UNIFORM APPROXIMATION OF φ-FUNCTIONS IN EXPONENTIAL INTEGRATORS BY A RATIONAL KRYLOV SUBSPACE METHOD WITH SIMPLE POLES

TANJA GÖCKLER AND VOLKER GRIMM*

Abstract. We consider the approximation of the matrix φ-functions that appear in exponential integrators for stiff systems of differential equations. For stiff systems, the field-of-values of the occurring matrices is large and lies somewhere in the left complex half-plane. In order to obtain an efficient method uniformly for all matrices with a field-of-values in the left complex half-plane, we consider the approximation by a rational Krylov subspace method with equidistant poles of order one on the line $\operatorname{Re} z = \gamma > 0$. We present error bounds that predict a faster convergence rate as for the resolvent Krylov subspace approximation using a single repeated pole at $\gamma > 0$. Poles of order one allow moreover for a parallel implementation of the corresponding rational Krylov subspace decomposition. We analyze the convergence of the proposed rational Krylov subspace method and present numerical experiments that illustrate our results.

Key words. Rational approximation, rational Krylov subspace method, exponential integrator, φ-functions, parallel method

AMS subject classifications. (2010) 65F60, 65M15, 65M22.

1. Introduction. Exponential integrators form an interesting class of numerical methods for the time integration of stiff ordinary differential equations of the form

$$y'(t) = Ay(t) + g(t, y(t)), \quad y(t_0) = y_0 \tag{1.1}$$

with a large matrix A. Such problems typically arise from a semi-discretization of a partial differential equation, for instance, by finite-difference, finite-element, or pseudospectral methods. The simplest exponential integrator is the exponential Euler method that reads

$$y(t_0 + \tau) \approx y_0 + \tau\varphi_1(\tau A)\big(Ay_0 + g(t_0, y_0)\big), \quad \varphi_1(z) = \frac{e^z - 1}{z},$$

where τ denotes the time step size and φ_1 is an entire function.

Stiff ordinary differential equations might be characterized by a huge field-of-values $W(A)$ of the matrix A in the left complex half-plane. Such matrices A we therefore call "stiff" in the following. Since explicit integrators have a small stability region in the left complex half-plane, they usually fail to integrate stiff differential equations unless very small time steps are used. In contrast, exponential integrators allow for the application of explicit schemes without a severe step size restriction. The basic idea of these integrators is to treat the linear part $Ay(t)$ in (1.1) exactly, and to use an appropriate approximation for the nonlinear remainder g. More information about the integration of stiff ordinary differential equations by exponential integrators can be found in the review [21] by Hochbruck and Ostermann.

For the application of more general exponential integrators than the exponential Euler method, the computation of $\varphi_\ell(\tau A)v$ is required, where v is a vector, A is a stiff matrix, and φ_ℓ is one of the so-called φ-functions. These matrix functions are given by

$$\varphi_\ell(\tau A) := \int_0^1 e^{(1-s)\tau A} \frac{s^{\ell-1}}{(\ell-1)!}\, ds = \frac{1}{\tau^\ell} \int_0^\tau e^{(\tau-s)A} \frac{s^{\ell-1}}{(\ell-1)!}\, ds, \quad \ell \geq 1. \tag{1.2}$$

The efficient and reliable computation of $\varphi_\ell(\tau A)v$ is an important ingredient in every exponential integrator. For stiff ordinary differential equations, one therefore needs to find methods to compute these matrix functions times a vector efficiently for all matrices with a field-of-values in the left complex half-plane. Recently, the use of rational Krylov subspaces for the approximation of

* Karlsruhe Institute of Technology (KIT), Institut für Angewandte und Numerische Mathematik, D–76128 Karlsruhe, Germany, {tanja.goeckler | volker.grimm}@kit.edu.

$f(A)v$, where f is a function defined for matrices and/or operators, has been studied by a number of papers, e.g. in [2, 3, 5–17, 22–28, 30–33, 35]. Rational Krylov subspaces are of the form

$$\mathcal{Q}_m(A,v) = \left\{ \frac{p_{m-1}(A)}{q_{m-1}(A)}\, v,\ p_{m-1} \in \mathcal{P}_{m-1} \right\},$$

where $\mathcal{P}_{m-1}$ designates the space of polynomials of degree smaller than or equal to $m-1$ and the denominator $q_{m-1} \in \mathcal{P}_{m-1}$ is a fixed chosen polynomial. The choice of q_{m-1} determines the properties of the rational Krylov subspace approximation. Of course, the roots of q_{m-1} must not coincide with eigenvalues of the matrix A and therefore have to be located in the right complex half-plane, since otherwise the matrix functions in the rational Krylov subspace were not defined for arbitrary matrices with a field-of-values in the left complex half-plane. With the help of the standard (polynomial) Krylov subspace

$$\mathcal{K}_m(A,v) = \{ p_{m-1}(A)v,\ p_{m-1} \in \mathcal{P}_{m-1} \},$$

the rational Krylov subspace can alternatively be expressed as

$$\mathcal{Q}_m(A,v) = q_{m-1}(A)^{-1}\mathcal{K}_m(A,v) = \mathcal{K}_m(A, q_{m-1}(A)^{-1}v).$$

In recent works on the approximation of matrix functions by Krylov subspace methods, it becomes more and more apparent that, for stiff matrices A, rational Krylov subspace methods work tremendously better than standard Krylov subspace methods. There is a strong analogy to the behavior of the explicit and implicit Euler method for stiff systems. This can be seen, if we consider the stiff problem

$$y'(t) = Ay(t) + \frac{t^{\ell-1}}{(\ell-1)!}\, v, \quad y(0) = 0, \tag{1.3}$$

whose exact solution at time τ is given as $y(\tau) = \tau^\ell \varphi_\ell(\tau A)v$. In the following numerical experiments, $A \in \mathbb{R}^{N^2 \times N^2}$ is the standard finite-difference discretization matrix for the two-dimensional Laplacian on the unit square with homogeneous Dirichlet boundary conditions, where we use a regular grid with N^2 inner discretization points and mesh size $\frac{1}{N+1}$. The vector v contains the evaluations of the function $30 \cdot x(1-x)y(1-y)$, $(x,y) \in [0,1]^2$, at the inner grid points. By applying the explicit Euler method with step size $\frac{\tau}{m}$, we approximate $y(\tau) = y(t_m)$ by

$$y_m = \frac{\tau}{m} \sum_{p=0}^{m-1} \left(I + \frac{\tau}{m}A\right)^p \cdot \frac{t_{m-p-1}^{\ell-1}}{(\ell-1)!}\, v \in \mathcal{K}_m(A,v), \quad t_k = k \cdot \frac{\tau}{m}, \quad k = 0, \ldots, m,$$

which can be interpreted as a fixed polynomial approximation to $\tau^\ell \varphi_\ell(\tau A)v$ in $\mathcal{K}_m(A,v)$. Since the terms $\|I + \frac{\tau}{m}A\|^p$, $p = 0, \ldots, m-1$, have to remain bounded for $m \to \infty$, we have to ensure that $\|I + \frac{\tau}{m}A\| \leq 1$. As a result, the explicit method requires very small time steps in order to achieve the desired accuracy and stability of the method. This can be seen in Figure 1.1. Where the red diamond-marked line appears, it corresponds to the obtained error for the approximation of $\varphi_1(\tau A)v$ for $\tau = 0.025$ versus the computation time. Where the red diamond-marked line does not appear, especially on the right-hand side of Figure 1.1, a reasonable approximation of the solution in the given computation time was not possible. The blue circle-marked line corresponds to the standard Krylov subspace approximation of $\varphi_1(\tau A)v$ in $\mathcal{K}_m(A,v)$. The approximation is significantly better for small matrices (e.g. [20]), but for only slightly larger, mildly stiff matrices, the method deteriorates and for stiff matrices an efficient approximation is not possible at all. The polynomial Krylov subspace approximation behaves analogously to the explicit Euler method for stiff matrices with respect to accuracy and computation time.

In contrast to this, the implicit Euler method, where the rational approximation

$$y_m = \frac{\tau}{m} \sum_{p=1}^{m} \left(I - \frac{\tau}{m}A\right)^{-p} \cdot \frac{t_{m-p+1}^{\ell-1}}{(\ell-1)!}\, v \in \mathcal{Q}_m(A,v), \quad t_k = k \cdot \frac{\tau}{m}, \quad k = 0, \ldots, m,$$

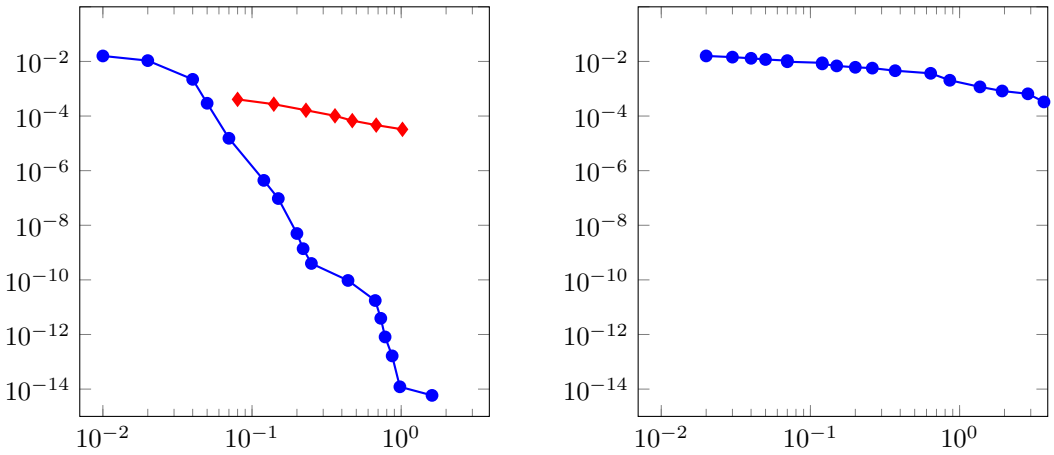

Fig. 1.1: Accuracy of the explicit Euler method (red diamond-marked line) and of the standard Krylov subspace method (blue circle-marked line) for a $3\,969 \times 3\,969$ - matrix (on the left-hand side) and a $65\,025 \times 65\,025$ - matrix (on the right-hand side) versus the computation time in seconds for $\tau = 0.025$ and $\ell = 1$.

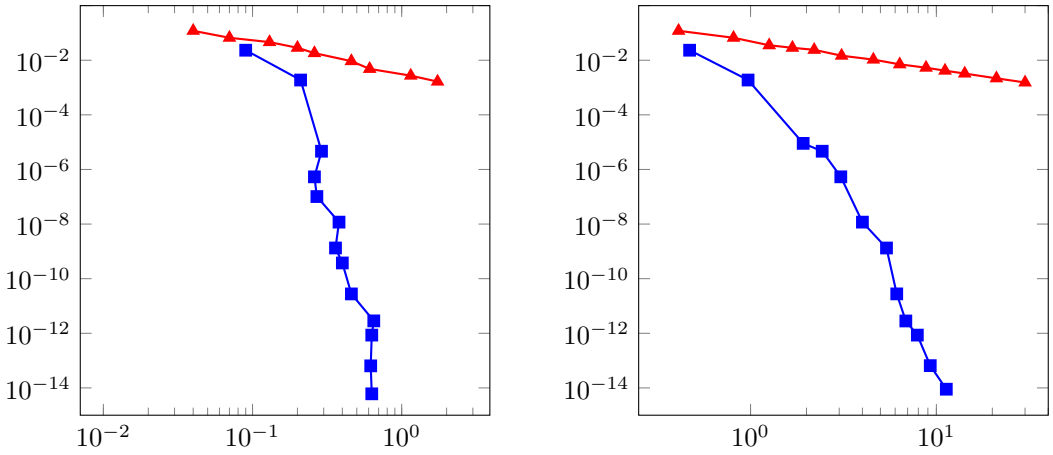

Fig. 1.2: Accuracy of the implicit Euler method (red triangle-marked line) and of the rational Krylov subspace method (blue square-marked line) for a $65\,025 \times 65\,025$ - matrix (on the left-hand side) and a $1\,046\,529 \times 1\,046\,529$ - matrix (on the right-hand side) versus the computation time in seconds for $\tau = 0.025$ and $\ell = 1$.

to $y(\tau)$ is used, works well even for much larger time steps in case of a stiff problem. The error curve of the implicit Euler method versus the computation time in seconds is shown with the red triangle-marked line in Figure 1.2. This time, the approximation lies in the rational Krylov subspace $\mathcal{Q}_m(A, v) = q_{m-1}(A)^{-1}\mathcal{K}_m(A, v)$ with $q_{m-1}(z) = (\gamma - z)^{m-1}$ and $\gamma = \frac{m}{\tau}$. With the choice $\gamma = \frac{m}{\tau}$, it would be necessary to build the whole space from scratch, if one wants to rise the dimension of the approximation space. In order to avoid this and to obtain a space that is augmented by just one vector from step to step, one usually chooses a fixed γ (e.g. [12, 13]). The performance of the approximation in the rational Krylov subspace with $\gamma = \frac{1}{\tau}$ corresponds to the blue square-marked line in Figure 1.2. That the rational Krylov method, which is inspired by the implicit Euler method, outperforms the implicit Euler method and is more suited to the computation of the φ-functions for stiff matrices can be clearly seen.

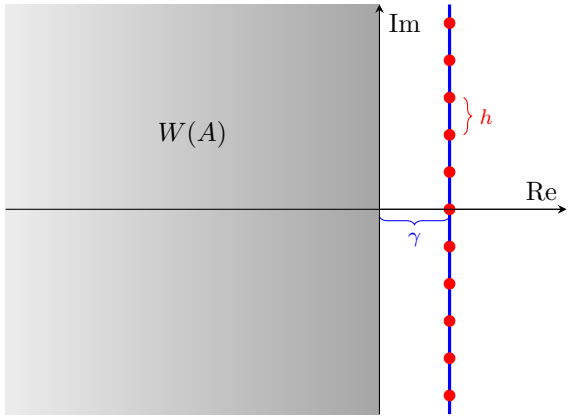

Fig. 1.3: Locus of the field-of-values $W(A)$ of the stiff matrix A, as well as the position of the simple roots of the denominator polynomial q_{2m+1} of the rational Krylov subspace $\mathcal{Q}_{2m+2}(A, v)$.

For the approximation to $\varphi_\ell(A)v$ uniformly for stiff matrices in the rational Krylov subspace $\mathcal{Q}_m(A, v) = (\gamma I - A)^{-(m-1)}\mathcal{K}_m(A, v)$ as above, the sublinear bound

$$\inf_{R \in \mathcal{Q}_m(A,v)} \|\varphi_\ell(A)v - R\| \leq \frac{C}{m^{\frac{\ell}{2}}} \|v\|, \quad \ell \geq 1$$

has been proved in [13]. In the following, we present a Krylov subspace method that allows for faster convergence rates, which are uniform for all stiff matrices according to our setting. Furthermore, the new method can be easily parallelized in time. In order to achieve these goals, we consider an approximation method based on the special rational Krylov subspace

$$\mathcal{Q}_{2m+2}(A, v) = \left\{ \frac{p_{2m+1}(A)}{q_{2m+1}(A)} v, \ p_{2m+1} \in \mathcal{P}_{2m+1} \right\},$$

where the denominator polynomial is chosen as

$$q_{2m+1}(z) = \prod_{k=-m}^{m} (z_k - z) \in \mathcal{P}_{2m+1}, \quad z_k = \gamma + ihk, \quad k = -m, \ldots, m, \quad \gamma > 0.$$

That is, the poles of the rational approximation are equidistant on the line $\operatorname{Re} z = \gamma > 0$ in the right complex half-plane, see Figure 1.3. Since the field-of-values of stiff matrices lies in the left complex half-plane, the rational matrix functions in $\mathcal{Q}_{2m+2}(A, v)$ are defined.

More exactly, we first obtain a uniform error estimate for the best approximation of $\varphi_\ell(A)$ in the rational matrix space

$$\mathcal{R}_{2m+1}(A) = \left\{ \frac{p_{2m}(A)}{q_{2m+1}(A)}, \ p_{2m} \in \mathcal{P}_{2m} \right\}. \tag{1.4}$$

This result can then be used to bound the error of the rational Krylov subspace approximation of $\varphi_\ell(A)v$ in $\mathcal{Q}_{2m+2}(A, v)$. Since the approximation of $\varphi_\ell(A)$ in $\mathcal{R}_{2m+1}(A)$ is uniform for stiff matrices, this holds also true for the rational Krylov subspace approximation of $\varphi_\ell(A)v$.

Applying partial fraction expansion, it turns out that this rational Krylov subspace can alternatively be written as

$$\mathcal{Q}_{2m+2}(A, v) = \operatorname{span}\left\{ v, \frac{1}{z_{-m} - A} v, \frac{1}{z_{-m+1} - A} v, \ldots, \frac{1}{z_m - A} v \right\}. \tag{1.5}$$

In contrast to rational Krylov subspace methods with one multiple pole, one can compute the basis of this rational Krylov subspace in parallel, by assigning to each node the solution of the

linear system $(z_k I - A)x = v$. If the stiff matrix A is large, which is the case for nearly all fine discretizations of evolution equations, the parallel computation of the basis of $\mathcal{Q}_{2m+2}(A, v)$ leads to a significant speed-up compared to standard methods.

The paper is organized as follows: After some preliminaries in Section 2, we discuss the approximation of the matrix φ-functions in the rational matrix subspace of the form (1.4) in Section 3. The approximation in the corresponding Krylov subspace is discussed in Section 4 and the choice of the free parameters in Section 5. Finally, numerical experiments are presented in Section 6.

2. Preliminaries. In this paper, we consider stiff matrices $A \in \mathbb{C}^{N\times N}$ of arbitrary dimension N with a field-of-values

$$W(A) = \{(Ax, x) \,|\, x \in \mathbb{C}^N, \|x\| = 1\}$$

in the left complex half-plane $H_0^- := \{z \in \mathbb{C} \mid \operatorname{Re} z \leq 0\}$. Hereby, $(\cdot\,,\cdot)$ is an inner product, the vector space $\mathbb{C}^N$ has been equipped with, and $\|\cdot\|$ designates its induced norm. As matrix norm, we always choose the induced matrix norm, which we again designate by $\|\cdot\|$. One might also characterize our matrices by

$$\operatorname{Re}(Ax, x) \leq 0\,, \quad \forall\, x \in \mathbb{C}^N\,.$$

Such matrices play an important role in the stability theory of differential equations. For simplicity, we only consider matrices, but all results in the paper can be generalized to operators A on Hilbert spaces and some results to operators A on Banach spaces.

We very often use notations that are common in the analysis of matrix functions. For example, we may write $\frac{1}{A}$, meaning the matrix A inserted in the function $\frac{1}{z}$, instead of A^{-1} provided that the considered function is defined on the spectrum of A. A comprehensive overview of the theory of matrix functions can be found in the book [19] by Higham.

A result that relates the field-of-values of A to the stability of the matrix exponential is recalled in the following lemma.

LEMMA 2.1. *Let A be a matrix with $W(A) \subseteq H_0^-$ and let $\tau \geq 0$ be arbitrarily chosen, then*

$$\|e^{\tau A}\| \leq 1\,.$$

Proof. For the convenience of the reader, we first proof that

$$\left\| \frac{1}{z - A} \right\| \leq \frac{1}{\operatorname{dist}(z, W(A))}\,, \quad \forall\, z \notin W(A)\,, \tag{2.1}$$

where $\operatorname{dist}(z, W(A))$ designates the minimum distance of z from $W(A)$ for a general inner product. For the Euclidean norm, estimate (2.1) is well-known (e.g. [29], Theorem 1.19) and can be easily transferred to general inner products. Let $z_0 \notin W(A)$ be arbitrarily chosen. Then, we have for any $0 \neq x \in \mathbb{C}^N$ that $(Ax, x)/(x, x) \in W(A)$ and therefore

$$\operatorname{dist}(z_0, W(A)) \leq \left| z_0 - \frac{(Ax, x)}{(x, x)} \right| = \frac{|((z_0 I - A)x, x)|}{\|x\|^2}$$

or

$$\operatorname{dist}(z_0, W(A)) \cdot \|x\|^2 \leq |((z_0 I - A)x, x)|\,. \tag{2.2}$$

Since $z_0 \notin W(A)$ and thus $z_0 \notin \Lambda(A) \subseteq W(A)$, we know that $z_0 I - A$ is invertible. If one chooses an arbitrary $0 \neq y \in \mathbb{C}^N$, then it follows by substituting $0 \neq x = (z_0 I - A)^{-1} y$ in (2.2) that

$$\operatorname{dist}(z_0, W(A)) \cdot \|(z_0 I - A)^{-1} y\|^2 \leq |(y, (z_0 I - A)^{-1} y)| \leq \|(z_0 I - A)^{-1} y\| \|y\|$$

which is equivalent to

$$\frac{\|(z_0 I - A)^{-1} y\|}{\|y\|} \leq \frac{1}{\operatorname{dist}(z_0, W(A))}\,.$$

Because $0 \neq y \in \mathbb{C}^N$ can be chosen arbitrarily, we also have

$$\|(z_0 I - A)^{-1}\| = \sup_{0 \neq y \in \mathbb{C}^N} \frac{\|(z_0 I - A)^{-1} y\|}{\|y\|} \leq \frac{1}{\operatorname{dist}(z_0, W(A))}\,.$$

With the help of estimate (2.1) we can conclude for $\tau > 0$, $m \in \mathbb{N}$, and $W(A) \subseteq H_0^-$ that

$$\left\| \frac{1}{1 - \frac{\tau}{m} A} \right\| = \frac{m}{\tau} \left\| \frac{1}{\frac{m}{\tau} - A} \right\| \leq 1 \quad \Rightarrow \quad \left\| \left(\frac{1}{1 - \frac{\tau}{m} A} \right)^m \right\| \leq 1\,. \tag{2.3}$$

The convergence of the implicit Euler method now gives

$$1 \geq \lim_{m \to \infty} \left\| \left(\frac{1}{1 - \frac{\tau}{m} A} \right)^m \right\| = \left\| \lim_{m \to \infty} \left(\frac{1}{1 - \frac{\tau}{m} A} \right)^m \right\| = \left\| e^{\tau A} \right\|\,,$$

where the first equality follows by the reverse triangle inequality. For $\tau = 0$, the assertion is obtained immediately from $e^{0 \cdot A} = I$. □

Inequality (2.3) shows, i.a., that the implicit Euler method is A-stable, since the whole left complex half-plane belongs to the stability region.

A simple integral transform of (1.2) shows that the matrix φ-functions can also be written as

$$\varphi_\ell(A) = \int_0^\infty e^{sA} \cdot \mathbf{1}_{[0,1]}(s) \frac{(1-s)^{\ell-1}}{(\ell-1)!}\, ds\,, \quad \ell \geq 1\,, \tag{2.4}$$

where $\mathbf{1}_{[0,1]}(s)$ denotes the indicator function that is equal to 1 for $s \in [0,1]$ and 0 elsewhere. The rational matrix functions $r(A) \in \mathcal{R}_{2m+1}(A)$, that we intend to use for the approximation, can be represented by a similar integral. Before we can state this integral representation, we need to remark that, via partial fraction expansion, the space $\mathcal{R}_{2m+1}(A)$, as given in (1.4), can be written in an alternative form.

LEMMA 2.2. *We have*

$$\mathcal{R}_{2m+1}(A) = \operatorname{span}\left\{ \frac{1}{z_{-m} - A}, \ldots, \frac{1}{z_m - A} \right\}\,, \quad z_k = \gamma + ihk\,, \quad k = -m, \ldots, m\,.$$

Proof. Since for stiff matrices the matrix functions are defined for both spaces, it remains to show that we have for $z \in \mathbb{C}$ the equality of the following two function sets

$$\frac{\mathcal{P}_{2m}}{q_{2m+1}} = \left\{ \frac{p_{2m}(z)}{q_{2m+1}(z)},\ p_{2m} \in \mathcal{P}_{2m} \right\} = \operatorname{span}\left\{ \frac{1}{z_{-m} - z}, \ldots, \frac{1}{z_m - z} \right\} := M\,, \tag{2.5}$$

where $q_{2m+1}(z) = \prod_{k=-m}^m (z_k - z)$. Let $r(z) = \sum_{k=-m}^m a_k \frac{1}{z_k - z} \in M$ be arbitrary. A simple calculation shows that

$$r(z) = \sum_{k=-m}^m a_k \frac{1}{z_k - z} = \frac{1}{q_{2m+1}(z)} \sum_{k=-m}^m a_k \prod_{\substack{l=-m \\ l \neq k}}^m (z_l - z) \in \frac{\mathcal{P}_{2m}}{q_{2m+1}}\,.$$

Vice versa, let $r(z) = p_{2m}(z)/q_{2m+1}(z)$. Partial fraction expansion yields

$$r(z) = -\sum_{k=-m}^m \frac{p_{2m}(z_k)}{q'_{2m+1}(z_k)} \cdot \frac{1}{z_k - z} = \sum_{k=-m}^m a_k \frac{1}{z_k - z} \quad \text{with} \quad a_k = -\frac{p_{2m}(z_k)}{q'_{2m+1}(z_k)}\,.$$

The validity of (2.5) can then be transferred from $z \in \mathbb{C}$ to matrices $A \in \mathbb{C}^{N\times N}$. ∎

Prepared by Lemma 2.2, we can state an integral representation for functions belonging to $\mathcal{R}_{2m+1}(A)$ similar to the representation of the matrix φ-functions in (2.4).

LEMMA 2.3. *The rational matrix functions that appear in the rational matrix subspace $\mathcal{R}_{2m+1}(A)$ for a matrix A with $W(A) \subseteq H_0^-$ can be written as*

$$\sum_{k=-m}^{m} a_k \frac{1}{\gamma + ihk - A} = \int_0^\infty e^{sA} \cdot \sum_{k=-m}^{m} a_k e^{-(\gamma+ihk)s}\, ds\,. \tag{2.6}$$

Proof. According to Lemma 2.2, $\mathcal{R}_{2m+1}(A)$ is the set of all linear combinations as they appear on the left-hand side of (2.6). We further obtain $\|e^{sA}e^{-(\gamma+ihk)s}\| \le |e^{-(\gamma+ihk)s}|\|e^{sA}\| \le e^{-\gamma s}$ for $s \ge 0$ by Lemma 2.1. Thus, the following improper Riemann integral exists and can be easily computed to

$$\int_0^\infty e^{sA}e^{-(\gamma+ihk)s}\, ds = \frac{1}{\gamma + ihk - A}\,.$$

Due to the linearity of the integration, the lemma is proved. ∎

The next lemma sounds simple, but it has the far-reaching consequence that all theorems that we prove for matrices A with $W(A) \subseteq H_0^-$ in the following hold true for all matrices τA, $\tau \ge 0$, with the same constants in the theorems.

LEMMA 2.4. *Let A be a matrix with $W(A) \subseteq H_0^-$, then also*

$$W(\tau A) \subseteq H_0^-, \quad \forall\, \tau \ge 0\,.$$

Proof. Let z be in $W(\tau A)$. Then, there exists an $x \in \mathbb{C}^N$ with $\|x\| = 1$, such that

$$z = ((\tau A)x, x) = \tau\,(Ax, x)\,.$$

Hence, $\operatorname{Re} z = \tau \operatorname{Re}(Ax, x) \le 0$ for $\tau \ge 0$. ∎

3. Approximation of the matrix φ-functions. We start with the question of how well the matrix φ-functions for stiff matrices can be approximated by rational matrix functions with poles of order one at $\gamma + ihk$, $k = -m, \dots, m$, on the line $\operatorname{Re} z = \gamma > 0$. More exactly, let $A \in \mathbb{C}^{N\times N}$ be a matrix with $W(A) \subseteq H_0^-$. We consider the approximation of $\varphi_\ell(A)$, $\ell \ge 1$, in the rational matrix subspace

$$\mathcal{R}_{2m+1}(A) = \left\{ \frac{p_{2m}(A)}{q_{2m+1}(A)},\ p_{2m} \in \mathcal{P}_{2m} \right\}, \quad q_{2m+1}(A) = \prod_{k=-m}^{m} (z_k I - A)\,, \quad z_k = \gamma + ihk\,.$$

This space can be written in short as $\mathcal{R}_{2m+1}(A) = \{r(A),\ r \in \mathcal{P}_{2m}/q_{2m+1}\}$.

We are interested in uniform bounds, i.e. bounds that are correct for all matrices A with a field-of-values in the left complex half-plane. The following theorem states a bound of that kind.

THEOREM 3.1. *Let A have a field-of-values in the left complex half-plane, i.e. $W(A) \subseteq H_0^-$, and choose the fixed denominator polynomial $q_{2m+1}(z) = \prod_{k=-m}^{m}(\gamma + ihk - z)$. Then, we have*

$$\inf_{r \in \frac{\mathcal{P}_{2m}}{q_{2m+1}}} \|\varphi_\ell(A) - r(A)\| \le C_1 \frac{e^{-\frac{\gamma\pi}{h}}}{1 - e^{-\frac{2\gamma\pi}{h}}} + C_2 \frac{1}{(hm)^\ell}\,, \quad \ell \ge 1\,, \tag{3.1}$$

where the coefficients C_1 and C_2 depend only on γ and ℓ.

The bound of the theorem well reflects the idea of the proof. The φ-functions along the imaginary axis are band-limited in Fourier space. They therefore allow for an approximation within a finite spectrum, which gives the second part of the error bound with constant C_2. In view of (2.6), the inverse $(z_k I - A)^{-1}$ can be seen as the Laplace transform of e^{sA}. Because of

the damping in the Laplace transform due to the shift $\gamma > 0$, the approximation in the remaining spectrum is exponentially damped, leading to the first part of the bound with constant C_1.

For the proof of Theorem 3.1, the concept of bounded variation is required. We say that a function f is of bounded variation on the unit circle $\mathbb{T} = [-\pi, \pi)$, consisting of $\mathbb{R}$ modulo 2π, if

$$\operatorname{Var}_{\mathbb{T}} f := \sup \sum_{k=1}^{n-1} |f(x_{k+1}) - f(x_k)| < \infty \,,$$

where the supremum is taken for all finite sequences $x_1 < x_2 < \ldots < x_n$, with $x_k \in \mathbb{T}$, $k = 1, \ldots, n$. Functions of bounded variation possess a countable number of discontinuities $\alpha_k \in \mathbb{T}$. In the following we also need a modified notion of Var given as $\operatorname{Var}^*_{\mathbb{T}} f := \operatorname{Var}_{\mathbb{T}} f^*$, where f^* is a correction of f such that $f^*(\alpha_k)$ is between $\lim_{s \nearrow \alpha_k} f(s)$ and $\lim_{s \searrow \alpha_k} f(s)$ for all interior discontinuities α_k (see [4], p. 17).

Proof. [of Theorem 3.1] Let A be an arbitrary matrix with a field-of-values in the left complex half-plane. Because of Lemma 2.2 above, we obtain

$$\inf_{r \in \frac{\mathcal{P}_{2m}}{q_{2m+1}}} \|\varphi_\ell(A) - r(A)\| = \inf_{a_{-m}, \ldots, a_m} \left\| \varphi_\ell(A) - \sum_{k=-m}^{m} a_k \, \frac{1}{\gamma + ihk - A} \right\| .$$

Due to the integral representations (2.4) and (2.6) and with the help of Lemma 2.1, we can reduce the approximation of $\varphi_\ell(A)$ to a one-dimensional approximation problem, that is

$$\begin{aligned} \left\| \varphi_\ell(A) - \sum_{k=-m}^{m} a_k \, \frac{1}{\gamma + ihk - A} \right\| &\le \int_0^\infty \|e^{sA}\| \left| \mathbf{1}_{[0,1]}(s) \, \frac{(1-s)^{\ell-1}}{(\ell-1)!} - \sum_{k=-m}^{m} a_k e^{-(\gamma+ihk)s} \right| ds \\ &\le \int_0^\infty \left| \mathbf{1}_{[0,1]}(s) \, \frac{(1-s)^{\ell-1}}{(\ell-1)!} - \sum_{k=-m}^{m} a_k e^{-(\gamma+ihk)s} \right| ds \,. \end{aligned}$$

We now assume without loss of generality that $h < \pi$, so that $\frac{\pi}{h} > 1$, and split the integral as

$$\begin{aligned} &\int_0^\infty \left| \mathbf{1}_{[0,1]}(s) \, \frac{(1-s)^{\ell-1}}{(\ell-1)!} - \sum_{k=-m}^{m} a_k e^{-(\gamma+ihk)s} \right| ds \qquad (3.2) \\ &= \int_0^{\frac{\pi}{h}} \left| \mathbf{1}_{[0,1]}(s) \, \frac{(1-s)^{\ell-1}}{(\ell-1)!} - \sum_{k=-m}^{m} a_k e^{-(\gamma+ihk)s} \right| ds + \sum_{l=1}^{\infty} \int_{(2l-1)\frac{\pi}{h}}^{(2l+1)\frac{\pi}{h}} \left| \sum_{k=-m}^{m} a_k e^{-(\gamma+ihk)s} \right| ds \,. \end{aligned}$$

In order to apply standard results from the theory of trigonometric approximation, we have to extend the first term containing the indicator function in a suitable manner to the interval $[-\frac{\pi}{h}, \frac{\pi}{h})$. Since we are interested in a best possible trigonometric approximation, we should ensure that the extended function has sufficient smoothness properties. For that purpose, we define a new function

$$g(s) := \begin{cases} C \displaystyle\int_{-1}^{s} e^{-\frac{1}{1-(2t+1)^2}} \, dt \,, & -1 < s < 0 \\ 1 \,, & s \ge 0 \\ 0 \,, & s \le -1 \end{cases} \qquad \text{with} \quad C = \left(\int_{-1}^{0} e^{-\frac{1}{1-(2t+1)^2}} \, dt \right)^{-1} , \qquad (3.3)$$

where $g \in C^\infty(\mathbb{R})$ and $|g(s)| \le 1$ for $s \in \mathbb{R}$. Furthermore, we set

$$f_\ell(s) := g(s) \cdot e^{\gamma s} \, \frac{(1-s)^{\ell-1}}{(\ell-1)!} \cdot \mathbf{1}_{[-1,1]}(s) \in C^{\ell-2}(\mathbb{R}) \,, \quad f_\ell \in C^\infty(\mathbb{R} \backslash \{1\}) \,.$$

One can check, that f_ℓ has a weak derivative of order $\ell - 1$ with jump discontinuity at the point 1.

Since the function g is defined such that $g(s) = 1$ for $s \geq 0$, we are now able to estimate

$$\begin{aligned}\int_0^{\frac{\pi}{h}} \left| \mathbf{1}_{[0,1]}(s)\, \frac{(1-s)^{\ell-1}}{(\ell-1)!} - \sum_{k=-m}^{m} a_k e^{-(\gamma+ihk)s} \right| ds &= \int_0^{\frac{\pi}{h}} e^{-\gamma s} \left| f_\ell(s) - \sum_{k=-m}^{m} a_k e^{-ihks} \right| ds \\ &\leq \int_0^{\frac{\pi}{h}} \left| f_\ell(s) - \sum_{k=-m}^{m} a_k e^{-ihks} \right| ds \\ &\leq \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \left| f_\ell(s) - \sum_{k=-m}^{m} a_k e^{-ihks} \right| ds\,.\end{aligned}$$

The coefficients a_k, $k = -m, \ldots, m$, are chosen such that

$$\begin{aligned}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \left| f_\ell(s) - \sum_{k=-m}^{m} a_k e^{-ihks} \right| ds &= \min_{b_{-m},\ldots,b_m} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \left| f_\ell(s) - \sum_{k=-m}^{m} b_k e^{-ihks} \right| ds \\ &= \frac{1}{h} \cdot \min_{b_{-m},\ldots,b_m} \int_{-\pi}^{\pi} \left| f_\ell\left(\frac{s}{h}\right) - \sum_{k=-m}^{m} b_k e^{-iks} \right| ds \\ &=: \frac{1}{h} \cdot E_m\left(f_\ell\left(\frac{\cdot}{h}\right)\right)_1\,,\end{aligned}$$

where $E_m(f)_1$ denotes the best trigonometric approximation of a periodic function f on $\mathbb{T}$ in the L_1-norm. Due to the fact that a best trigonometric approximation to the restriction $f_\ell\big|_{\mathbb{T}} : \mathbb{T} \to \mathbb{R}$ is a real trigonometric polynomial, the coefficients b_k can be restricted to coefficients that fulfill $b_{-k} = \overline{b_k}$. Consequently, the estimate of E_m is a standard real approximation problem for the function $f_\ell \in L_1(\mathbb{T})$, seen as a function on the torus. We will use the notations of the book [4] by DeVore and Lorentz. According to Theorem 2.3 in Chapter 7 on page 205 in [4], we have

$$E_m\left(f_\ell\left(\frac{\cdot}{h}\right)\right)_1 \leq C_\ell\, \omega_\ell\left(f_\ell\left(\frac{\cdot}{h}\right), \frac{1}{m}\right)_1$$

with a fixed constant C_ℓ that does not depend on f_ℓ, and $\omega_\ell(\cdot\,,\cdot)_1$ is the L_1-modulus of smoothness of order ℓ. According to Theorem 9.3 in Chapter 2 on page 53 and the definition of the modified variation Var^*, we further obtain

$$\omega_\ell\left(f_\ell\left(\frac{\cdot}{h}\right), \frac{1}{m}\right)_1 \leq \frac{1}{m^\ell} \cdot \mathrm{Var}^*_{\mathbb{T}}\, f_\ell^{(\ell-1)}\left(\frac{\cdot}{h}\right) \leq \frac{1}{m^\ell} \cdot \mathrm{Var}_{\mathbb{T}}\, u_\ell(\cdot)\,,$$

where $u_\ell(s) = f_\ell^{(\ell-1)}(\frac{s}{h}) = \frac{d^{\ell-1}}{ds^{\ell-1}}\left[f_\ell(\frac{s}{h})\right]$ for $s \in \mathbb{T}\backslash\{h\}$, and on the jump at h, $u_\ell(\cdot)$ is defined as the mean of the left and right limit of $f_\ell^{(\ell-1)}(\frac{\cdot}{h})$. The function u_ℓ is then of bounded variation on $\mathbb{T}$. We also define the transformed function $\tilde{u}_\ell(\cdot) = f_\ell^{(\ell-1)}(\cdot)$, where the function $\tilde{u}_\ell(\cdot)$ is again defined as the mean of the left and right limit at the jump at 1. Since $\tilde{u}_\ell(s) = 0$ for $s \leq -1$ and $s > 1$, it follows that

$$\mathrm{Var}_{\mathbb{T}}\, u_\ell(\cdot) = \frac{1}{h^{\ell-1}} \mathrm{Var}_{[-\frac{\pi}{h}, \frac{\pi}{h})}\, \tilde{u}_\ell(\cdot) = \frac{1}{h^{\ell-1}} \mathrm{Var}_{[-1,1]}\, \tilde{u}_\ell(\cdot)\,,$$

where the transformed function $\tilde{u}_\ell(\cdot)$ does not depend on h. Altogether, we have

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \left| f_\ell(s) - \sum_{k=-m}^{m} a_k e^{-ihks} \right| ds \leq C_2\, \frac{1}{(hm)^\ell}\,, \quad C_2 = C_\ell \mathrm{Var}_{[-1,1]}\, \tilde{u}_\ell(\cdot)$$

with the fixed constant C_2 which does depend only on ℓ and γ. Our first term in (3.2) is thus

bounded. For the second term, we obtain the immediate estimate

$$\sum_{l=1}^{\infty}\int_{(2l-1)\frac{\pi}{h}}^{(2l+1)\frac{\pi}{h}}\left|\sum_{k=-m}^{m}a_k e^{-(\gamma+ihk)s}\right|ds \le \sum_{l=1}^{\infty}e^{-\gamma(2l-1)\frac{\pi}{h}}\int_{(2l-1)\frac{\pi}{h}}^{(2l+1)\frac{\pi}{h}}\left|\sum_{k=-m}^{m}a_k e^{-ihks}\right|ds$$
$$= \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}\left|\sum_{k=-m}^{m}a_k e^{-ihks}\right|ds\cdot\frac{e^{-\gamma\frac{\pi}{h}}}{1-e^{-2\gamma\frac{\pi}{h}}}\,.$$

Since we take the coefficients a_k according to the best approximation, we can further conclude that

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}\left|\sum_{k=-m}^{m}a_k e^{-ihks}\right|ds \le \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}|f_\ell(s)|\,ds + \min_{b_{-m},\dots,b_m}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}\left|f_\ell(s)-\sum_{k=-m}^{m}b_k e^{-ihks}\right|ds$$
$$\le 2\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}|f_\ell(s)|\,ds\,,$$

by choosing $b_k = 0$ for $k = -m,\dots,m$. Finally, we have

$$\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}|f_\ell(s)|\,ds = \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}\left|g(s)\cdot e^{\gamma s}\,\frac{(1-s)^{\ell-1}}{(\ell-1)!}\cdot\mathbf{1}_{[-1,1]}(s)\right|ds \le e^{\gamma}\int_{-1}^{1}\left|\frac{(1-s)^{\ell-1}}{(\ell-1)!}\right|ds \le e^{\gamma}\,\frac{2^\ell}{\ell!}\,.$$

Hence, the second term in (3.2) is bounded by

$$\sum_{l=1}^{\infty}\int_{(2l-1)\frac{\pi}{h}}^{(2l+1)\frac{\pi}{h}}\left|\sum_{k=-m}^{m}a_k e^{-(\gamma+ihk)s}\right|ds \le C_1\,\frac{e^{-\gamma\frac{\pi}{h}}}{1-e^{-2\gamma\frac{\pi}{h}}}\,,\quad C_1 = e^{\gamma}\,\frac{2^{\ell+1}}{\ell!}\,,$$

where C_1 depends only on γ and ℓ. ∎

With Lemma 2.4, we instantly obtain the following corollary, which shows that the approximation works uniform in time and is therefore suitable for stiff problems in time integration.

COROLLARY 3.2. *Let A have a field-of-values in the left complex half-plane and take the fixed denominator polynomial $q_{2m+1}(z) = \prod_{k=-m}^{m}(\gamma+ihk-z)$. Then, we have for all $\tau \ge 0$*

$$\inf_{r\in\frac{\mathcal{P}_{2m}}{q_{2m+1}}}\|\varphi_\ell(\tau A)-r(\tau A)\| \le C_1\,\frac{e^{-\frac{\gamma\pi}{h}}}{1-e^{-\frac{2\gamma\pi}{h}}} + C_2\,\frac{1}{(hm)^\ell}\,,\quad \ell\ge 1\,,$$

where C_1 and C_2 depend only on γ and ℓ.

Similarly, all other theorems can be reformulated with the help of Lemma 2.4. Since these reformulations are obvious, we will not state them as corollaries in the following, and without loss of generality we will formulate our results for $\tau = 1$.

4. Rational Krylov subspace approximation. We now carry the results on the approximation of $\varphi_\ell(A)$ in the rational matrix subspace $\mathcal{R}_{2m+1}(A)$ over to the approximation of $\varphi_\ell(A)v$ in the rational Krylov subspace $\mathcal{Q}_{2m+2}(A,v)$. In Subsection 4.1, we state the resulting error bounds and we discuss the computation of the rational Krylov subspace approximation in Subsection 4.2.

4.1. Error bounds. We consider the approximation of $\varphi_\ell(A)v$ in the rational Krylov subspace

$$\mathcal{Q}_{2m+2}(A,v) = \left\{\frac{p_{2m+1}(A)}{q_{2m+1}(A)}\,v,\ p_{2m+1}\in\mathcal{P}_{2m+1}\right\},\quad q_{2m+1}(A) = \prod_{k=-m}^{m}(z_k I - A)$$

with $z_k = \gamma + ihk$ for $k = -m,\dots,m$. The basic idea is to restrict the large matrix $A\in\mathbb{C}^{N\times N}$ to the subspace $\mathcal{Q}_{2m+2}(A,v)$ by using the restriction $A_m = P_m A P_m$, where P_m is the orthogonal projection onto $\mathcal{Q}_{2m+2}(A,v)$. If this subspace has dimension $2m+2$, the matrix A_m has rank

$2m+2$ and can therefore be represented by a small $(2m+2)\times(2m+2)$-matrix S_m. The matrix φ-function times a vector is then approximated as

$$\varphi_\ell(A)v \approx \varphi_\ell(A_m)v\,,$$

where $\varphi_\ell(A_m)v$ can be computed more efficiently than $\varphi_\ell(A)v$, when A is a large matrix. In contrast to the previous section, we choose an approximation in the space $\mathcal{P}_{2m+1}/q_{2m+1}$ instead of $\mathcal{P}_{2m}/q_{2m+1}$. The additional vector v in the rational Krylov subspace $\mathcal{Q}_{2m+2}(A,v)$ compared to $\mathcal{R}_{2m+1}(A)v = p_{2m}(A)/q_{2m+1}(A)v$, $p_{2m}\in\mathcal{P}_{2m}$, is needed to ensure that $r(A)v = r(A_m)v$ for every rational function $r\in\mathcal{P}_{2m}/q_{2m+1}$. This condition requires $P_m v = v$ so that we have to include v in the rational Krylov subspace. Since $(\mathcal{P}_{2m}/q_{2m+1})(A)v + \operatorname{span}\{v\} = (\mathcal{P}_{2m+1}/q_{2m+1})(A)v = \mathcal{Q}_{2m+2}(A,v)$, we consider from now on the approximation of $\varphi_\ell(A)v$ in the rational Krylov subspace $\mathcal{Q}_{2m+2}(A,v)$. This is no restriction and allows nevertheless the application of the previous results for the approximation of the matrix φ-functions in the subspace $\mathcal{R}_{2m+1}(A)$ because of the inclusion $\mathcal{P}_{2m}/q_{2m+1}\subset\mathcal{P}_{2m+1}/q_{2m+1}$. We start with a lemma that shows that the locus of $W(A)$ in the left complex half-plane is preserved for the restriction $A_m = P_m A P_m$.

LEMMA 4.1. *Let $A\in\mathbb{C}^{N\times N}$ be such that $W(A)\subseteq H_0^-$ and let P_m be the orthogonal projection to the rational Krylov subspace $\mathcal{Q}_{2m+2}(A,v)$. Then, the restriction $A_m = P_m A P_m$ of A to $\mathcal{Q}_{2m+2}(A,v)$ belongs to our studied class of matrices, i.e. $W(A_m)\subseteq H_0^-$.*

Proof. Let $x\in\mathbb{C}^N$ be an arbitrary vector and set $y = P_m x$. Since P_m is an orthogonal projection and is therefore self-adjoint, we have

$$\operatorname{Re}(A_m x, x) = \operatorname{Re}(P_m A P_m x, x) = \operatorname{Re}(A P_m x, P_m x) = \operatorname{Re}(Ay, y)\le 0$$

by our assumption $W(A)\subseteq H_0^-$. This is equivalent to $W(A_m)\subseteq H_0^-$. ∎

Prepared with this lemma, the following theorem states the convergence rate for the approximation of $\varphi_\ell(A)v$ by $\varphi_\ell(A_m)v$ in the rational Krylov subspace $\mathcal{Q}_{2m+2}(A,v)$.

THEOREM 4.2. *Let A be an arbitrary matrix with $W(A)\subseteq H_0^-$ and let v be an arbitrary vector. Further, let $A_m = P_m A P_m$ be the restriction of A to the rational Krylov subspace $\mathcal{Q}_{2m+2}(A,v)$, where P_m is the orthogonal projection onto $\mathcal{Q}_{2m+2}(A,v)$. Then, we have the bound*

$$\|\varphi_\ell(A)v - \varphi_\ell(A_m)v\| \le 2\left[C_1\,\frac{e^{-\frac{\gamma\pi}{h}}}{1-e^{-\frac{2\gamma\pi}{h}}} + C_2\,\frac{1}{(hm)^\ell}\right]\|v\|\,,\quad \ell\ge 1\,, \tag{4.1}$$

where C_1 and C_2 depend only on γ and ℓ.

Proof. It is well-known (cf. [3], p. 21), that the rational Krylov subspace approximation in the space $\mathcal{Q}_{2m+2}(A,v)$ is exact for every rational function of the form $r\in\mathcal{P}_{2m+1}/q_{2m+1}$, that is $r(A)v = r(A_m)v$. Therefore, we can write

$$\|\varphi_\ell(A)v - \varphi_\ell(A_m)v\| \le \|\varphi_\ell(A) - r(A)\|\|v\| + \|\varphi_\ell(A_m) - r(A_m)\|\|v\|\,.$$

For the rational function $r\in\mathcal{P}_{2m+1}/q_{2m+1}$, we can in particular choose a function in the space $\mathcal{P}_{2m}/q_{2m+1}\subset\mathcal{P}_{2m+1}/q_{2m+1}$. Since we have $W(A)\subseteq H_0^-$ and also, due to Lemma 4.1, $W(A_m)\subseteq H_0^-$, we can then use the same estimates as in the proof of Theorem 3.1 on both parts of the sum to give

$$\|\varphi_\ell(A)v - \varphi_\ell(A_m)v\| \le 2\|v\|\int_0^\infty \left|\mathbf{1}_{[0,1]}(s)\,\frac{(1-s)^{\ell-1}}{(\ell-1)!} - \sum_{k=-m}^{m} a_k e^{-(\gamma+ihk)s}\right| ds\,.$$

From here, the proof proceeds analogously to the proof of Theorem 3.1 and we end up with

$$\|\varphi_\ell(A)v - \varphi_\ell(A_m)v\| \le 2\left[C_1\,\frac{e^{-\frac{\gamma\pi}{h}}}{1-e^{-\frac{2\gamma\pi}{h}}} + C_2\,\frac{1}{(hm)^\ell}\right]\|v\|\,,$$

where C_1 and C_2 are the same constants as in Theorem 3.1. ∎

Theorem 4.2 predicts a sublinear convergence that is independent of the norm of A. More precisely, we have obtained a uniform error estimate for arbitrary matrices with a large field-of-values somewhere in the left complex half-plane. In the case that the matrix A stems from a spatial discretization, we therefore have a grid-independent approximation of $\varphi_\ell(A)v$ in the rational Krylov subspace $\mathcal{Q}_{2m+2}(A, v)$. It should be noted that only the second term in the bound (4.1) decreases with the number m of iteration steps. For this reason, the free parameters γ and h must be suitably selected in order to obtain a reasonable error bound. We will come back to this issue in Section 5, where we discuss possible choices for these parameters.

4.2. Efficient computation of the rational Krylov subspace approximation. It remains to show that the rational Krylov subspace approximation $\varphi_\ell(A_m)v$ can be computed more efficiently than the original matrix function $\varphi_\ell(A)v$. The standard procedure is to use the Gram-Schmidt process to determine an orthonormal basis $V_m = [v_1\ v_2\ \cdots\ v_{2m+2}] \in \mathbb{C}^{N\times(2m+2)}$ of the rational Krylov subspace $\mathcal{Q}_{2m+2}(A, v)$ with respect to the chosen inner product $(\cdot\,,\cdot)$ on $\mathbb{C}^N$. This is realized by Algorithm 1 below. When an orthonormal basis V_m has been calculated, the orthogonal projection on the rational Krylov subspace is given by $P_m = V_m V_m^+$, where $V_m^+ : \mathbb{C}^N \to \mathbb{C}^{2m+2}$ is the Moore-Penrose pseudoinverse of $V_m : \mathbb{C}^{2m+2} \to \mathbb{C}^N$ satisfying

$$V_m V_m^+ V_m = V_m\,, \quad V_m^+ V_m V_m^+ = V_m^+\,, \quad (V_m V_m^+)^* = V_m V_m^+\,, \quad (V_m^+ V_m)^* = V_m^+ V_m\,.$$

With the notation $(\cdot)^*$, we denote the adjoint with respect to the inner product on $\mathbb{C}^N$ or, respectively, $\mathbb{C}^m$. Hereby, the space $\mathbb{C}^m$ is always endowed with the standard Euclidean inner product, but the space $\mathbb{C}^N$ might be equipped with another inner product. The adjoint with respect to the Euclidean inner product in both spaces is designated by $(\cdot)^H$, meaning conjugated and transposed as usual. With these notations, the rational Krylov subspace approximation reads

$$\varphi_\ell(A_m)v = V_m \varphi_\ell(S_m) V_m^+ v\,,$$

where $S_m = V_m^+ A V_m \in \mathbb{C}^{(2m+2)\times(2m+2)}$ is, in general, a small matrix. For the standard Euclidean inner product on $\mathbb{C}^N$ this simplifies to $P_m = V_m V_m^H$ and

$$\varphi_\ell(A_m)v = V_m \varphi_\ell(S_m) V_m^H v = \|v\| V_m \varphi_\ell(S_m) e_1$$

with $S_m = V_m^H A V_m$ and $V_m^H v = \|v\| e_1$, where e_1 is the first unit vector in $\mathbb{C}^{2m+2}$. Typically, $\varphi_\ell(A_m)v$ is a good approximation to $\varphi_\ell(A)v$ for $2m+2 \ll N$. To compute $\varphi_\ell(S_m)e_1$ for the small matrix S_m one can use known algorithms for dense matrices (see e.g. [1]).

5. On the choice of γ and h. In order to choose the two parameters γ and h appropriately, we have to know how the constant $C_2 = C_\ell \mathrm{Var}_{[-1,1]}\, \tilde{u}_\ell(\cdot)$ in the bound (3.1) depends on γ. Therefore, we will first establish an upper bound for the total variation of $\tilde{u}_\ell$ on the interval $[-1, 1]$. Remember that $\tilde{u}_\ell$ in the proof of Theorem 3.1 was defined as $\tilde{u}_\ell(\cdot) = f_\ell^{(\ell-1)}(\cdot)$ for $s \neq 1$ and as the mean of the left and right limit at $s = 1$, where

$$f_\ell(s) = g(s) \cdot e^{\gamma s}\, \frac{(1-s)^{\ell-1}}{(\ell-1)!} \cdot \mathbf{1}_{[-1,1]}(s)\,.$$

Consequently, the function $\tilde{u}_\ell$ is differentiable on $[-1, 1)$, has a jump at 1, and is equal to zero outside $[-1, 1]$. By the definition of the total variation, we thus can bound the variation of $\tilde{u}_\ell$ by

$$\mathrm{Var}_{[-1,1]}\, \tilde{u}_\ell(\cdot) \leq \int_{-1}^{1} |f_\ell^{(\ell)}(s)|\, ds + \sup_{s\in[-1,1)} |f_\ell^{(\ell-1)}(s)|\,.$$

The last term is needed to take the jump of $\tilde{u}_\ell$ at the point 1 into account. To estimate $|f_\ell^{(\ell)}(s)|$ and $|f_\ell^{(\ell-1)}(s)|$ for $s \in [-1, 1)$, we use the general Leibniz rule. We set

$$C_g := \max_{k=0,\ldots,\ell}\, \max_{s\in[-1,1]} |g^{(k)}(s)|\,,$$

where g is the function defined in (3.3). Since g does not depend on γ, the same holds true for the constant C_g, and we have

$$|f_\ell^{(\ell-1)}(s)| \le C_g \cdot \sum_{k=0}^{\ell-1} \binom{\ell-1}{k} \left| \left(e^{\gamma s} \, \frac{(1-s)^{\ell-1}}{(\ell-1)!} \right)^{(k)} \right| , \quad s \in [-1,1) \, .$$

Applying the Leibniz rule once more, we get

$$\left| \left(e^{\gamma s} \, \frac{(1-s)^{\ell-1}}{(\ell-1)!} \right)^{(k)} \right| \le \sum_{j=0}^{k} \binom{k}{j} \left| (e^{\gamma s})^{(k-j)} \right| \left| \left(\frac{(1-s)^{\ell-1}}{(\ell-1)!} \right)^{(j)} \right| \le 2e^{\gamma} (1+\gamma)^k \, ,$$

since the last factor in the sum can be bounded by 2 for all $j = 0, \ldots, k$, $\ell \ge 1$ and $s \in [-1,1)$. For the second term $|f_\ell^{(\ell)}(s)|$, one can proceed analogously. So, we conclude that

$$|f_\ell^{(\ell-1)}(s)| \le 2\, C_g e^{\gamma} (2+\gamma)^{\ell-1} \quad \text{and} \quad |f_\ell^{(\ell)}(s)| \le 2\, C_g e^{\gamma} (2+\gamma)^{\ell} \, .$$

Altogether, we obtain

$$\mathrm{Var}_{[-1,1]} \, \tilde{u}_\ell(\cdot) \le 6\, C_g e^{\gamma} (2+\gamma)^{\ell}$$

and therefore

$$C_2 \le \widetilde{C}_\ell e^{\gamma} (2+\gamma)^{\ell} \, ,$$

where $\widetilde{C}_\ell$ depends only on ℓ but not on γ and h.

To select the parameters γ and h in a suitable way, we now have to take the whole expression

$$C_1 \, \frac{e^{-\frac{\gamma\pi}{h}}}{1 - e^{-\frac{2\gamma\pi}{h}}} + C_2 \, \frac{1}{(hm)^\ell} \le \frac{2^{\ell+1}}{\ell!} \, \frac{e^{\gamma(1-\frac{\pi}{h})}}{1 - e^{-\frac{2\gamma\pi}{h}}} + \widetilde{C}_\ell \, e^{\gamma} (2+\gamma)^{\ell} \, \frac{1}{(hm)^\ell} \tag{5.1}$$

of our error bound into consideration. We can observe that the free parameter γ should neither be chosen too large nor too small to keep the two terms in (5.1) of moderate size.

Moreover, we have to deal with the question of a proper choice of the second parameter h. Since the first term in (5.1) does not depend on the iteration index m, a possible strategy could be to choose γ and h such that

$$C_1 \, \frac{e^{-\frac{\gamma\pi}{h}}}{1 - e^{-\frac{2\gamma\pi}{h}}} \le tol \, ,$$

where tol is a given tolerance, e.g. $tol = 10^{-4}$. Noting that the second term in (5.1) is of size $\mathcal{O}(\frac{1}{h^\ell})$, it is obvious that too small values of h should be avoided. Additionally, very small values of h can lead to instability problems in the parallel computation of the rational Krylov subspace decomposition. The stability problems can be cured by using Ruhe's serial rational Arnoldi orthogonalization procedure (cf. [31]), but this would then be a serial algorithm. On the other hand, h should not be chosen larger than one, since this would unfortunately result in large values for the prefactor $e^{\gamma}(2+\gamma)^{\ell}$ of the second term. We resume these correlations in Table 5.1 for the case $\ell = 1$.

A second possibility to determine the parameters γ and h is to fix γ and to choose h in such a way that both terms in (5.1) decrease equally with m, which leads to

$$e^{-\frac{\gamma\pi}{h}} = \frac{1}{(hm)^\ell} \quad \Longleftrightarrow \quad h = \frac{\gamma\pi}{\ell} \, \frac{1}{W\big(\frac{\gamma\pi m}{\ell}\big)} \, , \tag{5.2}$$

where $W(\cdot)$ denotes the Lambert W function, which fulfills the equation $z = W(z)e^{W(z)}$ for any

$C_1 \frac{e^{-\frac{\gamma\pi}{h}}}{1-e^{-\frac{2\gamma\pi}{h}}}$	h	γ	$e^\gamma(2+\gamma)$
10^{-4}	0.25	1	8.2
10^{-4}	0.5	2.1	33.5
10^{-4}	1	5	1038.9
10^{-4}	2	18.6	$2.5 \cdot 10^9$

Table 5.1: Correlation of γ and h for $tol = 10^{-4}$, $\ell = 1$.

$z \in \mathbb{C}$. Due to the well-known lower and upper bound for the Lambert W function (cf. [18])

$$\ln(z) - \ln(\ln(z)) \leq W(z) \leq \ln(z)\,, \quad z \geq e\,,$$

the choice (5.2) ensures that both terms in (5.1) decrease like

$$C_1 \frac{e^{-\frac{\gamma\pi}{h}}}{1-e^{-\frac{2\gamma\pi}{h}}} + C_2 \frac{1}{(hm)^\ell} \leq C_{\gamma,\ell} \left(\frac{\ln(m)}{m}\right)^\ell$$

with a constant $C_{\gamma,\ell}$ depending only on γ and ℓ.

6. Numerical experiments. In our first experiment in Subsection 6.1, we check the predicted convergence rates numerically. Subsection 6.2 studies the performance of the parallel version of our algorithm compared to the serial version. And finally, in Subsection 6.3, we present experiments with a finite-element discretization of a wave equation with homogeneous Neumann boundary conditions.

6.1. Convergence rate testing. We first consider a small test example, where we approximate $\varphi_\ell(A)v$ in the rational Krylov subspace $\mathcal{Q}_{2m+2}(A, v)$ for a $2\,000 \times 2\,000$-matrix A whose eigenvalues $\Lambda(A) = \{\lambda_1, \ldots, \lambda_{2\,000}\}$ lie on the boundary of the left semicircle around 0 with radius 100 in the left complex half-plane. For this, we build a diagonal matrix that contains the selected eigenvalues, and perform a similarity transform with an orthogonal test matrix Q from the Matlab gallery 'orthog' (type $k = 1$). This allows us to determine the exact matrix function by simply computing $\varphi_\ell(A) = Q^H \varphi_\ell(D) Q$, where D is the diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_{2\,000})$. By construction, A is normal and since the field-of-values is in this case the convex hull of its eigenvalues, $W(A)$ is located in the left complex half-plane and the matrix A therefore fits in our framework. The initial value v is chosen as a random vector and the free parameters are set to $\gamma = 2$ and $h = 0.5$ according to Table 5.1.

In Figure 6.1, we plot $\ln(E_m)/\ln(m)$ against the number m of iteration steps, where E_m designates the error for the approximation of $\varphi_\ell(A)v$ in the rational Krylov subspace $\mathcal{Q}_{2m+2}(A, v)$ given as $E_m = \|\varphi_\ell(A)v - \varphi_\ell(A_m)v\|$. The curves correspond for $\ell = 1, 2, 3, 4$ to the black solid, red dashed, blue dotted, and green dash-dotted line. As expected from our convergence analysis, $\ln(E_m)/\ln(m)$ tends to $-\ell$ for larger dimensions of the rational Krylov subspace.

6.2. A parallel test example. Parallel computing is mainly used either to treat extremely large problems, that cannot be handled on a single computer, or to save computation time by solving problems of medium or large size faster. Here, we direct our attention to the second case: Since we have $2m + 1$ different poles $z_k = \gamma + ihk$, $k = -m, \ldots, m$, and therefore $2m + 1$ independent linear system $(z_k I - A)^{-1} v$ to solve, our method is perfectly suited for a parallel implementation in time. By computing p single linear systems simultaneously on p kernels, a tremendous speed-up can be obtained.

Possible variants of a parallel rational Krylov algorithm are discussed by Skoogh in [34]. His general parallel algorithm is shown in Algorithm 1. This algorithm determines an orthonormal basis $V_m = [v_1\ v_2\ \cdots\ v_{2m+2}]$ of the rational Krylov subspace $\mathcal{Q}_{2m+2}(A, v)$ with the prescribed

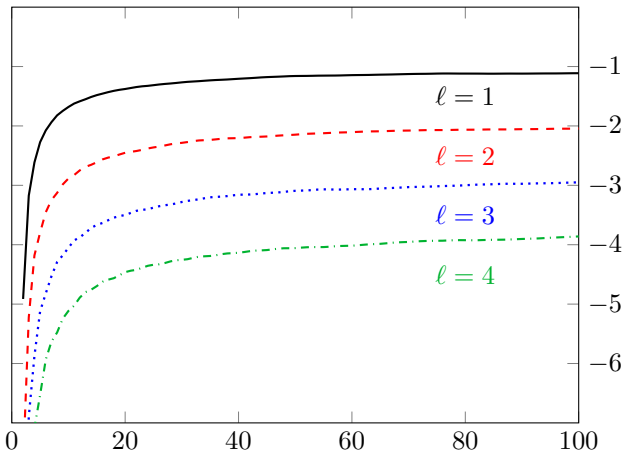

Fig. 6.1: Plot of $\ln(E_m)/\ln(m)$ versus m.

denominator polynomial $q_{2m+1}(z) = \prod_{k=-m}^{m}(z_k - z)$ by using the Gram-Schmidt process. After solving in each loop p linear systems $(z_k I - A)w_k = \tilde{w}_k$ on p different kernels, the resulting vectors are orthogonalized to each other. In our experiments, we only use one loop per time step and take $\tilde{w}_k = v$ for all k. A master program performs the program control, manages the communication and does all orthogonalizations, while the slave programs each solve one of the occurring linear systems.

Algorithm 1: Parallel rational Krylov algorithm

given: matrix $A \in \mathbb{C}^{N\times N}$, vector $v \in \mathbb{C}^N$,
set of poles $Z := \{z_k = \gamma + ihk,\ k = -m, \ldots, m\}$
$v_1 = v/\|v\|$
$j = 1$
while $j \leq 2m + 1$
  choose vectors $\tilde{w}_k,\ k = 1, \ldots, p$
  choose poles $z_k \in Z,\ k = 1, \ldots, p$
  compute $w_k = (z_k I - A)^{-1}\tilde{w}_k,\ k = 1, \ldots, p$
  for $k = 1, \ldots, p$ **do**
    for $i = 1, \ldots, j$
      $h_{i,j} = (w_k, v_i)$
    end for
    $w_k = w_k - \sum_{i=1}^{j} h_{i,j} v_i$
    $h_{j+1,j} = \|w_k\|$
    $v_{j+1} = w_k/h_{j+1,j}$
    $j = j + 1$
  end for
end while

To illustrate that a parallel implementation of the rational Krylov subspace method can significantly outperform the serial computation, we consider a full matrix $A \in \mathbb{C}^{1\,500\times 1\,500}$ with a field-of-values $[-1500, -1]$ on the negative real line and a random vector v of norm one. We approximate $\varphi_1(\tau A)v$ and $\varphi_4(\tau A)v$, $\tau = 0.05$, by applying a parallel and a serial version of our

rational Krylov subspace decomposition with $h = 0.25$ and $\gamma = 1$ in accordance with Table 5.1. We use reorthogonalization in both algorithms. In Figure 6.2, the approximation error is plotted against the computation time in seconds. The parallel version (red solid line) has been computed on a local cluster of 13 heterogeneous workstations using MPI and the serial version (blue dashed line) has been computed on one of the fastest workstations, both times using the C programming language. The network is not a high-performance network suited for parallel computation. Nevertheless, the parallel-in-time version of our rational Krylov subspace method is tremendously faster for our test problem, as can be seen in Figure 6.2. The smallest error corresponds to the dimension 100 of the Krylov subspace. The green dash-dotted line displays the error obtained by the implicit Euler method applied to the stiff equation (1.3) with solution $y(\tau) = \tau^\ell \varphi_\ell(\tau A)v$, where we scaled the result by $\tau^{-\ell}$. This error curve serves as a reference for standard stiff time integration methods. For comparable computation time, the φ_4-function can be approximated more accurately than the φ_1-function, which is in concordance with our theory. However, the implicit Euler method does not improve for higher φ-functions.

We also measured the error and time for a Krylov subspace of dimension 450. For this computation slightly more than one billion of floating point operations (1 Tflop) are necessary. The parallel variant computed an approximation of $\varphi_1(\tau A)v$ with an error of $1.044555 \cdot 10^{-9}$ in 33.53 seconds. The serial variant computed an approximation with the same error in 837.6 seconds, which corresponds to approximately 14 minutes.

In Figure 6.3, $\varphi_1(\tau A)v$ and $\varphi_4(\tau A)v$ have been computed on a single machine with 12 true kernels. The performance of the parallel version is again superior to the serial variant of the algorithm. The small jump in the red solid line for the parallel version in the approximation of the φ_1-function occurs exactly from the change from 12 to 13 threads.

In Figure 6.4, $\varphi_4(A)v$ has been computed for a full matrix $A \in \mathbb{C}^{1\,500 \times 1\,500}$ with a field-of-values $[-1\,500\,i, -i]$ on the imaginary axis and a random vector v. On the left-hand side of Figure 6.4, error versus computation time is shown for the local cluster of 13 heterogeneous workstations and on the right-hand side for the single machine with 12 true kernels. As before, the red solid line refers to the parallel implementation of the rational Krylov subspace method, the blue dashed line to the serial version of the rational Krylov subspace method, and the green dash-dotted line to the implicit Euler method. Again, the parallel implementation shows a significant speed-up.

Our test examples clearly demonstrate that a parallel rational Krylov subspace implementation can be significantly more efficient than a serial implementation.

6.3. Finite-element discretization. We discretize the wave equation

$$u'' = \triangle u - u\,, \quad u(0) = u_0\,, \quad u'(0) = u_0'$$

with homogeneous Neumann boundary conditions by linear finite elements on the domain Ω shown in Figure 6.5.

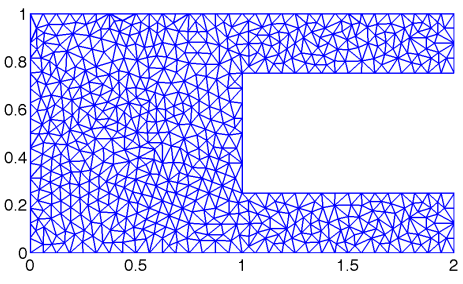

Fig. 6.5: Mesh for Ω with 667 nodes.

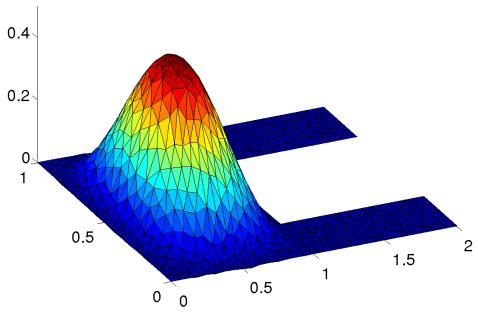

Fig. 6.6: Discrete initial value.

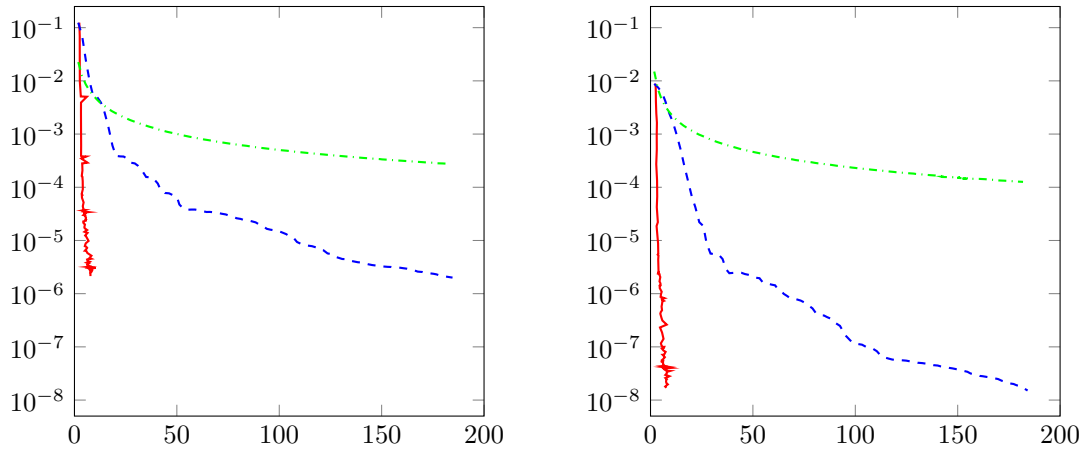

Fig. 6.2: Comparison of a serial (blue dashed line) and a parallel (red solid line) implementation on a local cluster of 13 heterogeneous workstations of the rational Krylov subspace method as well as the implicit Euler method (green dash-dotted line). The error of the approximation of $\varphi_1(\tau A)v$ (top) and $\varphi_4(\tau A)v$ (bottom) is plotted versus the computing time in seconds for a full matrix $A \in \mathbb{C}^{1\,500\times 1\,500}$ with field-of-values $[-1\,500, -1]$, $\tau = 0.05$, and a random vector $v \in \mathbb{C}^{1\,500}$.

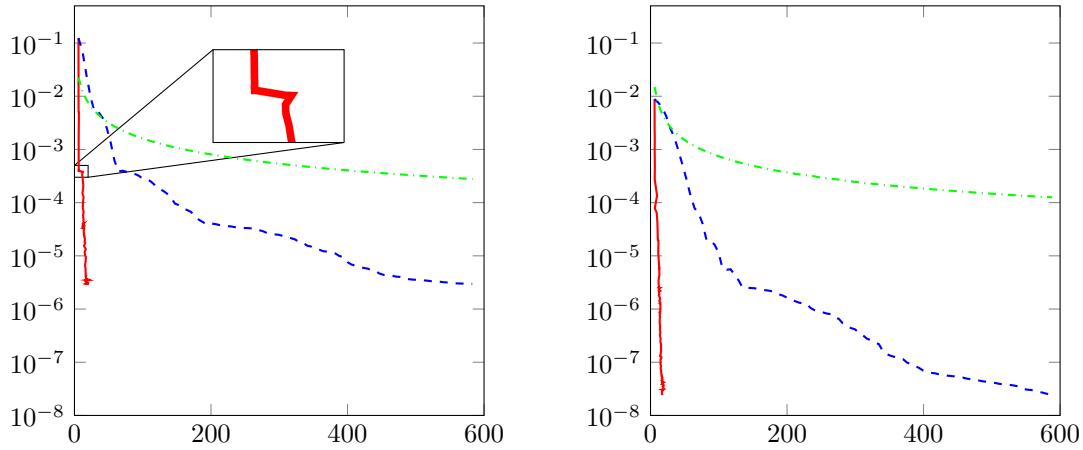

Fig. 6.3: Comparison of a serial (blue dashed line) and a parallel (red solid line) implementation on a machine with 12 true kernels of the rational Krylov subspace method as well as the implicit Euler method (green dash-dotted line). The error of the approximation of $\varphi_1(\tau A)v$ (top) and $\varphi_4(\tau A)v$ (bottom) is plotted versus the computing time in seconds for a full matrix $A \in \mathbb{C}^{1\,500\times 1\,500}$ with field-of-values $[-1\,500, -1]$, $\tau = 0.05$, and a random vector $v \in \mathbb{C}^{1\,500}$.

More exactly, we consider the first-order formulation

$$\begin{bmatrix} M & 0 \\ 0 & M \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix}' = \begin{bmatrix} 0 & M \\ -(K+M) & 0 \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix}, \quad \begin{bmatrix} q(0) \\ p(0) \end{bmatrix} = \begin{bmatrix} q_0 \\ p_0 \end{bmatrix},$$

where q and p are the discretizations of u and u', M is the standard mass matrix and the stiffness matrix $-(K+M)$ is the discretization of $\triangle u - u$ on the grid. Our chosen initial value u_0 can be seen in Figure 6.6. The velocity u_0' has been chosen to be zero. Since K and M are symmetric positive definite matrices, $K+M$ is positive definite and symmetric as well. With $y = [\,q^T\ p^T\,]^T$

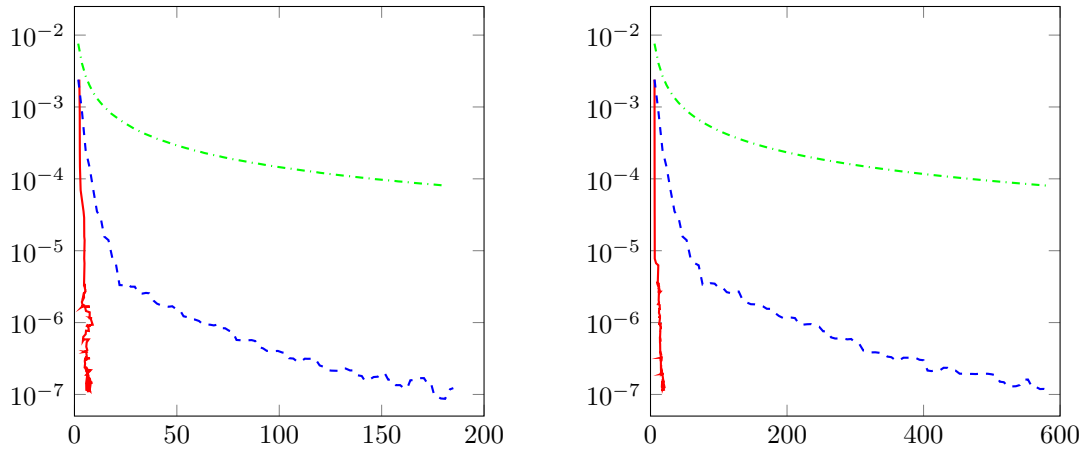

Fig. 6.4: Comparison of a serial (blue dashed line) and a parallel (red solid line) implementation of the rational Krylov subspace method as well as the implicit Euler method (green dash-dotted line). The error of the approximation of $\varphi_4(A)v$ is plotted versus the computing time in seconds for a full matrix $A \in \mathbb{C}^{1\,500\times 1\,500}$ with field-of-values $[-1\,500\,i, -i]$ and a random vector $v \in \mathbb{C}^{1\,500}$ On the left-hand side a local cluster of 13 heterogeneous workstations has been used, on the right-hand side a single machine with 12 true kernels.

we obtain

$$y' = \begin{bmatrix} 0 & I \\ -M^{-1}(K+M) & 0 \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix} =: Ay\,, \quad y(0) = \begin{bmatrix} q(0) \\ p(0) \end{bmatrix} = \begin{bmatrix} q_0 \\ p_0 \end{bmatrix} = y_0\,.$$

Let N be the number of nodes in the current finite-element grid. For the symmetric and positive definite matrix

$$B = \begin{bmatrix} K+M & 0 \\ 0 & M \end{bmatrix},$$

we equip the space $\mathbb{C}^{2N}$ with the inner product $(\cdot\,,\cdot)_B$ given as

$$\left(\begin{bmatrix} q \\ p \end{bmatrix}, \begin{bmatrix} \widetilde{q} \\ \widetilde{p} \end{bmatrix} \right)_B = \widetilde{q}^H (K+M)\, q + \widetilde{p}^H M p = \begin{bmatrix} \widetilde{q}^H & \widetilde{p}^H \end{bmatrix} \begin{bmatrix} K+M & 0 \\ 0 & M \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix}. \tag{6.1}$$

This inner product corresponds to the discretization of a suitable norm in the continuous case. That the discrete norm is properly chosen, is also reflected in the following lemma.

LEMMA 6.1. *With respect to the inner product* (6.1)*, we have*

$$\operatorname{Re}(Ay, y)_B \le 0\,, \quad \forall\, y \in \mathbb{C}^{2N}\,.$$

Proof. With a simple calculation, we have for an arbitrary $y = [\,q^T\ p^T\,]^T \in \mathbb{C}^{2N}$

$$\begin{aligned}(Ay, y)_B &= \left(\begin{bmatrix} 0 & I \\ -M^{-1}(K+M) & 0 \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix}, \begin{bmatrix} q \\ p \end{bmatrix} \right)_B = \left(\begin{bmatrix} p \\ -M^{-1}(K+M)\, q \end{bmatrix}, \begin{bmatrix} q \\ p \end{bmatrix} \right)_B \\ &= q^H (K+M)\, p - p^H (K+M)\, q = 2i \operatorname{Im}(q^H (K+M)\, p)\end{aligned}$$

and hence

$$\operatorname{Re}(Ay, y)_B = 0 \le 0\,,$$

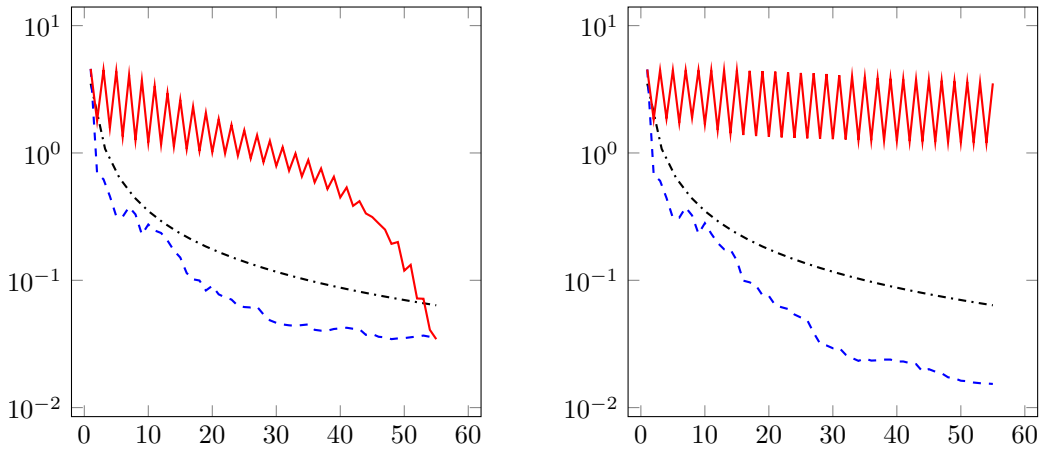

Fig. 6.7: Accuracy of the polynomial Krylov subspace method (red solid line) and of the parallel rational Krylov subspace method (blue dashed line) for the approximation of $\varphi_1(\tau A)Ay_0$ for a $1\,334 \times 1\,334$-matrix (on the left-hand side) and a $78\,862 \times 78\,862$-matrix (on the right-hand side) versus dimension of the Krylov subspaces. The black dash-dotted line has the predicted convergence order $\mathcal{O}\left(\frac{1}{m}\right)$ for the φ_1-function.

that is, the statement of our lemma. □

Lemma 6.1 shows that our theory is applicable. The choice of suitable norms is not only crucial for the continuous problem, but also for the discretized equations. At least, if one would like to have such desirable features as a grid-independent convergence. With respect to the rational Krylov approximation according to Subsection 4.2, the pseudoinverse V_m^+ reads $V_m^+ = V_m^H B$, the projection is $P_m = V_m V_m^H B$ and the small matrix $S_m = V_m^H B A V_m$. One can easily check that

$$S_m = V_m^H B A V_m = V_m^H \begin{bmatrix} 0 & K+M \\ -(K+M) & 0 \end{bmatrix} V_m \,,$$

so that the entry $M^{-1}(K+M)$ in A has not to be computed.

In Figure 6.7, we compare the approximation errors of the exponential Euler method $y(\tau) \approx y_0 + \tau\varphi_1(\tau A)Ay_0$ with $\tau = 0.5$, where the matrix φ_1-function of τA times Ay_0 has been approximated by the "non-stiff" standard polynomial Krylov subspace method (with properly chosen norms, of course) and with the "stiff" rational Krylov subspace method with $h = 0.5$, $\gamma = 1$, respectively. For the φ_1-function, our main theorem states a convergence rate of order $\mathcal{O}\left(\frac{1}{m}\right)$ up to a constant that is small enough for our choice of parameters. To illustrate that the error curve in our experiment behaves as predicted, we included a line of this order in the plot. For larger systems, it becomes obvious that the rational Krylov subspace method is more suited to the numerical solution of stiff problems.

7. Conclusion. We analyzed the convergence of a rational Krylov subspace method for the approximation of $\varphi_\ell(A)v$, $\ell \geq 1$, independent of the norm of the matrix A having a field-of-values in the left complex half-plane with respect to a general inner product. Up to an exponentially small term, the convergence is sublinear, $\mathcal{O}(m^{-\ell})$, for the φ_ℓ-functions. Due to the chosen simple poles, the method might be massively parallelized in time, which has the potential to render a significant speed-up that standard Runge-Kutta methods for stiff problems cannot possess due to their serial nature. The obtained error bounds are slightly improved in the current setting compared to the resolvent Krylov subspace approximation using a single repeated pole. We also emphasized the analogy of the standard Krylov subspace method to a non-stiff method and the rational Krylov subspace method to a stiff solver. The gain in efficiency of the "stiff" rational Krylov method with respect to the "non-stiff" standard Krylov subspace method is apparent.

Acknowledgements. This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) via GR 3787/1-1.

REFERENCES

[1] A. H. Al-Mohy and N. J. Higham. Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM J. Sci. Comput.*, 33(2):488–511, 2011.
[2] B. Beckermann and S. Güttel. Superlinear convergence of the rational Arnoldi method for the approximation of matrix functions. *Numer. Math.*, 121(2):205–236, 2012.
[3] B. Beckermann and L. Reichel. Error estimates and evaluation of matrix functions via the Faber transform. *SIAM J. Numer. Anal.*, 47(5):3849–3883, 2009.
[4] R. A. DeVore and G. G. Lorentz. *Constructive Approximation*. Springer-Verlag, Berlin Heidelberg, 1993.
[5] V. Druskin and L. Knizhnerman. Extended Krylov subspaces: approximation of the matrix square root and related functions. *SIAM J. Matrix Anal. Appl.*, 19(3):755–771, 1998.
[6] V. Druskin, L. Knizhnerman, and M. Zaslavsky. Solution of large scale evolutionary problems using rational Krylov subspaces with optimized shifts. *SIAM J. Sci. Comput.*, 31(5):3760–3780, 2009.
[7] V. Druskin, R. Remis, and M. Zaslavsky. An extended Krylov subspace model-order reduction technique to simulate wave propagation in unbounded domains. *J. Comput. Phys.*, 272:608–618, 2014.
[8] V. Druskin and M. Zaslavsky. On convergence of Krylov subspace approximations of time-invariant self-adjoint dynamical systems. *Linear Algebra Appl.*, 436(10):3883–3903, 2012.
[9] M. Eiermann, O. G. Ernst, and S. Güttel. Deflated restarting for matrix functions. *SIAM J. Matrix Anal. Appl.*, 32(2):621–641, 2011.
[10] A. Frommer and V. Simoncini. Error bounds for Lanczos approximations of rational functions of matrices. In *Numerical Validation in Current Hardware Architectures*, pages 203–216, 2008.
[11] A. Frommer and V. Simoncini. Stopping criteria for rational matrix functions of Hermitian and symmetric matrices. *SIAM J. Sci. Comput.*, 30(3):1387–1412, 2008.
[12] T. Göckler and V. Grimm. Convergence Analysis of an Extended Krylov Subspace Method for the Approximation of Operator Functions in Exponential Integrators. *SIAM J. Numer. Anal.*, 51(4):2189–2213, 2013.
[13] V. Grimm. Resolvent Krylov subspace approximation to operator functions. *BIT Numerical Mathematics*, 52(3):639–659, 2012.
[14] V. Grimm and M. Gugat. Approximation of semigroups and related operator functions by resolvent series. *SIAM J. Numer. Anal.*, 48(5):1826–1845, 2010.
[15] V. Grimm and M. Hochbruck. Rational approximation to trigonometric operators. *BIT Numerical Mathematics*, 48(2):215–229, 2008.
[16] S. Güttel. Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection. *GAMM Mitteilungen*, 36(1):8–31, 2013.
[17] S. Güttel and L. Knizhnerman. A black-box rational Arnoldi variant for Cauchy-Stieltjes matrix functions. *BIT Numerical Mathematics*, 53(3):595–616, 2013.
[18] M. Hassani. Approximation of the Lambert W Function. *RGMIA Research Report Collection*, 8(4), 2005.
[19] N. J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.
[20] M. Hochbruck and Ch. Lubich. On Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.*, 34:1911–1925, 1997.
[21] M. Hochbruck and A. Ostermann. Exponential integrators. *Acta Numer.*, 19:209–286, 2010.
[22] M. Hochbruck, T. Pažur, A. Schulz, E. Thawinan, and C. Wieners. Efficient time integration for discontinuous Galerkin approximations of linear wave equations. To appear in ZAMM, 2014.
[23] L. Knizhnerman and V. Simoncini. A new investigation of the extended Krylov subspace method for matrix function evaluations. *Numer. Linear Algebra Appl.*, 17(4):615–638, 2010.
[24] L. Lopez and V. Simoncini. Analysis of projection methods for rational function approximation to the matrix exponential. *SIAM J. Numer. Anal.*, 44(2):613–635, 2006.
[25] I. Moret. A note on Krylov Methods for Fractional Evolution Problems. *Numer. Funct. Anal. Optim.*, 34(5):539–556, 2013.
[26] I. Moret and P. Novati. RD-rational approximations of the matrix exponential. *BIT Numerical Mathematics*, 44:595–615, 2004.
[27] I. Moret and P. Novati. Interpolating functions of matrices on zeros of quasi-kernel polynomials. *Numer. Linear Algebra Appl.*, 12(4):337–353, 2005.
[28] P. Novati. Using the restricted-denominator rational Arnoldi method for exponential integrators. *SIAM J. Matrix Anal. Appl.*, 32(4):1537–1558, 2011.
[29] M. A. Olshanskii and E. E. Tyrtshnikov. *Iterative Methods for Linear Systems - Theory and Applications*. SIAM, 2014.
[30] M. Popolizio and V. Simoncini. Acceleration techniques for approximating the matrix exponential operator. *SIAM J. Matrix Anal. Appl.*, 30(2):657–683, 2008.
[31] A. Ruhe. Rational Krylov sequence methods for eigenvalue computation. *Linear Alg. Appl.*, 58:391–405, 1984.
[32] A. Ruhe. The rational Krylov algorithm for nonsymmetric eigenvalue problems. III. Complex shifts for real matrices. *BIT Numerical Mathematics*, 34(1):165–176, 1994.

[33] T. Schmelzer and L. N. Trefethen. Evaluating matrix functions for exponential integrators via Carathéodory-Fejér approximation and contour integrals. *Electronic Transactions on Numerical Analysis*, 29:1–18, 2007.

[34] D. Skoogh. An Implementation of a Parallel Rational Krylov Algorithm, PhD thesis, Göteborg University and Chalmers University of Technology, Sweden, 1996.

[35] J. van den Eshof and M. Hochbruck. Preconditioning Lanczos approximations to the matrix exponential. *SIAM J. Sci. Comp.*, 27(4):1438–1457, 2006.