SIAM J. NUMER. ANAL.
Vol. 44, No. 6, pp. 2481–2504

A RESTARTED KRYLOV SUBSPACE METHOD FOR THE EVALUATION OF MATRIX FUNCTIONS*

MICHAEL EIERMANN† AND OLIVER G. ERNST†

Abstract. We show how the Arnoldi algorithm for approximating a function of a matrix times a vector can be restarted in a manner analogous to restarted Krylov subspace methods for solving linear systems of equations. The resulting restarted algorithm reduces to other known algorithms for the reciprocal and the exponential functions. We further show that the restarted algorithm inherits the superlinear convergence property of its unrestarted counterpart for entire functions and present the results of numerical experiments.

Key words. matrix function, Krylov subspace approximation, Krylov projection method, restarted Krylov subspace method, linear system of equations, initial value problem

AMS subject classifications. 65F10, 65F99, 65M20

DOI. 10.1137/050633846

1. Introduction.

The evaluation of

$$f(A)\boldsymbol{b}, \qquad \text{where } A \in \mathbb{C}^{n\times n}, \boldsymbol{b} \in \mathbb{C}^n, \tag{1.1}$$

and $f : \mathbb{C} \supset D \to \mathbb{C}$ is a function for which $f(A)$ is defined, is a common computational task. Besides the solution of linear systems of equations, which involves the reciprocal function $f(\lambda) = 1/\lambda$, by far the most important application is the time evolution of a system under a linear operator, in which case $f(\lambda) = f_t(\lambda) = e^{t\lambda}$ and time acts as a parameter t. Other applications involving differential equations require the evaluation of (1.1) for the square root and trigonometric functions (see [8, 1]). Further applications include identification problems for semigroups involving the logarithm (see, e.g., [29]) and lattice quantum chromodynamics simulations requiring the evaluation of the matrix sign function (see [34] and the references therein).

In many of the applications mentioned above the matrix A is large and sparse or structured, typically resulting from discretization of an infinite-dimensional operator. In this case evaluating (1.1) by first computing $f(A)$ is usually unfeasible, so that most of the algorithms for the latter task (see, e.g., [18, 5]) cannot be used. The standard approach for approximating (1.1) directly is based on a Krylov subspace of A with initial vector $\boldsymbol{b}$ [8, 9, 28, 14, 17, 4, 19]. The advantage of this approach is that it requires A only for computing matrix-vector products and that, for smooth functions such as the exponential, it converges superlinearly [8, 28, 31, 17].

One shortcoming of the Krylov subspace approximation, however, lies in the fact that computing an approximation of (1.1) from a Krylov subspace $\mathcal{K}_m(A, \boldsymbol{b})$ of dimension m involves all the basis vectors of $\mathcal{K}_m(A, \boldsymbol{b})$, and hence these need to be stored. Memory constraints therefore often limit the size of the problem that can be solved, which is an issue especially when A is the discrete representation of a partial differential operator in three space dimensions. When A is Hermitian, the Hermitian Lanczos process allows the basis of $\mathcal{K}_m(A, \boldsymbol{b})$ to be constructed by a three-term recurrence. When solving linear systems of equations, this recurrence for the basis vectors

*Received by the editors June 17, 2005; accepted for publication (in revised form) June 13, 2006; published electronically December 1, 2006.
http://www.siam.org/journals/sinum/44-6/63384.html

†Institut für Numerische Mathematik und Optimierung, Technische Universität Bergakademie Freiberg, D-09596 Freiberg, Germany (eiermann@math.tu-freiberg.de, ernst@math.tu-freiberg.de).

immediately translates to efficient update formulas for the approximation. This, however, is a consequence of the simple form of the reciprocal function and such update formulas are not available for general (nonrational) functions.

When solving non-Hermitian linear systems of equations by Krylov subspace approximation, a common remedy is to limit storage requirements by restarting the algorithm each time the Krylov space has reached a certain maximal dimension [26, 10]. The subject of this work is the extension of this restarting approach to general functions. How such a generalization may be accomplished is not immediately obvious, since the restarting approach for linear systems is based on solving successive residual equations to obtain corrections to the most recent approximation. The availability of a residual, however, is another property specific to problems where $f(A)\boldsymbol{b}$ solves an (algebraic or differential) equation.

The remainder of this paper is organized as follows: Section 2 recalls the definition and properties of matrix functions and their approximation in Krylov spaces, emphasizing the role of Hermite interpolation, and closes with an error representation formula for Krylov subspace approximations. In section 3 we introduce a new restarted Krylov subspace algorithm for the approximation of (1.1) for general functions f. We derive two mathematically equivalent formulations of the restarted algorithm, the second of which, while slightly more expensive, was found to be more stable in the presence of rounding errors.

In section 4 we show that, for the reciprocal and exponential functions, our restarted method reduces to the restarted full orthogonalization method (FOM; see [27]) and is closely related to an algorithm by Celledoni and Moret [4], respectively. We further establish that, for entire functions of order one (such as the exponential function), the superlinear convergence property of the Arnoldi/Lanczos approximation of (1.1) is retained by our restarted method. In section 5 we demonstrate the performance of the restarted method for several test problems.

2. Matrix functions and their Krylov subspace approximation. In this section we fix notation, provide some background material on functions of matrices and their approximation using Krylov subspaces, highlight the connection with Hermite interpolation, and derive a new representation formula for the error of Krylov subspace approximations of $f(A)\boldsymbol{b}$.

2.1. Functions of matrices. We recall the definition of functions of matrices (as given, e.g., in Gantmacher [15, Chapter 5]): Let $\Lambda(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ denote the k distinct eigenvalues of $A \in \mathbb{C}^{n\times n}$ and let the minimal polynomial of A be given by

$$m_A(\lambda) = \prod_{j=1}^{k} (\lambda - \lambda_j)^{n_j} \in \mathcal{P}_K, \qquad \text{where } K = \sum_{j=1}^{k} n_j.$$

Given a complex-valued function f, the matrix $f(A)$ is defined if $f^{(r)}(\lambda_j)$ exists for $r = 0, 1, \ldots, n_j - 1$; $j = 1, 2, \ldots, k$. In this case $f(A) := q_{f,A}(A)$, where $q_{f,A} \in \mathcal{P}_{K-1}$ denotes the unique polynomial of degree at most $K-1$ which satisfies the K Hermite interpolation conditions

$$q_{f,A}^{(r)}(\lambda_j) = f^{(r)}(\lambda_j), \qquad r = 0, 1, \ldots, n_j - 1,\ j = 1, 2, \ldots, k. \tag{2.1}$$

In the remainder of the paper, we denote the unique polynomial q which interpolates f in the Hermite sense at a set of nodes $\{\vartheta_j\}_{j=1}^{k}$ with multiplicities n_j by $I_p f \in \mathcal{P}_{K-1}$,

$K = \sum_j n_j$, where $p \in \mathcal{P}_K$ is a (not necessarily monic) nodal polynomial with zeros ϑ_j of multiplicities n_j. In this notation, (2.1) reads

$$q_{f,A} = I_{m_A} f.$$

Our objective is the evaluation of $f(A)\boldsymbol{b}$ rather than $f(A)$, and this can possibly be achieved with polynomials of lower degree than $q_{f,A}$. To this end, let the minimal polynomial of $\boldsymbol{b} \in \mathbb{C}^n$ with respect to A be given by

$$m_{A,\boldsymbol{b}}(\lambda) = \prod_{j=1}^{\ell} (\lambda - \lambda_j)^{m_j} \in \mathcal{P}_L, \qquad \text{where } L = L(A,\boldsymbol{b}) = \sum_{j=1}^{\ell} m_j. \tag{2.2}$$

PROPOSITION 2.1. *Given a function f, a matrix $A \in \mathbb{C}^{n\times n}$ such that $f(A)$ is defined, and a vector $\boldsymbol{b} \in \mathbb{C}^n$ whose minimal polynomial with respect to A is given by* (2.2)*, there holds $f(A)\boldsymbol{b} = q_{f,A,\boldsymbol{b}}(A)\boldsymbol{b}$, where $q_{f,A,\boldsymbol{b}} := I_{m_{A,\boldsymbol{b}}} f \in \mathcal{P}_{L-1}$ denotes the unique Hermite interpolating polynomial determined by the conditions*

$$q_{f,A,\boldsymbol{b}}^{(r)}(\lambda_j) = f^{(r)}(\lambda_j), \qquad r = 0, 1, \ldots, m_j - 1,\ j = 1, 2, \ldots, \ell.$$

2.2. Krylov subspace approximations. We recall the definition of the mth Krylov subspace of $A \in \mathbb{C}^{n\times n}$ and $\mathbf{0} \neq \boldsymbol{b} \in \mathbb{C}^n$ given by

$$\mathcal{K}_m(A,\boldsymbol{b}) := \operatorname{span}\{\boldsymbol{b}, A\boldsymbol{b}, ..., A^{m-1}\boldsymbol{b}\} = \{q(A)\boldsymbol{b} \,:\, q \in \mathcal{P}_{m-1}\}.$$

By Proposition 2.1, $f(A)\boldsymbol{b}$ lies in $\mathcal{K}_L(A,\boldsymbol{b})$. The index $L = L(A,\boldsymbol{b}) \in \mathbb{N}$ (cf. (2.2)) is the smallest number for which $\mathcal{K}_L(A,\boldsymbol{b}) = \mathcal{K}_{L+1}(A,\boldsymbol{b})$. Note that for certain functions such as $f(\lambda) = 1/\lambda$, we have $f(A)\boldsymbol{b} \in \mathcal{K}_L(A,\boldsymbol{b}) \setminus \mathcal{K}_{L-1}(A,\boldsymbol{b})$; in general, however, $f(A)\boldsymbol{b}$ may lie in a space $\mathcal{K}_m(A,\boldsymbol{b})$ with $m < L$.[1]

In what follows, we consider a sequence of approximations $\boldsymbol{y}_m := q(A)\boldsymbol{b} \in \mathcal{K}_m(A,\boldsymbol{b})$ to $f(A)\boldsymbol{b}$ with polynomials $q \in \mathcal{P}_{m-1}$ which in some sense approximate f. The most popular of these approaches (see [28, 14, 17]), to which we shall refer as the *Arnoldi approximation*, is based on the Arnoldi decomposition of $\mathcal{K}_m(A,\boldsymbol{b})$,

$$AV_m = V_{m+1}\widetilde{H}_m = V_m H_m + \eta_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^{\top}. \tag{2.3}$$

Here, the columns of $V_m = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m]$ form an orthonormal basis of $\mathcal{K}_m(A,\boldsymbol{b})$ with $\boldsymbol{v}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|$, $\widetilde{H}_m = [\eta_{j,\ell}] \in \mathbb{C}^{(m+1)\times m}$ as well as $H_m := [I_m, \mathbf{0}]\,\widetilde{H}_m \in \mathbb{C}^{m\times m}$ are unreduced upper Hessenberg matrices, and $\boldsymbol{e}_m \in \mathbb{R}^m$ denotes the mth unit coordinate vector. The Arnoldi approximation to $f(A)\boldsymbol{b}$ is then defined by

$$\boldsymbol{f}_m := \beta V_m f(H_m)\boldsymbol{e}_1, \qquad \text{where } \beta = \|\boldsymbol{b}\|.$$

The rationale behind this approximation is that H_m represents the compression of A onto $\mathcal{K}_m(A,\boldsymbol{b})$ with respect to the basis V_m and that $\boldsymbol{b} = \beta V_m \boldsymbol{e}_1$.

The non-Hermitian (or two-sided) Lanczos algorithm is another procedure for generating a decomposition of the form (2.3). In that case the columns of V_m still form a basis of $\mathcal{K}_m(A,\boldsymbol{b})$, albeit one that is, in general, not orthogonal, and the upper Hessenberg matrices $\widetilde{H}_m$ are tridiagonal (or block tridiagonal if a look-ahead

[1] For the exponential function it was shown in [28, Theorem 3.6] that $e^{tA}\boldsymbol{b} \in \mathcal{K}_m(A,\boldsymbol{b})$ for all $t \in \mathbb{R}$ if and only if $m \geq L$.

technique is employed). The associated approximation to $f(A)\boldsymbol{b}$ is again defined by $\boldsymbol{f}_m := \beta V_m f(H_m)\boldsymbol{e}_1$ (see, e.g., [14, 28, 17]). Both approximations $\boldsymbol{f}_m$, based either on the Arnoldi or Lanczos decomposition, result from an interpolation procedure: If $q \in \mathcal{P}_{m-1}$ denotes the polynomial which interpolates f in the Hermite sense on the spectrum of H_m (counting multiplicities), then

$$\boldsymbol{f}_m = \beta V_m f(H_m)\boldsymbol{e}_1 = \beta V_m q(H_m)\boldsymbol{e}_1 = q(A)\boldsymbol{b}, \qquad m = 1, 2, \ldots, L$$

(see [28, Theorem 3.3]).

For later applications, next we show that similar results hold true for more general decompositions of $\mathcal{K}_m(A, \boldsymbol{b})$. To this end, we introduce a sequence of ascending (not necessarily orthonormal) basis vectors $\{\boldsymbol{w}_m\}_{m=1}^{L}$ such that

$$\mathcal{K}_m(A, \boldsymbol{b}) = \operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m\}, \qquad m = 1, 2, \ldots, L. \tag{2.4}$$

As is well known, there exists a unique unreduced upper Hessenberg matrix $H = [\eta_{j,m}] \in \mathbb{C}^{L\times L}$ such that, with $W := [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_L] \in \mathbb{C}^{n\times L}$, there holds $AW = WH$ and, for $m = 1, 2, \ldots, L-1$, we have

$$AW_m = W_{m+1}\widetilde{H}_m = W_m H_m + \eta_{m+1,m}\boldsymbol{w}_{m+1}\boldsymbol{e}_m^{\top}, \tag{2.5}$$

where $\widetilde{H}_m$ is the $(m+1)\times m$ leading submatrix of H, $H_m := [I_m, \boldsymbol{0}]\widetilde{H}_m$, and $W_m = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m]$. We shall refer to (2.5) as an *Arnoldi-like decomposition*[2] to distinguish it from a proper Arnoldi decomposition (2.3). We shall require the following lemma, which is a simple generalization of the corresponding result for (proper) Arnoldi decompositions (cf. [28, 23]).

LEMMA 2.2. *For any polynomial $q(\lambda) = \alpha_m\lambda^m + \cdots + \alpha_1\lambda + \alpha_0 \in \mathcal{P}_m$, the vector $q(A)\boldsymbol{b}$ may be represented in terms of the Arnoldi-like decomposition* (2.5) *as*

$$q(A)\boldsymbol{b} = \begin{cases} \beta\bigl[W_m q(H_m)\boldsymbol{e}_1 + \alpha_m\gamma_m\boldsymbol{w}_{m+1}\bigr], & m < L, \\ \beta W_L q(H_L)\boldsymbol{e}_1, & m \geq L, \end{cases} \tag{2.6}$$

where $\gamma_m := \prod_{j=1}^{m}\eta_{j+1,j}$ and $\beta\boldsymbol{w}_1 = \boldsymbol{b}$. In particular, for any $q \in \mathcal{P}_{m-1}$ there holds $q(A)\boldsymbol{b} = \beta W_m q(H_m)\boldsymbol{e}_1$.

The proof follows by verifying the assertion for monomials, taking account of the sparsity pattern of powers of a Hessenberg matrix (see, e.g., [12]).

We next introduce polynomial notation to describe Krylov subspaces. To each vector $\boldsymbol{w}_m$ of the nested basis (2.4) there corresponds a unique polynomial $w_{m-1} \in \mathcal{P}_{m-1}$ such that $\boldsymbol{w}_m = w_{m-1}(A)\boldsymbol{b}$. Via this correspondence, the Arnoldi-like recurrence (2.5) becomes

$$\begin{aligned}\lambda[w_0(\lambda), w_1(\lambda), \ldots, w_{m-1}(\lambda)] = {} & [w_0(\lambda), w_1(\lambda), \ldots, w_{m-1}(\lambda)]H_m \\ & + \eta_{m+1,m}[0, 0, \ldots, 0, w_m(\lambda)].\end{aligned} \tag{2.7}$$

From this equation it is evident that each zero of w_m is an eigenvalue of H_m. Moreover, by differentiating (2.7), one observes that zeros of multiplicity ℓ are eigenvalues of H_m with Jordan blocks of dimension ℓ. Since H_m is an unreduced Hessenberg matrix and

[2]We mention that the related term *Krylov decomposition* introduced by Stewart in [32] refers to a decomposition of the form (2.5) without the restriction that the basis be ascending and, consequently, to a matrix H which is not necessarily Hessenberg.

hence nonderogatory, we conclude that the zeros of w_m coincide with the eigenvalues of H_m counting multiplicity.

LEMMA 2.3. *Let H_m be the unreduced upper Hessenberg matrix in* (2.5) *and* (2.7) *and let f be a function such that $f(H_m)$ is defined. Then a polynomial $q_{m-1} \in \mathcal{P}_{m-1}$ satisfies*

$$q_{m-1}(H_m) = f(H_m)$$

if and only if $q_{m-1} = I_{w_m} f$, i.e., if q_{m-1} interpolates f in the Hermite sense at the eigenvalues of H_m.

Proof. The proof follows directly from the definition of $f(H_m)$ and the fact that the zeros of w_m are the eigenvalues of H_m with multiplicity. □

We summarize the contents of Lemmata 2.2 and 2.3 as follows.

THEOREM 2.4. *Given the Arnoldi-like decomposition* (2.5) *and a function f such that $f(A)$ as well as $f(H_m)$ are defined, we denote by $q \in \mathcal{P}_{m-1}$ the unique polynomial which interpolates f at the eigenvalues of H_m. Then there holds*

$$\boldsymbol{f}_m := \beta V_m f(H_m)\boldsymbol{e}_1 = \beta V_m q(H_m)\boldsymbol{e}_1 = q(A)\boldsymbol{b}. \tag{2.8}$$

We shall refer to (2.8) as *the Krylov subspace approximation of $f(A)\boldsymbol{b}$ associated with the Arnoldi-like decomposition* (2.5). Note that (2.8) is merely a computational device for generating the Krylov subspace approximation of $f(A)\boldsymbol{b}$ without explicitly carrying out the interpolation process. This is an advantage whenever $f(H_m)\boldsymbol{e}_1$ for $m \ll n$ can be evaluated efficiently.

Remark 2.5. We also point out the following—somewhat academic—detail regarding finite termination: While Krylov subspace approximations $q(A)\boldsymbol{b}$ are defined for polynomials q of any degree, Arnoldi-like decompositions, and hence (2.8), are only available for $1 \le m \le L = L(A, \boldsymbol{b})$. At index $m = L$, the characteristic polynomial of $H_L = H$ coincides with the minimal polynomial $m_{A,\boldsymbol{b}}$ of $\boldsymbol{b}$ with respect to A (see (2.2)). In view of (2.8) and Proposition 2.1, we then have $\boldsymbol{f}_L = f(A)\boldsymbol{b}$. In this sense, Krylov subspace approximations of the form (2.8) respect the spectral distribution of A relevant for $\boldsymbol{b}$ and, in exact arithmetic, possess the finite termination property. This is in contrast to other approaches such as those based on Chebyshev or Faber expansions (see below).

Besides those generated by the Arnoldi or Lanczos processes, any ascending basis $\{\boldsymbol{w}_m\}_{m=1}^L$ of $\mathcal{K}_L(A, \boldsymbol{b})$ or, equivalently, any sequence of polynomials $\{w_{m-1}\}_{m=1}^L$ of exact degree $m-1$ may be used in the Arnoldi-like decomposition (2.5) or its polynomial counterpart (2.7), provided a means for obtaining the matrix $\widetilde{H}_L$ of recurrence coefficients is available. One such example is the sequence of kernel/quasi-kernel polynomials associated with the Arnoldi/Lanczos decomposition (see [13]), where the corresponding Hessenberg matrix is easily constructed from that of the original decomposition. Approximations based on quasi-kernel polynomials are discussed in [20]. Yet another approach—one which emphasizes the interpolation aspect of the Krylov subspace approximation—fixes a sequence of nodes

$$\begin{matrix} \vartheta_1^{(1)} & & & \\ \vartheta_1^{(2)} & \vartheta_2^{(2)} & & \\ \vartheta_1^{(3)} & \vartheta_2^{(3)} & \vartheta_3^{(3)} & \\ \vdots & \vdots & & \ddots \end{matrix}$$

and chooses the basis vectors $\boldsymbol{w}_m = w_{m-1}(A)\boldsymbol{b}$ as the associated nodal polynomials

$$w_{m-1}(\lambda) = \omega_{m-1}(\lambda - \vartheta_1^{(m)})(\lambda - \vartheta_2^{(m)}) \dots (\lambda - \vartheta_m^{(m)}), \qquad \omega_m \neq 0.$$

One possible choice of such a node sequence is the zeros of Chebyshev polynomials, in which case the nested basis vectors correspond to Chebyshev polynomials. Other choices of node sequences are explored in [19, 22, 20]. Note that, in view of Remark 2.5, such a basis choice, which is independent of A and $\boldsymbol{b}$, will generally destroy the finite termination property.

We also point out that, when $f(A)$ is defined, this need not be so for $f(H_m)$ for $m < L$. For the Arnoldi approximation, a sufficient condition ensuring this is that f, as a scalar function, be analytic in a neighborhood of the field of values of A. As a case in point, consider the FOM for solving a nonsingular system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$. The solution is $f(A)\boldsymbol{b}$ with $f(\lambda) = 1/\lambda$ and, if the initial approximation is $\boldsymbol{x}_0 = \boldsymbol{0}$, the mth FOM iterate is simply the Arnoldi approximation $\boldsymbol{f}_m = \beta V_m H_m^{-1} \boldsymbol{e}_1$. There are well-known examples [3] in which $f(A)$, i.e., A^{-1}, is defined but for which H_m is singular for one or more of the indices $m = 1, \dots, L-1$, a phenomenon sometimes called a *Galerkin breakdown.*

2.3. An error representation. We conclude this section with a representation of the error of the Krylov subspace approximation of $f(A)\boldsymbol{b}$ based on any Arnoldi-like decomposition or, equivalently, any interpolatory approximation. We shall need the following notation: Given a function f and a set of nodes $\vartheta_1, \dots, \vartheta_m$ with associated nodal polynomial

$$p(\lambda) = (\lambda - \vartheta_1)(\lambda - \vartheta_2) \cdots (\lambda - \vartheta_m), \tag{2.9}$$

we denote the mth order divided difference of f with respect to the nodes $\{\vartheta_j\}_{j=1}^m$ by[3]

$$\Delta_p f := \frac{f - I_p f}{p}. \tag{2.10}$$

THEOREM 2.6. *Given $A \in \mathbb{C}^{n\times n}$, $\boldsymbol{b} \in \mathbb{C}^n$, and a function f, let* (2.5) *be an Arnoldi-like decomposition of $\mathcal{K}_m(A, \boldsymbol{b})$ and let $w_m \in \mathcal{P}_{m-1}$ be the associated polynomial; cf.* (2.7). *Then there holds*

$$f(A)\boldsymbol{b} - \beta W_m f(H_m)\boldsymbol{e}_1 = \beta\gamma_m [\Delta_{w_m} f](A)\boldsymbol{w}_{m+1} \tag{2.11}$$

with γ_m as in Lemma 2.2.

Proof. We consider first an arbitrary set of nodes $\vartheta_1, \dots, \vartheta_m$ with associated nodal polynomial p as in (2.9). From the definition (2.10), there holds $f(\lambda) = [I_p f](\lambda) + [\Delta_p f](\lambda)p(\lambda)$. Inserting A for λ in this identity and multiplying by $\boldsymbol{b}$, we obtain

$$f(A)\boldsymbol{b} = [I_p f](A)\boldsymbol{b} + [\Delta_p f](A)p(A)\boldsymbol{b}.$$

Since $I_p f \in \mathcal{P}_{m-1}$, Lemma 2.2 yields $[I_p f](A)\boldsymbol{b} = \beta W_m [I_p f](H_m)\boldsymbol{e}_1$ and, since $p \in \mathcal{P}_m$ is monic, $p(A)\boldsymbol{b} = \beta W_m p(H_m)\boldsymbol{e}_1 + \beta\gamma_m \boldsymbol{w}_{m+1}$, giving

$$f(A)\boldsymbol{b} - \beta W_m [I_p f](H_m)\boldsymbol{e}_1 = \beta [\Delta_p f](A)\big(W_m p(H_m)\boldsymbol{e}_1 + \gamma_m \boldsymbol{w}_{m+1}\big).$$

[3]The source of and justification for this notation can be found in [7].

Choosing p as the characteristic polynomial w_m of H_m, it follows that $w_m(H_m) = O$ by the Cayley–Hamilton theorem and, since $I_{w_m} f$ interpolates f at the eigenvalues of H_m, there also holds $[I_{w_m} f](H_m) = f(H_m)$ by Lemma 2.3. □

We interpret (2.11) as follows: Further improvement of a Krylov approximation $\boldsymbol{f}_m = \beta V_m f(H_m)\boldsymbol{e}_1$ could be achieved by approximating the error term

$$f(A)\boldsymbol{b} - \boldsymbol{f}_m = \widetilde{f}(A)\widetilde{\boldsymbol{b}}$$

with $\widetilde{f}(\lambda) := [\Delta_{w_m} f](\lambda)$ and $\widetilde{\boldsymbol{b}} := \beta\gamma_m \boldsymbol{w}_{m+1}$. Note that the modified function $\widetilde{f}$ has the same domain of analyticity as f and the vector $\widetilde{\boldsymbol{b}}$ points in the direction of the last vector in the Arnoldi-like decomposition.

3. A restarted Arnoldi approximation. For the remainder of the paper we shall restrict the discussion to the Arnoldi approximation of $f(A)\boldsymbol{b}$. To set this apart from a general Krylov subspace approximation (2.8) we denote the (orthonormal) Arnoldi basis vectors by $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_L$ and the Arnoldi decomposition by

$$AV_m = V_m H_m + \eta_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^\top, \qquad V_m = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m]$$

(cf. (2.3)). Our results apply to other Krylov subspace approximations with obvious modifications, some of which we shall point out.

3.1. Short recurrences are not enough. Besides the evaluation of $f(H_m)$, the computation of the mth Arnoldi approximation $\boldsymbol{f}_m = \beta V_m f(H_m)\boldsymbol{e}_1$ requires the Arnoldi basis V_m, which consists of m vectors of size n. As a consequence, even if the evaluation of $f(H_m)$ can be accomplished inexpensively, work and storage requirements incurred by V_m make this method impractical for moderate to large values of m. For $f(\lambda) = 1/\lambda$, i.e., when solving linear systems of equations, one can take advantage of the fact that the Arnoldi process reduces to the Hermitian Lanczos process when A is Hermitian. In this case the matrices H_m are Hermitian, hence tridiagonal, and three-term recurrence formulas can be derived for their characteristic polynomials w_m. (The same is true even in the non-Hermitian case when employing the non-Hermitian Lanczos process, possibly with look-ahead techniques.) If we interpolate $f(\lambda) = 1/\lambda$ at the zeros of the mth basis polynomial w_m, the resulting interpolating polynomial $q_{m-1} = I_{w_m} f$ satisfies

$$q_{m-1}(\lambda) = \frac{w_m(0) - w_m(\lambda)}{\lambda w_m(0)}, \tag{3.1}$$

and therefore q_{m-1} and hence also the approximation $\boldsymbol{f}_m$ obey a similar three-term recurrence. The relation (3.1) between the nodal and the interpolation polynomials can therefore be viewed as the basis for the efficiency of the conjugate gradient method and other polynomial acceleration methods such as Chebyshev iteration for solving linear systems of equations.

A relation analogous to (3.1) fails to hold for more complicated (nonrational) functions f such as the exponential function, and therefore short recurrences for the nodal polynomials do not translate into short recurrences for the interpolation polynomials. The computation of $\boldsymbol{f}_m$ therefore necessitates storing the full Arnoldi basis V_m also when A is Hermitian. It is therefore of interest to modify the Arnoldi approximation in a way that allows the construction of successively better approximations of $f(A)\boldsymbol{b}$

based on a sequence of Krylov spaces of small dimension.[4] Such *restarted Krylov subspace methods* are well known for the solution of linear systems of equations; see [26, 10].

3.2. Krylov approximation after Arnoldi restart. Consider two Krylov spaces of order m with Arnoldi decompositions

$$AV_m^{(1)} = V_m^{(1)} H_m^{(1)} + \eta_{m+1,m}^{(1)} \boldsymbol{v}_{m+1}^{(1)} \boldsymbol{e}_m^\top, \tag{3.2a}$$

$$AV_m^{(2)} = V_m^{(2)} H_m^{(2)} + \eta_{m+1,m}^{(2)} \boldsymbol{v}_{m+1}^{(2)} \boldsymbol{e}_m^\top, \tag{3.2b}$$

where $\boldsymbol{v}_1^{(1)} = \boldsymbol{b}/\beta$ and $\boldsymbol{v}_1^{(2)} = \boldsymbol{v}_{m+1}^{(1)}$, i.e., obtained from two cycles of the Arnoldi process applied to A, beginning with initial vector $\boldsymbol{b}$ and restarted after m steps with the last Arnoldi basis vector $\boldsymbol{v}_{m+1}^{(1)}$ from the first cycle. We note that the columns of $W_{2m} := [V_m^{(1)}, V_m^{(2)}]$ form a basis of $\mathcal{K}_{2m}(A, \boldsymbol{b})$, albeit not an orthonormal one, and we may combine the two proper Arnoldi decompositions (3.2) to the Arnoldi-like decomposition

$$AW_{2m} = W_{2m} H_{2m} + \eta_{m+1,m}^{(2)} \boldsymbol{v}_{m+1}^{(2)} \boldsymbol{e}_{2m}^\top, \tag{3.3}$$

where H_{2m} is the Hessenberg matrix

$$H_{2m} := \begin{bmatrix} H_m^{(1)} & O \\ \eta_{m+1,m}^{(1)} \boldsymbol{e}_1 \boldsymbol{e}_m^\top & H_m^{(2)} \end{bmatrix}. \tag{3.4}$$

Remark 3.1. We restart the Arnoldi process with $\boldsymbol{v}_{m+1}^{(1)}$, which is the most natural choice. We could, however, restart with any vector of the form

$$\widehat{\boldsymbol{v}}_{m+1} = V_m^{(1)} \boldsymbol{y} + y_{m+1} \boldsymbol{v}_{m+1}^{(1)} \in \mathcal{K}_{m+1}(A, \boldsymbol{b}) \setminus \mathcal{K}_m(A, \boldsymbol{b})$$

with a coefficient vector $\boldsymbol{y} = [y_1, y_2, \ldots, y_m]^\top \in \mathbb{C}^m$. In this case we must replace $H_m^{(1)}$ in (3.4) by the rank-one modification $H_m^{(1)} - (\eta_{m+1,m}^{(1)}/y_{m+1}) \boldsymbol{y} \boldsymbol{e}_m^\top$ and $\eta_{m+1,m}^{(1)}$ by $\eta_{m+1,m}^{(1)}/y_{m+1}$. It is conceivable that this could be used to emphasize certain directions such as Ritz approximations of certain eigenvectors as is done in popular restarting techniques for linear systems of equations [21] and eigenvalue calculations [30], but we shall not pursue this here.

Our objective is to compute the Krylov subspace approximation associated with (3.3) without reference to $V_m^{(1)}$. The former is defined as

$$\boldsymbol{f}_{2m} = [I_{w_{2m}} f](A)\boldsymbol{b} = \beta W_{2m} [I_{w_{2m}} f](H_{2m}) \boldsymbol{e}_1 = \beta W_{2m} f(H_{2m}) \boldsymbol{e}_1, \tag{3.5}$$

where w_{2m} is the nodal polynomial with zeros $\Lambda(H_m^{(1)}) \cup \Lambda(H_m^{(2)})$ with multiplicity. To evaluate the approximation (3.5), we note that $f(H_{2m})$ is of the form

$$f(H_{2m}) = \begin{bmatrix} f(H_m^{(1)}) & O \\ X_{2,1} & f(H_m^{(2)}) \end{bmatrix}, \qquad X_{2,1} \in \mathbb{C}^{m \times m}, \tag{3.6}$$

[4] Another remedy, well known from Lanczos-based eigenvalue computations (see [24, Chapter 13]), is to discard the basis vectors no longer needed in the recurrence and either recompute these or retrieve them from secondary storage when forming the approximation.

a consequence of the block triangular structure of H_{2m}, whereby (3.5) becomes

$$\boldsymbol{f}_{2m} = \beta V_m^{(1)} f(H_m^{(1)})\boldsymbol{e}_1 + \beta V_m^{(2)} X_{2,1}\boldsymbol{e}_1. \tag{3.7}$$

The first term on the right is the Arnoldi approximation with respect to $\mathcal{K}_m(A, \boldsymbol{b})$. If $X_{2,1}\boldsymbol{e}_1$ were computable, one could discard the basis vectors $V_m^{(1)}$ and use (3.7) to update the Arnoldi approximation, thus yielding the basis for a restarting scheme.

One conceivable approach is to observe that $X_{2,1}$ satisfies the Sylvester equation

$$H_m^{(2)} X_{2,1} - X_{2,1} H_m^{(1)} = \eta_{m+1,m}^{(1)} [f(H_m^{(2)})\boldsymbol{e}_1\boldsymbol{e}_m^\top - \boldsymbol{e}_1\boldsymbol{e}_m^\top f(H_m^{(1)})], \tag{3.8}$$

which follows from comparing the $(2,1)$ blocks of the identity $H_{2m} f(H_{2m}) = f(H_{2m})$ H_{2m}, and one could therefore proceed by solving (3.8). This approach, however, suffers from the shortcoming that the Sylvester equation (3.8) is only well conditioned if the spectra of $H_m^{(1)}$ and $H_m^{(2)}$ are well separated (cf. [16, section 15.3]). Since $H_m^{(1)}$ and $H_m^{(2)}$ are both compressions of the same matrix A, it is to be expected that at least some of their eigenvalues match very closely.

We shall instead derive a computable expression for $X_{2,1}\boldsymbol{e}_1$ directly by way of interpolation.

LEMMA 3.2. *Given two successive Arnoldi decompositions as in* (3.2), *let* $w_m^{(1)}$, $w_m^{(2)}$, *and* w_{2m} *denote the monic nodal polynomials associated with* $\Lambda(H_m^{(1)})$, $\Lambda(H_m^{(2)})$, *and* $\Lambda(H_{2m}) = \Lambda(H_m^{(1)}) \cup \Lambda(H_m^{(2)})$, *respectively, with* H_{2m} *the upper Hessenberg matrix of the combined Arnoldi-like decomposition* (3.3). *Then there holds*

$$[I_{w_{2m}} f](H_{2m})\boldsymbol{e}_1 = \begin{bmatrix} [I_{w_m^{(1)}} f]\big(H_m^{(1)}\big)\boldsymbol{e}_1 \\ \gamma_m^{(1)} [I_{w_m^{(2)}}(\mathbb{\Delta}_{w_m^{(1)}} f)]\big(H_m^{(2)}\big)\boldsymbol{e}_1 \end{bmatrix}, \tag{3.9}$$

where $\gamma_m^{(1)} = \prod_{j=1}^m \eta_{j+1,j}^{(1)}$ *(cf. Lemma* 2.2*).*

Proof. Due to the block triangular structure of H_{2m} as given in (3.6), there holds

$$[I_{w_{2m}} f]\left(\begin{bmatrix} H_m^{(1)} & O \\ \eta_{m+1,m}^{(1)}\boldsymbol{e}_1\boldsymbol{e}_m^\top & H_m^{(2)} \end{bmatrix}\right) = \begin{bmatrix} [I_{w_{2m}} f]\big(H_m^{(1)}\big) & O \\ X_{2,1} & [I_{w_{2m}} f]\big(H_m^{(2)}\big) \end{bmatrix} \tag{3.10}$$

with $X_{2,1}$ as in (3.6). Next, we establish the polynomial identity

$$[I_{w_{2m}} f] = I_{w_m^{(1)}} f + I_{w_m^{(2)}}(\mathbb{\Delta}_{w_m^{(1)}} f)\, w_m^{(1)}, \tag{3.11}$$

which can be seen by noting that both polynomials have the same degree $2m-1$ and interpolate f in the Hermite sense at the nodes $\Lambda(H_m^{(1)}) \cup \Lambda(H_m^{(1)})$. For nodes $\vartheta \in \Lambda(H_m^{(1)})$ this is so because $w_m^{(1)}(\vartheta) = 0$ and therefore

$$[I_{w_m^{(1)}} f](\vartheta) + [I_{w_m^{(2)}}(\mathbb{\Delta}_{w_m^{(1)}} f)](\vartheta)\, w_m^{(1)}(\vartheta) = [I_{w_m^{(1)}} f](\vartheta) = f(\vartheta) = [I_{w_{2m}} f](\vartheta).$$

For nodes $\vartheta \in \Lambda(H_m^{(2)})$ we have

$$\begin{aligned} [I_{w_m^{(1)}} f](\vartheta) + [I_{w_m^{(2)}}(\mathbb{\Delta}_{w_m^{(1)}} f)](\vartheta)\, w_m^{(1)}(\vartheta) &= [I_{w_m^{(1)}} f](\vartheta) + [\mathbb{\Delta}_{w_m^{(1)}} f](\vartheta)\, w_m^{(1)}(\vartheta) \\ &= f(\vartheta) = [I_{w_{2m}} f](\vartheta) \end{aligned}$$

with the second equality following from the definition (2.10), and (3.11) is established. The assertion on the first block of the vector (3.9) is now verified by inserting the

matrix $H_m^{(1)}$ into the polynomials on either side of (3.11), noting that $w_m^{(1)}(H_m^{(1)}) = O$, and multiplying both sides of (3.10) by e_1.

To verify the second block of (3.9), we use identity (3.11) to write

$$[I_{w_{2m}} f](H_{2m}) = M^{(1)} + M^{(2)} M^{(3)},$$

where

$$M^{(1)} := [I_{w_m^{(1)}} f](H_{2m}), \quad M^{(2)} := [I_{w_m^{(2)}}(\Delta_{w_m^{(1)}} f)](H_{2m}), \quad M^{(3)} := w_m^{(1)}(H_{2m}).$$

The block lower triangular structure of H_{2m} carries over to functions of H_{2m}, giving

$$M^{(i)} = \begin{bmatrix} M_{1,1}^{(i)} & O \\ M_{2,1}^{(i)} & M_{2,2}^{(i)} \end{bmatrix}, \qquad i = 1, 2, 3,$$

where in addition $M_{1,1}^{(3)} = w_m^{(1)}(H_m^{(1)}) = O$. In this notation the second block of (3.9) is given by

$$X_{2,1} e_1 = M_{2,1}^{(1)} e_1 + M_{2,2}^{(2)} M_{2,1}^{(3)} e_1. \tag{3.12}$$

For the first term on the right, we have $M_{2,1}^{(1)} e_1 = \mathbf{0}$ because, as the $(2,1)$-block of $M^{(1)} = [I_{w_m^{(1)}} f](H_{2m})$, a polynomial of degree $m-1$ in the Hessenberg matrix H_{2m}, $M_{2,1}^{(1)}$ has a zero first column. Next, again by the block lower triangular structure of H_{2m}, there holds $M_{2,2}^{(2)} = [I_{w_m^{(2)}}(\Delta_{w_m^{(1)}} f)](H_m^{(2)})$. Finally, we note that $M_{2,1}^{(3)} e_1 = \gamma_m^{(1)} e_1$. This follows in a similar way as the evaluation of $M_{2,1}^{(1)} e_1$, but here $M^{(3)} = w_m^{(1)}(H_{2m})$ is a polynomial of degree m in the $2m \times 2m$ upper Hessenberg matrix H_{2m}. Again by the sparsity structure of powers of Hessenberg matrices, the first column of $M_{2,1}^{(3)}$ is a multiple of e_1. Comparing coefficients reveals this multiple to be $\gamma_m^{(1)}$. Inserting these quantities in (3.12) establishes the second block of identity (3.9), and the proof is complete. ∎

Remark 3.3. We note that the same proof applies when the two Krylov spaces are of different dimensions m_1 and m_2.

Comparing coefficients in (3.7) and (3.9) reveals that $X_{2,1} e_1 = \gamma_m^{(1)} [\Delta_{w_m^{(1)}} f](H_m^{(2)}) e_1$, and we summarize the resulting basic restart step in the following theorem.

THEOREM 3.4. *The Krylov subspace approximation* (3.5) *based on the Arnoldi-like decomposition* (3.3) *is given by*

$$f_{2m} = \beta V_m^{(1)} f(H_m^{(1)}) e_1 + \beta \gamma_m^{(1)} V_m^{(2)} [\Delta_{w_m^{(1)}} f](H_m^{(2)}) e_1. \tag{3.13}$$

Proof. The proof follows immediately from (3.5) upon inserting the representation for $[I_{w_{2m}} f](H_{2m})$ given in Lemma 3.2: Starting with (3.5), we obtain

$$\begin{aligned} f_{2m} &= \beta W_{2m} f(H_{2m}) e_1 \\ &= \beta \left(V_m^{(1)} [I_{w_m^{(1)}} f](H_m^{(1)}) e_1 + V_m^{(2)} \gamma_m^{(1)} [I_{w_m^{(2)}}(\Delta_{w_m^{(1)}} f)](H_m^{(2)}) e_1 \right) \\ &= \beta V_m^{(1)} f(H_m^{(1)}) e_1 + \beta \gamma_m^{(1)} V_m^{(2)} [\Delta_{w_m^{(1)}} f](H_m^{(2)}) e_1, \end{aligned}$$

where the last equality follows from the interpolation properties of $I_{w_m^{(1)}}$ and $I_{w_m^{(2)}}$. ∎

3.3. The restarting algorithm. Theorem 3.4 suggests the following scheme for a Krylov approximation of $f(A)\boldsymbol{b}$ based on the restarted Arnoldi process with cycle length m: The first approximation $\boldsymbol{f}^{(1)}$ is simply the usual Arnoldi approximation with respect to the first Krylov space $\mathcal{K}_m(A,\boldsymbol{b})$, i.e., $\boldsymbol{f}^{(1)} = \beta V_m^{(1)} f(H_m^{(1)})\boldsymbol{e}_1$. The next Krylov space is generated with the initial vector $\boldsymbol{v}_{m+1}^{(1)}$ and, according to (3.13), the correction to $\boldsymbol{f}^{(1)}$ required to obtain the Krylov subspace approximation of $f(A)\boldsymbol{b}$ with respect to the Arnoldi-like decomposition (3.3) is given by

$$\boldsymbol{f}^{(2)} = \boldsymbol{f}^{(1)} + \beta\gamma_m^{(1)} V_m^{(2)} [\Delta_{w_m^{(1)}} f]\big(H_m^{(2)}\big)\boldsymbol{e}_1.$$

The effect of restarting is seen to be a modification of the function f to $\Delta_{w_m}^{(1)} f$ and a replacement of the vector $\boldsymbol{b}$ by $\beta\gamma_m^{(1)}\boldsymbol{v}_{m+1}^{(1)}$. Note that this is in line with the error representation (2.3) in that, after restarting, we are in fact approximating the error term and using this approximation as a correction. The computation of this update requires storing a representation of $\Delta_{w_m^{(1)}} f$ as well as the current approximation $\boldsymbol{f}^{(1)}$, but the Arnoldi basis $V_m^{(1)}$ can be discarded. Proceeding in this fashion, we arrive at the restarting scheme given in Algorithm 1.

ALGORITHM 1: RESTARTED ARNOLDI APPROXIMATION FOR $f(A)\boldsymbol{b}$

Given: $A, \boldsymbol{b}, f$
$f^{(0)} := f$, $\boldsymbol{f}^{(0)} := \boldsymbol{0}$, $\boldsymbol{b}^{(0)} := \boldsymbol{b}$, $\gamma^{(0)} := \|\boldsymbol{b}\|$.
for $k = 1, 2, \ldots$ *until convergence* **do**
  Compute the Arnoldi decomposition $AV_m^{(k)} = V_m^{(k)} H^{(k)} + \eta_{m+1,m}^{(k)} \boldsymbol{b}^{(k)} \boldsymbol{e}_m^\top$ of $\mathcal{K}_m(A, \boldsymbol{b}^{(k-1)})$.
  Update the approximation $\boldsymbol{f}^{(k)} := \boldsymbol{f}^{(k-1)} + \gamma^{(k-1)} V_m^{(k)} f^{(k-1)}(H_m^{(k)})\boldsymbol{e}_1$.
  $\gamma^{(k)} := \gamma^{(k-1)} \prod_{j=1}^m \eta_{j+1,j}^{(k)}$
  $f^{(k)} := \Delta_{w_m^{(k)}} f^{(k-1)}$, where $w_m^{(k)}$ is the characteristic polynomial of $H_m^{(k)}$.

Remark 3.5. Algorithm 1 is formulated for Krylov spaces of constant dimension m in each restart cycle, but this dimension can vary from cycle to cycle.

Although Algorithm 1 appears very attractive from a computational point of view, numerical experiments with a MATLAB implementation have revealed it to be afflicted with severe stability problems. The cause of this seems to be the difficulty of numerically computing interpolation polynomials of high degree (see also [33]).

We therefore turn to a slightly less efficient variant of our restarting scheme, which our numerical tests indicate to be free from these stability problems. The generic step of this second variant of the restarted Arnoldi algorithm proceeds as follows: After $k-1$ cycles of the algorithm, we may collect the entirety of Arnoldi decompositions in the $(k-1)$-fold Arnoldi-like decomposition

$$AW_{(k-1)m} = W_{(k-1)m} H_{(k-1)m} + \eta_{m+1,m}^{(k-1)} \boldsymbol{v}_{m+1}^{(k-1)} \boldsymbol{e}_{(k-1)m}^\top$$

with $W_{(k-1)m} = [V_m^{(1)} V_m^{(2)} \ldots V_m^{(k-1)}]$. Combining this with the Arnoldi decomposition

$$AV_m^{(k)} = V_m^{(k)} H_m^{(k)} + \eta_{m+1,m}^{(k)} \boldsymbol{v}_{m+1}^{(k)} \boldsymbol{e}_m^\top$$

of the next Krylov space $\mathcal{K}_m(A, \boldsymbol{v}_{m+1}^{(k-1)})$, we obtain the next Arnoldi-like decomposi-

tion

$$AW_{km} = W_{km}H_{km} + \eta_{m+1,m}^{(k)} v_{m+1}^{(k)} e_{km}^\top$$

with $W_{km} = [W_{(k-1)m}, V_m^{(k)}]$ and

$$H_{km} = \begin{bmatrix} H_{(k-1)m} & O \\ \eta_{m+1,m}^{(k-1)} e_1 e_{(k-1)m}^\top & H_m^{(k)} \end{bmatrix}.$$

Denoting by w_{km} the characteristic polynomial of H_{km}, formula (2.11) for the Krylov subspace approximation with respect to an Arnoldi-like decomposition gives

$$f^{(k)} = \beta W_{km} f(H_{km}) e_1 = f^{(k-1)} + \beta V_m^{(k)} [f(H_{km}) e_1]_{(k-1)m+1:km}, \tag{3.14}$$

where the subscript in the last term is meant to refer to the vector with the last m components of $f(H_{km})e_1$. (3.14) provides an alternative update formula for the restarted Arnoldi approximation. It is somewhat less efficient than that given in Algorithm 1 in that it requires storing H_{km} and the evaluation of $f(H_{km})$, but we have found it to be much stabler than the former. The second variant is summarized in Algorithm 2.

ALGORITHM 2: RESTARTED ARNOLDI APPROXIMATION FOR $f(A)b$ (VARIANT 2)

Given: A, b, f
$f^{(0)} := f$, $f^{(0)} := 0$, $b^{(0)} := b$, $\beta := \|b\|$.
for $k = 1, 2, \ldots$ *until convergence* **do**
  Compute the Arnoldi decomposition $AV_m^{(k)} = V_m^{(k)} H_m^{(k)} + \eta_{m+1,m}^{(k)} b^{(k)} e_m^\top$ of $\mathcal{K}_m(A, b^{(k-1)})$.
  if $k = 1$ **then**
    $H_{km} := H_m^{(1)}$
  else
    $H_{km} := \begin{bmatrix} H_{(k-1)m} & O \\ \eta_{m+1,m}^{(k-1)} e_1 e_{(k-1)m}^\top & H_m^{(k)} \end{bmatrix}$.
  Update the approximation $f^{(k)} := f^{(k-1)} + \beta V_m^{(k)} [f(H_{km}) e_1]_{(k-1)m+1:km}$.

4. Properties of the restarted Arnoldi algorithm.

4.1. Special cases. In this section we recover some known algorithms as special cases of the restarted Arnoldi approximation.

4.1.1. Linear systems of equations. We begin by showing that for $f(\lambda) = 1/\lambda$ we recover the well-known restarted FOM for solving linear systems of equations [27]. With $I_{w_m} f$ for this case given in (3.1), there results

$$[\Delta_{w_m} f](\lambda) = \frac{1}{w_m(0)} \frac{1}{\lambda},$$

so that the representation (2.11) becomes

$$A^{-1}b = \beta V_m H_m^{-1} e_1 + \frac{\beta\gamma_m}{w_m(0)} A^{-1} v_{m+1} = f_m + \frac{\beta\gamma_m}{w_m(0)} A^{-1} v_{m+1},$$

where $\boldsymbol{f}_m$ denotes the mth FOM iterate. The associated residual is therefore

$$\boldsymbol{r}_m = \boldsymbol{b} - A\boldsymbol{f}_m = \frac{\beta\gamma_m}{w_m(0)}\boldsymbol{v}_{m+1}, \tag{4.1}$$

which leads to

$$A^{-1}\boldsymbol{b} = \boldsymbol{f}_m + A^{-1}\boldsymbol{r}_m.$$

We conclude that in this case the exact correction $\boldsymbol{c}$ to the Arnoldi approximation $\boldsymbol{f}_m$ is the solution of the residual equation $A\boldsymbol{c} = \boldsymbol{r}_m$, leading to the problem of approximating $f(A)\boldsymbol{r}_m$, which in restarted FOM is carried out using a new Krylov space with initial vector $\boldsymbol{r}_m$. As an aside, we observe that (4.1) implies that the FOM residual norm can be expressed as

$$\|\boldsymbol{r}_m\| = \frac{\beta\gamma_m}{|w_m(0)|} = \frac{\beta\prod_{j=1}^{m}\eta_{j+1,j}}{|\det H_m|},$$

an expression first given in [4].

4.2. Initial value problems. We consider the initial value problem

$$\boldsymbol{y}'(t) = A\boldsymbol{y}(t), \quad \boldsymbol{y}(0) = \boldsymbol{b} \tag{4.2}$$

with $A \in \mathbb{C}^{n\times n}$, $\boldsymbol{b} \in \mathbb{C}^n$ (independent of t) with solution

$$\boldsymbol{y}(t) = f_t(A)\boldsymbol{b}, \qquad f_t(\lambda) = e^{t\lambda}. \tag{4.3}$$

The Arnoldi approximation of (4.3) with respect to (2.3) is given by

$$\boldsymbol{y}_m(t) = V_m\boldsymbol{u}(t), \qquad \boldsymbol{u}(t) = \beta e^{tH_m}\boldsymbol{e}_1, \quad \beta = \|\boldsymbol{b}\|. \tag{4.4}$$

As is easily verified, the associated approximation error $\boldsymbol{d}_m(t) := \boldsymbol{y}(t) - \boldsymbol{y}_m(t)$ as a function of t satisfies the initial value problem

$$(\partial_t - A)\boldsymbol{d}_m(t) = \boldsymbol{r}_m(t), \qquad \boldsymbol{d}_m(0) = \boldsymbol{0}, \tag{4.5}$$

in which the forcing term $\boldsymbol{r}_m(t)$, which plays the role of a residual, is given by

$$\boldsymbol{r}_m(t) := \eta_{m+1,m}\boldsymbol{e}_1^{\top}\boldsymbol{u}(t)\boldsymbol{v}_{m+1} = \beta\eta_{m+1,m}\boldsymbol{e}_m^{\top}e^{tH_m}\boldsymbol{e}_1\boldsymbol{v}_{m+1} =: \rho_m(t)\boldsymbol{v}_{m+1}. \tag{4.6}$$

In [4] (see also [19]) Celledoni and Moret propose a restarted Krylov subspace scheme for solving (4.2) based on the variation of constants formula

$$\boldsymbol{d}_m(t) = F_t(A)\boldsymbol{v}_{m+1}, \qquad F_t(\lambda) := \int_0^t e^{(t-s)\lambda}\rho_m(s)\,ds, \tag{4.7}$$

for the solution of the residual equation (4.5) using repeated Arnoldi approximations of $F_t(A)\boldsymbol{v}_{m+1}$ in a manner similar to Algorithm 1. We note that, in contrast to Algorithm 1, their method requires a time-stepping scheme in addition to the Krylov approximation. As the approximate solution (4.4) of (4.2) is an Arnoldi approximation, the error representation (4.7) must coincide with that given in (2.11). To provide more insight on the restarted Arnoldi approximation for solving initial value problems, we proceed to show explicitly that the two error representations are the same. The key is the proper treatment of the parameter t. Denoting the error representation (2.11) with $f = f_t$ by

$$\widetilde{\boldsymbol{d}}_m(t) = \beta\gamma_m[\Delta_{w_m}f_t](A)\boldsymbol{v}_{m+1}, \tag{4.8}$$

we prove the following result.

THEOREM 4.1. *The error representation* (4.8) *for the Arnoldi approximation of* (4.3) *as a function of* t *solves the initial value problem* (4.5).

Proof. The initial condition $\widetilde{\boldsymbol{d}}_m(0) = \mathbf{0}$ follows from the fact that $f_0 \equiv 1$ and, since this function is interpolated without error, the associated divided difference is zero.

To verify that $\widetilde{\boldsymbol{d}}_m$ solves the differential equation, note first that differentiating the interpolant of f_t with respect to the parameter t results in

$$\partial_t[I_{w_m} f_t] = I_{w_m}(\partial_t f_t). \tag{4.9}$$

This can be seen by writing the interpolant as

$$[I_{w_m} f_t](\lambda) = \sum_{j=1}^{k} \sum_{\ell=0}^{n_j-1} f_t^{(\ell)}(\vartheta_j)\, q_{j,\ell}(\vartheta_j), \qquad \sum_{j=1}^{k} n_j = m,$$

in terms of the Hermite basis polynomials $q_{j,\ell} \in \mathcal{P}_{m-1}$, characterized by

$$q_{j,\ell}^{(p)}(\vartheta_q) = \delta_{j,q}\,\delta_{\ell,p}, \qquad j,q = 1,2,\ldots,k, \qquad \ell,p = 0,1,\ldots,n_j-1,$$

and exchanging the order of differentiation. As a consequence of (4.9) and the fact that $(\partial_t f_t)(\lambda) = \lambda f_t(\lambda)$, we also have

$$\partial_t[\Delta_{w_m} f_t] = \Delta_{w_m}(\partial_t f_t) = \Delta_{w_m}(g f_t), \qquad \text{where } g(\lambda) = \lambda.$$

The product formula for divided differences (see, e.g., [25, Theorem 1.3.3]) now yields

$$\partial_t[\Delta_{w_m} f_t](\lambda) = \lambda[\Delta_{w_m} f_t](\lambda) + \pi_{m-1}(t), \tag{4.10}$$

where $\pi_{m-1}(t)$ is the leading coefficient of $I_{w_m} f_t$. Inserting A for λ in the scalar equation (4.10) and multiplying by $\boldsymbol{v}_{m+1}$ now gives us

$$\begin{aligned}(\partial_t - A)\widetilde{\boldsymbol{d}}_m(t) &= \beta\gamma_m \Big(A[\Delta_{w_m} f_t](A) + \pi_{m-1}(t) I - A[\Delta_{w_m} f_t](A) \Big) \boldsymbol{v}_{m+1} \\ &= \beta\gamma_m \pi_{m-1}(t) \boldsymbol{v}_{m+1}.\end{aligned}$$

A comparison with (4.6) reveals that what remains to be shown is that

$$\frac{\gamma_m}{\eta_{m+1,m}} \pi_{m-1}(t) = \boldsymbol{e}_m^{\top} e^{tH_m} \boldsymbol{e}_1.$$

The term on the right is the entry at the $(m,1)$ position of the matrix $e^{tH_m} = [I_{w_m} f_t](H_m)$. Due to the sparsity pattern of powers of an upper Hessenberg matrix, this entry is given by

$$\prod_{j=1}^{m-1} \eta_{j+1,j} \pi_{m-1}(t) = \frac{\gamma_m}{\eta_{m+1,m}} \pi_{m-1}(t)$$

and the proof is complete. □

The uniqueness of the solution of (4.5) together with Theorem 4.1 now imply once more that $\widetilde{\boldsymbol{d}}_m(t) = \boldsymbol{d}_m(t)$. We emphasize again that our restarted Arnoldi method approximates these error terms directly without recourse to a time-stepping scheme.

4.3. Convergence. The full Arnoldi approximation is known to converge superlinearly for the exponential function, as shown in, e.g., [17, 8]. For the case of solving linear systems of equations, i.e., the Arnoldi approximation for the function $f(\lambda) = 1/\lambda$, it is known that restarting the process can degrade or even destroy convergence. In this section we show that, for sufficiently smooth functions, restarting the Arnoldi approximation preserves its superlinear convergence. We state the following result for entire functions of order one (cf. [2, section 2.1]), a class which includes the exponential function, and note that the result generalizes to other orders with minor modifications.

THEOREM 4.2. *Given $A \in \mathbb{C}^{n\times n}$ and an entire function f of order one, let $\{\vartheta_j^{(m)}\}_{j=1}^m, m \geq 1$, denote an arbitrary node sequence contained in the field of values $W(A)$ of A with associated nodal polynomials $w_m \in \mathcal{P}_m$. Then there exist constants C and γ which are independent of m such that*

$$\|f(A)\boldsymbol{b} - [I_{w_m} f](A)\boldsymbol{b}\| \leq C\frac{\gamma^{m-1}}{(m-1)!}\|\boldsymbol{b}\| \qquad \text{for all } m. \tag{4.11}$$

Proof. We recall the well-known Hermite representation theorem for the interpolation error (cf. [6, Theorem 3.6.1]): Let $\Gamma \subset \mathbb{C}$ be a contour which contains $W(A)$, and hence also the interpolation nodes, in its interior, which we denote by Ω. Then for all $\lambda \in \Omega$ we have

$$f(\lambda) - [I_{w_m} f](\lambda) = \frac{1}{2\pi i}\int_\Gamma \frac{f(t)}{t-\lambda}\frac{w_m(\lambda)}{w_m(t)}\,dt. \tag{4.12}$$

By replacing f with $f - p$ in (4.12), we obtain for arbitrary polynomials $p \in \mathcal{P}_{m-1}$ the identity

$$f(\lambda) - [I_{w_m} f](\lambda) = \frac{1}{2\pi i}\int_\Gamma \frac{f(t)-p(t)}{t-\lambda}\frac{w_m(\lambda)}{w_m(t)}\,dt.$$

Inserting A for λ on both sides, multiplying with $\boldsymbol{b}$, and taking norms gives

$$\|f(A)\boldsymbol{b} - [I_{w_m} f](A)\boldsymbol{b}\| = \frac{1}{2\pi}\left\|\int_\Gamma [f(t)-p(t)](tI-A)^{-1}\frac{w_m(A)}{w_m(t)}\,dt\,\boldsymbol{b}\right\|. \tag{4.13}$$

We now bound each factor of the integrand. For any unit vector $\boldsymbol{u} \in \mathbb{C}^n$, we have $\boldsymbol{u}^H A\boldsymbol{u} \in W(A)$, and thus, for all $t \in \Gamma$,

$$\operatorname{dist}(\Gamma, W(A)) \leq |t - \boldsymbol{u}^H A\boldsymbol{u}| = |\boldsymbol{u}^H(tI-A)\boldsymbol{u}| \leq \|(tI-A)\boldsymbol{u}\|.$$

For arbitrary $\boldsymbol{v} \in \mathbb{C}^n$, it follows that $\|(tI-A)\boldsymbol{v}\| \geq \operatorname{dist}(\Gamma, W(A))\|\boldsymbol{v}\|$ and therefore

$$\|(tI-A)^{-1}\| \leq \frac{1}{\operatorname{dist}(\Gamma, W(A))}. \tag{4.14}$$

Similarly, since the nodes $\vartheta_j^{(m)}$ are contained in $W(A)$ by assumption, we have

$$|w_m(t)| = |(t-\vartheta_1^{(m)})(t-\vartheta_2^{(m)})\cdots(t-\vartheta_m^{(m)})| \geq \operatorname{dist}(\Gamma, W(A))^m, \quad t \in \Gamma. \tag{4.15}$$

Moreover, with $r(A) := \max\{|\lambda| : \lambda \in W(A)\}$ denoting the numerical radius of A, we may bound $\|w_m(A)\|$ by

$$\|w_m(A)\| \leq \prod_{j=1}^m \|A - \vartheta_j I\| \leq \prod_{j=1}^m (\|A\| + \vartheta_j) \leq [3r(A)]^m, \tag{4.16}$$

which follows from the well-known inequality $\|A\| \leq 2r(A)$ and since $\vartheta_j^{(m)} \in W(A)$.

Thus, from (4.13), (4.14), (4.15), (4.16), and the fact that $p \in \mathcal{P}_{m-1}$ was arbitrary, we obtain the bound

$$\|f(A)\boldsymbol{b} - [I_{w_m} f](A)\boldsymbol{b}\| \leq \frac{\ell(\Gamma)}{2\pi} \frac{\inf_{p\in\mathcal{P}_{m-1}} \|f-p\|_{\infty,\Omega}\,[3r(A)]^m}{\operatorname{dist}(\Gamma, W(A))^{m+1}} \|\boldsymbol{b}\|,$$

where $\ell(\Gamma)$ denotes the length of the contour Γ and $\|\cdot\|_{\infty,\Omega}$ denotes the supremum norm on Ω. The assertion now follows from the convergence rate of best uniform approximation of entire functions of order one by polynomials. In particular, it is known (see [11]) that there exist constants $\widetilde{C}$ and $\widetilde{\gamma}$ such that

$$\inf_{p\in\mathcal{P}_{m-1}} \|f-p\|_{\infty,\Omega} \leq \widetilde{C}\frac{\widetilde{\gamma}^{m-1}}{(m-1)!}. \qquad \square$$

COROLLARY 4.3. *The restarted Arnoldi approximation converges superlinearly for entire functions of order one.*

Proof. This follows from Theorem 4.2 by noting that, for the Arnoldi approximation, the set of interpolation nodes for each restart cycle are Ritz values of A and therefore contained in $W(A)$. $\square$

5. Numerical experiments. In this section we demonstrate the behavior of the restarted Arnoldi approximation for the exponential function using several examples from the literature. All computations were carried out in MATLAB version 7.0 (R14) on a 1.6 GHz Power Mac G5 computer with 1.5 GB of RAM.

5.1. Three-dimensional heat equation. Our first numerical experiment is based on one from [14]: Consider the initial boundary value problem

$$\dot{u} - \Delta u = 0 \quad \text{on } (0,1)^3 \times (0,T), \tag{5.1a}$$

$$u(x,t) = 0 \quad \text{on } \partial(0,1)^3 \text{ for all } t \in [0,T], \tag{5.1b}$$

$$u(x,0) = u_0(x), \quad x \in (0,1)^3. \tag{5.1c}$$

When the Laplacian is discretized by the usual seven-point stencil on a uniform grid involving n interior grid points in each Cartesian direction, problem (5.1) reduces to the initial value problem

$$\begin{aligned} \dot{\boldsymbol{u}}(t) &= A\boldsymbol{u}(t), \quad t \in (0,T), \\ \boldsymbol{u}(0) &= \boldsymbol{u}_0, \end{aligned}$$

with an $N \times N$ matrix A ($N = n^3$) and an initial vector $\boldsymbol{u}_0$ consisting of the values $u_0(x)$ at the grid points x, the solution of which is given by

$$\boldsymbol{u}(t) = f_t(A)\boldsymbol{u}_0 = e^{tA}\boldsymbol{u}_0. \tag{5.2}$$

As in [14], we give the initial vector in terms of its expansion in eigenfunctions of the discrete Laplacian as

$$\boldsymbol{u}_0^{i,j,k} = \sum_{i',j',k'} \frac{1}{i'+j'+k'} \sin(ii'\pi h)\sin(jj'\pi h)\sin(kk'\pi h).$$

Here $h = 1/(n+1)$ is the mesh size and the triple indexing is relative to the lexicographic ordering of the mesh points in the unit cube.

TABLE 5.1

The full Arnoldi approximation applied to the three-dimensional heat equation with $h = 1/36$ *and* $t = 0.1 = n_{\text{steps}}\Delta t$. *The dimension* m *of the Krylov spaces is chosen as the smallest to result in an error* $\|e\|_2$ *less than* 10^{-10} *at* $t = 0.1$.

Δt	n_{steps}	m	Time [s]	$\|e\|_2$
1e-1	1	72	12.0	7.76e-11
5e-2	2	51	10.5	8.47e-11
2e-2	5	36	13.7	8.54e-11
1e-2	10	29	18.3	2.09e-11
5e-3	20	22	22.7	5.13e-11
1e-3	100	13	42.4	1.55e-11
5e-4	200	11	62.6	5.36e-12
1e-4	1000	8	172.2	1.20e-12
5e-5	2000	7	299.3	1.72e-12

TABLE 5.2

The restarted Arnoldi approximation applied to the three-dimensional heat equation with $h = 1/36$ *and* $t = 0.1$. *The dimension* m *of the Krylov spaces is chosen to coincide with the runs in Table* 5.1 *and now the number of restarts* k *is chosen as the smallest to result in an error* $\|e\|_2$ *less than* 10^{-10} *at* $t = 0.1$.

Δt	k	m	Time [s]	$\|e\|_2$
1e-1	2	51	10.2	2.22e-17
1e-1	2	36	5.2	3.61e-12
1e-1	3	29	5.0	7.78e-15
1e-1	4	22	4.1	9.54e-15
1e-1	6	13	2.2	4.37e-11
1e-1	7	11	1.8	1.29e-11
1e-1	10	8	1.7	7.01e-11
1e-1	12	7	1.6	3.27e-11

We first consider the case $n = 35$ and repeat a calculation in [14], where (5.2) is approximated at $t = 0.1$ using the unrestarted Arnoldi approximation. Writing the solution in the form $\boldsymbol{u}(t) = (e^{\Delta t A})^k \boldsymbol{u}_0$, where $k\Delta t = t$, one can compute the solution using k applications of the Arnoldi approximations involving the matrix $\Delta t A$, which has a smaller spectral interval than A and hence results in faster convergence. There is thus a tradeoff between using Krylov spaces of small dimension and having to take a small number of time steps of length Δt. The results in Table 5.1 show the execution times which result from fixing the time step Δt and using the smallest Krylov subspace dimension m which results in a Euclidean norm of less that 10^{-10} for the error vector $\boldsymbol{e}$ of the unrestarted Arnoldi approximation. We observe that using smaller time steps does allow one to use smaller Krylov spaces, but at a higher cost in terms of execution time.

We next consider the same problem, but instead of taking several time steps with the full Arnoldi approximation, we reduce the size of the Krylov spaces by restarting after every m steps. The results are given in Table 5.2. The dimension m of the Krylov spaces is chosen to coincide with the corresponding runs from Table 5.1, but now the number of restarts k is chosen as the smallest to result in an error less than 10^{-10}. Again there is a tradeoff between the size of the Krylov space and the number of restarts required until convergence. In contrast to Table 5.1, however, we note that the total execution times decrease rather than increase when smaller Krylov spaces are employed, in spite of the fact that this requires more restart cycles. Moreover, the longest execution time of the restarted variant is less than half of the shortest execution time of any of the full Arnoldi approximation runs.

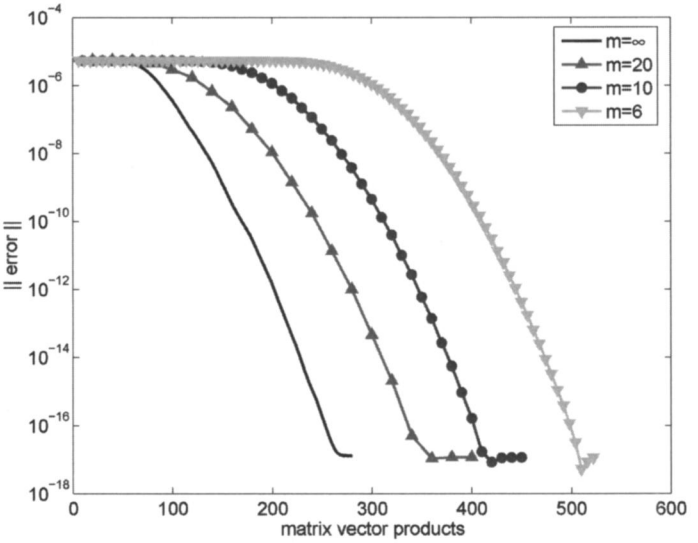

FIG. 5.1. *Error norm histories for the restarted Arnoldi approximation applied to the three-dimensional heat equation with* $n = 50$, *i.e.,* $N = 125\,000$, *for several restart lengths* m.

TABLE 5.3
Execution times for the runs depicted in Figure 5.1*:* m *denotes the restart length and* k *the number of restart cycles.*

m	k	Time [s]
∞	1	1948
20	20	206
10	45	146
6	87	153

Finally, we consider the same problem for a finer discretization with $n = 50$, resulting in a matrix of dimension $N = 125\,000$. We apply the restarted Arnoldi approximation with restart lengths $m = 6, 10$, and 20 using the full Arnoldi approximation ($m = \infty$) as a reference. Each iteration is run until the accuracy no longer improves. The resulting error curves are shown in Figure 5.1, and the corresponding execution times in Table 5.3. We observe here that the method requires successively more restart cycles to converge as the restart length is decreased. Convergence, however, is merely delayed and is maintained down to the smallest restart length $m = 6$. In terms of execution time, there appears to be a point of diminishing returns using shorter and shorter restart lengths, as the shortest execution time was obtained for $m = 10$.

5.2. Skew-symmetric problem. Our next example is taken from [17]. We consider a matrix A with 1001 equidistant eigenvalues in $[-20i, 20i]$. In contrast to [17], we choose A to be block diagonal and real (and not diagonal and complex) in order to avoid complex arithmetic, as follows:

$$
\begin{aligned}
&A = \operatorname{blockdiag}(B_0, B_1, \ldots, B_{500}) \in \mathbb{R}^{1001\times 1001},\\
&B_0 = 0,\\
&B_j = \tfrac{j}{25}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad j = 1, 2, \ldots, 500.
\end{aligned}
\tag{5.3}
$$

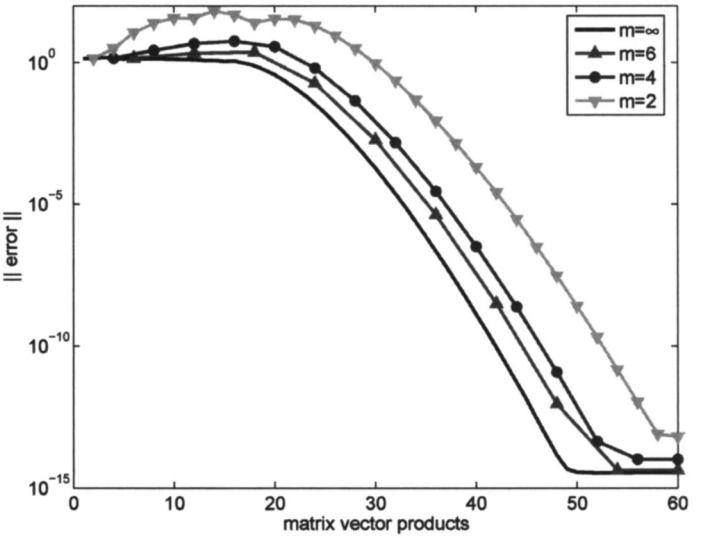

FIG. 5.2. *Error norm histories for the skew-symmetric problem of dimension* $n = 1001$.

The vector $\boldsymbol{b}$ is a random vector[5] of unit norm. The error curve of the full Arnoldi approximation ($m = \infty$) as well as those of the restarted Arnoldi approximation with restart lengths $m = 2, 4$, and 6 are shown in Figure 5.2.

We observe that the errors associated with the restarted Arnoldi approximations initially increase before tending to zero. We also observe that the final accuracy of the approximation deteriorates with decreasing restart length. This indicates that the restart length m is too small to "resolve" the spectral interval of A.

For an explanation, recall that the Arnoldi approximations $\boldsymbol{f}_m$ of $\exp(A)\boldsymbol{b}$ can be viewed as the result of an interpolation process: $\boldsymbol{f}_m = q_{m-1}(A)\boldsymbol{b}$, where q_{m-1} is an interpolating polynomial for the exponential function. For the unrestarted Arnoldi method, the interpolation nodes are the Ritz values of A with respect to $\mathcal{K}_m(A, \boldsymbol{b})$, which are approximately uniformly distributed over $[-20i, 20i]$ (cf. Figure 5.3, where the imaginary parts of the Ritz values are shown[6].) For the restarted Arnoldi method (with restart length m), however, the interpolation nodes are the collection of the Ritz values of A with respect to several Krylov spaces $\mathcal{K}_m\left(A, \boldsymbol{b}^{(j)}\right)$, $j = 0, 1, \ldots, k-1$ (after k restarts). These are far from uniformly distributed in $[-20i, 20i]$, but rather tend to accumulate at m discrete points (see Figure 5.3).

In the extreme case of restart length one, all interpolating nodes equal $\vartheta = 0$ (at least in exact arithmetic) and the interpolating polynomial $q_{k-1}(\lambda) = \sum_{j=0}^{k-1} \frac{1}{j!}\lambda^j$ is simply the truncated Taylor expansion of $\exp(\lambda)$. It is well known that, for $|\lambda| \gg 0$, intermediate partial sums are much larger than the final limit. An analogous statement holds for Hermite interpolating polynomials of the exponential function at too few nodes.

The phenomenon described above becomes more pronounced if we increase the spectral interval of A: Again, we consider the matrix A of (5.3), but now of dimension 10001 with equidistant eigenvalues in $[-200i, 200i]$. The resulting error curves for the restart lengths $m = 5, 10, 20$, and 40 are shown in Figure 5.4.

[5]Generated by the MATLAB syntax `randn('state',0); b = randn(1001,1)`

[6]Note that all Ritz values are purely imaginary because A is skew-symmetric and $\boldsymbol{b}$ is real.

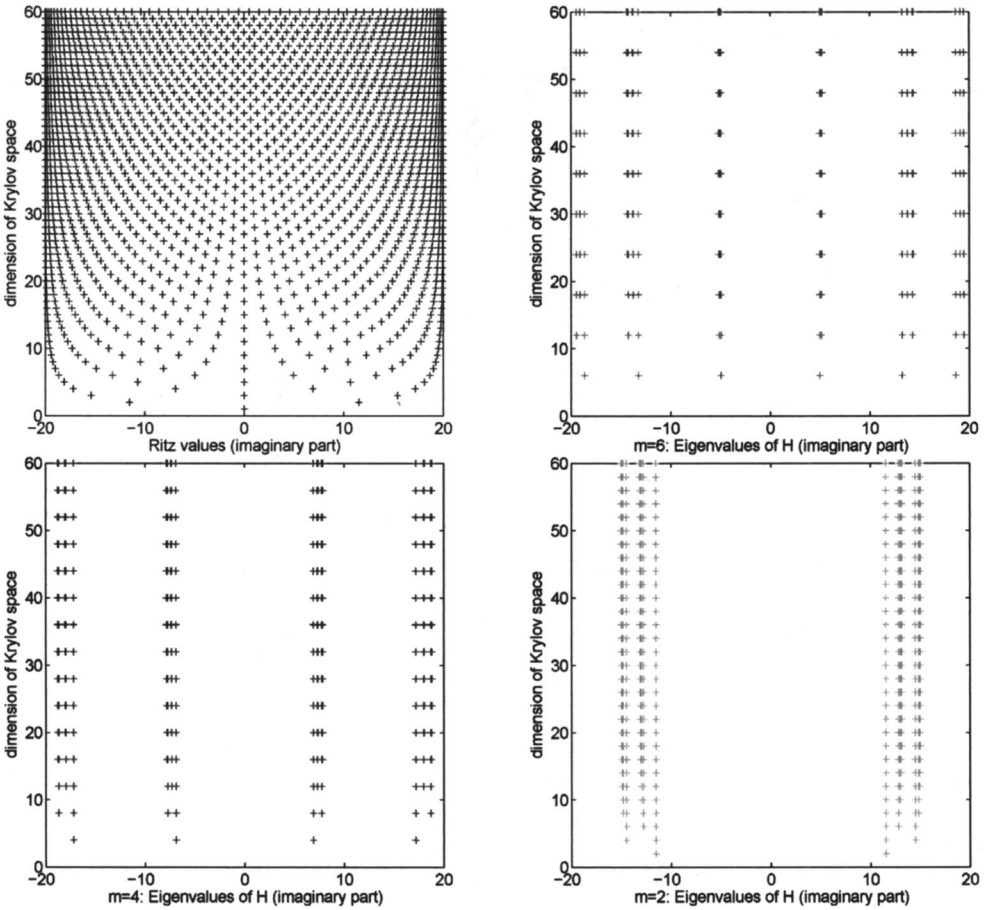

FIG. 5.3. *Interpolation nodes for the skew-symmetric problem of dimension* $n = 1001$.

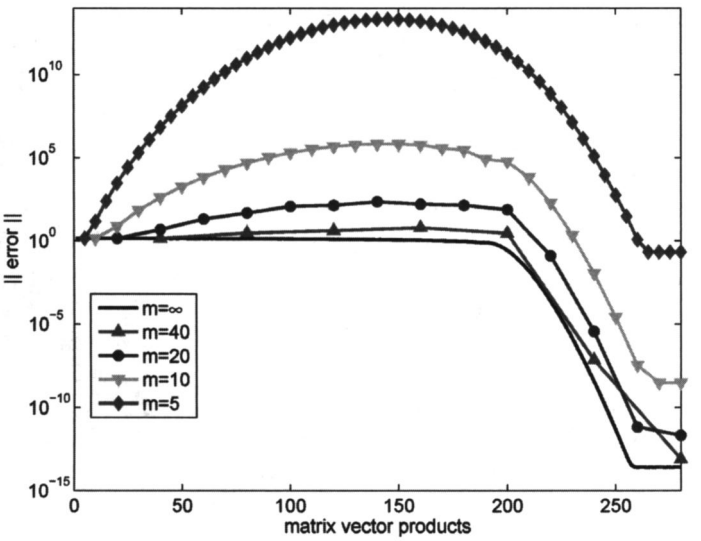

FIG. 5.4. *Error norm histories for the skew-symmetric problem of dimension* $n = 10001$.

TABLE 5.4
The restarted Arnoldi approximation applied to the skew-symmetric problem of dimension 10001, *for several restart lengths* m *(cf. Figure* 5.4*).*

m	Matrix vector products	Time[s]	Final accuracy	Largest error	Largest error / Final accuracy
∞	260	367	2.5e−14	1.4e+01 [1]	1.8e−14
40	280	48	7.8e−14	6.3e+01 [160]	1.2e−14
20	280	26	2.1e−12	2.3e+02 [140]	8.9e−15
10	270	16	2.9e−09	6.8e+05 [140]	4.3e−15
5	275	13	2.1e−01	2.2e+13 [145]	9.7e−15

Table 5.4 shows the number of matrix-vector products and the execution times which were required to reach a final accuracy for different restart length m. We also list the largest intermediate error (and after how many matrix-vector multiplications it is observed). Note that for every m the quotient of this largest error and the final accuracy approximately equals the machine precision of $2e-16$.

5.3. Convection-diffusion problem. Our final example is taken from [19, Example 6.1]: We consider the initial boundary value problem

$$\begin{aligned} \dot{u} - \Delta u + \tau_1 u_{x_1} + \tau_2 u_{x_2} &= 0 \quad \text{on } (0,1)^3 \times (0,T), \\ u(x,t) &= 0 \quad \text{on } \partial(0,1)^3 \text{ for all } t \in [0,T], \\ u(x,0) &= u_0(x), \quad x \in (0,1)^3. \end{aligned}$$

Discretizing the Laplacian by the usual seven-point stencil and the first-order derivatives, u_{x_1} and u_{x_2}, by central differences on a uniform grid with step size $h = 1/(n+1)$ leads—as in section 5.1—to an ordinary initial value problem

$$\begin{aligned} \dot{\boldsymbol{u}}(t) &= A\boldsymbol{u}(t), \quad t \in (0,T), \\ \boldsymbol{u}(0) &= \boldsymbol{u}_0 \end{aligned}$$

with the matrix

$$A = I_n \otimes [I_n \otimes C_1] + [B \otimes I_n + I_n \otimes C_2] \otimes I_n$$

of dimension $N = n^3$. Here,

$$B = \tfrac{1}{h^2}\text{tridiag}(1,-2,1), \quad C_j = \tfrac{1}{h^2}\text{tridiag}(1+\mu_j, -2, 1-\mu_j), \quad j = 1,2,$$

where $\mu_j = \tau_j h/2$. The nonsymmetric matrix A is a popular test matrix because its eigenvalues are explicitly known: If $|\mu_j| > 1$ (for at least one j), they are complex; more precisely (cf. [19]),

$$\begin{aligned} \Lambda(A) \subset\ & \tfrac{1}{h^2}\left[-6 - 2\cos(\pi h)\,\text{Re}(\theta),\ -6 + 2\cos(\pi h)\,\text{Re}(\theta)\right] \\ & \times \tfrac{1}{h^2}\left[-2i\cos(\pi h)\,\text{Im}(\theta),\ 2i\cos(\pi h)\,\text{Im}(\theta)\right] \end{aligned}$$

with $\theta = 1 + \sqrt{1-\mu_1^2} + \sqrt{1-\mu_2^2}$. As in [19], we choose $h = 1/16$, $\tau_1 = 96$, $\tau_2 = 128$ ($\tau_1 = \tau_2 = 320$) which leads to $\mu_1 = 3$, $\mu_2 = 4$ ($\mu_1 = \mu_2 = 10$), and approximate $e^{tA}\boldsymbol{b}$, where $t = h^2$ and $\boldsymbol{b} = [1,1,\ldots,1]^\top$. The resulting error norm histories are shown in Figure 5.5. In the second example we again observe transient error growth. We attribute this, as in the skew-symmetric example, to the sufficiently large imaginary parts of the eigenvalues of $h^2 A$, which lie in

$$[-8.0, -4.0] \times i[-13.1, 13.1] \quad \text{for} \quad \mu_1 = 3, \mu_2 = 4$$

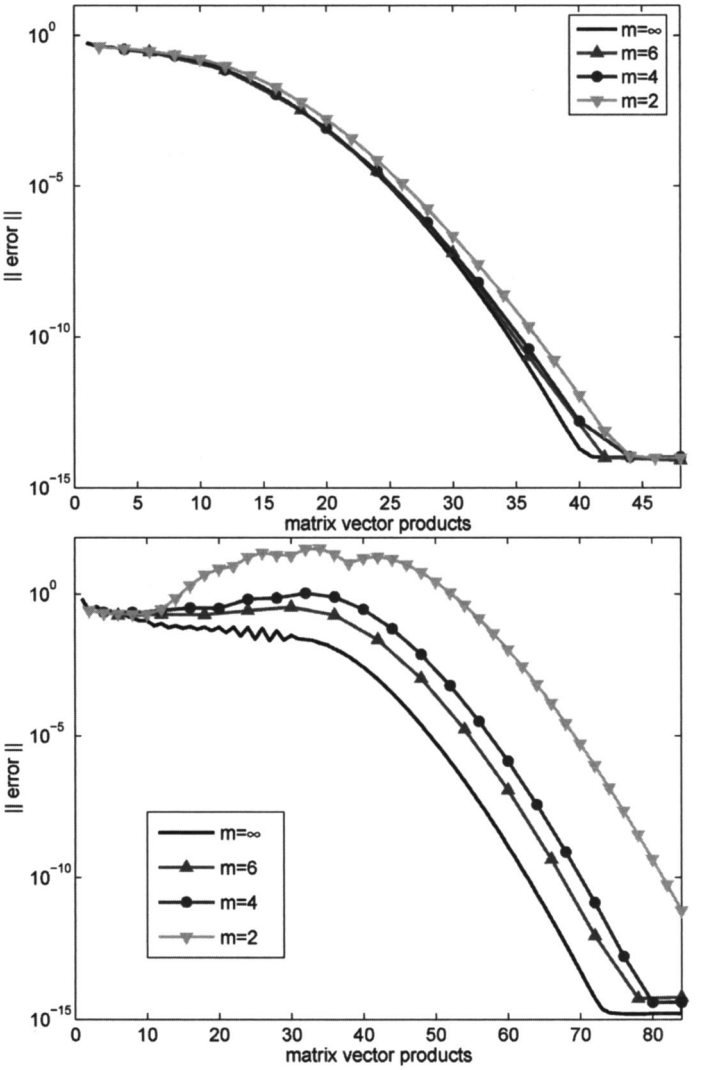

FIG. 5.5. *Error norm histories for the restarted Arnoldi approximation applied to the convection-diffusion problem with* $n = 15$, *i.e.,* $N = 3375$ *for several restart lengths* m. *As in* [19], *we chose* $\mu_1 = 3$, $\mu_2 = 4$ *(top), and* $\mu_1 = \mu_2 = 10$ *(bottom).*

and

$$[-8.0, -4.0] \times i[-39.0, 39.0] \quad \text{for} \quad \mu_1 = \mu_2 = 10,$$

respectively.

6. Conclusions. We have shown how Krylov subspace methods for approximating $f(A)\boldsymbol{b}$ may be restarted. This permits the application of schemes like the Arnoldi approximation to very large matrices using a fixed amount of storage space. For functions f which are entire of order one, the restarted method retains the superlinear convergence property of the unrestarted method. In addition, we have identified the relationship of the restarted method to known algorithms in the cases $f(\lambda) = 1/\lambda$ and

$f_t(t\lambda) = e^{t\lambda}$. Moreover, we have demonstrated that the method performs well on several numerical examples from the literature. Related issues such as characterizing the convergence of the Arnoldi approximation using potential theoretic methods as well as yet more efficient implementations of the restarted algorithm will be the subject of future research.

REFERENCES

[1] E. J. Allen, J. Baglama, and S. K. Boyd, *Numerical approximation of the product of the square root of a matrix with a vector*, Linear Algebra Appl., 310 (2000), pp. 167–181.

[2] R. P. Boas, *Entire Functions*, Academic Press, New York, 1954.

[3] P. N. Brown, *A theoretical comparison of the Arnoldi and GMRES algorithms*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 58–78.

[4] E. Celledoni and I. Moret, *A Krylov projection method for systems of ODEs*, Appl. Numer. Math., 24 (1997), pp. 365–378.

[5] P. I. Davies and N. J. Higham, *A Schur–Parlett algorithm for computing matrix functions*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 464–485.

[6] P. J. Davis, *Interpolation and Approximation*, Dover, New York, 1975.

[7] C. de Boor, *Divided differences*, Surv. Approximation Theory, 1 (2005), pp. 46–69.

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

[9] V. L. Druskin and L. A. Knizhnerman, *Krylov subspace approximation of eigenpairs and matrix functions in exact and computer arithmetic*, Numer. Linear Algebra Appl., 2 (1995), pp. 205–217.

[10] M. Eiermann, O. G. Ernst, and O. Schneider, *Analysis of acceleration strategies for restarted minimal residual methods*, J. Comput. Appl. Math., 123 (2000), pp. 261–292.

[11] M. Freund and E. Görlich, *Polynomial approximation of entire functions and rate of growth of Taylor coefficients*, Proc. Edinb. Math. Soc., 28 (1985), pp. 341–348.

[12] R. W. Freund and M. Hochbruck, *Gauss-quadratures associated with the Arnoldi process and the Lanczos algorithm*, in Linear Algebra for Large-Scale and Real-Time Applications, M. S. Moonen, G. H. Golub, and B. L. R. de Moor, eds., Kluwer Academic Publishers, Dordrecht, 1993, pp. 377–380.

[13] R. W. Freund, *Quasi-kernel polynomials and their use in non-Hermitian matrix iterations*, J. Comput. Appl. Math., 43 (1992), pp. 135–158.

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

[15] F. R. Gantmacher, *The Theory of Matrices*, Vol. I, AMS Chelsea, Providence, RI, 1959.

[16] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, 1996.

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

[18] C. Moler and C. F. van Loan, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*, SIAM Rev., 45 (2003), pp. 3–49.

[19] I. Moret and P. Novati, *An interpolatory approximation of the matrix exponential based on Faber polynomials*, J. Comput. Appl. Math., 131 (2001), pp. 361–380.

[20] I. Moret and P. Novati, *Interpolating functions of matrices on zeros of quasi-kernel polynomials*, Numer. Linear Algebra Appl., 12 (2005), pp. 337–353.

[21] R. B. Morgan, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1154–1171.

[22] P. Novati, *A method based on Fejér points for the computation of functions of nonsymmetric matrices*, Appl. Numer. Math., 44 (2003), pp. 201–224.

[23] C. C. Paige, B. Parlett, and H. A. van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numer. Linear Algebra Appl., 2 (1995), pp. 115–133.

[24] B. N. Parlett, *The Symmetric Eigenvalue Problem*, SIAM, Philadelphia, 1997.

[25] G. M. Phillips, *Interpolation and Approximation by Polynomials*, Springer-Verlag, New York, 2003.

[26] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

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

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

[29] B. SINGER AND S. SPILERMAN, *The representation of social processes by Markov models*, Amer. J. Sociology, 8 (1976), pp. 1–54.

[30] D. C. SORENSEN, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.

[31] D. E. STEWART AND T. S. LEYK, *Error estimates for Krylov subspace approximations of matrix exponentials*, J. Comput. Appl. Math., 72 (1996), pp. 359–369.

[32] G. W. STEWART, *A Krylov–Schur algorithm for large eigenproblems*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 601–614.

[33] H. TAL-EZER, *High degree polynomial interpolation in Newton form*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 648–667.

[34] J. VAN DEN ESHOF, A. FROMMER, T. LIPPERT, AND H. A. VAN DER VORST, *Numerical methods for the QCD overlap operator. I. Sign-function and error bounds*, Comput. Phys. Comm., 146 (2002), pp. 203–224.