To appear in SIAM J. Numer. Anal. Revised version, January 1996

ON KRYLOV SUBSPACE APPROXIMATIONS TO THE MATRIX EXPONENTIAL OPERATOR

MARLIS HOCHBRUCK* AND CHRISTIAN LUBICH*

Abstract. Krylov subspace methods for approximating the action of matrix exponentials are analyzed in this paper. We derive error bounds via a functional calculus of Arnoldi and Lanczos methods that reduces the study of Krylov subspace approximations of functions of matrices to that of linear systems of equations. As a side result, we obtain error bounds for Galerkin-type Krylov methods for linear equations, namely the biconjugate gradient method and the full orthogonalization method. For Krylov approximations to matrix exponentials, we show superlinear error decay from relatively small iteration numbers onwards, depending on the geometry of the numerical range, the spectrum or the pseudospectrum. The convergence to $\exp(\tau A)v$ is faster than that of corresponding Krylov methods for the solution of linear equations $(I-\tau A)x = v$, which usually arise in the numerical solution of stiff ordinary differential equations. We therefore propose a new class of time integration methods for large systems of nonlinear differential equations which use Krylov approximations to the exponential function of the Jacobian instead of solving linear or nonlinear systems of equations in every time step.

Key words. Krylov subspace methods, conjugate gradient type methods, Arnoldi method, Lanczos method, matrix exponential function, superlinear convergence, matrix-free time integration methods.

AMS(MOS) subject classifications. 65F10, 65L05, 65M15.

1. Introduction.

In this article we study Krylov subspace methods for the approximation of $\exp(\tau A)v$ when A is a matrix of large dimension, v is a given vector, and $\tau > 0$ is a scaling factor which may be associated with the step size in a time integration method. Such Krylov approximations were apparently first used in Chemical Physics [20, 22, 17] and were more recently studied by Gallopoulos and Saad [10, 24]; see also their account of related previous work. They present Krylov schemes for exponential propagation, discuss the implementation, report excellent numerical results, and give some theoretical error bounds. As they also mention, these bounds are however too pessimistic to explain the numerically observed error reductions. Moreover, their error bounds do not make evident that — or when and why — Krylov methods perform far better than standard explicit time stepping methods in stiff problems. A further open question concerns the relationship between the convergence properties of Krylov subspace methods for exponential operators and those for the linear systems of equations arising in implicit time integration methods. In the present paper we intend to clear up the error behavior.

When we wrote this paper, we were unaware of the important previous work by Druskin and Knizhnerman [3, 4, 14, 15] who use a different approach to the analysis. We will comment on the relationship of some of their results to ours in a note at the end of this paper.

Our error analysis is based on a functional calculus of Arnoldi and Lanczos methods which reduces the study of approximations of $\exp(\tau A)v$ to that of the corresponding iterative methods for linear systems of equations. Somewhat oversimplified, it may be said that the error of the mth iterate for $\exp(\tau A)v$ behaves like the minimum,

* Mathematisches Institut, Universität Tübingen, Auf der Morgenstelle 10, D–72076 Tübingen, Germany. E-mail: `marlis@na.uni-tuebingen.de, lubich@na.uni-tuebingen.de`

taken over all $\alpha > 0$, of e^α multiplied with the error of the mth iterate for the solution of $(\alpha I - \tau A)x = v$ by the same Krylov subspace method. This minimum is usually attained far from $\alpha = 1$, especially for large iteration numbers. Unless a good preconditioner for $I - \tau A$ is available, the iteration for $\exp(\tau A)v$ converges therefore faster than that for $(I - \tau A)x = v$. We do not know, however, of a way to "precondition" the iteration for $\exp(\tau A)v$.

Gallopoulos and Saad showed that the error of the mth iterate in the approximation of $\exp(\tau A)v$ has a bound proportional to $\|\tau A\|^m/m!$, which gives superlinear convergence for $m \gg \|\tau A\|$. In many cases, however, superlinear error decay begins for much smaller iteration numbers. For example, we will show that for symmetric negative definite matrices A this occurs already for $m \geq \sqrt{\|\tau A\|}$, whereas for skew-Hermitian matrices with uniformly distributed eigenvalues substantial error reduction begins in general only for m near $\|\tau A\|$. We will obtain rapid error decay for $m \ll \|\tau A\|$ also for a class of sectorial, non-normal matrices. Convergence within the required tolerance for $m < \|\tau A\|$ ensures that the methods become superior to standard explicit time stepping methods for large systems. For $m \geq \|\tau A\|$, our error bounds improve upon those of [10] and [24] typically by a factor $2^{-m}e^{-c\|\tau A\|}$ with a $c > 0$. The analysis explains how the error depends on the geometry of critical sets in the complex plane, namely the numerical range of A for Arnoldi-based approximations, and the location of the spectra or pseudospectra of A and the Krylov-Galerkin matrix H_m for both Lanczos- and Arnoldi-based approximations. In our framework, it is also easily seen that clustering of eigenvalues has similar beneficial effects in the Krylov subspace approximation of $\exp(\tau A)v$ as in the iterative solution of linear systems of equations.

As mentioned above, $\exp(\tau A)v$ can often be computed faster than $(I - \tau A)^{-1}v$ by Krylov subspace methods. This fact has implications in the time integration of very large systems of ordinary differential equations arising, e.g., in many-particle simulations and from spatial discretizations of time-dependent partial differential equations. It justifies renewed interest in ODE methods that use the exponential or related functions of the Jacobian instead of solving linear or nonlinear systems of equations in every time step. Methods of this type in the literature include the exponential Runge-Kutta methods of Lawson [18] and Friedli [8], the adaptive Runge-Kutta methods of Strehmel and Weiner [27] in their non-approximated form, exponentially fitted methods of [5], and the exponential multistep methods of [9].

In the last section of this paper, we propose a promising new class of "exponential" integration methods. With Krylov approximations, substantial savings can be expected for large, moderately stiff systems of ordinary differential equations which are routinely solved by explicit time-stepping methods despite stability restrictions of the step size, or when implicit methods require prohibitively expensive Jacobians and linear algebra.

The paper is organized as follows: In Section 2, we describe the general framework and derive a basic error bound for the Arnoldi method. In Section 3, this leads to specific error bounds for the approximation of $\exp(\tau A)v$ for various classes of matrices A. Lanczos methods are studied in Section 4, which contains also error bounds for BiCG and FOM. In Section 5, we introduce a class of time-stepping methods for large systems of ODEs which replace the solution of linear systems of equations by multiplication with $\varphi(\tau A)$, where $\varphi(z) = (e^z - 1)/z$, whose Krylov subspace approximations converge as fast as those for $\exp(\tau A)v$.

Throughout the paper, $\|\cdot\|$ is the Euclidean norm or its induced matrix norm.

2. Arnoldi-based approximation of functions of matrices. In the sequel, let A be a complex square matrix of (large) dimension N, and $v \in \mathbf{C}^N$ a given vector of unit length, $\|v\| = 1$. The *Arnoldi process* generates an orthonormal basis $V_m = [v_1, \ldots, v_m]$ of the Krylov space $K_m = \text{span}(v, Av, \ldots, A^{m-1}v)$ and a Hessenberg matrix H_m of dimension m (which is the upper left part of its successor H_{m+1}) such that

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T \ , \tag{2.1}$$

where e_i is the ith unit vector in $\mathbf{R}^m$. By induction this clearly implies, as noted in [3, Theorem 2] and [24, Lemma 3.1],

$$q_{m-1}(A)v = V_m \, q_{m-1}(H_m) e_1 \quad \text{for all polynomials } q_{m-1} \text{ of degree } \leq m-1 \ . \tag{2.2}$$

A standard use of the Arnoldi process is in the solution of linear equations [23], where one approximates

$$(\lambda I - A)^{-1} v \approx V_m (\lambda I - H_m)^{-1} e_1 \tag{2.3}$$

when λ is not an eigenvalue of A, and hopefully not of H_m. The latter condition is always satisfied when λ is outside the *numerical range*

$$\mathcal{F}(A) = \{x^* A x : x \in \mathbf{C}^N, \|x\| = 1\},$$

since (2.1) implies $H_m = V_m^* A V_m$ and therefore

$$\mathcal{F}(H_m) \subset \mathcal{F}(A) \ . \tag{2.4}$$

We now turn to the approximation of functions of A. Let f be analytic in a neighborhood of $\mathcal{F}(A)$. Then

$$f(A)v = \frac{1}{2\pi i} \int_\Gamma f(\lambda) \, (\lambda I - A)^{-1} v \, d\lambda \ , \tag{2.5}$$

where Γ is a contour that surrounds $\mathcal{F}(A)$. In view of (2.3), we are led to replace this by

$$\frac{1}{2\pi i} \int_\Gamma f(\lambda) \, V_m (\lambda I - H_m)^{-1} e_1 \, d\lambda = V_m f(H_m) e_1 \ , \tag{2.6}$$

so that we approximate

$$f(A)v \approx V_m f(H_m) e_1 \ . \tag{2.7}$$

Such an approximation was proposed previously [22, 33, 3, 10], with different derivations.

In practice, we are then left with the task of computing the lower-dimensional expression $f(H_m)e_1$, which for $m \ll N$ is usually much easier to compute than $f(A)v$, e.g., by diagonalization of H_m. The above derivation of (2.7) also indicates how to obtain error bounds: Study the error in the Arnoldi approximation (2.3) of linear systems and integrate their error bounds, multiplied with $|f(\lambda)|$, for λ varying along a suitable contour Γ. This will actually be done in the present paper.

Our error bounds are based on Lemma 1 below. To prepare its setting, let E be a convex, closed bounded set in the complex plane. Let ϕ be the conformal mapping that carries the exterior of E onto the exterior of the unit circle $\{|w| > 1\}$, with $\phi(z) = z/\rho + O(1)$ as $z \to \infty$ for a $\rho > 0$. We note that ρ is the logarithmic capacity of E. Finally, let Γ be the boundary curve of a piecewise smooth, bounded region G that contains E, and assume that f is analytic in G and continuous on the closure of G.

Lemma 1. *Under the above assumptions, and if the numerical range of A is contained in E, we have for every polynomial q_{m-1} of degree at most $m-1$*

$$\|f(A)v - V_m f(H_m)e_1\| \leq \frac{M}{2\pi} \cdot \int_\Gamma |f(\lambda) - q_{m-1}(\lambda)| \cdot |\phi(\lambda)|^{-m} \cdot |d\lambda| \tag{2.8}$$

with $M = \ell(\partial E)/[d(\partial E)\cdot d(\Gamma)]$, where $\ell(\partial E)$ is the length of the boundary curve ∂E of E, and where $d(S)$ is the minimal distance between $\mathcal{F}(A)$ and a subset S of the complex plane. If E is a straight line segment or a disk, then (2.8) *holds with $M = 6/d(\Gamma)$.*

Remark. It will be useful to choose the integration contour dependent on m, in order to balance the decay of ϕ^{-m} away from E against the growth of f outside E. For entire functions f, such as the exponential function studied in detail below, this will ultimately yield *superlinear* convergence. On the other hand, the liberty in choosing the polynomial q_{m-1} will not materialize in the study of the exponential function.

Proof. (a) We begin by studying the error of (2.3) and consider a fixed $\lambda \in \Gamma$ for the moment. Our argumentation in this part of the proof is inspired by [25]. Using $e_1 = V_m^* v$, we rewrite the error as

$$(\lambda I - A)^{-1}v - V_m(\lambda I - H_m)^{-1}e_1 = \Delta_m v$$

with $\Delta_m = (\lambda I - A)^{-1} - V_m(\lambda I - H_m)^{-1}V_m^*$. By (2.1) and the orthogonality of $V_{m+1} = [V_m, v_{m+1}]$, we have $V_m^*(\lambda I - A)V_m = \lambda I - H_m$ and therefore

$$\Delta_m(\lambda I - A)V_m = 0.$$

Hence we have for arbitrary $y_m \in \mathbf{C}^m$

$$(\lambda I - A)^{-1}v - V_m(\lambda I - H_m)^{-1}e_1 = \Delta_m\left(v - (\lambda I - A)V_m y_m\right) . \tag{2.9}$$

We note that

$$v - (\lambda I - A)V_m y_m = p_m(A)v ,$$

where p_m is a polynomial of degree $\leq m$ with $p_m(\lambda) = 1$. Conversely, for every such polynomial, $p_m(A)v$ is of the above form. To bound Δ_m, we recall $\|V_m\| = 1$ and use the estimates $\|(\lambda I - A)^{-1}\| \leq \mathrm{dist}(\lambda, \mathcal{F}(A))^{-1}$ and $\|(\lambda I - H_m)^{-1}\| \leq \mathrm{dist}(\lambda, \mathcal{F}(A))^{-1}$ which follow from [26, Thm.4.1] and (2.4). We thus obtain

$$\|(\lambda I - A)^{-1}v - V_m(\lambda I - H_m)^{-1}e_1\| \leq 2\, d(\Gamma)^{-1} \cdot \|p_m(A)\| \tag{2.10}$$

for every polynomial p_m of degree at most m with $p_m(\lambda) = 1$.
(b) It remains to bound $p_m(A)$. Since

$$p_m(A) = \frac{1}{2\pi i}\int_{\partial E} p_m(z)\,(zI - A)^{-1}\,dz ,$$

we have

$$\|p_m(A)\| \le \frac{\ell(\partial E)}{2\pi d(\partial E)} \max_{z\in E} |p_m(z)| \ .$$

For the special case when E is a line segment, we have A of the form $A = \alpha I + \beta B$ with a Hermitian B and complex coefficients α, β, so that then

$$\|p_m(A)\| \le \max_{z\in E} |p_m(z)| \ .$$

When E is a disk $|z-\mu| \le \rho$, then $p_m(z)$ will be chosen as a multiple of $(z-\mu)^m$, and an inequality of Berger (see [1, p. 3]) then tells us that

$$\|p_m(A)\| \le 2 \max_{z\in E} |p_m(z)| \ .$$

In all these cases we thus have

$$\|p_m(A)\| \le \tfrac{1}{3} d(\Gamma) \cdot M \cdot \max_{z\in E} |p_m(z)| \tag{2.11}$$

with M as stated in Lemma 1.
(c) To proceed in the proof, we use near-optimality properties of Faber polynomials. These have been employed previously in analyses of iterative methods by Eiermann [6] and Nevanlinna [21]. Let $\phi_m(z)$ denote the Faber polynomial of degree m associated with the region E. This is defined as the polynomial part of $\phi(z)^m$, i.e., $\phi(z)^m = \phi_m(z) + O(z^{-1})$ as $z \to \infty$. We now choose the polynomial $p_m(z)$ with the normalization $p_m(\lambda) = 1$ as

$$p_m(z) = [\phi_m(z) - \phi_m(\lambda) + \phi(\lambda)^m]/\phi(\lambda)^m \ , \tag{2.12}$$

cf. [21, p.76]. A theorem of Kövari and Pommerenke [16, Thm.2] provides us with the bound

$$|\phi_m(z) - \phi(z)^m| \le 1 \qquad \text{for} \quad z \in \mathbf{C}\backslash E \ . \tag{2.13}$$

This implies $\max_{z\in E} |\phi_m(z)| = \max_{z\in\partial E} |\phi_m(z)| \le 2$ and

$$\max_{z\in E} |p_m(z)| \le 3\, |\phi(\lambda)|^{-m} \ . \tag{2.14}$$

(d) Combining inequalities (2.10),(2.11), and (2.14) gives us

$$\|(\lambda I - A)^{-1} v - V_m(\lambda I - H_m)^{-1} e_1\| \le M \cdot |\phi(\lambda)|^{-m} \ . \tag{2.15}$$

The proof is now completed by inserting this bound into the difference of formulas (2.5) and (2.6) and taking account of (2.2). □

Remark. Part (c) of the above proof, combined with Cauchy's integral formula, shows that there exists a polynomial $\Pi_{m-1}(z)$ of degree at most $m-1$ such that

$$\max_{z\in E} |f(z) - \Pi_{m-1}(z)| \le \frac{3}{2\pi\delta} \int_\Gamma |f(\lambda)| \cdot |\phi(\lambda)|^{-m} \cdot |d\lambda| \ , \tag{2.16}$$

where δ is the minimal distance between Γ and E. This holds for $\Pi_{m-1}(z) = \frac{1}{2\pi i} \int_\Gamma f(\lambda)(1-p_m(z,\lambda))/(\lambda - z)\, d\lambda$, with $p_m(z,\lambda)$ of (2.12). Polynomial approximation bounds of this type are closely tied to Bernstein's theorem [19, Thm.III.3.19].

3. Approximation of the matrix exponential operator. In this section we give error bounds for the Arnoldi approximation of $e^{\tau A}v$ for various classes of matrices A. We may restrict our attention to cases where the numerical range of A is contained in the left half-plane, so that $\|e^{\tau A}\|$ and, in view of (2.4), also $\|e^{\tau H_m}\|$ are bounded by unity. This assumption entails no loss of generality, since a shift from A to $A+\alpha I$ changes both $e^{\tau A}v$ and its approximation by a factor $e^{\tau\alpha}$.

The same bounds as in Theorems 2 to 6 below (even slightly more favorable bounds) are valid also for Krylov subspace approximations of $\varphi(\tau A)v$, with $\varphi(z) = (e^z-1)/z$.

Theorem 2. *Let A be a Hermitian negative semi-definite matrix with eigenvalues in the interval $[-4\rho, 0]$. Then the error in the Arnoldi approximation of $e^{\tau A}v$, i.e., $\varepsilon_m := \|e^{\tau A}v - V_m e^{\tau H_m} e_1\|$, is bounded in the following ways:*

$$\varepsilon_m \le 10\, e^{-m^2/(5\rho\tau)}, \qquad \sqrt{4\rho\tau} \le m \le 2\rho\tau \tag{3.1}$$

$$\varepsilon_m \le 10\, (\rho\tau)^{-1} e^{-\rho\tau} \left(\frac{e\rho\tau}{m}\right)^m, \qquad m \ge 2\rho\tau . \tag{3.2}$$

Remark. It is instructive to compare the above error bounds with that of the conjugate gradient method applied to the linear system $(I-\tau A)x = v$, which is given by

$$\|x - x_m\| \le 2\sqrt{1+4\rho\tau}\left(1 - \frac{2}{\sqrt{1+4\rho\tau}+1}\right)^m .$$

This bound becomes small for $m \gg \sqrt{\rho\tau}$ but only with a linear decay.

Proof. We use Lemma 1 with $E = [-4\rho, 0]$. Then the conformal mapping is

$$\phi(z) = 1 + \frac{z}{2\rho} + \sqrt{\left(1+\frac{z}{2\rho}\right)^2 - 1}.$$

We start by applying the linear transformation $\mu = 1 + \lambda/(2\rho)$, which maps E to the interval $[-1, 1]$. As contour Γ we choose the parabola with right-most point γ that is mapped to the parabola Π given by the parametrization

$$\mu = (1+\epsilon)(1 - \tfrac{1}{2}\theta^2) + i\sqrt{2\epsilon+\epsilon^2}\,\theta , \qquad -\infty < \theta < +\infty ,$$

where $\epsilon = \gamma/(2\rho)$. This parabola osculates to the ellipse $\mathcal{E}$ with foci ± 1 and major semi-axis $1+\epsilon$. Lemma 1 gives us the error bound

$$\varepsilon_m \le \frac{1}{\gamma}\int_\Gamma |e^{\lambda\tau}| \cdot |\phi(\lambda)|^{-m} \cdot |d\lambda| = \frac{1}{\epsilon}\int_\Pi |e^{2\rho\tau(\mu-1)}| \cdot |\Phi(\mu)|^{-m} \cdot |d\mu| , \tag{3.3}$$

where $\Phi(\mu) = \mu + \sqrt{\mu^2-1}$. The absolute value of $\Phi(\mu)$ is constant along every ellipse with foci ± 1. Since the parabola Π is located outside the ellipse $\mathcal{E}$, we have along Π

$$|\Phi(\mu)| \ge \Phi(1+\epsilon) =: r.$$

Hence we obtain from (3.3)

$$\begin{aligned}
\varepsilon_m &\le \epsilon^{-1} e^{2\rho\tau\epsilon} r^{-m} \cdot 2\int_0^\infty e^{-\rho\tau\theta^2}\left((1+\epsilon)\theta + \sqrt{2\epsilon+\epsilon^2}\right) d\theta \\
&= e^{2\rho\tau\epsilon} r^{-m}\left(\frac{1+\epsilon}{\rho\tau\epsilon} + \sqrt{\frac{(2+\epsilon)\pi}{\rho\tau\epsilon}}\right) . \qquad (3.4)
\end{aligned}$$

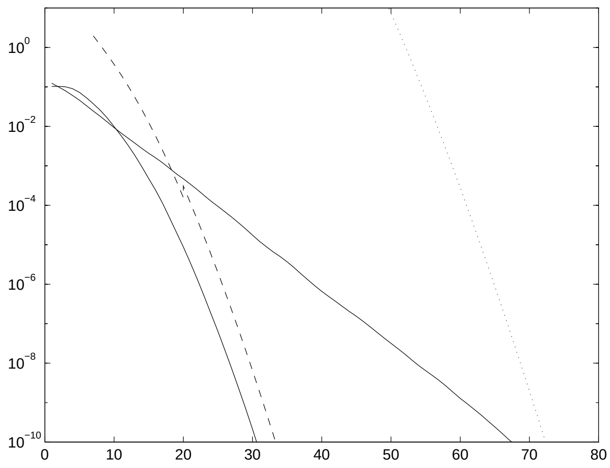

FIG. 3.1. *Errors and error bounds for the symmetric example*

Moreover, we have $r \geq e^{\alpha\sqrt{2\epsilon}}$ with $\alpha > 0.96$ for $\epsilon \leq 1/2$ (and $\alpha > 0.98$ for $\epsilon \leq 1/4$). Minimizing $e^{2\rho\tau\epsilon - m\sqrt{2\epsilon}}$ with respect to ϵ yields

$$\epsilon = \frac{m^2}{8(\rho\tau)^2} .$$

Inserting this ϵ in (3.4) results in the bound (with $\beta = 2\alpha - 1 > 0.92$)

$$\varepsilon_m \leq \left(12\frac{\rho\tau}{m^2} + 8\frac{\sqrt{\rho\tau}}{m}\right) e^{-\beta m^2/(4\rho\tau)} , \tag{3.5}$$

which together with $\varepsilon_m \leq 2$ is a sharper version of (3.1). The condition $\epsilon \leq 1/2$ is equivalent to $m \leq 2\rho\tau$.

To obtain the bound (3.2), we note that $1 + \epsilon = (r + r^{-1})/2$ and insert in (3.4) $r = m/(\rho\tau)$, which is close to the minimum for $m \gg \rho\tau$. This yields for $m \geq 2\rho\tau$

$$\varepsilon_m \leq \left(5(\rho\tau)^{-1} + 3\sqrt{\pi}(\rho\tau)^{-1/2}\right) e^{(\rho\tau)^2/m} e^{-2\rho\tau} \left(\frac{e\rho\tau}{m}\right)^m , \tag{3.6}$$

which is a sharper version of (3.2).

Finally we remark, in view of the proof of Theorem 3 below, that the bounds (3.1) and (3.2) are also obtained when Γ is chosen as a composition of the part of the above parabola contained in the right half-plane and two rays on the imaginary axis. □

To give an illustration of our error bounds we consider the diagonal matrix A with equidistantly spaced eigenvalues in the interval $[-40, 0]$ and a random unit vector v of dimension 1001. Fig. 3.1 shows the errors of the approximation to $\exp(A)v$ and those of the cg approximation to $(I - A)^{-1}v$, which gives nearly a straight line. Moreover, the dashed line shows the error bounds (3.5) and (3.6), while the dotted line corresponds to $2\|\frac{1}{2}A\|^m/m!$, which is the error bound of [24, Corollary 4.6] for symmetric, negative semi-definite matrices A.

It is well known that Krylov subspace methods for the solution of linear systems of equations benefit from a clustering of the eigenvalues. The same is true also for the Krylov subspace approximation of $e^{\tau A}v$. This is actually not surprising in view of the Cauchy integral representations (2.5), (2.6). As an example of such a result, we state the following theorem. This might be generalized in various directions for different types of clusterings and different types of matrices, but we will not pursue this further.

Theorem 3. *Let A be a Hermitian negative semi-definite matrix with eigenvalues contained in $\{\lambda_1\} \cup [-4\rho, 0]$ with $\lambda_1 < -4\rho$. Then the $(m+1)$st error ε_{m+1} in the Arnoldi approximation of $e^{\tau A}v$ is bounded by the right-hand sides of* (3.1) *and* (3.2).

Proof. The result is proved by using the polynomial

$$p_m(z) = \frac{z - \lambda_1}{\lambda - \lambda_1} \left[\phi_{m-1}(z) - \phi_{m-1}(\lambda) + \phi(\lambda)^{m-1}\right] / \phi(\lambda)^{m-1}$$

instead of (2.12). The absolute value of the first factor is bounded by unity for $z \in [-4\rho, 0]$ and $\mathrm{Re}\lambda \geq 0$. Hence we obtain the same error bounds as in Theorem 2 with m replaced by $m-1$. $\square$

For skew-Hermitian matrices A (with uniformly distributed eigenvalues) we cannot show superlinear error decay for $m < \rho\tau$. The reason is, basically, that here the conformal mapping ϕ maps the vertical line $\mathrm{Re}\lambda = \epsilon\rho$ onto a contour with $|\phi(\lambda)| \geq 1 + \epsilon$, whereas in the symmetric negative definite case we have $|\phi(\lambda)| \geq 1 + \sqrt{\epsilon}$. This behavior affects equally the convergence of Krylov subspace methods for the solution of linear systems $(I - \tau A)x = v$. For skew-Hermitian A, there is, in general, no substantial error reduction for $m < \rho\tau$, and convergence is linear with a rate like $(1 + 1/\rho\tau)^{-1}$ for $m \gg \rho\tau$.

Theorem 4. *Let A be a skew-Hermitian matrix with eigenvalues in an interval on the imaginary axis of length 4ρ. Then the error in the Arnoldi approximation of $e^{\tau A}v$ is bounded by*

$$\varepsilon_m \leq 12\, e^{-(\rho\tau)^2/m} \left(\frac{e\rho\tau}{m}\right)^m, \quad m \geq 2\rho\tau.$$

Proof. We use Lemma 1 with $E = i[\alpha - 2\rho, \alpha + 2\rho]$. Then the conformal mapping is

$$\phi(z) = i\left(\frac{z - i\alpha}{2\rho} + \sqrt{\left(\frac{z - i\alpha}{2\rho}\right)^2 + 1}\right).$$

After applying the linear transformation $\mu = i(\lambda - i\alpha)/(2\rho)$, we choose the integration contour as an ellipse with foci ± 1 and minor semiaxis $b = \gamma/(2\rho)$, $\gamma > 0$. The major semiaxis is then $a = \sqrt{1 + b^2}$ and the length of the contour is bounded by $2\pi a$. In addition, we have $d(\Gamma) = 2\rho(a - 1)$. The absolute value $r = |\phi(\lambda)| = |\mu + \sqrt{\mu^2 + 1}|$ is constant along the ellipse. With Lemma 1, we get for the error

$$\varepsilon_m \leq \frac{12\rho}{2\pi d(\Gamma)} 2\pi\sqrt{1 + b^2} e^{\gamma\tau} r^{-m} \leq \frac{6}{1 - 1/\sqrt{1 + b^2}} e^{2\rho\tau b} r^{-m},$$

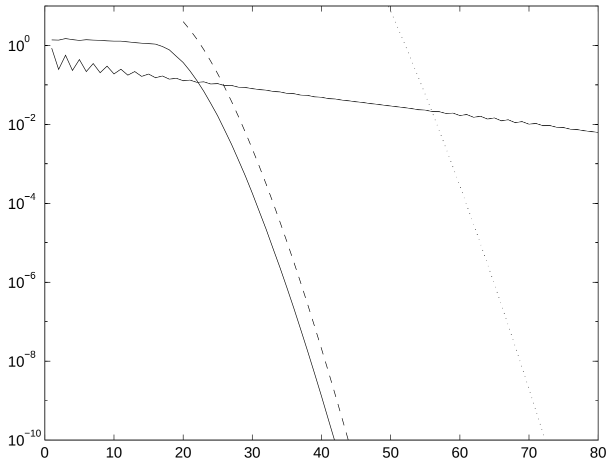

FIG. 3.2. *Errors and error bounds for the skew-Hermitian example*

with $b = (r - r^{-1})/2$. Inserting $r = m/(\rho\tau)$ gives the stated error bound. A sharper bound for $\rho\tau > \frac{1}{2}$, which is obtained by integrating over the parabola that osculates to the above ellipse at the right-most point, reads

$$\varepsilon_m \leq \tfrac{1}{3}\left(4(\rho\tau)^{-1} + 11(\rho\tau)^{-1/2}\right) e^{-(\rho\tau)^2/m} \left(\frac{e\rho\tau}{m}\right)^m, \quad m \geq 2\rho\tau. \tag{3.7}$$ □

For a numerical illustration we choose the diagonal matrix A with 1001 equidistant eigenvalues in $[-20i, 20i]$, and a random vector v of unit length. Fig. 3.2 shows the errors of the approximation to $\exp(A)v$ and those of the BiCG approximation to $(I - A)^{-1}v$, which again form nearly a straight line. The dashed line shows the error bound (3.7), the dotted line corresponds to $2\|A\|^m/m!$, which is the bound given in [10, Corollary 2.2].

Theorem 5. *Let A be a matrix with numerical range contained in the disk $|z+\rho| \leq \rho$. Then the error in the Arnoldi approximation of $e^{\tau A}v$ is bounded by*

$$\varepsilon_m \leq 12\, e^{-\rho\tau} \left(\frac{e\rho\tau}{m}\right)^m, \quad m \geq 2\rho\tau\ .$$

Proof. We use Lemma 1 with $E = \{|z+\rho| \leq \rho\}$ and Γ a circle with radius $r\rho$ centered at $-\rho$. Lemma 1 gives the bound

$$\varepsilon_m \leq 6 \frac{r}{r-1} e^{(r-1)\rho\tau} r^{-m}.$$

Setting $r = m/(\rho\tau)$ gives the stated result. □

The following is a worst-case example which shows nearly no error reduction for $m \leq \rho\tau$.

Example. Let A be the bidiagonal matrix of dimension N that has -1 on the diagonal and $+1$ on the subdiagonal. The numerical range of A is then contained in the disk $|z+1| \leq 1$. For $v = e_1$, we have

$$e^{\tau A}v = e^{-\tau}(1, \tau, \tau^2/2!, \ldots, \tau^N/N!)^T \ .$$

The Arnoldi process gives $V_m = (I_m, 0)^T$ and H_m as the m-dimensional version of A, so that

$$V_m e^{\tau H_m} e_1 = e^{-\tau}(1, \tau, \tau^2/2!, \ldots, \tau^m/m!, 0, 0, \ldots, 0)^T \ .$$

The error vector thus contains the entries $e^{-\tau}\tau^k/k!$ for $k > m$. The largest of these is close to $(2\pi\tau)^{-1/2}$ by Stirling's formula, if $m < \tau < N$. □

Similar to Theorem 2, the onset of superlinear convergence begins already for $m \ll \rho\tau$ when $\mathcal{F}(A)$ is contained in a wedge-shaped set. In particular, consider the conformal mapping

$$\psi(w) = \left(1 - \frac{1}{w}\right)^{2-\theta} w \ , \quad |w| \geq 1$$

for $0 < \theta < 1$, which maps the exterior of the unit disk onto the exterior of the bounded sectorial set in the left half-plane

$$S_\theta = \mathbf{C} \backslash \{z = \psi(w) : |w| > 1\} \ .$$

S_θ has a corner at 0 with opening angle $\theta\pi$ and is symmetric with respect to the real axis.

Theorem 6. *For some $\rho > 0$ and $0 < \theta < 1$, let the numerical range of A be contained in $\rho \cdot S_\theta$. Then the error in the Arnoldi approximation of $e^{\tau A}v$ is bounded by*

$$\varepsilon_m \leq C\left(\frac{\rho\tau}{m}\right)^\gamma e^{-c[m/(\rho\tau)^\alpha]^\beta} \ , \quad (\rho\tau)^\alpha \leq m \leq 2\rho\tau \tag{3.8}$$

$$\varepsilon_m \leq C m\, e^{-(r-\psi(r))\rho\tau}\left(\frac{e\rho\tau}{m}\right)^m \ , \quad r = \frac{m}{\rho\tau} \geq 2 \tag{3.9}$$

with $\alpha = \frac{1}{2-\theta}$, $\beta = \frac{2-\theta}{1-\theta}$, $\gamma \leq \frac{2}{1-\theta}$. The constants C and $c > 0$ depend only on θ.

Proof. In the course of this proof, C denotes a generic constant which takes on different values on different occurrences. After the transformation $\mu = \lambda/\rho$, we use Lemma 1 with $\rho^{-1}E = \mathbf{C}\backslash\{\psi(w) : |w| > 1+\delta\}$ and $\rho^{-1}\Gamma = \{\mu = \psi(w) : |w| = r\}$ for $r = 1 + (2\epsilon)^\alpha$ and suitable $\delta > 0$, $\epsilon > 0$. For $\epsilon \leq 1$, we choose δ such that $r/(1+\delta) = \exp(\epsilon^\alpha)$ and note that then $\rho^{-1}d(\partial E) \geq \delta \geq C^{-1}\epsilon^\alpha$ and also $\rho^{-1}d(\Gamma) \geq C^{-1}\epsilon^\alpha$. The right-most point of the integration contour $\rho^{-1}\Gamma$ is $\psi(r) < 2\epsilon$. Hence we have from Lemma 1

$$\varepsilon_m \leq C \cdot \epsilon^{-2\alpha} e^{2\epsilon\rho\tau} e^{-m\epsilon^\alpha} \ .$$

For $m \gg (\rho\tau)^\alpha$, the right-hand side is minimized near

$$\epsilon = \left(\frac{\alpha m}{2\rho\tau}\right)^{\frac{1}{1-\alpha}} \ .$$

Inserting this ϵ gives (3.8) with $c = 2^{-\frac{\alpha}{1-\alpha}}\left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right) > 0$.

For any $r \geq 2$ and $\delta > 0$, Lemma 1 gives

$$\varepsilon_m \leq \frac{C}{\delta} e^{\psi(r)\rho\tau} \left(\frac{r}{1+\delta}\right)^{-m} ,$$

which becomes (3.9) upon choosing $r = (m+1)/(\rho\tau)$ and $\delta = 1/m$. □

4. Lanczos-based approximation of functions of matrices. The Arnoldi method unfortunately requires long recurrences for the construction of the Krylov basis. The *Lanczos method* overcomes this difficulty by computing an auxiliary basis $W_m = [w_1, \dots, w_m]$ which spans the Krylov subspace with respect to A^* and w_1. The Lanczos vectors v_j and w_j are constructed such that they satisfy a biorthogonality condition, or block biorthogonality in case of the look-ahead version [7, 28], i.e., $D_m := W_m^* V_m$ is block diagonal. The look-ahead process ensures that D_m is well conditioned when the index m terminates a block, which will be assumed of m in the sequel. The Lanczos vectors can be constructed by short (mostly three-term) recurrences. This results again in a matrix representation (2.1), but now with a block tridiagonal matrix $H_m = D_m^{-1} W_m^* A V_m$. However, unlike the Arnoldi case, neither V_m nor W_m are orthogonal matrices. It is usual to scale the Lanczos vectors to have unit norm, in which case the norms of V_m and W_m are bounded by $\sqrt{m}$. Since H_m is now an oblique projection of A, the numerical range of H_m is in general not contained in $\mathcal{F}(A)$. Variants of Lemma 1, which apply in this situation, are given in the following two lemmas. For the exponential function, Lemmas 7 and 8 lead to essentially the same error bounds as given for the Arnoldi method in Theorems 5 and 6, except for different constants. In Theorems 2, 3, and 4, Arnoldi and Lanczos approximations coincide.

The first lemma works with the *ϵ-pseudospectrum* of A [32], defined by

$$\Lambda_\epsilon(A) = \{\lambda \in \mathbf{C} :\ \|(\lambda I - A)^{-1}\| \geq \epsilon^{-1}\} , \qquad \epsilon > 0.$$

Otherwise, the setting is again the one described before Lemma 1.

Lemma 7. *If $\Lambda_\epsilon(A) \subset E$ and $\Lambda_\gamma(A) \cup \Lambda_\gamma(H_m) \subset G$, then the error of the Lanczos approximation of $f(A)v$ is bounded by* (2.8) *with* $M = \frac{1}{2}(1 + \|V_m\| \cdot \|D_m^{-1} W_m^*\|)\ell(\partial E)/(\epsilon\gamma)$.

Proof. The proof modifies the proof of Lemma 1. For the Lanczos process we have $D_m^{-1} W_m^* A V_m = H_m$ and $D_m^{-1} W_m^* V_m = I$, and therefore

$$\left[(\lambda I - A)^{-1} - V_m(\lambda I - H_m)^{-1} D_m^{-1} W_m^*\right](\lambda I - A) V_m = 0 .$$

Noting $e_1 = D_m^{-1} W_m^* v$, we thus obtain

$$\begin{aligned}(\lambda I - A)^{-1} v - V_m(\lambda I - H_m)^{-1} e_1 \\ = \left[(\lambda I - A)^{-1} - V_m(\lambda I - H_m)^{-1} D_m^{-1} W_m^*\right] p_m(A) v\end{aligned} \tag{4.1}$$

for every polynomial p_m of degree $\leq m$ with $p_m(\lambda) = 1$. By assumption we have that the norms of both $(\lambda I - A)^{-1}$ and $(\lambda I - H_m)^{-1}$ are bounded by γ^{-1} for $\lambda \in \Gamma$. Using further $\|p_m(A)\| \leq \ell(\partial E)/(2\pi\epsilon) \cdot \max_{z \in E} |p_m(z)|$ leads to

$$\begin{aligned}&\|(\lambda I - A)^{-1} v - V_m(\lambda I - H_m)^{-1} e_1\| \\ &\quad \leq \frac{\ell(\partial E)}{2\epsilon\gamma}(1 + \|V_m\| \cdot \|D_m^{-1} W_m^*\|) \cdot |\phi(\lambda)|^{-m},\end{aligned} \tag{4.2}$$

which in turn yields the estimate stated in the lemma. ◻

For a diagonalizable matrix A we let

$$\kappa_e(A) = \|X\| \cdot \|X^{-1}\|,$$

where X is the matrix that contains the eigenvectors of A in its columns.

The following lemma involves only the spectrum $\Lambda(A)$ of A and uses once more the setting of Lemma 1.

Lemma 8. *Let A be diagonalizable and assume that $\Lambda(A) \subset E$, $\Lambda(H_m) \subset G$, and $\|(\lambda I - H_m)^{-1}\| \le \gamma^{-1} < \infty$ for $\lambda \in \Gamma$. Then the Lanczos approximation of $f(A)v$ satisfies* (2.8) *with $M = 3\kappa_e(A)\left(\delta^{-1} + \|V_m\| \cdot \|D_m^{-1}W_m^*\|\,\gamma^{-1}\right)$, where δ is the minimal distance between $\Lambda(A)$ and Γ.*

Proof. The result follows from (4.1) along the lines of parts (c) and (d) of the proof of Lemma 1. ◻

Remarks. (a) It is known that in generic situations, extreme eigenvalues of A are well approximated by those of H_m for sufficiently large m [34]. For a contour Γ that is bounded away from $\Lambda(A)$, one can thus expect that usually $\|(\lambda I - H_m)^{-1}\|$ is uniformly bounded along Γ.

(b) Lemmas 7 and 8 apply also to the Arnoldi method, where $D_m = I$, $W_m = V_m$, and $\|V_m\| = 1$.

(c) The convexity assumption about E can be removed at the price of a larger factor M. For E a continuum containing more than one point, one can use instead of inequality (2.13) the estimate in the lemma on pp. 107f. in Volume III of [19].

The proofs of Lemmas 1, 7, and 8 provide error bounds for iterative methods for the solution of linear systems of equations whose iterates are defined by a Galerkin condition (2.3). This gives new error bounds for the biconjugate gradient method, where the Krylov basis is constructed via the Lanczos process, and for the full orthogonalization method, which is based on the Arnoldi process. The proofs can be extended to give similar error bounds also for the (quasi-) minimization methods QMR and GMRES, see [13].

5. A class of integration methods for large systems of ODEs. In the numerical integration of very large stiff systems of ordinary differential equations $y' = f(y)$, Krylov subspace methods have been used successfully for the solution of the linear systems of equations arising in fully or linearly implicit integration schemes [11, 2, 25]. These linear systems are of the form $(I - \gamma h A)x = v$, where A is the Jacobian of f evaluated near the current integration point, h is the step size, and γ is a method parameter. The attraction with Krylov subspace methods lies in the fact that they require only the computation of matrix-vector products Aw. When it is convenient, these can be approximated as directional derivatives $Aw \doteq \left(f(y+\epsilon w) - f(y)\right)/\epsilon$, so that the Jacobian A need never be formed explicitly. Our theoretical results as well as computational experiments indicate that Krylov subspace approximations of $e^{\gamma h A}v$ or $\varphi(\gamma h A)v$, with

$$\varphi(z) = (e^z - 1)/z\ ,$$

converge faster than the corresponding iterations for $(I - \gamma hA)^{-1}v$, at least unless a good preconditioner is at hand. This suggests the use of the following class of integration schemes, in which the linear systems arising in a linearly implicit method of Rosenbrock-Wanner type are replaced by multiplication with $\varphi(\gamma hA)$. Starting from $y_0 \approx y(t_0)$, the scheme computes an approximation y_1 of $y(t_0+h)$ via

$$k_i = \varphi(\gamma hA)\left(f(u_i) + hA\sum_{j=1}^{i-1}\gamma_{ij}k_j\right), \qquad i = 1,\dots,s \tag{5.1}$$

$$u_i = y_0 + h\sum_{j=1}^{i-1}\alpha_{ij}k_j \tag{5.2}$$

$$y_1 = y_0 + h\sum_{i=1}^{s}\beta_i k_i\ . \tag{5.3}$$

Here $A = f'(y_0)$, and $\gamma, \gamma_{ij}, \alpha_{ij}, \beta_i$ are the coefficients that determine the method. The internal stages $u_1,\dots,u_s$ are computed one after the other, with one multiplication by $\varphi(\gamma hA)$ and a function evaluation at each stage. The simplest method of this type is the well-known exponentially fitted Euler method

$$y_1 = y_0 + h\varphi(hA)f(y_0)\ , \tag{5.4}$$

which is of order 2 and exact for linear differential equations $y' = Ay + b$ with constant A and b. It appears well suited as a basis for Richardson extrapolation. Here is another example of such a method:

Theorem 9. *The two-stage methods with coefficients $\gamma = 1/2$, $\alpha_{21} = \alpha$ (a free parameter), $\gamma_{21} = \frac{3}{4}\alpha^2 - \alpha$, $\beta_1 = 1 - 1/(3\alpha^2)$, $\beta_2 = 1/(3\alpha^2)$ are of order 3. For arbitrary step sizes, they provide the exact solution for every linear system of differential equations $y' = Ay + b$ with constant matrix A and constant inhomogeneity b.*

Proof. Taylor expansion in h of the exact and the numerical solutions shows that the order conditions up to order 3, which correspond to the elementary differentials $f, f'f, f''(f,f), f'f'f$, are given by

$$\sum_i \beta_i = 1 \tag{5.5}$$

$$\sum_{i,j}\beta_i(\alpha_{ij}+\gamma_{ij}) = \tfrac{1}{2}(1-\gamma) \tag{5.6}$$

$$\sum_i \beta_i(\sum_j \alpha_{ij})^2 = \tfrac{1}{3} \tag{5.7}$$

$$\sum_{i,j,k}\beta_i(\alpha_{ij}+\gamma_{ij})(\alpha_{jk}+\gamma_{jk}) = \tfrac{1}{3}(\tfrac{1}{2}-\gamma)(1-\gamma)\ . \tag{5.8}$$

Here all sums extend from 1 to s, and we have set $\alpha_{ij} = \gamma_{ij} = 0$ for $i \le j$. Cf. with the order conditions for Rosenbrock methods in [12], p.116, which differ from the present order conditions only in the right-hand side polynomials in γ.

For $\gamma = 1/2$, the right-hand side of the last order condition vanishes, and hence this condition is automatically satisfied for every two-stage method with $\gamma = 1/2$. With $\alpha_{21} = \alpha$ as a free parameter, the remaining three equations yield the stated method coefficients. Direct calculation shows that the method applied to $y' = Ay + b$, $y(t_0) = y_0$ gives

$$y_1 = e^{hA} y_0 + h\varphi(hA)b = y(t_0 + h) \ ,$$

which is the claimed property. $\square$

Remarks. (a) With $\alpha = 3/4$, the method satisfies the order condition $\beta_2 \alpha_{21}^3 = 1/4$, which corresponds to the fourth-order elementary differential $f'''(f, f, f)$. The order conditions corresponding to $f'f'f'f$ and $f'f''(f, f)$ are satisfied independently of α, so that the order condition corresponding to $f''(f, f'f)$ is then the only fourth-order condition that remains violated.

(b) For non-autonomous problems $y' = f(t, y)$, it is useful to rewrite the equation in autonomous form by adding the trivial equation $t' = 1$ and taking the Jacobian

$$\widetilde{A} = \begin{pmatrix} 0 & 0 \\ f_t & f_y \end{pmatrix} \ .$$

In particular, the method is then exact for every linear equation of the form $y' = Ay + b + tc$, since this is rewritten as

$$\begin{pmatrix} t \\ y \end{pmatrix}' = \begin{pmatrix} 0 & 0 \\ c & A \end{pmatrix} \begin{pmatrix} t \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ b \end{pmatrix} \ ,$$

which is again a linear system with constant inhomogeneity.

An efficient implementation and higher-order methods are currently under investigation.

Note added in the revised version. After finishing this paper we learned that Druskin and Knizhnerman [3, 4] previously obtained an estimate similar to (3.5) for the symmetric case, using a different proof. They give the asymptotic estimate

$$\varepsilon_m \le \left[\sqrt{2\pi} + O\left(\frac{m}{a}\right) \right] \cdot \frac{\sqrt{a}}{m} \cdot \exp\left[-\frac{m^2}{2a} + O\left(\frac{m^4}{a^3}\right) \right] \ , \qquad m \le a \ ,$$

with $a = 2\rho\tau$, which they prove using the Chebyshev series expansion of the exponential function. In an extension of this technique to the non-Hermitian case, Knizhnerman [14] derived error bounds in terms of Faber series for the Arnoldi method (2.7). He showed

$$\varepsilon_m \le \text{const} \sum_{k=m}^{\infty} |f_k| k^{\alpha}, \tag{5.9}$$

where f_k are the Faber series coefficients of f and the exponent α depends on the numerical range of A. As one referee emphasizes, the Faber series approach could be put to similar use as our Lemma 1. In fact, Leonid Knizhnerman showed to us in a personal communication how it would become possible to derive a result of the type of our Theorem 6 using (5.9). Our approach via Lemma 1 makes it more obvious to see how the geometry of the numerical range comes into play. An example similar to that after Theorem 5 is given in [15, §3]. We thank Anne Greenbaum and two referees for pointing out these references and Leonid Knizhnerman for providing a commented version of the Russian paper [14]. Error bounds via Chebyshev and Faber series, for the related problem of approximating matrix functions by methods that generalize semi-iterative methods for linear systems, were given by Tal-Ezer [29, 30, 31].

Acknowledgement. We are grateful to Peter Leinen and Harry Yserentant for providing the initial motivation for this work.

REFERENCES

[1] F.F. Bonsall and J. Duncan, *Numerical Ranges of Operators on Normed Spaces and of Elements of Normed Algebras*, Cambridge Univ. Press, Cambridge, 1971.

[2] P.N. Brown and A.C. Hindmarsh, *Matrix-free methods for stiff systems of ODEs*, SIAM J. Numer. Anal. 23 (1986), 610-638.

[3] V.L. Druskin and L.A. Knizhnerman, *Two polynomial methods of calculating functions of symmetric matrices*, U.S.S.R Comput. Maths. Math. Phys. 29 (1989), 112-121.

[4] V.L. Druskin and L.A. Knizhnerman, *Krylov subspace approximations of eigenpairs and matrix functions in exact and computer arithmetic*, Numer. Lin. Alg. Appl. 2 (1995), 205–217.

[5] E.P. Doolan, J.J.H. Miller, and W.H.A. Schilders, *Uniform Numerical Methods for Problems with Initial and Boundary Layers*, Boole Press, Dublin, 1980.

[6] M. Eiermann, *On semiiterative methods generated by Faber polynomials*, Numer. Math. 56 (1989), 139–156.

[7] R.W. Freund, M.H. Gutknecht, and N.M. Nachtigal, *An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices*, SIAM J. Sci. Comput. 14 (1993), 137–158.

[8] A. Friedli, *Verallgemeinerte Runge-Kutta Verfahren zur Lösung steifer Differentialgleichungen*, in Springer Lecture Notes in Mathematics 631, Springer-Verlag, Berlin, 1978, 35–50.

[9] R.A. Friesner, L.S. Tuckerman, B.C. Dornblaser, and T.V. Russo, *A method of exponential propagation of large systems of stiff nonlinear differential equations*, J. Sci. Comp. 4 (1989), 327–354.

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

[11] C.W. Gear and Y. Saad, *Iterative solution of linear equations in ODE codes*, SIAM J. Sci. Stat. Comp. 4 (1983), 583–601.

[12] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II*, Springer-Verlag, Berlin, 1991.

[13] M. Hochbruck and Ch. Lubich, *Error analysis of Krylov methods in a nutshell*, SIAM J. Sci. Comput., to appear.

[14] L.A. Knizhnerman, *Computation of functions of unsymmetric matrices by means of Arnoldi's method*, J. Comput. Math. and Math. Phys., 1991, v. 31, N 1, p. 5-16 (pages in the Russian issue).

[15] L.A. Knizhnerman, *Error bounds in Arnoldi's method: The case of a normal matrix*, Comput. Maths. Math. Phys. 32 (1992), 1199–1211.

[16] T. Kövari and Ch. Pommerenke, *On Faber polynomials and Faber expansions*, Math. Zeitschr. 99 (1967), 193–206.

[17] R. Kosloff, *Propagation methods for quantum molecular dynamics*, Annu. Rev. Phys. Chem. 45 (1994), 145–178.

[18] J.D. Lawson, *Generalized Runge-Kutta processes for stable systems with large Lipschitz constants*, SIAM J. Numer. Anal. 4 (1967), 372–380.

[19] A.I. Markushevich, *Theory of Functions of a Complex Variable*, Chelsea Publishing Company, New York, 2nd ed., 1977.

[20] A. Nauts and R.E. Wyatt, *New approach to many-state quantum dynamics: The recursive-residue-generation method*, Phys. Rev. Lett., 51 (1983), 2238–2241.

[21] O. Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.

[22] T.J. Park and J.C. Light, *Unitary quantum time evolution by iterative Lanczos reduction*, J. Chem. Phys. 85 (1986), 5870–5876.

[23] Y. Saad, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comp 37 (1981), 105–126.

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

[25] B.A. Schmitt and R. Weiner, *Matrix-free W-methods using a multiple Arnoldi iteration*, Appl. Numer. Math. 18 (1995), 307–320.

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

[27] K. Strehmel and R. Weiner, *Linear-implizite Runge-Kutta Methoden und ihre Anwendung*, Teubner, Stuttgart-Leipzig, 1992.

[28] D.R. Taylor, *Analysis of the look-ahead Lanczos algorithm*, Ph.D. Dissertation, University of California at Berkeley, 1982.

[29] H. Tal-Ezer, *Spectral methods in time for hyperbolic problems*, SIAM J. Numer. Anal. 23 (1986), 11–26.

[30] H. Tal-Ezer, *Spectral methods in time for parabolic problems*, SIAM J. Numer. Anal. 26 (1989), 1–11.

[31] H. Tal-Ezer, *Polynomial approximation of functions of matrices and applications*, J. Sci. Comput. 4 (1989), 25–60.

[32] L.N. Trefethen, *Pseudospectra of matrices*, in D.F. Griffiths and G.A. Watson, eds., Numerical Analysis 1991, Longman, Harlow, UK, 1992.

[33] H.A. van der Vorst, *An iterative solution method for solving $f(A)x = b$, using Krylov subspace information obtained for the symmetric positive definite matrix A*, J. Comput. Appl. Math. 18 (1987), 249–263.

[34] Q. Ye, *A convergence analysis for nonsymmetric Lanczos algorithms*, Math. Comp. 56 (1991), 677–691.