SIAM J. MATRIX ANAL. APPL.
Vol. 33, No. 3, pp. 958–978

ANY RITZ VALUE BEHAVIOR IS POSSIBLE FOR ARNOLDI AND FOR GMRES*

JURJEN DUINTJER TEBBENS† AND GÉRARD MEURANT‡

Abstract. We show that arbitrary convergence behavior of Ritz values is possible in the Arnoldi method, and we give two parametrizations of the class of matrices with initial Arnoldi vectors that generate prescribed Ritz values (in all iterations). The second parametrization enables us to prove that any GMRES residual norm history is possible with any prescribed Ritz values (in all iterations), provided that we treat the stagnation case appropriately.

Key words. Ritz values, Arnoldi process, Arnoldi method, GMRES method, prescribed convergence, interlacing properties

AMS subject classifications. 65F15, 65F10, 65F18, 15A18

DOI. 10.1137/110843666

1. Introduction.

Let A be a nonsingular matrix of order n and b a nonzero n-dimensional vector. The Arnoldi process [3] reduces A to upper Hessenberg form by a particular type of Gram–Schmidt orthogonalization for the vectors $b, Ab, A^2b, \ldots$. At each step of the process, one matrix-vector multiplication with A is performed, and one row and one column are appended to the previous Hessenberg matrix. The process is well suited to iterative methods with large sparse matrices A. Two popular methods for extracting approximate solutions from the generated Hessenberg matrices are the generalized minimal residual (GMRES) method [40] for solving the linear system $Ax = b$ and the Arnoldi method (see, e.g., [38, 39]) for computing the eigenvalues and eigenvectors of A.

The Arnoldi process can be seen as a generalization to non-Hermitian matrices of the Lanczos process for tridiagonalization of Hermitian matrices [24]. The Lanczos process is at the basis of the conjugate gradients (CG) method [23, 25] for Hermitian positive definite linear systems and of the Lanczos method for Hermitian eigenproblems [24]. In this sense GMRES is a generalization of CG (even though the l_2 norm of the residual is not minimized in CG), and the Arnoldi method is a generalization of the Lanczos method. As convergence of the CG and Lanczos methods are well understood, it is natural to take the convergence theory of these methods as a starting point for explaining the behavior of the GMRES and Arnoldi methods. In the CG method, the convergence behavior is dictated by the distribution of the eigenvalues of the matrix. In practice, the same is often observed for the GMRES method, but, with possibly nonnormal input matrices, the situation becomes more subtle. For example, Greenbaum and Strakoš [22] proved that if a residual norm convergence

*Received by the editors August 8, 2011; accepted for publication (in revised form) by Q. Ye May 3, 2012; published electronically September 5, 2012.
http://www.siam.org/journals/simax/33-3/84366.html

†Institute of Computer Science, Academy of Sciences of the Czech Republic, 18 207 Praha 8-Libeň, Czech Republic (duintjertebbens@cs.cas.cz). This author's work is part of the Institutional Research Plan AV0Z10300504 and was supported by project IAA100300802 of the Grant Agency of the ASCR and by project M100300901 of the institutional support of the ASCR.

‡30 rue du sergent Bauchat, 75012 Paris, France (gerard.meurant@gmail.com). The work on this paper was started in 2010 during this author's visit to the Nečas Center of Charles University in Prague supported by a grant from the Jindrich Nečas Center for Mathematical Modeling, project LC06052, financed by MSMT.

curve is generated by GMRES, the same curve can be obtained with a matrix having prescribed nonzero eigenvalues (see [12, Lemma 6.9] for an analogue on prescribed nonzero singular values). Greenbaum, Pták, and Strakoš [21] complemented their result by proving that *any* nonincreasing sequence of residual norms can be given by GMRES (a similar result for residual norms at the end of restart cycles in the restarted GMRES method can be found in [47]). Furthermore, in Arioli, Pták, and Strakoš [2] a complete parametrization was given of all pairs $\{A, b\}$ generating a prescribed residual norm convergence curve and such that A has a prescribed spectrum. The results in these papers show that the GMRES residual norm convergence need not, in general, depend on the eigenvalues of A alone. Other objects, mostly closely related to eigenvalues, have been considered to explain convergence, for example, the pseudospectrum [44], the field of values [11], or the numerical polynomial hull [20]. In [46] it was suggested that convergence of the eigenvalues of the Hessenberg matrices generated in the Arnoldi process (the so-called Ritz values) to eigenvalues of A will often explain the acceleration of convergence of GMRES.

A fundamental tool in the convergence analysis of the Lanczos method for Hermitian eigenproblems is the interlacing property for the eigenvalues of the subsequently generated tridiagonal matrices. It enables one to prove, among other things, the persistence theorem on stabilization of Ritz values (see, e.g., [32, 33, 34] or [31]). There are several generalizations of the interlacing property to normal matrices; see, e.g., [16, 1] or the papers [27, 14] with geometric interpretations. However, just as for GMRES, potentially nonnormal input matrices make convergence analysis of the Arnoldi method delicate. There is no interlacing property for the principal submatrices of general nonnormal matrices; see [42] for a thorough discussion on this topic and its relation to the field of Lie algebra. In [9, 10] one finds a sufficient and necessary condition for prescribing arbitrary eigenvalues of (not necessarily principal) submatrices of general non-Hermitian matrices. For a detailed spectral analysis of nonnormal Hessenberg matrices and their principal submatrices, see also [49].

Since the GMRES and the Arnoldi methods are closely related through the Arnoldi orthogonalization process, a naturally arising question is whether a result, similar to the results of Arioli, Greenbaum, Pták, and Strakoš, on arbitrary convergence behavior of the Arnoldi method can be proved. By arbitrary convergence behavior of the Arnoldi method, we mean the ability to prescribe *all* Ritz values from the very first until the very last iteration (we do not consider convergence to eigenvectors). In this paper we will give a parametrization of the class of all matrices and initial Arnoldi vectors that generates prescribed Ritz values. Besides this result on arbitrary convergence behavior of the Arnoldi method, we derive a parametrization that allows us to characterize all pairs $\{A, b\}$ generating arbitrary convergence behavior of *both* GMRES and Arnoldi. The Ritz values generated in the GMRES method therefore do not, in general, have any influence on the generated residual norms.

The paper is organized as follows: In the remainder of this section we introduce some notation, in particular the notation used in [2], which we adopt, and we recall the parametrization given in [2]. In section 2 we give a parametrization of the class of matrices and initial Arnoldi vectors that generates prescribed Ritz values. Section 3 reformulates the parametrization in order to parametrize the pairs $\{A, b\}$ generating arbitrary behavior of GMRES and Arnoldi at the same time. We close with a brief discussion of our results and some words on future work.

1.1. Notation. We will use the following parametrization of matrices and right-hand sides giving prescribed spectrum and prescribed convergence of the GMRES method (see Theorem 2.1 and Corollary 2.4 of [2]).

THEOREM 1.1. *Assume that we are given n positive numbers*

$$f(0) \geq f(1) \geq \cdots \geq f(n-1) > 0$$

and n complex numbers $\lambda_1, \ldots, \lambda_n$ all different from 0. Let A be a matrix of order n and b an n-dimensional vector. The following assertions are equivalent:

1. *The spectrum of A is $\{\lambda_1, \ldots, \lambda_n\}$, and GMRES applied to A and b with zero initial guess yields residuals $r^{(k)}$, $k = 0, \ldots, n-1$, such that*

$$\|r^{(k)}\| = f(k), \quad k = 0, \ldots, n-1.$$

2. *The matrix A is of the form*

$$A = WYC^{(n)}Y^{-1}W^*$$

and $b = Wh$, where W is a unitary matrix; Y is given by

$$Y = \begin{bmatrix} h & R \\ & 0 \end{bmatrix}, \tag{1.1}$$

with R being a nonsingular upper triangular matrix of order $n-1$ and h a vector such that

$$h = [\eta_1, \ldots, \eta_n]^T, \qquad \eta_k = (f(k-1)^2 - f(k)^2)^{1/2}, \quad k < n, \quad \eta_n = f(n-1); \tag{1.2}$$

and $C^{(n)}$ is the companion matrix corresponding to the polynomial $q(\lambda)$ defined as

$$q(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n) = \lambda^n + \sum_{j=0}^{n-1} \alpha_j \lambda^j,$$

$$C^{(n)} = \begin{bmatrix} 0 & & -\alpha_0 \\ & \ddots & \vdots \\ & I_{n-1} & -\alpha_{n-1} \end{bmatrix}.$$

Furthermore, we will denote by e_j the jth column of the identity matrix of appropriate order. For a matrix M, the leading principal submatrix of order k will be denoted by M_k. With "the subdiagonal" and "subdiagonal entries" we mean the (entries on the) first diagonal under the main diagonal. Throughout the paper we assume exact arithmetics, and we also assume that the investigated Arnoldi processes do not terminate before the nth iteration. This means that the input matrix must be nonderogatory. Note that Theorem 1.1 assumes this situation. The case of early termination will be treated in a forthcoming paper.

2. Prescribed convergence of Ritz values in Arnoldi's method. Consider the kth iteration of an Arnoldi process with a matrix A and initial vector b where an upper Hessenberg matrix H_k (with entries $h_{i,j}$) is generated satisfying

$$AV^{(k)} = V^{(k)}H_k + h_{k+1,k}\, v_{k+1} e_k^T, \quad k < n, \tag{2.1}$$

with ${V^{(k)}}^* V^{(k)} = I_k$, $V^{(k)} e_1 = b/\|b\|$, and ${V^{(k)}}^* v_{k+1} = 0$, $V^{(k)}$ being the matrix whose columns are the basis vectors $v_1, \ldots, v_k$ of the kth Krylov subspace $\mathcal{K}_k(A, b) \equiv \text{span}\{b, Ab, \ldots, A^{k-1}b\}$. The eigenvalues of H_k give the k-tuple

$$\mathcal{R}^{(k)} = (\rho_1^{(k)}, \ldots, \rho_k^{(k)})$$

of the k (not necessarily distinct) Ritz values generated at the kth iteration of Arnoldi's method. We denote by $\mathcal{R}$ the set

$$\mathcal{R} \equiv \{\mathcal{R}^{(1)}, \mathcal{R}^{(2)}, \ldots, \mathcal{R}^{(n)}\}$$

representing all $(n+1)n/2$ generated Ritz values. We also use $\mathcal{S}$ for the *strict* Ritz values without the spectrum of A, i.e.,

$$\mathcal{S} \equiv \mathcal{R} \setminus \mathcal{R}^{(n)},$$

and we will denote the (not necessarily distinct) eigenvalues of the input matrix by $\lambda_1, \ldots, \lambda_n$, i.e.,

$$\mathcal{R}^{(n)} = (\lambda_1, \ldots, \lambda_n).$$

In this section we investigate whether the Arnoldi method can generate arbitrary Ritz values in all iterations. The Ritz values in the Arnoldi method are eigenvalues of the leading principal submatrices of upper Hessenberg matrices with positive real subdiagonal entries. Prescribing the set $\mathcal{R}$ is possible only if there exist, at all, Hessenberg matrices with positive subdiagonal entries where the eigenvalues of all the leading principal submatrices can be prescribed. Parlett and Strang proved that there is a unique upper Hessenberg matrix with the entry one along the subdiagonal such that all leading principal submatrices have arbitrary prescribed eigenvalues; see [36, Theorem 3]. We give here a characterization of this unique matrix, which we denote with $H(\mathcal{R})$, that shows how it is constructed from the prescribed Ritz values.

PROPOSITION 2.1. *Let the set*

$$\begin{aligned}\mathcal{R} = \{ & \rho_1^{(1)}, \\ & (\rho_1^{(2)}, \rho_2^{(2)}), \\ & \vdots \\ & (\rho_1^{(n-1)}, \ldots, \rho_{n-1}^{(n-1)}), \\ & (\lambda_1, \ldots\ldots\ldots, \lambda_n) \}\end{aligned}$$

represent any choice of $n(n+1)/2$ complex Ritz values, and denote the $k \times k$ companion matrix of the polynomial with roots $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ by $C^{(k)}$. If we define the kth column of the unit upper triangular matrix $U(\mathcal{S})$ through

$$U(\mathcal{S})\, e_1 = e_1, \qquad U(\mathcal{S})\, e_k = \begin{bmatrix} -e_1^T C^{(k-1)} e_{k-1} \\ \vdots \\ -e_{k-1}^T C^{(k-1)} e_{k-1} \\ 1 \\ 0 \\ \vdots \end{bmatrix}, \qquad k = 2, \ldots, n, \tag{2.2}$$

then the unique upper Hessenberg matrix $H(\mathcal{R})$ with the entry one along the subdiagonal and with the spectrum $\lambda_1, \ldots, \lambda_n$ such that the kth leading principal submatrix has eigenvalues $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ for all $k = 1, \ldots, n-1$ is

$$H(\mathcal{R}) = U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}). \tag{2.3}$$

Proof. We will show that the spectrum of the $k \times k$ leading principal submatrix of $H(\mathcal{R})$ is $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ (uniqueness of $H(\mathcal{R})$ was shown in [36] and will also be proved later). Let U_k denote the $k \times k$ leading principal submatrix of $U(\mathcal{S})$, and let, for $j > k$, $\tilde{u}_j$ denote the vector of the first k entries of the jth column of $U(\mathcal{S})^{-1}$. The spectrum of the $k \times k$ leading principal submatrix of $H(\mathcal{R})$ is the spectrum of

$$[I_k, 0]\, U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) \begin{bmatrix} I_k \\ 0 \end{bmatrix} = [U_k^{-1}, \tilde{u}_{k+1}, \ldots, \tilde{u}_n] \begin{bmatrix} 0 \\ U_k \\ 0 \end{bmatrix} = [U_k^{-1}, \tilde{u}_{k+1}] \begin{bmatrix} 0 \\ U_k \end{bmatrix}.$$

It is also the spectrum of the matrix

$$U_k [U_k^{-1}, \tilde{u}_{k+1}] \begin{bmatrix} 0 \\ U_k \end{bmatrix} U_k^{-1} = [I_k, U_k \tilde{u}_{k+1}] \begin{bmatrix} 0 \\ I_k \end{bmatrix},$$

which is a companion matrix with last column $U_k \tilde{u}_{k+1}$. From

$$\begin{aligned} e_{k+1} = U_{k+1} U_{k+1}^{-1} e_{k+1} &= \begin{bmatrix} U_k & -C^{(k)} e_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} U_k^{-1} & \tilde{u}_{k+1} \\ 0 & 1 \end{bmatrix} e_{k+1} \\ &= \begin{bmatrix} U_k \tilde{u}_{k+1} - C^{(k)} e_k \\ 1 \end{bmatrix} \end{aligned}$$

we obtain $U_k \tilde{u}_{k+1} = C^{(k)} e_k$. ∎

Note that (2.3) represents a similarity transformation separating the spectrum of $H(\mathcal{R})$ from the strict Ritz values $\mathcal{S}$ of $H(\mathcal{R})$. The matrix $U(\mathcal{S})$ transforms the companion matrix whose strict Ritz values are all zero to a Hessenberg matrix with arbitrary Ritz values, and it is itself composed of (parts of) companion matrices. We will call $U(\mathcal{S})$, for lack of a better name, the *Ritz value companion transform.*

Clearly, the Ritz values generated in the Arnoldi method can exhibit any convergence behavior: It suffices to apply the Arnoldi process with the initial Arnoldi vector e_1 and the matrix $H(\mathcal{R})$ with arbitrary prescribed Ritz values from Proposition 2.1. Then the method generates the Hessenberg matrix $H(\mathcal{R})$ itself. If the prescribed Ritz values occur in complex conjugate pairs, then the Ritz value companion transform $U(\mathcal{S})$ and the Hessenberg matrix $H(\mathcal{R})$ in (2.3) are real, and the Arnoldi process runs without complex arithmetics.

We next look for a parametrization of the class of all matrices and initial Arnoldi vectors generating given Ritz values. From $H(\mathcal{R})$ we can easily obtain an upper Hessenberg matrix whose leading principal submatrices have the same prescribed eigenvalues but with arbitrary positive values along the subdiagonal. Let $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$ be given positive real numbers, and consider the similarity transformation

$$H \equiv \operatorname{diag}(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1} \sigma_j)\, H(\mathcal{R}) \left(\operatorname{diag}(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1} \sigma_j)\right)^{-1}.$$

Then the subdiagonal of H has the entries $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$, and all leading principal submatrices of H are similar to the corresponding leading principal submatrices of $H(\mathcal{R})$. The following theorem shows the uniqueness of H.

THEOREM 2.2. *Let the set*

$$\begin{aligned}\mathcal{R} = \{\, & \rho_1^{(1)}, \\ & (\rho_1^{(2)}, \rho_2^{(2)}), \\ & \vdots \\ & (\rho_1^{(n-1)}, \ldots, \rho_{n-1}^{(n-1)}), \\ & (\lambda_1, \ldots\ldots\ldots, \lambda_n)\,\}\end{aligned}$$

represent any choice of $n(n+1)/2$ *complex Ritz values, and let*

$$D_\sigma = \operatorname{diag}(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1}\sigma_j),$$

where $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$ *are* $n-1$ *positive real numbers. Then*

$$H = D_\sigma\, H(\mathcal{R}) D_\sigma^{-1}$$

is the unique Hessenberg matrix H *with subdiagonal entries*

$$h_{k+1,k} = \sigma_k, \qquad k = 1, \ldots, n-1,$$

with eigenvalues $\lambda_1, \ldots, \lambda_n$ *and with* $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ *being the eigenvalues of its* k*th leading principal submatrix for all* $k = 1, \ldots, n-1$.

Proof. We have already explained that H has the desired Ritz values and subdiagonal entries. It remains to show uniqueness. For this we need a recursion for the characteristic polynomials of the leading submatrices H_k of H. We denote the prescribed characteristic polynomial of H_k by $p_k(\lambda)$, and by $\sigma^{k,i}$ we denote the product of prescribed subdiagonal entries

$$\sigma^{k,i} = \prod_{\ell=i}^{k} \sigma_\ell.$$

We also define the polynomial $p_0(\lambda) \equiv 1$. Using expansion along the last column to compute the determinant of $H_k - \lambda I$, we get

$$\begin{aligned}\det(H_k - \lambda I) = {} & (-1)^{k-1} h_{1,k}\sigma^{k-1,1} + (-1)^{k-2} h_{2,k} p_1(\lambda)\sigma^{k-1,2} \\ & + (-1)^{k-3} h_{3,k} p_2(\lambda)\sigma^{k-1,3} + \cdots + (h_{k,k} - \lambda) p_{k-1}(\lambda),\end{aligned}$$

and hence we have the recursion

$$p_k(\lambda) = (h_{kk} - \lambda) p_{k-1}(\lambda) + \sum_{i=1}^{k-1} (-1)^{k-i} h_{ik}\sigma^{k-1,i} p_{i-1}(\lambda), \quad 1 \le k \le n. \tag{2.4}$$

Now assume that both H and $\tilde{H}$ have the desired Ritz values and subdiagonal entries, and let us prove that $H = \tilde{H}$ by induction for all subsequent leading principal submatrices. Clearly, $h_{1,1} = \tilde{h}_{1,1} = \rho_1^{(1)}$, and if the claim is valid for all leading principal submatrices of dimension at most $k-1$, then the entries of H_k and $\tilde{H}_k$ can differ only in the last column. By comparing the coefficients (subsequently before λ^k until λ^0) of the polynomial $p_k(\lambda)$ as given in (2.4) with the coefficients given by

$$p_k(\lambda) = (\tilde{h}_{kk} - \lambda) p_{k-1}(\lambda) + \sum_{i=1}^{k-1} (-1)^{k-i} \tilde{h}_{ik}\sigma^{k-1,i} p_{i-1}(\lambda),$$

we obtain $h_{ik} = \tilde{h}_{ik}$ subsequently for $i = k, k-1, \ldots, 1$. ∎

Theorem 2.2 immediately leads to a parametrization of the matrices and initial Arnoldi vectors that generate a given set of Ritz values $\mathcal{R}$. In addition, the subdiagonal of the generated Hessenberg matrix can be prescribed.

COROLLARY 2.3. *Assume that we are given a set of tuples of complex numbers*

$$\begin{aligned}\mathcal{R} = \{&\rho_1^{(1)},\\ &(\rho_1^{(2)}, \rho_2^{(2)}),\\ &\vdots\\ &(\rho_1^{(n-1)}, \ldots, \rho_{n-1}^{(n-1)}),\\ &(\lambda_1, \ldots\ldots\ldots, \lambda_n)\}\end{aligned}$$

and $n-1$ positive real numbers $\sigma_1, \ldots, \sigma_{n-1}$. If A is a matrix of order n and b a nonzero n-dimensional vector, then the following assertions are equivalent:

1. *The Hessenberg matrix generated by the Arnoldi process applied to A and initial Arnoldi vector b has eigenvalues $\lambda_1, \ldots, \lambda_n$ and subdiagonal entries $\sigma_1, \ldots, \sigma_{n-1}$, and $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ are the eigenvalues of its kth leading principal submatrix for all $k = 1, \ldots, n-1$.*
2. *The matrix A is of the form*

$$A = V D_\sigma U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) D_\sigma^{-1} V^*$$

and $b = \|b\| V e_1$, where V is a unitary matrix, D_σ is the diagonal matrix

$$D_\sigma = \operatorname{diag}(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1}\sigma_j),$$

$U(\mathcal{S})$ is the Ritz value companion transform in (2.2), and $C^{(n)}$ is the companion matrix of the polynomial with roots $\lambda_1, \ldots, \lambda_n$.

Corollary 2.3 is an analogue of Theorem 1.1 on arbitrary convergence of the GMRES method. Here we prescribe k values (the k Ritz values) in the kth iteration, whereas Theorem 1.1 prescribes one value (the kth residual norm); the spectrum of A is prescribed in both results. Note that in [43] it was shown that if the Arnoldi method produces a *particular* sequence of $n(n+1)/2$ Ritz values, the same sequence can be generated by a whole class of matrices together with initial Arnoldi vectors. The paper also gives a description of this class. It can be seen as an analogue of the earlier result of Greenbaum and Strakoš [22], showing that if a residual norm convergence curve is generated by GMRES, the same curve can be obtained by a whole class of matrices together with right-hand sides. Our corollary shows, surprisingly, that for general nonnormal matrices the distribution of the Ritz values generated in the Arnoldi method can be arbitrary and fully independent of the spectrum. We remark that there exist some results on the distribution of Ritz values for specific nonnormal matrices, for example, for Jordan blocks and block diagonal matrices with a simple normal eigenvalue; see [7].

The given parametrization may give some additional insight into the convergence behavior of versions of Arnoldi used in practice, e.g., implicitly restarted Arnoldi with polynomial shifts [4, 5]; in particular it may help one to better understand cases where Arnoldi with exact shifts fails; see, e.g., [13]. As Ritz values are contained in the field of values, it may also have implications for field of values–based analysis of iterative methods.

We deal here with the problem of constructing both an input matrix and an initial vector to produce prescribed Ritz values. In Corollary 2.3 the initial vector

$b = \|b\|Ve_1$ could be chosen arbitrarily if we define A appropriately, since the only requirement for the matrix V is to be unitary. When the matrix A is given, changing b will, of course, change the Ritz values. Constructing an initial vector to produce prescribed Ritz values was done for the Lanczos method in [41]. If a Hermitian matrix has distinct eigenvalues, that paper shows how to construct a perverse initial vector such that the Ritz values in the next-to-last iteration are as far from the eigenvalues as allowed by the interlacing property (see [14] for a generalization to the normal case).

Another consequence of Corollary 2.3 is that the Ritz values in the Arnoldi method are in general independent of the subdiagonal elements $h_{k+1,k}$ of the generated Hessenberg matrix. This is not that strange if one realizes that $h_{k+1,k}$ is not an element of the matrix H_k used to extract the current Ritz values. But, on the other hand, the independency from $h_{k+1,k}$ is still surprising in view of the fact that one is used to regarding the residual norm

$$\|AV^{(k)}y - \rho^{(k)}V^{(k)}y\| = h_{k+1,k}|e_k^T y| \tag{2.5}$$

for an eigenpair $(\rho^{(k)}, y)$ of H_k (see (2.1)) as a measure of the quality of the approximate Ritz value-vector pair $(\rho^{(k)}, V^{(k)}y)$. Corollary 2.3 shows that, in theory, any small nonzero value of $h_{k+1,k}$ is possible with $\rho^{(k)}$ arbitrarily far from the eigenvalues of A. And conversely, all eigenvalues of H_k may coincide with eigenvalues of A with an arbitrarily large value of $h_{k+1,k}$. Though it is known that the residual norm is not always indicative for the quality of the Ritz values (see, e.g., [8, 18]), one might expect that in such counterintuitive cases, the misleading behavior of $h_{k+1,k}$ is compensated for by $|e_k^T y|$ in (2.5). But consider the following: Let A be parametrized as $A = VH(\mathcal{R})V^*$ and $b = Ve_1$, and let for an approximate Ritz value-vector pair $(\rho^{(k)}, V^{(k)}y)$ the residual norm in (2.5) be $|e_k^T y|$ (all subdiagonal entries $h_{k+1,k}$ of $H(\mathcal{R})$ are one), where

$$H(\mathcal{R})_k y = \rho^{(k)} y.$$

For any choice of small nonzero entries $\sigma_1, \ldots, \sigma_{n-1}$, the matrix $VD_\sigma H(\mathcal{R})D_\sigma^{-1}V^*$ with $D_\sigma = \mathrm{diag}(1, \sigma_1, \ldots, \Pi_{j=1}^{n-1}\sigma_j)$ generates the same Ritz value $\rho^{(k)}$, but the residual norm in (2.5) will change as $\sigma_k|e_k^T y_s|$, where

$$\left(D_{\sigma_k} H(\mathcal{R})_k D_{\sigma_k}^{-1}\right) y_s = \rho^{(k)} y_s$$

with $D_{\sigma_k} = \mathrm{diag}(1, \sigma_1, \ldots, \Pi_{j=1}^{k-1}\sigma_j)$. However, the eigenvector y_s is nothing but a scaling of y because

$$\left(D_{\sigma_k} H(\mathcal{R})_k D_{\sigma_k}^{-1}\right)(D_{\sigma_k} y) = \rho^{(k)}(D_{\sigma_k} y),$$

i.e., $y_s = D_{\sigma_k} y$. This means that, with appropriate subdiagonal entries, the value $|e_k^T y_s|$ can be small *too* (even if y_s is normalized) and does not compensate for a small σ_k, in spite of a possibly diverging Ritz value $\rho^{(k)}$. Something similar can be said about cases where all eigenvalues of H_k coincide with eigenvalues of A for arbitrarily large values of σ_k.

3. Prescribed convergence behavior of the Arnoldi and the GMRES methods for the same pair $\{A, b\}$. The diagonal matrix D_σ with positive entries in Corollary 2.3 contains the subdiagonal entries of the generated Hessenberg matrix, and it can be chosen arbitrarily for any prescribed Ritz values. Because the values of these subdiagonal entries influence the residual norms generated by the GMRES

method applied to the corresponding linear system, there is a chance we can modify the behavior of GMRES while maintaining the prescribed Ritz values. This is what we will investigate next. Rather than directly choosing the diagonal matrix D_σ to control GMRES convergence, we will derive an alternative parametrization of the matrices and initial Arnoldi vectors that generate a given set of Ritz values. This parametrization will reveal the relation with the parametrization in Theorem 1.1 and thus might enable us to combine prescribing Ritz values with prescribing GMRES residual norms.

The parametrization in Corollary 2.3 is based on a unitary matrix V whose columns span the nth Krylov subspace $\mathcal{K}_n(A, b)$, whereas the parametrization in Theorem 1.1 works with a unitary matrix W whose columns span $A\mathcal{K}_n(A, b)$. To better understand the relation between Corollary 2.3 and Theorem 1.1, we will translate the former parametrization in terms of the latter. To achieve this, we will use two factorizations of the Krylov matrix

$$K \equiv \left[b, Ab, A^2b, \ldots, A^{n-1}b\right],$$

one with V and one with W. The first factorization is nothing but the QR decomposition

$$K = VU \tag{3.1}$$

of K. By *the* QR decomposition we will always mean the unique QR decomposition whose upper triangular factor has positive real main diagonal. The upper triangular factor U is related to the generated Ritz values as follows.

LEMMA 3.1. *Let H be the Hessenberg matrix generated by an Arnoldi process terminating at the nth iteration applied to A and b, and let $U(\mathcal{S})$ be the Ritz value companion transform in* (2.2) *corresponding to the generated strict Ritz values. Then the upper triangular factor U of the QR factorization* (3.1) *of the Krylov matrix K is*

$$U = \|b\| \operatorname{diag}\left(1, h_{2,1}, h_{2,1}h_{3,2}, \ldots, \Pi_{j=1}^{n-1} h_{j+1,j}\right) U(\mathcal{S})^{-1}.$$

Proof. Any Arnoldi process (terminating at the nth iteration) can be written according to the parametrization of Corollary 2.3 with $D_\sigma = \operatorname{diag}\left(1, h_{2,1}, \ldots, \Pi_{j=1}^{n-1} h_{j+1,j}\right)$. Then in the Krylov matrix

$$K = \left[b, Ab, \ldots, A^{n-1}b\right]$$

we can take $\|b\|V$ out of the brackets to factor it since

$$\begin{aligned}
b &= \|b\| V e_1, \\
Ab &= \|b\| V D_\sigma U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) D_\sigma^{-1} e_1, \\
A^2b &= \|b\| V \left(D_\sigma U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) D_\sigma^{-1}\right)^2 e_1, \\
&\cdots = \cdots \\
A^{n-1}b &= \|b\| V \left(D_\sigma U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) D_\sigma^{-1}\right)^{n-1} e_1.
\end{aligned}$$

Therefore

$$K = \|b\| V \left[e_1, D_\sigma U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) D_\sigma^{-1} e_1, \ldots, \left(D_\sigma U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) D_\sigma^{-1}\right)^{n-1} e_1\right].$$

Now we would like to show that the last matrix on the right-hand side is just $D_\sigma U(\mathcal{S})^{-1}$. The first entry of the diagonal matrix D_σ being one, we have $U(\mathcal{S})D_\sigma^{-1}e_1 = e_1$. Obviously we have $(D_\sigma U(\mathcal{S})^{-1}C^{(n)}U(\mathcal{S})D_\sigma^{-1})^j = (D_\sigma U(\mathcal{S})^{-1}(C^{(n)})^j U(\mathcal{S})D_\sigma^{-1})$. Hence $\left(D_\sigma U(\mathcal{S})^{-1}C^{(n)}U(\mathcal{S})D_\sigma^{-1}\right)^j e_1 = D_\sigma U(\mathcal{S})^{-1}(C^{(n)})^j e_1$. It is straightforward to see that $(C^{(n)})^j e_1 = e_{j+1}$. This yields

$$\left(D_\sigma U(\mathcal{S})^{-1}C^{(n)}U(\mathcal{S})D_\sigma^{-1}\right)^j e_1 = D_\sigma U(\mathcal{S})^{-1}e_{j+1}, \quad j = 0, \ldots, n-1,$$

and hence we have the factorization $K = \|b\| V D_\sigma U(\mathcal{S})^{-1}$. On the other hand, $K = VU$. The uniqueness of the QR factorization gives $U = \|b\| D_\sigma U(\mathcal{S})^{-1}$. ∎

A similar result is proved in [28, Proposition 3.1]. The second factorization of K which we need involves the unitary factor W. We prove the following result in the same way as the previous lemma; it was also proved in [2] in a different way.

LEMMA 3.2. *Consider a matrix A with initial Arnoldi vector b such that the Arnoldi process does not terminate before iteration n. If $A = WYC^{(n)}Y^{-1}W^*$ and $b = Wh$ according to Theorem* 1.1*, then we have*

$$K = WY.$$

Proof. With Theorem 1.1 the Krylov matrix is defined as

$$K = \left[Wh, AWh, A^2Wh, \ldots, A^{n-1}Wh\right].$$

We wish to take W out of the brackets to factor K. This can be done since

$$\begin{aligned} AW &= WYC^{(n)}Y^{-1}, \\ A^2W &= W(YC^{(n)}Y^{-1})^2, \\ \cdots &= \cdots \\ A^{n-1}W &= W(YC^{(n)}Y^{-1})^{n-1}. \end{aligned}$$

Therefore

$$K = W\left[h, YC^{(n)}Y^{-1}h, \ldots, (YC^{(n)}Y^{-1})^{n-1}h\right].$$

Now we would like to show that the last matrix on the right-hand side is just Y. The vector h being the first column of Y, we have $h = Ye_1$. Obviously we have $(YC^{(n)}Y^{-1})^j = Y(C^{(n)})^j Y^{-1}$. Hence $(YC^{(n)}Y^{-1})^j h = Y(C^{(n)})^j e_1$. As before, $(C^{(n)})^j e_1 = e_{j+1}$. This yields

$$(YC^{(n)}Y^{-1})^j h = Ye_{j+1}, \quad j = 0, \ldots, n-1,$$

and this proves the result. ∎

With the two factorizations $K = VU = WY$ we are ready for a second parametrization, formulated with the notation of Theorem 1.1 and based on the unitary matrix W, of the pairs $\{A, b\}$ generating arbitrary Ritz values.

THEOREM 3.3. *Assume that we are given a set of tuples of complex numbers*

$$\begin{aligned} \mathcal{R} = \{&\rho_1^{(1)}, \\ &(\rho_1^{(2)}, \rho_2^{(2)}), \\ &\vdots \\ &(\rho_1^{(n-1)}, \ldots, \rho_{n-1}^{(n-1)}), \\ &(\lambda_1, \ldots\ldots\ldots, \lambda_n)\}, \end{aligned}$$

such that $(\lambda_1, \ldots, \lambda_n)$ *contains only nonzero numbers, and* $n-1$ *positive real numbers* $\sigma_1, \ldots, \sigma_{n-1}$. *If* A *is a matrix of order* n *and* b *a nonzero* n*-dimensional vector, then the following assertions are equivalent:*

1. *The Hessenberg matrix generated by the Arnoldi process applied to* A *and initial Arnoldi vector* b *has eigenvalues* $\lambda_1, \ldots, \lambda_n$ *and subdiagonal entries* $\sigma_1, \ldots, \sigma_{n-1}$, *and* $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ *are the eigenvalues of its* k*th leading principal submatrix for all* $k = 1, \ldots, n-1$.
2. *The matrix* A *is of the form*

$$A = WYC^{(n)}Y^{-1}W^*$$

and $b = Wh$, *where* W *is a unitary matrix,* $C^{(n)}$ *is the companion matrix corresponding to the eigenvalues* $\lambda_1, \ldots, \lambda_n$, *and* Y *is of the form*

$$Y = \begin{bmatrix} h & R \\ & 0 \end{bmatrix}.$$

R is the upper triangular matrix

$$R = \Gamma L^* T \tag{3.2}$$

of order $n-1$, *where* T *is the trailing principal submatrix in the partitioning*

$$\|b\| \operatorname{diag}(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1}\sigma_j) U(\mathcal{S})^{-1} = \begin{bmatrix} \|b\| & t^* \\ 0 & T \end{bmatrix} \tag{3.3}$$

of the scaled inverse of the Ritz value companion transform $U(\mathcal{S})$ *in* (2.2) *and* L *is the lower triangular factor in the Cholesky decomposition*

$$LL^* = I_{n-1} + T^{-*}tt^*T^{-1}. \tag{3.4}$$

The diagonal matrix Γ *with unit modulus entries is such that*

$$e_k^T \Gamma L^{-1} T^{-*} t \geq 0, \quad k = 1, \ldots, n-1, \tag{3.5}$$

and the entries of $h = [\eta_1, \ldots, \eta_n]^T$ *satisfy*

$$[\eta_1, \ldots, \eta_{n-1}]^T = \|b\| \Gamma L^{-1} T^{-*} t, \qquad \eta_n = \|b\| \sqrt{1 - \|L^{-1}T^{-*}t\|^2}. \tag{3.6}$$

Proof. First we prove the implication $1 \to 2$. Because the Arnoldi process does not stop before the last iteration, GMRES applied to the linear system with matrix A, right-hand side b, and zero initial guess does not stop before the last iteration, and we can write $A = WYC^{(n)}Y^{-1}W^*$ and $b = Wh$ according to Theorem 1.1. From Lemma 3.2, the factorization (3.1), and Lemma 3.1, we have

$$K^*K = Y^*W^*WY = Y^*Y, \qquad K^*K = U^*V^*VU = \|b\|^2 U(\mathcal{S})^{-*} D_\sigma^T D_\sigma U(\mathcal{S})^{-1}.$$

Hence the matrix Y from the parametrization must satisfy

$$Y^*Y = \|b\|^2 U(\mathcal{S})^{-*} D_\sigma^T D_\sigma U(\mathcal{S})^{-1}.$$

Let $\hat{h} = [\eta_1, \ldots, \eta_{n-1}]^T$ be the vector of the first $n-1$ components of h from (1.2). Then from (1.1) we have

$$Y^*Y = \begin{bmatrix} \|h\|^2 & \hat{h}^*R \\ R^*\hat{h} & R^*R \end{bmatrix}. \tag{3.7}$$

Comparing (3.7) with $\|b\|^2 U(\mathcal{S})^{-*} D_\sigma^T D_\sigma U(\mathcal{S})^{-1}$ and using the partitioning (3.3), we obtain for R and $\hat{h}$ the conditions

$$R^*R = T^*T + tt^*, \qquad \hat{h} = \|b\| R^{-*} t. \tag{3.8}$$

Furthermore, we have the conditions $\eta_k \geq 0$, $k = 1, \ldots, n-1$, because all entries of $\hat{h}$ correspond to entries describing the GMRES convergence curve according to (1.2).

Let L be the lower triangular factor in the Cholesky decomposition

$$LL^* = I_{n-1} + T^{-*} t t^* T^{-1},$$

let Γ be a diagonal matrix with unit modulus entries, and let $R = \Gamma L^* T$. Then

$$R^*R = T^* L \Gamma^* \Gamma L^* T = T^*(I_{n-1} + T^{-*} t t^* T^{-1}) T = T^*T + tt^*$$

is always satisfied and Γ can be chosen such that

$$e_k^T \Gamma L^{-1} T^{-*} t \geq 0, \quad k = 1, \ldots, n-1.$$

It follows that

$$\hat{h} = \|b\| R^{-*} t = \|b\| \, \Gamma L^{-1} T^{-*} t,$$

and with $\|h\| = \|W^* b\| = \|b\|$ we obtain

$$\eta_n = \sqrt{\|h\|^2 - \|\hat{h}\|^2} = \|b\| \sqrt{1 - \|L^{-1} T^{-*} t\|^2}.$$

For the implication $2 \to 1$, let $A = WYC^{(n)}Y^{-1}W^*$ be the parametrization of A given in assertion 2, and let $b = Wh$. By Lemma 3.2, $K = WY$; let $K = V\tilde{U}$ be the QR factorization of the Krylov matrix K. We first show that $\tilde{U} = \|b\| D_\sigma U(\mathcal{S})^{-1}$.

In the QR decomposition $K = V\tilde{U}$ we have $Ve_1 = b/\|b\|$, and therefore we can partition $\tilde{U}$ as

$$\tilde{U} = \begin{bmatrix} \|b\| & \tilde{t}^* \\ 0 & \tilde{T} \end{bmatrix}. \tag{3.9}$$

With the first part of the proof

$$R^*R = \tilde{T}^*\tilde{T} + \tilde{t}\tilde{t}^*, \qquad \hat{h} = \|b\| R^{-*} \tilde{t},$$

(see (3.8)), i.e.,

$$\tilde{t} = \frac{R^*\hat{h}}{\|b\|}, \qquad \tilde{T}^*\tilde{T} = R^*R - \frac{R^*\hat{h}\hat{h}^*R}{\|b\|^2}.$$

But by assumption, we have for t and T from (3.4) and (3.6) the same equalities,

$$t = \frac{T^* L \Gamma^* \hat{h}}{\|b\|} = \frac{R^*\hat{h}}{\|b\|},$$

$$T^*T = T^*(LL^* - T^{-*} t t^* T^{-1}) T = T^* L \Gamma^* \Gamma L^* T - tt^* = R^*R - \frac{R^*\hat{h}\hat{h}^*R}{\|b\|^2}.$$

The matrix $R^*R - \frac{R^*\hat{h}\hat{h}^*R}{\|b\|^2}$ is positive definite since it is the Schur complement of $\|h\|^2$ in Y^*Y, which is positive definite. Therefore the Cholesky decomposition of the matrix $R^*R - \frac{R^*\hat{h}\hat{h}^*R}{\|b\|^2}$ exists, and $\tilde{T} = T$ is the unique Cholesky factor. Together with $\tilde{t} = t = \frac{R^*\hat{h}}{\|b\|}$ we have

$$\tilde{U} = \|b\| D_\sigma U(\mathcal{S})^{-1}.$$

Because of $K = WY = V\tilde{U}$ and with (2.3) it follows that

$$\begin{aligned} A &= WYC^{(n)}Y^{-1}W^* = V\tilde{U}C^{(n)}\tilde{U}^{-1}V^* \\ &= VD_\sigma U(\mathcal{S})^{-1}C^{(n)}U(\mathcal{S})D_\sigma^{-1}V^* = VD_\sigma H(\mathcal{R})D_\sigma^{-1}V^*. \end{aligned}$$

The upper Hessenberg matrix $D_\sigma H(\mathcal{R})D_\sigma^{-1}$ generated by the Arnoldi method therefore has the prescribed Ritz values and subdiagonal entries. ∎

Note that Theorem 3.3 and Corollary 2.3 are not fully equivalent. In Theorem 3.3 we must assume, for reasons of compatibility with Theorem 1.1, that the spectrum of A does not contain the origin. In Corollary 2.3 the only free parameters are a unitary matrix and the norm of the initial Arnoldi vector. In Theorem 3.3 there appears to be slightly more freedom because a unit modulus entry of Γ can lie anywhere on the unit circle if the corresponding entry of $L^{-1}T^{-*}t$ is zero; see (3.5). There is of course much less freedom in Theorem 3.3 than there is in the parametrization of Theorem 1.1 when prescribing a GMRES convergence curve.

We see that by modifying the choice of the subdiagonal entries $\sigma_1, \ldots, \sigma_{n-1}$ in Theorem 3.3, we might modify the vector h representing the GMRES convergence curve generated with A and b while maintaining the prescribed Ritz values, i.e., while leaving the Ritz value companion transform $U(\mathcal{S})$ in (3.3) unchanged. Does this mean we can force any GMRES convergence speed with arbitrary Ritz values? There is one situation where this is certainly not possible: When there is a zero Ritz value in some iteration, this implies a singular Hessenberg matrix and corresponds to an indefinable iterate in the full orthogonalization method, which is equivalent to stagnation in the parallel GMRES process; see, e.g., [6, 19]. Hence zero Ritz values are equivalent with GMRES stagnation. For completeness, we give another proof of this well-known fact, formulated with the notation of Theorem 3.3.

LEMMA 3.4. *With the notation of Theorem* 3.3 *and for* $1 \le k \le n-1$, *the* k*-tuple* $(\rho_1^{(k)}, \ldots, \rho_k^{(k)})$ *contains a zero Ritz value if and only if* $\eta_k = 0$ *in* (3.6).

Proof. Denote by $U(\mathcal{S})$ the Ritz value companion transform in (2.2), and let it be partitioned according to (3.3) as

$$U(\mathcal{S}) = \|b\| D_\sigma \begin{bmatrix} \|b\| & t^* \\ 0 & T \end{bmatrix}^{-1} = \|b\| D_\sigma \begin{bmatrix} \frac{1}{\|b\|} & \frac{-t^*T^{-1}}{\|b\|} \\ 0 & T^{-1} \end{bmatrix},$$

where $D_\sigma = \operatorname{diag}(1, \sigma_1, \sigma_1\sigma_2, \ldots, \Pi_{j=1}^{n-1}\sigma_j)$. By definition of $U(\mathcal{S})$, the k-tuple $(\rho_1^{(k)}, \ldots, \rho_k^{(k)})$ contains a zero Ritz value if and only if $t^*T^{-1}e_k = 0$. It can easily be checked that the lower triangular factor L in the Cholesky decomposition

$$LL^* = I_{n-1} + T^{-*}tt^*T^{-1}$$

has its kth row and column zero, except for the diagonal entry, if and only if $t^*T^{-1}e_k = 0$. Then the vector $\hat{h}$, being the solution of the lower triangular system

$$L\Gamma^*\hat{h} = T^{-*}t,$$

has kth entry zero if and only if $t^*T^{-1}e_k = 0$. ∎

Thus GMRES residual norms cannot be fully independent of Ritz values. However, we will show that the *only* restriction Ritz values put on GMRES residual norms is precisely that zero Ritz values imply stagnation. Otherwise, any GMRES behavior is possible with arbitrary prescribed Ritz values. Before proving this, we need the following auxiliary result.

LEMMA 3.5. *Consider n positive real numbers*

$$f(0) \geq f(1) \geq \cdots \geq f(n-1) > 0,$$

and define

$$\eta_k = (f(k-1)^2 - f(k)^2)^{1/2}, \quad k < n, \quad \eta_n = f(n-1), \qquad \hat{h} = [\eta_1, \ldots, \eta_{n_1}]^T.$$

If we denote by R_h the upper triangular factor of the Cholesky decomposition

$$R_h^T R_h = I_{n-1} - \frac{\hat{h}\hat{h}^T}{f(0)^2},$$

then we have

$$e_k^T R_h^{-T} \hat{h} = 0 \quad \Leftrightarrow \quad f(k-1) = f(k), \qquad k = 1, \ldots, n-1.$$

Proof. The entries of R_h^T are

(3.10)

$$(R_h^T)_{i,k} = \frac{-\eta_i \eta_k}{\sqrt{\eta_{k+1}^2 + \cdots + \eta_n^2}\sqrt{\eta_k^2 + \cdots + \eta_n^2}}, \quad (R_h^T)_{k,k} = \frac{\sqrt{\eta_{k+1}^2 + \cdots + \eta_n^2}}{\sqrt{\eta_k^2 + \cdots + \eta_n^2}};$$

see [17] on the Cholesky decomposition of a rank-one updated identity matrix, or also [29, Theorem 4.2]. Therefore, if $\eta_k = 0$ for some $k \leq n-1$, then the kth row and kth column of R_h^T are zero except for the main diagonal entry. It is easily seen from solving the lower triangular system $R_h^T x = \hat{h}$ with forward substitution that $x = R_h^{-T}\hat{h}$ is zero only where $\hat{h}$ is zero. ∎

THEOREM 3.6. *Consider a set of tuples of complex numbers*

$$\begin{aligned} \mathcal{R} = \{ & \rho_1^{(1)}, \\ & (\rho_1^{(2)}, \rho_2^{(2)}), \\ & \vdots \\ & (\rho_1^{(n-1)}, \ldots, \rho_{n-1}^{(n-1)}), \\ & (\lambda_1, \ldots\ldots\ldots, \lambda_n)\}, \end{aligned}$$

such that $(\lambda_1, \ldots, \lambda_n)$ contains no zero number, and n positive numbers

$$f(0) \geq f(1) \geq \cdots \geq f(n-1) > 0,$$

such that $f(k-1) = f(k)$ if and only if the k-tuple $(\rho_1^{(k)}, \ldots, \rho_k^{(k)})$ contains a zero number. Let A be a square matrix of size n, and let b be a nonzero n-dimensional vector. The following assertions are equivalent:

1. *The GMRES method applied to A and right-hand side b with zero initial guess yields residuals $r^{(k)}$, $k = 0, \ldots, n-1$, such that*

$$\|r^{(k)}\| = f(k), \quad k = 0, \ldots, n-1,$$

A has eigenvalues $\lambda_1, \ldots, \lambda_n$, and $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ are the eigenvalues of the kth leading principal submatrix of the generated Hessenberg matrix for all $k = 1, \ldots, n-1$.

2. *The matrix A is of the form*

$$A = WYC^{(n)}Y^{-1}W^*$$

and $b = Wh$, where W is a unitary matrix and $C^{(n)}$ is the companion matrix corresponding to the polynomial with roots $\lambda_1, \ldots, \lambda_n$. Y is given by

$$Y = \begin{bmatrix} h & R \\ 0 & \end{bmatrix},$$

h being the vector

$$h = [\eta_1, \ldots, \eta_n]^T, \qquad \eta_k = (f(k-1)^2 - f(k)^2)^{1/2}, \quad k < n, \quad \eta_n = f(n-1),$$

and R being the nonsingular upper triangular matrix of order $n-1$

$$R = R_h^{-1} D_c^{-*} C^{-1}, \tag{3.11}$$

where C is the trailing principal submatrix in the partitioning

$$U(\mathcal{S}) = \begin{bmatrix} 1 & c^* \\ 0 & C \end{bmatrix} \tag{3.12}$$

of the Ritz value companion transform $U(\mathcal{S})$ for $\mathcal{R}$ defined in (2.2). *R_h is the upper triangular factor of the Cholesky decomposition*

$$R_h^T R_h = I_{n-1} - \frac{\hat{h}\hat{h}^T}{f(0)^2}$$

for $\hat{h} = [\eta_1, \ldots, \eta_{n-1}]^T$, and D_c is a nonsingular diagonal matrix such that

$$R_h^{-T}\hat{h} = -f(0)^2 D_c\, c. \tag{3.13}$$

Proof. Because of Theorem 1.1 it is clear that the parametrization given here generates the prescribed GMRES residual norms and vice versa. Hence it suffices to show that the given parametrization generates the prescribed Ritz values and vice versa. For this we will use the parametrization of Theorem 3.3 and prove that the matrix R in (3.11) satisfies the same conditions as the upper triangular R in (3.2) in Theorem 3.3.

First we show that the nonsingular diagonal matrix D_c used to define R in (3.11) exists. With the assumed partitioning (3.12) of $U(\mathcal{S})$ and by the definition of $U(\mathcal{S})$, the entries of c are zero precisely at positions corresponding to iterations with a zero Ritz value. By assumption, $\hat{h}$ is zero at exactly these positions and so is $R_h^{-T}\hat{h}$ with Lemma 3.5. Thus we can always define a nonsingular diagonal matrix D_c such that

$$R_h^{-T}\hat{h} = -f(0)^2 D_c c.$$

Now with the definition (3.11) of R we have

$$R^*\hat{h} = -f(0)^2 C^{-*} c.$$

Next, in analogy with (3.3), consider the partitioning

$$\operatorname{diag}(f(0), D_c^{-*})\, U(\mathcal{S})^{-1} = \begin{bmatrix} f(0) & t^* \\ 0 & T \end{bmatrix} \tag{3.14}$$

of a diagonal scaling of $U(\mathcal{S})^{-1} = \left[\begin{smallmatrix} 1 & -c^*C^{-1} \\ 0 & C^{-1} \end{smallmatrix}\right]$. It follows that

$$t = -f(0)C^{-*}c = \frac{R^*\hat{h}}{f(0)}$$

and

$$T = D_c^{-*} C^{-1}.$$

To prove that the matrix R in (3.11) satisfies the same conditions as the upper triangular R in (3.2) in Theorem 3.3, it remains to show that $R_h^{-1} = L^*$, $\Gamma = I_{n-1}$, where L and Γ are the matrices defined in the second assertion of Theorem 3.3. We have

$$\begin{aligned} I_{n-1} + T^{-*}tt^*T^{-1} &= I_{n-1} + D_c C^* \frac{R^*\hat{h}}{f(0)} \left(D_c C^* \frac{R^*\hat{h}}{f(0)} \right)^* \\ &= I_{n-1} + \frac{R_h^{-T}\hat{h}}{f(0)} \left(\frac{R_h^{-T}\hat{h}}{f(0)} \right)^* = R_h^{-T} \left(R_h^T R_h + \frac{\hat{h}\hat{h}^*}{f(0)^2} \right) R_h^{-1} \\ &= R_h^{-T} R_h^{-1} \end{aligned}$$

and with $\Gamma = I_{n-1}$

$$e_k^T R_h^T T^{-*} t = e_k^T R_h^T \frac{R_h^{-T}\hat{h}}{f(0)} = \frac{\eta_k}{f(0)} \geq 0, \quad k = 1, \ldots, n-1.$$

Together with

$$\begin{aligned} \eta_n = f(n-1) &= \sqrt{f(0)^2 - (f(0)^2 - f(1)^2) - \cdots - (f(n-2)^2 - f(n-1)^2)} \\ &= f(0)\sqrt{1 - \frac{\|\hat{h}\|^2}{f(0)^2}}, \end{aligned}$$

we have that matrices of the form

$$W \begin{bmatrix} h & R \\ 0 & \end{bmatrix} C(\mathcal{R}^{(n)}) \begin{bmatrix} h & R \\ 0 & \end{bmatrix}^{-1} W^*$$

and right-hand sides Wh generate the prescribed Ritz values and vice versa; see Theorem 3.3. ∎

The only freedom we have to prescribe both Ritz values and GMRES residual norms is in the unitary matrix W and in those entries of the diagonal matrix D_c on positions corresponding to iterations with a zero Ritz value or, equivalently, on positions corresponding to iterations where GMRES stagnates. On these positions D_c may have arbitrary values. In this sense we have exhausted all the degrees of

freedom; GMRES and Arnoldi are invariant under unitary transformation, and more values than Ritz values and residual norms cannot be prescribed for the same Arnoldi process.

Theorem 3.6 says that one can construct matrices and right-hand sides for which converged Ritz values need not imply accelerated convergence speed in the GMRES method, as is the case for the CG method for Hermitian positive definite matrices [45]. The only restriction Ritz values put on GMRES is that a zero Ritz value leads to stagnation in the corresponding iteration. A restricted role of Ritz values for GMRES may be expected in view of the fact that the Ritz values are *not* the roots of the polynomials GMRES generates to compute its residuals. These roots are the harmonic Ritz values [35, 19]. Although harmonic Ritz values generated in the Arnoldi procedure might be prescribed in a way similar to what we did for ordinary Ritz values in the previous section [30], it is not clear whether this is possible with given GMRES residual norms. Nonetheless, the extent to which ordinary Ritz values and residual norms are independent is astonishing. Note, for example, that for matrices close to normal the bounds derived in [46] suggest that as soon as eigenvalues of such matrices are sufficiently well approximated by Ritz values, GMRES from then on converges at least as fast as for a related system in which these eigenvalues are missing. This may be surprising, but it is not contradictory.

Note that we also could have formulated the second assertion in the previous theorem analogously to the second assertion in Theorem 3.3. Then the diagonal scaling matrix in (3.3) takes the form of the diagonal matrix in (3.14); otherwise the assertion need not be changed. Translated in the notation of Corollary 2.3, this gives the following alternative parametrization.

COROLLARY 3.7. *Assume that we are given a set of tuples of complex numbers*

$$\begin{aligned}\mathcal{R} = \{&\rho_1^{(1)},\\ &(\rho_1^{(2)}, \rho_2^{(2)}),\\ &\vdots\\ &(\rho_1^{(n-1)}, \ldots, \rho_{n-1}^{(n-1)}),\\ &(\lambda_1, \ldots\ldots\ldots, \lambda_n)\},\end{aligned}$$

such that $(\lambda_1, \ldots, \lambda_n)$ *contains no zero number, and* n *positive real numbers*

$$f(0) \geq f(1) \geq \cdots \geq f(n-1) > 0,$$

such that $f(k-1) = f(k)$ *if and only if the* k*-tuple* $(\rho_1^{(k)}, \ldots, \rho_k^{(k)})$ *contains a zero number. If* A *is a matrix of order* n *and* b *a nonzero* n*-dimensional vector, then the following assertions are equivalent:*

1. *The GMRES method applied to* A *and right-hand side* b *with zero initial guess yields residuals* $r^{(k)}$, $k = 0, \ldots, n-1$, *such that*

$$\|r^{(k)}\| = f(k), \quad k = 0, \ldots, n-1,$$

   A has eigenvalues $\lambda_1, \ldots, \lambda_n$, *and* $\rho_1^{(k)}, \ldots, \rho_k^{(k)}$ *are the eigenvalues of the* k*th leading principal submatrix of the generated Hessenberg matrix for all* $k = 1, \ldots, n-1$.
2. *The matrix A is of the form*

$$A = V \operatorname{diag}(f(0), D_c^{-*})\, U(\mathcal{S})^{-1} C^{(n)} U(\mathcal{S}) \operatorname{diag}(f(0)^{-1}, D_c^{*}) V^*$$

and $b = \|b\|Ve_1$, where V is a unitary matrix, $U(\mathcal{S})$ is the Ritz value companion transform for $\mathcal{R}$ defined in (2.2), and $C^{(n)}$ is the companion matrix of the polynomial with roots $\lambda_1, \ldots, \lambda_n$. D_c is a nonsingular diagonal matrix such that

$$R_h^{-T}\hat{h} = -f(0)^2 D_c c$$

with $\hat{h}$ being the vector

$$\hat{h} = [\eta_1, \ldots, \eta_{n-1}]^T, \qquad \eta_k = (f(k-1)^2 - f(k)^2)^{1/2},$$

R_h being the upper triangular factor of the Cholesky decomposition

$$R_h^T R_h = I_{n-1} - \frac{\hat{h}\hat{h}^T}{f(0)^2},$$

and c being the first row of $U(\mathcal{S})$ without its diagonal entry.

This parametrization is based on unitary matrices V spanning $\mathcal{K}_n(A, b)$ instead of unitary matrices W spanning $A\mathcal{K}_n(A, b)$ and is therefore closer to the actual Arnoldi process which is run in standard implementations of the GMRES and Arnoldi methods. On the other hand, the parametrization in Theorem 3.6 reveals more clearly the relation with the prescribed residual norms. Note that we can easily change Corollary 3.7 to yield a "V-based" analogue of Theorem 1.1; it suffices to consider $U(\mathcal{S})$ as a free parameter matrix. Corollary 3.7 also shows how to define the subdiagonal entries $h_{k+1,k}$ of a Hessenberg matrix with prescribed Ritz values in order to obtain prescribed GMRES residual norms: They follow from the equality

$$f(0)\,\mathrm{diag}\left(1, h_{2,1}, h_{2,1}h_{3,2}, \ldots, \Pi_{j=1}^{n-1} h_{j+1,j}\right) = \mathrm{diag}\,(f(0), D_c^{-*}).$$

4. Conclusions and future work. The Arnoldi orthogonalization process is a cornerstone of several successful Krylov subspace methods for non-Hermitian matrices. Nevertheless, two of the most popular methods based on it, the GMRES and the Arnoldi methods, can exhibit counterintuitive convergence behavior. For GMRES it has been known for some time that any nonincreasing convergence curve is possible and can be generated with any spectrum [21]; the fact that *all* Ritz values formed by the Arnoldi method can be prescribed appears not to have been noticed so far. The present paper also shows that arbitrary convergence of GMRES is possible not only with any spectrum, but even with any Ritz values for all iterations (provided that we treat the stagnation case correctly).

Given the success of (modified versions of) the GMRES and Arnoldi methods for a large variety of problems, the situations described in our theoretical results may occur rarely in solving practical problems in scientific computing. For example, in the Arnoldi method, cases of Ritz values diverging further away from the spectrum in every iteration are possible, as we proved in section 2, but they happen for particular matrices only in combination with particular initial Arnoldi vectors. As one normally chooses the initial Arnoldi vector randomly, the chances that this vector produces diverging Ritz values may be small, and in practice one can easily rerun the process with a different random initial Arnoldi vector. In the GMRES method, however, one is stuck with a given right-hand side, and applications exist where the pathological cases described in [21] occur. An example is given by convection-diffusion problems; see, e.g., [37] or [26, Figures 3.10 and 3.11]. This type of problem also contains an illustration of our results of section 3: In the convection dominated case, system matrices are

often close to transposed Jordan blocks (i.e., upper Hessenberg matrices with identical Ritz values for all iterations), and, for certain boundary conditions, right-hand sides are close to the first unit vector [26]. Hence we have almost converged Ritz values from the very start, but this does not mean that GMRES converges rapidly as one would expect. On the contrary, it is known that these problems give very slow, nearly stagnating GMRES residual norms during the initial phase of convergence [15, 26].

It is often assumed that counterintuitive GMRES behavior, i.e., spectral information which is misleading for residual norms, is possible in the highly nonnormal case only, and one may expect the counterintuitive results of this paper to be restricted to the highly nonnormal case, too. Neither of the two statements is entirely correct; for instance, arbitrary GMRES convergence curves are possible for such nice normal matrices as are the perfectly conditioned unitary matrices; see [22, section 3.1] and [21]. As for our results on the Arnoldi method, certainly prescribed Ritz values outside the convex hull of the eigenvalues are possible with nonnormal matrices only, and probably the further one prescribes Ritz values away from the convex hull, the more nonnormal the constructed input matrix must be. On the other hand, divergence *inside* the convex hull might still be possible with some normal but non-Hermitian matrices. Very little appears to the authors to be known on this topic (for general normal matrices of size three, see, e.g., [7]). Although there are generalized interlacing properties for normal matrices, they cannot be exploited because the leading principal submatrices of normal Hessenberg matrices need not be normal. Let us also recall that the Ritz values generated in the Lanczos method in the next-to-last iteration can be as far from the eigenvalues as allowed by the interlacing property [41].

Our results are of a theoretical nature and may give additional insight into the properties of the GMRES and the Arnoldi methods. An important issue related to our results is how to detect, a priori, whether a matrix with initial vector will lead to diverging Ritz value behavior in Arnoldi or to stagnation in GMRES. For GMRES, work on complete or partial stagnation was done, for example, in [48] or, recently, in [29], where the results are linked with the parametrization in Theorem 1.1. More generally, the question is whether our theory gives some insight into what is a good Arnoldi starting vector, respectively, right-hand side b. Work for the near future includes modifications of our results for popular restarted versions of Arnoldi or GMRES which may enhance theoretical insight into the behavior of strategies that are frequently used in practice.

Software. At http://www.cs.cas.cz/duintjertebbens/duintjertebbens_soft.html the reader can find MATLAB subroutines to create matrices and initial vectors with the parametrizations in this paper.

Acknowledgments. The authors are indebted to Zdeněk Strakoš for initiating their work on this topic. They thank the anonymous referees for their comments, and they thank Russel Carden for pointing out reference [43].

REFERENCES

[1] G. S. Ammar and C. Y. He, *On an inverse eigenvalue problem for unitary Hessenberg matrices*, Linear Algebra Appl., 218 (1995), pp. 263–271.

[2] M. Arioli, V. Pták, and Z. Strakoš, *Krylov sequences of maximal length and convergence of GMRES*, BIT, 38 (1998), pp. 636–643.

[3] W. E. Arnoldi, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.

[4] C. Beattie, M. Embree, and J. Rossi, *Convergence of restarted Krylov subspaces to invariant subspaces*, SIAM J. Matrix Anal. Appl., 25 (2004), pp. 1074–1109.
[5] C. A. Beattie, M. Embree, and D. C. Sorensen, *Convergence of polynomial restart Krylov methods for eigenvalue computations*, SIAM Rev., 47 (2005), pp. 492–515.
[6] P. N. Brown, *A theoretical comparison of the Arnoldi and GMRES algorithms*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 58–78.
[7] R. Carden, *Ritz Values and Arnoldi Convergence for Non-Hermitian Matrices*, Ph.D. thesis, Rice University, Houston, TX, 2011.
[8] F. Chatelin, *Eigenvalues of Matrices*, John Wiley & Sons, Chichester, UK, 1993.
[9] G. N. de Oliveira, *Matrices with prescribed characteristic polynomial and a prescribed submatrix*. I, Pacific J. Math. 29 (1969), pp. 653–661.
[10] G. N. de Oliveira, *Matrices with prescribed characteristic polynomial and a prescribed submatrix*. II, Pacific J. Math. 29 (1969), pp. 663–667.
[11] M. Eiermann, *Fields of values and iterative methods*, Linear Algebra Appl., 180 (1993), pp. 167–197.
[12] M. Eiermann and O. G. Ernst, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numer., 10 (2001), pp. 251–312.
[13] M. Embree, *The Arnoldi eigenvalue iteration with exact shifts can fail*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 1–10.
[14] T. Ericsson, *On the Eigenvalues and Eigenvectors of Hessenberg Matrices*, Numerical Analysis Group, Göteborg, Report 10, Chalmers University of Technology and the University of Göteborg, Department of Computer Sciences, Göteborg, Sweden, 1990; available online at http://www.math.chalmers.se/Math/Research/NumericalAnalysis/num_analysis/reports/ericsson_Hessenberg_matrices.ps.gz.
[15] O. G. Ernst, *Residual-minimizing Krylov subspace methods for stabilized discretizations of convection-diffusion equations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1079–1101.
[16] K. Fan and G. Pall, *Imbedding conditions for Hermitian and normal matrices*, Canad. J. Math., 9 (1957), pp. 298–304.
[17] P. E. Gill, G. H. Golub, W. Murray, and M. A. Saunders, *Methods for modifying matrix factorizations*, Math. Comp., 28 (1974), pp. 505–535.
[18] S. Godet-Thobie, *Eigenvalues of Large Highly Nonnormal Matrices*, Ph.D. thesis, University Paris IX, Dauphine, Paris, France, 1993.
[19] S. Goossens and D. Roose, *Ritz and harmonic Ritz values and the convergence of FOM and GMRES*, Numer. Linear Algebra Appl., 6 (1999), pp. 281–293.
[20] A. Greenbaum, *Generalizations of the field of values useful in the study of polynomial functions of a matrix*, Linear Algebra Appl., 347 (2002), pp. 233–249.
[21] 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.
[22] A. Greenbaum and Z. Strakoš, *Matrices that generate the same Krylov residual spaces*, in Recent Advances in Iterative Methods, IMA Vol. Math. Appl. 60, Springer, New York, 1994, pp. 95–118.
[23] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436.
[24] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards, 45 (1950), pp. 255–282.
[25] C. Lanczos, *Solution of systems of linear equations by minimized iterations*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 33–53.
[26] J. Liesen and Z. Strakoš, *GMRES convergence analysis for a convection-diffusion model problem*, SIAM J. Sci. Comput., 26 (2005), pp. 1989–2009.
[27] S. M. Malamud, *Inverse spectral problem for normal matrices and the Gauss-Lucas theorem*, Trans. Amer. Math. Soc., 357 (2005), pp. 4043–4064.
[28] G. Meurant, *GMRES and the Arioli, Pták and Strakoš parametrization*, BIT, 52 (2012), pp. 687–702.
[29] G. Meurant, *Necessary and sufficient conditions for GMRES complete and partial stagnation*, submitted.
[30] G. Meurant, *Notes on GMRES Convergence* (15): *The Matrix H and the Ritz Values*, private communication, 2011.
[31] G. Meurant and Z. Strakoš, *The Lanczos and conjugate gradient algorithms in finite precision arithmetic*, Acta Numer., 15 (2006), pp. 471–542.
[32] C. C. Paige, *The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices*, Ph.D. thesis, University of London, London, UK, 1971.
[33] C. C. Paige, *Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix*, J. Inst. Math. Appl., 18 (1976), pp. 341–349.

[34] C. C. Paige, *Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem*, Linear Algebra Appl., 34 (1980), pp. 235–258.
[35] C. C. Paige, B. N. Parlett, and H. A. van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numer. Linear Algebra Appl., 2 (1995), pp. 115–133.
[36] B. N. Parlett and G. Strang, *Matrices with prescribed Ritz values*, Linear Algebra Appl., 428 (2008), pp. 1725–1739.
[37] S. C. Reddy and L. N. Trefethen, *Pseudospectra of the convection-diffusion operator*, SIAM J. Appl. Math., 54 (1994), pp. 1634–1649.
[38] Y. Saad, *Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices*, Linear Algebra Appl., 34 (1980), pp. 269–295.
[39] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Algorithms Archit. Adv. Sci. Comput., Manchester University Press, Manchester, UK, 1992.
[40] 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.
[41] D. S. Scott, *How to make the Lanczos algorithm converge slowly*, Math. Comp., 33 (1979), pp. 239–247.
[42] N. Shomron and B. N. Parlett, *Linear algebra meets Lie algebra: The Kostant-Wallach theory*, Linear Algebra Appl., 431 (2009), pp. 1745–1767.
[43] P. Smit, *Generating Identical Ritz Values*, Research Memorandum FEW 696, Faculty of Economics and Business Administration, Tilburg University, Tilburg, The Netherlands, 1995.
[44] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra*, Princeton University Press, Princeton, NJ, 2005.
[45] A. van der Sluis and H. A. van der Vorst, *The rate of convergence of conjugate gradients*, Numer. Math., 48 (1986), pp. 543–560.
[46] H. A. van der Vorst and C. Vuik, *The superlinear convergence behaviour of GMRES*, J. Comput. Appl. Math., 48 (1993), pp. 327–341.
[47] 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.
[48] I. Zavorin, D. P. O'Leary, and H. Elman, *Complete stagnation of GMRES*, Linear Algebra Appl., 367 (2003), pp. 165–183.
[49] J.-P. M. Zemke, *Hessenberg eigenvalue-eigenmatrix relations*, Linear Algebra Appl., 414 (2006), pp. 589–606.