A MOMENT-MATCHING ARNOLDI ITERATION FOR LINEAR COMBINATIONS OF φ FUNCTIONS

ANTTI KOSKELA[*] AND ALEXANDER OSTERMANN[†]

Abstract. The action of the matrix exponential and related φ functions on vectors plays an important role in the application of exponential integrators to ordinary differential equations. For the efficient evaluation of linear combinations of such actions we consider a new Krylov subspace algorithm. By employing Cauchy's integral formula an error representation of the numerical approximation is given. This is used to derive a priori error bounds that describe well the convergence behavior of the algorithm. Further, an efficient a posteriori estimate is constructed. Numerical experiments illustrating the convergence behavior are given in MATLAB.

Key words. error bounds, exponential integrators, phi functions, Krylov subspace methods

AMS subject classifications. 65F60, 65F10, 65L04

1. Introduction.

Matrix functions are of central importance in different fields of science and engineering. The so called φ functions arise in the formulation of exponential integrators [21] when applied to semilinear differential equations

$$u'(t) = Au(t) + g(t, u(t)), \quad u(0) = u_0, \tag{1.1}$$

where $A \in \mathbb{C}^{n\times n}$, $u(t) \in \mathbb{C}^n$, and g is a (non)linear function of t and u. Ordinary differential equations (ODEs) of this form typically arise from spatial semidiscretization of semilinear partial differential equations. Exponential integrators are particularly efficient in the case where the stiffness of the equation (1.1) comes from the linear part, which means that the Lipschitz constant of g is considerably smaller than the norm of A.

By applying exponential integrators to semilinear ODEs, one has to evaluate terms of the form $f(A)b$, i.e., products of matrix functions and vectors. There are several equivalent definitions of matrix functions, see [11], [14], [18], [22]. For exponential integrators the required matrix functions are the so-called φ functions. These are entire functions, closely related to the exponential function, as can be seen from the series representation (3.3) below.

When A comes from the spatial discretization of a differential operator, it is usually sparse and very large. Then, instead of computing the matrix $f(A)$ explicitly, an efficient alternative is to project the problem onto a smaller dimensional subspace, namely onto a *Krylov subspace*. The effectiveness of this approach comes from the fact that generating Krylov subspaces is primarily based on operations of the form $b \to Ab$, which is a cheap operation for sparse A. In fact, by using Krylov subspace methods it was first shown in [20] that exponential integrators can be competitive compared to classical stiff solvers. The approximation of products of the form $f(A)b$ by using Krylov subspace methods has recently been a very active research topic: we mention the work on classical Krylov subspaces [6], [12], [23], [27], extended Krylov subspaces [7], [8], and rational Krylov subspaces [3], [9], [26]. We note, however, that the use of

[*] KTH Royal Institute of Technology, Lindstedtvägen 25, 10044 Stockholm, Sweden (akoskela@kth.se). The work of this author was supported by the Austrian Science Fund (FWF) doctoral program 'Computational Interdisciplinary Modelling' W1227.

[†] Department of Mathematics, University of Innsbruck, Technikerstraße 13, A-6020 Innsbruck, Austria (alexander.ostermann@uibk.ac.at).

extended and rational Krylov subspaces requires the solution of linear systems. This will not be considered here.

More specifically, when employing exponential integrators for the numerical solution of (1.1), one has to compute linear combinations of products of the form $\varphi_\ell(A)w_\ell$, where ℓ denotes the order of the φ function. These linear combinations appear, for instance, in the stages of exponential Runge–Kutta methods [21] and in the time step formulas of exponential Taylor methods [24]. Instead of computing these products separately, an attractive alternative is to compute the linear combination at once using a single product of the form $\exp(\widetilde{A})\widetilde{w}$, where the matrix $\widetilde{A}$ is constructed by augmenting A with the coefficient vectors w_ℓ and one Jordan block (see [1]).

However, even when A is structured (e.g. Hermitian negative semidefinite), the augmented matrix will be unstructured and non-normal, in general, having a field of values considerably larger than that of A. As the convergence analysis of Krylov approximations is based on the field of values of the operator, it becomes difficult to predict how fast the convergence of Krylov approximations for $\exp(\widetilde{A})\widetilde{w}$ will be compared to Krylov approximations for the products $\varphi_\ell(A)w_\ell$. The bounds based on the field of values of $\widetilde{A}$ give largely pessimistic a priori bounds for the convergence of the iteration. Moreover, in order to make the Krylov iteration numerically stable, the augmented matrix needs an additional scaling, which requires some further effort.

The subject of this paper is to propose a new Krylov subspace algorithm to overcome these problems. The convergence analysis of the new algorithm depends only on the field of values of A as can be seen from the obtained convergence bounds. Moreover, the new algorithm avoids the scaling that is needed for the augmented matrix. Finally an effective a posteriori estimate can be obtained based on the notion of the residual.

The rest of the paper is organized as follows. In Section 2 we review the polynomial Krylov approximation of matrix functions and emphasize the approach based on Cauchy's integral formula (originally considered in [20]). In Section 3 we consider the evaluation of linear combinations of products of the form $\varphi_\ell(A)w_\ell$. We give a short proof for the Cauchy integral representation of φ functions and, based on this, discuss the representation of linear combinations of the products $\varphi_\ell(A)w_\ell$ by a single product $\exp(\widetilde{A})\widetilde{w}$, as stated in [1]. Our proof also gives an integral representation for this linear combination, which will be needed later in the error analysis. Further the scaling of the augmented matrix is discussed for the case when the linear combination is computed using the product $\exp(\widetilde{A})\widetilde{w}$. In Section 4 we give a vector-valued Taylor series representation for linear combinations of the products $\varphi_\ell(A)w_\ell$ and introduce a projection based approach to approximate this series. We formulate an Arnoldi-like algorithm to obtain this approximation and show that the approximation gives a Padé-type approximant of an order of the size of the underlying Krylov subspace. The main result of Section 5 is Theorem 5.1, which gives a bound for the error of the approximation based on the field values of the operator A. The derivation is based on an exact representation of the error with the help of Cauchy's integral formula. The proof closely follows the analysis carried out in [20] for the approximation of the matrix exponential. As a special case, we derive analytical bounds for the case of Hermitian negative semidefinite A. Section 6 is devoted to a posteriori error estimates. In Theorem 6.1 we show that the error can be represented as the exact solution of an ODE. The error is based on the notion of a residual that can be obtained from the iteration. As a consequence of this theorem, we derive an a posteriori error estimate for the approximation. Section 7 concludes with numerical experiments that illustrate the effectiveness of the approximation for several test problems.

2. Preliminaries.

The research carried out in this article is motivated by questions arising in the time integration of stiff problems of the form (1.1). Exponential integrators for such problems are constructed and analyzed with the help of the variation-of-constants formula

$$u(t) = \mathrm{e}^{(t-t_0)A} u_0 + \int_{t_0}^{t} \mathrm{e}^{(t-\tau)A} g(\tau, u(\tau))\, \mathrm{d}\tau. \tag{2.1}$$

As an example we give the simplest exponential integrator, the exponential Euler method, which is obtained by interpolating the integrand in (2.1) using the known value $u(t_0) = u_0$. The method is given by

$$u_1 = \mathrm{e}^{hA} u_0 + h\varphi_1(hA) g(t_0, u_0),$$

where h denotes the step size and φ_1 the entire function

$$\varphi_1(z) = \frac{\mathrm{e}^z - 1}{z}.$$

The efficient computation of the action of this and similar φ functions is *the* central problem in the implementation of exponential integrators. In this work, Krylov subspace methods will be used for that purpose.

2.1. Krylov subspace methods.

Krylov subspace methods are based on the idea of projecting a high dimensional problem involving a matrix $A \in \mathbb{C}^{n\times n}$ and a vector $b \in \mathbb{C}^n$ onto a low dimensional subspace $\mathcal{K}_k(A, b)$, which is defined by

$$\mathcal{K}_k(A, b) = \operatorname{span}\{b, Ab, A^2 b, \ldots, A^{k-1} b\}.$$

If the generating vectors are linearly independent, this subspace has dimension k and the well-known Arnoldi iteration provides an orthonormal basis $\{q_1, \ldots, q_k\}$ of $\mathcal{K}_k(A, b)$, which is usually written as a matrix $Q_k = [q_1, \ldots, q_k] \in \mathbb{C}^{n\times k}$, and a Hessenberg matrix $H_k = Q_k^* A Q_k \in \mathbb{C}^{k\times k}$, which represents the action of A in the subspace $\mathcal{K}_k(A, b)$. If A is Hermitian, then H_k is tridiagonal and we get the Lanczos iteration. Moreover, the recursion

$$AQ_k = Q_k H_k + h_{k+1,k} q_{k+1} e_k^{\mathsf{T}} \tag{2.2}$$

holds, where $h_{k+1,k}$ denotes the corresponding entry in H_{k+1} and e_k is the kth standard basis vector in $\mathbb{C}^k$.

The numerical solution of linear systems $Ax = b$ constitutes a typical application for Krylov subspace methods. There, one way to determine the Krylov approximation $x_k \in \mathcal{K}_k(A, b)$ to the solution x is to use the Galerkin condition $Ax_k - b \perp \mathcal{K}_k(A, b)$, which implies

$$x_k = Q_k H_k^{-1} e_1 \|b\|_2. \tag{2.3}$$

2.2. Matrix functions.

Cauchy's integral formula states that any analytic function f defined on a domain $D \subset \mathbb{C}$ satisfies

$$f(z) = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma} \frac{f(\lambda)}{\lambda - z}\, \mathrm{d}\lambda,$$

where Γ is an appropriate contour in D enclosing z. This formula can be used to define matrix functions. Replacing z by a square matrix A and considering a path Γ inside D that encloses the spectrum $\sigma(A)$, the matrix function $f(A)$ is defined by

$$f(A) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma f(\lambda)(\lambda I - A)^{-1} \, \mathrm{d}\lambda. \tag{2.4}$$

For the equivalence of various definitions of matrix functions, we refer to [22] and [18].

A Krylov subspace approximation for the product $f(A)b$ is obtained by inserting the approximate value (see (2.3))

$$(\lambda I - A)^{-1} b \approx Q_k (\lambda I - H_k)^{-1} e_1 \|b\|_2$$

into Cauchy's integral formula (2.4). More precisely, we apply (2.4) to the vector b and choose a contour that encloses the *field of values* of A:

$$\mathcal{F}(A) = \{x^* A x \, : \, x \in \mathbb{C}^n, \|x\|_2 = 1\}.$$

This choice is motivated by the requirement that the contour has to enclose both $\sigma(A)$ and $\sigma(H_k)$. Obviously, $\mathcal{F}(A)$ contains the spectrum $\sigma(A)$, and since the columns of Q_k are orthonormal, we have the inclusions

$$\sigma(H_k) \subset \mathcal{F}(H_k) = \mathcal{F}(Q_k^* A Q_k) \subset \mathcal{F}(A).$$

This justifies the sought-after approximation

$$f(A)b \approx \frac{1}{2\pi \mathrm{i}} \int_\Gamma f(\lambda) Q_k (\lambda I - H_k)^{-1} e_1 \|b\|_2 \, \mathrm{d}\lambda = Q_k f(H_k) e_1 \|b\|_2. \tag{2.5}$$

For example, the product $\mathrm{e}^A b$ is approximated by $\mathrm{e}^A b \approx \|b\|_2 Q_k \mathrm{e}^{H_k} e_1$.

3. Computing series of φ functions.

Throughout this paper, we consider the entire functions

$$\varphi_0(z) = \mathrm{e}^z, \qquad \varphi_\ell(z) = \int_0^1 \mathrm{e}^{(1-\theta)z} \frac{\theta^{\ell-1}}{(\ell-1)!} \, \mathrm{d}\theta, \quad \ell \geq 1 \tag{3.1}$$

which play a crucial role in exponential integrators. They satisfy $\varphi_\ell(0) = \frac{1}{\ell!}$ and the recurrence relation

$$\varphi_{\ell+1}(z) = \frac{\varphi_\ell(z) - \varphi_\ell(0)}{z}, \quad \ell \geq 0. \tag{3.2}$$

From (3.2) it follows that their series representation is given by

$$\varphi_\ell(z) = \sum_{j=0}^{\infty} \frac{z^j}{(j+\ell)!}. \tag{3.3}$$

Our goal is a fast approximation of expressions of the form

$$u(h) = \sum_{\ell=0}^{p} h^\ell \varphi_\ell(hA) w_\ell, \tag{3.4}$$

where $A \in \mathbb{C}^{n \times n}$, $h \in \mathbb{R}$, and $w_\ell \in \mathbb{C}^n$, $0 \leq \ell \leq p$. Such expressions play a fundamental role in exponential integrators.

We first give an auxiliary result, which was already stated in [30] with a different proof.

LEMMA 3.1. *Let Γ be a closed contour encircling the points 0 and $z \in \mathbb{C}$ with winding number one. Then*

$$\varphi_\ell(z) = \frac{1}{2\pi i} \int_\Gamma \frac{\mathrm{e}^\lambda}{\lambda^\ell} \frac{1}{\lambda - z} \, \mathrm{d}\lambda. \tag{3.5}$$

Proof. The lemma is proved by induction. For $\ell = 0$, relation (3.5) clearly holds. For $\ell > 0$, we use the decomposition

$$\frac{1}{\lambda^{\ell+1}(\lambda - z)} = \frac{1}{z} \left(\frac{1}{\lambda^\ell(\lambda - z)} - \frac{1}{\lambda^{\ell+1}} \right),$$

and the recursion (3.2). □

For the fast computation of (3.4), the following lemma will be of importance. The result itself is not new and can be found, e.g., in [1, Thm. 2.1]. Our proof, however, is different and provides an integral representation for (3.4) which will be needed in the error analysis below.

LEMMA 3.2. *Let $A \in \mathbb{C}^{n\times n}$, $W = [w_p, w_{p-1}, \ldots, w_1] \in \mathbb{C}^{n\times p}$, $h \in \mathbb{R}$ and*

$$\widetilde{A} = \begin{bmatrix} A & W \\ 0 & J \end{bmatrix} \in \mathbb{C}^{(n+p)\times(n+p)}, \quad J = \begin{bmatrix} 0 & I_{p-1} \\ 0 & 0 \end{bmatrix}, \quad \widetilde{w}_0 = \begin{bmatrix} w_0 \\ e_p \end{bmatrix} \in \mathbb{C}^{n+p} \tag{3.6}$$

with $e_p = [0, \ldots, 0, 1]^\mathsf{T}$ and I_p the identity matrix of size p. Then we have

$$u(h) = \sum_{\ell=0}^{p} h^\ell \varphi_\ell(hA) w_\ell = \begin{bmatrix} I_n & 0 \end{bmatrix} \mathrm{e}^{h\widetilde{A}} \widetilde{w}_0. \tag{3.7}$$

Proof. From (3.6) we see that $\sigma(h\widetilde{A}) = \sigma(hA) \bigcup \{0\}$. Now let Γ be a contour in $\mathbb{C}$ that encircles $\sigma(h\widetilde{A})$. Using Cauchy's integral formula, we get the representation

$$\mathrm{e}^{h\widetilde{A}} \widetilde{w}_0 = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \mathrm{e}^\lambda (\lambda I - h\widetilde{A})^{-1} \widetilde{w}_0 \, \mathrm{d}\lambda. \tag{3.8}$$

One can easily verify the following representation for the resolvent $(\lambda I - h\widetilde{A})^{-1}$:

$$(\lambda I - h\widetilde{A})^{-1} = \begin{bmatrix} (\lambda I - hA)^{-1} & (\lambda I - hA)^{-1} W h (\lambda I - hJ)^{-1} \\ 0 & (\lambda I - hJ)^{-1} \end{bmatrix}.$$

As $h(\lambda I - hJ)^{-1}$ is explicitly given by

$$h(\lambda I - hJ)^{-1} = \frac{h}{\lambda}\Big(I - \frac{h}{\lambda} J\Big)^{-1} = \begin{bmatrix} \frac{h}{\lambda} & \frac{h^2}{\lambda^2} & \frac{h^3}{\lambda^3} & \cdots & \frac{h^p}{\lambda^p} \\ & \frac{h}{\lambda} & \frac{h^2}{\lambda^2} & & \\ & & \frac{h}{\lambda} & \ddots & \vdots \\ & & & \ddots & \frac{h^2}{\lambda^2} \\ & & & & \frac{h}{\lambda} \end{bmatrix}, \tag{3.9}$$

we infer that

$$\begin{bmatrix} I_n & 0 \end{bmatrix} (\lambda I - h\widetilde{A})^{-1} \widetilde{w}_0 = (\lambda I - hA)^{-1} \sum_{\ell=0}^{p} \frac{h^\ell}{\lambda^\ell} w_\ell.$$

By inserting this result into (3.8) and employing Lemma 3.1, we finally get

$$\begin{aligned}\begin{bmatrix} I_n & 0 \end{bmatrix} \mathrm{e}^{h\widetilde{A}} \widetilde{w}_0 &= \frac{1}{2\pi \mathrm{i}} \int_\Gamma \mathrm{e}^{\lambda} (\lambda I - hA)^{-1} \sum_{\ell=0}^{p} \frac{h^\ell}{\lambda^\ell} w_\ell \, \mathrm{d}\lambda \\ &= \sum_{\ell=0}^{p} h^\ell \varphi_\ell(hA) w_\ell,\end{aligned} \tag{3.10}$$

which is the desired result. □

With the help of Lemma 3.2, the vector $u(h)$ can be computed as the action of the exponential function of an enlarged matrix. This approach, combined with a scaling and squaring, was pursued in [1].

Note that the very structure of (3.7) easily allows one to scale the vectors w_j. For an invertible matrix $S \in \mathbb{C}^{p \times p}$, the scaled version of (3.7) is given by

$$\begin{aligned} u(h) &= \begin{bmatrix} I_n & 0 \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & S^{-1} \end{bmatrix} \mathrm{e}^{h\widetilde{A}} \begin{bmatrix} I_n & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} w_0 \\ S^{-1} e_p \end{bmatrix} \\ &= \begin{bmatrix} I_n & 0 \end{bmatrix} \exp\left(h \begin{bmatrix} A & WS \\ 0 & S^{-1}JS \end{bmatrix} \right) \begin{bmatrix} w_0 \\ S^{-1} e_p \end{bmatrix}. \end{aligned} \tag{3.11}$$

This scaling of $\widetilde{A}$ affects the accuracy of (3.7) when the matrix exponential is computed with a scaling and squaring strategy; see [1], where the authors propose the simple scaling $S = \eta I_p$ for some suitably chosen η to avoid large $\|W\|$. More precisely, they took $\eta = 2^{-\lceil \log_2(\|W\|_1) \rceil}$. This scaling also affects the propagation of the round-off errors in the Arnoldi iteration when using the approximation (2.5) for the product (3.7). This will also be illustrated in the numerical experiments.

The above scaling can further be used to efficiently compute linear combinations of the form

$$\widehat{u}(h) = \mathrm{e}^{hA} w_0 + h \sum_{\ell=1}^{p} \varphi_\ell(hA) \widehat{w}_\ell \tag{3.12}$$

which arise in exponential integrators, see [21]. By choosing

$$S = \operatorname{diag}(h^{p-1}, \ldots, h, 1),$$

one obtains the representation

$$\widehat{u}(h) = \begin{bmatrix} I_n & 0 \end{bmatrix} \exp\left(h \begin{bmatrix} A & \widehat{w}_p, \ldots, \widehat{w}_1 \\ 0 & \frac{1}{h} J \end{bmatrix} \right) \begin{bmatrix} w_0 \\ e_p \end{bmatrix}, \tag{3.13}$$

which avoids the badly scaled vectors $h^{1-\ell} \widehat{w}_\ell$ that would arise when using (3.7).

4. A moment-matching Arnoldi iteration. By inserting the Taylor series representation (3.3) of the φ functions into (3.4) we obtain

$$u(h) = \sum_{\ell=0}^{p} h^\ell \sum_{j=0}^{\infty} \frac{h^j}{(j+\ell)!} A^j w_\ell = \sum_{j=0}^{\infty} \sum_{\ell=0}^{p} \frac{h^{j+\ell}}{(j+\ell)!} A^j w_\ell.$$

This can be expressed as a Taylor series

$$u(h) = \sum_{\nu=0}^{\infty} \frac{h^\nu}{\nu!} m_\nu \tag{4.1}$$

with moments $m_\nu = \sum_{\ell=0}^{\min(\nu,p)} A^{\nu-\ell} w_\ell$ that satisfy the recursion

$$m_0 = w_0, \tag{4.2a}$$
$$m_\nu = A m_{\nu-1} + w_\nu, \qquad \text{for } 1 \le \nu \le p, \tag{4.2b}$$
$$m_\nu = A m_{\nu-1}, \qquad \text{for } \nu > p. \tag{4.2c}$$

This shows that $u(h)$ lies in the closure of span$\{m_0, m_1, \ldots\}$ and consequently leads us to define the *enriched Krylov subspace*

$$\mathcal{K}_k(A, w_0, \ldots, w_p) = \operatorname{span}\{m_0, m_1, \ldots, m_{k-1}\}, \tag{4.3}$$

where $A \in \mathbb{C}^{n\times n}$, and the basis vectors m_i are given by the recursion (4.2).

The main idea here is to perform an Arnoldi-like iteration for obtaining an orthonormal basis $Q_k \in \mathbb{C}^{n\times k}$ of the subspace $\mathcal{K}_k(A, w_0, \ldots, w_p)$, and to approximate $u(h)$ in this lower dimensional space by using the projection of A onto this subspace,

$$F_k = Q_k^* A Q_k. \tag{4.4}$$

To this end, we set $W = [w_p, \ldots, w_1]$ and J as in (3.6),

$$V = [v_p, \ldots, v_1] = Q_k^* W, \quad v_0 = Q_k^* w_0, \quad \widetilde{F}_k = \begin{bmatrix} F_k & V \\ 0 & J \end{bmatrix} \quad \text{and} \quad \widetilde{v}_0 = \begin{bmatrix} v_0 \\ e_p \end{bmatrix}.$$

The sought-after approximation $u_k(h) \approx u(h)$ is given by

$$u_k(h) = Q_k v_k(h), \tag{4.5a}$$

where, due to Lemma 3.2,

$$v_k(h) = \sum_{\ell=0}^{p} h^\ell \varphi_\ell(hF_k) v_\ell = \begin{bmatrix} I_k & 0 \end{bmatrix} \mathrm{e}^{h\widetilde{F}_k} \widetilde{v}_0. \tag{4.5b}$$

This approximation has the following property.

Lemma 4.1. *The approximation* (4.5) *gives a kth order approximant for $u(h)$, i.e., it holds*

$$u(h) - u_k(h) = \mathcal{O}(h^k).$$

Proof. From (4.5), we see that

$$u_k(h) = Q_k \sum_{j=0}^{\infty} \frac{h^j}{j!} \widetilde{m}^j,$$

where the vectors $\widetilde{m}_j$ satisfy the recursion (4.2) with A replaced by F_k and w_ν replaced by v_ν. We will show by induction that

$$\widetilde{m}_j = Q_k^* m_j, \qquad 0 \le j < k.$$

For $j = 0$, this relation clearly holds. For $1 \le j \le p$, we have by induction

$$\widetilde{m}_j = F_k \widetilde{m}_{j-1} + v_j = Q_k^* A Q_k Q_k^* m_{j-1} + Q_k^* w_j.$$

Since $m_j \in \mathcal{R}(Q_k)$ for $j < k$, we have $Q_k Q_k^* m_j = m_j$ and consequently by (4.2b)

$$\widetilde{m}_j = Q_k^* \left(A m_{j-1} + w_j \right) = Q_k^* m_j.$$

For $j > p$, the result follows from (4.2c), since then

$$\widetilde{m}_j = F_k \widetilde{m}_{j-1} = Q_k^* A Q_k Q_k^* m_{j-1} = Q_k^* A m_{j-1} = Q_k^* m_j.$$

Thus,

$$u_k(h) = Q_k \sum_{\ell=0}^{k-1} \frac{h^\ell}{\ell!} Q_k^* m_\ell + \mathcal{O}(h^k) = \sum_{\ell=0}^{k-1} \frac{h^\ell}{\ell!} m_\ell + \mathcal{O}(h^k), \tag{4.6}$$

which implies the assertion of the lemma. □

Similar methods for matching moments of transfer functions using Krylov subspace techniques can be found in the context of reduced-order modeling of large-scale systems, see e.g. [2] and [10].

An interesting characterization, similar to the polynomial approximation property of the Krylov subspaces, is the following: for every polynomial q of degree at most $k-1$ it holds that

$$q(A) w_0 + q^{(1)}(A) w_1 + \cdots + q^{(p)}(A) w_p \in \mathcal{K}_k(A, w_0, \ldots, w_p),$$

where

$$q^{(\ell)}(z) = \frac{q^{(\ell-1)}(z) - q^{(\ell-1)}(0)}{z}, \quad 1 \le \ell \le p, \quad q^{(0)}(z) = q(z).$$

This follows from the definition of the subspace (4.3) and the recursion (4.2).

From the recursion (4.2b) we see that the enriched subspace (4.3) can also be viewed as an *augmented* Krylov subspace

$$\mathcal{K}_k(A, w_0, \ldots, w_p) = \mathcal{K}_{k-p}(A, m_p) + \mathcal{W}, \tag{4.7}$$

where $\mathcal{W} = \operatorname{span}\{m_0, \ldots, m_{p-1}\}$. In augmented Krylov subspace methods (see e.g. [13], [28]) the Krylov subspace is commonly enriched with eigenvectors or *almost* eigenvectors of A. For example, augmenting with almost eigenvectors corresponding to the smallest eigenvalues of A may speed up the convergence of iterative solvers of linear systems. For theoretical results, see [28].

In our situation, the vectors m_k spanning $\mathcal{W}$ are not almost eigenvectors of A, in general. However, an orthonormal basis Q_k of $\mathcal{K}_k(A, w_0, \ldots, w_p)$ can be obtained by performing an Arnoldi iteration for $\mathcal{K}_{k-p}(A, m_p)$ and then augmenting the produced basis by $\mathcal{W}$. Note that in this way short orthogonalization recursions are possible in case A is (skew-) Hermitian. A drawback, however, is that the explicit computation of the vector m_p may make the algorithm numerically unstable. Moreover, an Arnoldi-like relation (4.11) does not hold, which makes the derivation of a posteriori estimates more difficult.

In the next subsection we derive an iteration which gives an almost Arnoldi-like relation (4.10) employable for error estimates, and which avoids the explicit computation of the vectors $m_1, \ldots, m_p$. The decomposition (4.7) also shows up in the error analysis of Section 5, where the error estimation is carried out using polynomial approximation techniques in the complex plane.

We further note that the series (3.4) could also be approximated by using *block* Krylov subspace methods (see e.g. [16], [25], [29]), since

$$\mathcal{K}_k(A, w_0, \dots, w_p) \subset \mathcal{K}_k(A, w_0) + \mathcal{K}_{k-1}(A, w_1) + \cdots + \mathcal{K}_{k-p}(A, w_p).$$

This can be again seen from the recursion (4.2). However, since running a block Krylov method needs p times more matrix-vector multiplications than our approach, we will not consider block Krylov methods here.

4.1. Arnoldi-like iteration for the enriched Krylov subspace. In this subsection we describe a Krylov iteration for constructing an orthonormal basis of the enriched Krylov subspace (4.3). This is achieved by using the QR decomposition of the matrix

$$K_{k+1} = [m_0, m_1, \dots, m_k]. \tag{4.8}$$

In the kth step, we start from the unique QR decomposition

$$K_k = Q_k R_k,$$

where the diagonal elements of R_k are all positive, and the columns of Q_k constitute an orthonormal basis of $\mathcal{K}_k(A, w_0, \dots, w_p)$. The vector q_{k+1} is obtained by orthogonalizing m_k against the columns of Q_k. If K_{k+1} has full rank, this vector is unique upon choosing $r_{k+1,k+1} > 0$. In this way, we obtain

$$K_{k+1} = Q_{k+1} R_{k+1}, \tag{4.9a}$$

where the matrices Q_{k+1} and R_{k+1} have the following structure:

$$Q_{k+1} = \begin{bmatrix} Q_k & q_{k+1} \end{bmatrix}, \qquad R_{k+1} = \begin{bmatrix} R_k & \times \\ 0 & r_{k+1,k+1} \end{bmatrix}. \tag{4.9b}$$

The next lemma shows the existence of a relation similar to (2.2) for this Arnoldi-like iteration.

LEMMA 4.2. *For $A \in \mathbb{C}^{n\times n}$ and $\{w_0, ..., w_p\} \subset \mathbb{C}^n$ there exists a matrix $F_k \in \mathbb{C}^{k\times k}$ such that*

$$AQ_k = Q_k F_k - (I - Q_k Q_k^*)\, \widehat{W}_k\, R_k^{-1} + h_{k+1,k} q_{k+1} e_k^{\mathsf{T}} \tag{4.10}$$

with Q_{k+1} and R_k as in (4.9). The matrix $\widehat{W}_k \in \mathbb{C}^{n\times k}$ is defined as

$$\widehat{W}_k = \begin{cases} [w_1, w_2, \dots, w_k] & \textit{for } k \le p, \\ [w_1, w_2, \dots, w_p, 0, \dots, 0] & \textit{for } k > p, \end{cases}$$

and $h_{k+1,k}$ is related to the entry $f_{k+1,k} = (F_{k+1})_{k+1,k}$ as

$$h_{k+1,k} = f_{k+1,k} + q_{k+1}^{\mathsf{T}} \widehat{W}_k R_k^{-1} e_k.$$

Proof. From the definition of the moments (4.2), we deduce the recursion

$$K_{k+1} = \begin{bmatrix} w_0 & AK_k \end{bmatrix} + \begin{bmatrix} 0 & \widehat{W}_k \end{bmatrix}.$$

Let

$$t_1 = [r_{11}, 0, \dots, 0]^\mathsf{T} \in \mathbb{C}^k, \qquad t_2 = [0, \dots, 0, r_{k+1,k+1}]^\mathsf{T} \in \mathbb{C}^k.$$

Using the QR decomposition of K_{k+1} we get

$$\begin{aligned} K_{k+1} = Q_{k+1} R_{k+1} &= \begin{bmatrix} Q_k & q_{k+1} \end{bmatrix} \begin{bmatrix} t_1 & \widetilde{H}_k \\ 0 & t_2 \end{bmatrix} \\ &= \begin{bmatrix} w_0 & A Q_k R_k \end{bmatrix} + \begin{bmatrix} 0 & \widehat{W}_k \end{bmatrix}, \end{aligned}$$

where $\widetilde{H}_k = R_{k+1}(1:k, 2:k+1) \in \mathbb{C}^{k \times k}$ is a Hessenberg matrix. From this we infer the following relation

$$A Q_k R_k + \widehat{W}_k = Q_k \widetilde{H}_k + r_{k+1,k+1} q_{k+1} e_k^\mathsf{T},$$

and furthermore

$$A Q_k + \widehat{W}_k R_k^{-1} = Q_k H_k + h_{k+1,k} q_{k+1} e_k^\mathsf{T}, \tag{4.11}$$

where H_k is the Hessenberg matrix $H_k = \widetilde{H}_k R_k^{-1}$, and $h_{k+1,k} = r_{k+1,k+1}/r_{k,k}$ the corresponding entry of H_{k+1}. The $(k \times k)$ matrix (4.4) is now given by

$$F_k = Q_k^* A Q_k = H_k - Q_k^* \widehat{W}_k R_k^{-1}. \tag{4.12}$$

Substituting (4.12) to (4.11), we finally get (4.10). □

4.2. The algorithm. Multiplying (4.11) by e_k from the right-hand side yields

$$A q_k + \widehat{W}_k R_k^{-1} e_k = Q_k H_k e_k + h_{k+1,k} q_{k+1}, \tag{4.13}$$

which directly gives an algorithm to update the basis Q_k for the enriched Krylov subspace $\mathcal{K}_k(A, w_0, \dots, w_p)$:

1. Compute the vector $s := A q_k + \widehat{W}_k R_k^{-1} e_k$;
2. Orthogonalize s against Q_k to obtain $h_{k+1,k} q_{k+1}$ and the kth column of H_k;
3. Normalize $h_{k+1,k} q_{k+1}$;
4. Update the $(k+1)$st column of R_{k+1} with help of the relations

   $$R_{k+1}(1:k, k+1) = \widetilde{H}_k e_k = \widetilde{H}_k R_k^{-1} R_k e_k = H_k R_k e_k$$

   and $h_{k+1,k} = r_{k+1,k+1}/r_{k,k}$.

A pseudocode for this algorithm is provided in Algorithm 1. We recall that the matrices R_{k+1} and H_k are defined recursively by (4.9) and (4.12).

The $(k \times k)$ matrix F_k can be obtained by the multiplication $Q_k^* A Q_k$ at the end, which is a relatively cheap operation if A is sparse. This approach is used in all of the numerical experiments.

However, when orthogonalizing the vectors $A q_k$ and $\widehat{W}_k R_k^{-1} e_k$ separately in the Gram–Schmidt procedure, also the upper-triangular part of F_k is obtained. If A is Hermitian, then also is F_k, and the whole F_k is obtained. This option doubles the number of inner-products during the iteration, but avoids the matrix multiplication

Algorithm 1 Moment-matching Arnoldi iteration for the enriched Krylov subspace $\mathcal{K}_k(A, w_0, ..., w_p)$ using the modified Gram–Schmidt orthogonalization, see also (4.13)

```
$r_{1,1} \leftarrow \|w_0\|$
$q_1 \leftarrow w_0/r_{1,1}$
**for** $i = 1, 2, \ldots, k-1$ **do**
    $l \leftarrow \min(i, p)$
    $s \leftarrow R_i^{-1} e_i$
    $s \leftarrow Aq_i + \widehat{W}_l s(1:l)$
    **for** $j = 1$ to $i$ **do**
        $h_{j,i} \leftarrow (s, q_j)$
        $s \leftarrow s - h_{j,i} q_j$
    **end for**
    $R_{i+1}(1:i, i+1) \leftarrow H_i R_i(1:i, i)$
    $h_{i+1,i} \leftarrow \|s\|_2$
    $q_{i+1} \leftarrow s/h_{i+1,i}$
    $r_{i+1,i+1} \leftarrow r_{i,i} h_{i+1,i}$
**end for**
```

$Q_k^* A Q_k$. In the non-Hermitian case, the lower-triangular part of F_k can be computed also by saving the vectors AQ_k and performing the missing inner-products at the end of the iteration.

To evaluate the small dimensional expression in (4.5),

$$\sum_{\ell=0}^{p} h^\ell \varphi_\ell(hF_k) v_\ell = \begin{bmatrix} I_k & 0 \end{bmatrix} \mathrm{e}^{h\widetilde{F}_k} \widetilde{v}_0,$$

we have used the Padé approximation of the matrix exponential with scaling and squaring (command `expm` in MATLAB [19]). Another possibility would be to evaluate the terms $\varphi_\ell(hF_k)v_\ell$ separately using Padé approximations of the φ functions [18, Thm. 10.31].

As the matrix A may have a very large norm (e.g. when A is coming from a spatial discretization of a differential operator), the norms of the columns of the matrix K_k in (4.8) grow rapidly. This results in an ill-conditioned upper triangular matrix R_k of the QR decomposition of K_k and in stagnation of the iteration. This is a drawback compared to the classical Arnoldi iteration, where the upper-triangular matrix R_k of the QR decomposition of the Krylov matrix does not appear explicitly. However, this problem can be circumvented by employing higher precision for R_k, as R_k is always of relatively moderate size. This is done in the numerical experiments of Section 7. Moreover, when using double precision, this stagnation was found to happen at Krylov subspace sizes sufficient for accurate time integration of PDEs. In each numerical experiment we point out the threshold caused by the ill-conditioning of R_k.

5. An a priori error representation. Our error analysis of the moment-matching Arnoldi iteration is strongly guided by the ideas presented in [20]. However, we note that other techniques could be used to show the a priori superlinear convergence of the method, see e.g. [3].

Let Γ be a closed contour that encircles the numerical range of hA with winding number one. The application of Lemma 3.1 to (3.4) and (4.5), respectively, provides us with the following integral representations of the solution

$$u(h) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \mathrm{e}^\lambda (\lambda I - hA)^{-1} \sum_{\ell=0}^{p} \frac{h^\ell}{\lambda^\ell} w_\ell \, \mathrm{d}\lambda$$

and the numerical approximation

$$u_k(h) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \mathrm{e}^\lambda Q_k (\lambda I - hF_k)^{-1} Q_k^* \sum_{\ell=0}^{p} \frac{h^\ell}{\lambda^\ell} w_\ell \, \mathrm{d}\lambda.$$

The error $\varepsilon_k(h)$ of the moment-matching approximation (4.5) after $k \geq p$ steps is thus given by

$$\varepsilon_k(h) = u(h) - u_k(h) = \frac{1}{2\pi \mathrm{i}} \int\limits_\Gamma \mathrm{e}^\lambda \Delta_{k,\lambda} \sum_{\ell=0}^{p} \frac{h^\ell}{\lambda^\ell} w_\ell \, \mathrm{d}\lambda, \tag{5.1a}$$

where

$$\Delta_{k,\lambda} = (\lambda I - hA)^{-1} - Q_k (\lambda I - hF_k)^{-1} Q_k^*. \tag{5.1b}$$

To estimate the right-hand side of (5.1a), we first note that

$$\Delta_{k,\lambda} (\lambda I - hA) K_k = 0$$

and recall from (4.2b) the relation

$$AK_p = K_{p+1} \begin{bmatrix} 0 \\ I_p \end{bmatrix} - \widehat{W}_p.$$

With $x = \left[\frac{1}{\lambda}, \frac{h}{\lambda^2}, \ldots, \frac{h^{p-1}}{\lambda^p} \right]^\mathsf{T}$ we find that

$$\sum_{\ell=0}^{p} \frac{h^\ell}{\lambda^\ell} w_\ell - (\lambda I - hA) K_p x = \frac{h^p}{\lambda^p} m_p,$$

and consequently

$$\Delta_{k,\lambda} \sum_{\ell=0}^{p} \frac{h^\ell}{\lambda^\ell} w_\ell = \Delta_{k,\lambda} \left(\frac{h^p}{\lambda^p} m_p + (\lambda I - hA) K_k \begin{bmatrix} 0 \\ y \end{bmatrix} \right)$$

for all $k > p$ and all $y \in \mathbb{C}^{k-p}$. We have thus shown that

$$\Delta_{k,\lambda} \sum_{\ell=0}^{p} \frac{h^\ell}{\lambda^\ell} w_\ell = h^p \Delta_{k,\lambda} p_{k-p}(hA) m_p$$

for every polynomial p_{k-p} of degree at most $k-p$, normalized by $p_{k-p}(\lambda) = \lambda^{-p}$. Our approach to bound the error $\varepsilon_k(h)$ is to use this freedom in (5.1a) to choose different polynomial p_{k-p} for each $\lambda \in \Gamma$ on some contour $\Gamma \subset \mathbb{C}$ and to estimate the representation

$$\varepsilon_k(h) = \frac{h^p}{2\pi \mathrm{i}} \int_\Gamma \mathrm{e}^\lambda \Delta_{k,\lambda} p_{k-p}(hA) m_p \, \mathrm{d}\lambda. \tag{5.2}$$

From [31, Thm. 5.1] we infer that

$$\|\Delta_{k,\lambda}\| \le 2\,\mathrm{dist}\big(\Gamma, \mathcal{F}(A)\big)^{-1}.$$

If A is normal, the estimate

$$\|p_{k-p}(hA)\| \le \max_{z\in\sigma(A)} |p_{k-p}(z)|$$

is straightforward. For arbitrary A, also a theorem by Crouzeix [5] can be used, which states that

$$\|p(A)\| \le 11.08 \max_{z\in\mathcal{F}(A)} |p(z)|$$

for every polynomial p and matrix $A \in \mathbb{C}^{n\times n}$. In our analysis, Cauchy's integral formula is used to bound $\|p_{k-p}(hA)\|$.

We are now in the position to state our first error bound, which is a slight adaptation of [20, Lemma 1].

Theorem 5.1. *Let E be a convex and compact subset of $\mathbb{C}$ containing the field of values of hA, let ϕ be the conformal mapping that maps the exterior of E onto the exterior of the unit circle (with $\phi(z) = z/\varrho + O(1)$ for $z \to \infty$, where ρ is the logarithmic capacity of E), and let Γ be the boundary curve of a piecewise smooth, bounded region G that contains E. Then*

$$\|\varepsilon_k(h)\|_2 \le \frac{M\|m_p\|_2 h^p}{2\pi} \int_\Gamma \left|\mathrm{e}^\lambda\right| \cdot |\phi(\lambda)|^{-(k-p)} \cdot |\lambda|^{-p} \cdot |\mathrm{d}\lambda|\,, \tag{5.3}$$

where $M = \mathrm{length}(\partial E)/\big(\mathrm{dist}\big(\partial E, \mathcal{F}(hA)\big) \cdot \mathrm{dist}\big(\Gamma, \mathcal{F}(hA)\big)\big)$. If E is a straight line segment, then (5.3) *holds with $M = 6/\,\mathrm{dist}\big(\Gamma, \mathcal{F}(hA)\big)$.*

Proof. The proof is very similar to that of [20, Lemma 1]. Taking into account (5.2), we choose now the polynomial

$$p_{k-p}(z) = \frac{\phi_{k-p}(z) - \phi_{k-p}(\lambda) + \phi(\lambda)^{k-p}}{\phi(\lambda)^{k-p}\lambda^p},$$

where $\phi_{k-p}(z)$ is the Faber polynomial of degree $k-p$ related to $\phi(z)$. This choice implies that

$$\max_{z\in E} |p_{k-p}(z)| \le \frac{3}{|\phi(\lambda)^{k-p}| \cdot |\lambda^p|},$$

and the result now follows literally as in the proof of [20, Lemma 1]. □

5.1. Hermitian negative semidefinite case. In the case of a Hermitian matrix hA with eigenvalues in the interval $E = [-4\rho, 0]$, we proceed as in [20]. The affine transformation

$$\lambda \mapsto \mu = 1 + \frac{\lambda}{2\rho}$$

maps E to $[-1, 1]$ and reveals that $\phi(\lambda) = \Phi(\mu) = \mu + \sqrt{\mu^2 - 1}$. In these new coordinates, we choose as contour the parabola with parametrization

$$\Pi : \theta \mapsto \mu = (1+\epsilon)\Big(1 - \frac{1}{2}\theta^2\Big) + \mathrm{i}\theta\sqrt{2\epsilon + \epsilon^2}, \quad -\infty < \theta < \infty.$$

This parabola is osculating to the ellipse with vertex $\mu = 1+\epsilon$ and foci ± 1. Therefore, its minimal distance to $[-1,1]$ is ϵ, i.e. $\lambda = 2\rho\epsilon$ in the original coordinates. From (5.3) we thus get the estimate

$$\varepsilon_k(h) \leq \frac{3h^p \|m_p\|_2}{\pi\epsilon(2\rho)^p} \int_\Pi \left| \mathrm{e}^{2\rho(\mu-1)} \right| \cdot |\Phi(\mu)|^{-(k-p)} \cdot |\mu - 1|^{-p} \cdot |\mathrm{d}\mu| . \tag{5.4}$$

For $\mu = \mu(\theta)$ on the parabola, we have the following estimates:

$$\begin{aligned} &\Phi(1+\epsilon) \leq |\Phi(\mu)|, && |\mu - 1|^{-1} \leq a(\theta^2 + c)^{-1}, \\ &\mathrm{Re}(\mu - 1)2\rho \leq 2\epsilon\rho - \rho\theta^2, && |\mathrm{d}\mu| \leq \Big((1+\epsilon)\,|\theta| + \sqrt{2\epsilon + \epsilon^2} \Big)\, |\mathrm{d}\theta| \end{aligned}$$

with $a = \frac{2}{1+\epsilon}$, $c = \frac{2\epsilon}{(1+\epsilon)^2}$. Inserting theses estimates into (5.4) gives us the following result.

Theorem 5.2. *Let A be a Hermitian negative semidefinite matrix such that hA has its eigenvalues in the interval $[-4\rho, 0]$. Then the error of the moment-matching Arnoldi iteration* (4.5) *after $k \geq p$ steps satisfies the bound*

$$\varepsilon_k(h) \leq \frac{3h^p \mathrm{e}^{2\rho\epsilon} \|m_p\|_2}{\pi\epsilon\rho^p \big(\Phi(1+\epsilon)\big)^{k-p} (1+\epsilon)^p} \Big(\sqrt{2\epsilon + \epsilon^2}\, \mathcal{I}_p + 2(1+\epsilon)\mathcal{J}_p \Big) \tag{5.5a}$$

with

$$\mathcal{I}_p = \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-\rho\theta^2}}{(\theta^2 + c)^p} \mathrm{d}\theta \quad \textit{and} \quad \mathcal{J}_p = \int_0^{\infty} \frac{\theta \mathrm{e}^{-\rho\theta^2}}{(\theta^2 + c)^p} \mathrm{d}\theta. \tag{5.5b}$$

The parameter $\epsilon > 0$ in this bound can be chosen freely for each k.

Integration by parts shows that the integrals (5.5b) satisfy the following recurrence relations (for $p \geq 1$):

$$\begin{aligned} &\mathcal{I}_{p+1} = \frac{2p - 1 - 2c\rho}{2pc} \mathcal{I}_p + \frac{\rho}{pc} \mathcal{I}_{p-1}, \qquad \mathcal{I}_0 = \sqrt{\frac{\pi}{\rho}}, \quad \mathcal{I}_1 = \frac{\pi \mathrm{e}^{\rho c}}{\sqrt{c}} \big(1 - \mathrm{erf}(\sqrt{\rho c})\big), \\ &\mathcal{J}_{p+1} = \frac{1}{2pc^p} - \frac{\rho}{p} \mathcal{J}_p, \qquad \mathcal{J}_0 = \frac{1}{2\rho}, \quad \mathcal{J}_1 = -\frac{\mathrm{e}^{\rho c}}{2} \mathrm{Ei}(-\rho c), \end{aligned}$$

where $\mathrm{Ei}(x)$ denotes the exponential integral (see [15, formula 8.2111]):

$$\mathrm{Ei}(-x) = -\int_x^\infty \frac{\mathrm{e}^{-t}}{t} \mathrm{d}t, \quad x > 0.$$

The result for $\mathcal{I}_1$ is formula 3.4661 in [15], that for $\mathcal{J}_1$ follows from

$$\int_0^\infty \frac{\theta \mathrm{e}^{-\rho\theta^2}}{\theta^2 + c} \mathrm{d}\theta = \frac{1}{2} \int_0^\infty \frac{\mathrm{e}^{-\rho s}}{s + c} \mathrm{d}s = \frac{1}{2} \int_{\rho c}^\infty \frac{\mathrm{e}^{\rho c - t}}{t} \mathrm{d}t.$$

Figure 7.1 depicts on the right the upper bound (5.5a) compared to the actual convergence for the symmetric problem of subsection 7.1. The bound is optimized numerically with respect to the free parameter ϵ.

6. An a posteriori error estimate. In this section we follow the ideas presented in [4] and [9], where a residual notion for the approximation of the matrix exponential is used. The following theorem gives the residual (as defined in [4]) for the moment-matching approximation.

THEOREM 6.1. *The error $\varepsilon_k(t) = u(t) - u_k(t)$ of the moment-matching approximation is the solution of the ODE*

$$\varepsilon_k'(t) = A\varepsilon_k(t) + r_k(t), \qquad \varepsilon_k(0) = 0,$$

where the residual $r_k(t)$ is given by

$$r_k(t) = h_{k+1,k}q_{k+1}e_k^{\mathsf{T}}v_k(t) + (I - Q_kQ_k^*)\Bigg(\sum_{\ell=1}^{p} \frac{t^{\ell-1}}{(\ell-1)!}w_\ell - \widehat{W}_kR_k^{-1}v_k(t)\Bigg). \tag{6.1}$$

Proof. With the help of the variation-of-constants formula (2.1) and the definition of the φ functions (3.1), we see that $u(t)$ as given in (3.4) is the solution of the inhomogeneous differential equation

$$u'(t) = Au(t) + \sum_{\ell=1}^{p} \frac{t^{\ell-1}}{(\ell-1)!}w_\ell, \qquad u(0) = w_0. \tag{6.2}$$

Likewise, we see from (4.5) that $u_k(t)$ solves the differential equation

$$u_k'(t) = Q_kv_k'(t) = Q_kF_kv_k(t) + Q_kQ_k^*\sum_{\ell=1}^{p} \frac{t^{\ell-1}}{(\ell-1)!}w_\ell, \qquad v_k(0) = Q_k^*w_0. \tag{6.3}$$

Substituting the relation (4.10) to (6.3), we find that $u_k(t)$ satisfies

$$u_k'(t) = Au_k(t) + (I - Q_kQ_k^*)\widehat{W}_kR_k^{-1}v_k(t) - h_{k+1,k}q_{k+1}e_k^{T}v_k(t) + Q_kQ_k^*\frac{t^{\ell-1}}{(\ell-1)!}w_\ell.$$

Consequently, the error $\varepsilon_k(t) = u(t) - u_k(t)$ is the solution of

$$\varepsilon_k'(t) = A\varepsilon_k(t) + r_k(t), \qquad \varepsilon_k(0) = 0,$$

with $r_k(t)$ given by (6.1). □

The variation-of-constants formula (2.1) yields the representation

$$\varepsilon_k(h) = \int_0^h \mathrm{e}^{(h-s)A}r_k(s)\,\mathrm{d}s. \tag{6.4}$$

Note that all quantities in (6.1) are computable. Therefore, (6.4) (or an appropriate approximation to it) can be used as an error estimate.

A directly computable estimate for (6.4) is obtained by making the approximation (see (4.6) and (4.8))

$$u_k(t) \approx \sum_{\ell=0}^{k-1} \frac{t^\ell}{\ell!}m_\ell = K_k\begin{bmatrix}1 & t & \dots & \frac{t^{\ell-1}}{(\ell-1)!}\end{bmatrix}^{\mathsf{T}},$$

i.e.,

$$R_k^{-1}v_k(t) \approx \begin{bmatrix}1 & t & \dots & \frac{t^{\ell-1}}{(\ell-1)!}\end{bmatrix}^{\mathsf{T}}, \tag{6.5}$$

as

$$u_k(t) = Q_k v_k(t) = Q_k R_k R_k^{-1} v_k(t) = K_k R_k^{-1} v_k(t).$$

Substituting (6.5) to (6.1) gives

$$r_k(t) \approx h_{k+1,k} q_{k+1} e_k^{\mathsf{T}} v_k(t). \tag{6.6}$$

Substituting (6.6) into (6.4), and recalling (4.5), we obtain an approximation

$$\varepsilon_k(h) \approx h_{k+1,k} \int_0^h \mathrm{e}^{(h-s)A} q_{k+1} e_k^{\mathsf{T}} \begin{bmatrix} I_k & 0 \end{bmatrix} \mathrm{e}^{s\widetilde{F}_k} \widetilde{v}_0 \, \mathrm{d}s, \tag{6.7}$$

where

$$\widetilde{F}_k = \begin{bmatrix} F_k & V \\ 0 & J \end{bmatrix} \quad \text{and} \quad \widetilde{v}_0 = \begin{bmatrix} v_0 \\ e_p \end{bmatrix}.$$

Expressing the matrix exponential $\mathrm{e}^{(h-s)A}$ in terms of the φ_1 function shows that

$$\mathrm{e}^{(h-s)A} q_{k+1} = q_{k+1} + (h-s)\varphi_1((h-s)A) A q_{k+1}.$$

A cheaply computable estimate is obtained by neglecting the second term. We thus define

$$err_k = h_{k+1,k}\, q_{k+1}\, h\, e_k^{\mathsf{T}} \begin{bmatrix} I_k & 0 \end{bmatrix} \varphi_1(h\widetilde{F}_k)\widetilde{v}_0. \tag{6.8}$$

Figure 7.1 shows on the right the efficiency of the estimate (6.8) when the iteration is applied to the symmetric problem of subsection 7.1.

7. Numerical experiments.

To illustrate the behavior of the new algorithm, we compare it with the standard Arnoldi iteration for some test problems. We also illustrate the effect of the scaling (3.11) for the augmented matrix $\widetilde{A}$. Finally, we illustrate the benefit of using the moment-matching Arnoldi iteration for the implementation of exponential Runge–Kutta methods.

The experiments are carried out on a desktop machine using MATLAB. No comparisons of execution times are made. We simply note that using the current implementation, the moment-matching Arnoldi iteration is about twice as slow as the Arnoldi iteration applied to the exponential of the augmented matrix $\widetilde{A}$ (see (3.7)).

7.1. Illustration of convergence and error estimates.

In the first numerical experiment we take a diagonal matrix $A \in \mathbb{R}^{200\times 200}$ and vectors $w_0, ..., w_5 \in \mathbb{R}^{200}$ such that the elements $w_{k,i}$ are independently normally distributed with variance 10^{2k}, $0 \le k \le 5$. To illustrate the convergence, we compare the moment-matching Arnoldi approximation of the sum (3.4) to the Arnoldi approximation of the product $\mathrm{e}^{hA} w_0$.

In the first comparison the diagonal elements of A are set as

$$a_{ii} = -4\rho \sin^2\left(\frac{i\pi}{2(n+1)}\right), \quad 1 \le i \le n \le 200,$$

which gives a Hermitian matrix. In the second experiment, we choose

$$a_{ii} = -4\mathrm{i}\rho \sin^2\left(\frac{i\pi}{2(n+1)}\right), \quad 1 \le i \le n \le 200,$$

which gives a skew-Hermitian matrix. Note that the above entries are the eigenvalues of the central difference discretization of the one-dimensional Laplacian Δ and $\mathrm{i}\,\Delta$, respectively, with homogeneous Dirichlet boundary conditions. We set $\rho = 80$, and $h = 0.1$. Then $\|hA\|_2 \approx 32$, and $\|h^k w_k\|_2 \approx 13$, $1 \leq k \leq 5$, which corresponds to a stiff nonlinearity g in (1.1).

The left part of Figure 7.1 shows that in both cases the relative error of the moment-matching approximation for (3.4) is very close to that of the Arnoldi approximation for $\mathrm{e}^{hA} w_0$. In the skew-symmetric case the convergence is considerably slower than in the symmetric case, as can be expected from the analysis of [20].

To illustrate the a priori error bound (5.5) and the a posteriori bound (6.8), we take the Hermitian case from above with $h = 0.05$. The results are shown on the right of Figure 7.1. We note that in double precision the algorithm stagnated at around Krylov subspace size 30.

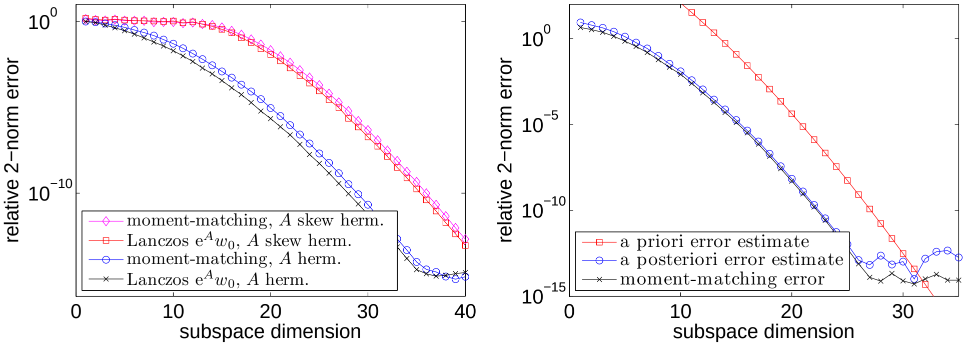

Fig. 7.1. *Left: Convergence of the moment-matching iteration for* (3.4) *and of the Arnoldi iteration for* $\mathrm{e}^A w_0$*, both for the Hermitian and skew-Hermitian test cases. Right: Error bounds and the actual convergence of the moment-matching iteration for the Hermitian test case.*

7.2. Effect of scaling of $\widetilde{A}$ on the Arnoldi iteration and selection of an optimal scaling parameter. The following numerical example illustrates the effect of the scaling (3.11) on the convergence of the Arnoldi approximation of $u(h) = \begin{bmatrix} I_n & 0 \end{bmatrix} \mathrm{e}^{h\widetilde{A}} \widetilde{w}_0$, when the simple scaling $S = \eta I$ is used.

We take a matrix $A \in \mathbb{R}^{100 \times 100}$ with elements A_{ij} that are normally distributed with variance $\sigma^2 = 100$, and set the vectors $w_0, \ldots, w_5$ such that the elements of w_i, $0 \leq i \leq 5$, are normally distributed with variance $\sigma^2 = 5000^{2i}$. The time step h is set to 0.25. This is an extremal case as the norms of the vectors $h^k w_k$ are considerably larger than that of hA. This corresponds for example to one step of a high-order exponential Taylor method [24] applied to a semilinear equation (1.1), where the stiffness arises from the nonlinear part.

Figure 7.2 shows the relative errors for the approximations of (3.4) obtained with the moment-matching iteration and the Arnoldi iteration applied to the exponential of the augmented matrix with scalings $\eta = 10^{-5}$, 10^{-10}, 10^{-15}, 10^{-20}. We note that in double precision the moment matching iteration did not stagnate.

We see that the scaling has a considerable effect on the numerical stability of the Arnoldi iteration, and that with the scaling $\eta = 10^{-20}$ the convergence of the Arnoldi method and moment-matching iteration are almost identical. In [1] the scaling $S = \eta I$ with $\eta = 2^{-\lceil \log_2(\|W\|_1) \rceil}$ is used, which scales the (1,2)-block of the augmented matrix

$\widetilde{A}$ down to an order of one and speeds up the Taylor method [1] for computing $\mathrm{e}^{h\widetilde{A}}\widetilde{u}_0$. In this case $\|W\|_2 \approx 2 \cdot 10^{20}$, so that $\eta = \|W\|_2^{-1}$ gives likewise a good scaling factor when using the Arnoldi iteration to approximate $\mathrm{e}^{h\widetilde{A}}\widetilde{u}_0$.

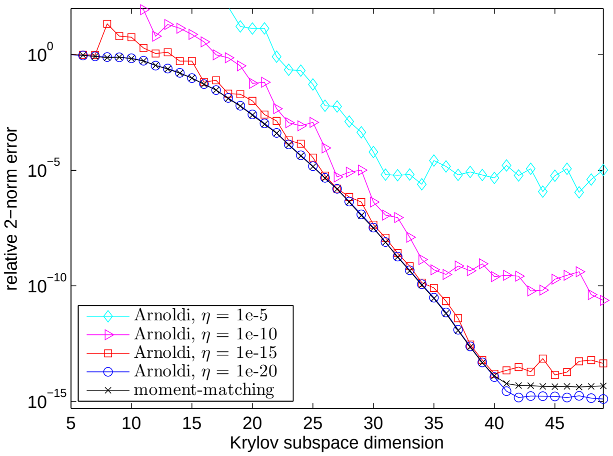

FIG. 7.2. *Relative approximation errors for the Arnoldi iteration with different scalings η and for the moment-matching method for the example of subsection 7.2*

7.3. Diffusion-reaction equation and a 4th order exponential Runge–Kutta method. In this example, we consider a finite difference spatial discretization of the semilinear problem

$$\partial_t u = \partial_{xx} u + \gamma u(1-u), \quad x \in \big[-2, 2\big], \tag{7.1}$$

subject to homogeneous Dirichlet boundary conditions, and with initial value

$$u_0(x) = \sin\left(\tfrac{\pi}{4}x\right) - 0.5x.$$

The number of spatial discretization points is set to 800, resulting in a system of stiff ODEs

$$u'(t) = Au(t) + g(u(t)), \quad u(t_0) = u_0,$$

where A denotes the discretized second derivative. We perform one time step using a $4th$ order exponential integrator given by the Butcher tableau (see [21, Sec. 2.3])

$$\begin{array}{c|ccccc}
0 & & & & & \\
\frac{1}{2} & \frac{1}{2}\varphi_{1,2} & & & & \\
\frac{1}{2} & \frac{1}{2}\varphi_{1,3} - \varphi_{2,3} & \varphi_{2,3} & & & \\
1 & \varphi_{1,4} - 2\varphi_{2,4} & \varphi_{2,4} & \varphi_{2,4} & & \\
\frac{1}{2} & \frac{1}{2}\varphi_{1,5} - a_{5,2} - \frac{1}{4}\varphi_{2,5} & a_{5,2} & a_{5,2} & \frac{1}{4}\varphi_{2,5} - a_{5,2} & \\
\hline
 & \varphi_1 - 3\varphi_2 + 4\varphi_3 & 0 & 0 & -\varphi_2 + 4\varphi_3 & 4\varphi_2 - 8\varphi_3
\end{array}$$

where

$$a_{5,2} = \frac{1}{2}\varphi_{2,5} - \varphi_{3,4} + \frac{1}{4}\varphi_{2,4} - \frac{1}{2}\varphi_{3,5}.$$

We compare the computation of the final stage of the integrator

$$u_1 = \mathrm{e}^{hA}u_0 + h\sum_{i=1}^{5} b_i(hA)G_{0,i}, \tag{7.2}$$

using the moment-matching iteration and the Arnoldi iteration for the augmented matrix $\widetilde{A}$ (with and without scaling). The vectors $G_{0,i}$ coming from the stages of the method are computed to machine precision. To get (7.2) into a suitable form for these iterations, we rewrite it as (3.12).

The time step is set to $h = 2.0 \cdot 10^{-3}$, and we compare the algorithms for the cases $\gamma = 200$ and 1000. The scaling (3.11) of the augmented matrix $\widetilde{A}$ is done using $S = \eta^{-1}$, where $\eta = \|W\|_2$, which was found to be a good choice in the numerical experiment of subsection 7.2. In this case $\|hA\|_2 \approx 320$, and $\|hg'(u_0)\| \approx 0.4$ for $\gamma = 200$ while $\|hg'(u_0)\| \approx 2$ for $\gamma = 1000$. For comparison we compute in addition the Arnoldi approximation of the product $\mathrm{e}^{hA}u_0$. In double precision arithmetic the algorithm stagnated at a Krylov subspace size of around 30.

From Figure 7.3 we infer that the moment-matching approximation gives in both cases a smaller error for a given subspace size than the Arnoldi approximation of the augmented matrix $\widetilde{A}$. We also note that the augmented Krylov subspace method, as explained in Section 4, converged to a relative 2-norm error of 10^{-13} and was considerably faster than applying the Arnoldi iteration to the augmented matrix $\widetilde{A}$.

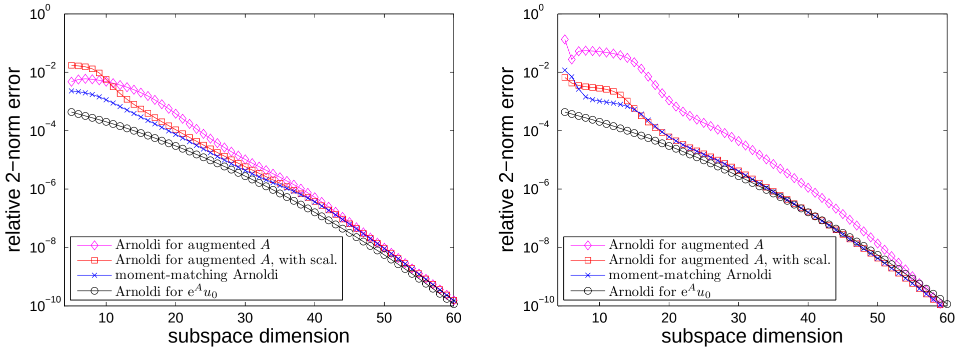

FIG. 7.3. *Convergence of different methods for the approximation of the final stage* (7.2) *with* $\gamma = 200$ *(left) and* $\gamma = 1000$ *(right). The Arnoldi approximation of the augmented matrix is performed with and without scaling.*

Acknowledgments. The authors thank Marlis Hochbruck, Daniel Kressner and Volker Mehrmann for several valuable comments.

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 (2010), pp. 488–511.

[2] Z. Bai, *Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems*, Appl. Numer. Math., 43 (2002), pp. 9–44.

[3] B. Beckermann and L. Reichel, *Error estimates and evaluation of matrix functions via the Faber transform*, SIAM J. Numer. Anal., 47 (2009), pp. 3849–3883.

[4] M.A. Botchev, V. Grimm, and M. Hochbruck, *Residual, restarting, and Richardson iteration for the matrix exponential*, SIAM J. Sci. Comput., 35 (2013), pp. A1376–A1397.

[5] M. Crouzeix, *Numerical range and functional calculus in Hilbert space*, J. Funct. Anal., 244 (2007), pp. 668–690.

[6] V.L. Druskin and L.A. Knizhnerman, *Two polynomial methods of calculating functions of symmetric matrices*, USSR Comput. Math. Math. Phys., 29 (1989), pp. 112–121.

[7] V.L. Druskin and L.A. Knizhnerman, *Extended Krylov subspaces: approximation of the matrix square root and related functions*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 755–771.

[8] V.L. Druskin and V. Simoncini, *A new investigation of the extended Krylov subspace method for matrix function evaluations*, Numer. Linear Algebra Appl., 17 (2010), pp. 615–638.

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

[10] P. Feldmann and R.W. Freund, *Efficient linear circuit analysis by Padé approximation via the Lanczos process*, IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., 14 (1995), pp. 639–649.

[11] A. Frommer and V. Simoncini, *Matrix functions*, in: Model order reduction: theory, research aspects and applications (W.H.A. Schilders, H.A. van den Vorst, J. Rommers, eds.), Springer, New York, 2008, pp. 275–303.

[12] E. Gallopoulos and Y. Saad, *Efficient solution of parabolic equations by Krylov approximation methods*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 1236–1264.

[13] A. Gaul, M.H. Gutknecht, J. Liesen, and R. Nabben, *A framework for deflated and augmented Krylov subspace methods*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 495–518.

[14] G. Golub and C. Van Loan, *Matrix computations*, 3rd edition, The Johns Hopkins University Press, Baltimore, 2012.

[15] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*, German edition, Verlag Harry Deutsch, Thun, 1981.

[16] M.H. Gutknecht, *Block Krylov space methods for linear systems with multiple right-hand sides: an introduction*, Modern Mathematical Models, Methods and Algorithms for Real World Systems (A.H. Siddiqi, I.S. Duff, and O. Christensen, eds.), Anamaya Publishers, New Delhi, 2005, pp. 420–447.

[17] S. Güttel, *Rational Krylov approximation of matrix functions: numerical methods and optimal pole selection*, GAMM Mitteilungen, 36 (2013), pp. 8–31.

[18] N.J. Higham, *Functions of matrices: theory and computation.* SIAM, Philadelphia, 2008.

[19] N.J. Higham, *The scaling and squaring method for the matrix exponential revisited.* SIAM J. Matrix Anal. Appl., 26 (2005), pp. 1179-1193.

[20] M. Hochbruck and C. Lubich, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.

[21] M. Hochbruck and A. Ostermann, *Exponential integrators*, Acta Numer. 19 (2010), pp. 209–286.

[22] R.A. Horn and C.R. Johnson, *Topics in matrix analysis*, Cambridge University Press, 1991.

[23] L. Knizhnerman, *Calculation of functions of unsymmetric matrices using Arnoldi's method*, Comput. Math. Math. Phys., 31 (1992), pp. 1–9.

[24] A. Koskela and A. Ostermann, *Exponential Taylor methods: analysis and implementation*, Comput. Math. Appl., 65 (2013), pp. 487-499.

[25] L. Lopez and V. Simoncini, *Preserving geometric properties of the exponential matrix by block Krylov subspace methods*, BIT, 46 (2006), pp. 813–830.

[26] I. Moret and P. Novati, *RD-rational approximations of the matrix exponential*, BIT 44 (2004), pp. 595–615.

[27] Y. Saad, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 29 (1992), pp. 209–228.

[28] Y. Saad, *Analysis of augmented Krylov subspace methods*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 435–449.

[29] Y. Saad, *Iterative methods for sparse linear systems*, SIAM, 2003.

[30] T. Schmelzer and L.N. Trefethen, *Evaluating matrix functions for exponential integrators via Carathéodory–Fejér approximation and contour integrals*, Electron. Trans. Numer. Anal., 29 (2007), pp. 1–18.

[31] M.N. Spijker, *Numerical ranges and stability estimates*, Appl. Numer. Math., 13 (1993), pp. 241–249.