\Sigma$ of the condenser $(-\Sigma, \Sigma)$. On the left of Figure 2 we show the convergence of the rational Arnoldi method for 17 logspaced time parameters $\tau \in T$, together with the expected convergence slope given by (20) (and (19)). Note that we have approximated all $\exp(-\tau A_1)\boldsymbol{b}$ from the same rational Krylov space, which has been constructed only once, and we could use this space to cheaply extract approximations for much more than 17 time parameters in a larger time interval. On the right we plot level lines of the rational nodal function $|s_{20}|$, together with the rational Ritz values of order 20 and the first 19 poles of the rational Krylov space. This plot confirms that $|s_m|$ is uniformly small on $\Sigma = [1, 1000]$ (relative to $|s_m|$ on $-\Sigma$) and approximately constant on the imaginary axis.

**Remark 4.3** Rational Krylov methods can be generalized to problems where both the matrix $A = A(\tau)$ and/or the vector $\boldsymbol{b} = \boldsymbol{b}(\tau)$ may be parameter-dependent. The corresponding methods go under the names *nonlinear Krylov subspace method* [105, 106], *interpolatory projection method* [107, 108], or *parameter-dependent Krylov subspace method* [109]. Such methods can be applied, e.g., for the solution of general time-invariant dynamical systems, and a potential-theoretic approach to pole selection for such problems has been given in [110].

### 4.3  Markov functions

We finally discuss the selection of optimal poles for rational Krylov methods when approximating Markov functions $f$ defined by (10). Such functions are analytic in the complex plane with the exception of the set $\Gamma = [-\infty, 0]$, the support of the generating measure $\gamma$. It is therefore natural to take $\Xi = \Gamma$ as the set of allowed poles. The choice of alternating poles 0 and $\infty$ used in the extended Krylov subspace method has already been discussed in Subsection 3.2. Asymptotically optimal poles for the rational Arnoldi method have been studied by Beckermann & Reichel [41]. These authors give upper and lower bounds for the rational Arnoldi
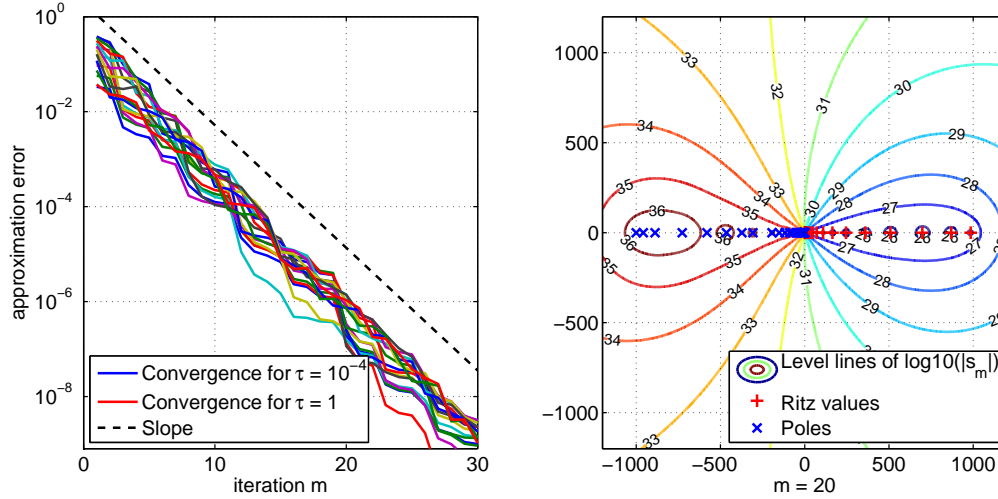
**Fig. 2** Left: Convergence of the rational Arnoldi method for approximating $\exp(-\tau A_1)\boldsymbol{b}$ for 17 parameters $\tau \in [10^{-4}, 1]$. The dashed line shows the expected convergence slope given by (20) (and (19)). Right: Level lines of the rational nodal function $|s_{20}|$, together with the rational Ritz values of order 20 on $\Sigma = [1, 1000]$ (red +) and the 19 poles of the associated rational Krylov space on $-\Sigma$ (blue ×).

approximation error using techniques based on the Faber transform, and relate their results to work by Hale, Higham & Trefethen [36], who computed explicit rational approximants for $f$ via Talbot quadrature formulas.

In the following we consider the approximation of $f(A)\boldsymbol{b}$ for a Hermitian matrix $A$ having eigenvalues in $\Sigma = [\lambda_{\min}, \lambda_{\max}] > 0$. It can be shown that the condenser $(\Sigma, \Xi) = ([\lambda_{\min}, \lambda_{\max}], [-\infty, 0])$ has Riemann modulus

$$R = \exp\left(\frac{\pi}{2}\frac{K'(\mu)}{K(\mu)}\right) \simeq \exp\left(\frac{\pi^2}{2\log(4/\delta)}\right), \quad \mu = \frac{1-\delta}{1+\delta}, \quad \delta = \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}. \quad (21)$$

(See [36, Fig. 3] for a geometrical sketch of the conformal map of the complement of $\Sigma \cup \Xi$ onto the annulus $\mathbb{A}_R$.) As explained at the beginning of this section, rational interpolation of $f$ with generalized Leja points as interpolation nodes and poles on $(\Sigma, \Xi)$ will converge at the rate $R$. Also the rational Arnoldi method will converge at least at this rate due to its near-optimality. We confirm this numerically in the left of Figure 3, showing the convergence of the generalized Leja point and the (standard) rational Arnoldi methods for approximating $f(A_2)\boldsymbol{b}$ with the "impedance function"

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

and the matrix $A_2$ and vector $\boldsymbol{b}$ of Example 3.5. In both cases, the selected poles are generalized Leja points. On the right of Figure 3 we show the relative error curves $|1 - x^{1/2}r_m(x)|$ of the rational interpolants $r_m^{\mathrm{RA}}$ and $r_m^{\mathrm{GL}}$ underlying both approximations, for $m = 15$. These
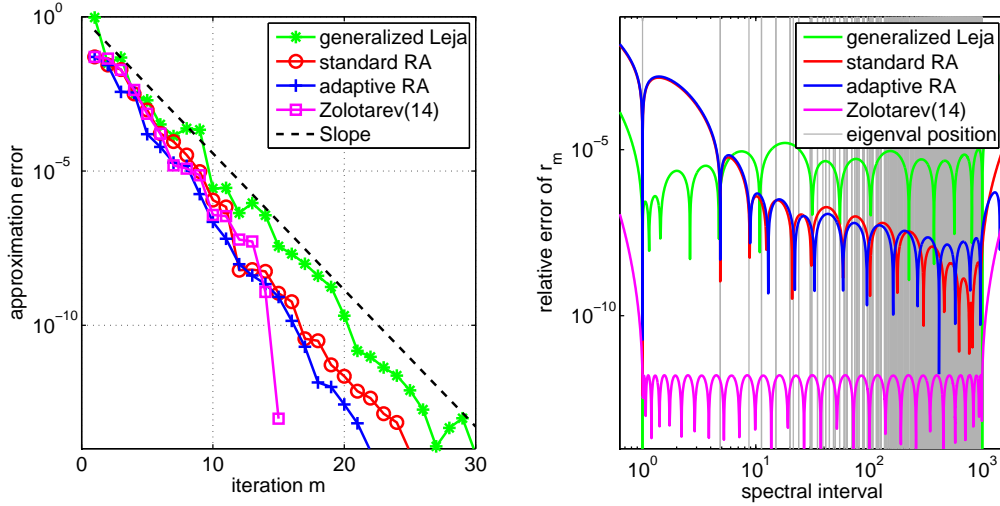
**Fig. 3** Left: Convergence of various rational Krylov methods for approximating $A_2^{-1/2}\boldsymbol{b}$ for a matrix $A_2$ with spectral interval $\Sigma = [1, 1000]$. The dashed line shows the expected convergence slope given by (21). Right: Relative error curves $|1 - x^{1/2}r_m(x)|$ of the rational functions $r_m$ underlying the rational Krylov approximants of order $m = 15$. The grey vertical lines indicate the positions of the eigenvalues of $A_2$.

are rational functions of type [14/14]. Note how the error curve for $r_m^{\mathrm{GL}}$ is uniformly small on the spectral interval $\Sigma$, and the error of $r_m^{\mathrm{RA}}$ is less uniform but almost zero very close to some of the left-most eigenvalues of $A$ (indicated by the vertical grey lines). This is an indication of spectral adaption of the rational Arnoldi method.

We also show the convergence curve for the adaptive rational Arnoldi method proposed in [111]. This method does not require the explicit computation of optimal poles; instead near-optimal poles are determined in the course of iterating. This method can therefore be seen as a black-box method. Finally, we show the convergence of rational Arnoldi where we have supplied the poles of Zolotarev's best rational approximation for $z^{-1/2}$ of degree 14 to the method (in Leja ordering, see [112] for more on best rational approximants). Note how the approximation error suddenly drops at iteration $m = 15$, which is when all 14 poles are present in the rational Krylov space. An accuracy of about $10^{-13}$ is achieved with these optimal poles about twice as fast as with the generalized Leja points (see Figure 3, left). This is Remark 4.1 in action: rational interpolants with free poles may converge about twice as fast as generalized Leja interpolants. The relative error curve of the Zolotarev rational function shows the well-known equioscillation property indicating optimality.

# References

[1] G. H. Golub and C. F. Van Loan, Matrix Computations, 3rd edition (Johns Hopkins University Press, Baltimore, MD, 1996).

[2] R. A. Horn and C. R. Johnson, Topics in Matrix Analysis (Cambridge University Press, 1991).

[3] N. J. Higham, Functions of Matrices. Theory and Computation (SIAM, Philadelphia, PA, 2008).

[4] A. Frommer and V. Simoncini, Matrix functions, in: Model Order Reduction: Theory, Research Aspects and Applications, edited by W. H. A. Schilders et al. (Springer-Verlag, Berlin, 2008), pp. 275–303.

[5] K. Gallivan, E. Grimme, and P. Van Dooren, Numer. Algorithms **12**, 33–63 (1996).

[6] R. U. Börner, O. G. Ernst, and K. Spitzer, Geophys. J. Int. **173**, 766–780 (2008).

[7] L. A. Knizhnerman, V. L. Druskin, and M. Zaslavsky, SIAM J. Numer. Anal. **47**, 953–971 (2009).

[8] M. Hochbruck and C. Lubich, SIAM J. Numer. Anal. **34**, 1911–1925 (1997).

[9] M. Hochbruck, C. Lubich, and H. Selhofer, SIAM J. Sci. Statist. Comput. **19**, 1552–1574 (1998).

[10] S. M. Cox and P. C. Matthews, J. Comput. Phys. **176**, 430–455 (2002).

[11] A. K. Kassam and L. N. Trefethen, SIAM J. Sci. Comput. **26**, 1214–1233 (2005).

[12] B. V. Minchev and W. M. Wright, A review of exponential integrators for first order semi-linear problems, Preprint Numerics No. 2/2005, Norges Teknisk-Naturvitenskapelige Universitet, Trondheim, Norway, 2005.

[13] T. Schmelzer and L. N. Trefethen, Electron. Trans. Numer. Anal. **29**, 1–18 (2007).

[14] M. Hochbruck and A. Ostermann, Acta Numerica **19**, 209–286 (2010).

[15] B. Philippe and R. B. Sidje, Transient solutions of Markov processes by Krylov subspaces, Research Report RR-1989, INRIA, Rennes, France, 1993.

[16] Y. Saad, Preconditioned Krylov subspace methods for the numerical solution of Markov chains, in: Computations with Markov Chains, edited by W. J. Stewart (Kluwer Academic Publishers, Boston, MA, 1995).

[17] R. B. Sidje and W. J. Stewart, Comput. Statist. Data Anal. **29**, 345–368 (1999).

[18] E. Estrada, N. Hatano, and M. Benzi, to appear in Physics Reports (2012).

[19] M. Benzi, E. Estrada, and C. Klymko, Ranking hubs and authorities using matrix functions, Technical Report TR-2012-003, Emory University, Georgia, USA, 2012.

[20] I. Najfeld and T. F. Havel, Adv. in Appl. Math. **16**, 321–375 (1995).

[21] A. Nauts and R. E. Wyatt, Phys. Rev. Lett. **51**, 2238–2241 (1983).

[22] V. L. Druskin and L. A. Knizhnerman, Radio Science **29**, 937–953 (1994).

[23] V. L. Druskin, L. A. Knizhnerman, and M. Zaslavsky, SIAM J. Sci. Comput. **31**, 3760–3780 (2009).

[24] V. Druskin, C. Lieberman, and M. Zaslavsky, SIAM J. Sci. Comput. **32**, 2485–2496 (2010).

[25] V. Grimm and M. Hochbruck, BIT **48**, 215–229 (2008).

[26] E. J. Allen, Dyn. Contin. Discrete Impuls. Systems **5**, 271–281 (1999).

[27] K. Burrage, N. Hale, and D. Kay, SIAM J. Sci. Comp. (2012), to appear.

[28] W. D. Sharp and E. J. Allen, Annals of Nuclear Energy **27**, 99–116 (2000).

[29] E. J. Allen, J. Baglama, and S. K. Boyd, Linear Algebra Appl. **310**, 167–181 (2000).

[30] M. Arioli and D. Loghin, Matrix square-root preconditioners for the Steklov–Poincaré operator, Technical Report RAL-TR-2008-003, Rutherford Appleton Laboratory, Didcot, UK, 2008.

[31] J. van den Eshof, A. Frommer, T. Lippert, K. Schilling, and H. van der Vorst, Comput. Phys. Commun. **146**, 203–224 (2002).

[32] J. C. R. Bloch and S. Heybrock, Comput. Phys. Commun. **182**, 878–889 (2011).

[33] W. J. Cody, G. Meinardus, and R. S. Varga, J. Approx. Theory **2**, 50–65 (1969).

[34] L. N. Trefethen, J. A. C. Weideman, and T. Schmelzer, BIT **46**, 653–670 (2006).

[35] T. Schmelzer and L. N. Trefethen, SIAM J. Numer. Anal. **45**, 558–571 (2007).

[36] N. Hale, N. J. Higham, and L. N. Trefethen, SIAM J. Numer. Anal. **46**, 2505–2523 (2008).

[37] V. Grimm and M. Gugat, SIAM J. Numer. Anal. **48**, 1826–1845 (2010).

[38] S. Güttel, Rational Krylov Methods for Operator Functions, PhD thesis, Institut für Numerische Mathematik und Optimierung der Technischen Universität Bergakademie Freiberg, 2010.

[39] A. Ruhe, Linear Algebra Appl. **58**, 391–405 (1984).

[40] A. Ruhe, IMA Vol. Math. Appl. **60**, 149–164 (1994).

[41] B. Beckermann and L. Reichel, SIAM J. Numer. Anal. **47**, 3849–3883 (2009).

[42] K. Deckers and A. Bultheel, Rational Krylov sequences and orthogonal rational functions, Tech. Rep. TW499, Katholieke Universiteit Leuven, Department of Computer Science, Leuven, Belgium, 2007.

[43] V. L. Druskin and L. A. Knizhnerman, USSR Comput. Maths. Math. Phys. **29**, 112–121 (1989).

[44] T. Ericsson, Computing functions of matrices using Krylov subspace methods, Technical Report, Chalmers University of Technology, Department of Computer Science, Göteborg, Sweden, 1990.

[45] Y. Saad, SIAM J. Numer. Anal. **29**, 209–228 (1992).

[46] V. L. Druskin and L. A. Knizhnerman, SIAM J. Matrix Anal. Appl. **19**, 775–771 (1998).

[47] R. W. Freund and M. Hochbruck, Gauss quadratures associated with the Arnoldi process and the Lanczos algorithm, in: Linear Algebra for Large Scale and Real-Time Applications, edited by M. S. Moonen et al. (Kluwer Academic Publishers, Boston, MA, 1993), pp. 377–380.

[48] G. H. Golub and G. Meurant, Matrices, Moments and Quadrature with Applications (Princeton University Press, Princeton, 2010).

[49] G. L. Lagomasino, L. Reichel, and L. Wunderlich, Linear Algebra Appl. **429**, 2540–2554 (2008).

[50] K. Deckers, Orthogonal Rational Functions: Quadrature, Recurrence and Rational Krylov, PhD thesis, Departement Computerwetenschappen, Katholieke Universiteit Leuven, 2009.

[51] C. Jagels and L. Reichel, Linear Algebra Appl. **434**, 1716–1732 (2011).

[52] M. Crouzeix, J. Funct. Anal. **244**, 668–690 (2007).

[53] B. Beckermann, S. Güttel, and R. Vandebril, SIAM J. Matrix Anal. Appl. **31**, 1740–1774 (2010).

[54] B. Beckermann and S. Güttel, Numer. Math. (2011), to appear.

[55] A. B. J. Kuijlaars, SIAM J. Matrix Anal. Appl. **22**, 306–321 (2000).

[56] B. Beckermann and A. B. J. Kuijlaars, SIAM J. Numer. Anal. **39**, 300–329 (2001).

[57] V. Simoncini, SIAM J. Sci. Comput. **29**, 1268–1288 (2007).

[58] C. Jagels and L. Reichel, Linear Algebra Appl. **431**, 441–458 (2009).

[59] L. A. Knizhnerman and V. Simoncini, Numer. Linear Algebra Appl. **17**, 615–638 (2010).

[60] V. Druskin, A. Greenbaum, and L. Knizhnerman, SIAM J. Sci. Comput. **19**, 38–54 (1998).

[61] I. Moret and P. Novati, BIT **44**, 595–615 (2004).

[62] J. Eshof and M. Hochbruck, SIAM J. Sci. Comput. **27**, 1438–1457 (2006).

[63] M. Caliari, M. Vianello, and L. Bergamaschi, J. Comput. Appl. Math. **172**, 79–99 (2004).

[64] L. Bergamaschi, M. Caliari, and M. Vianello, The ReLPM exponential integrator for FE discretizations of advection–diffusion equations, in: Computational Science – ICCS 2004, edited by M. Bubak et al., Lecture Notes in Computer Science Vol. 3039 (Springer-Verlag, Berlin, 2004), pp. 434–442.

[65] M. Caliari, M. Vianello, and L. Bergamaschi, J. Comput. Appl. Math. **210**, 56–63 (2007).

[66] M. Eiermann and O. G. Ernst, SIAM J. Numer. Anal. **44**, 2481–2504 (2006).

[67] M. Afanasjew, M. Eiermann, O. G. Ernst, and S. Güttel, Electron. Trans. Numer. Anal. **28**, 206–222 (2008).

[68] M. Afanasjew, M. Eiermann, O. G. Ernst, and S. Güttel, Linear Algebra Appl. **429**, 2293–2314 (2008).

[69] M. Eiermann, O. G. Ernst, and S. Güttel, SIAM J. Matrix Anal. Appl. **32**, 621–641 (2011), SIAM J. Matrix Anal. Appl.

[70] T. Ransford, Potential Theory in the Complex Plane (Cambridge University Press, Cambridge, UK, 1995).

[71] E. B. Saff and V. Totik, Logarithmic Potentials with External Fields (Springer-Verlag, Berlin, 1997).

[72] E. Levin and E. B. Saff, Potential theoretic tools in polynomial and rational approximation, in: Harmonic Analysis and Rational Approximation, edited by J.-D. Fournier et al., Lecture Notes in Control and Information Sciences Vol. 327 (Springer-Verlag, Berlin, 2006), pp. 71–94.

[73] T. Bagby, J. Math. Mech. **17**, 315–329 (1967).

[74] A. A. Gonchar, Math. USSR Sb. **7**, 623–635 (1969).

[75] A. L. Levin and E. B. Saff, Constr. Approx. **10**, 235–273 (1994).

[76] E. I. Zolotarev, Zap. Imp. Akad. Nauk St. Petersburg **30**, 1–59 (1877), In Russian.

[77] J. Todd, Applications of transformation theory: A legacy from Zolotarev (1847–1878), in: Approximation Theory and Spline Functions, edited by S. P. Singh (D. Reidel Publishing, Dordrecht, Netherlands, 1984), pp. 207–245.

[78] P. Henrici, Applied and Computational Complex Analysis, Volume 1 (John Wiley & Sons, New York, 1988).

[79] T. Bagby, Duke Math. J. **36**, 95–104 (1969).

[80] O. G. Parfenov, Math. USSR Sb. **59**, 497–514 (1988).

[81] V. A. Prokhorov, Sb. Math. **184**, 3–32 (1993).

[82] V. A. Prokhorov, J. Approx. Theory **133**, 284–296 (2005).

[83] A. A. Gonchar, Math. USSR Sb. **34**, 131–145 (1978).

[84] H. Stahl, General convergence results for rational approximants, in: Approximation Theory VI, edited by C. K. Chui, L. L. Schumaker, and J. D. Ward (Academic Press, Boston, MA, 1989), pp. 605–634.

[85] V. I. Lebedev, USSR Comput. Maths. Math. Phys. **17**, 58–76 (1977).

[86] N. S. Ellner and E. L. Wachspress, SIAM J. Numer. Anal. **28**, 859–870 (1991).

[87] N. Levenberg and L. Reichel, Numer. Math. **66**, 215–233 (1993).

[88] G. Starke, SIAM J. Numer. Anal. **28**, 1431–1445 (1991).

[89] Z. Bai, Appl. Numer. Math. **43**, 9–44 (2002).

[90] R. W. Freund, Acta Numer. **12**, 267–319 (2003).

[91] A. Ruhe, BIT **34**, 165–176 (1994).

[92] O. Axelsson and A. Kucherov, Numer. Linear Algebra Appl. **7**, 197–218 (2000).

[93] B. Le Bailly and J. P. Thiran, SIAM J. Numer. Anal. **38**, 1409–1424 (2000).

[94] S. Gugercin, A. C. Antoulas, and C. Beattie, SIAM J. Matrix Anal. Appl. **30**, 609–638 (2008).

[95] V. Druskin and V. Simoncini, Systems & Control Letters **60**, 546–560 (2011).

[96] E. B. Saff, A. Schönhage, and R. S. Varga, Numer. Math. **25**, 307–322 (1976).

[97] T. C. Y. Lau, BIT **17**, 191–199 (1977).

[98] S. P. Nørsett and A. Wolfbrandt, BIT **17**, 200–208 (1977).

[99] S. M. Serbin, SIAM J. Sci. Stat. Comput. **13**, 449–458 (1992).

[100] P. B. Borwein, J. Approx. Theory **38**, 279–283 (1983).

[101] J. E. Andersson, J. Approx. Theory **32**, 85–95 (1981).

[102] M. Popolizio and V. Simoncini, SIAM J. Matrix Anal. Appl. **30**, 657–683 (2008).

[103] V. Grimm, BIT (2011).

[104] P. Novati, SIAM J. Matrix Anal. Appl. **32**, 1537–1558 (2011).

[105] H. Voss, BIT Numer. Math. **44**, 387–401 (2004).

[106] E. Jarlebring and H. Voss, Appl. Math. **50**, 543–554 (2005).

[107] C. Beattie and S. Gugercin, Systems & Control Letters **58**, 225–232 (2009).

[108] U. Baur, C. A. Beattie, P. Benner, and S. Gugercin, SIAM J. Sci. Comput. **33**, 2489–2518 (2011).

[109] M. Zaslavsky and V. Druskin, J. Comput. Phys. **229**, 4831–4839 (2010).

[110] V. Druskin and M. Zaslavsky, Linear Algebra Appl. **436**, 38833903 (2012).

[111] S. Güttel and L. Knizhnerman, A black-box rational Arnoldi variant for the approximation of Cauchy–Stieltjes matrix functions, Tech. Rep. 12/03, University of Oxford, 2012, submitted to SIAM J. Sci. Comput.

[112] L. N. Trefethen, Approximation Theory and Approximation Practice (SIAM, to appear 2013).