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 generates 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 we treat the stagnation case appropriately.

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 is appended to the previous Hessenberg matrix. The process is well suited to iterative methods with large sparse matrices A. Two popular methods extracting approximate solutions from the generated Hessenberg matrices are the Generalized Minimal Residual (GMRES) method [34] for solving the linear system $Ax = b$ and the Arnoldi method (see, e.g., [32, 33]) 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 [20]. The Lanczos process is at the basis of the Conjugate Gradients (CG) method [19, 21] for hermitian positive definite linear systems and of the Lanczos method for hermitian eigenproblems [20]. 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 a generalization of the Lanczos method. As convergence of the CG and Lanczos methods are well understood, it was 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 eigenvalues of the matrix. In practice, the same is often observed for the GMRES method, but, with possibly non-normal input matrices, the situation becomes more subtle. For example, Greenbaum and Strakoš [18] proved that if a residual norm convergence curve is generated by GMRES, the same curve can be obtained with a matrix having prescribed eigenvalues. Greenbaum, Pták and Strakoš [17] complemented this result by proving that *any* nonincreasing sequence of residual norms can be given by GMRES (a similar result for prescribing the norm of the residual at restarting iterations for the restarted GMRES method can be found in [40]). 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 prescribed spectrum. The results in these papers show that the GMRES residual norm convergence needs not, in general, depend on the eigenvalues of A. Other objects, mostly closely related to eigenvalues, have been considered to explain convergence, for example the pseudo-spectrum [37], the field of values [9] or the numerical polynomial hull [16]. In [39] 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 accelerate convergence of GMRES.

A fundamental tool in the convergence analysis of the Lanczos method for hermitian eigen-problems is the interlacing property for the eigenvalues of the subsequently generated tridiagonal

*Institute of Computer Science, Academy of Sciences of the Czech Republic. Pod Vodárenskou věží 2, 18 207 Praha 8 - Libeň, Czech Republic (duintjertebbens@cs.cas.cz). The work of J. Duintjer Tebbens is a part of the Institutional Research Plan AV0Z10300504 and it was supported by the 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 a visit of G. Meurant to the Nečas Center of Charles University in Prague supported by the grant Jindrich Nečas Center for mathematical modeling, project LC06052 financed by MSMT.

matrices. It enables to prove, among others, the persistence theorem or stabilization of Ritz values (see, e.g., [27, 28, 29] or [25]). There are several generalizations of the interlacing property to normal matrices, see e.g. [12, 1], or the papers [22, 11] with geometric interpretations. However, just as for GMRES, potentially non-normal input matrices make convergence analysis of the Arnoldi method delicate. There is no interlacing property for the principal submatrices of general non-normal matrices, see [36] for a thorough discussion on this topic and its relation to the field of Lie algebra's. In [26] one finds a sufficient and necessary condition for prescribing arbitrary eigenvalues of (not necessarily principal) submatrices of general non-hermitian matrices. It follows from [31] that the principal submatrices of non-normal Hessenberg matrices do not satisfy an interlacing property either. For a detailed spectral analysis of non-normal Hessenberg matrices and their principal submatrices, see also [42].

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 of eigenvectors). Note that this involves many more conditions than prescribing one residual norm per GMRES iteration and the spectrum of the input matrix. 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 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 some words on future work.

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

THEOREM 1.1. *Assume 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

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

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

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

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

$$q(\lambda) = (\lambda - \lambda_1)\cdots(\lambda-\lambda_n) = \lambda^n + \sum_{j=0}^{n-1} \alpha_j \lambda^j, \qquad C^{(n)} = \begin{bmatrix} 0 & -\alpha_0 \\ I_{n-1} & \vdots \\ & -\alpha_{n-1} \end{bmatrix}.$$

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. 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

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

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 $\mathcal{K}_k(A, b)$. 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 lower subdiagonal entries. Prescribing the set $\mathcal{R}$ is only possible if there exist, at all, Hessenberg matrices with positive subdiagonal entries where the eigenvalues of all the leading principal submatrices can be prescribed. In the paper [31] it was 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 [31, Theorem 3]. We here give 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*

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

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

$$U(\mathcal{S}) = I_n - \left[\begin{array}{cccc} 0 & C^{(1)}e_1 & \vdots & \vdots \\ & 0 & C^{(2)}e_2 & \vdots \\ & & 0 & C^{(n-1)}e_{n-1} \\ & & & 0 \end{array}\right], \tag{2.2}$$

then the unique upper Hessenberg matrix $H(\mathcal{R})$ with the entry one along the lower subdiagonal and with the spectrum $\lambda_1, \dots, \lambda_n$ such that the kth leading principal submatrix has eigenvalues $\rho_1^{(k)}, \dots, \rho_k^{(k)}$ for all $k = 1, \dots, n-1$ 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 $\rho_1^{(k)}, \dots, \rho_k^{(k)}$ (uniqueness of $H(\mathcal{R})$ was shown in [31] 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

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

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

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

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 arbitrarily 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 lower 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 lower 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 lower 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 kth leading principal submatrix for all $k = 1, \ldots, n-1$.

Proof. We 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 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

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

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 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. Denote the characteristic polynomial of $\tilde{H}_k$ by $\tilde{p}_k(\lambda)$. By comparing the coefficients (subsequently before λ^k until λ^0) of the polynomial $p_k(\lambda)$ in (2.4) and of the polynomial

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

which must be identical with $p_k(\lambda)$ by assumption, 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 lower subdiagonal of the generated Hessenberg matrix can be prescribed.

Corollary 2.3. *Assume 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 $\lambda_1, \ldots, \lambda_n$, lower 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

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

$U(\mathcal{S})$ 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. It is a stronger result in the sense that we prescribe k values (the k Ritz values) in the kth iteration, whereas Theorem 1.1 prescribes only one value (the kth residual norm); the spectrum of A is prescribed in both results. The corollary shows, surprisingly, that for general non-normal matrices the distribution of the Ritz values generated in the Arnoldi method can be arbitrary and fully independent on the spectrum. Note that there exist some results on the distribution of Ritz values for specific non-normal matrices, for example for Jordan blocks and block diagonal matrices with a simple normal eigenvalue, see [7].

The given parametrization may be useful for convergence analysis of versions of Arnoldi used in practice, e.g. implicitly restarted Arnoldi with polynomial shifts [4, 5]; in particular it may help to better understand (and avoid) cases where Arnoldi with exact shifts fails, see, e.g. [10]. As Ritz values are contained in the field of values, it may also have implications for analysis of iterative methods based on the field of values.

Of course, we here deal with the problem of constructing both an input matrix and an initial vector to produce prescribed Ritz values. With the matrix given, constructing an initial vector to produce prescribed Ritz values was done in [35] for a hermitian matrix. If it has distinct eigenvalues, the paper shows how to construct a perverse initial vector such that the Ritz values in the one but last iteration are as far from the eigenvalues as allowed by the interlacing property (see [11] 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 regard 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 for the quality of the approximate Ritz value-vector pair $(\rho^{(k)}, V^{(k)}y)$. Corollary 2.3 shows that 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, 14], one might expect that in such counterintuitive cases, the misleading behavior of $h_{k+1,k}$ is compensated by $|e_k^T y|$ in (2.5). But consider the following: Let A be parameterized 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 arbitrarily small nonzero entries $\sigma_1, \dots, \sigma_{n-1}$, the matrix $VD_\sigma H(\mathcal{R})D_\sigma^{-1}V^*$ with $D_\sigma = \mathrm{diag}(1, \sigma_1, \dots, \Pi_{j=1}^{n-1}\sigma_j)$ generates the same Ritz values, 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 $D_{\sigma_k} = \mathrm{diag}(1, \sigma_1, \dots, \Pi_{j=1}^{k-1}\sigma_j)$. 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 small enough subdiagonal entries, the value $|e_k^T y_s|$ is small *too* (even if y_s is normalized) and does not compensate for a small σ_k, in spite of possibly diverging

Ritz values. 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.

Let us conclude this section with an example of prescribing Ritz values with the parametrization in Corollary 2.3. As we can prescribe any behavior of Ritz values, we can also prescribe Ritz values that grow further away from the eigenvalues of A in every iteration. We illustrate this behavior for a small example with nice numbers. We prescribe the Ritz values

$$\begin{aligned}\mathcal{R} = \{&1,\\ &(0,2)\,,\\ &(-1,1,3)\,,\\ &(-2,0,2,4)\,,\\ &(1,1,1,1,1)\}\,.\end{aligned}$$

Then the Ritz value companion transform $U(\mathcal{S})$ is

$$U(\mathcal{S}) = \begin{bmatrix} 1 & -1 & 0 & 3 & 0 \\ & 1 & -2 & -1 & 16 \\ & & 1 & -3 & -4 \\ & & & 1 & -4 \\ & & & & 1 \end{bmatrix}.$$

With the companion matrix $C^{(5)}$ for the spectrum $\{1\}$, the matrix $H(\mathcal{R})$ corresponding to the prescribed Ritz values with unit lower subdiagonal is

$$H(\mathcal{R}) = U(\mathcal{S})^{-1} C^{(5)} U(\mathcal{S}) = \begin{bmatrix} 1 & 1 & 0 & -3 & 0 \\ 1 & 1 & 3 & 0 & -31 \\ & 1 & 1 & 6 & 0 \\ & & 1 & 1 & -10 \\ & & & 1 & 1 \end{bmatrix}.$$

With Corollary 2.3, for any unitary matrix V of size five and any diagonal matrix D_σ of size five with positive entries, the matrix

$$V D_\sigma \begin{bmatrix} 1 & 1 & 0 & -3 & 0 \\ 1 & 1 & 3 & 0 & -31 \\ & 1 & 1 & 6 & 0 \\ & & 1 & 1 & -10 \\ & & & 1 & 1 \end{bmatrix} D_\sigma^{-1} V^*$$

with initial Arnoldi vector $\beta V e_1$ for a nonzero β generates the prescribed "diverging" Ritz values. The residual norms (2.5) with $D_\sigma = I_5$ in the above parametrization are

$$\begin{aligned}\{\ & 1,\\ &(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}})\,,\\ &(\tfrac{1}{\sqrt{6}}, \tfrac{1}{\sqrt{10}}, \tfrac{1}{\sqrt{6}})\,,\\ &(\tfrac{1}{\sqrt{19}}, \tfrac{1}{\sqrt{91}}, \tfrac{1}{\sqrt{91}}, \tfrac{1}{\sqrt{19}})\,\}\,.\end{aligned}$$

The same "diverging" Ritz values are generated with the exponentially decreasing values $2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}$ on the lower subdiagonal of the generated Hessenberg matrix: With $D_\sigma = \mathrm{diag}(1, 2^{-1}, 2^{-3}, 2^{-6}, 2^{-10})$ we have the parametrization

$$V \begin{bmatrix} 1 & 2 & 0 & -192 & 0 \\ 0.5 & 1 & 12 & 0 & -15872 \\ & 0.25 & 1 & 48 & 0 \\ & & 0.125 & 1 & -160 \\ & & & 0.0625 & 1 \end{bmatrix} V^* \tag{2.6}$$

of the input matrix. The rounded residual norms (2.5) with this choice of lower subdiagonal entries are

$$\begin{aligned} \{ & \frac{1}{2}, \\ & (0.1118, 0.1118)\,, \\ & (0.011, 0.0052, 0.011)\,, \\ & (0.0006, 0.0001, 0.0001, 0.0006)\,\}\,, \end{aligned}$$

even though the corresponding Ritz values are not converging. Note the large numbers in the last columns of the previous Hessenberg matrix; its condition number is about $1.37 \cdot 10^8$, the condition number of its eigenvector matrix is about $4.19 \cdot 10^{15}$. With $D_\sigma = I_5$ the condition number of the Hessenberg matrix is about 1390, but the condition number of the corresponding eigenvector matrix is still about $2.61 \cdot 10^{13}$.

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 lower 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 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 first parametrization in terms of the second one. 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 $D_\sigma = \operatorname{diag}(1, h_{2,1}, \ldots, \Pi_{j=1}^{n-1} h_{j+1,j})$. 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

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

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

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

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

The previous lemma immediately leads to a relation between U and H, $H = UC^{(n)}U^{-1}$. A similar result is proven in [23]. The second factorization of K 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 we write A as $A = WYC^{(n)}Y^{-1}W^*$ and b as $b = Wh$ 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 $(YC^{(n)}Y^{-1})^j = Y(C^{(n)})^jY^{-1}$. Hence $(YC^{(n)}Y^{-1})^jh = Y(C^{(n)})^je_1$. As we have seen before, $(C^{(n)})^je_1 = e_{j+1}$. This yields

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

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 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* $\lambda_1, \ldots, \lambda_n$, *lower 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*

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

R is the upper triangular matrix

$$R = \Gamma R_t T, \tag{3.2}$$

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 R_t is the upper triangular factor in the Cholesky decomposition

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

The diagonal matrix Γ with unit modulus entries is such that

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

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

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

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 R_t be the upper triangular factor in the Cholesky decomposition

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

let Γ be a diagonal matrix with unit modulus entries and let $R = \Gamma R_t T$. Then

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

is always satisfied and Γ can be chosen such that

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

It follows that

$$\hat{h} = \|b\| R^{-*} t = \|b\| \, \Gamma (R_t T)^{-*} t$$

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

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

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, $K = WY$ and 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 equalities

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

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

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

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

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. $\square$

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 $(R_tT)^{-*}t$ 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. Note that every choice of $\sigma_1, \ldots, \sigma_{n-1}$ in Theorem 3.3 uniquely determines the vector h representing the GMRES convergence curve generated with A and b.

Let us illustrate Theorem 3.3 with the example we used earlier. With the prescribed Ritz values

$$\begin{aligned}\mathcal{R} = \{&1,\\ &(0,2)\,,\\ &(-1,1,3)\,,\\ &(-2,0,2,4)\,,\\ &(1,1,1,1,1)\}\,,\end{aligned}$$

a unit initial vector b and $D_\sigma = \operatorname{diag}(1, 2^{-1}, 2^{-3}, 2^{-6}, 2^{-10})$, the partitioning (3.3) in Theorem 3.3 is

$$D_\sigma U(\mathcal{S})^{-1} = \begin{bmatrix} 1 & 1 & 2 & 4 & 8 \\ & \frac{1}{2} & 1 & 3.5 & 10 \\ & & \frac{1}{4} & 0.375 & 2 \\ & & & \frac{1}{64} & \frac{1}{8} \\ & & & & \frac{1}{1024} \end{bmatrix} \equiv \begin{bmatrix} 1 & t^* \\ 0 & T \end{bmatrix}.$$

Then

$$T^{-*}t = \begin{bmatrix} 2 \\ 0 \\ -192 \\ 0 \end{bmatrix}$$

and the upper triangular factor R_t of the Cholesky decomposition (3.4) is

$$R_t = \begin{bmatrix} \sqrt{5} & 0 & \frac{-384}{\sqrt{5}} & 0 \\ & 1 & 0 & 0 \\ & & \sqrt{\frac{36869}{5}} & 0 \\ & & & 1 \end{bmatrix}. \tag{3.10}$$

Because of

$$(R_tT)^{-*}t = \begin{bmatrix} \sqrt{\frac{4}{5}} \\ 0 \\ -\sqrt{\frac{1}{5} - (0.0052)^2} \\ 0 \end{bmatrix}$$

we can define $\Gamma = \operatorname{diag}(1, 1, -1, -1)$, see (3.5), giving the upper triangular matrix R with rounded entries

$$R = \Gamma R_t T = \begin{bmatrix} 1.118 & 2.2361 & 5.143 & 4.6276 \\ & 0.125 & 0.375 & 2 \\ & & -1.3417 & -5.3669 \\ & & & \frac{-1}{1024} \end{bmatrix}.$$

The vector h in (3.6) and the corresponding GMRES convergence curve, see (1.2), are

$$h = \begin{bmatrix} \sqrt{\frac{4}{5}} \\ 0 \\ \sqrt{\frac{1}{5} - (0.0052)^2} \\ 0 \\ 0.0052 \end{bmatrix},$$

$$f(0) = \|r^{(0)}\| = 1, \quad f(1) = \|r^{(1)}\| = \sqrt{\frac{1}{5}}, \quad f(2) = \|r^{(2)}\| = \sqrt{\frac{1}{5}},$$
$$f(3) = \|r^{(3)}\| = 0.0052, \quad f(4) = \|r^{(4)}\| = 0.0052.$$

In this example we prescribed rapidly decreasing lower subdiagonal entries of the Hessenberg matrix. As might be expected, this leaded to relatively fast GMRES convergence, in spite of completely diverging Ritz values. Can we force any GMRES convergence speed with arbitrary Ritz values ?

The previous example indicates there is a relation between zero Ritz values and stagnation in GMRES. This is well-known: A singular Hessenberg matrix corresponds to an undefinable iterate in the FOM method, which is equivalent to stagnation in the parallel GMRES process, see e.g. [6, 15]. For completeness, we give another proof of this result, formulated with the notation of Theorem 3.3.

Lemma 3.4. *With the notation of Theorem 3.3 and for* $1 \leq k \leq n-1$, *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 upper triangular factor R_t in the Cholesky decomposition

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

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 upper triangular system

$$(\Gamma R_t)^* \hat{h} = T^{-*}t$$

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

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

then we have

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

Proof. The entries of R_h^T are

$$(R_h^T)_{i,k} = -\frac{\eta_i \eta_k}{\sqrt{\eta_{k+1}^2 + \cdots + \eta_n^2}\sqrt{\eta_k^2 + \cdots + \eta_n^2}}, \quad (R_h^T)_{k,k} = \frac{\sqrt{\eta_{k+1}^2 + \cdots + \eta_n^2}}{\sqrt{\eta_k^2 + \cdots + \eta_n^2}}, \tag{3.11}$$

see [13] on the Cholesky decomposition of a rank-one updated identity matrix, or also [24, 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 there 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 any unitary matrix and $C^{(n)}$ is the companion matrix corresponding to the polynomial with roots $\lambda_1, \ldots, \lambda_n$. Y is given by

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

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 $n-1$

$$R \equiv R_h^{-1} D_c^{-*} C^{-1}, \tag{3.12}$$

where C is the trailing principal submatrix in the partitioning

$$U(\mathcal{S}) = \begin{bmatrix} 1 & c^* \\ 0 & C \end{bmatrix}. \tag{3.13}$$

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

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

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

$$R_h^{-T}\hat{h} = -f(0)^2 D_c\, c. \tag{3.14}$$

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 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 (3.12) 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 (3.12) exists. With the assumed partitioning (3.13) 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 (3.12) of R we have

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

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

$$\operatorname{diag}(f(0), D_c^{-*})\, U(\mathcal{S})^{-1} = \begin{bmatrix} f(0) & t^* \\ 0 & T \end{bmatrix}, \tag{3.15}$$

of a diagonal scaling of $U(\mathcal{S})^{-1} = \begin{bmatrix} 1 & -c^*C^{-1} \\ 0 & C^{-1} \end{bmatrix}$. 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 (3.12) satisfies the same conditions as the upper triangular R in (3.2) in Theorem 3.3, it remains to show that $R_h^{-1} = R_t$, $\Gamma = I$, where R_t and Γ are the matrices defined in the second assertion of Theorem 3.3. We have

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

and with $\Gamma = I$

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

Together with

$$\eta_n = f(n-1) = \sqrt{f(0)^2 - (f(0)^2 - f(1)^2) - \ldots - (f(n-2)^2 - f(n-1)^2)} = f(0)\sqrt{1 - \frac{\|\hat{h}\|^2}{f(0)^2}},$$

we have that matrices of the form

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

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

The only freedom 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 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, in general, converging 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 [38]. 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, see [30, 15]. Nevertheless, the extent to which Ritz values and residual norms are independent is astonishing. Note, for example, that for matrices close to normal the bounds derived in [39] 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 could also 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 (3.15); otherwise the assertion needs not be changed. Translated in the notation of Corollary 2.3, this gives the following alternative parametrization.

Corollary 3.7. *Assume 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 $b = \|b\| V e_1$, *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

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

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

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

We conclude this section with an example where all generated Ritz values coincide with the spectrum of the input matrix, but GMRES is nearly stagnating. Note that the previous example demonstrated the opposite behavior: Ritz values where diverging, but GMRES converged rapidly. We now prescribe the Ritz values

$$\begin{aligned}\mathcal{R} \;=\; \{&1,\\ &(1,1)\,,\\ &(1,1,1)\,,\\ &(1,1,1,1)\,,\\ &(1,1,1,1,1)\}\end{aligned}$$

and prescribe the nearly stagnating GMRES residual norms

$$\|r^{(0)}\| = 1, \quad \|r^{(1)}\| = \frac{1}{2}, \quad \|r^{(2)}\| = \frac{1}{3}, \quad \|r^{(3)}\| = \frac{1}{4}, \quad \|r^{(4)}\| = \frac{1}{5}, \quad \|r^{(5)}\| = 0.$$

The prescribed Ritz values give the Ritz value companion transform

$$U(\mathcal{S}) = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 \\ & 1 & -2 & 3 & -4 \\ & & 1 & -3 & 6 \\ & & & 1 & -4 \\ & & & & 1 \end{bmatrix}.$$

From the partitioning (3.13) we have

$$c = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \qquad C^{-1} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ & 1 & 3 & 6 \\ & & 1 & 4 \\ & & & 1 \end{bmatrix}.$$

The prescribed residual norms give $h = [\sqrt{3}/2, \sqrt{5}/6, \sqrt{7}/12, 3/20, 1/5]^T$ and the Cholesky factor R_h of $I - \hat{h}\hat{h}^T$ with $\hat{h} = [\sqrt{3}/2, \sqrt{5}/6, \sqrt{7}/12, 3/20]^T$ takes the form

$$R_h = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{15}}{6} & -\frac{\sqrt{21}}{12} & -\frac{3\sqrt{3}}{20} \\ & \frac{2}{3} & -\frac{\sqrt{35}}{12} & -\frac{3\sqrt{5}}{20} \\ & & \frac{3}{4} & -\frac{3\sqrt{7}}{20} \\ & & & \frac{4}{5} \end{bmatrix},$$

see the formulaes (3.11). Then

$$R_h^{-T}\hat{h} = \begin{bmatrix} \sqrt{3} \\ \sqrt{5} \\ \sqrt{7} \\ 3 \end{bmatrix}, \qquad \text{i.e.} \quad D_c = \operatorname{diag}(\sqrt{3}, -\sqrt{5}, \sqrt{7}, -3),$$

see (3.14). This gives in Theorem 3.6 the matrix $R = R_h^{-1}D_c^{-1}C^{-1}$ with rounded entries

$$R = \begin{bmatrix} 1.1547 & 1.4434 & 1.7321 & 2.0207 \\ & -0.6708 & -1.6398 & -2.9069 \\ & & 0.504 & 1.7953 \\ & & & -0.4167 \end{bmatrix}$$

and the input matrix A is of the form

$$A = W \begin{bmatrix} 1.25 & 0.1076 & 0.0636 & 0.0433 & 0.0577 \\ -0.5809 & 1.1944 & 0.115 & 0.0783 & 0.1043 \\ & -0.7512 & 1.1181 & 0.0803 & 0.1071 \\ & & -0.8268 & 1.0775 & 0.1033 \\ & & & -0.48 & 0.36 \end{bmatrix} W^*$$

for an arbitrary unitary matrix W. The rounded entries in the previous parametrization of A are of moderate size; the condition number of A is 4.34. However, the matrix of eigenvectors of A has condition number around 10^{11}. Note also that the Arnoldi process with this parametrization yields the upper Hessenberg matrix with rounded entries

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5774 & 1 & 0 & 0 & 0 \\ & 0.7746 & 1 & 0 & 0 \\ & & 0.8452 & 1 & 0 \\ & & & 0.8819 & 1 \end{bmatrix}.$$

All its leading principal submatrices H_k are defective with the sole eigenvector e_k. Therefore, the residual norms (2.5) for the corresponding Ritz value-vector pairs are, subsequently,

$$0.5774, \quad 0.7746, \quad 0.8452, \quad 0.8819,$$

i.e. they grow with the iteration number whereas all Ritz values have converged since the very start of the Arnoldi process (this need not be true for the Ritz vectors).

4. Conclusion 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 arbitrarily bad convergence behavior. For GMRES it is known for some time that any non-increasing convergence curve can be generated with any spectrum [17]; 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 we treat the stagnation case correctly).

Seen the success of (modified versions of) the GMRES and Arnoldi methods, a question which arises from the mentioned results is their importance for practice. First let us remark that we assumed exact aritmetics throughout the paper; with large dimensions it might become very difficult to generate prescribed convergence behavior in finite precision. Leaving aside the influence of finite precision, for the GMRES method, it is more or less assumed that the pathological cases arise in the highly non-normal case only. Here we have to distinguish two phenomena: The ability to prescribe arbitrary residual norm curves and the independence of residual norms from eigenvalues. Arbitrary GMRES convergence curves are possible for normal matrices, even for unitary matrices [18, 17]. On the other hand, a sufficient condition for convergence to be dominated by eigenvalues (and Ritz values), is normality of the system matrix. But this is not a necessary condition. For a highly non-normal counterexample, consider GMRES applied to a Jordan block. If the right-hand side is chosen appropriately, convergence is determined by the eigenvalue of the block only. As a consequence of [17] and our results, any GMRES convergence analysis for a significantly non-normal problem based on eigenvalues and Ritz values only, should be justified by a detailed investigation of the involved data structure, in particular the interplay between system matrix and right-hand side. Similarly, versions of restarted GMRES like deflation-based techniques, based on modifying the spectrum of the system matrix or of generated Hessenberg matrices, can be guaranteed to work only for particular classes of problems, for example with normal matrices.

The influence of normal input matrices on the behavior of the Arnoldi method is more confusing. It follows from the uniqueness of Hessenberg matrices with prescribed Ritz values and lower subdiagonal, see Theorem 2.2, that not every set of Ritz values can lead to a normal Hessenberg matrix. The characterization of the Ritz values that can be generated by a normal matrix, apart from their location inside the convex hull of the eigenvalues, appears to the authors to be an open problem. 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. Thus cases of rather pathological Ritz value behavior with normal matrices might still be possible. It seems that, even more than for the GMRES method, tools for convergence analysis in the Arnoldi method must be developed individually for every (class of) problem(s).

The main issue following from 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 [41] or, recently, in [24], where the results are linked with the parametrization in Theorem 1.1. More generally, the question is whether our theory gives some insight on what is a good Arnoldi starting vector, respectively, right-hand side b. Related questions are: Given a matrix A and a residual norm convergence curve does there exist a right-hand side b for which GMRES gives the prescribed norms? Or given A and b, what are the possible convergence curves? Particularly relevant for practice are the questions whether our results are valid for popular restarted versions of Arnoldi or GMRES and whether the case of early termination in the Arnoldi process can be incorporated in the theory. The last two questions are the subject of a forthcoming paper.

Software. At the link 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.

Acknowledgements. The authors are indebted to Zdeněk Strakoš for initiating their work on this topic.

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 (electronic).

[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 nonsymmetric matrices*, Master thesis, Rice University, 2009.

[8] F. Chatelin, *Eigenvalues of matrices*, John Wiley & Sons Ltd., Chichester, 1993. With exercises by Mario Ahués and the author, Translated from the French and with additional material by Walter Ledermann.

[9] M. Eiermann, *Fields of values and iterative methods*, Linear Algebra Appl., 180 (1993), pp. 167–197.

[10] M. Embree, *The Arnoldi eigenvalue iteration with exact shifts can fail*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 1–10.

[11] 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, S-412 96 Göteborg, Sweden, available online at http://www.cs.chalmers.se/pub/num_analysis/reports/, June 1990.

[12] K. Fan and G. Pall, *Imbedding conditions for Hermitian and normal matrices*, Canad. J. Math., 9 (1957), pp. 298–304.

[13] P. E. Gill, G. H. Golub, W. Murray, and M. A. Saunders, *Methods for modifying matrix factorizations*, Math. Comp., 28 (1974), pp. 505–535.

[14] S. Godet-Thobie, *Eigenvalues of Large Highly Nonnormal Matrices*, PhD thesis, University Paris IX, Dauphine, Paris, France, 1993.

[15] 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.

[16] 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.

[17] 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.

[18] A. Greenbaum and Z. Strakoš, *Matrices that generate the same Krylov residual spaces*, in Recent advances in iterative methods, vol. 60 of IMA Vol. Math. Appl., Springer, New York, 1994, pp. 95–118.

[19] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436 (1953).

[20] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Research Nat. Bur. Standards, 45 (1950), pp. 255–282.

[21] ———, *Solution of systems of linear equations by minimized iterations*, J. Research Nat. Bur. Standards, 49 (1952), pp. 33–53.

[22] S. M. Malamud, *Inverse spectral problem for normal matrices and the Gauss-Lucas theorem*, Trans. Amer. Math. Soc., 357 (2005), pp. 4043–4064.

[23] G. Meurant, *GMRES and the Arioli, Pták and Strakoš factorization*, submitted.

[24] ———, *Necessary and sufficient conditions for GMRES complete and partial stagnation*, submitted.

[25] G. Meurant and Z. Strakoš, *The Lanczos and conjugate gradient algorithms in finite precision arithmetic*, Acta Numerica, 15 (2006), pp. 471–542.

[26] G. N. de Oliveira, *Matrices with prescribed characteristic polynomial and a prescribed submatrix. I, II*, Pacific J. Math. 29 (1969), 653-661; ibid., 29 (1969), pp. 663–667.

[27] C. C. Paige, *The computation of eigenvalues and eigenvectors of very large sparse matrices*, PhD thesis, University of London, 1971.

[28] ———, *Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix*, J. Inst. Math. Appl., 18 (1976), pp. 341–349.

[29] ———, *Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem*, Linear Algebra Appl., 34 (1980), pp. 235–258.

[30] 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.

[31] B. Parlett and G. Strang, *Matrices with prescribed Ritz values*, Linear Algebra Appl., 428 (2008), pp. 1725–1739.

[32] Y. Saad, *Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices*, Linear Algebra Appl., 34 (1980), pp. 269–295.

[33] Y. Saad, *Numerical methods for large eigenvalue problems*, Algorithms and Architectures for Advanced Scientific Computing, Manchester University Press, Manchester, 1992.

[34] 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.

[35] D. S. Scott, *How to make the Lanczos algorithm converge slowly*, Math. Comp., 33 (1979), pp. 239–247.

[36] N. Shomron and B. N. Parlett, *Linear algebra meets Lie algebra: the Kostant-Wallach theory*, Linear Algebra Appl., 431 (2009), pp. 1745–1767.

[37] L. N. Trefethen and M. Embree, *Spectra and pseudospectra*, Princeton University Press, Princeton, NJ, 2005.

[38] A. van der Sluis and H. A. van der Vorst, *The rate of convergence of conjugate gradients*, Numer. Math., 48 (1986), pp. 543–560.

[39] H. A. van der Vorst and C. Vuik, *The superlinear convergence behaviour of GMRES*, J. Comput. Appl. Math., 48 (1993), pp. 327–341.

[40] E. Vecharynski and J. Langou, *Any admissible cycle-convergence behavior is possible for restarted GMRES at its initial cycles*, Num. Lin. Algebr. Appl., 18 (2011), pp. 499–511.

[41] I. Zavorin, D. P. O'Leary, and H. Elman, *Complete stagnation of GMRES*, Linear Algebra Appl., 367 (2003), pp. 165–183.

[42] J.-P. M. Zemke, *Hessenberg eigenvalue-eigenmatrix relations*, Linear Algebra Appl., 414 (2006), pp. 589–606.