ASC Report No. 23/2018

Computable strict upper bounds for Krylov approximations to a class of matrix exponentials and ϕ-functions

T. Jawecki, W. Auzinger, O. Koch

Institute for Analysis and Scientific Computing
Vienna University of Technology — TU Wien
www.asc.tuwien.ac.at ISBN 978-3-902627-00-1

Most recent ASC Reports

22/2018 *A. Jüngel, and O. Leingang*
Convergence of an implicit Euler Galerkin scheme for Poisson-Maxwell-Stefan systems

21/2018 *J. Kraus, C.-M. Pfeiler, D. Praetorius, M. Ruggeri, and B. Stiftner*
Iterative solution and preconditioning for the tangent plane scheme in computational micromagnetics

20/2018 *T. Führer, G. Gantner, D. Praetorius, and S. Schimanko*
Optimal additive Schwarz preconditioning for adaptive 2D IGA boundary element methods

19/2018 *A. Arnold, C. Klein, and B. Ujvari*
WKB-method for the 1D Schrödinger equation in the semi-classical limit: enhanced phase treatment

18/2018 *A. Bespalov, T. Betcke, A. Haberl, and D. Praetorius*
Adaptive BEM with optimal convergence rates for the Helmholtz equation

17/2018 *C. Erath and D. Praetorius*
Optimal adaptivity for the SUPG finite element method

16/2018 *M. Fallahpour, S. McKee, and E.B. Weinmüller*
Numerical simulation of flow in smectic liquid crystals

15/2018 *A. Bespalov, D. Praetorius, L. Rocchi, and M. Ruggeri*
Goal-oriented error estimation and adaptivity for elliptic PDEs with parametric or uncertain inputs

14/2018 *J. Burkotova, I. Rachunkova, S. Stanek, E.B. Weinmüller, S. Wurm*
On nonsingular BVPs with nonsmooth data. Part 1: Analytical results

13/2018 *J. Gambi, M.L. Garcia del Pino, J. Mosser, and E.B. Weinmüller*
Post-Newtonian equations for free-space laser communications between space-based systems

Institute for Analysis and Scientific Computing
Vienna University of Technology
Wiedner Hauptstraße 8–10
1040 Wien, Austria

E-Mail: admin@asc.tuwien.ac.at
WWW: http://www.asc.tuwien.ac.at
FAX: +43-1-58801-10196

ISBN 978-3-902627-00-1

Computable strict upper bounds for Krylov approximations to a class of matrix exponentials and ϕ-functions

Tobias Jawecki · **Winfried Auzinger** · **Othmar Koch**

September 10, 2018

Abstract An a posteriori bound for the error of a standard Krylov approximation to the exponential of nonexpansive matrices is derived. This applies for instance to skew-Hermitian matrices which appear in the numerical treatment of Schrödinger equations. It is proven that this defect-based bound is strict in contrast to existing approximations of the error, and it can be computed economically in the underlying Krylov space. In view of time-stepping applications, assuming that the given matrix is scaled by a time step, it is shown that the bound is asymptotically correct (with an order related to the dimension of the Krylov space) for the time step tending to zero. Furthermore, this result is extended to Krylov approximations of ϕ-functions and to improved versions of such approximations. The accuracy of the derived bounds is demonstrated by examples and compared with different variants known from the literature, which are also investigated more closely. Also other error bounds are tested on examples, in particular a version based on the concept of effective order. For the case where the matrix exponential is used in time integration algorithms, a step size selection strategy is proposed and illustrated by experiments.

Keywords matrix exponential · Krylov approximation · a posteriori error estimation · strict upper bound · Schrödinger equation

Mathematics Subject Classification (2010) 15A16, 65F15, 65F60

1 Introduction

We consider Krylov approximations to the matrix exponential function for the purpose of the solution of a linear, homogeneous system of differential equations

$$\psi'(t) = M\psi(t), \quad \psi(0) = \psi_0, \qquad \psi(t) = \mathrm{e}^{tM}\psi_0. \tag{1.1}$$

Our main focus is on equations of Schrödinger type, where $M = -\mathrm{i}H$ with an Hermitian matrix H, which commonly arise after spatial discretization of linear Schrödinger equations.

In this work we present new results for precise a posteriori error estimation, which also extend to the evaluation of so-called ϕ-functions. The application of these estimates for the purpose of time propagation is also discussed and illustrated.

Tobias Jawecki · Winfried Auzinger
Institut für Analysis und Scientific Computing, Technische Universität Wien
Wiedner Hauptstrasse 8–10/E101, A-1040 Wien, Austria
E-mail: tobias.jawecki@tuwien.ac.at, w.auzinger@tuwien.ac.at

Othmar Koch
Fakultät für Mathematik, Universität Wien
Oskar-Morgenstern-Platz 1, A-1090 Wien, Austria
E-mail: othmar@othmar-koch.org

Overview on existing approaches and results. The approximate evaluation of large matrix exponential functions is a topic which has been extensively treated in the numerical analysis literature, for basic reference see e.g. [12, 26]. A standard approach is to project the given matrix M to a low-dimensional Krylov space via Lanczos iteraton, and to directly exponentiate the projected small matrix. A first mention of the Lanczos approach can be found in [29], where it is also recognized that for the method to perform satisfactorily, the time-steps have to be controlled. However, the control mechanism from [29] is not very elaborate and is based on a series expansion of the error, which is only valid in the asymptotic regime, see for instance [28]. For discretizations of parabolic problems, [11] uses an error estimator to choose the step-size, this approach is improved in [34] and has been generalized in [25]. Notably, in the latter reference a strict error bound is used to estimate the time-step instead of asymptotic techniques. It is furthermore observed that in contrast, strategies based on Taylor expansion tend to overestimate the error. Hence the strategy from [25] performs better than [24] and better in turn than [29].

A first systematic study of Krylov-based methods for the matrix exponential function was given in [31]. The error is analyzed theoretically, yielding both a priori and computable a posteriori estimates. The analysis there relies on approximation theory and yields a priori error bounds which are asymptotically optimal in the dimension of the Krylov subspace in important situations. The analysis moreover implies correction schemes to lift the convergence order which are cheap to compute based on the already available information. The error expansion also suggests a posteriori error estimators resorting to the leading error term. This approach relies on the assumption of the sufficiently rapid decay of the asymptotic series. A recent generalization of this work together with a more rigorous justification is given in [21].

A thorough theoretical analysis of the error of Krylov methods for the exponential of an Hermitian or skew- (anti-) Hermitian matrix was given in [16]. The analysis derives an asymptotic error expansion and shows superlinear error decay in the dimension m of the approximation subspace for sufficiently large m. In [16], a posteriori error estimation is also discussed. This topic is furthermore addressed in [22]. There, the Krylov approximation method is interpreted as a Galerkin method, whence an error bound can be obtained from an error representation for this variational approximation. This yields a computable estimate via a quadrature approximation of the error integral involving the defect of the numerical approximation. The a priori error analysis reveals a step-size restriction for the convergence of the method, which is less stringent when the subspace dimension is larger.

Further work in the direction of controlling the Lanczos process through information gained from the defect is given in [5]. The defect is a scalar multiple of the successive Krylov vector arising in the iteration and can be evaluated efficiently. If the error is approximated by a Galerkin approach, the resulting estimator corresponds to the difference of two Lanczos iterates. For the purpose of practical error estimation, in [5] it is seen as preferable to continue the original Krylov process. Some other defect-based upper bounds for the error of the matrix exponential are given in [21], but they still require some information on the matrix spectrum. [21] also shows that the simple error estimates of [31] are reliable up to a constant factor, depending on the spectrum of the matrix.

Various other improved methods for computing the matrix exponential function are given in the literature, for example restarted methods, deflated restarting methods or quadrature based restarting methods, see [1], [7], and [10].

It has also been advocated in [8] to use a preconditioning in the Lanczos method by a shifted inverse in order to get a good approximation of the leading invariant subspaces, where the choice of the shift is critical for the success of this procedure. This strategy amounts to a transformation of the spectrum which grants a convergence speed which is independent of the norm of the given matrix. In [8], a posteriori error estimation based on the asymptotical expansion of the error is advocated as well. We note that our results do not immediately carry over to the shift-and-invert approach, see Remark 2.

Overview on present work. In Section 2 we introduce the Krylov approximation and the integral representation of the approximation error in terms of its defect. In Section 3 we derive a new computable upper bound for the error, i.e., we construct a strict upper limit by using data available from the Krylov process with negligible additional computational effort (Theorem 1). This upper bound is cheap to evaluate and update on the fly during the Lanczos iteration. It is also asymptotically correct, i.e., for

$t \to 0$ the error of the error estimator tends to zero faster asymptotically than the error itself. In Section 4 these results are extended to the case where the Krylov approach is employed to approximate the ϕ-functions of matrices (generalizing the exponential function), see Theorem 2. In Section 5, improved approximations derived from an extended (respectively *corrected*) Krylov process [31] are discussed, and corresponding error estimators are analyzed, including an asymptotically correct strict upper bound on the error (Theorem 3). This approach can be used to increase the order, but it has the drawback of violating mass conservation. In Proposition 6 error estimates are particularized to the Hermitian case. Another view on defect-based error estimation is presented in Section 6.

Section 7 is devoted to practical application of the various error estimators for the control of the time steps t including smaller substeps Δt if it appears indicated. In Section 8 we present numerical results for a finite difference discretization of the free Schrödinger equation, the heat equation, and a Hubbard model of solar cells, illustrating our theoretical results. Additional practical aspects are also investigated: A priori estimates and the role of restarting are discussed in particular in the context of practical step-size adaptation. Finally, we demonstrate the computational efficiency of our adaptive strategy.

2 Problem setting, Krylov approximation, and defect-based representation of approximation error

We discuss the approximation of the exponential of a nonexpansive matrix σA,

$$\begin{aligned} &E(t)v = \mathrm{e}^{\sigma t A}v, \quad A \in \mathbb{C}^{n\times n}, \ \sigma \in \mathbb{C}, \ \text{with} \\ &|\sigma| = 1 \ \text{ and numerical range } W(\sigma A) \subseteq \mathbb{C}_-, \end{aligned} \tag{2.1}$$

with step size t, applied to an initial vector $v \in \mathbb{C}^n$. To simplify the notation we assume $\|v\|_2 = 1$ without loss of generality. The exponential $E(t) = \mathrm{e}^{\sigma t A}$ satisfies the matrix differential equation

$$E'(t) = \sigma A E(t), \quad E(0) = I.$$

By assumption on A we have[1] $\|E(t)\|_2 \le 1$.

Krylov subspaces and associated identities. The numerical approximation of (2.1) considered here (see (2.7) below) is based on the conventional Krylov subspace

$$\mathscr{K}_m(A,v) = \operatorname{span}\{v, Av, \ldots, A^{m-1}v\} \subseteq \mathbb{C}^n.$$

First, a unitary basis of $\mathscr{K}_m(A,v)$ is obtained by the well-known Arnoldi iteration, see [32]. This produces a basis matrix $V_m \in \mathbb{C}^{n\times m}$ satisfying $V_m^* V_m = I_{m\times m}$, and an upper Hessenberg matrix $T_m \in \mathbb{C}^{m\times m}$ such that the Krylov identity[2]

$$A V_m = V_m T_m + \tau_{m+1,m}\, v_{m+1} e_m^* \tag{2.2}$$

is valid, with $\tau_{m+1,m} \in \mathbb{R}_+$ and $v_{m+1} \in \mathbb{C}^n$ with $\|v_{m+1}\|_2 = 1$.

Remark 1 We are assuming that the Arnoldi iteration is executed until the desired dimension m. Then, by construction, all lower diagonal entries of T_m are positive [32]. If this is not the case, i.e., if a breakdown occurs, it is known that this breakdown is *lucky,*, i.e., the approximation (2.7) below obtained in the step before breakdown is already exact, see [31].

The parameter σ is introduced to benefit from cheaper construction of the Krylov subspace for the case of an Hermitian matrix A. The skew-Hermitian case[3] with $\sigma = -\mathrm{i}$ is the most prominent one,

$$E(t)v = \mathrm{e}^{-\mathrm{i}tH}v, \quad H \in \mathbb{C}^{n\times n} \ \text{ Hermitian.} \tag{2.3}$$

[1] The assumption $W(A) \subseteq \mathbb{C}_-$ implies no loss of generality if a priori information is available to shift the spectrum of σA: $\mathrm{e}^{\sigma t A} = \mathrm{e}^{\sigma t \alpha}\mathrm{e}^{\sigma t(A-\alpha I)}$ for $0 < (\sigma\alpha) \in \mathbb{R}$.

[2] Here, $e_m = (0,\ldots,0,1)^* \in \mathbb{C}^m$, and in the sequel we also denote $e_1 = (1,0,\ldots,0)^* \in \mathbb{C}^m$.

[3] In this case the matrix A is usually named H, typically for equations of Schrödinger type.

In this case, $E(t)$ represents a unitary evolution, i.e., $E^{-1}(t) = E^*(t)$.

By the separation of σ and a symmetric matrix A the Krylov subspace can be constructed using the Lanczos iteration which is a special case of the Arnoldi iteration, resulting in a tridiagonal matrix $T_m \in \mathbb{R}^{m\times m}$. In the following we discuss the general case and comment on the case of an Hermitian matrix A whenever appropriate.

The following identities hold true due to the upper Hessenberg [tridiagonal] structure of T_m together with (2.2):

$$e_m^* T_m^j e_1 = 0 \quad \text{for } \; j = 0,\dots,m-2, \tag{2.4}$$

and

$$A^{m-1} v = V_m T_m^{m-1} e_1. \tag{2.5}$$

Furthermore, let

$$\gamma_m = e_m^* T_m^{m-1} e_1 = \prod_{j=1}^{m-1} (T_m)_{j+1,j}\,, \tag{2.6}$$

where the claimed identity also follows from the upper Hessenberg [tridiagonal] structure of T_m.

Krylov approximation. The standard Krylov approximation to $E(t)v$ is

$$S_m(t)v = V_m \mathrm{e}^{\sigma t T_m} V_m^* v = V_m \mathrm{e}^{\sigma t T_m} e_1. \tag{2.7}$$

We denote the corresponding error operator by $L_m(t)$, with

$$L_m(t) = E(t) - S_m(t) \in \mathbb{C}^{n\times n}. \tag{2.8}$$

Defect-based integral representation of the approximation error. We define the *defect* operator $D_m(t)$ of $S_m(t)$ by

$$D_m(t) = \sigma A S_m(t) - S_m'(t) \in \mathbb{C}^{n\times n}. \tag{2.9}$$

Then, $L_m(t)v$ and $D_m(t)v$ are related via the differential equation

$$L_m'(t)v = \sigma A L_m(t)v + D_m(t)v, \quad L_m(0)v = 0, \tag{2.10a}$$

whence

$$L_m(t)v = \int_0^t E(t-s)\, D_m(s)v \,\mathrm{d}s. \tag{2.10b}$$

An explicit representation for $D_m(s)v$ is obtained from (2.2),

$$\begin{aligned} D_m(s)v &= \sigma A V_m \mathrm{e}^{\sigma s T_m} e_1 - \sigma V_m T_m \mathrm{e}^{\sigma s T_m} e_1 = \sigma \left(A V_m - V_m T_m\right) \mathrm{e}^{\sigma s T_m} e_1 \\ &= \sigma \tau_{m+1,m} \left(e_m^* \mathrm{e}^{\sigma s T_m} e_1\right) v_{m+1}. \end{aligned} \tag{2.11}$$

Asymptotically for $t \to 0$,

$$D_m(t)v = \sigma \tau_{m+1,m} \gamma_m \frac{(\sigma t)^{m-1}}{(m-1)!} v_{m+1} + \mathscr{O}(t^m), \tag{2.12}$$

which follows from the Taylor series representation for $\mathrm{e}^{\sigma t T_m}$ together with (2.4) and (2.6). Thus,

$$\|D_m(t)v\| = \mathscr{O}(t^{m-1}), \quad \text{whence} \quad \|L_m(t)v\| = \mathscr{O}(t^m).$$

We can also characterize the asymptotically leading term of the error:

Proposition 1 *For any $A \in \mathbb{C}^{m\times m}$ the error $L_m(t)v$ satisfies the asymptotic relation*

$$L_m(t)v = \tau_{m+1,m}\gamma_m \frac{(\sigma t)^m}{m!} v_{m+1} + \mathscr{O}(t^{m+1}) \tag{2.13}$$

for $t \to 0$.

Proof Due to $L_m(t)v = \mathcal{O}(t^m)$ we have

$$L_m(t)v = E(t)v - S_m(t)v = \sum_{k=0}^{\infty} \frac{(\sigma tA)^k}{k!} v - V_m \sum_{k=0}^{\infty} \frac{(\sigma tT_m)^k}{k!} e_1 = \sum_{k=m}^{\infty} \frac{(\sigma t)^k (A^k v - V_m T_m^k e_1)}{k!}.$$

Furthermore, multiplication of the identity (2.5) by H and using (2.2) gives

$$\begin{aligned} A^m v &= A V_m T_m^{m-1} e_1 = (V_m T_m + \tau_{m+1,m} v_{m+1} e_m^*) T_m^{m-1} e_1 \\ &= V_m T_m^m e_1 + \tau_{m+1,m} v_{m+1} e_m^* T_m^{m-1} e_1 = V_m T_m^m e_1 + \tau_{m+1,m} \gamma_m v_{m+1}, \end{aligned}$$

which implies (2.13). □

3 An asymptotically correct upper error bound for the case (2.1)

For the case (2.1) we have $\|E(t-s)\|_2 \le 1$ for $0 \le s \le t$, and (2.10b) implies

$$\|L_m(t)v\|_2 = \left\| \int_0^t E(t-s)\, D_m(s)v \, \mathrm{d}s \right\|_2 \le \int_0^t \|D_m(s)v\|_2 \, \mathrm{d}s.$$

With $\|v_{m+1}\|_2 = 1$, and

$$\delta_m(s) = e_m^* \mathrm{e}^{\sigma s T_m} e_1 = \left(\mathrm{e}^{\sigma s T_m}\right)_{m,1}, \tag{3.1a}$$

together with (2.11) we obtain

$$\|L_m(t)v\|_2 \le \tau_{m+1,m} \int_0^t |\delta_m(s)| \, \mathrm{d}s. \tag{3.1b}$$

This estimate is also given in [22, Section III.2]. Of course, (3.1b) cannot be directly evaluated, because $D_m(s)v$ is only directly available for the actual step size used, i.e., for $s = t$ (or, for instance $s = t/2$ if two half-steps are performed). In [22] it is proposed to use numerical quadrature[4] to approximate the integral in (3.1b). Our aim here is to derive a strict, computable upper bound. We proceed in two steps.[5]

Analytic matrix function via interpolation on the spectrum. To approximate the error integral in (3.1b) we use the representation of matrix exponentials via Hermite interpolation of the scalar exponential function on the spectrum of the matrix T_m, see [15, Chap. 1]: If $\mu_1, \ldots, \mu_r$ ($r \le m$) denote the distinct eigenvalues of T_m and n_j is the dimension of the largest Jordan block associated with μ_j, then

$$\mathrm{e}^{\sigma t T_m} = p_t(T_m), \tag{3.2a}$$

where $p_t(\lambda)$ is the Hermite interpolant of degree $\le m-1$ of the function

$$f_t(\lambda) = \mathrm{e}^{\sigma t \lambda} \tag{3.2b}$$

over the nodes $\mu_1, \ldots, \mu_m$ in the sense of [15, (1.7)],

$$p_t^{(\ell)}(\mu_j) = f_t^{(\ell)}(\mu_j), \quad j = 1, \ldots, r, \quad \ell = 0, \ldots, n_j - 1.$$

For a general matrix, the degree of p_t may be smaller than $m-1$. However, in our context a special case occurs: Since the lower diagonal entries of T_m do not vanish, T_m is nonderogatory, i.e., for each eigenvalue μ_j the associated eigenspace is one-dimensional, see [18, Section 3.1]. Then, $\sum_{j=1}^r n_j = m$, which implies that the degree of p_t is exactly $m-1$.

In the following we denote the full sequence of the m eigenvalues of T_m by $\lambda_1, \ldots, \lambda_m$. By applying basic properties of the Krylov decomposition and imposed conditions on the numerical range of A we obtain

$$\mathrm{spec}(\sigma T_m) \subseteq W(\sigma T_m) \subseteq W(\sigma A) \subseteq \mathbb{C}_-.$$

The following proposition is partially related to [6, Chap. 3] or [35], respectively. Here, divided differences have to be understood in the general sense, i.e., in the confluent sense if multiple eigenvalues occur; for the detailed definition and properties see [15, Section B.16].

[4] See also Section 6 below.

[5] In the sequel, the argument of $\delta_m(\cdot)$ is again denoted by t instead of s.

Proposition 2 *Let $T_m \in \mathbb{C}^{m\times m}$ be an upper Hessenberg matrix with eigenvalues $\lambda_1,\dots,\lambda_m$ and spec$(\sigma T_m) \subseteq \mathbb{C}_-$. Then the function $\delta_m(t)$ defined as in* (3.1a)*, i.e.,*

$$\delta_m(t) = e_m^* \mathrm{e}^{\sigma t T_m} e_1 = \left(\mathrm{e}^{\sigma t T_m}\right)_{m,1},$$

satisfies

$$\delta_m(t) = f_t[\lambda_1,\dots,\lambda_m]\gamma_m \le \frac{t^{m-1}}{(m-1)!}\,\gamma_m, \tag{3.3}$$

with γ_m from (2.6) *and where $f_t[\lambda_1,\dots,\lambda_m]$ is the $(m-1)$-th divided difference over* spec(T_m) *of the function f_t defined in* (3.2b)*.*

Proof We proceed from the Newton representation of the interpolant $p_t(\lambda)$ from (3.2),

$$p_t(\lambda) = \sum_{j=0}^{m-1} f_t[\lambda_1,\dots,\lambda_{j+1}]\,\omega_j(\lambda),$$

with $\omega_j(\lambda) = (\lambda-\lambda_1)\cdots(\lambda-\lambda_j)$. From (2.4) and by definition of γ_m, it is obvious that the ω_j satisfy

$$e_m^*\,\omega_j(T_m)\,e_1 = \begin{cases} 0, & j = 0,\dots,m-2,\\ \gamma_m, & j = m-1.\end{cases}$$

Together with (3.2) this shows that the identity claimed in (3.3) is valid:

$$\delta_m(t) = e_m^*\mathrm{e}^{\sigma t T_m}\,e_1 = e_m^*\,p_t(T_m)e_1 = \sum_{j=0}^{m-1} f_t[\lambda_1,\dots,\lambda_{j+1}]e_m^*\,\omega_j(T_m)e_1 = f_t[\lambda_1,\dots,\lambda_m]\gamma_m.$$

According to [15, (B.28)] the divided difference can be estimated by

$$|f_t[\lambda_1,\dots,\lambda_m]| \le \frac{\max_{z\in\Omega} D^{(m-1)}f_t(z)}{(m-1)!}$$

for convex $\Omega \subseteq \mathbb{C}$ which contains all eigenvalues λ_j.

With $D^{(m-1)}f_t(\lambda) = (\sigma t)^{m-1}\mathrm{e}^{\sigma t\lambda}$, $|\sigma| = 1$ and $\mathrm{Re}(\sigma\lambda_j) \le 0$ we obtain

$$|f_t[\lambda_1,\dots,\lambda_m]| \le \frac{t^{m-1}}{(m-1)!},$$

which implies the estimate (3.3) for $\delta_m(t)$. □

Error estimate and asymptotical correctness. Now we apply Proposition 2 in the context of our Krylov approximation.

Theorem 1 (Computable upper bound)
The error $L_m(t)v$ of the Krylov approximation (2.7) *to $E(t)v$ satisfies*

$$\|L_m(t)v\|_2 \le \tau_{m+1,m}\gamma_m\,\frac{t^m}{m!} \tag{3.4}$$

with $\tau_{m+1,m}$ from (2.2) *and γ_m from* (2.6)*.*

Proof We proceed from (3.1). For δ_m defined in (3.1a), Proposition 2 implies

$$|\delta_m(s)| \le \frac{s^{m-1}}{(m-1)!}\gamma_m,$$

and this gives an upper bound for the error integral (3.1b):

$$\|L_m(t)v\|_2 \le \tau_{m+1,m}\gamma_m\int_0^t \frac{s^{m-1}}{(m-1)!}\,\mathrm{d}s = \tau_{m+1,m}\gamma_m\frac{t^m}{m!},$$

which completes the proof. □

The upper bound (3.4) corresponds to the 2-norm of the leading error term (2.13) according to Proposition 1. It is easily computable from the Lanczos decomposition (2.2). We denote the error estimate given by (3.4) as

$$\mathrm{Err}_a = \tau_{m+1,m}\gamma_m \frac{t^m}{m!}. \tag{Err$_a$}$$

We will refer to Err_a as the asymptotic error estimate.

Proposition 3 (Asymptotical correctness)
The upper bound (3.4) *is asymptotically correct for* $t \to 0$*, i.e.,*

$$\|L_m(t)v\|_2 = \tau_{m+1,m}\gamma_m \frac{t^m}{m!} + \mathscr{O}(t^{m+1}). \tag{3.5}$$

Proof The asymptotic estimate

$$\begin{aligned}\left| \|L_m(t)v\|_2 - \tau_{m+1,m}\gamma_m \frac{t^m}{m!} \right| &= \left| \|L_m(t)v\|_2 - \left\| \tau_{m+1,m}\gamma_m \frac{(\sigma t)^m}{m!} v_{m+1} \right\|_2 \right| \\ &\le \left\| L_m(t)v - \tau_{m+1,m}\gamma_m \frac{(\sigma t)^m}{m!} v_{m+1} \right\|_2 = \mathscr{O}(t^{m+1})\end{aligned}$$

is valid due to Proposition 1, and this proves (3.5). □

Remark 2 In [8, Section 4] a defect-based error formulation is given for the shift-and-invert Krylov approximation of the matrix exponential function. In contrast to the standard Krylov method, the defect is not of order $m-1$ for $t \to 0$ there. Hence, our new results do not directly apply to shift-and-invert Krylov approximations. A study of a posteriori error estimates for the shift-and-invert approach is a topic of future investigations.

4 Krylov approximation to ϕ-functions.

As another application we consider the so-called ϕ-functions, with power series representation

$$\phi_p(z) = \sum_{k=0}^{\infty} \frac{z^k}{(k+p)!}, \quad p \ge 0. \tag{4.1a}$$

We have $\phi_0(z) = \mathrm{e}^z$, and

$$\phi_p(z) = \frac{1}{(p-1)!} \int_0^1 (1-\theta)^{p-1} \mathrm{e}^{\theta z} \mathrm{d}\theta, \quad p \ge 1. \tag{4.1b}$$

As the matrix exponential, ϕ-functions of matrices also appear in a wide range of applications, such as exponential integrators, see for instance [2, 17, 28, 33]. Krylov approximation is a common technique to evaluate ϕ-functions of matrices applied to a starting vector,

$$\phi_p(\sigma t A)v \approx V_m \phi_p(\sigma t T_m)e_1, \quad p \ge 0. \tag{4.2}$$

Since ϕ-functions are closely related to the matrix exponential, our ideas can be applied to these as well. We use the following notation for the error of the ϕ-functions:

$$L_m^p(t)v = \phi_p(\sigma t A)v - V_m \phi_p(\sigma t T_m)e_1. \tag{4.3}$$

With (4.3) we generalize the previously used notation: $L_m(t) = L_m^0(t)$.

Theorem 2 *The error of the Krylov approximation* (4.2) *to* $\phi_p(\sigma tA)v$ *with* $p \ge 0$ *satisfies*

$$L_m^p(t)v = \tau_{m+1,m}\,\gamma_m \frac{(\sigma t)^m}{(m+p)!}\, v_{m+1} + \mathscr{O}(t^{m+1}). \tag{4.4a}$$

Furthermore, its norm is strictly bounded by

$$\|L_m^p(t)v\|_2 \le \tau_{m+1,m}\,\gamma_m \frac{t^m}{(m+p)!}, \tag{4.4b}$$

and this bound is asymptotically correct for $t \to 0$*.*

Proof For $p = 0$ the result directly follows from Propositions 1, 3 and Theorem 1. We now assume $p \geq 1$. Via the series representation (4.1a) of ϕ_p we can determine the leading term of the error in an analogous way as in Proposition 1:

$$\begin{aligned}\phi_p(\sigma tA)v - V_m\phi_p(\sigma tT_m)e_1 &= \sum_{k=m}^{\infty} \frac{(\sigma t)^k\,(A^k v - V_m T_m^k e_1)}{(k+p)!} \\ &= \tau_{m+1,m}\,\gamma_m \frac{(\sigma t)^m}{(m+p)!}\,v_{m+1} + \mathscr{O}(t^{m+1}),\end{aligned}$$

which proves (4.4a).

Furthermore, proceeding from (4.1b) we obtain

$$\begin{aligned}\phi_p(\sigma tA)v - V_m\phi_p(\sigma tT_m)e_1 &= \frac{1}{(p-1)!}\int_0^1 (1-\theta)^{p-1}\left(\mathrm{e}^{\sigma\theta tA}v - V_m\,\mathrm{e}^{\sigma\theta tT_m}e_1\right)\mathrm{d}\theta \\ &= \frac{1}{(p-1)!}\int_0^1 (1-\theta)^{p-1} L_m(\theta t)v\,\mathrm{d}\theta,\end{aligned}$$

with the error $L_m(t)v$ for the matrix exponential case. Now we apply Theorem 1 to obtain

$$\begin{aligned}\|\phi_p(\sigma tA)v - V_m\phi_p(\sigma tT_m)e_1\|_2 &\leq \frac{1}{(p-1)!}\int_0^1 (1-\theta)^{p-1}\|L_m(\theta t)v\|_2\,\mathrm{d}\theta \\ &\leq \tau_{m+1,m}\,\gamma_m\,\frac{t^m}{(p-1)!\,m!}\int_0^1 (1-\theta)^{p-1}\theta^m\,\mathrm{d}\theta \\ &= \tau_{m+1,m}\,\gamma_m\,\frac{t^m}{(m+p)!},\end{aligned}$$

which proves (4.4b). □

5 Corrected Krylov approximation for the exponential and ϕ-functions.

Let us recall the well-known error representation given in [31].

Proposition 4 see [31, Theorem 5.1] *With the ϕ-functions defined in* (4.1)*, the error* (2.8) *can be represented in the form*

$$L_m(t)v = \tau_{m+1,m}\,\sigma t \sum_{j=1}^{\infty} e_m^*\phi_j(\sigma tT_m)\,e_1(\sigma tA)^{j-1}v_{m+1}. \tag{5.1}$$

In [31] it is stated that, typically, the first term of the sum given in Proposition 4, formula (Err_1), is already a good approximation to $L_m(t)v$. Analogously to [31, Section 5.2] we use the notation Err_1 for the norm of this term,

$$\mathrm{Err}_1 = \tau_{m+1,m}\,t\,|e_m^*\phi_1(\sigma tT_m)e_1|. \tag{Err_1}$$

In [21] it is even shown that Err_1 is an upper bound up to a factor depending on spectral properties of the matrix A. For the case of Hermitian σA we show $\|L_m(t)v\|_2 \leq \mathrm{Err}_1$ in Proposition 6 below.

In Remark 3 below we show that Err_1 is also an asymptotically correct approximation for the error norm (in the sense of Proposition 3). Furthermore, the error estimate Err_1 is computable at nearly no extra cost, see [31, Proposition 2.1].

According to [31, Proposition 2.1], $\phi_1(\sigma tT_m)$ can be computed from the extended matrix

$$T_m^+ = \begin{bmatrix} T_m & 0 \\ \tau_{m+1,m}\,e_m^* & 0 \end{bmatrix} \in \mathbb{C}^{(m+1)\times(m+1)} \tag{5.2a}$$

as

$$\mathrm{e}^{\sigma tT_m^+}e_1 = \begin{bmatrix} \mathrm{e}^{\sigma tT_m}e_1 \\ \tau_{m+1,m}\,\sigma t\,(e_m^*\,\phi_1(\sigma tT_m)\,e_1) \end{bmatrix} \in \mathbb{C}^{m+1}. \tag{5.2b}$$

Equation (5.2b) can be used to evaluate the error estimate Err_1 or a corrected Krylov approximation in the form

$$S_m^+(t)v = V_m^+ \,\mathrm{e}^{\sigma t T_m^+} e_1 \quad \text{with} \quad V_m^+ = \left[V_m \,\middle|\, v_{m+1} \right] \in \mathbb{C}^{n\times(m+1)}, \tag{5.3}$$

for which the first term of the error expansion according to Proposition 4 vanishes, see [31]. For the error of the corrected Krylov approximation we use the notation

$$L_m^+(t)v = E(t)v - S_m^+(t)v.$$

For general ϕ-functions we obtain an error representation similar to Proposition 4 and a corrected Krylov approximation for ϕ-functions. The corrected Krylov approximation to $\phi_p(\sigma t A)v$ is given in [33, Theorem 2]:

$$\phi_p(\sigma t A)v \approx V_m^+ \phi_p(\sigma t\, T_m^+)e_1$$

with T_m^+ and V_m^+ given in (5.2a) and (5.3). The error of the corrected Krylov approximation is denoted by

$$L_m^{p,+}(t)v = \phi_p(\sigma t A)v - V_m^+ \phi_p(\sigma t\, T_m^+)e_1. \tag{5.4}$$

Proposition 5 (see [33, Theorem 2]) *The error of the Krylov approximation $L_m^p(t)v$, see* (4.3)*, satisfies*

$$L_m^p(t)v = \tau_{m+1,m}\sigma t \sum_{j=p+1}^{\infty} \left(e_m^* \phi_j(\sigma t\, T_m)e_1\right) (\sigma t A)^{j-p-1} v_{m+1}. \tag{5.5a}$$

The error of the corrected Krylov approximation $L_m^{p,+}(t)v$, see (5.4)*, is given by*

$$L_m^{p,+}(t)v = \tau_{m+1,m}\sigma t \sum_{j=p+2}^{\infty} \left(e_m^* \phi_j(\sigma t\, T_m)e_1\right) (\sigma t A)^{j-p-1} v_{m+1}. \tag{5.5b}$$

The following remark will be used later on.

Remark 3 From the representation (4.1a) for the ϕ_j together with (2.4) and (2.6) we observe

$$\begin{aligned} e_m^* \phi_j(\sigma t T_m)e_1 &= \sum_{k=m-1}^{\infty} \frac{(\sigma t)^k e_m^* T_m^k e_1}{(k+j)!} = \frac{(\sigma t)^{m-1} e_m^* T_m^{m-1} e_1}{(m-1+j)!} + \mathscr{O}(t^m) \\ &= \gamma_m \frac{(\sigma t)^{m-1}}{(m-1+j)!} + \mathscr{O}(t^m). \end{aligned} \tag{5.6}$$

By (5.6) we observe $e_m^* \phi_j(\sigma t T_m)\, e_1 = \mathscr{O}(t^{m-1})$ for $j \geq 0$ and we conclude that the asymptotically leading order term of $L_m^p(t)v$ for $t \to 0$ is obtained by the leading term ($j = p+1$) of the series (5.5a):

$$L_m^p(t)v = \tau_{m+1,m}\sigma t (e_m^* \phi_{p+1}(\sigma t\, T_m)e_1) v_{m+1} + \mathscr{O}(t^m). \tag{5.7a}$$

Analogously we obtain the asymptotically leading order term of $L_m^{p,+}(t)v$ for $t \to 0$ by the leading term ($j = p+2$) of the series (5.5b):

$$L_m^{p,+}(t)v = \tau_{m+1,m}(\sigma t)^2 (e_m^* \phi_{p+2}(\sigma t\, T_m)e_1) A v_{m+1} + \mathscr{O}(t^{m+1}). \tag{5.7b}$$

The asymptotically leading terms in (5.7a) and (5.7b) can be used as error estimators:

$$\|L_m^p(t)v\|_2 \approx \tau_{m+1,m}\, t |e_m^* \phi_{p+1}(\sigma t\, T_m)e_1| \tag{5.8a}$$

and

$$\|L_m^{p,+}(t)v\|_2 \approx \|A v_{m+1}\|_2 \tau_{m+1,m}\, t^2 |e_m^* \phi_{p+2}(\sigma t\, T_m)e_1|. \tag{5.8b}$$

The error estimators (5.8a) and (5.8b) are already suggested in [33, 28]. We will refer to them as Err_1 in the context of the ϕ-functions with standard and corrected Krylov approximation, generalizing the corresponding quantities for the exponential case $p = 0$.

We also obtain strict upper bounds for the matrix exponential ($p = 0$) and general ϕ-functions with $p \geq 1$.

Theorem 3 (Asymptotic error estimate for corrected Krylov approximation, ϕ-functions) *The error of the corrected Krylov approximation* (5.4) *to $\phi_p(\sigma tA)v$ with $p \geq 0$ satisfies*

$$L_m^{p,+}(t)v = \tau_{m+1,m}\,\gamma_m \frac{(\sigma t)^{m+1}}{(m+p+1)!} Av_{m+1} + \mathscr{O}(t^{m+2}). \tag{5.9a}$$

Furthermore, its norm is strictly bounded by

$$\|L_m^{p,+}(t)v\|_2 \leq \|Av_{m+1}\|_2\,\tau_{m+1,m}\,\gamma_m \frac{t^{m+1}}{(m+p+1)!}, \tag{5.9b}$$

and this bound is asymptotically correct for $t \to 0$.

Proof Applying (5.6) (with $j = p+2$) to (5.7b) shows (5.9a):

$$L_m^{p,+}(t)v = \tau_{m+1,m}\,\gamma_m \frac{(\sigma t)^{m+1}}{(m+p+1)!} Av_{m+1} + \mathscr{O}(t^{m+2}).$$

From Proposition 5 we observe

$$L_m^{p,+}(t)v = \sigma tA\, L_m^{p+1}(t)v.$$

Using the integral representation analogously as in the proof of Theorem 2 for $L_m^{p+1}(t)v$ and formula (2.10b) for $L_m(t)v$, we obtain

$$\begin{aligned} L_m^{p,+}(t)v = \sigma tA\, L_m^{p+1}(t)v &= \tau_{m+1,m}\sigma t \frac{1}{p!}\int_0^1 (1-\theta)^p A L_m(\theta t)v\,\mathrm{d}\theta \\ &= \tau_{m+1,m}\sigma t \frac{1}{p!}\int_0^1 (1-\theta)^p \int_0^{\theta t} \mathrm{e}^{\sigma(\theta t - s)A} Av_{m+1}\,\delta_m(s)\,\mathrm{d}s\,\mathrm{d}\theta. \end{aligned}$$

With norm inequalities and Proposition 2 we obtain

$$\begin{aligned} \|L_m^{p,+}(t)v\|_2 &\leq \tau_{m+1,m} t \|Av_{m+1}\|_2 \frac{1}{p!}\int_0^1 (1-\theta)^p \int_0^{\theta t} |\delta_m(s)|\,\mathrm{d}s\,\mathrm{d}\theta \\ &\leq \|Av_{m+1}\|_2 \tau_{m+1,m}\gamma_m t^{m+1} \frac{1}{p!\,m!}\int_0^1 (1-\theta)^p \theta^m\,\mathrm{d}s\,\mathrm{d}\theta \\ &= \|Av_{m+1}\|_2 \tau_{m+1,m}\gamma_m \frac{t^{m+1}}{(m+p+1)!}, \end{aligned}$$

which proves (5.9b). □

If the error estimate (5.9b) is to be evaluated, the effort of the computation of $\|Av_{m+1}\|_2$ is comparable to one additional step of the Arnoldi iteration.

As mentioned before, we also can show that for Hermitian σA the estimate Err_1 gives a strict upper bound:

Proposition 6 *For the case $\sigma = 1$ and an Hermitian, nonexpansive A we obtain*

$$|\delta_m(t)| = \delta_m(t) > 0 \quad \textit{for } t > 0.$$

This leads to the following strict upper bounds for the errors L_m^p and $L_m^{p,+}$ with $p \geq 0$:

$$\|L_m^p(t)v\|_2 \leq \tau_{m+1,m}\, t \underbrace{e_m^* \phi_{p+1}(tT_m) e_1}_{\geq 0} \tag{5.10a}$$

and

$$\|L_m^{p,+}(t)v\|_2 \leq \|Av_{m+1}\|_2\,\tau_{m+1,m}\, t^2 \underbrace{e_m^* \phi_{p+2}(tT_m) e_1}_{\geq 0}. \tag{5.10b}$$

Proof For an Hermitian A we obtain a symmetric, tridiagonal matrix T_m with distinct, real eigenvalues via Lanczos approximation, see [18, Chap. 3.1]. By Proposition 2 we observe

$$\delta_m(t) = f_t[\lambda_1, \ldots, \lambda_m]\gamma_m$$

with $f_t(\lambda) = \mathrm{e}^{t\,\lambda}$ for the case $\sigma = 1$. For divided differences of real-valued functions over real nodes we obtain $f_t[\lambda_1, \ldots, \lambda_m] \in \mathbb{R}$ and

$$f_t[\lambda_1, \ldots, \lambda_m] = \frac{D^{(m-1)} f_t(\xi)}{(m-1)!} = \frac{t^{m-1}\mathrm{e}^{t\xi}}{(m-1)!} \quad \text{for } \xi \in [\lambda_1, \lambda_m]. \tag{5.11}$$

Equation (5.11) shows $f_t[\lambda_1, \ldots, \lambda_m] > 0$ and with $\gamma_m > 0$ we conclude

$$\delta_m(t) > 0, \quad \text{and} \quad |\delta_m(t)| = \delta_m(t).$$

We continue with (5.10a) in the case $p = 0$:

$$\|L_m(t)v\|_2 \le \tau_{m+1,m} \int_0^t |\delta_m(s)|\,\mathrm{d}s = \tau_{m+1,m} \int_0^t e_m^* \mathrm{e}^{sT_m} e_1\,\mathrm{d}s = \tau_{m+1,m} t\, e_m^* \phi_1(tT_m)\, e_1.$$

For the case $p \ge 1$ we start analogously to Theorem 2. Using definition (4.1b) for the ϕ-functions and resorting to the case $p = 0$ we find

$$\begin{aligned}
\|L_m^p(t)v\|_2 &= \frac{1}{(p-1)!} \int_0^1 (1-\theta)^{p-1} \|L_m(\theta t)v\|_2\,\mathrm{d}\theta \\
&\le \tau_{m+1,m}\, t\, \frac{1}{(p-1)!} \int_0^1 (1-\theta)^{p-1}\, \theta\, e_m^* \phi_1(\theta t T_m)\, e_1\,\mathrm{d}\theta.
\end{aligned}$$

Evaluation of the integral yields

$$\begin{aligned}
\|L_m^p(t)v\|_2 &\le \tau_{m+1,m}\, t\, \frac{1}{(p-1)!} \int_0^1 (1-\theta)^{p-1}\, \theta\, e_m^* \phi_1(\theta t T_m)\, e_1\,\mathrm{d}\theta \\
&= \tau_{m+1,m}\, t \sum_{k=0}^{\infty} \frac{e_m^* (t T_m)^k e_1}{(p-1)!\,(k+1)!} \int_0^1 (1-\theta)^{p-1} \theta^{k+1}\,\mathrm{d}\theta \\
&= \tau_{m+1,m}\, t \sum_{k=0}^{\infty} \frac{e_m^* (t T_m)^k e_1}{(p+k+1)!} \\
&= \tau_{m+1,m}\, t\, e_m^* \phi_{p+1}(t T_m)\, e_1.
\end{aligned} \tag{5.12}$$

This shows (5.10a). To show (5.10b) we start analogously to Theorem 3:

$$\|L_m^{p,+}(t)v\|_2 = \|t A L_m^{p+1}(t)v\|_2 \le \|Av_{m+1}\|_2\, \tau_{m+1,m}\, t\, \frac{1}{p!} \int_0^1 (1-\theta)^p \int_0^{\theta t} |\delta_m(s)|\,\mathrm{d}s\,\mathrm{d}\theta.$$

Using $|\delta_m(s)| = \delta_m(s)$ and evaluating the inner integral by the ϕ_1 function, we obtain

$$\|L_m^{p,+}(t)v\|_2 \le \|Av_{m+1}\|_2\, \tau_{m+1,m}\, t^2\, \frac{1}{p!} \int_0^1 (1-\theta)^p\, \theta\, e_m^* \phi_1(\theta t T_m)\, e_1\,\mathrm{d}\theta.$$

Evaluation of the integral analogously to (5.12),

$$\|L_m^{p,+}(t)v\|_2 \le \|Av_{m+1}\|_2\, \tau_{m+1,m}\, t^2\, e_m^* \phi_{p+2}(t T_m)\, e_1,$$

completes the proof. □

6 Defect-based quadrature error estimates revisited

The term on the right-hand side of (2.13) is a computable error estimate, which has been investigated more closely in Section 3. It can also be interpreted in an alternative way. To this end we again proceed from the integral representation (2.10b),

$$L_m(t)v = \int_0^t \underbrace{E(t-s)\,D_m(s)}_{=:\,\Theta_m(s,t)} v\,\mathrm{d}s. \tag{6.1}$$

Due to $\|D_m(t)v\| = \mathscr{O}(t^{m-1})$,

$$\tfrac{\mathrm{d}^j}{\mathrm{d}s^j} D_m(s)v\big|_{s=0} = 0, \quad j = 0,\dots,m-2,$$

and the same is true for the integrand in (6.1),

$$\tfrac{\partial^j}{\partial s^j} \Theta_m(s,t)v\big|_{s=0} = 0, \quad j = 0,\dots,m-2.$$

Analogously as in [3], this allows us to approximate (6.1) by an Hermite-type quadrature formula in the form

$$\int_0^t \Theta_m(s,t)v\,\mathrm{d}s \approx \frac{t}{m}\,\Theta_m(t,t)v = \frac{t}{m}\,D_m(t)v. \tag{6.2}$$

From (2.12),

$$\frac{t}{m}\,D_m(t)v = \tau_{m+1,m}\,\gamma_m \frac{(\sigma t)^m}{m!}\,v_{m+1} + \mathscr{O}(t^{m+1}),$$

which is the same as (2.13). This means that the quadrature approximation (6.2) approximates the leading error term in an asymptotically correct way. From (6.2), (2.11) and (3.1a) we obtain

$$\|L_m(t)v\|_2 \approx \tau_{m+1,m}\frac{t}{m}|\delta_m(t)|. \tag{6.3}$$

The quadrature error in (6.2) is $\mathscr{O}(t^{m+1})$. It is useful to argue this also in a direct way: By construction, the Hermite quadrature formula underlying (6.2) is of order m, and its error has the Peano representation (cf. also [3])

$$\frac{t}{m}\,\Theta_m(t,t) - \int_0^t \Theta_m(s,t)v\,\mathrm{d}s = \int_0^t \frac{s\,(t-s)^{m-1}}{m!}\,\tfrac{\partial^m}{\partial s^m}\,\Theta_m(s,t)v\,\mathrm{d}s. \tag{6.4}$$

Here, $\frac{\partial^m}{\partial s^m}\,\Theta_m(s,t)v = \mathscr{O}(1)$, because $\frac{\mathrm{d}^m}{\mathrm{d}s^m}\,D_m(s)v = \mathscr{O}(1)$ which follows from $D_m(s)v = \mathscr{O}(s^{m-1})$. This shows that, indeed, the quadrature error (6.4) is $\mathscr{O}(t^{m+1})$. Furthermore, a quadrature formula of order $m+1$ can be constructed by including an additional evaluation of

$$\tfrac{\partial}{\partial s}\,\Theta_m(s,t)v\big|_{s=t} = D_m^{[2]}(t)v, \quad \text{with} \quad D_m^{[2]}(t) = \tfrac{\mathrm{d}}{\mathrm{d}t}D_m(t) - \sigma A\,D_m(t).$$

A routine calculation shows

$$\int_0^t \Theta_m(s,t)v\,\mathrm{d}s = \frac{2t}{m+1}\,D_m(t)v - \frac{t^2}{m(m+1)}\,D_m^{[2]}(t)v + \mathscr{O}(t^{m+2}), \tag{6.5}$$

where the error depends on $\frac{\mathrm{d}^{m+1}}{\mathrm{d}s^{m+1}}\,D_m(s)v = \mathscr{O}(1)$. This may be considered as an improved error estimate[6] which can be evaluated using

$$\tfrac{\mathrm{d}}{\mathrm{d}t}D_m(t)v = \sigma^2\tau_{m+1,m}\,e_m^*(T_m\,\mathrm{e}^{\sigma t\,T_m})e_1 v_{m+1}.$$

With the solution in the Krylov subspace, $\mathrm{e}^{\sigma t\,T_m}e_1$ with $e_m^*\mathrm{e}^{\sigma t\,T_m}e_1 = (\mathrm{e}^{\sigma t\,T_m}e_1)_m$, we can compute the derivative of the defect at $\mathscr{O}(1)$ cost,

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}D_m(t)v &= \sigma^2\tau_{m+1,m}\,e_m^*(T_m\,\mathrm{e}^{\sigma t\,T_m})e_1 v_{m+1}\\ &= \sigma^2\tau_{m+1,m}\left((T_m)_{m,m}\,(\mathrm{e}^{\sigma t\,T_m}e_1)_m + (T_m)_{m,m-1}\,(\mathrm{e}^{\sigma t\,T_m}e_1)_{m-1}\right)v_{m+1}.\end{aligned}$$

[6] In the setting of [3] (higher-order splitting methods) such an improved error estimate was not taken into account since it cannot be evaluated with reasonable effort in that context.

Also longer expansions may be considered, for instance

$$\int_0^t \Theta_m(s,t)v\,\mathrm{d}s = \frac{3t}{m+2}D_m(t)v - \frac{3t^2}{(m+1)(m+2)}D_m^{[2]}(t)v + \frac{t^3}{m(m+1)(m+2)}D_m^{[3]}(t)v + \mathscr{O}(t^{m+3}), \quad \text{with} \quad D_m^{[3]}(t) = \tfrac{\mathrm{d}}{\mathrm{d}t}D_m^{[2]}(t) - AD_m^{[2]}(t),$$

etc. This alternative way of computing improved error estimates is worth investigating but will not be pursued further here.

Quadrature estimate for (3.1b) *revisited.* In [22] it is suggested to use the trapezoidal rule as a practical approximation to the integral (3.1b),

$$\|L_m(t)v\|_2 \le \tau_{m+1,m}\int_0^t |\delta_m(s)|\,\mathrm{d}s \approx \tau_{m+1,m}\tfrac{t}{2}\,|\delta_m(t)|, \tag{6.6}$$

or alternatively the Simpson rule. Applying Hermite-type quadrature in (3.1b) also directly leads to the error estimate (6.3). In contrast to (6.3), the error estimate (6.6) is not asymptotically correct for $t \to 0$, but an upper bound.

For a better understanding of the approximation (6.6) we consider the effective order of $|\delta_m(t)|$ as a function of t. Let us denote $f(t) := |\delta_m(t)|$ and assume $f(t) > 0$ in a sufficiently small interval $(0,T]$. For the Hermitian case this assumption is fulfilled for all $t > 0$, see Proposition 6.

The effective order of the function $f(t)$ can be understood as the slope of the double-logarithmic function

$$\xi(\tau) = \ln(f(\mathrm{e}^\tau)) \quad \text{with} \quad \tau = \ln t,$$
$$\text{with derivative} \quad \xi'(\tau) = \frac{f'(\mathrm{e}^\tau)\,\mathrm{e}^\tau}{f(\mathrm{e}^\tau)}.$$

We denote the order by

$$\rho(t) = \frac{f'(t)\,t}{f(t)}, \tag{6.7}$$

and obtain

$$f(t) = \frac{f'(t)\,t}{\rho(t)}.$$

Integration and application of the mean value theorem shows the existence of $t^* \in [0,t]$ with

$$\int_0^t f(s)\,\mathrm{d}s = \frac{1}{\rho(t^*)}\int_0^t f'(s)\,s\,\mathrm{d}s,$$

and integration by parts gives

$$\int_0^t |\delta_m(s)|\,\mathrm{d}s = \frac{t\,|\delta_m(t)|}{1+\rho(t^*)}.$$

With the plausible assumption that the order is bounded by $1 \le \tilde{m} \le \rho(t) \le m-1 = \rho(0+)$ for $t \in [0,T]$, we obtain

$$\tfrac{t}{m}\,|\delta_m(t)| \le \int_0^t |\delta_m(s)|\,\mathrm{d}s \le \tfrac{t}{\tilde{m}+1}\,|\delta_m(t)| \le \tfrac{t}{2}\,|\delta_m(t)|. \tag{6.8}$$

The inequalities (6.8) show that the error estimate based on the trapezoidal quadrature (6.6) leads to an upper bound of the error. The error estimate based on Hermite-type quadrature (6.3) leads to a lower bound of the integral (3.1b), which not necessarily leads to a lower bound for the norm of the error.

Remark 4 With $\rho(0+) = m-1$ and the assumptions that the effective order is slowly decreasing locally at $t = 0$ and sufficiently smooth, we suggest choosing $\tilde{m} = \rho(t)$ for a step of size t improve the quadrature based estimate.

$$\|L_m(t)v\|_2 \le \tau_{m+1,m}\int_0^t |\delta_m(s)|\,\mathrm{d}s \approx \tau_{m+1,m}\tfrac{t}{\rho(t)+1}\,|\delta_m(t)|, \tag{6.9}$$

We will refer to this as *effective order quadrature* estimate. In the limit $t \to 0$ this choice of quadrature is equivalent to the Hermite quadrature and, therefore, asymptotically correct.

Up to now we did refer to the effective order of the defect $|\delta_m(t)|$. For $t \to 0$ the effective order of the error is given by $\rho(t)+1$.

7 The matrix exponential as a time integrator.

We recall from [16] that superlinear convergence as a function of m, the dimension of the underlying Krylov space, sets in for

$$t\,\|A\|_2 \lessapprox m. \tag{7.1}$$

This relation also affects the error considered as a function of time t. Equation (7.1) can be seen as a very rough estimate for a choice of t which leads to a systematic error and convergence behavior. Only for special classes of problems as for instance symmetric negative definite matrices, the relation (7.1) can be weakened, see [16] for details.

In general a large time step t would necessitate large m or a restart of the Krylov method. However, choosing m too large can lead to a deviation from orthogonality of the Krylov basis due to round-off effects and is not recommended in general. Restarting the Krylov method is thus preferable. For the matrix exponential seen as a time propagator, a simple restart is possible. The following procedure has been introduced in [33] and is recapitulated here to fix the notation.

We split the time range $[0,t]$ into N subintervals,

$$\begin{aligned} 0 = t_0 < t_1 < \ldots < t_N = t, \\ \text{with step sizes } \ \Delta t_j = t_j - t_{j-1}, \quad j = 1,\ldots,N. \end{aligned}$$

The exact solution at time t_j is denoted by $v^{[j]}$, whence

$$v^{[j]} = E(\Delta t_j)v^{[j-1]} = E(t_j)v, \quad \text{with} \quad v^{[0]} = v.$$

For simplicity we assume that the dimension m of the Krylov subspace is fixed over the substeps. We obtain approximations $w^{[j]}$ to $v^{[j]}$ by applying multiple restarted Krylov steps, with orthonormal bases $V_m^{[j]}$ and upper Hessenberg matrices $T_m^{[j]}$. Starting from $w^{[0]} = v$, for $j = 1,\ldots,N$,

$$w^{[j]} := S_m^{[j]}(\Delta t_j)w^{[j-1]} = V_m^{[j]}\mathrm{e}^{\sigma\,\Delta t_j T_m^{[j]}}\left(V_m^{[j]}\right)^* w^{[j-1]} = V_m^{[j]}\mathrm{e}^{\sigma\,\Delta t_j T_m^{[j]}}e_1.$$

The error matrix in the j-th step is denoted by

$$L_m^{[j]}(\Delta t_j) := E(\Delta t_j) - S_m^{[j]}(\Delta t_j),$$

and the accumulated error by

$$L_m^\star(t)v = v^{[N]} - w^{[N]}. \tag{7.2}$$

With

$$\begin{aligned} v^{[j]} - w^{[j]} &= E(\Delta t_j)v^{[j-1]} - S_m^{[j]}(\Delta t_j)w^{[j-1]} \\ &= E(\Delta t_j)(v^{[j-1]} - w^{[j-1]}) + L_m^{[j]}(\Delta t_j)w^{[j-1]} \end{aligned}$$

we obtain

$$L_m^\star(t)v = \sum_{j=1}^{N} E(\Delta t_N)\cdots E(\Delta t_{j+1})\,L_m^{[j]}(\Delta t_j)w^{[j-1]}.$$

Recall our premise that $E(\cdot)$ is nonexpansive and assume that the local error is bounded by

$$\|L_m^{[j]}(\Delta t_j)w^{[j]}\|_2 \le \mathrm{tol}\cdot\Delta t_j. \tag{7.3}$$

Then,

$$\|L_m^\star(t)v\|_2 \le \sum_{j=1}^{N}\|L_m^{[j]}(\Delta t_j)w^{[j-1]}\|_2 \le \mathrm{tol}\sum_{j=1}^{N}\Delta t_j = \mathrm{tol}\cdot t.$$

The term $\|L_m^{[j]}(\Delta t_j)w^{[j]}\|_2$ denotes the truncation error of a single substep and is studied in the first part of this paper. We now apply local error estimates to predict acceptable time steps.

Step size control. For a single substep, the error estimate (Err_a) suggests a step size to satisfy a given error tolerance tol as

$$\Delta t_j = \left(\frac{\mathrm{tol}\, m!}{\tau_{m+1,m}^{[j]} \gamma_m^{[j]}} \right)^{1/m}, \quad j = 1, \ldots, N. \tag{7.4}$$

For a local error as in (7.3), we replace tol by $(\Delta t_j\, \mathrm{tol})$ in (7.4) and obtain

$$\Delta t_j = \left(\frac{\mathrm{tol}\, m!}{\tau_{m+1,m}^{[j]} \gamma_m^{[j]}} \right)^{1/(m-1)}, \quad j = 1, \ldots, N. \tag{7.5}$$

We will refer to the step size given by (7.5) as asymptotic step size. We remark that Δt_j can be computed together with the construction of the Krylov subspace, therefore, $\tau_{m+1,m}^{[j]}$ and $\gamma_m^{[j]}$ are known values at this point. For the corrected Krylov approximation $S_m^+(t)v^{[j]}$, see (5.3), the error estimate given in (5.9b) ($p = 0$) suggests a local step size of

$$\Delta t_j = \left(\frac{\mathrm{tol}\, (m+1)!}{\left\| A v_{m+1}^{[j]} \right\|_2 \tau_{m+1,m}^{[j]} \gamma_m^{[j]}} \right)^{1/m}, \quad j = 1, \ldots, N. \tag{7.6}$$

The error estimator Err_1 and estimates given in Section 6 cannot be inverted directly to predict the step size. Computing a feasible step size is still possible via heuristic step size control. This approach will be formulated for the error estimate Err_1 but can also be used in conjunction with all the error estimates given in Section 6. Ideas of heuristic step size control are given in [13] in general and [33] or [28] for a Krylov approximation of the matrix exponential. For a step, with step size Δt_{j-1} and estimated error $\mathrm{Err}^{[j-1]}$, a reasonable size for the subsequent step can be chosen as

$$\Delta t_j = \left(\frac{\Delta t_{j-1}\, \mathrm{tol}}{\mathrm{Err}^{[j-1]}} \right)^{1/m} \Delta t_{j-1}, \quad j = 2, \ldots, N. \tag{7.7}$$

In (7.7) we only need the evaluation of the error estimate for the previously computed step with step size Δt_{j-1}, and we can substitute Err_1 for $\mathrm{Err}^{[j-1]}$. The heuristic step size control always requires information from the previous step, therefore, heuristic step size control cannot be used to compute the first time step Δt_1. The first step is then usually based on a priori estimates, which in general do not provide very sharp results. The first step size in [33] is chosen as

$$\Delta t_1 = \frac{1}{\|H\|_\infty} \left(\frac{\mathrm{tol}\, ((m+1)/\mathrm{e})^{m+1} \sqrt{2\pi (m+1)}}{4 \|H\|_\infty} \right)^{1/m}. \tag{7.8}$$

Remark 5 In the case of A Hermitian the matrix T_m is symmetric, tridiagonal and real-valued which allows cheap and robust computation of its eigenvalue decomposition. The eigenvalue decomposition of T_m is independent of the step size Δt and allows cheap evaluation of Err_1 for multiple choices of Δt. In this case we can cheaply adapt the choice of Δt_j in an iterative manner before continuing to time step $j+1$:

$$\begin{aligned} &\Delta t_{j,1} := \Delta t_{j-1} \ \text{ or result of (7.5)}, \\ &\Delta t_{j,l} := \left(\frac{\Delta t_{j,l-1}\, \mathrm{tol}}{\mathrm{Err}^{[j,l-1]}} \right)^{1/m} \Delta t_{j,l-1}, \quad l = 2, \ldots, N_j, \\ &\Delta t_j := \Delta t_{j,N_j}. \end{aligned} \tag{7.9}$$

By choosing $\mathrm{Err}^{[j,l-1]}$ as an error estimate for the Krylov approximation of the j-th step with time step $\Delta t_{j,l-1}$. For $\mathrm{Err}^{[j,l-1]}$ we can use Err_1 or estimates given in Section 6. The aim of the iteration (7.9) is to determine a step size $\Delta t_{j,\infty}$ with $\mathrm{Err}^{[j,\infty]} = \mathrm{tol}$. The convergence behavior of iteration (7.9) depends on the structure of the error estimate Err_1. In our practical examples this iteration converges monotonically in a small number of steps.

This step size iteration can also be used for the case of non-Hermitian A with the drawback of higher computational cost for the error estimate $\mathrm{Err}^{[j,l-1]}$ for every j and l.

8 Numerical considerations and examples

In this section we first give an illustration of our theoretical results. Then we demonstrate the performance our error estimate (3.4) for practical step size control. To show that this error estimate is efficient in practice we compare it with results delivered by the standard package Expokit [33]. For our tests we use different types of matrices.

Free Schrödinger equation and heat equation. We consider

$$H = \frac{1}{4}\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{8.1}$$

with dimension $n = 10000$. The matrix H corresponds to the symmetric second order finite difference discretization of the one-dimensional negative Laplacian operator. By rescaling time t we can rescale the matrix H arbitrarily. Therefore, without loss of generality, we are using the scalar factor $1/4$ to scale the eigenvalues of H such that $\lambda \in (0,1)$ for all $\lambda \in \mathrm{spec}(H)$.

With $A = H$ and $\sigma = -\mathrm{i}$ in (2.1) we obtain the free Schrödinger equation. The eigenvalue decomposition of H is well known, and we can use the discrete sine transform with high precision arithmetic in Matlab to compute the exact solution $E(t)v$, see (2.1). The starting vector v is chosen randomly. To compute the Krylov subspace approximation $S_m(t)v$, see (2.7), we use the eigenvalue decomposition of the tridiagonal matrix T_m. To obtain the heat equation in (2.1) we choose $A = H$ and $\sigma = -1$.

Discrete Hubbard model. For the description of the Hubbard model we employ a self-contained notation. The Hubbard model first appears in [19] and was further used in many papers and books, e.g. [23, 30]. The Hubbard model is used to describe electron density on a given number of sites, which correspond to Wannier discretization of orbitals, and spin up or down. We consider the following Hubbard Hamiltonian, in second quantization and without chemical potential:

$$H = \frac{1}{2}\sum_{i,j,\sigma} v_{ij} c^\dagger_{j\sigma} c_{i\sigma} + \sum_{j,\sigma} U \hat{n}_{j\sigma}\hat{n}_{j\sigma'}, \tag{8.2}$$

where i, j sum over the number of sites n_{sites} and the spins $\sigma, \sigma' \in \{\uparrow, \downarrow\}$ where σ' is the opposite spin to σ. The entries v_{ij} with $i, j = 1, \ldots, n_{\text{sites}}$ describe electron hopping from site i to j. In (8.2), the notation $c^\dagger_{j\sigma} c_{i\sigma}$ describes the 2nd quantization operator and $\hat{n}_{j\sigma} = c^\dagger_{j\sigma} c_{j\sigma}$ the occupation number operator. For details on the notation in (8.2) we can recommend several references, e.g. [19, 20, 23, 30].

For our tests we model 8 electrons at 8 sites ($n_{\text{sites}} = 8$) with spin up and down for each site, this leads to 16 possible states for electrons. Such an electron distribution is also referred to as half-filled in the literature. We also restrict our model by considering the number of electrons with spin up and down to be fixed as $n_{\text{sites}}/2$. This leads to $n = (\text{binomial}(8,4))^2 = 4900$ considered occupation states which create a discrete basis. For the numerical implementation of the basis we consider 16-bit integers for which each bit describes a position which is occupied in case the bit is equal to 1 or empty otherwise. The set of occupation states can be ordered by the value of the integers which leads to a unique representation of the Hubbard Hamiltonian (8.2) by a matrix $H \in \mathbb{C}^{n\times n}$. Such an implementation of the Hubbard Hamiltonian is also described in [20, Section 3].

In our test setting we use $U = 5$ and parameter-dependent values for electron hopping $v_{ij} = v_{ij}(\omega) \in \mathbb{C}$ with $\omega \in (0, 2\pi]$:

$$\begin{aligned} &v_{11} = v_{88} = -1.75, \quad v_{jj} = -2 \ \text{ for } \ j = 2, \ldots, 7, \\ &v_{j,j+1} = \bar{v}_{j+1,j} = -\cos\omega + \mathrm{i}\sin\omega \ \text{ for } \ j = 1, \ldots, 7 \ \text{ and } \ v_{ij} = 0 \quad \text{otherwise.} \end{aligned}$$

For this choice of $v_{ij}(\omega)$ we obtain an Hermitian matrix $H_\omega \in \mathbb{C}^{n\times n}$ with 43980 nonzero entries (for a general choice of ω) and $\mathrm{spec}(H_\omega) \subset (-19.1, 8.3)$. The spectrum of H_ω is independent of ω.

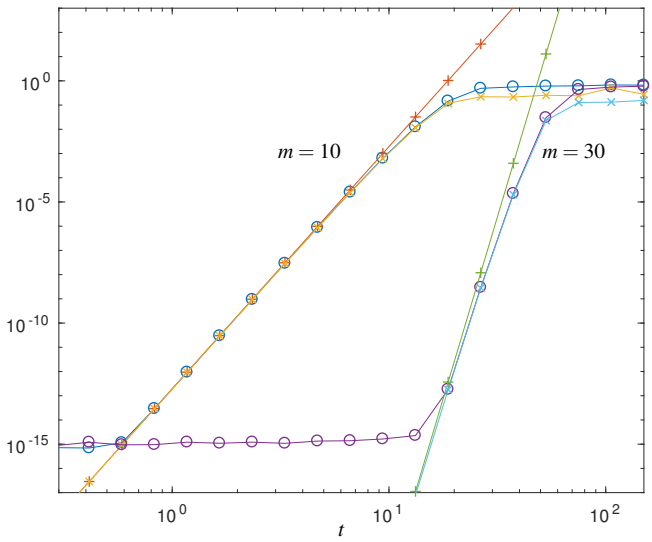


Fig. 8.1 Error $\|L_m(t)v\|_2$ ($\circ$) and the error estimates Err_1 ($\times$) and Err_a ($+$) for the free Schrödinger problem and Krylov subspace dimensions $m = 10$ and $m = 30$. Err_a is an upper bound for the error, and both estimates show the correct asymptotical behavior. Due to round-off error, for $m = 30$ the observed effective order is less clear than for $m = 10$.

A relevant application where the Hubbard Hamiltonian (8.2) is of importance is the simulation of oxide solar cells with the goal of finding candidates for new materials promising a gain in the efficiency of the solar cell, see [9, 14]. The study of solar cells considers time-dependent electron hoppings $v_{ij} = v_{ij}(t)$ to model time-dependent potentials which lead to Hamiltonian matrices $H(t)$. Time propagation of a linear, non-autonomous ODE system can be approximated by Magnus-type integrators which are based on one or more evaluations of matrix exponentials applied to different starting vectors at several times t, see for instance [4]. With different choices of ω the matrices H_ω cover the time-dependent Hamiltonians $H(t)$ given in [9] with corresponding choices of t. Our test setting for the Hubbard Hamiltonian with arbitrary ω is then obtained by (2.1) with the matrix $A = H_\omega$ as described above and $\sigma = -\mathrm{i}$.

In the following Subsection 8.1 we focus on the skew-Hermitian case. For tests on the Hermitian case see Subsection 8.2 below.

8.1 The skew-Hermitian case

Verification of upper bound. In the following Figures 8.1 and 8.2 we compare the error $\|L_m(t)v\|_2$ with the error estimates Err_1 and Err_a. Figure 8.1 refers to the matrix (8.1) of the free Schrödinger problem and Figure 8.2 to the Hubbard Hamiltonian (8.2) with $\omega = 0.123$. For both cases we show results with Krylov subspace dimensions $m = 10$ and $m = 30$, respectively.

We observe that the error estimate Err_1 is a good approximation, but it is not an upper bound in general. In contrast, Err_a is a strict upper bound. Up to round-off error, for $m = 10$ we observe the correct asymptotic behavior of Err_a and Err_1. For larger choices of m the asymptotic regime starts at time steps for which the error is already close to round-off precision. Therefore, for larger choices of m, the Krylov approximation, as a time integrator, cannot achieve its full order for typical time steps in double precision.

The matrix (8.1) of the free Schrödinger problem has been scaled such that $\mathrm{spec}(H) \subset [0,1]$. In accordance with (7.1) stagnation of the error is observed for times $t \lessapprox m$, see Figure 8.1.

We verify the error estimates in the skew-Hermitian setting of the free Schrödinger equation (8.1) for the standard Krylov approximation of the ϕ_1 function in Figure 8.3 and the corrected Krylov approximation of the matrix exponential function in Figure 8.4. In Figure 8.3 the error estimator Err_1 refers to formula (5.8a) and Err_a shows the asymptotically correct strict upper bound (4.4b) from

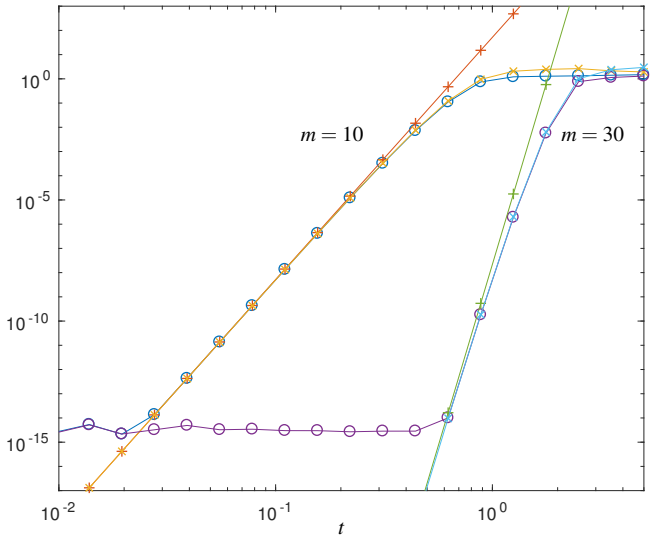


Fig. 8.2 Error $\|L_m(t)v\|_2$ ($\circ$) and the error estimates Err_1 ($\times$) and Err_a ($+$) for the Hubbard Hamiltonian with $\omega = 0.123$ and Krylov subspace dimensions $m = 10$ and $m = 30$. This shows the same behavior as in Figure 8.1.

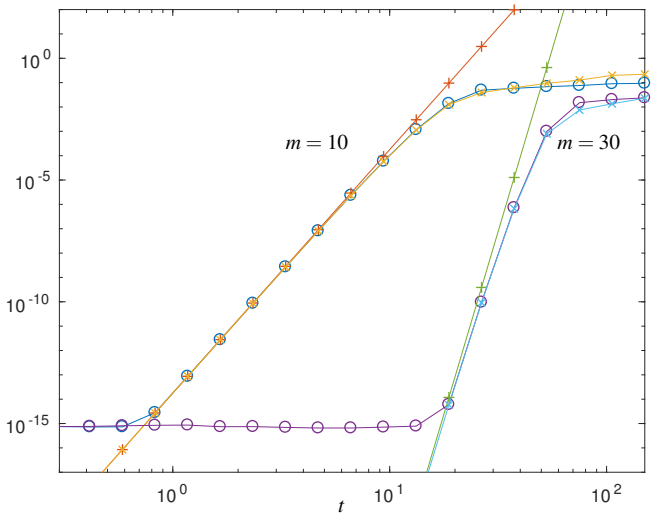


Fig. 8.3 Error $\|L^1_m(t)v\|_2$ ($\circ$) and the error estimates Err_1 ($\times$) and Err_a ($+$) for the free Schrödinger problem and Krylov subspace dimension $m = 10$ and $m = 30$.

Theorem 2, both for the case $p = 1$. In Figure 8.4, Err_1 is from formula (5.8b) and Err_a denotes the asymptotically correct strict upper bound (5.9b) from Theorem 3, both for the case $p = 0$.

Illustration of defect-based quadrature error estimates from Section 6. We first illustrate the performance of the estimates based on Hermite quadrature according to (6.3) and improved Hermite quadrature according to (6.5) for the Hubbard model, see Figure 8.5. Both estimates are asymptotically correct, whereas the improved quadrature (6.5) is slightly better for larger time steps t, with the drawback of one additional matrix-vector multiplication. (See Remark 7 below for cost efficiency of more expensive error estimates.)

Figure 8.6 refers to estimates based on the trapezoidal rule (6.6), the effective order quadrature according to Remark 4, and the Hermite quadrature (6.3). For our test problems the assumptions from Section 6 on the defect and its effective order are satisfied for a significant range of values of t. We also

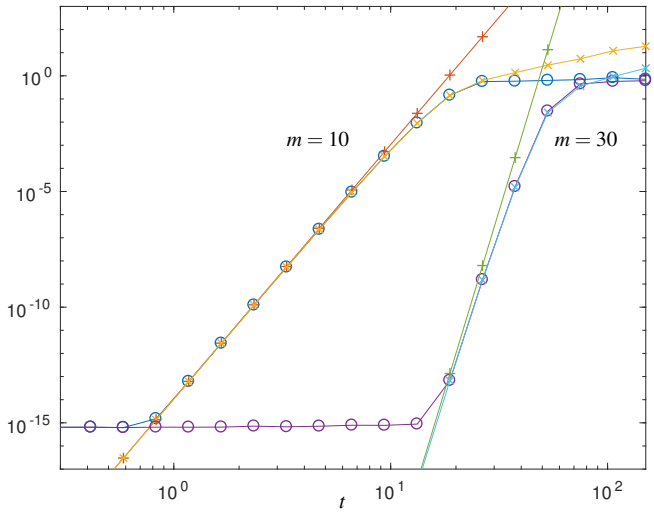


Fig. 8.4 Error $\|L_m^+(t)v\|_2$ ($\circ$) and the error estimates Err_1 ($\times$) and Err_a ($+$) for the free Schrödinger problem and Krylov subspace dimension $m = 10$ and $m = 30$.

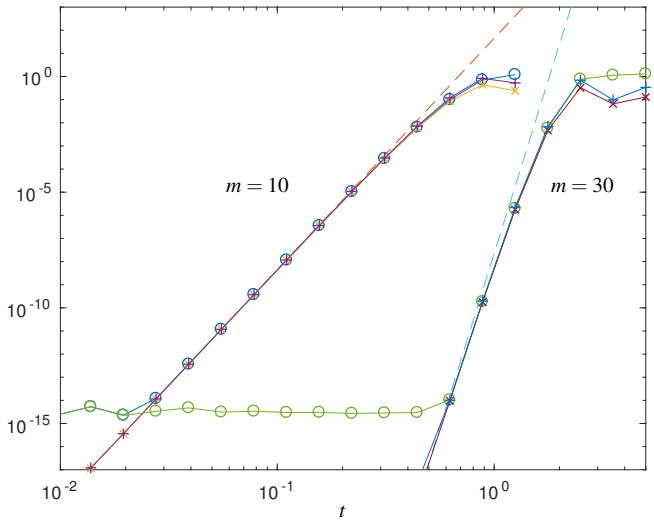


Fig. 8.5 Error $\|L_m^+(t)v\|_2$ ($\circ$) and the error estimates based on the Hermite quadrature ($\times$) and improved Hermite quadrature ($+$), see (6.3) and (6.5), for the Hubbard Hamiltonian with $m = 10$ and $m = 30$. The dashed lines show the error estimate Err_a.

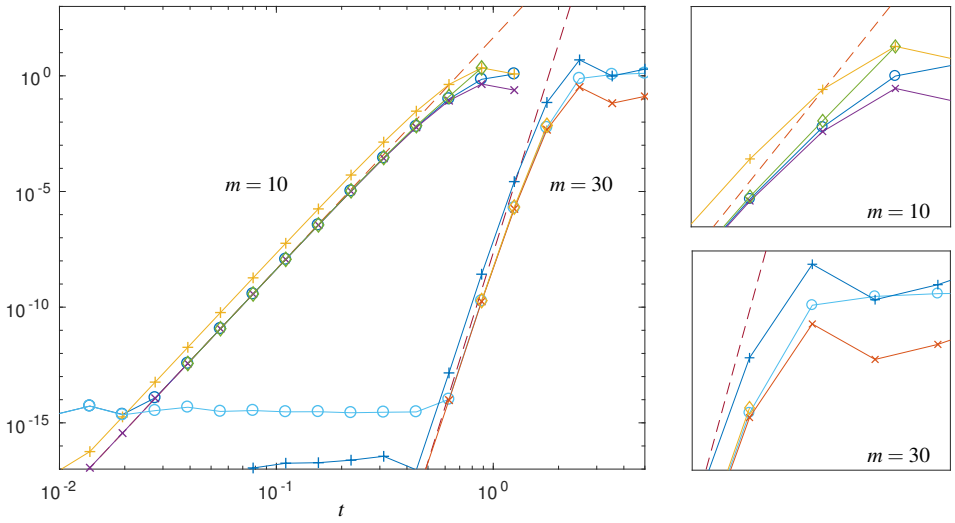
observe that the inequalities (6.8) are satisfied. The effective order and Hermite quadrature estimates behave in an asymptotically correct way, while the trapezoidal rule estimate leads to an upper bound which is not sharp for $t \to 0$.

For the skew-Hermitian case $\sigma = -\mathrm{i}$ and $A = H$ we obtain

$$|\delta_m(t)| = \left((\mathrm{e}^{\mathrm{i}tT_m}e_1)_m(\mathrm{e}^{-\mathrm{i}tT_m}e_1)_m\right)^{1/2} \quad \text{and effective order } \rho(t), \text{ see (6.7)},$$

$$\rho(t) = \frac{t\left(|\delta_m(t)|\right)'}{|\delta_m(t)|} = \frac{\mathrm{i}t\,(T_m)_{m-1,m}}{2}\left(\frac{(\mathrm{e}^{\mathrm{i}tT_m}e_1)_{m-1}}{(\mathrm{e}^{\mathrm{i}tT_m}e_1)_m} - \frac{(\mathrm{e}^{-\mathrm{i}tT_m}e_1)_{m-1}}{(\mathrm{e}^{-\mathrm{i}tT_m}e_1)_m}\right).$$

For computing the effective order we only consider time steps for which the defect is not too close to round-off precision, $\rho(t) > 0$, and where ρ appears indeed to be monotonically decreasing over the computed discrete time steps. This restriction is compatible with our assumptions in Section 6.



Computed effective order of the defect for $m = 10$:

t	$3.9 \cdot 10^{-2}$	$5.5 \cdot 10^{-2}$	$7.8 \cdot 10^{-2}$	$1.1 \cdot 10^{-1}$	$1.5 \cdot 10^{-1}$	$2.2 \cdot 10^{-1}$	$3.1 \cdot 10^{-1}$	$4.4 \cdot 10^{-1}$	$6.3 \cdot 10^{-1}$	$8.8 \cdot 10^{-1}$
$\rho(t)$	8.99	8.98	8.95	8.90	8.81	8.63	8.24	7.44	5.66	1.00

and $m = 30$:

t	$8.8 \cdot 10^{-1}$	$1.2 \cdot 10^{0}$	$1.8 \cdot 10^{0}$
$\rho(t)$	26.68	24.18	18.42

Fig. 8.6 The upper left plot shows the Error $\|L_m^+(t)v\|_2$ ($\circ$) and the error estimates based on the Hermite quadrature (6.3) ($\times$), the trapezoidal rule (6.6) ($+$) and the effective order quadrature (6.9) ($\diamond$) for the Hubbard Hamiltonian with $m = 10$ and $m = 30$. The dashed lines show the error estimate Err_a. On the right-hand side the graphics show a detail from the error plots to illustrate the inequalities (6.8). The table on the bottom shows the computed effective order of the defect for $m = 10$ and $m = 30$ which is used for the effective order quadrature.

Corrected Krylov approximation, mass conservation and loss of orthogonality. We remark that error estimates for the corrected Krylov approximation usually require one additional matrix-vector multiplication, and applying a standard Krylov approximation of dimension $m+1$ seems to be a more favorable choice in our approach to error estimation.

The Krylov approximation of the matrix exponential conserves the mass for the skew-Hermitian case in contrast to the corrected Krylov approximation. Whether this is a real drawback of the corrected Krylov approximation depends on the emphasis placed on mass conservation. In the following examples we focus on the standard Krylov approximation, with some exceptions which serve for comparisons with the original Expokit code, which is based on the corrected Krylov approximation.

In exact arithmetic we obtain mass conservation for the skew-Hermitian case: For the case $\|v\|_2 = 1$ and standard Krylov approximation $S_m(t)v$ we have

$$\|S_m(t)v\|_2 = \|V_m \mathrm{e}^{-\mathrm{i}tT_m} e_1\|_2 = e_1^* \mathrm{e}^{\mathrm{i}tT_m} V_m^* V_m \mathrm{e}^{-\mathrm{i}tT_m} e_1 = 1. \tag{8.3}$$

The requirement $V_m^* V_m = I$ is essential to obtain mass conservation in (8.3). In inexact arithmetic with larger choices of m the loss of orthogonality of the Krylov basis V_m is well known and also observed for the Hubbard Hamiltonian, see Table 8.1. It was shown in [27] that the Lanczos approximation does not suffer critically from round-off errors for larger m. Concerning mass conservation we are not aware of such stability results, but practical examples still show relatively good mass conservation also for larger dimensions m.

Numerical tests for step size control. The idea of choosing discrete time steps for the Krylov approximation is described in Section 7. The following tests are applied to the matrix exponential of the Hubbard Hamiltonian. We first clarify the notation used for our test setting.

Table 8.1 Loss of orthogonality of the Krylov basis for the Hubbard model.

m	10	20	30	40	50	60	70	80
$\lfloor \log_{10}(\|V_m^* V_m - I\|_2) \rfloor$	−14	−14	−13	−11	−10	−8	−6	−5

Expokit and Expokit*. The original Expokit code uses the corrected Krylov approximation with heuristic step size control and an error estimator which is based on the error expansion (5.1), see [33, Algorithm 3.2] for details. Since the standard Krylov approximation is not part of the Expokit package, we have slightly adapted the code and its error estimate such that the standard Krylov approximation is used. We refer to the adapted package as Expokit*. With Expokit* our comparison can be drawn with the standard Krylov approximation which may in some cases be the method of choice as discussed above.

Step size based on Err_a. In another test code the asymptotic error estimate Err_a from Theorem 1 is used. With Err_a we obtain strict upper bounds on the error and reliable step sizes (7.5).

By *tr.quad*, *eff.o.quad*, and *Err*$_1$ we refer to the trapezoidal rule (6.6), the effective order quadrature (6.9), and (Err_1), respectively. Because these error estimates cannot be inverted directly we need to apply heuristic ideas for the step size control, see (7.7). In addition, we use the iteration (7.9) given in Remark 5 to improve step sizes. Monitoring the effective order $\rho(t)$ of the defect (see Section 6) the heuristic step size control can be improved. For the test problems we have solved, iteration (7.9) converges in less than 2 iterations for $m = 10$ or less than 5 iterations for $m = 30$. We simply choose $N_j = 5$ for our tests.

The *a priori estimates* (7.8)*, [16, Theorem 4] and [25, eq. (20)]* are given in the corresponding references. Formula (7.8) taken from the Expokit code directly provides a step size. In [25, eq. (20)] the computation of the step size is described. For the error estimate given in [16, Theorem 4] we apply Newton iteration to determine an appropriate step size. For tests on the Hubbard model we use $(\lambda_{\max} - \lambda_{\min}) = 27.4$ as suggested in the description of the Hubbard Hamiltonian.

In Remark 7 below we also investigate the following variants:

Step size based on Err_a^+. By Err_a^+ we denote the asymptotically correct error estimate for the corrected Krylov approximation as given in Theorem 3 with $p = 0$. The corresponding step size is given by (7.6).

By *i.H.quad* we refer to the improved Hermite quadrature (6.5). Similary to other quadrature error estimates we use heuristic step size control and iteration (7.9) to determine adequate step sizes.

Remark 6 In the Expokit code the step sizes are rounded to 2 digits in every step. Rounding the step size can give too large errors in some steps. This makes it necessary to include safety parameters in Expokit which on the other hand slow down the performance of the code. It seems advisable to avoid any kind of rounding of step sizes.

In Table 8.2 we compare the total time step t for the Krylov approximation with $m = 10$ and $m = 30$ after $N = 10$ steps obtained with the different step size control strategies. For the local error we choose the tolerance $\mathrm{tol} = 10^{-8}$. The original Expokit code seems to give larger step sizes, but it also uses a larger number of matrix-vector multiplications, see Remark 7. The error estimate Err_a leads to optimal step sizes for $m = 10$ and close to optimal step sizes for $m = 30$. For any choice of m the error estimate Err_a gives reliable step sizes. The trapezoidal quadrature rule overestimates the error and, therefore, step sizes are smaller. The effective order quadrature and Err_1 give optimal step sizes. With the assumptions of Section 6 (which apply to our test examples), the trapezoidal and effective order quadrature give reliable step sizes. For the error estimate Err_1 we do not have results on the reliability of the step sizes since the error estimate Err_1 does not lead to an upper bound of the error. The tested a priori estimates (7.8), [16, Th. 4], and [25, (20)] overestimate the error and lead to precautious step size choices. For all the tested versions the accumulated error $L_m^\star$ (see (7.2)) satisfies $\|L_m^\star v\|_2/t \leq \mathrm{tol}$.

Apart from step size control, the cheap asymptotic error estimate Err_a can be used on the fly to test if the dimension of the Krylov subspace is already sufficiently large to solve the problem in a

Table 8.2 The displayed step size t is the sum of $N = 10$ substeps computed by different versions of step size control, as described above. In the top table we show the results for $m = 10$, in the bottom table for $m = 30$, both for tolerance tol $= 10^{-8}$, for the Hubbard Hamiltonian.

$m = 10$	Expokit	Expokit*	Err_a	tr.quad	eff.o.quad	Err_1	(7.8)	[16, Th. 4]	[25, (20)]
t	0.8930	0.6850	0.8422	0.7058	0.8443	0.8444	0.1918	0.4918	0.6879
N	10	10	10	10	10	10	10	10	10
# m-v	110	100	100	100	100	100	100	100	100
$\|L_m^\star v\|_2/t$	$3.4\cdot10^{-09}$	$3.1\cdot10^{-09}$	$9.8\cdot10^{-09}$	$2.0\cdot10^{-09}$	$1.0\cdot10^{-08}$	$1.0\cdot10^{-08}$	$2.4\cdot10^{-14}$	$7.8\cdot10^{-11}$	$1.6\cdot10^{-09}$

$m = 30$	Expokit	Expokit*	Err_a	tr.quad	eff.o.quad	Err_1	(7.8)	[16, Th. 4]	[25, (20)]
t	8.5700	8.2500	9.7361	9.2582	10.2243	10.2338	2.1131	8.2642	8.8111
N	10	10	10	10	10	10	10	10	10
# m-v	310	300	300	300	300	300	300	300	300
$\|L_m^\star v\|_2/t$	$2.4\cdot10^{-10}$	$2.8\cdot10^{-10}$	$2.6\cdot10^{-09}$	$7.0\cdot10^{-10}$	$9.5\cdot10^{-09}$	$9.7\cdot10^{-09}$	$2.9\cdot10^{-15}$	$3.3\cdot10^{-11}$	$1.9\cdot10^{-10}$

Table 8.3 With a test setting similar to Table 8.2, we now compute up to a fixed time $t = 0.3$ and choose the number N of steps according to the step size control. We use a tolerance tol $= 10^{-8}$ and $m = 30$. For this problem we see a significant reduction in the number of matrix-vector multiplications used for the estimate Err_a by the stopping critera described in the text.

$m = 10$	Expokit	Expokit*	Err_a	tr.quad	Err_1	(7.8)	[16, Th. 4]	[25, (20)]
t	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
N	2	2	1	1	1	2	1	1
# m-v	62	60	17	30	30	60	30	30
$\|L_m^\star v\|_2/t$	$8.4\cdot10^{-15}$	$8.4\cdot10^{-15}$	$1.0\cdot10^{-09}$	$9.7\cdot10^{-15}$	$9.7\cdot10^{-15}$	$1.0\cdot10^{-14}$	$9.7\cdot10^{-15}$	$9.7\cdot10^{-15}$

single time step with the required accuracy. For our test problems this stopping criterion is applied to the Err_a estimate, but could be added to any variant of Lanczos. We refer to Table 8.3, in which we observe the Lanczos method with error estimate Err_a to stop after 17 steps instead of computing the full Krylov subspace of dimension 30. In comparison, the original Expokit package needs a total of 62 matrix-vector multiplications.

Remark 7 Error estimates for the corrected Krylov approximation or improved error estimates such as the improved Hermite quadrature (6.5) require additional matrix-vector multiplications. Instead of investing computational effort in improving the error estimate, one may as well increase the dimension of the standard Krylov subspace. For comparison we test the original Expokit code, the corrected Krylov approximation with error estimate Err_a^+ and the improved Hermite quadrature (6.5) with Krylov subspace $m-1$. Table 8.4 shows that a standard Krylov approximation with dimension m leads to better results, although all considered versions use the same number of matrix-vector multiplications. Since the reliability of error estimates such as Err_a has been demonstrated earlier, it appears that additional cost to improve the error estimate is not justified.

8.2 The Hermitian case

To obtain a more complete picture, we also briefly consider the case of an Hermitian matrix $A = H$ with $\sigma = 1$ in (2.1). Such a model is typical of the discretization of a parabolic PDE. Thus, the result may depend on the regularity of the initial data, which is chosen to be random in our experiments.

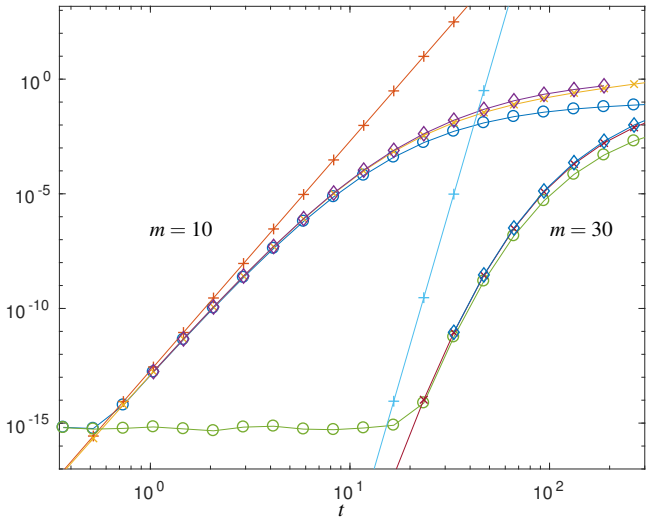
For the heat equation, H given in (8.1), we can also verify the error estimates, see Figure 8.7. In comparison to the skew-Hermitian case we do not observe a large time regime for which the error is of the asymptotic order m. As shown in Proposition 6 we do obtain a strict upper bound using Err_1 for the heat equation. However, the evolution is not unitary but non-expansive in this case, whence the asymptotics are not observed as clearly here.

Similar to the skew-Hermitian case, we can also apply the effective order quadrature according to Remark 4 to the Hermitian case. For the Hermitian case results of Proposition 6 can be applied to

Table 8.4 All variants shown use exactly m matrix-vector multiplications. Whereas Expokit, improved Hermite quadrature (i.H.quad) and Err_a^+ imply higher cost for the error estimate, the other codes Err_a, effective order quadrature (eff.o.quad) and Err_1 use standard Krylov subspaces and do not spend additional matrix-vector multiplications on error estimates.

$m = 10$	Expokit	Err_a^+	i.H.quad	Err_a	eff.o.quad	Err_1
t	0.6620	0.7828	0.5863	0.8346	0.8366	0.8368
N	10	10	10	10	10	10
# m-v	100	100	100	100	100	100
$\|L_m^\star v\|_2/t$	$4.1\cdot10^{-09}$	$8.8\cdot10^{-09}$	$1.0\cdot10^{-08}$	$9.8\cdot10^{-09}$	$1.0\cdot10^{-08}$	$1.0\cdot10^{-08}$

$m = 30$	Expokit	Err_a^+	i.H.quad	Err_a	eff.o.quad	Err_1
t	8.1900	9.5763	9.6591	9.7482	10.2378	10.2473
N	10	10	10	10	10	10
# m-v	100	100	100	100	100	100
$\|L_m^\star v\|_2/t$	$3.6\cdot10^{-10}$	$2.7\cdot10^{-09}$	$9.2\cdot10^{-09}$	$2.6\cdot10^{-09}$	$9.5\cdot10^{-09}$	$9.7\cdot10^{-09}$



Computed effective order of the defect for $m = 10$ (partly):

t	$1.0\cdot10^{0}$	$1.5\cdot10^{0}$	$2.1\cdot10^{0}$	$2.9\cdot10^{0}$	$4.1\cdot10^{0}$	$5.9\cdot10^{0}$	$8.3\cdot10^{0}$	$1.2\cdot10^{1}$	$1.7\cdot10^{1}$	$2.3\cdot10^{1}$	$3.3\cdot10^{1}$
$\rho(t)$	8.50	8.30	8.02	7.64	7.14	6.48	5.65	4.66	3.58	2.52	1.60

and $m = 30$:

t	$3.3\cdot10^{1}$	$4.7\cdot10^{1}$	$6.6\cdot10^{1}$	$9.4\cdot10^{1}$	$1.3\cdot10^{2}$	$1.9\cdot10^{2}$	$2.7\cdot10^{2}$
$\rho(t)$	16.60	13.47	10.33	7.48	5.15	3.36	2.07

Fig. 8.7 Error $\|L_m(t)v\|_2$ ($\circ$), the error estimates Err_1 ($\times$) and Err_a ($+$) and the error estimate based on the effective order quadrature (6.9) ($\diamond$) for the heat equation with $m = 10$ and $m = 30$. The tabular on the bottom shows some of the computed values for the effective order.

obtain

$$|\delta_m(t)| = \delta_m(t) = (\mathrm{e}^{\mathrm{i}tT_m}e_1)_m \quad \text{and effective order } \rho(t), \text{ see (6.7)},$$
$$\rho(t) = \frac{t\left(|\delta_m(t)|\right)'}{|\delta_m(t)|} = t\left(\frac{(T_m)_{m,m-1}\,(\mathrm{e}^{tT_m}e_1)_{m-1}}{(\mathrm{e}^{tT_m}e_1)_m} + (T_m)_{m,m}\right).$$

For computing the effective order we only consider time steps for which the defect is not too close to round-off precision, $\rho(t) > 0$, and where ρ appears indeed to be monotonically decreasing over the computed discrete time steps. This restriction is compatible with our assumptions in Section 6.

9 Summary and outlook.

We have studied new reliable error estimates for Krylov approximations to the matrix exponential and ϕ-functions. An asymptotically correct error estimate gives a strict upper bound to the error, and it can be computed on the fly at nearly no additional cost. The Lanczos process can be stopped as soon as the error estimate satisfies a given tolerance.

Step size control for a simple restarted scheme is an important application. The asymptotic error estimate is an appropriate tool for this task, since the optimal step size for a given tolerance can be computed directly. This is not the case for other error estimates for the Krylov approximation, which usually employ heuristic schemes to compute optimal step sizes in the restarting approach. Also the use of a priori bounds is not optimal in most cases. Comparing our asymptotic step size control with heuristic step size control shows that our procedure allows larger step sizes using reliable error estimates. In addition to better performance, we can avoid safety parameters and assumptions on the convergence of an error expansion. This makes the asymptotic step size control a reliable method to find the optimal step size.

Numerical examples illustrate our theoretical results and show that the asymptotic error estimate is sharp in the case of smaller Krylov subspace dimensions and tolerances. For larger dimensions we observe reduction in the local order. We observed that for t small enough, the local order is monotonically decreasing. This loss of local order will be the topic of further investigation and can possibly lead to enhanced defect-based error estimates.

The shift-and-invert method is a also a relevant approach which deserves further investigations.

Acknowledgements. This work was supported by the Doctoral College TU-D, Technische Universität Wien, and by the Austrian Science Fund (FWF) under the grant P 30819-N32.

References

1. M. Afanasjew, M. Eiermann, O. 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. A.H. Al-Mohy and N.J. Higham. Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM J. Sci. Comput.*, 33(2):488–511, 2011.
3. W. Auzinger, O. Koch, and M. Thalhammer. Defect-based local error estimators for splitting methods, with application to Schrödinger equations, Part II: Higher-order methods for linear problems. *J. Comput. Appl. Math.*, 255:384–403, 2013.
4. S. Blanes and P.C. Moan. Fourth- and sixth-order commutator-free Magnus integrators for linear and non-linear dynamical systems. *Appl. Numer. Math.*, 56:1519–1537, 2005.
5. M. A. Botchev, V. Grimm, and M. Hochbruck. Residual, restarting and Richardson iteration for the matrix exponential. *SIAM J. Sci. Comput.*, 35:A1376–A1397, 2013.
6. E. Celledoni and I. Moret. A Krylov projection method for systems of ODEs. *Appl. Numer. Math.*, 24, 1997.
7. M. Eiermann, O. Ernst, and S. Güttel. Deflated restarting for matrix functions. *SIAM J. Matrix Anal. Appl.*, 32(2):621–641, 2011.
8. J. van den Eshof and M. Hochbruck. Preconditioning Lanczos approximations to the matrix exponential. *SIAM J. Sci. Comput.*, 27:1438–1457, 2006.
9. M. Innerberger et al. Title to be announced, 2018. In preparation.
10. A. Frommer, S. Güttel, and M. Schweitzer. Efficient and stable Arnoldi restarts for matrix functions based on quadrature. *SIAM J. Matrix Anal. Appl.*, 35(2):661–683, 2014.
11. E. Gallopoulos and Y. Saad. Efficient solution of parabolic equations by Krylov approximation methods. *SIAM J. Sci. Statist. Comput.*, 13:1236–1264, 1992.
12. G. H. Golub and C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, 2nd edition, 1989.
13. K. Gustafsson. Control theoretic techniques for stepsize selection in explicit Runge–Kutta methods. *ACM Trans. Math. Software*, 17:533–554, 1991.

14. K. Held. Electronic structure calculations using dynamical mean field theory. *Advances in Physics*, 56:829–926, 2007.
15. N. Higham. *Functions of Matrices. Theory and Computations*. SIAM, Philadelphia, PA, 2008.
16. M. Hochbruck and Ch. Lubich. On Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.*, 34:1911–1925, 1997.
17. M. Hochbruck, Ch. Lubich, and H. Selhofer. Exponential integrators for large systems of differential equations. *SIAM J. Sci. Comput.*, 19:1552–1574, 1998.
18. R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, Cambridge, 1985.
19. J. Hubbard. Electron correlations in narrow energy bands. *Proc. Roy. Soc. London A*, 276:238–257, 1963.
20. S. Jafari. Introduction to Hubbard model and exact diagonalization. *Iranian Journal of Physics Research*, 8, 2008. Available from `http://ijpr.iut.ac.ir/article-1-279-en.pdf`.
21. Zh. Jia and H. Lv. A posteriori error estimates of Krylov subspace approximations to matrix functions. *Numer. Algorithms*, 69:1–28, 2015.
22. Ch. Lubich. *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis*. Zurich Lectures in Advanced Mathematics. European Mathematical Society, Zurich, 2008.
23. G.D. Mahan. *Many-particle physics*. Physics of solids and liquids. Plenum Press, New York, 2nd edition, 1993.
24. N. Mohankumar and S.M. Auerbach. On time-step bounds in unitary quantum evolution using the Lanczos method. *Comput. Phys. Commun.*, 175:473–481, 2006.
25. N. Mohankumar and T. Carrington. A new approach for determining the time step when propagating with the Lanczos algorithm. *Comput. Phys. Commun.*, 181:1859–1861, 2010.
26. C. Moler and C.F. Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev.*, 45(1):3–000, 2003.
27. C. Musco, Ch. Musco, and A. Sidford. Stability of the Lanczos method for matrix function approximation. In A. Czumaj, editor, *Proceeding SODA '18 — Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1605–1624, 2018. Preprint available from `http://arxiv.org/abs/1708.07788`.
28. J. Niesen and W. Wright. Algorithm 919: A Krylov subspace algorithm for evaluating the φ-functions appearing in exponential integrators. *ACM Trans. Math. Software*, 38:22, 2012.
29. T.J. Park and J.C. Light. Unitary quantum time evolution by iterative Lanczos reduction. *J. Chem. Phys.*, 85:5870–5876, 1986.
30. E. Pavarini, E. Koch, J. van den Brink, and G. Sawatzky. *Quantum Materials: Experiments and Theory*, volume 6 of *Modeling and Simulation*. Forschungszentrum Jülich, Jülich, Sep 2016.
31. Y. Saad. Analysis of some Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.*, 29(1):209–228, 1992.
32. Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia, PA, USA, 2nd edition, 2003.
33. R. Sidje. Expokit: A software package for computing matrix exponentials. *ACM Trans. Math. Software*, 24(1):130–156, 1998.
34. D.E. Stewart and T.S. Leyk. Error estimates for Krylov subspace approximations of matrix exponentials. *J. Comput. Appl. Math.*, 72:359–369, 1996.
35. J. Zhongxiao and L. Hui. A posteriori error estimates of Krylov subspace approximations to matrix functions. *Numer. Algorithms*, 69(1):1–28, 2015.