

CONVERGENCE ANALYSIS OF AN EXTENDED KRYLOV SUBSPACE METHOD FOR THE APPROXIMATION OF OPERATOR FUNCTIONS IN EXPONENTIAL INTEGRATORS

TANJA GÖCKLER AND VOLKER GRIMM *

Abstract. We analyze the convergence of an extended Krylov subspace method for the approximation of operator functions that appear in exponential integrators. For operators, the size of the polynomial part of the extended Krylov subspace is restricted according to the smoothness of the initial data. This restriction for the continuous operator has a significant influence on the approximation of matrix functions evaluated for matrices stemming from space discretizations of the continuous operator. We prove convergence of the method for the continuous operator and, in the discrete case, this leads to a convergence independent of the norm of the discretized operator uniformly over all possible grids. The analysis is illustrated by numerical experiments.

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

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

1. Introduction. Exponential integrators form an interesting class of numerical methods for the time integration of evolution equations. The favorable properties of exponential integrators are usually shown by studying exponential integrators for strongly continuous semigroups with infinitesimal generator A . For finite-difference, pseudospectral, finite-element, or discontinuous Galerkin discretizations of such operators, the operator A turns into a large matrix. The norm of this matrix grows with finer space discretizations. However, the properties of exponential integrators that hold true in the continuous case of an operator carry over to results that are independent of the norm of the discretization matrix. These results are therefore independent of the refinement of the space discretization. A review of exponential integrators and their properties is given in [26] by Hochbruck and Ostermann.

For the application of these integrators, the computation of $\varphi(A)v$ is required, where v is a vector and φ is one of the so-called φ -functions. These operator functions are given by

$$\varphi_k(\tau A) := \int_0^1 e^{(1-s)\tau A} \frac{s^{k-1}}{(k-1)!} ds = \frac{1}{\tau^k} \int_0^\tau e^{(\tau-s)A} \frac{s^{k-1}}{(k-1)!} ds, \quad k \geq 1. \quad (1.1)$$

With a suitable functional calculus, they can be seen as the operator A being inserted in the functions defined by the recursion $\varphi_0(z) = e^z$,

$$\varphi_{k+1}(z) = \frac{\varphi_k(z) - \varphi_k(0)}{z}, \quad \varphi_k(0) = \frac{1}{k!}, \quad k \geq 0. \quad (1.2)$$

The efficient and reliable computation of $\varphi(A)v$ is an important ingredient in every exponential integrator.

Recently, the use of rational Krylov subspaces for the approximation of $f(A)v$, where f is a function defined for matrices and/or operators, has been studied by a number of papers (e.g. [1–4, 7–11, 14, 15, 19–21, 27–31, 33, 34, 36]). A promising method is the extended Krylov subspace method that has been suggested by Druskin and Knizhnerman in [7] to approximate matrix functions for invertible matrices in the space

$$\mathcal{K}_0^{q,n}(A, v) = \text{span}\{A^{q-1}v, \dots, Av, v, A^{-1}v, \dots, A^{-n+1}v\}, \quad n \geq 1, q \geq 1.$$

We will study a slight variation of this method for operators, in order to obtain error bounds for the approximation of the φ -functions that are independent of the norm of the matrix of the discretized operator A , uniformly over all possible grids in space. This is the crucial property to be preserved

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

for a successful application of the extended Krylov subspace method in exponential integrators. Without loss of generality, we study bounded strongly continuous semigroups on a Banach space X in this paper. (The results for a general semigroup can be obtained via the common rescaling procedure, e.g. page 60 in [12].) In contrast to matrices, the generator A of a C_0 -semigroup might only be applicable to a dense subset $\mathcal{D}(A)$ of the Banach space X . With the definition $\mathcal{D}(A^0) := X$, where $A^0 := I$, and the recursive definition $\mathcal{D}(A^n) := \{v \in \mathcal{D}(A^{n-1}) \mid A^{n-1}v \in \mathcal{D}(A)\}$ for $n \in \mathbb{N}$, we define smoother and smoother spaces. It is clear, that the operator A can be applied at most q times to vectors in $\mathcal{D}(A^q)$. The generators A of bounded semigroups are not invertible in general, but the resolvent $(\gamma I - A)^{-1}$, with $\gamma > 0$, that is the inverse of the operator $\gamma I - A$, exists and is a bounded operator that pushes $v \in X$ to $(\gamma I - A)^{-1}v \in \mathcal{D}(A)$. The iterated resolvent $(\gamma I - A)^{-n}$ can thus be applied to all vectors $v \in X$ for all $n \geq 0$. In short, the resolvent is a “smoothing” operator in contrast to A . Therefore, we study the approximation of operator functions times a vector $v \in \mathcal{D}(A^q)$ in the space

$$\mathcal{K}_\gamma^{q,n}(A, v) = \text{span} \{A^{q-1}v, \dots, Av, v, (\gamma I - A)^{-1}v, \dots, (\gamma I - A)^{-n+1}v\}, \quad n \geq 1, \quad (1.3)$$

with $\gamma > 0$. The space could be extended by $A^q v$ “to the left”, but the analysis shows that this would give no improvement over our derived bounds despite the fact that $v \in \mathcal{D}(A^q)$ is used in the error analysis.

In contrast to other applications of the extended Krylov subspace, our index q is fixed and given by the abstract smoothness of the initial data. While for $q, n \geq 1$, the meaning of the above definition is clear, we also use

$$\mathcal{K}_\gamma^{0,n}(A, v) = \text{span} \{(\gamma I - A)^{-1}v, \dots, (\gamma I - A)^{-n+1}v\}, \quad n \geq 2, \quad (1.4)$$

where the superscript 0 that refers to the polynomial part of the Krylov space indicates that the space is purely rational, and

$$\mathcal{K}_\gamma^{q,0}(A, v) = \text{span} \{A^{q-1}v, \dots, Av, v\}, \quad q \geq 1, \quad (1.5)$$

designates the standard polynomial Krylov subspace.

Due to the well-known shift invariance of the standard Krylov subspace, the space (1.3) is identical to the extended Krylov subspace with the shifted operator/matrix $\gamma I - A$. We therefore also call methods based on this subspace extended Krylov subspace methods. We will illustrate that the concept of smoothness plays an important role for the discretized operators and that the restriction of the polynomial part of (1.3) with respect to the smoothness of the initial data is well considered.

The paper is organized as follows: After the introduction in this section, we motivate the use of the extended Krylov subspace method and the importance of our continuous analysis for the approximation of matrix functions for matrices stemming from discretized evolution operators in Section 2. After some preliminaries in Section 3, we study the approximation of the semigroup and the φ -functions in the extended Krylov subspace for infinitesimal generators of semigroups in Banach spaces in Section 4. In Section 5, we present a good choice of the free parameter $\gamma > 0$ that rises the asymptotic convergence. The extended Krylov subspace approximation in Hilbert spaces is studied in Section 6. We illustrate our analysis by numerical experiments in Section 7, before giving a brief conclusion in Section 8.

2. Motivation. The main motivation to examine rational Krylov subspace methods is the favorable property that the convergence of these methods does not deteriorate for matrices with large norm in contrast to the standard Krylov subspace method (e.g. [16], [25]). Numerical evidence that this is also true for the method discussed here is given in Section 7 on numerical experiments. The purpose of this section is to motivate why and how the abstract smoothness of the value v can be used to speed up the grid-independent approximation of operator functions, that is, to motivate the special Krylov subspace (1.3). For the illustration, we use the one-dimensional wave equation written as a first order system on the Hilbert space $L^2(0, 1) \times L^2(0, 1)$. This representation of

the wave equation is equivalent to the more common representation as a first order system on the Hilbert space $H_0^1(0, 1) \times L^2(0, 1)$ equipped with its graph norm (cf. Theorem 5.5 in [13]).

We choose the operator $Bu = -u_{xx}$ on the interval $\Omega = (0, 1)$ for the Hilbert space $H = L^2(\Omega)$ with homogeneous Dirichlet boundary conditions. B is a positive, self-adjoint operator on H with a compact resolvent, hence we have a complete orthonormal basis of eigenfunctions e_k and positive eigenvalues λ_k ,

$$e_k(x) = \sqrt{2} \sin(k\pi x), \quad \lambda_k = (k\pi)^2, \quad k = 1, 2, \dots$$

For a function f bounded on $[0, \infty)$, one can define $f(B)v$ as

$$f(B)v := \sum_{k=1}^{\infty} f(\lambda_k) \mu_k e_k, \quad v = \sum_{k=1}^{\infty} \mu_k e_k \in H.$$

Fractional powers B^α are defined on appropriate domains

$$\mathcal{D}(B^\alpha) := \{v \in H \mid \|B^\alpha v\| < \infty\}.$$

Let now $X = H \times H$, and define $A : \mathcal{D}(A) \rightarrow H \times H$ by

$$A = \begin{bmatrix} 0 & B^{\frac{1}{2}} \\ -B^{\frac{1}{2}} & 0 \end{bmatrix},$$

where the domain is given by $\mathcal{D}(A) = \mathcal{D}(B^{\frac{1}{2}}) \times \mathcal{D}(B^{\frac{1}{2}})$, and the domains of the iterated operators are $\mathcal{D}(A^q) = \mathcal{D}(B^{\frac{q}{2}}) \times \mathcal{D}(B^{\frac{q}{2}})$, respectively. In our example, we are dealing with the Sobolev space $\mathcal{D}(B^{\frac{1}{2}}) = H_0^1(0, 1)$. We now consider the evolution equation

$$y'(\tau) = \begin{bmatrix} q(\tau) \\ p(\tau) \end{bmatrix}' = \begin{bmatrix} 0 & B^{\frac{1}{2}} \\ -B^{\frac{1}{2}} & 0 \end{bmatrix} \begin{bmatrix} q(\tau) \\ p(\tau) \end{bmatrix} = Ay(\tau), \quad y(0) = y_0 = \begin{bmatrix} q_0 \\ p_0 \end{bmatrix}. \quad (2.1)$$

For $q = 2, 4, 6, 8$, we use four different initial values

$$y_0^q = \begin{bmatrix} q_0^q \\ p_0^q \end{bmatrix}, \quad q_0^q = 0, \quad p_0^q = \frac{\tilde{p}_0^q}{\|\tilde{p}_0^q\|}, \quad \tilde{p}_0^q = \begin{cases} [0, 1] & \rightarrow \mathbb{R} \\ \tilde{p}_0^q(x) & = x^q(1-x)^q \end{cases}.$$

With this choice, we have $y_0^q \in \mathcal{D}(A^q)$ and $y_0^q \notin \mathcal{D}(A^{q+1})$. That is, q might be called the maximal index of smoothness for the initial values y_0^q with respect to the operator A . The solution of the initial value problem is

$$y^q(\tau) = e^{\tau A} y_0^q = \begin{bmatrix} \cos(\tau B^{\frac{1}{2}}) & \sin(\tau B^{\frac{1}{2}}) \\ -\sin(\tau B^{\frac{1}{2}}) & \cos(\tau B^{\frac{1}{2}}) \end{bmatrix} \begin{bmatrix} q_0^q \\ p_0^q \end{bmatrix}, \quad q = 2, 4, 6, 8.$$

Galerkin discretization with respect to the subspace $V_N = \text{span}\{e_1, \dots, e_N\}$ leads to the matrix

$$A_N = \begin{bmatrix} 0 & B_N^{\frac{1}{2}} \\ -B_N^{\frac{1}{2}} & 0 \end{bmatrix}, \quad B_N^{\frac{1}{2}} = \text{diag}(\pi, 2\pi, \dots, N\pi),$$

and the initial condition

$$y_{0,N}^q = \begin{bmatrix} q_{0,N}^q \\ p_{0,N}^q \end{bmatrix}, \quad q_{0,N}^q = 0, \quad p_{0,N}^q = \tilde{p}_{0,N}^q,$$

where $\tilde{p}_{0,N}^q$ can be computed exactly by partial integration or approximated with the discrete sine transform. The solution of the initial value problem is

$$y_N^q(\tau) = e^{\tau A_N} y_{0,N}^q = \begin{bmatrix} \cos(\tau B_N^{\frac{1}{2}}) & \sin(\tau B_N^{\frac{1}{2}}) \\ -\sin(\tau B_N^{\frac{1}{2}}) & \cos(\tau B_N^{\frac{1}{2}}) \end{bmatrix} \begin{bmatrix} q_{0,N}^q \\ p_{0,N}^q \end{bmatrix}, \quad q = 2, 4, 6, 8.$$

In Figure 2.1, we show the errors of the approximation to $e^{\tau A_N} y_{0,N}^q$ in the three Krylov subspaces versus the dimension n of these subspaces. More exactly, the subplots of Figure 2.1 refer to the approximation of $e^{\tau A_N} y_{0,N}^q$, with $q = 2, 4, 6, 8$, from top left to bottom right with $\tau = 1/4$ and $N = 255$. The approximation in the standard Krylov subspace $\mathcal{K}_\gamma^{n,0}(A_N, y_{0,N}^q)$ is marked by crosses and a dash-dotted line, the approximation in the purely rational Krylov subspace $\mathcal{K}_\gamma^{0,n}(A_N, y_{0,N}^q)$, with $\gamma = 1$, by stars and a dashed line, and the approximation in the extended Krylov subspace $\mathcal{K}_\gamma^{q,n-q}(A_N, y_{0,N}^q)$, with $\gamma = 1$, by circles and a solid line. Errors are measured in the standard Euclidean norm that corresponds to the L^2 -norm in this example.

A first thing to observe is that the methods approximate $e^{\tau A_N} y_{0,N}^q$ the better the smoother the original continuous initial value y_0^q is, that is, the higher the number q with $y_0^q \in \mathcal{D}(A^q)$ is. Hence, the smoothness of the continuous initial data has a significant effect on the question how well the matrix functions with the discretization matrix times the discretized initial value can be approximated in the corresponding Krylov subspaces. While this is to be expected for very fine space discretizations, it is surprising that it starts with very coarse space discretizations with $N = 255$ and small matrices in our test problem.

The performance of the standard Krylov subspace method (cross-marked and dash-dotted line) deteriorates exactly as soon as the space uses vectors $A_N^l y_{0,N}^q$ with $l \geq q$. The transition from the standard Krylov subspace $\mathcal{K}_\gamma^{q,0}(A_N, y_{0,N}^q)$ that does not use $A_N^q y_{0,N}^q$ to $\mathcal{K}_\gamma^{q+1,0}(A_N, y_{0,N}^q)$ that does use $A_N^q y_{0,N}^q$ is marked by a vertical gray line. This behavior in our test problem perfectly fits the theory presented later and motivates the proposed restriction of the polynomial part of the extended Krylov subspace method.

The purely rational method (star-marked and dashed line) gives the steepest descent in the error. However, the rational method requires the solution of a linear system with a large matrix $\gamma I - A_N$ in every step for general discretization matrices. While it is clear that this can pay off for the “right of the gray line”, the standard Krylov method that only uses matrix multiplications is clearly faster as long as the index is smaller than the smoothness of the data. It is therefore reasonable to assume that the most effective method with respect to error versus numerical work should use the polynomial subspace as long as possible and then switch to the rational method. This is exactly what an approximation in the extended Krylov subspace (1.3) (circle-marked and solid line) does.

Numerical examples, like the one just discussed, show that the smoothness of the initial data is important for an effective approximation. The smoothness of the initial data should therefore be taken into account. The fact that the discretized operators, mostly represented by matrices, behave more and more like the continuous operator for finer space discretizations, motivates the use and the study of the Krylov subspace (1.3). The discussion on the continuous operator level guarantees that the obtained error bounds are independent of the space discretization.

3. Preliminaries. We briefly review a functional calculus that might be seen as a slightly simpler version of the functional calculus of Hille and Phillips in [24]. More details about our functional calculus can be found in [18].

The Lebesgue spaces of complex-valued functions defined on \mathbb{R} are denoted by $L^q(\mathbb{R})$ with norm $\|\cdot\|_q$. Besides

$$C(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{C} \mid f \text{ is continuous on } \mathbb{R}\},$$

let

$$\mathcal{M}_+ = \{f \in C(\mathbb{R}) \mid \mathcal{F}f \in L^1(\mathbb{R}) \text{ and } \text{supp}(\mathcal{F}f) \subseteq [0, \infty)\}, \quad (3.1)$$

where $\mathcal{F}f$ is the Fourier transform of f , i.e.

$$\mathcal{F}f(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixs} f(x) dx.$$

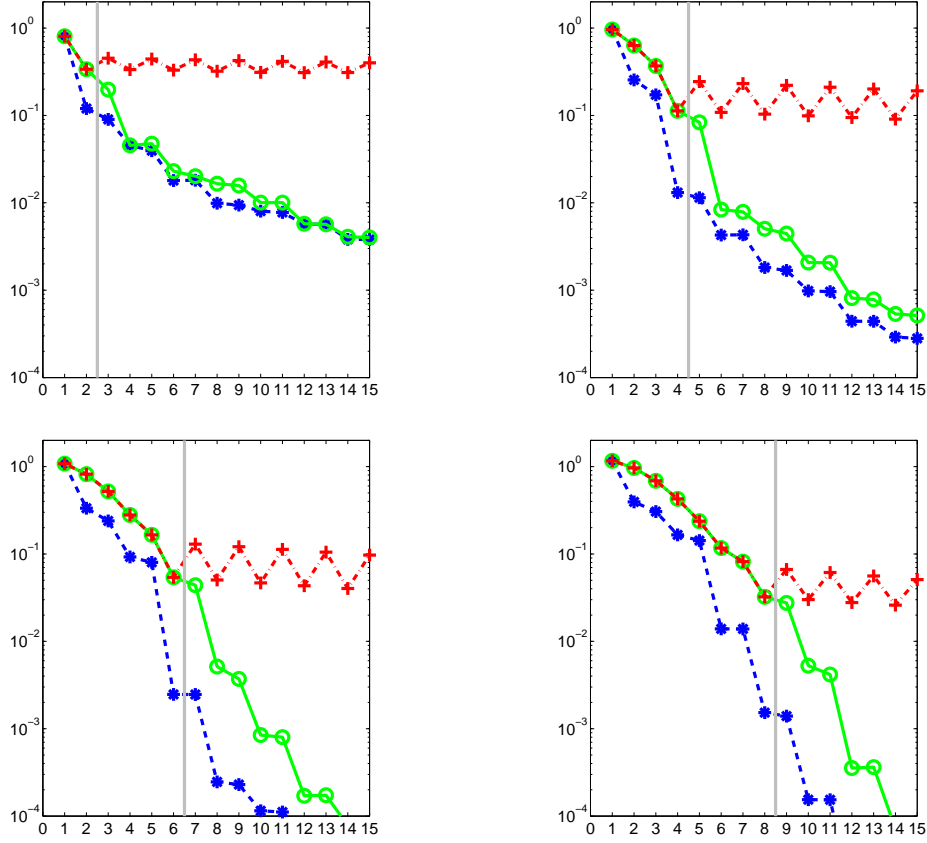


FIG. 2.1. Plot of error vs. dimension of the Krylov space for the standard Krylov subspace $\mathcal{K}_\gamma^{n,0}(A_N, y_{0,N}^q)$ (cross-marked and dash-dotted line), the purely rational Krylov subspace $\mathcal{K}_\gamma^{0,n}(A_N, y_{0,N}^q)$ (star-marked and dashed line) and the extended Krylov subspace $\mathcal{K}_\gamma^{q,n-q}(A_N, y_{0,N}^q)$ (circle-marked and solid line) for $N = 255$ and initial values $y_{0,N}^q$ that are discretized initial values from smoother and smoother data $y_0^q \in \mathcal{D}(A^q)$, $q = 2, 4, 6, 8$, from top left to bottom right.

For each function holomorphic in the left complex half-plane, we denote by $f_{(0)} : \mathbb{R} \rightarrow \mathbb{C}$ the restriction of f to $\text{Re } z = 0$, so that $f_{(0)}(\xi) = f(i\xi)$, and define the algebra

$$\widetilde{\mathcal{M}} := \{f \text{ holomorphic and bounded in } \text{Re } z \leq 0 \mid f_{(0)} \in \mathcal{M}_+\}.$$

Let now A generate a bounded strongly continuous semigroup on some Banach space X , that is $\|e^{\tau A}\| \leq N$. For functions $f \in \widetilde{\mathcal{M}}$, we define a functional calculus via

$$f(A) = \int_0^\infty e^{sA} \mathcal{F}f_{(0)}(s) ds.$$

This defines a bounded linear operator $f(A)$ satisfying

$$\|f(A)\| \leq N \|\mathcal{F}f_{(0)}\|_1.$$

The functional calculus so far is suitable to treat the φ_k -functions for $k \geq 1$. We will need another extension in order to include the semigroup, i.e. the φ_0 -function. Let

$$\mathcal{M}_0 := \{f \text{ holomorphic for } \text{Re } z \leq 0 \mid \exists n \in \mathbb{N}_0 : \frac{f(z)}{(1-z)^n} \in \widetilde{\mathcal{M}}\}.$$

For $f \in \mathcal{M}_0$, we now define

$$f(A) := (I - A)^n \left(\frac{f(z)}{(1-z)^n} \right) (A),$$

where $n \in \mathbb{N}_0$ is such that $\frac{f(z)}{(1-z)^n} \in \widetilde{M}$. Note that the definition does not depend on the choice of n . The definition results in a closed operator on X . Finally, we define

$$\widetilde{\mathcal{M}} \subseteq \mathcal{M} := \{f \in \mathcal{M}_0 \mid f(A) : X \rightarrow X \text{ is bounded}\}.$$

The following lemma can be found as Proposition 1.12 in [22].

LEMMA 3.1. *The mapping $f \rightarrow f(A)$ is a homomorphism of \mathcal{M} into the algebra of bounded linear operators on X .*

With this definition of the functional calculus, the functions $\varphi_k(z)$, $k \geq 0$, used with the functional calculus defined above coincide with the definition (1.1) of the bounded operators $\varphi_k(\tau A)$ given in the literature about exponential integrators (cf. Lemma 3.2 and Lemma 3.5 in [18]). Also, for $f(z) = (z_0 - z)^{-1}$ with $\operatorname{Re} z_0 > 0$, we have, by elementary semigroup theory,

$$\left(\frac{1}{z_0 - z}\right)(A) = \int_0^\infty e^{sA} e^{-sz_0} ds = (z_0 I - A)^{-1},$$

that is, the definition coincides with the definition in terms of the resolvent. We will also need the following lemma of Brenner and Thomée (cf. Lemma 4 in [5]).

LEMMA 3.2. *Let f and g be such that $f, g \in \mathcal{M}$, and $f(z) = z^l g(z)$ for $\operatorname{Re} z \leq 0$ and some $l > 0$. Then we have*

$$f(A)v = g(A)A^l v, \quad \text{for } v \in \mathcal{D}(A^l).$$

For all functions relevant to our discussion, the functional calculus coincides with the definitions in semigroup theory. We therefore do not use different notations in the following and simply write $f(A)$ for a function f of an operator A with respect to our functional calculus. For a better readability and for consistency, we will use $\frac{1}{\gamma - A}$ instead of the equivalent expression $(\gamma I - A)^{-1}$ at some places.

4. Extended Krylov subspace approximation. Let A be the infinitesimal generator of a C_0 -semigroup on a Banach space. We consider the approximation of $\varphi_k(\tau A)v$, $k \geq 0$, for $v \in \mathcal{D}(A^q)$ in the extended Krylov subspace

$$\mathcal{K}_\gamma^{q,n}(A, v) = \operatorname{span} \{A^{q-1}v, \dots, Av, v, (\gamma I - A)^{-1}v, \dots, (\gamma I - A)^{-n+1}v\},$$

with $\gamma > 0$. The following theorem shows that the convergence to $\varphi_k(\tau A)v$ is the faster the higher the index k of the φ_k -function and the smoother the initial data is.

THEOREM 4.1. *Let A generate a bounded C_0 -semigroup. Then, for $v \in \mathcal{D}(A^q)$ and an arbitrary φ_k -function, $k \geq 0$, we have*

$$\inf_{z \in \mathcal{K}_\gamma^{q,n}(A, v)} \|\varphi_k(A)v - z\| \leq C \frac{1}{n^{\frac{q+k}{2}}} \|A^q v\|,$$

where the constant C only depends on γ , q and k .

Without loss of generality, we state our theorems for $\tau = 1$. If A is an infinitesimal generator of a bounded C_0 -semigroup, then $\tilde{A} = \tau A$, for $\tau > 0$, is an infinitesimal generator of a C_0 -semigroup bounded by the same constants relevant to our analysis.

For the proof of our main theorem, we need the following lemma.

LEMMA 4.2. *Let A generate a bounded C_0 -semigroup. Then, for $v \in \mathcal{D}(A^q)$ and an arbitrary φ_k -function, $k \geq 0$, we have*

$$\inf_{z \in \mathcal{K}_\gamma^{q,n}(A, v)} \|\varphi_k(A)v - z\| \leq \inf_{y \in \mathcal{K}_\gamma^{0,n}(A, A^q v)} \|\varphi_{q+k}(A)A^q v - y\|.$$

Proof. The statement directly follows from

$$\{\|\varphi_{q+k}(A)A^q v - y\| \mid y \in \mathcal{K}_\gamma^{0,n}(A, A^q v)\} \subseteq \{\|\varphi_k(A)v - z\| \mid z \in \mathcal{K}_\gamma^{q,n}(A, v)\}. \quad (4.1)$$

In order to show this inclusion, we have to show that we can find for any given b_1, \dots, b_{n-1} coefficients $\tilde{a}_0, \dots, \tilde{a}_{q-1}$ and a_1, \dots, a_{n-1} such that

$$\varphi_{q+k}(A)A^q v - \sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A)^l} A^q v = \varphi_k(A)v - \sum_{l=0}^{q-1} \tilde{a}_l A^l v - \sum_{l=1}^{n-1} a_l \frac{1}{(\gamma - A)^l} v. \quad (4.2)$$

With the help of Lemma 3.2, we obtain by the definition of the φ_k -functions

$$\varphi_{q+k}(A)A^q v = \varphi_k(A)v - \sum_{l=0}^{q-1} \frac{1}{(l+k)!} A^l v. \quad (4.3)$$

Now, we show that

$$\sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A)^l} A^q v = \sum_{l=1}^{n-1} a_l \frac{1}{(\gamma - A)^l} v + p_{q-1}(A)v \quad (4.4)$$

for a polynomial p_{q-1} of degree $q-1$. Here, we assume that $q \leq n-1$ (otherwise, one proceeds analogously). Then, we can split the left-hand side in the following way

$$\sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A)^l} A^q v = \sum_{l=1}^q b_l \frac{1}{(\gamma - A)^l} A^q v + \sum_{l=q+1}^{n-1} b_l \frac{1}{(\gamma - A)^l} A^q v. \quad (4.5)$$

For the case $q < l$, we can conclude with

$$\begin{aligned} \frac{A^q}{(\gamma - A)^l} &= \left(\frac{\gamma}{\gamma - A} - 1 \right)^q \frac{1}{(\gamma - A)^{l-q}} = \sum_{i=0}^q \binom{q}{i} \left(\frac{\gamma}{\gamma - A} \right)^{q-i} (-1)^i \frac{1}{(\gamma - A)^{l-q}} \\ &= \sum_{i=0}^q c_i \frac{1}{(\gamma - A)^{l-i}}, \quad \text{where } c_i = \binom{q}{i} \gamma^{q-i} (-1)^i, \end{aligned} \quad (4.6)$$

that the second sum on the right-hand side of (4.5) has the form

$$\sum_{l=q+1}^{n-1} b_l \frac{1}{(\gamma - A)^l} A^q v = \sum_{l=1}^{n-1} \tilde{b}_l \frac{1}{(\gamma - A)^l} v.$$

For the first sum on the right-hand side of (4.5), we have

$$\sum_{l=1}^q b_l \frac{1}{(\gamma - A)^l} A^q v = \frac{p_{2q-1}(A)}{(\gamma - A)^q} v,$$

where p_{2q-1} is a polynomial of degree $2q-1$. Using polynomial long division, we write this expression in the divisor-quotient form

$$\frac{p_{2q-1}(A)}{(\gamma - A)^q} v = p_{q-1}(A)v + \frac{\tilde{p}_{q-1}(A)}{(\gamma - A)^q} v,$$

where p_{q-1} and \tilde{p}_{q-1} are polynomials with $\deg(p_{q-1}) = q-1$ and $\deg(\tilde{p}_{q-1}) \leq q-1$. With relation (4.6), one can see in an analogous way as above that

$$\frac{\tilde{p}_{q-1}(A)}{(\gamma - A)^q} v = \sum_{l=1}^q \hat{b}_l \frac{1}{(\gamma - A)^l} v.$$

Altogether, this yields the validity of (4.4) with $a_l = \tilde{b}_l + \hat{b}_l$ for $l \leq q$ and $a_l = \tilde{b}_l$ for $l > q$. With (4.3) and (4.4), we finally obtain (4.2). \square

Proof. [of Theorem 4.1] With the help of Lemma 4.2, we have

$$\inf_{z \in \mathcal{K}_\gamma^{q,n}(A,v)} \|\varphi_k(A)v - z\| \leq \inf_{y \in \mathcal{K}_\gamma^{0,n}(A,A^q v)} \|\varphi_{q+k}(A)A^q v - y\|.$$

For an arbitrary

$$y = \sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A)^l} A^q v \in \mathcal{K}_\gamma^{0,n}(A, A^q v),$$

we can conclude that

$$\begin{aligned} \|\varphi_{q+k}(A)A^q v - y\| &= \left\| \varphi_{q+k}(A)A^q v - \sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A)^l} A^q v \right\| \\ &\leq \left\| \varphi_{q+k}(A) - \sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A)^l} \right\| \|A^q v\| \leq C \frac{1}{n^{\frac{q+k}{2}}} \|A^q v\| \end{aligned}$$

by Theorem 4.2 in [17]. This yields

$$\inf_{z \in \mathcal{K}_\gamma^{q,n}(A,v)} \|\varphi_k(A)v - z\| \leq \inf_{y \in \mathcal{K}_\gamma^{0,n}(A,A^q v)} \|\varphi_{q+k}(A)A^q v - y\| \leq C \frac{1}{n^{\frac{q+k}{2}}} \|A^q v\|, \quad (4.7)$$

and our theorem is proved. \square

5. On the choice of γ . In Lemma 4.2, we have seen that the approximation of $\varphi_k(A)v$ in the extended Krylov subspace $\mathcal{K}_\gamma^{q,n}(A, v)$ is related to the approximation of $\varphi_{q+k}(A)A^q v$ in the purely rational Krylov space $\mathcal{K}_\gamma^{0,n}(A, A^q v)$ via

$$\inf_{z \in \mathcal{K}_\gamma^{q,n}(A,v)} \|\varphi_k(A)v - z\| \leq \inf_{y \in \mathcal{K}_\gamma^{0,n}(A,A^q v)} \|\varphi_{q+k}(A)A^q v - y\|. \quad (5.1)$$

As before, the index q is given by the smoothness of the initial vector $v \in \mathcal{D}(A^q)$. The relation (5.1) implies that an improvement of the approximation of the φ_k -functions, $k \geq 0$, in the purely rational Krylov space by an appropriate choice of γ leads to an improvement for the extended Krylov subspace method. To obtain a better upper bound for the approximation error, it thus suffices to restrict our studies to the purely rational case. We will see that the convergence rate for the best rational approximation of the φ_k -function in the so-called resolvent subspace $\mathcal{R}_n := \text{span}\{(\gamma I - A)^{-1}, (\gamma I - A)^{-2}, \dots, (\gamma I - A)^{-n}\}$ which is of order (cf. [17])

$$\inf_{R \in \mathcal{R}_n} \|\varphi_k(A) - R\| = \mathcal{O}\left(\frac{1}{n^{\frac{k}{2}}}\right)$$

can be improved to

$$\inf_{R \in \mathcal{R}_n} \|\varphi_k(A) - R\| = \mathcal{O}\left(\frac{1}{n^{\frac{k}{2}(1 + \frac{r-k}{r+k})}}\right), \quad r < n - 1.$$

As an adequate measure for the convergence rate of the best approximation, moduli of smoothness play a basic and important role in the area of approximation theory. For the improvement of the rational approximation, we have to analyze, in our case, the so-called weighted ϕ -modulus of smoothness ω_ϕ^r . This modulus was introduced by De Bonis, Mastroianni and Viggiano in [6] for the characterization of the K -functional of a function f by its structural properties.

With the Laguerre weight function $w(x) = e^{-x}$, $x > 0$, we denote by L_w^1 the set of all functions for which the relation

$$\|fw\|_1 = \int_0^\infty |w(x)f(x)|dx < \infty$$

holds true. Let $0 < t \leq 1$, $\phi(x) = \sqrt{x}$ and set $I_{rh} = [4r^2h^2, \frac{1}{h^2}]$ with $h > 0$. Then, the main part of the weighted ϕ -modulus of smoothness is defined by

$$\Omega_\phi^r(f, t) := \sup_{0 < h \leq t} \|w \Delta_{h\phi}^r f\|_{L^1(I_{rh})},$$

where $\Delta_{h\phi}^r f$ designates the r th symmetric difference of a function f given by

$$\Delta_{h\phi}^r f(x) = \sum_{j=0}^r (-1)^j \binom{r}{j} f\left(x + \frac{h\phi(x)}{2}(r-2j)\right).$$

The complete modulus $\omega_\phi^r(f, t)$ is now composed of three particular parts in the following way

$$\omega_\phi^r(f, t) := \Omega_\phi^r(f, t) + \inf_{P \in \mathcal{P}_{r-1}} \|w(f - P)\|_{L^1(0, 4r^2t^2)} + \inf_{Q \in \mathcal{P}_{r-1}} \|w(f - Q)\|_{L^1(\frac{1}{t^2}, \infty)},$$

where $0 < t \leq 1$ and \mathcal{P}_{r-1} denotes the set of all algebraic polynomials of maximum degree $r-1$.

Using this weighted ϕ -modulus of smoothness, De Bonis, Mastroianni and Viggiano (cf. [6]) could prove that for all $n, r \in \mathbb{N}$ and $f \in L_w^1$ there exists a constant $C > 0$, independent of n and f , such that the best weighted polynomial approximation is bounded by

$$\inf_{P \in \mathcal{P}_n} \|w(f - P)\|_1 \leq C \omega_\phi^r\left(f, \frac{1}{\sqrt{n}}\right), \quad r < n.$$

This result was used in [17] to obtain an upper bound for the best approximation of $f(A)$, $f \in \widetilde{\mathcal{M}}$, in the resolvent subspace \mathcal{R}_n . For the φ_k -functions, $k \geq 1$, this bound is

$$\inf_{R \in \mathcal{R}_n} \|\varphi_k(A) - R\| \leq \frac{C}{\gamma} \omega_\phi^r\left(\tilde{\varphi}_k, \frac{1}{\sqrt{n-1}}\right) = \mathcal{O}\left(\frac{1}{n^{\frac{k}{2}}}\right), \quad r < n-1, \quad (5.2)$$

where the generic constant C only depends on k and r , and

$$\tilde{\varphi}_k(s) := e^s \mathcal{F}\varphi_{k,(0)}\left(\frac{s}{\gamma}\right) = e^s \mathbf{1}_{[0,1]}\left(\frac{s}{\gamma}\right) \frac{(1 - \frac{s}{\gamma})^{k-1}}{(k-1)!} \in L_w^1.$$

We now search for an optimal shift $\gamma > 0$ depending on the size n of the resolvent subspace \mathcal{R}_n . A slight modification of the occurring constants enables us to examine the modulus $\omega_\phi^r\left(\tilde{\varphi}_k, \frac{1}{\sqrt{n}}\right)$ instead of $\omega_\phi^r\left(\tilde{\varphi}_k, \frac{1}{\sqrt{n-1}}\right)$. We set $\gamma := n^\alpha$ with $\alpha \in \mathbb{R}$ and estimate for arbitrary $0 < t \leq 1$ separately the three particular parts of the r th weighted ϕ -modulus of smoothness for $\tilde{\varphi}_k$, that is $\omega_\phi^r(\tilde{\varphi}_k, t)$. If we finally set $t = n^{-1/2}$ in order to get an upper bound for the three parts of $\omega_\phi^r\left(\tilde{\varphi}_k, \frac{1}{\sqrt{n}}\right)$, we obtain for $r < n-1$ the following results

$$(i) \quad \frac{1}{\gamma} \inf_{P \in \mathcal{P}_{r-1}} \|w(\tilde{\varphi}_k - P)\|_{L^1(0, 4r^2t^2)} \leq \frac{C_1}{n^{\alpha+r+1}} \quad (5.3)$$

$$(ii) \quad \frac{1}{\gamma} \inf_{Q \in \mathcal{P}_{r-1}} \|w(\tilde{\varphi}_k - Q)\|_{L^1(\frac{1}{t^2}, \infty)} = 0 \quad \text{for } \gamma \leq \frac{1}{t^2} \quad (5.4)$$

$$(iii) \quad \frac{1}{\gamma} \Omega_\phi^r(\tilde{\varphi}_k, t) \leq \frac{C_2}{n^{\frac{k}{2}(\alpha+1)}} + \frac{C_3}{n^{\frac{r}{2}(1-\alpha)}} \quad (5.5)$$

with constants C_1, C_2 and C_3 that only depend on r and k . We will see below that with our choice of γ the condition $\gamma \leq \frac{1}{t^2} = \frac{1}{n}$ in (5.4) is always fulfilled. In the following we sketch the basic ideas to show (5.3) – (5.5):

- (i) We choose $P \in \mathcal{P}_{r-1}$ as the $(r-1)$ th Taylor polynomial T_{r-1} of $\hat{\varphi}_k(s) := e^{s \frac{(1-\frac{s}{\gamma})^{k-1}}{(k-1)!}}$ around the point 0 such that

$$\hat{\varphi}_k(s) - T_{r-1}(s) = \frac{\hat{\varphi}_k^{(r)}(\xi)}{r!} s^r = \sum_{l=1}^{r+1} \binom{r}{l-1} \frac{1}{(k-l)!} \frac{(-1)^{l-1}}{\gamma^{l-1}} e^{\xi} \left(1 - \frac{\xi}{\gamma}\right)^{k-l} \frac{s^r}{r!},$$

where $\xi \in [0, s]$. If we substitute this expression in the integral representation of the norm $\|w(\tilde{\varphi}_k - P)\|_{L^1(0, 4r^2 t^2)}$ and use $\xi \leq s \leq 4r^2 t^2$, we obtain the desired estimate.

- (ii) For $\gamma \leq \frac{1}{t^2}$ the intersection of $\text{supp}\{\tilde{\varphi}_k\} = [0, \gamma]$ and the interval $(\frac{1}{t^2}, \infty)$ of integration is empty. This yields

$$\inf_{Q \in \mathcal{P}_{r-1}} \|w(\tilde{\varphi}_k - Q)\|_{L^1(\frac{1}{t^2}, \infty)} \stackrel{Q=0}{\leq} \|w\tilde{\varphi}_k\|_{L^1(\frac{1}{t^2}, \infty)} = 0.$$

- (iii) The tedious procedure to estimate the main part of the ϕ -modulus is as follows. First, one has to consider which s fulfill the condition of the occurring indicator function, that is

$$\frac{1}{\gamma} \left(s + \frac{h\sqrt{s}}{2}(r-2j) \right) \in [0, 1], \quad j \in \{0, 1, \dots, r\}.$$

Then, one can split the integral in Ω_ϕ^r in a suitable way and bound the different parts separately.

Looking at the upper bounds in (5.3), (5.4) and (5.5), we notice that the two summands in the last equation are the most restrictive. Hence, we can conclude that a well-chosen $\gamma = n^\alpha$ should fulfill the condition

$$\frac{k}{2}(\alpha + 1) = \frac{r}{2}(1 - \alpha) \iff \alpha = \frac{r-k}{r+k}$$

in order to get a shift that is as optimal as possible. This choice of γ leads to a convergence rate of order

$$\mathcal{O}\left(\frac{1}{n^{\frac{k}{2}(1+\frac{r-k}{r+k})}}\right) = \mathcal{O}\left(\frac{1}{n^{\frac{rk}{r+k}}}\right). \quad (5.6)$$

By way of illustration, let us consider the following simple example: If we want to approximate the φ_1 -function in the resolvent subspace \mathcal{R}_n and choose for instance $r = 3$, we obtain $\inf_{R \in \mathcal{R}_n} \|\varphi_1(A) - R\| = \mathcal{O}\left(n^{-\frac{3}{4}}\right)$ instead of $\mathcal{O}\left(n^{-\frac{1}{2}}\right)$ predicted in (5.2).

On the one hand, (5.6) shows that the convergence is getting faster and faster for larger values of r . Keep in mind that by assumption the condition $r < n - 1$ must be fulfilled, cf. (5.2). On the other hand, one can calculate that large values of r result unfortunately in ever increasing constants C_1 , C_2 and C_3 in (5.3) and (5.5).

As described at the beginning of this section, the transition to the extended Krylov space requires a modification of our shift γ depending on the maximal index q of smoothness for the initial value. In the extended case the parameter α in the exponent of $\gamma = n^\alpha$ should therefore be chosen as

$$\alpha = \frac{r - (q+k)}{r + (q+k)}.$$

6. Extended Krylov subspace method. Let A be the operator on a Hilbert space with inner product (\cdot, \cdot) . Assume that

$$\text{Re}(Av, v) \leq 0, \quad \forall v \in \mathcal{D}(A), \quad (6.1)$$

and that $\text{Range}(\lambda I - A) = H$ for some λ with $\text{Re} \lambda > 0$. Then, A is the generator of a C_0 -semigroup with

$$\|e^{\tau A}\| \leq 1, \quad \forall \tau \geq 0,$$

that is $N = 1$ in Section 3. The theory of the previous sections applies and we consider the approximation of the semigroup $e^{\tau A}$ in the extended Krylov subspace (1.3)

$$\mathcal{K}_\gamma^{q,n}(A, v) = \text{span} \{ A^{q-1}v, \dots, Av, v, (\gamma I - A)^{-1}v, \dots, (\gamma I - A)^{-n+1}v \},$$

with $\gamma > 0$ as before and $v \in \mathcal{D}(A^q)$. Since we are in a Hilbert space, we designate by P_n the orthogonal projection to $\mathcal{K}_\gamma^{q,n}(A, v)$ and by $A_n := P_n A P_n$ the restriction of A to this subspace. Then, A_n also satisfies

$$\text{Re}(A_n x, x) \leq 0, \quad \forall x \in \mathcal{K}_\gamma^{q,n}(A, v), \quad (6.2)$$

and therefore generates a C_0 -semigroup with the same constant $N = 1$ as above and our functional calculus can be applied again. Since we can now make sense to the operator functions with A and A_n , the following theorem describes the relation between the two.

THEOREM 6.1. *Let $A_n = P_n A P_n$, where P_n is the orthogonal projection onto the extended Krylov subspace (1.3). For $v \in \mathcal{D}(A^q)$ and a function φ_k , $k \geq 0$, we then have*

$$\|\varphi_k(A)v - \varphi_k(A_n)v\| \leq C \frac{1}{n^{\frac{q+k}{2}}} \|A^q v\|,$$

where C only depends on γ , q and k .

Proof. Let k be arbitrarily chosen but fixed. Under our assumptions, we have

$$p((\gamma I - A)^{-1})v = p((\gamma I - A_n)^{-1})v, \quad \forall p \in \mathcal{P}_{n-1},$$

and therefore, for arbitrary a_l , $l = 1, \dots, n-1$,

$$\sum_{l=1}^{n-1} a_l \frac{1}{(\gamma - A)^l} v = \sum_{l=1}^{n-1} a_l \frac{1}{(\gamma - A_n)^l} v.$$

It is easy to see that

$$A^l v = A_n^l v \quad \text{for } 0 \leq l \leq q-1,$$

so that altogether the equation

$$\sum_{l=0}^{q-1} \tilde{a}_l A^l v + \sum_{l=1}^{n-1} a_l \frac{1}{(\gamma - A)^l} v = \sum_{l=0}^{q-1} \tilde{a}_l A_n^l v + \sum_{l=1}^{n-1} a_l \frac{1}{(\gamma - A_n)^l} v$$

can be obtained. Since a_1, \dots, a_{n-1} and $\tilde{a}_0, \dots, \tilde{a}_{q-1}$, have been chosen arbitrarily, we have

$$\mathcal{K}_\gamma^{q,n}(A_n, v) = \mathcal{K}_\gamma^{q,n}(A, v).$$

Let now $y \in \mathcal{K}_\gamma^{q,n}(A_n, v) = \mathcal{K}_\gamma^{q,n}(A, v)$ be arbitrary such that b_1, \dots, b_{n-1} exist with

$$\varphi_k(A)v - y = \varphi_{q+k}(A)A^q v - \sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A)^l} A^q v,$$

cf. (4.2). We then have

$$\begin{aligned} \|\varphi_k(A)v - \varphi_k(A_n)v\| &\leq \|\varphi_k(A)v - y\| + \|\varphi_k(A_n)v - y\| \\ &= \left\| \varphi_{q+k}(A)A^q v - \sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A)^l} A^q v \right\| + \left\| \varphi_{q+k}(A_n)A_n^q v - \sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A_n)^l} A_n^q v \right\| \\ &\leq \left\| \varphi_{q+k}(A) - \sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A)^l} \right\| \|A^q v\| + \left\| \varphi_{q+k}(A_n) - \sum_{l=1}^{n-1} b_l \frac{1}{(\gamma - A_n)^l} \right\| \|A_n^q v\|. \end{aligned}$$

If we use the fact that the rational Krylov subspace $\mathcal{K}_\gamma^{0,n}(A, v)$ can also be written as

$$\mathcal{K}_\gamma^{0,n}(A, v) = \text{span} \left\{ \frac{1}{\gamma - A}v, \frac{A}{(\gamma - A)^2}v, \dots, \frac{A^{n-2}}{(\gamma - A)^{n-1}}v \right\},$$

and proceed similar as in the proof of Theorem 4.1 in [17], one can turn the two problems of approximating $\varphi_{q+k}(A)$ and $\varphi_{q+k}(A_n)$ in the corresponding resolvent subspaces into one single approximation problem in the L^1 -norm. By using our functional calculus, we obtain

$$\|\varphi_{q+k}(A) - s_{n-1}(A)\| \leq \|\mathcal{F}(\varphi_{q+k,(0)} - s_{n-1,(0)})\|_1$$

and

$$\|\varphi_{q+k}(A_n) - s_{n-1}(A_n)\| \leq \|\mathcal{F}(\varphi_{q+k,(0)} - s_{n-1,(0)})\|_1,$$

with

$$s_{n-1}(z) = \sum_{l=1}^{n-1} c_l \frac{z^{l-1}}{(\gamma - z)^l}, \quad \mathcal{F}(s_{n-1,(0)})(s) = \sum_{l=1}^{n-1} c_l (-1)^{l-1} e^{-\gamma s} L_{l-1}(\gamma s), \quad s \geq 0,$$

where L_{l-1} denotes the standard Laguerre polynomial of degree $l-1$. If we choose y from above such that the infimum of the L^1 -problem is attained, which is possible due to (4.2), we get by using Theorem 4.2 in [17]

$$\|\varphi_k(A)v - \varphi_k(A_n)v\| \leq C \frac{1}{n^{\frac{q+k}{2}}} \|A^q v\| + C \frac{1}{n^{\frac{q+k}{2}}} \|A_n^q v\|.$$

The calculation

$$A_n^q v = P_n A P_n A_n^{q-1} v = P_n A P_n \underbrace{A^{q-1} v}_{\in \mathcal{K}_\gamma^{q,n}(A, v)} = P_n A A^{q-1} v = P_n A^q v$$

yields

$$\|A_n^q v\| = \|P_n A^q v\| \leq \|A^q v\|,$$

and therefore

$$\|\varphi_k(A)v - \varphi_k(A_n)v\| \leq C \frac{2}{n^{\frac{q+k}{2}}} \|A^q v\|.$$

□

In contrast to the unknown best approximation in the general case, the approximation $\varphi_k(A_n)v$ to $\varphi_k(A)v$ in the extended Krylov subspace $\mathcal{K}_\gamma^{q,n}(A, v)$ can be efficiently computed. For this purpose, we first determine an orthonormal basis with the following algorithm. Thereby, the algorithm is given in a good human readable form. Obvious improvements with respect to stability

and efficiency should be performed.

```

 $v \in \mathcal{D}(A^q)$ , compute  $\beta = \sqrt{(v, v)} = \|v\|$ 
 $v_1 = v/\beta$ 
for  $m = 1, 2, \dots$  do
  if  $m < q$  do
    for  $j = 1, \dots, m$  do
       $h_{j,m} = (Av_m, v_j)$ 
    end for
     $\tilde{v}_{m+1} = Av_m - \sum_{j=1}^m h_{j,m}v_j$ 
     $h_{m+1,m} = \|\tilde{v}_{m+1}\|$ 
     $v_{m+1} = \tilde{v}_{m+1}/h_{m+1,m}$ 
  else
    for  $j = 1, \dots, m$  do
       $h_{j,m} = ((\gamma I - A)^{-1}v_m, v_j)$ 
    end for
     $\tilde{v}_{m+1} = (\gamma I - A)^{-1}v_m - \sum_{j=1}^m h_{j,m}v_j$ 
     $h_{m+1,m} = \|\tilde{v}_{m+1}\|$ 
     $v_{m+1} = \tilde{v}_{m+1}/h_{m+1,m}$ 
  end if
end for

```

(6.3)

Note that the algorithm first computes a basis of the polynomial part. However, it is easy to see that the vectors v_1, \dots, v_m form an orthonormal basis of $\mathcal{K}_\gamma^{q,n}(A, v)$ for $m = q + n - 1$.

When an orthonormal basis is known, the subspace projection operator and the approximation of an operator function via subspace projection can be formulated conveniently with the help of quasi-matrices (cf. Lecture 5, [35]). In our case, we consider the quasi-matrix

$$V_n := [v_1 \ v_2 \ \cdots \ v_{q+n-1}]$$

whose entries form an orthonormal basis of $\mathcal{K}_\gamma^{q,n}(A, b)$. According to the definitions in [35], we can write our projection operator as $P_n = V_n V_n^H$ and we have

$$P_n v = V_n V_n^H v = V_n \begin{bmatrix} (v, v_1) \\ \vdots \\ (v, v_{q+n-1}) \end{bmatrix} = \sum_{j=0}^{q+n-1} (v, v_j) v_j$$

for an arbitrary vector $v \in H$. The approximation to $f(A)v$ is given by

$$f(A_n)v = f(P_n A P_n)v = f(V_n V_n^H A V_n V_n^H)v = V_n f(V_n^H A V_n) V_n^H v = V_n f(S_n) V_n^H v,$$

(as in the case of matrices), where

$$S_n = \begin{bmatrix} (Av_1, v_1) & \cdots & (Av_m, v_1) \\ \vdots & & \vdots \\ (Av_1, v_m) & \cdots & (Av_m, v_m) \end{bmatrix} \in \mathbb{C}^{m \times m} \quad \text{and} \quad V_n^H v = \begin{bmatrix} (v, v_1) \\ \vdots \\ (v, v_m) \end{bmatrix} \in \mathbb{C}^m$$

is a matrix and a vector, respectively, and $m = q + n - 1$. Hence, after computing the matrix function $f(S_n)$ for the small matrix S_n multiplied by the vector $V_n^H v$, i.e.

$$g_n = f(S_n) V_n^H v,$$

we can compute our approximation as

$$f(A_n)v = V_n f(S_n) V_n^H v = V_n g_n = \sum_{j=0}^{q+n-1} g_j v_j \in H.$$

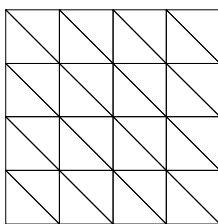


FIG. 7.1. *Regular mesh.*

7. Numerical experiments. In this section, we illustrate our analysis with different approximations of evolution equations. In our first experiment, we use a finite-difference discretization on the unit square that is equivalent to a regular triangulation with linear finite elements and mass lumping. This educational example shows the important fact that the rational Krylov subspace method can be more efficient, despite the fact that the polynomial Krylov method only uses sparse matrix multiplications and the rational Krylov method needs to solve linear systems in every step. Furthermore, the experiments with this approximation are easy to understand and to repeat. In our second experiment, we consider discontinuous Galerkin approximations to Maxwell's equations on a more complex domain. Both experiments clearly show that a grid-independent convergence can be obtained.

7.1. Finite-difference discretization and finite-element discretization. We refer to our motivation in Section 2, but this time we consider the (real) Schrödinger equation on the unit square $\Omega = (0, 1)^2$ with homogeneous Dirichlet boundary conditions. This leads to the operator

$$A = \begin{bmatrix} 0 & B \\ -B & 0 \end{bmatrix},$$

with $H = L^2(\Omega) \times L^2(\Omega)$ and $Bu = -\Delta u$. The domain is given by $\mathcal{D}(A) = \mathcal{D}(B) \times \mathcal{D}(B)$ with $\mathcal{D}(B) = H_0^1(\Omega) \cap H^2(\Omega)$, and the domains of iterated operators are $\mathcal{D}(A^q) = \mathcal{D}(B^q) \times \mathcal{D}(B^q)$, respectively.

We consider the evolution equation

$$y'(\tau) = Ay(\tau), \quad y(0) = y_0 = \begin{bmatrix} q_0 \\ p_0 \end{bmatrix},$$

where we use initial values y_0^q , $q = 2, 4$, with

$$y_0^q = \begin{bmatrix} q_0^q \\ p_0^q \end{bmatrix}, \quad q_0^q = p_0^q = \frac{\tilde{p}_0^q}{\|\tilde{p}_0^q\|}, \quad \tilde{p}_0^q = \begin{cases} [0, 1]^2 & \rightarrow \mathbb{R} \\ \tilde{p}_0^q(x, y) & = x^{2q}(1-x)^{2q}y^{2q}(1-y)^{2q} \end{cases}.$$

With this choice, we have $y_0^q \in \mathcal{D}(A^q)$ and $y_0^q \notin \mathcal{D}(A^{q+1})$. Discretization with finite differences on the standard grid (ih, jh) , $i, j = 1, \dots, d$, with $h = \frac{1}{d+1}$ for a fixed integer d , leads with $N = d^2$ to the $2N \times 2N$ -matrix

$$A_N = \begin{bmatrix} 0 & B_N \\ -B_N & 0 \end{bmatrix}, \quad B_N = \frac{1}{h^2} (T \otimes I_N + I_N \otimes T), \quad T = \text{tridiag}(-1, 2, -1),$$

where \otimes is the Kronecker product. B_N is the standard discretization with the five-point stencil for the negative Laplacian. The inner product is $(u, v) = h^2 v^H u$, $u, v \in \mathbb{C}^N$, and the norm therefore $\|u\|_h = h \|u\|_2$, where $\|\cdot\|_2$ is the standard Euclidean norm. The discretized initial values $y_{0,N}^q$ are just the functions evaluated at the grid points.

The same discretization arises from a regular triangulation with linear finite elements and mass lumping. We will briefly explain how one can see this. We use a regular triangulation with the same nodes as the grid in the finite-difference approximation as shown in Figure 7.1. Then,

one uses the standard N^2 linear ansatz functions. This leads to the system of ordinary differential equations

$$M_N y'(\tau) = \tilde{A}_N y(\tau), \quad y(0) = y_0 = \begin{bmatrix} q_0 \\ p_0 \end{bmatrix}, \quad (7.1)$$

where M_N is the mass matrix and \tilde{A}_N is the stiffness matrix. The idea of mass lumping is not to use the standard inner product, but a discrete one that arises by replacing the integrals over a triangle K by the quadrature formula

$$\int_K f(x) dx = \frac{1}{3} |K| (f(P_1) + f(P_2) + f(P_3)),$$

where $|K|$ designates the area of the triangle and P_i , $i = 1, 2, 3$, are the corners of the triangle. With respect to this inner product, the linear ansatz functions are orthogonal and we have the mass matrix

$$M_N = h^2 \cdot I_{2N^2}, \quad I_{2N^2} \text{ identity matrix of dimension } 2N^2,$$

and the stiffness matrix

$$\tilde{A}_N = \begin{bmatrix} 0 & \tilde{B}_N \\ -\tilde{B}_N & 0 \end{bmatrix}, \quad \tilde{B}_N = (T \otimes I_N + I_N \otimes T), \quad T = \text{tridiag}(-1, 2, -1).$$

The mass matrix determines the inner product with $(u, v)_{M_N} = v^H M_N u = h^2 v^H u$, $u, v \in \mathbb{C}^N$, that is, exactly the same inner product as for the finite-difference approximation. By multiplication from the left-hand side with the mass matrix, the ordinary differential equation (7.1) turns to

$$y'(\tau) = M_N^{-1} \tilde{A}_N y(\tau), \quad y(0) = y_0 = \begin{bmatrix} q_0 \\ p_0 \end{bmatrix}, \quad M_N^{-1} \tilde{A}_N = A_N,$$

where A_N is the matrix for the finite differences. Generally, in the case (7.1), the algorithm (6.3) reads

```

β = ||v||MN, v1 = v/β
for m = 1, 2, ... do
  if m < q do
    for j = 1, ..., m do
      hj,m = (MN-1  $\tilde{A}_N$  vm, vj)MN = vjH  $\tilde{A}_N$  vm
    end for
     $\tilde{v}_{m+1}$  = MN-1  $\tilde{A}_N$  vm - ∑j=1m hj,m vj
    hm+1,m = || $\tilde{v}_{m+1}$ ||MN
    vm+1 =  $\tilde{v}_{m+1}$ /hm+1,m
  else
    for j = 1, ..., m do
      hj,m = ((γI - MN-1  $\tilde{A}_N$ )-1 vm, vj)MN
    end for
     $\tilde{v}_{m+1}$  = (γI - MN-1  $\tilde{A}_N$ )-1 vm - ∑j=1m hj,m vj
    hm+1,m = || $\tilde{v}_{m+1}$ ||MN
    vm+1 =  $\tilde{v}_{m+1}$ /hm+1,m
  end if
end for

```

The algorithm is presented in a good human readable form. And, apart from using modified Gram-Schmidt orthogonalization, we implemented the algorithm as given above. For a general purpose

application of the algorithm, a reorthogonalization might be necessary for large dimensions of the Krylov subspaces and several further improvements with respect to robustness and efficiency should be performed.

The projection $P_n = V_n V_n^H M_N$ to the extended Krylov subspace $\mathcal{K}_\gamma^{q,n}(M_N^{-1} \tilde{A}_N, v)$ with inner product $(u, v)_{M_N} = v^H M_N u$ yields

$$A_n = P_n M_N^{-1} \tilde{A}_N P_n = V_n S_n V_n^H M_N,$$

where $S_n = V_n^H \tilde{A}_N V_n$ and V_n is the matrix that contains the computed orthonormal basis of $\mathcal{K}_\gamma^{q,n}(M_N^{-1} \tilde{A}_N, v)$. Then, the extended Krylov approximation is given by

$$\varphi_k(A_n)v = \|v\|_{M_N} V_n \varphi_k(S_n) e_1,$$

where e_1 is the first unit vector.

As a first experiment, we check, whether the smoothness of the initial data also plays such an important role for finer discretizations than in our motivation example in Section 2. We choose the initial values $y_0^q \in \mathcal{D}(A^q)$ and $y_0^q \notin \mathcal{D}(A^{q+1})$ for $q = 2, 4$, as defined above, and plot the error of the Krylov methods versus the dimension of the Krylov subspaces for step size $\tau = 0.005$, $\gamma = 1$, and $N = 1\,046\,529$, that is A_N is a $2\,093\,058 \times 2\,093\,058$ -matrix. In Figure 7.2, one can see that the standard Krylov method is useful according to the smoothness of the initial data. The gray line indicates, as before, the point where the standard Krylov method starts to use approximations to $A^l y_0^q$ terms with $l \geq q$. The smoothness seems to play exactly the same important role. The starting polynomial Krylov steps in the extended Krylov method might not lead to progress if $\tau^q \|A^q y_0^q\|$ is of order $\mathcal{O}(1)$. The quality of the polynomial approximation is mainly measured via Taylor expansion with this term in the remainder. But if $\tau^q \|A^q y_0^q\|$ is sufficiently small, that is sufficiently small for the continuous data, then the polynomial part in the extended Krylov subspace method can save computation time independent of the space discretization.

The choice of γ can significantly improve the approximation of the exponential in the rational and extended Krylov subspaces. This is illustrated in Figure 7.3, where one can see the improvement of the rational Krylov subspace method with $\gamma = 11^{\frac{1}{3}}$ (diamond-marked and dashed line) over the rational Krylov subspace method with $\gamma = 1$ (star-marked and dashed line) on the left-hand side of Figure 7.3. The shift γ has been chosen according to Section 5 with the highest possible rate for 15 steps and $r = 2q$ for $q = 4$ according to the smoothness of the initial data. Remember that the theory in Section 5 gives results for the purely rational method as well as the extended method. The improvement for the extended method is also significant, what can be seen on the right-hand side of Figure 7.3, where the square-marked and solid line represents the extended Krylov method with $\gamma = 11^{\frac{1}{3}}$ and the circle-marked and solid line represents the extended Krylov method with $\gamma = 1$.

We choose this discretization, because the polynomial Krylov subspace method only needs multiplications with a sparse matrix and is computationally cheap compared to one step of the rational method where a linear system needs to be solved. If one would use linear finite elements without mass lumping, the application of the polynomial Krylov method would also require the solution of a linear system with the mass matrix. Then, the polynomial and rational Krylov subspace method need comparable numerical work in every step. For higher-order finite elements, when no good mass lumping is known, this is anyway true. If the rational variants can be faster in overall computing time for our discretization that favors the polynomial Krylov method, one can dare to say that this will be true for many large systems of discretized partial differential equations that fit into the presented framework.

In Figure 7.4, we compare the performance of the three Krylov methods with respect to error versus computing time. The computation has been conducted in the software environment Matlab, Release R2012a, under Ubuntu, Release 10.4, on a dual core processor with frequency 3GH on a desktop machine. In this example, one step of the polynomial Krylov method is just multiplication with the sparse matrix A_N and this is much cheaper than the computation of $(\gamma I - A_N)^{-1}$ in the rational and the extended Krylov method. $(\gamma I - A_N)^{-1}$ has been computed with a multigrid method without further preconditioning. Nevertheless, the rational and the extended Krylov

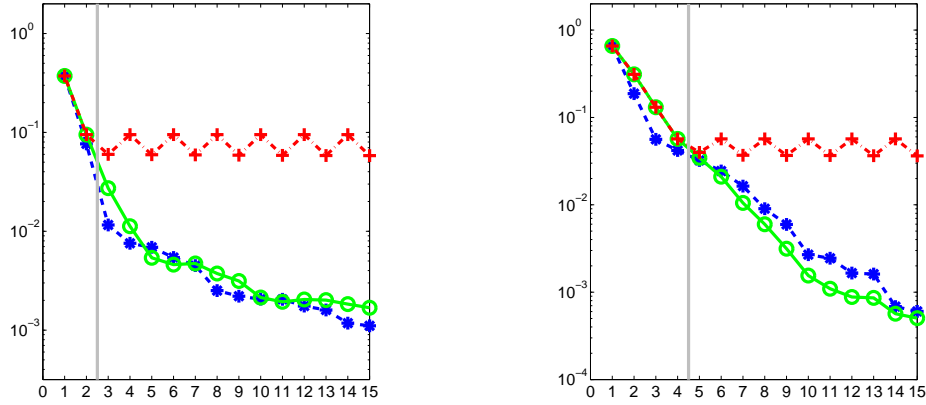


FIG. 7.2. Plot of error vs. dimension of the Krylov space for the standard Krylov subspace $\mathcal{K}_\gamma^{n,0}(A_N, y_{0,N}^q)$ (cross-marked and dash-dotted line), the purely rational Krylov subspace $\mathcal{K}_\gamma^{0,n}(A_N, y_{0,N}^q)$ (star-marked and dashed line) and the extended Krylov subspace $\mathcal{K}_\gamma^{q,n-q}(A_N, y_{0,N}^q)$ (circle-marked and solid line) for $N = 1046529$ and initial values $y_{0,N}^q$ that are discretized initial values of data $y_0^q \in \mathcal{D}(A^q)$ but $y_0^q \notin \mathcal{D}(A^{q+1})$ for $q = 2, 4$.

method are more efficient for the computation of the matrix exponential. As already mentioned, the performance of the standard Krylov method can be estimated with a simple Taylor expansion as long as the smoothness of the initial data allows. If τ becomes larger, then the polynomial part of the extended Krylov method might not lead to a gain in computing time. This is illustrated in Figure 7.5. On the left-hand side of Figure 7.5, the Krylov methods are compared for $N = 65026$, $\gamma = 1$ fixed as before, $y_0^6 \in \mathcal{D}(A^6)$ and $\tau = 0.001$. The step size is smaller and the Taylor expansion suggests a better approximation in the first six polynomial Krylov steps what can be seen in the figure. On the right-hand side, we use the same data but $\tau = 10$. For this τ the remainder in the Taylor expansion is of order $\mathcal{O}(1)$ and it can be clearly observed that the standard Krylov method and the starting procedure in the extended Krylov method make no progress. Since we did not adapt γ , the performance of the rational method is also significantly slower. In the theoretical estimate a term $\tau^6 = 10^6$ appears. This experiment also demonstrates that the analysis that we conducted and which essentially plays the convergence of the extended Krylov subspace method back to the purely rational Krylov subspace method is not an artefact of the analysis. There might be no gain in the first few polynomial Krylov steps. But these steps are cheap and for more realistic (large) time steps, as on the left-hand side of Figure 7.5, the polynomial Krylov part according to the smoothness of the initial data clearly pays off.

In a last experiment, we answer the question, whether one could detect the index of smoothness q automatically in our experiment. We use $y_0^q \in \mathcal{D}(A^q)$ and $y_0^q \notin \mathcal{D}(A^{q+1})$, $q = 2, 4$, as initial value and plot the error of approximation of the matrix exponential for step size $\tau = 0.005$ and $N = 1046529$ versus the dimension of the standard Krylov subspace in Figure 7.6. This is the same initial data as used in Figure 7.2. Besides this, we show

$$\|e^{\tau A_n} y_{0,N}^q - e^{\tau A_{n-1}} y_{0,N}^q\|_{M_N}$$

versus $n \geq 2$, the number of standard Krylov steps taken, as a simple heuristic method to detect the smoothness of the vector. One can clearly see, that this simple idea would have allowed us to detect the index of smoothness of the initial data without prior knowledge of the smoothness of the initial data at the cost of few additional matrix-vector multiplications.

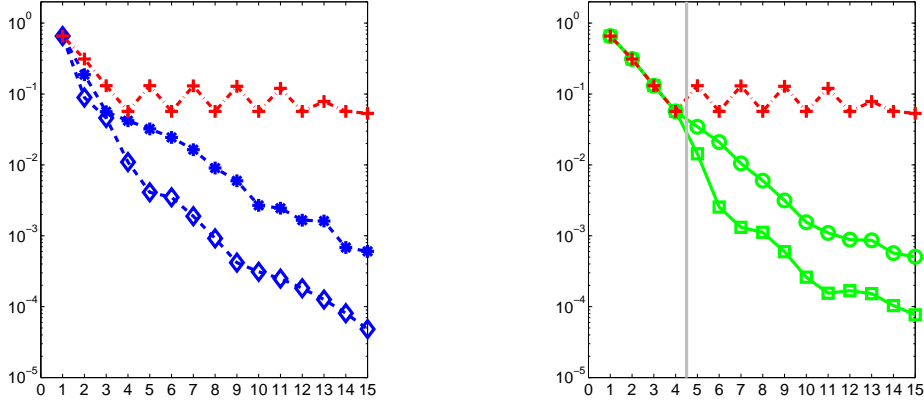


FIG. 7.3. Plot of error vs. dimension of the Krylov space for the standard Krylov subspace $\mathcal{K}_\gamma^{n,0}(A_N, y_{0,N}^4)$ (cross-marked and dash-dotted line), the purely rational Krylov subspace $\mathcal{K}_\gamma^{0,n}(A_N, y_{0,N}^4)$ with $\gamma = 1$ (star-marked and dashed line) and the purely rational Krylov subspace $\mathcal{K}_\gamma^{0,n}(A_N, y_{0,N}^4)$ with $\gamma = 11^{\frac{1}{3}}$ (diamond-marked and dashed line) on the left-hand side and the extended Krylov subspace $\mathcal{K}_\gamma^{4,n-4}(A_N, y_{0,N}^4)$ with $\gamma = 1$ (circle-marked and solid line) and the extended Krylov subspace $\mathcal{K}_\gamma^{4,n-4}(A_N, y_{0,N}^4)$ with $\gamma = 11^{\frac{1}{3}}$ (square-marked and solid line) on the right-hand side for $N = 1\,046\,529$ and initial value $y_{0,N}^4$.

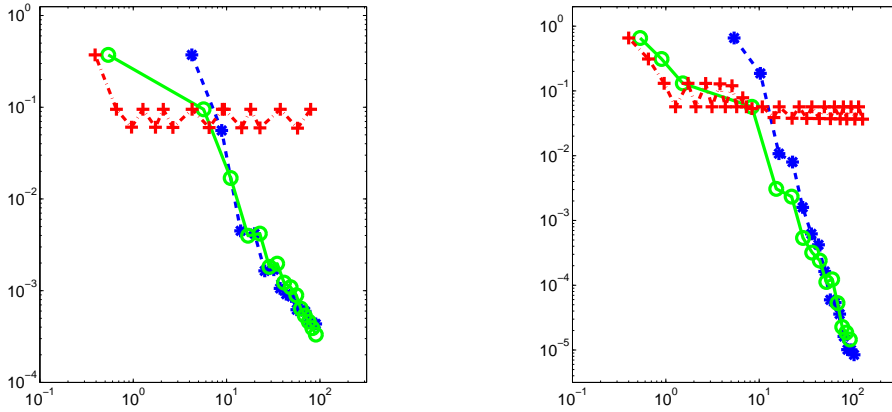


FIG. 7.4. Plot of error vs. time for the standard Krylov subspace $\mathcal{K}_\gamma^{n,0}(A_N, y_{0,N}^q)$ (cross-marked and dash-dotted line), the purely rational Krylov subspace $\mathcal{K}_\gamma^{0,n}(A_N, y_{0,N}^q)$ (star-marked and dashed line) and the extended Krylov subspace $\mathcal{K}_\gamma^{q,n-q}(A_N, y_{0,N}^q)$ (circle-marked and solid line) for $N = 1\,046\,529$ and initial values $y_{0,N}^q$ that are discretized initial values of data $y_0^q \in \mathcal{D}(A^q)$ but $y_0^q \notin \mathcal{D}(A^{q+1})$ for $q = 2, 4$.

7.2. Discontinuous Galerkin discretization. To check the relevance of our analysis for less regular domains, we use normalized Maxwell's equations in transverse magnetic form

$$\begin{aligned} \frac{\partial H^x}{\partial t} &= -\frac{\partial E^z}{\partial y} \\ \frac{\partial H^y}{\partial t} &= \frac{\partial E^z}{\partial x} \\ \frac{\partial E^z}{\partial t} &= \frac{\partial H^y}{\partial x} - \frac{\partial H^x}{\partial y}, \end{aligned}$$

as used in Hesthaven and Warburton [23] on the domain Ω shown in Figure 7.7 with a perfectly electrically conducting boundary. We follow the presentation in [23] closely and use the free Matlab-codes that accompany the book in order to build the stiffness matrix \tilde{A}_N and the mass matrix M_N in equation (7.1) for our experiment. Since we use the central flux in the discontinuous

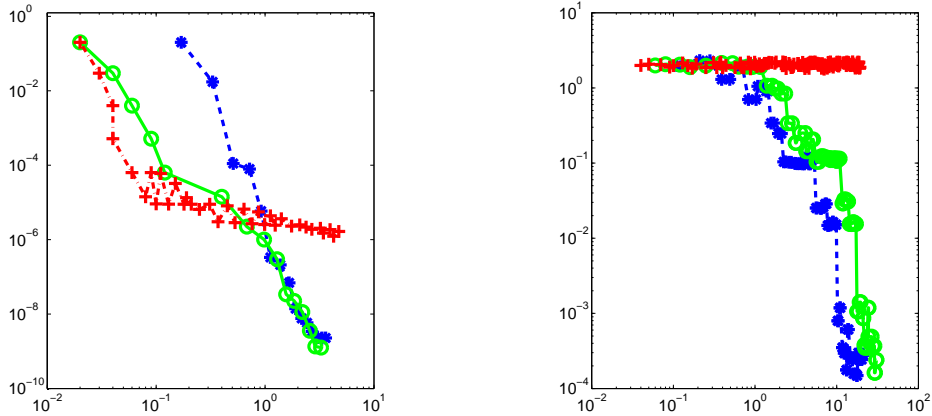


FIG. 7.5. Plot of error vs. time for the standard Krylov subspace $\mathcal{K}_\gamma^{n,0}(A_N, y_{0,N}^q)$ (cross-marked and dash-dotted line), the purely rational Krylov subspace $\mathcal{K}_\gamma^{0,n}(A_N, y_{0,N}^q)$ (star-marked and dashed line) and the extended Krylov subspace $\mathcal{K}_\gamma^{q,n-q}(A_N, y_{0,N}^q)$ (circle-marked and solid line) for $N = 65\,025$ and initial value $y_{0,N}^6$ which is the discretized initial value of data $y_0^6 \in \mathcal{D}(A^6)$ but $y_0^6 \notin \mathcal{D}(A^7)$ for $\tau = 0.001$ on the left-hand side and $\tau = 10$ on the right-hand side.

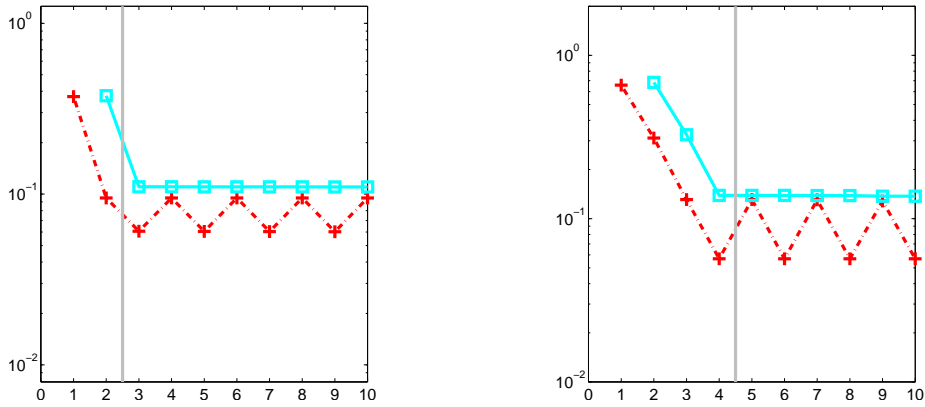


FIG. 7.6. Plot of error vs. dimension of the Krylov space for the standard Krylov subspace $\mathcal{K}_\gamma^{n,0}(A_N, y_{0,N}^q)$ (cross-marked and dash-dotted line) and of $\|e^{\tau A_n} y_{0,N}^q - e^{\tau A_{n-1}} y_{0,N}^q\|_{M_N}$ (square-marked and solid line) as an indicator of smoothness in the standard Krylov subspace method for $N = 1\,046\,529$ and initial values $y_{0,N}^q$ that are discretized initial values of data $y_0^q \in \mathcal{D}(A^q)$ but $y_0^q \notin \mathcal{D}(A^{q+1})$ for $q = 2, 4$.

Galerkin discretization, the eigenvalues of the stiffness matrix are on the imaginary axis, reflecting the hyperbolic nature of the problem. As initial condition, we use a peak in a square in the domain Ω for E^z and zero for H^x and H^y that we evolve with scaled time 1 with the equations and that leads to the initial values shown in Figure 7.8. The chosen peak is in $\mathcal{D}(A^3)$ for the semigroup. Now, we use three grids, one coarser (66 nodes) as the grid shown in Figure 7.7 with 211 nodes and one finer as the grid shown in Figure 7.7 with 935 nodes. The meshes have been generated by DistMesh (cf. [32]). We also use higher order elements on the finer grids. More exactly, first order elements on the coarsest grid, fourth order elements on the second grid, and six order elements on the third and finest grid. As a reference solution, we have used the solution provided by the codes of Hesthaven and Warburton for a small step size. In Figure 7.9, we show the error of our three Krylov subspace methods for the coarsest and the finest grid for a time step $\tau = 0.1$. For the small discretization matrix that belongs to the coarsest grid, the superlinear convergence of the standard Krylov method can be clearly seen and the smoothness of the data does not seem to be important. On the finest grid where the matrix is a $139\,956 \times 139\,956$ -matrix, and for the time-evolved initial value in $\mathcal{D}(A^3)$ on a non-standard domain, the methods behave exactly

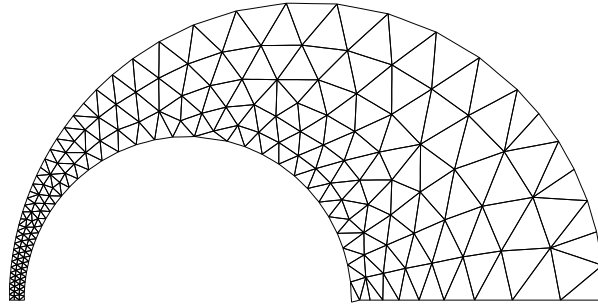


FIG. 7.7. Second mesh with 211 nodes.

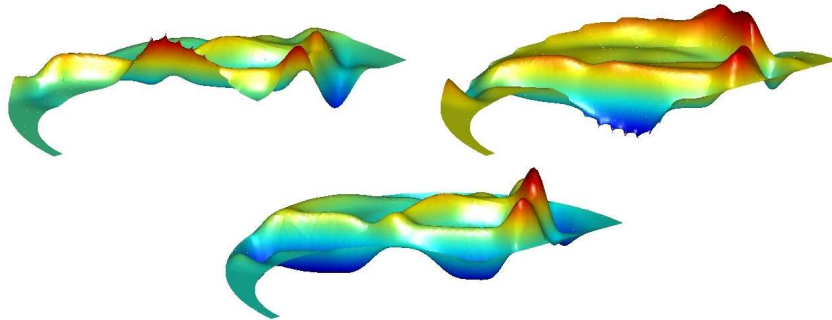


FIG. 7.8. Initial values H^x (top left), H^y (top right), and E^z (bottom).

according to the theory described in this work.

8. Conclusion. We have analyzed the convergence of approximations to the φ -functions by the extended Krylov subspace method, where the size of the polynomial part of the extended Krylov subspace is restricted according to the smoothness of the initial data. We could prove a sublinear convergence rate of the extended Krylov subspace method for the continuous operator. The convergence rate depends on the smoothness of the initial data. The analysis also shows a grid-independent convergence for the discretized operator. Furthermore, we suggested possible choices of the parameter γ that improve the convergence rate for the extended Krylov subspace as well as for the purely rational Krylov subspace. The relevance of the analysis has been illustrated by several evolution equations and for different discretizations.

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

REFERENCES

- [1] M. Afanasjew, M. Eiermann, O. G. Ernst, and S. Güttel. Implementation of a restarted Krylov subspace method for the evaluation of matrix functions. *Linear Algebra Appl.*, 429(10):2293–2314, 2008.
- [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] L. Bergamaschi and M. Vianello. Efficient computation of the exponential operator for large, sparse, symmetric matrices. *Numer. Linear Algebra Appl.*, 7(1):27–45, 2000.
- [5] P. Brenner and V. Thomée. On rational approximations of semigroups. *SIAM J. Numer. Anal.*, 16(4):683–694, 1979.
- [6] M. C. De Bonis, G. Mastroianni, and M. Viggiano. K -functionals, moduli of smoothness and weighted best approximation on the semiaxis. In *Functions, series, operators (Budapest, 1999)*, pages 181–211. János Bolyai Math. Soc., Budapest, 2002.
- [7] 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.

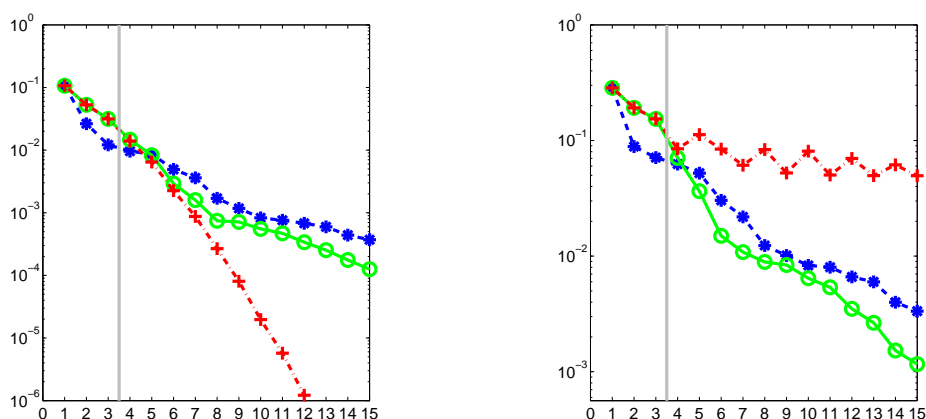


FIG. 7.9. Plot of error vs. dimension of the Krylov space for the standard Krylov subspace $\mathcal{K}_\gamma^{n,0}(A_N, y_{0,N}^a)$ (cross-marked and dash-dotted line), the purely rational Krylov subspace $\mathcal{K}_\gamma^{0,n}(A_N, y_{0,N}^a)$ (star-marked and dashed line) and the extended Krylov subspace $\mathcal{K}_\gamma^{q,n-q}(A_N, y_{0,N}^a)$ (circle-marked and solid line) for $N = 765$ on the left-hand side and for $N = 139356$ on the right-hand side.

- [8] 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.
- [9] 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.
- [10] M. Eiermann and O. G. Ernst. A restarted Krylov subspace method for the evaluation of matrix functions. *SIAM J. Numer. Anal.*, 44(6):2481–2504, 2006.
- [11] M. Eiermann, O. G. Ernst, and S. Güttel. Deflated restarting for matrix functions. *SIAM J. Matrix Anal. Appl.*, 32(2):621–641, 2011.
- [12] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer-Verlag, New York Berlin Heidelberg, 2000.
- [13] H. O. Fattorini. *Second order linear differential equations in Banach spaces*, volume 108 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1985.
- [14] 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.
- [15] 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.
- [16] E. Gallopoulos and Y. Saad. Efficient solution of parabolic equations by krylov approximation methods. *SIAM J. Sci. Statist. Comput.*, 13:1236–1264, 1992.
- [17] V. Grimm. Resolvent Krylov subspace approximation to operator functions. *BIT Numerical Mathematics*, 52(3):639–659, 2012.
- [18] V. Grimm and M. Gugat. Approximation of semigroups and related operator functions by resolvent series. *SIAM J. Numer. Anal.*, 48(5):1826–1845, 2010.
- [19] V. Grimm and M. Hochbruck. Rational approximation to trigonometric operators. *BIT Numerical Mathematics*, 48(2):215–229, 2008.
- [20] S. Güttel. Rational Krylov Methods for Operator Functions, PhD thesis, TU Bergakademie Freiberg, Germany, 2010.
- [21] S. Güttel. Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection. Preprint, 2012.
- [22] M. Haase. The Functional Calculus for Sectorial Operators and Similarity Methods, PhD thesis, University of Ulm, Germany, 2003.
- [23] J. S. Hesthaven and T. Warburton. *Nodal discontinuous Galerkin methods*, volume 54 of *Texts in Applied Mathematics*. Springer, New York, 2008.
- [24] E. Hille and R.S. Phillips. *Functional Analysis and Semi-Groups*. American Mathematical Society, Providence, 1957.
- [25] M. Hochbruck and Ch. Lubich. On Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.*, 34:1911–1925, 1997.
- [26] M. Hochbruck and A. Ostermann. Exponential integrators. *Acta Numer.*, 19:209–286, 2010.
- [27] 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.
- [28] 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.
- [29] I. Moret and P. Novati. RD-rational approximations of the matrix exponential. *BIT Numerical Mathematics*, 44:595–615, 2004.

- [30] I. Moret and P. Novati. Interpolating functions of matrices on zeros of quasi-kernel polynomials. *Numer. Linear Algebra Appl.*, 12(4):337–353, 2005.
- [31] P. Novati. Using the restricted-denominator rational Arnoldi method for exponential integrators. *SIAM J. Matrix Anal. Appl.*, 32(4):1537–1558, 2011.
- [32] P.-O. Persson and G. Strang. A simple mesh generator in Matlab. *SIAM Rev.*, 46(2):329–345, 2004.
- [33] M. Popolizio and V. Simoncini. Acceleration techniques for approximating the matrix exponential operator. *SIAM J. Matrix Anal. Appl.*, 30(2):657–683, 2008.
- [34] A. Ruhe. Rational Krylov sequence methods for eigenvalue computation. *Linear Alg. Appl.*, 58:391–405, 1984.
- [35] G. W. Stewart. *Afternotes goes to graduate school : lectures on advanced numerical analysis*. SIAM, Philadelphia, 1998.
- [36] J. van den Eshof and M. Hochbruck. Preconditioning Lanczos approximations to the matrix exponential. *SIAM J. Sci. Comp.*, 27(4):1438–1457, 2006.