Electronic Transactions on Numerical Analysis.
Volume 45, pp. 133–145, 2016.

ISSN 1068–9613.

ETNA
Kent State University
http://etna.math.kent.edu

ANY FINITE CONVERGENCE CURVE IS POSSIBLE IN THE INITIAL ITERATIONS OF RESTARTED FOM*

MARCEL SCHWEITZER†

Abstract. We investigate the possible convergence behavior of the restarted full orthogonalization method (FOM) for non-Hermitian linear systems $A\boldsymbol{x} = \boldsymbol{b}$. For the GMRES method, it is known that any nonincreasing sequence of residual norms is possible, independent of the eigenvalues of $A \in \mathbb{C}^{n\times n}$. For FOM, however, there has not yet been any literature describing similar results. This paper complements the results for (restarted) GMRES by showing that any finite sequence of residual norms is possible in the first n iterations of restarted FOM, where by finite we mean that we only consider the case that all FOM iterates are defined, and thus no “infinite” residual norms occur. We discuss the relation of our results to known results on restarted GMRES and give a new result concerning the possible convergence behavior of restarted GMRES for iteration counts exceeding the matrix dimension n. In addition, we give a conjecture on an implication of our result with respect to the convergence of the restarted Arnoldi approximation for $g(A)\boldsymbol{b}$, the action of a matrix function on a vector.

Key words. linear systems, restarted Krylov subspace methods, full orthogonalization method, restarted Arnoldi method for matrix functions, GMRES method

AMS subject classifications. 65F10, 65F50, 65F60

1. Introduction.

For solving a linear system

$$A\boldsymbol{x} = \boldsymbol{b} \tag{1.1}$$

with a large, sparse, non-Hermitian matrix $A \in \mathbb{C}^{n\times n}$ and a vector $\boldsymbol{b} \in \mathbb{C}^n$ one often uses a *Krylov subspace method*. One possible choice is the *full orthogonalization method (FOM)*; see, e.g., [16, 17, 19]. Given an initial guess $\boldsymbol{x}_0$, one computes the residual $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$ and then generates the Arnoldi decomposition

$$AV_j = V_jH_j + h_{j+1,j}\boldsymbol{v}_{j+1}\boldsymbol{e}_j^H, \tag{1.2}$$

where the columns of $V_j = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j] \in \mathbb{C}^{n\times j}$ form an orthonormal basis of the jth Krylov subspace $\mathcal{K}_j(A, \boldsymbol{r}_0) = \operatorname{span}\{\boldsymbol{r}_0, A\boldsymbol{r}_0, \ldots, A^{j-1}\boldsymbol{r}_0\}$, the matrix $H_j \in \mathbb{C}^{j\times j}$ is unreduced upper Hessenberg and $\boldsymbol{e}_j \in \mathbb{C}^j$ denotes the jth canonical unit vector. The jth FOM iterate for the linear system (1.1) is then given by

$$\boldsymbol{x}_j = \boldsymbol{x}_0 + \|\boldsymbol{r}_0\|_2 V_j H_j^{-1}\boldsymbol{e}_1, \tag{1.3}$$

provided that H_j is nonsingular, and can be characterized by the variational condition

$$\boldsymbol{b} - A\boldsymbol{x}_j \perp \mathcal{K}_j(A, \boldsymbol{r}_0).$$

Note that the decomposition (1.2) is unique up to scaling of the columns of V_j by scalars of modulus one (and scaling of the corresponding entries of H_j; see, e.g., [20, Chapter 5, Theorem 1.3]). Therefore, if the subdiagonal entries of H_j are prescribed to be real and positive (as it is always the case when (1.2) is computed by the Arnoldi process), the decomposition is unique.

*Received June 18, 2014. Accepted February 12, 2016. Published online on May 4, 2016. Recommended by Z. Strakos. This work was supported by Deutsche Forschungsgemeinschaft through Collaborative Research Centre SFB TRR55 “Hadron Physics from Lattice QCD”.

†Department of Mathematics, Bergische Universität Wuppertal, 42097 Wuppertal, Germany (schweitzer@math.uni-wuppertal.de)

For larger values of j, the computational cost of constructing the orthonormal basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j$ as well as the cost of computing $H_j^{-1}\boldsymbol{e}_1$ grows. In addition, all basis vectors $\boldsymbol{v}_i, i = 1, \ldots, j$ need to be stored to evaluate (1.3). For Hermitian A, these problems do not occur because short recurrences for the basis vectors can be used (which also translate into short recurrences for the iterates $\boldsymbol{x}_j$), leading to the *conjugate gradient method (CG)* [15] when A is positive definite. In the non-Hermitian case, a typical remedy is *restarting*. After a fixed (small) number m of iterations (the *restart length*), one computes a first approximation

$$\boldsymbol{x}_m^{(1)} = \boldsymbol{x}_0 + \|\boldsymbol{r}_0\|_2 V_m^{(1)} \left(H_m^{(1)}\right)^{-1} \boldsymbol{e}_1$$

and then uses the fact that the error $\boldsymbol{d}_m^{(1)} = \boldsymbol{x}^* - \boldsymbol{x}_m^{(1)}$, where $\boldsymbol{x}^*$ is the exact solution of (1.1), satisfies the residual equation

$$A\boldsymbol{d}_m^{(1)} = \boldsymbol{r}_m^{(1)}, \text{ where } \boldsymbol{r}_m^{(1)} = \boldsymbol{b} - A\boldsymbol{x}_m^{(1)}, \tag{1.4}$$

so that $\boldsymbol{d}_m^{(1)}$ can be approximated by another m iterations of FOM for the linear system (1.4) without needing to store the quantities $V_m^{(1)}, H_m^{(1)}$ from the first m iterations. The resulting approximation $\widetilde{\boldsymbol{d}}_m^{(1)}$ is then used as an additive correction to the iterate $\boldsymbol{x}_m^{(1)}$, i.e.,

$$\boldsymbol{x}_m^{(2)} = \boldsymbol{x}_m^{(1)} + \widetilde{\boldsymbol{d}}_m^{(1)}.$$

This approach can be continued until the resulting iterate $\boldsymbol{x}_m^{(k)}$ fulfills a prescribed stopping criterion (e.g., a residual norm below some given tolerance). In the following we refer to the resulting iterative method as restarted FOM with restart length m, or, as a shorthand, FOM(m). An *iteration of FOM(m)* refers to advancing from iterate $\boldsymbol{x}_{j-1}^{(k)}$ to $\boldsymbol{x}_j^{(k)}$ (an operation which involves one matrix-vector multiplication with A), while a *cycle of FOM(m)* refers to the m iterations necessary to advance from $\boldsymbol{x}_m^{(k-1)}$ to $\boldsymbol{x}_m^{(k)}$.

While FOM(m) is simple to understand and implement, it is not at all clear whether the iterates $\boldsymbol{x}_m^{(k)}$ will converge to $\boldsymbol{x}^*$ for $k \to \infty$, even when all iterates are defined (i.e., when $H_m^{(k)}$ is nonsingular for all k). In this paper we show that the residual norms produced by restarted FOM can attain any finite values in the first n iterations, for a matrix A with any prescribed nonzero eigenvalues, showing that a convergence analysis for restarted FOM based exclusively on spectral information is not possible in general. This result is similar to results from [2, 7–9, 13, 21] for (restarted) GMRES, but there are also some differences: Due to the minimizing property of GMRES (see, e.g., [18]) the GMRES convergence curve is always nonincreasing. The results in [7, 13] only consider full GMRES (without restarts) while in [21] only the residual norms at the end of each restart cycle are prescribed, instead of the residual norms after each individual iteration.

The remainder of this paper is organized as follows. In Section 2, we present our main result and its constructive proof. A discussion of the relation and differences of our result to those on (restarted) GMRES is presented in Section 3. In Section 4 we briefly discuss the approximation of $g(A)\boldsymbol{b}$—the action of a matrix function on a vector—by the restarted Arnoldi method and give a conjecture on the arbitrary convergence behavior of this method, motivated by our results on FOM(m). Concluding remarks are given in Section 5.

2. Any finite convergence curve is possible in the first n iterations of FOM(m). For the sake of simplicity, we only consider the case of constant restart length m across all restart cycles, keeping in mind that all results below generalize in a straightforward manner to the case of varying restart lengths. We conveniently number the residuals corresponding to iterates from different restart cycles consecutively, i.e., $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m$ are the residuals of the iterates from the first restart cycle, $\boldsymbol{r}_{m+1}, \ldots, \boldsymbol{r}_{2m}$ are the residuals of the iterates from the second restart cycle and so on.

THEOREM 2.1. *Let $m, n, q \in \mathbb{N}$ with $m \leq n-1$ and $q \leq n$; let $f_1, \ldots, f_q \in \mathbb{R}_0^+$ be given with $f_1, \ldots, f_{q-1} > 0$ and $f_q \geq 0$; and let $\mu_1, \ldots, \mu_n \in \mathbb{C} \setminus \{0\}$, not necessarily pairwise distinct. Then there exists a matrix $A \in \mathbb{C}^{n \times n}$ with $\operatorname{spec}(A) = \{\mu_1, \ldots \mu_n\}$ (where $\operatorname{spec}(A)$ denotes the set of all eigenvalues of A) and vectors $\boldsymbol{b}, \boldsymbol{x}_0 \in \mathbb{C}^n$ such that the residuals $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_q$ generated by q iterations of FOM(m) for $A\boldsymbol{x} = \boldsymbol{b}$ with initial guess $\boldsymbol{x}_0$ satisfy*

$$\|\boldsymbol{r}_j\|_2 = f_j \qquad \text{for } j = 1, \ldots, q.$$

For ease of presentation, we will first assume $q = n$ and $f_n > 0$ in the following. Afterwards, we briefly comment on the modifications necessary in the general case. The proof of Theorem 2.1 is constructive in nature and based on investigating properties of matrices of the form

$$A(\boldsymbol{d}, \boldsymbol{s}) = \begin{bmatrix} d_1 & 0 & \cdots & 0 & s_n \\ s_1 & d_2 & 0 & \cdots & 0 \\ 0 & s_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & d_{n-1} & 0 \\ 0 & \cdots & 0 & s_{n-1} & d_n \end{bmatrix} \tag{2.1}$$

defined by two vectors $\boldsymbol{d}, \boldsymbol{s} \in \mathbb{C}^n$. Before proceeding, we explain why it is reasonable to investigate these kinds of matrices in our setting. First, notice that we want to be able to prescribe up to $2n$ values (the n eigenvalues of A and up to n residual norms). Thus, assuming that the initial residual $\boldsymbol{r}_0$ is fixed (which will be the case in our construction), we should have at least $2n$ degrees of freedom when choosing the matrix A. The matrix $A(\boldsymbol{d}, \boldsymbol{s})$ exactly fulfills this minimal requirement. The other important property is that the structure of A is chosen such that unrestarted FOM and FOM(m) for any restart length behave exactly the same in the first $n-1$ iterations when the initial residual is a canonical unit vector, so that we do not have to take the restart length into account and can perform most of the analysis as if we were dealing with unrestarted FOM. This is proven in Lemma 2.4.

We begin our analysis by proving the following proposition, which characterizes the result of j iterations of FOM for the matrix $A(\boldsymbol{d}, \boldsymbol{s})$ when started with a (multiple of a) canonical unit vector.

PROPOSITION 2.2. *Let $A(\boldsymbol{d}, \boldsymbol{s}) \in \mathbb{C}^{n \times n}$ be of the form (2.1), let $j \leq n-1$, $\xi_0 \in \mathbb{C}$ with $|\xi_0| = 1$, let $c > 0$ and let $\mathbf{e}_i$ denote the ith canonical unit vector. Let $\boldsymbol{x}_0, \boldsymbol{b} \in \mathbb{C}^n$ be given such that the residual $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$ satisfies $\boldsymbol{r}_0 = \xi_0 c \mathbf{e}_i$. Then the basis V_{j+1} generated by j iterations of FOM for $A(\boldsymbol{d}, \boldsymbol{s})$ and $\boldsymbol{b}$ with initial guess $\boldsymbol{x}_0$ is given by*

$$V_{j+1} = [\xi_0 \mathbf{e}_i, \xi_1 \mathbf{e}_{i+1}, \ldots, \xi_j \mathbf{e}_{i+j}]$$

(where, like everywhere in the following, for ease of notation, the indices are to be understood cyclically, i.e., $\mathbf{e}_{n+1} := \mathbf{e}_1, \mathbf{e}_{n+2} := \mathbf{e}_2, \ldots$), with

$$\xi_\ell = \frac{s_{i+\ell-1}\xi_{\ell-1}}{|s_{i+\ell-1}|}, \quad \ell = 1, \ldots, j.$$

ETNA
Kent State University
http://etna.math.kent.edu

The corresponding upper Hessenberg matrix is given by

$$H_j = \begin{bmatrix} d_i & 0 & \cdots & 0 & 0 \\ |s_i| & d_{i+1} & 0 & \cdots & 0 \\ 0 & |s_{i+1}| & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & d_{i+j-2} & 0 \\ 0 & \cdots & 0 & |s_{i+j-2}| & d_{i+j-1} \end{bmatrix}, \quad h_{j+1,j} = |s_{i+j-1}|.$$

Proof. The result follows by direct verification of the Arnoldi relation (1.2). ∎
Of course, it is also possible to state the result of Proposition 2.2 by just using $\zeta \boldsymbol{e}_i$, with an arbitrary complex scalar ζ, as initial residual. We distinguish between the complex scalar ξ_0 of modulus one and the real, positive scalar c in Proposition 2.2 mainly for ease of notation and a clearer presentation. It allows us to give an easy explicit recursion for the values $\xi_\ell, \ell = 1, \dots, j$, appearing in the Arnoldi basis and to relate the value c to the residual norms to be prescribed later on (one directly sees that in Proposition 2.2 we have $\|\boldsymbol{r}_0\|_2 = c$).

The following result is now easily provable by using Proposition 2.2.

PROPOSITION 2.3. *Let the assumptions of Proposition 2.2 hold. Then the residual generated by $j \le n-1$ iterations of FOM is given by*

$$\boldsymbol{r}_j = (-1)^j \xi_j c \frac{|s_i \cdot s_{i+1} \cdots s_{i+j-1}|}{d_i \cdot d_{i+1} \cdots d_{i+j-1}} \boldsymbol{e}_{i+j}.$$

In particular,

$$\|\boldsymbol{r}_j\|_2 = \left| \frac{s_{i+j-1}}{d_{i+j-1}} \right| \cdot \|\boldsymbol{r}_{j-1}\|_2.$$

Proof. The FOM residual satisfies $\boldsymbol{r}_j = -h_{j+1,j} \|\boldsymbol{r}_0\|_2 \left(\boldsymbol{e}_j^H H_j^{-1} \boldsymbol{e}_1 \right) \boldsymbol{v}_{j+1}$; see [17]. In our setting we have

$$\|\boldsymbol{r}_0\|_2 = c, \quad h_{j+1,j} = |s_{i+j}|, \quad \boldsymbol{v}_{j+1} = \xi_j \boldsymbol{e}_{i+j},$$

and

$$\boldsymbol{e}_j^H H_j^{-1} \boldsymbol{e}_1 = \frac{|s_{i+1}| \cdots |s_{i+j-1}|}{d_{i+1} \cdots d_{i+j}}$$

due to the simple, bidiagonal structure of H_j. ∎

Proposition 2.2 and 2.3 now allow to relate the behavior of FOM(m) for $A(\boldsymbol{d}, \boldsymbol{s})$ to the behavior of unrestarted FOM.

LEMMA 2.4. *Let the assumptions of Proposition 2.2 hold and let $m \le n-1$. Then the residuals produced by the first $n-1$ iterations of FOM(m) and unrestarted FOM coincide.*

Proof. For simplicity, we only consider the situation after one restart, the result for an arbitrary number of restarts follows analogously. Let $1 \le j \le m$ be such that $m + j \le n - 1$. According to Proposition 2.3, the residual norm after $m + j$ iterations of unrestarted FOM is given by

$$\boldsymbol{r}_{m+j} = (-1)^{m+j} \xi_{m+j} c \frac{|s_i \cdot s_{i+1} \cdots s_{i+m+j-1}|}{d_i \cdot d_{i+1} \cdots d_{i+m+j-1}} \boldsymbol{e}_{i+m+j}. \tag{2.2}$$

ETNA
Kent State University
http://etna.math.kent.edu

In the same way, after the first cycle of FOM(m), the residual—and thus the right hand side for the second cycle—is given by

$$\boldsymbol{r}_m = (-1)^m \xi_m c \frac{|s_i \cdot s_{i+1} \cdots s_{i+m-1}|}{d_i \cdot d_{i+1} \cdots d_{i+m-1}} \boldsymbol{e}_{i+m}. \tag{2.3}$$

The result of Proposition 2.3 for j iterations of FOM with initial residual (2.3) yields exactly the same result as (2.2). ◻

We are now in a position to prove our main result. We first note that it would also be possible to derive Theorem 2.1 by exploiting known results for the convergence of GMRES together with the relationship between FOM residual norms and GMRES residual norms (cf. also the discussion in Section 3), but we prefer to give a constructive proof of our result, as it gives additional insight into the behavior of the method which will be useful for further considerations in the remaining sections of this paper.

Assume that a sequence $f_1, \ldots, f_m$ of positive scalars is given. Based on Proposition 2.3, we can derive conditions on $\boldsymbol{d}$ and $\boldsymbol{s}$, such that FOM(m) for $A(\boldsymbol{d}, \boldsymbol{s})$ and $\boldsymbol{b} = \boldsymbol{e}_1$ with initial guess $\boldsymbol{x}_0 = \boldsymbol{0}$ produces the residual norms $f_1, \ldots, f_m$. Setting $f_0 = \|\boldsymbol{r}_0\|_2 = 1$, the conditions

$$s_j = \frac{f_j}{f_{j-1}} d_j, \quad j = 1, \ldots, m \tag{2.4}$$

guarantee that the desired residual norm sequence is obtained. Therefore, for any fixed nonzero choice of $d_1, \ldots, d_m$, there exist coefficients $s_1, \ldots, s_m$ such that the first m iterations of FOM produce the desired residual norms $f_1, \ldots, f_m$. The freedom in the choice of the coefficient vector $\boldsymbol{d}$ can be used to prescribe the eigenvalues of $A(\boldsymbol{d}, \boldsymbol{s})$. Before describing this in detail, we consider the situation after restarting the method.

After m iterations of FOM started with $\boldsymbol{x}_0 = \boldsymbol{0}$ for $A(\boldsymbol{d}, \boldsymbol{s})$ and $\boldsymbol{e}_1$, the residual $\boldsymbol{r}_m = \boldsymbol{e}_1 - A\boldsymbol{x}_m$ is, according to Proposition 2.3, given by

$$\boldsymbol{r}_m = (-1)^m \xi_m \frac{|s_1 \cdot s_2 \cdots s_m|}{d_1 \cdot d_2 \cdots d_m} \boldsymbol{e}_{m+1}.$$

Therefore, the situation after restarting the method with new initial guess $\boldsymbol{x}_m$ is covered by Proposition 2.3 as well, where

$$c = \left| \frac{s_1 \cdots s_{m-1}}{d_1 \cdots d_{m-1}} \right| = f_m,$$

cf. also the proof of Lemma 2.4. It is immediately clear that choosing the next values in $\boldsymbol{d}$ and $\boldsymbol{s}$ (analogously to (2.4)) as

$$s_j = \frac{f_j}{f_{j-1}} d_j, \quad j = m+1, \ldots, \min\{2m, n\}$$

produces the residual norms $f_{m+1}, \ldots, f_{\min\{2m,n\}}$ in the next cycle of FOM(m) (or in the first $n - m$ iterations of this cycle if $2m > n$). This construction can be continued for a total of n iterations, until all values in $\boldsymbol{s}$ are fixed.

We now describe how to prescribe the eigenvalues of $A(\boldsymbol{d}, \boldsymbol{s})$ by choosing the coefficients in $\boldsymbol{d}$ accordingly. The characteristic polynomial of $A(\boldsymbol{d}, \boldsymbol{s})$ is given by

$$\chi_{A(\boldsymbol{d},\boldsymbol{s})}(\lambda) = (\lambda - d_1) \cdots (\lambda - d_n) - s_1 \cdots s_n. \tag{2.5}$$

ETNA
Kent State University
http://etna.math.kent.edu

To eliminate the dependency of the characteristic polynomial on $\boldsymbol{s}$, we note that multiplying all equations in (2.4) (and its counterparts in later restart cycles) yields

$$s_1 \cdots s_n = f_n \cdot d_1 \cdots d_n. \tag{2.6}$$

Therefore, we may rewrite the characteristic polynomial of a matrix $A(\boldsymbol{d}, \boldsymbol{s})$ generating the prescribed residual norm sequence $f_1, \ldots, f_n$ as

$$\chi_{A(\boldsymbol{d},\boldsymbol{s})}(\lambda) = (\lambda - d_1) \cdots (\lambda - d_n) - f_n \cdot d_1 \cdots d_n. \tag{2.7}$$

Prescribing the eigenvalues of $A(\boldsymbol{d}, \boldsymbol{s})$ therefore means choosing the values $d_1, \ldots, d_n$ such that the zeros of (2.7) are $\mu_1, \ldots, \mu_n$. This can be done as follows. Writing

$$(\lambda - \mu_1) \cdots (\lambda - \mu_n) = \lambda^n + \beta_{n-1}\lambda^{n-1} + \cdots + \beta_1\lambda + \beta_0,$$

we choose the components d_i of $\boldsymbol{d}$ as the n roots of the polynomial

$$\lambda^n + \beta_{n-1}\lambda^{n-1} + \cdots \beta_1\lambda + \widetilde{\beta}_0 \text{ with } \widetilde{\beta}_0 = \frac{\beta_0}{1 + (-1)^{n+1} f_n}, \tag{2.8}$$

assuming for the moment that $f_n \neq (-1)^n$, so that $\widetilde{\beta}_0$ is defined. Then, $(-1)^n d_1 \cdots d_n = \widetilde{\beta}_0$ and we have

$$\begin{aligned}
\chi_{A(\boldsymbol{d},\boldsymbol{s})}(\lambda) &= \lambda^n + \beta_{n-1}\lambda^{n-1} + \cdots + \beta_1\lambda + \widetilde{\beta}_0 - f_n \cdot d_1 \cdots d_n \\
&= \lambda^n + \beta_{n-1}\lambda^{n-1} + \cdots + \beta_1\lambda + \widetilde{\beta}_0 + (-1)^{n+1} f_n \widetilde{\beta}_0 \\
&= \lambda^n + \beta_{n-1}\lambda^{n-1} + \cdots + \beta_1\lambda + \beta_0 \left(\frac{1}{1 + (-1)^{n+1} f_n} + \frac{(-1)^{n+1} f_n}{1 + (-1)^{n+1} f_n} \right) \\
&= \lambda^n + \beta_{n-1}\lambda^{n-1} + \cdots + \beta_1\lambda + \beta_0,
\end{aligned}$$

showing that $A(\boldsymbol{d}, \boldsymbol{s})$ has the desired eigenvalues. In addition, the above construction implies that $d_i \neq 0, i = 1, \ldots, n$, so that all Hessenberg matrices $H_j^{(k)}$ are nonsingular and all FOM iterates are defined. This proves Theorem 2.1 for the case $q = n$ and $f_n > 0$, $f_n \neq (-1)^n$. We conclude the proof by commenting on how the special cases excluded so far can be handled.

The case $f_n = (-1)^n$ can be handled as follows: As a first step, replace the sequence $f_1, \ldots, f_n$ by the new sequence

$$\widetilde{f}_1 = \frac{f_1}{2}, \quad \widetilde{f}_2 = \frac{f_2}{2}, \quad \ldots, \quad \widetilde{f}_n = \frac{f_n}{2}. \tag{2.9}$$

With $f_n = (-1)^n$ replaced by $\widetilde{f}_n$, the value $\widetilde{\beta}_0$ in (2.8) is guaranteed to exist and one can construct the matrix $A(\boldsymbol{d}, \boldsymbol{s})$ as described in the previous part of the proof. When solving the system $A(\boldsymbol{d}, \boldsymbol{s})\boldsymbol{x} = \boldsymbol{e}_1$, with the matrix $A(\boldsymbol{d}, \boldsymbol{s})$ constructed for the modified sequence (2.9), the generated residual norms are

$$\|\boldsymbol{r}_j\|_2 = \widetilde{f}_j = \frac{f_j}{2}.$$

Thus, the residual norm sequence $f_1, \ldots, f_n$ can be generated with this matrix by using the right-hand side $\boldsymbol{b} = 2\boldsymbol{e}_1$ instead (implying $\|\boldsymbol{r}_0\|_2 = 2$, as $\boldsymbol{x}_0 = \boldsymbol{0}$).

For $q < n$ and $f_q > 0$ we can use exactly the same construction, setting the "unused" coefficients $s_{q+1}, \ldots, s_n$ to arbitrary values in such a way that (2.6) still holds (with f_n replaced by f_q).

ETNA
Kent State University
http://etna.math.kent.edu

If $f_q = 0$, then $s_q = 0$ and the characteristic polynomial (2.5) of $A(\boldsymbol{d}, \boldsymbol{s})$ simplifies to

$$\chi_{A(\boldsymbol{d},\boldsymbol{s})}(\lambda) = (\lambda - d_1)\cdots(\lambda - d_n),$$

such that the eigenvalues of $A(\boldsymbol{d}, \boldsymbol{s})$ are just the entries of $\boldsymbol{d}$, again allowing to freely prescribe them. The remaining entries $s_{q+1}, \ldots, s_n$ can be chosen arbitrarily (e.g., all equal to zero), as FOM(m) terminates after the qth iteration in this case. This concludes the proof of Theorem 2.1.

REMARK 2.5. According to Theorem 2.1, we can prescribe the residual norms produced by FOM(m) for (at most) n iterations. While not being able to prescribe residual norms in further iterations, we do indeed have full information on the convergence behavior of FOM(m) in later iterations when considering the matrices $A(\boldsymbol{d}, \boldsymbol{s})$ from (2.1), $\boldsymbol{b} = \boldsymbol{e}_1$ and $\boldsymbol{x}_0 = \boldsymbol{0}$. For the sake of simplicity, we again only consider the case that the first n residual norms are prescribed, and that $f_n \neq 0$. Proposition 2.3 also applies to the situation in which there have been more than n iterations, as the residual is still a multiple of a canonical unit vector then. Therefore, the residual norm for iteration j (possibly larger than n) fulfills

$$\|\boldsymbol{r}_j\|_2 = \frac{f_{j \bmod n}}{f_{j-1 \bmod n}}\|\boldsymbol{r}_{j-1}\|_2$$

i.e., the ratios between consecutive residuals are repeated cyclically. This information about later iterations is in contrast to the approach of [8, 21] for restarted GMRES, where nothing is known about the behavior of the method after n iterations. An interesting interpretation of the behavior of FOM(m) for iteration numbers exceeding n is that it behaves exactly as unrestarted FOM applied to the (infinite) bidiagonal matrix

$$\begin{bmatrix}
d_1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
s_1 & d_2 & \ddots & & & & & \\
0 & s_2 & \ddots & \ddots & & & & \\
\vdots & \ddots & \ddots & d_{n-1} & \ddots & & & \\
\vdots & & \ddots & s_{n-1} & d_n & \ddots & & \\
\vdots & & & \ddots & s_n & d_1 & \ddots & \\
\vdots & & & & \ddots & s_1 & d_2 & \ddots \\
\vdots & & & & & \ddots & \ddots & \ddots
\end{bmatrix},$$

which results from "gluing together" copies of $A(\boldsymbol{d}, \boldsymbol{s})$ with the entries s_n moved from the upper right corner of $A(\boldsymbol{d}, \boldsymbol{s})$ to the subdiagonal entry "connecting" two copies of $A(\boldsymbol{d}, \boldsymbol{s})$.

In light of Lemma 2.4—the relation between FOM(m) and unrestarted FOM for the matrices $A(\boldsymbol{d}, \boldsymbol{s})$—the result of Theorem 2.1 also holds for unrestarted FOM (with the obvious difference that $\|\boldsymbol{r}_n\|_2$ must be zero due to the finite termination property of FOM).

COROLLARY 2.6. *Let $n \in \mathbb{N}$, $1 \leq q \leq n$, $f_1, \ldots, f_{q-1} \in \mathbb{R}^+$, $f_q = 0$ and let $\mu_1, \ldots, \mu_n \in \mathbb{C} \setminus \{0\}$, not necessarily pairwise distinct. Then there exist a matrix $A \in \mathbb{C}^{n\times n}$ with* $\operatorname{spec}(A) = \{\mu_1, \ldots \mu_n\}$ *and vectors $\boldsymbol{b}, \boldsymbol{x}_0 \in \mathbb{C}^n$ such that the residuals $\boldsymbol{r}_j$ generated by j iterations of FOM for $A\boldsymbol{x} = \boldsymbol{b}$ with initial guess $\boldsymbol{x}_0$ satisfy*

$$\|\boldsymbol{r}_j\|_2 = f_j \ \text{ for } j = 1, \ldots, q.$$

ETNA
Kent State University
http://etna.math.kent.edu

3. Relation and differences to results on (restarted) GMRES. In this section, we briefly comment on the relation of Theorem 2.1 to similar results concerning the convergence of (restarted) GMRES from [2, 7, 8, 13, 21]. In [7, 13] it is shown that arbitrary (nonincreasing) residual norm sequences can be prescribed for unrestarted GMRES, for a matrix with any desired eigenvalues. In [9], the authors extend these results also to the case of breakdown in the underlying Arnoldi process.

There exists a well-known relation between the residual norms generated by FOM and those generated by GMRES. Precisely, provided that the jth FOM iterate for $A\boldsymbol{x} = \boldsymbol{b}$ is defined, it holds

$$\|\boldsymbol{r}_j^F\|_2 = \frac{\|\boldsymbol{r}_j^G\|_2}{\sqrt{1 - (\|\boldsymbol{r}_j^G\|_2 / \|\boldsymbol{r}_{j-1}^G\|_2)^2}}, \tag{3.1}$$

where $\boldsymbol{r}_j^F$ and $\boldsymbol{r}_j^G$ denote the residual of the jth FOM and GMRES iterate, respectively; see, e.g., [4–6]. Using this relation, Corollary 2.6 can be derived as a corollary of the results from [7, 9, 13] for GMRES as follows: Given values $f_1^F, \ldots, f_q^F$ to be prescribed as FOM residual norms, define the quantities

$$f_j^G := \frac{f_j^F}{\sqrt{1 + (f_j^F / f_{j-1}^G)^2}} \text{ with } f_0^G = 1. \tag{3.2}$$

Now construct a matrix A (with any desired eigenvalues) and a vector $\boldsymbol{b}$ which generate the sequence $f_1^G, \ldots, f_q^G$ of GMRES residual norms using the techniques from [7, 9, 13]. Then, by (3.1), A and $\boldsymbol{b}$ will generate the sequence $f_1^F, \ldots, f_q^F$ of FOM residual norms.

In [2], the authors give a complete parameterization of all matrices with prescribed eigenvalues and corresponding right-hand sides which exhibit a certain (unrestarted) GMRES convergence curve. Obviously, the matrices $A(\boldsymbol{d}, \boldsymbol{s})$ from (2.1) thus must belong to this parameterized class of matrices, corresponding to eigenvalues $\mu_1, \ldots \mu_n$ and GMRES residual norms (3.2). This gives rise to another alternative proof of our main result, first showing that $A(\boldsymbol{d}, \boldsymbol{s})$ belongs to this class of matrices and then using the result of Lemma 2.4 to conclude that the convergence curve of restarted FOM is the same as that of unrestarted FOM. Again, this approach would not have given the information on iterations exceeding n given in Remark 2.5.

In much the same way, the result on restarted GMRES from [21] can be transferred to a result on restarted FOM. In [21], the authors construct linear systems for which the residual norm at the end of the first $k \leq \frac{n}{m}$ cycles of restarted GMRES can be prescribed. Again using (3.2), one can easily construct systems for which the residual norms at the end of the first $\lfloor \frac{n}{m} \rfloor$ cycles of restarted FOM can be prescribed. Our construction, however, allows for prescribing the residual norms in *all* iterations of restarted FOM, not just at the end of each cycle. Therefore, Theorem 2.1 cannot be derived from the results of [21].

In fact, using (3.1), one can use our construction for generating matrices with arbitrary nonzero spectrum which produce a prescribed, decreasing convergence curve of restarted GMRES in the first n iterations (the case of stagnation needs to be handled separately, as it corresponds to a FOM iterate not being defined, see, e.g., [4], a case we do not consider here). In the work [8]—which was published as a preprint simultaneously to this paper—another construction for prescribing the residual norms in the first n iterations of restarted GMRES is presented, which also deals with stagnation. In addition, in [8], it is investigated which convergence curves cannot be exhibited by restarted GMRES (so-called *inadmissible convergence curves*), as it is no longer true that any nonincreasing convergence curve is possible as soon as stagnation is allowed. Therefore, considering the behavior of restarted GMRES (or restarted FOM, although the connection is not pointed out in [8]), the approach from [8] gives more general results than what we presented here. However, our construction gives rise to a further result on restarted GMRES, cf. Theorem 3.1, which cannot be derived from the results of [8], and thus complements the analysis presented in [8].

Due to the direct relation between the residual norms generated by FOM and GMRES, Remark 2.5 also applies (in a modified form) to restarted GMRES. We can use this to answer an open question raised in the conclusions section of [21]. There, the authors ask whether it is possible to give convergence estimates based on spectral information for *later cycles* of restarted GMRES, i.e., for $k > \frac{n}{m}$. By using our construction, and in particular the statement of Remark 2.5, we can construct a matrix A for which the ratio $\|\boldsymbol{r}_\ell^G\|_2/\|\boldsymbol{r}_0^G\|_2$ can be arbitrarily close to one, for any $\ell \in \mathbb{N}$ and with any prescribed eigenvalues. Therefore, it is impossible to use spectral information alone to give bounds for the residual norms generated by either FOM(m) or GMRES(m) in "later" restart cycles. We end this section by stating the precise result for GMRES(m) in the following theorem (a corresponding result for FOM(m) directly follows from Remark 2.5).

THEOREM 3.1. *Let $n, m, \ell \in \mathbb{N}$, $m \leq n-1$, let $\mu_1, \ldots, \mu_n \in \mathbb{C} \setminus \{0\}$, not necessarily distinct, and let $0 \leq \delta < 1$. Then there exist a matrix $A \in \mathbb{C}^{n\times n}$ with $\operatorname{spec}(A) = \{\mu_1, \ldots \mu_n\}$ and vectors $\boldsymbol{x}_0, \boldsymbol{b} \in \mathbb{C}^n$ such that the residual $\boldsymbol{r}_\ell^G = \boldsymbol{b} - A\boldsymbol{x}_\ell^G$ generated by ℓ (with ℓ possibly larger than n) iterations of GMRES(m) for $A\boldsymbol{x} = \boldsymbol{b}$ with initial guess $\boldsymbol{x}_0$ satisfies*

$$\|\boldsymbol{r}_\ell^G\|_2/\|\boldsymbol{r}_0^G\|_2 \geq \delta.$$

Proof. By Theorem 2.1 there exist a matrix $A \in \mathbb{C}^{n\times n}$ with $\operatorname{spec}(A) = \{\mu_1, \ldots, \mu_n\}$ and vectors $\boldsymbol{b}, \boldsymbol{x}_0 \in \mathbb{C}^n$ such that the residuals $\boldsymbol{r}_j^F, j = 1, \ldots, n$, produced by the first n iterations of FOM(m) fulfill

$$\|\boldsymbol{r}_j^F\|_2 = \rho^j \text{ with } \rho = \frac{\delta^{1/\ell}}{(1-\delta^{2/\ell})^{1/2}} \text{ for } j = 1, \ldots, n. \tag{3.3}$$

By Remark 2.5, we then have that (3.3) also holds for $j > n$. We can rephrase this as

$$\|\boldsymbol{r}_{j-1}^F\|_2 = \frac{1}{\rho}\|\boldsymbol{r}_j^F\|_2 \text{ for all } j \in \mathbb{N}. \tag{3.4}$$

By relation (3.1), two consecutive residual norms generated by GMRES(m) for A, $\boldsymbol{b}$ and $\boldsymbol{x}_0$ fulfill

$$\begin{aligned}
\frac{\|\boldsymbol{r}_j^G\|_2}{\|\boldsymbol{r}_{j-1}^G\|_2} &= \frac{\|\boldsymbol{r}_j^F\|_2}{\|\boldsymbol{r}_{j-1}^G\|_2\sqrt{1+(\|\boldsymbol{r}_j^F\|_2/\|\boldsymbol{r}_{j-1}^G\|_2)^2}} \\
&= \frac{\|\boldsymbol{r}_j^F\|_2}{\sqrt{\|\boldsymbol{r}_{j-1}^G\|_2^2+\|\boldsymbol{r}_j^F\|_2^2}} = \frac{\|\boldsymbol{r}_j^F\|_2}{\sqrt{\frac{\|\boldsymbol{r}_{j-1}^F\|_2^2}{1+(\|\boldsymbol{r}_{j-1}^F\|_2/\|\boldsymbol{r}_{j-2}^G\|_2)^2}+\|\boldsymbol{r}_j^F\|_2^2}} \\
&\geq \frac{\|\boldsymbol{r}_j^F\|_2}{\sqrt{\|\boldsymbol{r}_{j-1}^F\|_2^2+\|\boldsymbol{r}_j^F\|_2^2}}\,.
\end{aligned} \tag{3.5}$$

Inserting (3.4) into the right hand side of (3.5), we find

$$\frac{\|\boldsymbol{r}_j^G\|_2}{\|\boldsymbol{r}_{j-1}^G\|_2} \geq \frac{\|\boldsymbol{r}_j^F\|_2}{\sqrt{\|\boldsymbol{r}_j^F\|_2^2+\frac{1}{\rho^2}\|\boldsymbol{r}_j^F\|_2^2}} = \frac{1}{\sqrt{1+\frac{1}{\rho^2}}}\,. \tag{3.6}$$

By repeated application of (3.6) for all $j \leq \ell$, we find

$$\|\boldsymbol{r}_\ell^G\|_2/\|\boldsymbol{r}_0^G\|_2 = \left(\|\boldsymbol{r}_\ell^G\|_2/\|\boldsymbol{r}_{\ell-1}^G\|_2\right)\cdots\left(\|\boldsymbol{r}_1^G\|_2/\|\boldsymbol{r}_0^G\|_2\right) \geq \frac{1}{(1+\frac{1}{\rho^2})^{\ell/2}}. \tag{3.7}$$

The result follows from (3.7) by noting that $(1+\frac{1}{\rho^2})^{\ell/2} = \frac{1}{\delta}$. $\square$

4. Approximating $g(A)\boldsymbol{b}$ by the restarted Arnoldi method. Restarted FOM is rarely used in practice (although there exist situations where it is considered useful, e.g., when solving families of shifted linear systems; see [19]) as restarted GMRES is typically the method of choice for non-Hermitian linear systems. However, the (restarted) Arnoldi method for approximating $g(A)\boldsymbol{b}$, the action of a matrix function on a vector (see, e.g., [1, 10, 11]) can be interpreted as implicitly performing (restarted) FOM for families of shifted linear systems if the function g has an integral representation involving a resolvent function. This is, e.g., the case for *Stieltjes functions* [3, 14] defined by the Riemann–Stieltjes integral

$$g(z) = \int_0^\infty \frac{1}{z+t}\,\mathrm{d}\alpha(t),$$

where α is a monotonically increasing, nonnegative function. Examples of Stieltjes functions include $g(z) = z^{-\sigma}$ for $\sigma \in (0,1]$ or $g(z) = \log(1+z)/z$. One can show that the restarted Arnoldi approximation (after k cycles with restart length m) for $g(A)\boldsymbol{b}$ is given as

$$\boldsymbol{g}_m^{(k)} = \int_0^\infty \boldsymbol{x}_m^{(k)}(t)\,\mathrm{d}\alpha(t), \tag{4.1}$$

when g is a Stieltjes function, where $\boldsymbol{x}_m^{(k)}(t)$ denotes the iterate obtained from k cycles of FOM(m) for the shifted linear system

$$(A+tI)\boldsymbol{x}(t) = \boldsymbol{b} \tag{4.2}$$

with initial guess $\boldsymbol{x}_0(t) = \boldsymbol{0}$; see [10, 12]. In [12], a convergence analysis of the restarted Arnoldi method for A Hermitian positive definite and g a Stieltjes function is given. There it is proved that the method always converges to $g(A)\boldsymbol{b}$, independent of the restart length m and that the asymptotic convergence factor of the method depends on the condition number, the ratio of the largest and smallest eigenvalues, of A.

Motivated by the result of Theorem 2.1, we conjecture that for non-normal A it is not possible to analyze the behavior of the restarted Arnoldi method based solely on spectral information.

CONJECTURE 4.1. *Let g be a Stieltjes function, $m, n \in \mathbb{N}$ with $m \leq n-1$ and let $\mu_1, \ldots, \mu_n \in \mathbb{C} \setminus \mathbb{R}_0^-$, not necessarily distinct. Then there exist a matrix $A \in \mathbb{C}^{n\times n}$ and a vector $\boldsymbol{b} \in \mathbb{C}^n$ such that* $\mathrm{spec}(A) = \{\mu_1, \ldots, \mu_n\}$ *and the iterates of the restarted Arnoldi method with restart length m do not converge to $g(A)\boldsymbol{b}$.*

Intuitively, the statement of Conjecture 4.1 is plausible in light of the analysis presented for FOM(m) in Section 2 and the characterization (4.1) of the restarted Arnoldi approximations $\boldsymbol{g}_m^{(k)}$. It is easily possible to construct the matrix A in such a way that the (implicitly generated) FOM(m) iterates $\boldsymbol{x}_m^{(k)}(t)$ for t in some interval $[t_1, t_2]$ diverge (by prescribing increasing residual norms), and one would surely expect the corresponding Arnoldi approximation $\boldsymbol{g}_m^{(k)}$ for $g(A)\boldsymbol{b}$ to inherit this behavior and diverge in this case as well. We give a small numerical example which illustrates this (a similar example was presented in [12]). We construct a matrix $A(\boldsymbol{d}, \boldsymbol{s}) \in \mathbb{C}^{6\times 6}$ with $\mathrm{spec}(A(\boldsymbol{d}, \boldsymbol{s})) = \{1, 2, \ldots, 6\}$ such that the residual norms

ETNA
Kent State University
http://etna.math.kent.edu

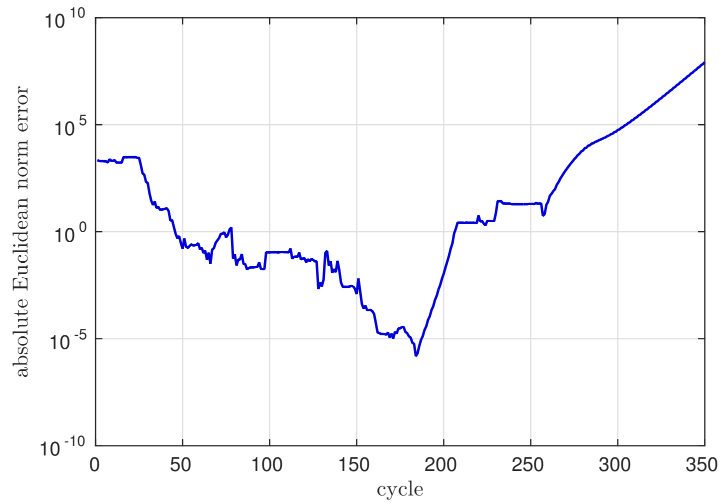

FIG. 4.1. *Convergence curve for approximating* $A(\boldsymbol{d},\boldsymbol{s})^{-1/2}\boldsymbol{e}_1$*, where* $A(\boldsymbol{d},\boldsymbol{s}) \in \mathbb{C}^{6\times 6}$ *with* $\operatorname{spec}(A(\boldsymbol{d},\boldsymbol{s})) = \{1,2,\ldots,6\}$ *is constructed such that the FOM(*m*) residual norms increase by a factor of two from one iteration to the next. The restart length is chosen as* $m = 5$*.*

produced by restarted FOM increase by a factor of two from one cycle to the next. The resulting convergence curve when approximating $A^{-1/2}\boldsymbol{e}_1$ by the restarted Arnoldi method with restart length $m = 5$ is depicted in Figure 4.1. The observed behavior can be explained as follows: As the factor by which the residual norm increases depends continuously on the values in $\boldsymbol{d}$, there exists an interval $[0, t_0]$ such that the residual norms for the shifted linear systems

$$(A(\boldsymbol{d},\boldsymbol{s}) + tI)\boldsymbol{x}(t) = \boldsymbol{b}, \quad t \in [0, t_0]$$

are also increasing. One can expect FOM(m) to converge for those linear systems corresponding to large values of t, as those are close to trivial. The error norm is therefore decreasing initially, until the FOM iterates for the underlying (implicitly solved) linear systems with large t all have converged. From this moment on, the divergence of the FOM iterates for the systems corresponding to small shifts becomes visible and the error norm begins to increase. Thus, the convergence curve shown in Figure 4.1 is in complete agreement with what one would expect motivated by our theory.

The difficulty in proving Conjecture 4.1 in the setting of this paper is the following: We only made statements about FOM *residual norms*, but not about the actual error vectors. When approximating $g(A)\boldsymbol{b}$, it immediately follows from (4.1) that

$$g(A)\boldsymbol{b} - \boldsymbol{g}_m^{(k)} = \int_0^\infty \boldsymbol{d}_m^{(k)}(t)\,\mathrm{d}\alpha(t),$$

where $\boldsymbol{d}_m^{(k)}(t) = \boldsymbol{x}^*(t) - \boldsymbol{x}_m^{(k)}(t)$ are the errors of the FOM(m) iterates for the systems (4.2). Surely, if $\|\boldsymbol{r}_m^{(k)}(t)\|_2 \to \infty$ as $k \to \infty$ for $t \in [t_1, t_2]$, it follows $\|\boldsymbol{d}_m^{(k)}(t)\|_2 \to \infty$ as $k \to \infty$. However, this does *not* imply that

$$\left\| \int_{t_1}^{t_2} \boldsymbol{d}_m^{(k)}(t)\,\mathrm{d}\alpha(t) \right\|_2 \to \infty \text{ as } k \to \infty,$$

as we do not have any information about the entries of $\boldsymbol{d}_m^{(k)}(t)$, and their integral might be zero despite all vectors being nonzero (and of possibly very large norm). Therefore, one needs additional information on the entries of the error vectors, or some completely different approach, for proving the conjecture.

5. Conclusions. We have shown that (and how) it is possible to construct a matrix $A \in \mathbb{C}^{n\times n}$ with arbitrary nonzero eigenvalues and vectors $\boldsymbol{b}, \boldsymbol{x}_0 \in \mathbb{C}^n$ such that the norms of the first n residuals of FOM(m) for $A\boldsymbol{x} = \boldsymbol{b}$ with initial guess $\boldsymbol{x}_0$ attain any desired (finite) values, indicating that convergence analysis of FOM(m) based solely on spectral information is not possible for non-normal A. In addition, we have pointed out the connection of our results to results on (restarted) GMRES and addressed the open question whether a convergence analysis based on spectral information is possible for restarted GMRES in "later iterations" (exceeding the matrix dimension). While not being able to freely prescribe residual norms for later iterations, our construction gives full information on these norms and allows us to find matrices having any prescribed nonzero eigenvalues for which the reduction of the residual norm is arbitrarily small throughout any number of iterations (exceeding n), so that also in this setting no (nontrivial) convergence estimates can be given based on spectral information. We also briefly commented on extending our result to the approximation of $g(A)\boldsymbol{b}$, the action of a Stieltjes matrix function on a vector. Intuition and numerical evidence suggest that a similar result, presented as a conjecture, also holds in this case.

Acknowledgment. The author wishes to thank Zdeněk Strakoš and an anonymous referee for their careful reading of the manuscript and their valuable suggestions which helped to greatly improve the presentation and led to the inclusion of Section 3 as well as the refinement of several results presented in the paper.

REFERENCES

[1] M. AFANASJEW, M. EIERMANN, O. G. ERNST, AND S. GÜTTEL, *Implementation of a restarted Krylov subspace method for the evaluation of matrix functions*, Linear Algebra Appl., 429 (2008), pp. 2293–2314.
[2] M. ARIOLI, V. PTÁK, AND Z. STRAKOŠ, *Krylov sequences of maximal length and convergence of GMRES*, BIT, 38 (1998), pp. 636–643.
[3] C. BERG AND G. FORST, *Potential Theory on Locally Compact Abelian Groups*, Springer, New York, 1975.
[4] P. N. BROWN, *A theoretical comparison of the Arnoldi and GMRES algorithms*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 58–78.
[5] J. CULLUM, *Iterative methods for solving $Ax = b$, GMRES/FOM versus QMR/BiCG*, Adv. Comput. Math., 6 (1996), pp. 1–24.
[6] J. CULLUM AND A. GREENBAUM, *Relations between Galerkin and norm-minimizing iterative methods for solving linear systems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 223–247.
[7] J. DUINTJER TEBBENS AND G. MEURANT, *Any Ritz value behavior is possible for Arnoldi and for GMRES*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 958–978.
[8] ———, *On the admissible convergence curves for restarted GMRES*, Tech. Report NCMM/2014/23, Nečas Center for Mathematical Modeling, Prague, Czech Republic, 2014.
[9] ———, *Prescribing the behavior of early terminating GMRES and Arnoldi iterations*, Numer. Algorithms, 65 (2014), pp. 69–90.
[10] M. EIERMANN AND O. G. ERNST, *A restarted Krylov subspace method for the evaluation of matrix functions*, SIAM J. Numer. Anal., 44 (2006), pp. 2481–2504.
[11] A. FROMMER, S. GÜTTEL, AND M. SCHWEITZER, *Efficient and stable Arnoldi restarts for matrix functions based on quadrature*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 661–683.
[12] ———, *Convergence of restarted Krylov subspace methods for Stieltjes functions of matrices*, SIAM. J. Matrix Anal. Appl., 35 (2014), pp. 1602–1624.
[13] A. GREENBAUM, V. PTÁK, AND Z. STRAKOŠ, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469.
[14] P. HENRICI, *Applied and Computational Complex Analysis. Vol. 2*, Wiley, New York, 1977.
[15] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436.

[16] Y. SAAD, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comp., 37 (1981), pp. 105–126.
[17] Y. SAAD, *Iterative Methods for Sparse Linear Systems, 2nd Edition*, SIAM, Philadelphia, 2003.
[18] 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.
[19] V. SIMONCINI, *Restarted full orthogonalization method for shifted linear systems*, BIT, 43 (2003), pp. 459–466.
[20] G. W. STEWART, *Matrix Algorithms Volume II: Eigensystems*, SIAM, Philadelphia, 2001.
[21] E. VECHARYNSKI AND J. LANGOU, *Any admissible cycle-convergence behavior is possible for restarted GMRES at its initial cycles*, Numer. Linear Algebra Appl., 18 (2011), pp. 499–511.