Noname manuscript No.
(will be inserted by the editor)

The role eigenvalues play in forming GMRES residual norms with non-normal matrices

Gérard Meurant · Jurjen Duintjer Tebbens

Received: date / Accepted: date

Abstract In this paper we give explicit expressions for the norms of the residual vectors generated by the GMRES algorithm applied to a non-normal matrix. They involve the right-hand side of the linear system, the eigenvalues, the eigenvectors and, in the non-diagonalizable case, the principal vectors. They give a complete description of how eigenvalues contribute in forming residual norms and offer insight in what quantities can prevent GMRES from being governed by eigenvalues.

Keywords GMRES convergence · non-normal matrix · eigenvalues · residual norms

1 Introduction

We consider the convergence of GMRES (the Generalized Minimal RESidual method) for solving linear systems with complex nonsingular matrices A of size n and n-dimensional right-hand sides b; see e.g. [38] or [37] for a description of the algorithm. The kth GMRES iterate x_k minimizes, with $x_0 = 0$, the norm of the kth residual vector $r_k = b - Ax_k$ over all vectors in the kth Krylov subspace $\mathcal{K}_k(A, b) \equiv \text{span}\{b,\ Ab,\ \dots,\ A^{k-1}b\}$. Therefore, residual norms are non-increasing and satisfy

$$\|r_k\| = \min_{p \in \pi_k} \|p(A)b\|,$$

G. Meurant
30 rue du sergent Bauchat, 75012 Paris, France.
E-mail: gerard.meurant@gmail.com

J. Duintjer Tebbens
Institute of Computer Science, Academy of Sciences of the Czech Republic. Pod Vodárenskou věží 2, 18 207 Praha 8 - Libeň and Charles University in Prague, Faculty of Pharmacy in Hradec Králové, Heyrovského 1203, 500 05 Hradec Králové.

where π_k is the set of polynomials of degree k with the value one at the origin and $\|\cdot\|$ denotes the 2-norm. If the Jordan canonical form of A is denoted by $A = XJX^{-1}$, then

$$\|r_k\| = \min_{p\in\pi_k} \|X\,p\,(J)X^{-1}b\|. \tag{1}$$

In this paper we focus on how convergence of the GMRES residual norms is influenced by the entirety of spectral properties of A, that is, by the eigenvalues contained in J and by the eigenvectors or principal vectors contained in X.

If A is Hermitian, the orthogonality of the eigenvectors results in a predominant influence of the eigenvalues on convergence. For example, in Hermitian counterparts of GMRES like the MINRES method [34] or the Conjugate Gradients method [19], clustering of eigenvalues stimulates convergence, eigenvalues close to zero hamper convergence and the eigenvalue distribution decides about the rate of convergence (for a survey, see, e.g., [27]). In addition, there exist for these methods sharp upper bounds consisting of a min-max problem which depends on the spectrum only. For instance, in MINRES the residual norms satisfy

$$\frac{\|r_k\|}{\|b\|} \leq \min_{p\in\pi_k} \max_{i=1,\ldots,n} |p_k(\lambda_i)|, \tag{2}$$

with λ_i denoting the eigenvalues of A (see, e.g., [37]) and for every k there exists a right-hand side (depending on k) such that equality holds. If A is a normal matrix, the residual norms generated by GMRES satisfy the same inequality and in this case, GMRES convergence is governed by eigenvalues as well. Moreover, from (1) we have for a normal matrix

$$\|r_k\| = \min_{p\,\in\pi_k} \|p\,(J)X^*b\|, \tag{3}$$

with J being a diagonal matrix of eigenvalues. This shows that the residual norms are fully determined by two quantities: eigenvalues and components of the right-hand side in the eigenvector basis. A closed-form expression for the kth GMRES residual norm in terms of these quantities (in fact of the moduli of the components of the right-hand side in the eigenvector basis), i.e. the solution of (3), was presented in [10] and in an unpublished report Bellalij and Sadok (A new approach to GMRES convergence, 2011).

When A is not normal, the predominant role of the eigenvalues can be lost. For diagonalizable non-normal matrices, the upper bound (2) is multiplied with the condition number $\kappa(X)$ of the eigenvector matrix, which may be large. We refer to [26, Section 3.1] for a detailed discussion of other difficulties with interpreting this bound in the non-normal case. The probably most convincing results showing that GMRES needs not be governed only by eigenvalues can be found in a series of papers by Arioli, Greenbaum, Pták and Strakoš [18, 17, 1]. They show that for any prescribed sequence of n non-increasing residual norms, there exists a class of right-hand sides and matrices, whose nonzero eigenvalues can be chosen *arbitrarily*, giving residual norms that coincide with the given non-increasing sequence. In this sense, GMRES convergence curves (with respect to residual norms) are independent from the eigenvalues of A. It was shown in [8] that convergence curves do not even depend on the Ritz values generated during all iterations of the GMRES process. The strong potential independence from eigenvalues inspired many papers that look for some approaches other than eigenvalue analysis to explain GMRES convergence. They include pseudospectra [33,44], the field of values [11], the polynomial numerical hull [16], potential theory [23], decomposition in normal plus low-rank [20] or comparison with GMRES for non-Euclidean inner products [36]. Though they can be very suited to explain convergence for particular problems, none of the approaches seems to represent a universal tool for GMRES analysis.

Nevertheless for many practical problems, eigenvalues seem to influence convergence behavior strongly. This follows for instance from the fact that slow convergence can often be successfully cured by eliminating particular convergence hampering eigenvalues with a so-called deflation strategy; see, to mention just some of a large number of proposed techniques, for instance [30, 22,7,12,2,24,31,32,5,35,29,14,6,15]. This is not surprising since residual vectors are formed from a matrix polynomial times the right-hand side and matrix polynomials are naturally related to eigenvalues. It is often assumed that the situation where the behavior of GMRES is not or little governed by eigenvalues occurs only for matrices that are far from normal. However, even such a highly non-normal matrix as a Jordan block can yield GMRES convergence curves that are dominated by the size of the involved eigenvalue (this will also be discussed in Section 3 of this paper). In fact, Arioli, Greenbaum, Pták and Strakoš never wrote in [18,17,1] that GMRES convergence does not depend on the eigenvalues. The results in [18,17,1] merely show that there are sets of matrices with different (arbitrary) eigenvalue distributions and right-hand sides giving the same GMRES residual norms. In view of (1) this means that if one modifies eigenvalues, then in order to have the same residual norms, the eigenvectors and/or principal vectors and the right-hand side must and can be modified appropriately.

In this paper we address the interplay of eigenvalues, eigenvectors and the right-hand side with respect to convergence. In the first place, our goal is to show as precisely as possible, how eigenvalues contribute to the computation of residual norms. To this end, we derive closed-form expressions for the residual norms. In the second place, we use these expressions in an attempt to enhance insight in when convergence can be suspected to be dominated by the spectrum and when not. We discuss several interpretations of departure from normality, the role of the right-hand side and the frequently observed convergence hampering influence of eigenvalues close to the origin.

The contents of the paper are as follows. In Section 2 we give an expression of the GMRES residual norms for diagonalizable matrices. Section 3 generalizes the ideas of the previous section for matrices with one Jordan block and Section 4 treats the more general case when the matrix A is not diagonalizable. We formulate some conclusions in the last section. Throughout the paper we will use the phrase ,,convergence is governed by eigenvalues" when convergence depends only on eigenvalues and on components of the right-hand side in the eigenvector basis; eigenvectors and right-hand side do not influence convergence curves otherwise. This is the case for GMRES applied to normal matrices, see (3), for the MINRES method, and, with respect to the norm of the A-error, the Conjugate Gradients method. We will assume that GMRES does not terminate before iteration n. Hence, the Krylov subspaces are of full dimension and their orthogonal bases constructed using the Gram-Schmidt algorithm are well defined. For the sake of simplicity we choose $x_0 = 0$ and we normalize the right-hand side b such that $\|r_0\| = \|b\| = 1$. The vector e_i will denote the ith column of the identity matrix (of appropriate order). The entry on the ith row and in the jth column of a matrix X is denoted by $X_{i,j}$ and $X_{i:j,k:\ell}$ denotes the submatrix of X with rows from i to j and columns from k to ℓ. $X_{i:j,:}$ denotes the submatrix with rows from i to j and with all columns of X.

2 GMRES convergence for diagonalizable matrices

In this section we look for the solution of the minimization problem (1) in terms of J, X and $X^{-1}b$ when A is diagonalizable with spectral factorization $X\Lambda X^{-1}$ where the eigenvalues are contained in $\Lambda = J = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. To this end, we generalize the results in [10] and in the unpublished report Bellalij and Sadok (A new approach to GMRES convergence, 2011) that solved the minimization problem (3) for normal matrices. The next sections will address the non-diagonalizable case.

Let

$$K = \begin{pmatrix} b & Ab & A^2b & \cdots & A^{n-1}b \end{pmatrix},$$

be the Krylov matrix whose first k columns are the natural basis vectors of the Krylov subspace $\mathcal{K}_k(A, b)$ for $1 \le k \le n$ and let $c = X^{-1}b$. Then the Krylov matrix K can be written as $K = X \begin{pmatrix} c & \Lambda c & \cdots & \Lambda^{n-1}c \end{pmatrix}$ and let us define the moment matrix

$$M = K^*K = \begin{pmatrix} c & \Lambda c & \cdots & \Lambda^{n-1}c \end{pmatrix}^* X^*X \begin{pmatrix} c & \Lambda c & \cdots & \Lambda^{n-1}c \end{pmatrix}. \tag{4}$$

For all Krylov subspaces to have full dimension we need the eigenvalues to be distinct and c to have no zero entries. We remark that it is easily seen from the parametrizations in [1] and [9] that any non-increasing GMRES convergence curve is possible for diagonalizable matrices with any distinct eigenvalues. We now try to show how eigenvectors and components of the right-hand side must be modified if we wish to generate the same residual norms with different distinct eigenvalues.

The residual norms in GMRES are given by

$$\|r_k\|^2 = \frac{1}{e_1^T M_{k+1}^{-1} e_1}, \qquad k = 1, \ldots, n-1, \tag{5}$$

where M_{k+1} is the leading principal submatrix of order $k+1$ of M. This result has been proved independently in several papers; see [45, Theorem 4.1], [21, Theorem 2.1] where the result is formulated differently using a pseudo-inverse and [39, Lemma 1] where it is given for real matrices. In [25, Theorem 2.1] and the remarks thereafter it is pointed out that the formula goes back to [40, Section 3 and 4]. As in [10] and in the unpublished report Bellalij and Sadok (A new approach to GMRES convergence, 2011), the $(1,1)$ entry of M_{k+1}^{-1} in (5) will be calculated using Cramer's rule:

$$(M_{k+1}^{-1})_{1,1} = \frac{\det(M_{2:k+1,2:k+1})}{\det(M_{k+1})}. \tag{6}$$

With D_c denoting the diagonal matrix whose diagonal entries c_i are the components of c and with

$$\mathcal{V}_{k+1} = \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^k \\ 1 & \lambda_2 & \cdots & \lambda_2^k \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \cdots & \lambda_n^k \end{pmatrix}, \tag{7}$$

an $n \times (k+1)$ matrix, we see that M_{k+1} in (6) can be written as

$$M_{k+1} = \mathcal{V}_{k+1}^* D_{\bar{c}} X^* X D_c \mathcal{V}_{k+1}. \tag{8}$$

If R^*R is the Cholesky decomposition of X^*X and $F \equiv RD_c\mathcal{V}_{k+1}$, then M_{k+1} is the product F^*F of two rectangular matrices. To compute the determinants of M_{k+1} and $M_{2:k+1,2:k+1}$ in (6) we will use the Cauchy-Binet formula for determinants of products of rectangular matrices: For the product of a $(k \times n)$ matrix G with an $(n \times k)$ matrix H there holds

$$\det(GH) = \sum_{I_k} \det(G_{:,I_k}) \det(H_{I_k,:}).$$

The notation used here is clear from the following definitions, which we will need in the sequel.

Definition 1 With I_k (or J_k) we denote sets of k ordered indices $i_1, \dots, i_k$ such that $1 \le i_1 < \cdots < i_k \le n$. With $\sum_{I_k}$ we denote summation over all such possible ordered index sets. With $\sum_{J_k \ge I_k}$ we denote summation over the ordered index sets J_k that are greater than or equal to a given index set I_k, where greater is understood with respect to lexicographic ordering. With X_{I_k,J_k} we denote the square $k \times k$ submatrix of X whose row and column indices of entries are defined respectively by I_k and J_k. With $\prod_{j_\ell < j_p \in J_k}$ we denote the product over all pairs of indices j_ℓ, j_p in the ordered index set J_k such that $j_\ell < j_p$.

Having outlined the main proof ingredients, we now give the resulting expressions of the residual norm for GMRES processes that do not terminate before iteration n. The next theorem does not contain very elegant formulaes, but it gives the solution of (1) in the case where J is a diagonal matrix.

Theorem 1 *Let A be a diagonalizable matrix with a spectral factorization $X\Lambda X^{-1}$ where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ contains the distinct eigenvalues. Let b be a vector of unit norm such that $c = X^{-1}b$ has no zero entries and let R^*R be the Cholesky decomposition of X^*X. When solving $Ax = b$ with $x_0 = 0$, the GMRES residual norm at iteration $k < n$ satisfies*

$$\|r_k\|^2 = \sigma^N_{k+1}/\sigma^D_k,$$

where

$$\sigma^N_{k+1} = \sum_{I_{k+1}} \left| \sum_{J_{k+1} \geq I_{k+1}} \det(R_{I_{k+1},J_{k+1}})\, c_{j_1} \cdots c_{j_{k+1}} \prod_{j_\ell < j_p \in J_{k+1}} (\lambda_{j_p} - \lambda_{j_\ell}) \right|^2,$$

$\sigma^D_1 = \sum_{i=1}^n \left| \sum_{j \geq i} R_{i,j}\, c_j\, \lambda_j \right|^2$, *and for $k \geq 2$*

$$\sigma^D_k = \sum_{I_k} \left| \sum_{J_k \geq I_k} \det(R_{I_k,J_k})\, c_{j_1} \cdots c_{j_k}\, \lambda_{j_1} \cdots \lambda_{j_k} \prod_{j_\ell < j_p \in J_k} (\lambda_{j_p} - \lambda_{j_\ell}) \right|^2.$$

Proof We apply Cramer's rule (6) to compute the $(1, 1)$ entry of the inverse of M_{k+1}. Let us first consider the determinant of M_{k+1}. By the Cauchy-Binet formula,

$$\det(M_{k+1}) = \sum_{I_{k+1}} |\det(F_{I_{k+1},:})|^2.$$

Thus we have to compute the determinant of $F_{I_{k+1},:}$, a matrix which consists of rows $i_1, \ldots, i_{k+1}$ of $RD_c\mathcal{V}_{k+1}$. It is the product of a $(k + 1) \times n$ matrix that we can write as $(RD_c)_{I_{k+1},:}$ by the $n \times (k + 1)$ matrix $\mathcal{V}_{k+1}$. Once again we can use the Cauchy-Binet formula. Let

$$\mathcal{V}(\lambda_{j_1}, \ldots, \lambda_{j_{k+1}}) = \begin{pmatrix} 1 & \lambda_{j_1} & \cdots & \lambda^k_{j_1} \\ 1 & \lambda_{j_2} & \cdots & \lambda^k_{j_2} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_{j_{k+1}} & \cdots & \lambda^k_{j_{k+1}} \end{pmatrix}$$

which is a square Vandermonde matrix of order $k+1$. Then, taking into account that R is upper triangular,

$$\det(F_{I_{k+1},:}) = \sum_{J_{k+1} \geq I_{k+1}} \det(R_{I_{k+1},J_{k+1}}) c_{j_1} \cdots c_{j_{k+1}} \det(\mathcal{V}(\lambda_{j_1}, \ldots, \lambda_{j_{k+1}})).$$

Moreover, we have (see, e.g. [13])

$$\det(\mathcal{V}(\lambda_{j_1}, \ldots, \lambda_{j_{k+1}})) = \prod_{j_\ell < j_p \in J_{k+1}} (\lambda_{j_p} - \lambda_{j_\ell}).$$

Finally, the determinant of M_{k+1} is

$$\sigma_{k+1}^N = \sum_{I_{k+1}} \left| \sum_{J_{k+1} \geq I_{k+1}} \det(R_{I_{k+1},J_{k+1}}) c_{j_1} \cdots c_{j_{k+1}} \prod_{j_\ell < j_p \in J_{k+1}} (\lambda_{j_p} - \lambda_{j_\ell}) \right|^2 .$$

Let us now consider the determinant of $M_{2:k+1,2:k+1}$ which is a matrix of order k. The computation is essentially the same, except that we have to consider the rows and columns 2 to $k+1$. Therefore, it is not $\mathcal{V}_k$ which is involved any longer but a matrix that can be written as $\Lambda \mathcal{V}_k$. We have

$$M_{2:k+1,2:k+1} = \mathcal{V}_k^* \Lambda^* D_{\bar{c}} R^* R D_c \Lambda \mathcal{V}_k.$$

Then, we have some additional factors arising from the diagonal matrix Λ and we have to consider only sets of k indices I_k and J_k. The determinant of $M_{2:k+1,2:k+1}$ is obtained, for $k > 1$, as

$$\sigma_k^D = \sum_{I_k} \left| \sum_{J_k \geq I_k} \det(R_{I_k,J_k}) c_{j_1} \cdots c_{j_k} \, \lambda_{j_1} \cdots \lambda_{j_k} \prod_{j_\ell < j_p \leq J_k} (\lambda_{j_p} - \lambda_{j_\ell}) \right|^2 .$$

Noting that for $k=1$, the matrix $\mathcal{V}_{I_k}$ reduces to the number one, we have

$$\begin{aligned} \sigma_1^D &= \sum_{I_1} \sum_{J_1 \geq I_1} |\det(R_{I_1,J_1}) c_{j_1} \cdots c_{j_1} \, \lambda_{j_1} \cdots \lambda_{j_1} \, \det(\mathcal{V}_{I_1})|^2 \\ &= \sum_{i=1}^{n} \left| \sum_{j \geq i} R_{i,j} \, c_j \, \lambda_j \det(1) \right|^2 \end{aligned}$$

The residual norm squared is finally given as $\|r_k\|^2 = \sigma_{k+1}^N / \sigma_k^D$.

Theorem 1 shows in what manner the norm of the residual vector depends on the eigenvalues (through eigenvalue products and products of eigenvalue differences), on the eigenvectors (through determinants of submatrices of the Cholesky factor of X^*X) and on $c = X^{-1}b$ (through products of its entries). Theorem 1 seems to support the frequently observed fact that eigenvalues close to the origin tend to hamper convergence. The common explanation for this behavior is that it is difficult for GMRES to construct, when it terminates, a polynomial with the value one in the origin which is zero in an eigenvalue close to zero. Theorem 1 shows that, with diagonalizable matrices, a spectrum close to the origin may cause many terms in the denominators σ_k^D to be close to zero and may give relatively large residual norms. Of course, the papers [18], [17], [1] proved that small eigenvalues need not hamper convergence in general.

As we mentioned in the introduction, a standard upper bound for GMRES residual norms with diagonalizable matrices is

$$\frac{\|r_k\|}{\|b\|} \leq \kappa(X) \min_{p \in \pi_k} \max_{i=1,\dots,n} |p_k(\lambda_i)|, \tag{9}$$

see, e.g., [38]. This bound suggests that the condition number $\kappa(X)$ of the eigenvector matrix plays an important role for convergence behavior. But according to Theorem 1, GMRES residual norms are not explicitly dependent on $\kappa(X)$. The eigenvector matrix X has a large impact, but its inverse is present only through the entries of $c = X^{-1}b$ (which is also clear from (1)). With an appropriate right-hand side, the influence of a large value of $\|X^{-1}\|$ can be eliminated and give a vector c with entries of moderate size.

Note that in Theorem 1 we could have replaced the matrix R by the matrix of the eigenvectors X. In that case, the summations $\sum_{J_{k+1} \geq I_{k+1}}$ and $\sum_{J_k \geq I_k}$ must be replaced by summation over all possible index sets, i.e. with $\sum_{J_{k+1}}$ and $\sum_{J_k}$, respectively. The matrix R was introduced because when the matrix A is normal, we have $X^*X = I$ and hence $R = I$. Then the sums over $J_k \geq I_k$ and $J_{k+1} \geq I_{k+1}$ reduce to only one term ($J_k = I_k$, respectively $J_{k+1} = I_{k+1}$) and we recover the formula in [10] and in the unpublished report Bellalij and Sadok (A new approach to GMRES convergence, 2011).

Theorem 2 *Let A be a normal matrix with distinct eigenvalues and the spectral factorization $X\Lambda X^*$ where $\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$, $X^*X = XX^* = I$. Let b be a vector of unit norm such that all entries of the vector $c = X^*b$ are nonzero. When solving $Ax = b$ with $x_0 = 0$, the GMRES residual norm at iteration $k = 1$ satisfies*

$$\|r_1\|^2 = \frac{\sum_{I_2} \omega_{i_1}\omega_{i_2} \prod_{i_\ell < i_j \in I_2} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{i=1}^{n} \omega_i |\lambda_i|^2}, \tag{10}$$

and for $k = 2, \ldots, n-1$,

$$\|r_k\|^2 = \frac{\sum_{I_{k+1}} \left[\prod_{j=1}^{k+1} \omega_{i_j}\right] \prod_{i_\ell < i_j \in I_{k+1}} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{I_k} \left[\prod_{j=1}^{k} \omega_{i_j} |\lambda_{i_j}|^2\right] \prod_{i_\ell < i_j \in I_k} |\lambda_{i_j} - \lambda_{i_\ell}|^2}, \tag{11}$$

where $\omega_{i_j} = |e_{i_j}^T c|^2$.

We remark that equations (10) and (11) were derived in [28, Theorem 2.1] for $k = n - 1$ and that the residual norms generated by the MINRES method satisfy exactly the same equations (for all k).

When A is normal, GMRES residual norms depend on the eigenvectors and the right-hand side *only* through the sizes ω_i of the squared components of the right-hand side in the eigenvector basis (which is also clear from (3)). Therefore, the role of eigenvalues is much more pronounced than in the non-normal case. If A is close to normal in the sense that $X^*X \approx I$, the Cholesky factor R may be a small perturbation of the identity matrix. Then in the numerators σ_{k+1}^N and denominators σ_k^D of Theorem 1 the involved determinants of submatrices of R may be small except for the choices $J_{k+1} = I_{k+1}$, respectively $J_k = I_k$, but this has to be investigated further. We can, however, derive bounds from Theorem 1 that involve the conditioning of X. We derive them with the help of the following bounds that can be found in [3].

Lemma 1 *Let G and H be two matrices of sizes $n \times (k+1)$ and $n \times n$ respectively, $k \leq n-1$. If the matrix G is of full rank,*

$$\frac{\sigma_{min}(H)^2}{e_1^T(G^*G)^{-1}e_1} \leq \frac{1}{e_1^T(G^*(H^*H)G)^{-1}e_1} \leq \frac{\sigma_{max}(H)^2}{e_1^T(G^*G)^{-1}e_1}. \tag{12}$$

Proposition 1 *Let A be a matrix with distinct eigenvalues and the spectral factorization $X\Lambda X^{-1}$ where $\Lambda = \operatorname{diag}(\lambda_1,\ldots,\lambda_n)$. Let b be a vector of unit norm such that all entries of the vector $c \equiv X^{-1}b$ are nonzero. When solving $Ax=b$ with $x_0=0$, the GMRES residual norm at iteration $k=1$ satisfies*

$$\|r_1\|^2 \geq \sigma_{min}(X)^2 \frac{\sum_{I_2} \omega_{i_1}\omega_{i_2} \prod_{i_\ell < i_j \in I_2} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{i=1}^n \omega_i |\lambda_i|^2},$$

$$\|r_1\|^2 \leq \|X\|^2 \frac{\sum_{I_2} \omega_{i_1}\omega_{i_2} \prod_{i_\ell < i_j \in I_2} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{i=1}^n \omega_i |\lambda_i|^2},$$

and for $k = 2,\ldots,n-1$,

$$\|r_k\|^2 \geq \sigma_{min}(X)^2 \frac{\sum_{I_{k+1}} \left[\prod_{j=1}^{k+1} \omega_{i_j}\right] \prod_{i_\ell < i_j \in I_{k+1}} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{I_k} \left[\prod_{j=1}^{k} \omega_{i_j} |\lambda_{i_j}|^2\right] \prod_{i_\ell < i_j \in I_k} |\lambda_{i_j} - \lambda_{i_\ell}|^2},$$

$$\|r_k\|^2 \leq \|X\|^2 \frac{\sum_{I_{k+1}} \left[\prod_{j=1}^{k+1} \omega_{i_j}\right] \prod_{i_\ell < i_j \in I_{k+1}} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{I_k} \left[\prod_{j=1}^{k} \omega_{i_j} |\lambda_{i_j}|^2\right] \prod_{i_\ell < i_j \in I_k} |\lambda_{i_j} - \lambda_{i_\ell}|^2},$$

where $\omega_{i_j} = |e_{i_j}^T c|^2$.

Proof Because of (5) and (8), we have

$$\|r_k\|^2 = \frac{1}{e_1^T(\mathcal{V}_{k+1}^* D_{\bar{c}}(X^*X)D_c\mathcal{V}_{k+1})^{-1}e_1}.$$

Applying Lemma 1 with $G \equiv D_c\mathcal{V}_{k+1}$ and $H \equiv X$ we obtain

$$\frac{\sigma_{min}(X)^2}{e_1^T(\mathcal{V}_{k+1}^* D_{\bar{c}}D_c\mathcal{V}_{k+1})^{-1}e_1} \leq \|r_k\|^2 \leq \frac{\|X\|^2}{e_1^T(\mathcal{V}_{k+1}^* D_{\bar{c}}D_c\mathcal{V}_{k+1})^{-1}e_1}.$$

The claim follows by realizing that the value $1/e_1^T(\mathcal{V}_{k+1}^* D_{\bar{c}}D_c\mathcal{V}_{k+1})^{-1}e_1$ is precisely the squared residual norm for a linear system with normal matrix having eigenvalues $\lambda_1,\ldots,\lambda_n$ and such that $c = X^{-1}b$.

The bounds in the previous proposition are attained if $\kappa(X)=1$ and are in some sense a two-sided alternative to (9). They show that if $\sigma_{min}(X)$ is close to $\sigma_{max}(X)$, then residual norms behave essentially as in the normal case and are governed by eigenvalues. However, the opposite needs not be true. If $\kappa(X)$

is large, the question whether convergence is dominated by the spectrum of A will depend on the interplay with the entries of $c = X^{-1}b$ and determinants of X. If we wish to derive bounds similar to those in Proposition 1 where the eigenvalues are fully separated from eigenvectors and right-hand side, this can be done as follows.

Proposition 2 *Let A be a matrix with distinct eigenvalues and the spectral factorization $X\Lambda X^{-1}$ where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Let b be a vector of unit norm such that all entries of the vector $c \equiv X^{-1}b$ are nonzero and let D_c denote the diagonal matrix whose diagonal entries c_i are the components of c. When solving $Ax = b$ with $x_0 = 0$, the GMRES residual norm at iteration $k = 1$ satisfies*

$$\|r_1\|^2 \geq \sigma_{min}(XD_c)^2 \frac{\sum_{I_2} \prod_{i_\ell < i_j \in I_2} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{i=1}^n |\lambda_i|^2},$$
$$\|r_1\|^2 \leq \|XD_c\|^2 \frac{\sum_{I_2} \prod_{i_\ell < i_j \in I_2} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{i=1}^n |\lambda_i|^2},$$

and for $k = 2, \ldots, n-1$,

$$\|r_k\|^2 \geq \sigma_{min}(XD_c)^2 \frac{\sum_{I_{k+1}} \prod_{i_\ell < i_j \in I_{k+1}} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{I_k} \left[\prod_{j=1}^k |\lambda_{i_j}|^2\right] \prod_{i_\ell < i_j \in I_k} |\lambda_{i_j} - \lambda_{i_\ell}|^2},$$
$$\|r_k\|^2 \leq \|XD_c\|^2 \frac{\sum_{I_{k+1}} \prod_{i_\ell < i_j \in I_{k+1}} |\lambda_{i_j} - \lambda_{i_\ell}|^2}{\sum_{I_k} \left[\prod_{j=1}^k |\lambda_{i_j}|^2\right] \prod_{i_\ell < i_j \in I_k} |\lambda_{i_j} - \lambda_{i_\ell}|^2}.$$

Proof Because of (5) and (8), we have

$$\|r_k\|^2 = \frac{1}{e_1^T(\mathcal{V}_{k+1}^* D_{\bar{c}}(X^*X)D_c\mathcal{V}_{k+1})^{-1}e_1}.$$

Applying Lemma 1 with $G \equiv \mathcal{V}_{k+1}$ and $H \equiv XD_c$ we obtain

$$\frac{\sigma_{min}(XD_c)^2}{e_1^T(\mathcal{V}_{k+1}^*\mathcal{V}_{k+1})^{-1}e_1} \leq \|r_k\|^2 \leq \frac{\|XD_c\|^2}{e_1^T(\mathcal{V}_{k+1}^*\mathcal{V}_{k+1})^{-1}e_1}.$$

The claim follows in the same way as in the proof of Proposition 1.

The bounds in this proposition may be tight even if the condition number of the eigenvector matrix X is large: $D_c = \operatorname{diag}(c)$ may represent a favorable scaling of the eigenvector matrix. In fact, as D_c contains X^{-1} through $c = X^{-1}b$, in some particular cases the influence of X^{-1} in the product XD_c might be cancelled out by X. For other bounds that incorporate the right-hand side through $X^{-1}b$ we refer to [43], where the scaling of X is also discussed.

Because for diagonalizable matrices, "departure from normality" can be translated to "size of the condition number of the eigenvector matrix", we conclude that GMRES for diagonalizable matrices close to normal will be governed by the spectrum. With a more important departure from normality, the degree to which eigenvalues govern GMRES will depend upon the interplay with determinants of X and entries of $X^{-1}b$; even with a high condition number $\kappa(X)$, GMRES behavior can be governed by the spectrum in particular cases.

3 One Jordan block

We start our investigation of how Theorem 1 can be extended to the non-diagonalizable case by considering the situation where the Jordan canonical form of A has one Jordan block only. Let A have the Jordan form XJX^{-1} with $J = \text{bidiag}\,(\lambda, 1)$ for a nonzero eigenvalue λ and let b be a vector of unit norm such that the last entry of $c = X^{-1}b$ is nonzero (otherwise GMRES terminates before the nth iteration). Then the moment matrix M is

$$M = K^*K = (\,c \quad Jc \quad \cdots \quad J^{n-1}c\,)^*\, X^*X\,(\,c \quad Jc \quad \cdots \quad J^{n-1}c\,)\,.$$

In contrast with the Krylov matrix $(\,c \quad \Lambda c \quad \cdots \quad \Lambda^{n-1}c\,) = D_c\mathcal{V}_n$ in the previous section (see (4) and (7)), the Krylov matrix $(\,c \quad Jc \quad \cdots \quad J^{n-1}c\,)$ cannot be written as the product of a diagonal matrix containing the entries of c with a Vandermonde matrix. Instead, it can be decomposed as

$$(\,c \quad Jc \quad \cdots \quad J^{n-1}c\,) \quad = CE \equiv \tag{13}$$

$$\begin{pmatrix} c_1 & c_2 & \cdots & \cdots & c_n \\ c_2 & c_3 & \cdots & c_n & \\ c_3 & \cdots & c_n & & \\ \vdots & c_n & & & \\ c_n & & & & \end{pmatrix} \begin{pmatrix} 1 & \lambda & \lambda^2 & \cdots & \lambda^{n-1} \\ 0 & 1 & 2\lambda & \cdots & \binom{n-1}{1}\lambda^{n-2} \\ 0 & 0 & 1 & \cdots & \binom{n-1}{2}\lambda^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

where the matrix C is a Hänkel "anti upper triangular" matrix defined by $c_1, \ldots, c_n$,
$0, \cdots, 0$. Here is a small example for illustration: Let $n = 5$ and let all entries of $c = X^{-1}b$ be nonzero. Then the Krylov matrix $(\,c \quad Jc \quad \cdots \quad J^4c\,)$ is

$$\begin{pmatrix} c_1 & \lambda c_1 + c_2 & \lambda^2 c_1 + 2\lambda c_2 + c_3 & \lambda^3 c_1 + 3\lambda^2 c_2 + 3\lambda c_3 + c_4 & \lambda^4 c_1 + 4\lambda^3 c_2 + 6\lambda^2 c_3 + 4\lambda c_4 + c_5 \\ c_2 & \lambda c_2 + c_3 & \lambda^2 c_2 + 2\lambda c_3 + c_4 & \lambda^3 c_2 + 3\lambda^2 c_3 + 3\lambda c_4 + c_5 & \lambda^4 c_2 + 4\lambda^3 c_3 + 6\lambda^2 c_4 + 4\lambda c_5 \\ c_3 & \lambda c_3 + c_4 & \lambda^2 c_3 + 2\lambda c_4 + c_5 & \lambda^3 c_3 + 3\lambda^2 c_4 + 3\lambda c_5 & \lambda^4 c_3 + 4\lambda^3 c_4 + 6\lambda^2 c_5 \\ c_4 & \lambda c_4 + c_5 & \lambda^2 c_4 + 2\lambda c_5 & \lambda^3 c_4 + 3\lambda^2 c_5 & \lambda^4 c_4 + 4\lambda^3 c_5 \\ c_5 & \lambda c_5 & \lambda^2 c_5 & \lambda^3 c_5 & \lambda^4 c_5 \end{pmatrix}$$

with the factorization

$$(\,c \quad Jc \quad \cdots \quad J^4c\,) = \begin{pmatrix} c_1 & c_2 & c_3 & c_4 & c_5 \\ c_2 & c_3 & c_4 & c_5 & \\ c_3 & c_4 & c_5 & & \\ c_4 & c_5 & & & \\ c_5 & & & & \end{pmatrix} \begin{pmatrix} 1 & \lambda & \lambda^2 & \lambda^3 & \lambda^4 \\ & 1 & 2\lambda & 3\lambda^2 & 4\lambda^3 \\ & & 1 & 3\lambda & 6\lambda^2 \\ & & & 1 & 4\lambda \\ & & & & 1 \end{pmatrix}.$$

The $(k+1)$st leading principal submatrix M_{k+1} of M is given by

$$M_{k+1} = (\, c \quad Jc \quad \cdots \quad J^k c\,)^* \, X^* X \, (\, c \quad Jc \quad \cdots \quad J^k c\,)\,.$$

With (13) and defining

$$Y \equiv XC,$$

we have

$$M_{k+1} = (E_{:,1:k+1})^* (XC)^* XCE_{:,1:k+1} = (E_{:,1:k+1})^* Y^* Y E_{:,1:k+1},$$

which can be written as the product $M_{k+1} = F^* F$ of two rectangular matrices where $F \equiv YE_{:,1:k+1}$. The matrix $E_{:,1:k+1}$ depends only on the eigenvalue, the matrix Y contains all information from the principal vectors and the right-hand side. Using exactly the same proof technique as for Theorem 1, we obtain for a single Jordan block the following.

Corollary 1 *Let A be a nonsingular matrix with a single eigenvalue λ and with Jordan form XJX^{-1} where $J = \mathrm{bidiag}\,(\lambda, 1)$. Let b be a vector of unit norm such that the last entry of $c = X^{-1}b$ is nonzero, let E be the eigenvalue matrix defined by (13) and let $Y = XC$, where C is the Hänkel matrix defined in (13). When solving $Ax = b$ with $x_0 = 0$, the GMRES residual norm at iteration $k < n$ satisfies*

$$\|r_k\|^2 = \frac{\sum_{I_{k+1}} \left| \sum_{J_{k+1}} \det(Y_{I_{k+1},J_{k+1}}) \det(E_{J_{k+1},1:k+1}) \right|^2}{\sum_{I_k} \left| \sum_{J_k} \det(Y_{I_k,J_k}) \det(E_{J_k,2:k+1}) \right|^2}. \tag{14}$$

Corollary 1 shows an interplay between eigenvalues, principal vectors and right-hand side which is similar to the interplay between eigenvalues, eigenvectors and right-hand side in Theorem 1. GMRES residual norms are formed from polynomials in the eigenvalue on the one hand and from determinants of the principal vector matrix multiplied with a matrix containing the entries of $X^{-1}b$ on the other hand. The inverse X^{-1} of the matrix of principal vectors X appears only in combination with the right-hand side through the vector $c = X^{-1}b$ and as before, possible ill-conditioning of X does not necessarily have a significant influence on convergence behavior.

One can prove an analogue of Proposition 1 by applying Lemma 1 with $G \equiv CE$ and $H \equiv X$. It would show that if $\kappa(X) = 1$, the behavior of GMRES applied to a very defective matrix is still governed by the eigenvalue. This would correspond to the special and somewhat superficial situation where A has a single Jordan block and where the matrix X is unitary, i.e. the Jordan form of A is $A = XJX^*$. For example, GMRES for a single, plain Jordan block is, in general, strongly governed by the eigenvalue (see, e.g., the results for a single Jordan block in [26] and [42]). Matrices of the form $A = XJX^*$ are far from normal in the sense of being maximally defective. Clearly, this type of departure from normality of A does not decide upon whether GMRES is governed by eigenvalues. As in the previous section, the departure from *orthogonality* of the eigenvector or principal vectors tells us something. If $\kappa(X)$ is large, the degree to which the spectrum governs convergence behavior is influenced by the entries of X and $c = X^{-1}b$ (an analogue of Proposition 2 for one Jordan block is possible too).

Compared with Theorem 1, we see that the summations over J_k and J_{k+1} in Corollary 1 involve all possible index sets, not only those larger or equal I_k, respectively I_{k+1}. This is because Y is not upper triangular. Like in the section on diagonalizable matrices, we could have introduced the Cholesky decomposition R^*R of X^*X and then have defined Y as $Y = RC$. Then Y would be, like C, Hänkel "anti upper triangular" and we can eliminate some terms in the summations exploiting this structure. We did not pursue this idea, because more useful and elegant simplifications of the expression (14) are given by the next lemmas.

The numerator of $\|r_k\|^2$ contains the determinants of $E_{J_{k+1},1:k+1}$ for all index sets J_{k+1}. Their values are given in the following result.

Lemma 2 *For all the sets of $k+1$ indices J_{k+1} in the numerator of (14), the only determinant of $E_{J_{k+1},1:k+1}$ which is non-zero is* $\det(E_{1:k+1,1:k+1}) = 1$.

Proof. We have to consider all the sets of indices j_ℓ such that $1 \le j_1 < \ldots < j_{k+1} \le n$. Since E is upper triangular, all the determinants involving a row of index larger than $k+1$ are zero. The only set of indices J_{k+1} without a row of index larger than $k+1$ is $\{1, 2, \ldots, k+1\}$. The corresponding submatrix is triangular with ones on the diagonal. □

From Lemma 2 there is only one term for the sum over J_{k+1} in the numerator σ^N_{k+1} in (14) and

$$\sigma^N_{k+1} = \sum_{I_{k+1}} \left|\det(Y_{I_{k+1},1:k+1})\right|^2 .$$

We remark that in this case the numerator does not depend on the eigenvalue. For the denominator in (14) we are interested in the determinants of $E_{J_k,2:k+1}$. They are characterized in the following lemma.

Lemma 3 *The k+1 non-zero determinants of $E_{J_k,2:k+1}$ are obtained for the sets of indices J_k not containing an index strictly larger than $k+1$. If those sets are enumerated in lexicographic order, the determinants are respectively $\lambda^k, \lambda^{k-1}, \ldots, \lambda, 1$. Moreover, the denominator σ^D_k for $\|r_k\|^2$ in (14) is*

$$\sigma^D_k = \sum_{I_k} \left|\lambda^k \det(Y_{I_k,\mathcal{I}_1}) + \quad \cdots \quad + \lambda \det(Y_{I_k,\mathcal{I}_k}) + \det(Y_{I_k,\mathcal{I}_{k+1}})\right|^2 ,$$

where $\mathcal{I}_j$, $j = 1, \ldots, k+1$, are the sets of indices with k elements in the ordered combinations of $k+1$ elements enumerated in lexicographic ordering.

Proof. The first claim is obvious since if there is a row index strictly larger than $k+1$ in J_k then there is a zero row in the matrix $E_{J_k,2:k+1}$ and the determinant is zero. The proof of the second claim is by induction on k. For $k=1$ the only nonzero determinants of $E_{J_1,2}$ are, in lexicographical order, $\det(E_{1,2}) = E_{1,2} = \lambda$ and $\det(E_{2,2}) = E_{2,2} = 1$. Let us assume that the claim is true for $k-1$. We have to consider the determinants of submatrices of order k of the $n \times k$ matrix

$$E_{:,2:k+1} = \begin{pmatrix} \lambda & \lambda^2 & \cdots & \lambda^{k-1} & \lambda^k \\ 1 & 2\lambda & \cdots & \binom{k-1}{1}\lambda^{k-2} & \binom{k}{1}\lambda^{k-1} \\ 0 & 1 & \cdots & \binom{k-1}{2}\lambda^{k-3} & \binom{k}{2}\lambda^{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \binom{k}{k-1}\lambda \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

In lexicographic order the first set of indices J_k is $\{1, 2, \ldots, k\}$. We have to consider the determinant of the matrix $E^{(k)}$ obtained from the first k rows of $E_{:,2:k+1}$. Let us compute this determinant using the last column. It is equal to

$$(-1)^{k+1}[\lambda^k \det(E^{(k)}_{-1,1:k-1}) \quad - \quad \binom{k}{1}\lambda^{k-1}\det(E^{(k)}_{-2,1:k-1})$$
$$- \quad \cdots \quad + (-1)^{k-1}\binom{k}{k-1}\lambda \det(E^{(k)}_{-k,1:k-1})],$$

where $\det(E^{(k)}_{-j,1:k-1})$ denotes the determinant of the square submatrix of order $k-1$ of $E^{(k)}$ from columns 1 to $k-1$ with row j removed. Those determinants are given by our induction hypothesis (in reverse order); they are 1, λ, $\ldots$, λ^{k-1}. Therefore we can factor λ^k and we obtain

$$\lambda^k[1 - \binom{k}{1} + \binom{k}{2} - \quad \cdots \quad + (-1)^{k-1}\binom{k}{k-1}].$$

One can see that the sum within brackets is equal to 1 and the determinant we were looking for is λ^k. The proof for the other sets of indices J_k is along the same lines. □

Combining Lemmas 3 and 2 with Corollary 1, we obtain the next theorem. Note that if the given right-hand side is sparse this may influence the nonzero pattern of Y and cause the annihilation of some further determinants.

Theorem 3 *Let A be a nonsingular matrix with a single eigenvalue λ and with Jordan form XJX^{-1} where $J = \mathrm{bidiag}\,(\lambda, 1)$. Let b be a vector of unit* *norm such that the last entry of $c = X^{-1}b$ is nonzero, let E be the eigenvalue matrix defined by (13) and let $Y = XC$, where C is as defined in (13). When solving $Ax = b$ with $x_0 = 0$, the GMRES residual norm at iteration $k < n$ satisfies*

$$\|r_k\|^2 = \frac{\sum_{I_{k+1}} |\det(Y_{I_{k+1},1:k+1})|^2}{\sum_{I_k} |\lambda^k \det(Y_{I_k,\mathcal{I}_1}) + \cdots + \lambda \det(Y_{I_k,\mathcal{I}_k}) + \det(Y_{I_k,\mathcal{I}_{k+1}})|^2}, \tag{15}$$

where $\mathcal{I}_j$, $j = 1, \ldots, k+1$ are the sets of indices with k elements in the ordered combinations of $k+1$ elements enumerated in lexicographic ordering.

Another result for the residual norms generated by GMRES applied to a Jordan block was given in [21]. The expression in that paper contains constants whose values are generally unknown.

We observe from Theorem 3 an interesting, slightly enhanced independence from the spectrum in comparison with diagonalizable matrices: The numerator is fully independent from the eigenvalue and so are the summands $\det(Y_{I_k,\mathcal{I}_{k+1}})$ in the denominator. In the expression for residual norms of Theorem 1 all summands in both numerator and denominator are depending on eigenvalues.

We next consider a very small convection-diffusion model problem where matrices close to a single Jordan block arise. The choice of the number of inner nodes for discretization and of the source term are physically somewhat artificial but we made these choices for the sake of showing that the formulas for the residual norm can be useful.

Consider the one-dimensional convection-diffusion problem on the unit interval $[0, 1]$

$$-\nu u'' + u' = f, \qquad u(0) = u(1) = 0,$$

discretized with finite differences on a regular grid with n inner nodes using upwind differences for the convective term. This gives a linear system where the system matrix A is tridiagonal with entries

$$A = h^{-2}\,\mathrm{tridiag}(-\nu - h, 2\nu + h, -\nu),$$

see, e.g. [41, Section 4]. In the convection dominated case, $\nu \ll h^2$ and A is close to a scaled transposed Jordan block. Let the source term be nonzero only around the first inner node $1/(n+1)$, with the value $(\nu + h)/(-h^2)$ in that node. Then the right-hand side b is a multiple of e_1 and GMRES applied to the pair (A, b) gives the same residual norms as GMRES applied to the pair

$$\left(\frac{-h^2}{\nu + h} I^- A I^-, \frac{-h^2}{\nu + h} I^- b\right), \tag{16}$$

where I^- denotes the (unitary) antidiagonal reversion matrix with ones on the antidiagonal. The matrix $\frac{-h^2}{\nu+h} I^- A I^-$ is a near Jordan block with the eigenvalue $\lambda = -(2\nu + h)/(\nu + h)$, the right-hand side is e_n.

In the left part of Figure 1 we show the GMRES residual norms generated with the pair (16), where $n = 4$ and $\nu = 0.01$ (dashed lines). We also show the convergence curve for the same pair, except that the lower subdiagonal entries of A have been put to zero to obtain a true Jordan block (dotted lines). Clearly, the convergence of GMRES applied to the pair (A, b) is very close to that for a Jordan block with eigenvalue $\lambda = -(2\nu+h)/(\nu+h) = -1.0476$ and right-hand side e_n. Below we give explicit formulaes for the residual norms generated with this Jordan block using Theorem 3. Note that in this example $Y = C = I^-$.

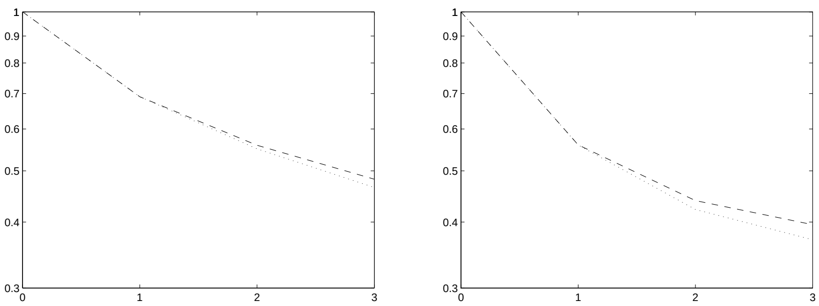

Fig. 1 GMRES residual norm curves for a one-dimensional convection-diffusion model problem with near Jordan block (dashed lines) and with true Jordan block (dotted lines). In the left part the right hand side is e_n, in the right part it is $e_1 + e_n$.

- For $k = 1$, with Lemma 2, the numerator in (15) is

$$\sum_{I_2} |\det(C_{I_2,1:2})|^2.$$

There are six terms for I_2 : $\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}$, with only the last one giving the nonzero determinant $\det(C_{\{3,4\},\{1,2\}}) = -1$. For the denominator in (15) we sum over the trivial index sets $\{1\},\{2\},\{3\}$,$\{4\}$ and $\mathcal{I}_1 = \{1\},\{2\}$. We obtain nonzero values for the index sets $\{3\}$,$\{4\}$ only:

$$|\lambda \det(C_{\{3\},\{1\}}) + \det(C_{\{3\},\{2\}})|^2 = 1, \; |\lambda \det(C_{\{4\},\{1\}}) + \det(C_{\{4\},\{2\}})|^2 = |\lambda|^2.$$

The first residual norm satisfies

$$\|r_1\|^2 = \frac{1}{1+|\lambda|^2}.$$

- For $k = 2$ the numerator in (15) is computed by summation over the sets of ordered indices $\{1,2,3\}$, $\{1,2,4\}$, $\{1,3,4\}$, $\{2,3,4\}$ with only the last one giving the nonzero determinant $\det(C_{\{2,3,4\},1:3}) = -1$.
For the denominator, we have $\mathcal{I}_2 = \{1,2\},\{1,3\},\{2,3\}$, and the outer summation is over the index sets $\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}$. From

these, only those not containing the index 1 lead to non-zero summands (the first three entries of the first row are all zero). Thus

$$\begin{aligned}
&\textstyle\sum_{I_2} \left|\lambda^2 \det(C_{I_2,\{1,2\}}) + \lambda \det(C_{I_2,\{1,3\}}) + \det(C_{I_2,\{2,3\}})\right|^2 \\
&= \left|\lambda^2 \det(C_{\{2,3\},\{1,2\}}) + \lambda \det(C_{\{2,3\},\{1,3\}}) + \det(C_{\{2,3\},\{2,3\}})\right|^2 \\
&+ \left|\lambda^2 \det(C_{\{2,4\},\{1,2\}}) + \lambda \det(C_{\{2,4\},\{1,3\}}) + \det(C_{\{2,4\},\{2,3\}})\right|^2 \\
&+ \left|\lambda^2 \det(C_{\{3,4\},\{1,2\}}) + \lambda \det(C_{\{3,4\},\{1,3\}}) + \det(C_{\{3,4\},\{2,3\}})\right|^2 \\
&= 1 + |\lambda|^2 + |\lambda|^4.
\end{aligned}$$

The square of the norm of the residual at iteration 2 is

$$\|r_2\|^2 = \frac{1}{1 + |\lambda|^2 + |\lambda|^4}.$$

- For $k = 3$ we have only one set of indices for I_4 that is, $\{1, 2, 3, 4\}$. Therefore,

$$\sum_{I_4} |\det(C_{I_4,1:4})|^2 = |\det(C)|^2 = |\det(I^-)|^2 = 1.$$

For the denominator in (15) we have $\mathcal{I}_3 = \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\} = I_3$. It yields

$$\begin{aligned}
&\textstyle\sum_{I_3} |\lambda^3 \det(C_{I_3,\{1,2,3\}}) + \lambda^2 \det(C_{I_3,\{1,2,4\}}) + \lambda \det(C_{I_3,\{1,3,4\}}) + \det(C_{I_3,\{2,3,4\}})|^2 \\
&= |\det(C_{\{1,2,3\},\{2,3,4\}})|^2 + |\lambda \det(C_{\{1,2,4\},\{1,3,4\}})|^2 + |\lambda^2 \det(C_{\{1,3,4\},\{1,2,4\}})|^2 \\
&+ |\lambda^3 \det(C_{\{2,3,4\},\{1,2,3\}})|^2 = 1 + |\lambda|^2 + |\lambda|^4 + |\lambda|^6
\end{aligned}$$

and the last non-zero residual norm satisfies

$$\|r_3\|^2 = \frac{1}{1 + |\lambda|^2 + |\lambda|^4 + |\lambda|^6}.$$

We can easily obtain formulaes for a right-hand side with more nonzero entries. For instance with a source term having the value $(\nu+h)/(-h^2)$ also in the last inner node $n/(n+1)$, we obtain a linear system with a near Jordan block and right-hand side $e_1 + e_n$. The convergence curves for GMRES applied to this system and applied to the same system where the nonzero lower subdiagonal entries have been replaced by zeros, are displayed in the right part of Figure 1. They are very close. Using Theorem 3 we obtain the exact residual norms for the latter system (for which $Y = C$ is the matrix $I^- + e_1 e_1^T$.)

- For $k = 1$, in comparison with the case $b = e_n$, the numerator in (15) contains the additional nonzero determinant $\det(C_{\{1,3\},\{1,2\}}) = b_1 = 1$. For the denominator in (15) we have an additonal nonzero value for the index sets $\{1\}$: $|\lambda \det(C_{\{1\},\{1\}}) + \det(C_{\{1\},\{2\}})|^2 = |\lambda b_1|^2 = |\lambda|^2$. The squared first residual norm is

$$\|r_1\|^2 = \frac{2}{1 + 2|\lambda|^2}.$$

– For $k = 2$, in comparison with the case $b = e_n$, the numerator in (15) also contains the nonzero determinant $\det(C_{\{1,2,3\},1:3}) = -b_1$. For the denominator, the outer summation is over the index sets $\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}$ where $\{1, 2\}, \{1, 3\}$ lead to the additional non-zero summands $|\lambda b_1|^2$ and $|\lambda^2 b_1|^2$, respectively. The square of the norm of the residual at iteration 2 is

$$\|r_2\|^2 = \frac{2}{1 + 2|\lambda|^2 + 2|\lambda|^4}.$$

– For $k = 3$, the numerator in (15) is $\sum_{I_4} |\det(C_{I_4,1:4})|^2 = |\det(C)|^2 = |\det(I^- + e_1 e_1^T)|^2 = 1$. For the denominator, the outer summand for the index set $\{1, 2, 3\}$ takes the value $|\lambda^3 b_1 + 1|^2$ and the remaining summands are unchanged. The last non-zero residual norm satisfies

$$\|r_3\|^2 = \frac{1}{|\lambda^3 + 1|^2 + |\lambda|^2 + |\lambda|^4 + |\lambda|^6}.$$

We see that for these right-hand sides we would have good convergence if the modulus of lambda is large, as one would expect. Nevertheless, it is in general not true that an eigenvalue close to zero hampers convergence for matrices with one Jordan block. If $\lambda \to 0$, then for a given k both the numerator and denominator in (15) go to values independent from λ. The speed of convergence it then fully determined by the entries of X and $X^{-1}b$ and needs not be slow. This is a nice illustration of the limited role of the eigenvalue, i.e. of the theory in the series of papers [18, 17, 1].

4 GMRES for non-diagonalizable matrices

The generalization of Section 3 to multiple Jordan blocks is straightforward. Let A have the Jordan form XJX^{-1} and let it have m ($m \le n$) distinct eigenvalues denoted as $\lambda_1, \lambda_2, \ldots, \lambda_m$. We assume A is non-derogatory because we consider GMRES processes that do not terminate before iteration n. Let the size of the Jordan block J_i corresponding to λ_i be n_i, i.e. $\sum_{i=1}^m n_i = n$, and let us denote by s_i, $i = 1, \ldots, m$ the index of the row where the block J_i starts, to which we add $s_{m+1} = n + 1$. The block J_i goes from row s_i to row $s_{i+1} - 1$. To avoid early termination, we also assume that the right-hand side b is a vector of unit norm such that the entries on positions $s_{i+1} - 1$, $1 \le i \le m$, of $c = X^{-1}b$ are nonzero.

As before, we have

$$M = K^*K = (\,c \quad Jc \quad \cdots \quad J^{n-1}c\,)^*\, X^*X\, (\,c \quad Jc \quad \cdots \quad J^{n-1}c\,)\,.$$

For multiple Jordan blocks, the decomposition (13) can be modified as follows. If we define the rows s_i to $s_{i+1}-1$ of E corresponding to the eigenvalue λ_i as

$$E_{s_i:s_{i+1}-1,:} \equiv \begin{pmatrix} 1 & \lambda_i & \lambda_i^2 & \cdots & \lambda_i^{n_i-1} & \cdots & \lambda_i^{n-1} \\ 0 & 1 & 2\lambda_i & \cdots & \binom{n_i-1}{1}\lambda_i^{n_i-2} & \cdots & \binom{n-1}{1}\lambda_i^{n-2} \\ 0 & 0 & 1 & \cdots & \binom{n_i-1}{2}\lambda_i^{n_i-3} & \cdots & \binom{n-2}{2}\lambda_i^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & \binom{n-1}{n_i-1}\lambda_i^{n-n_i} \end{pmatrix}$$

and the corresponding diagonal block of C as

$$C_{s_i:s_{i+1}-1,s_i:s_{i+1}-1} \equiv \begin{pmatrix} c_{s_i} & c_{s_i+1} & \cdots & \cdots & c_{s_{i+1}-1} \\ c_{s_i+1} & c_{s_i+2} & \cdots & c_{s_{i+1}-1} & \\ c_{s_i+2} & \cdots & c_{s_{i+1}-1} & & \\ \vdots & c_{s_{i+1}-1} & & & \\ c_{s_{i+1}-1} & & & & \end{pmatrix},$$

then

$$\begin{pmatrix} c & Jc & \cdots & J^{n-1}c \end{pmatrix} = CE.$$

The matrix C is block diagonal with Hänkel anti-upper triangular diagonal blocks of order n_i. We again give an example to illustrate.

Consider a matrix $A = XJX^{-1}$ of order 5 with J defined as

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & & \\ & & & \mu & 1 \\ & & & & \mu \end{pmatrix}, \tag{17}$$

where λ and μ ($\lambda \neq \mu$) are given complex numbers different from 0. Let $c = X^{-1}b$, where b is the right-hand side, and let c have no zero entries. Then the Krylov matrix $\begin{pmatrix} c & Jc & \cdots & J^{n-1}c \end{pmatrix}$ is

$$\begin{pmatrix} c_1 & \lambda c_1 + c_2 & \lambda^2 c_1 + 2\lambda c_2 + c_3 & \lambda^3 c_1 + 3\lambda^2 c_2 + 3\lambda c_3 & \lambda^4 c_1 + 4\lambda^3 c_2 + 6\lambda^2 c_3 \\ c_2 & \lambda c_2 + c_3 & \lambda^2 c_2 + 2\lambda c_3 & \lambda^3 c_2 + 3\lambda^2 c_3 & \lambda^4 c_2 + 4\lambda^3 c_3 \\ c_3 & \lambda c_3 & \lambda^2 c_3 & \lambda^3 c_3 & \lambda^4 c_3 \\ c_4 & \mu c_4 + c_5 & \mu^2 c_4 + 2\mu c_5 & \mu^3 c_4 + 3\mu^2 c_5 & \mu^4 c_4 + 4\mu^3 c_5 \\ c_5 & \mu c_5 & \mu^2 c_5 & \mu^3 c_5 & \mu^4 c_5 \end{pmatrix}$$

and can be factorized as

$$\begin{pmatrix} c & Jc & \cdots & J^{n-1}c \end{pmatrix} = \begin{pmatrix} c_1 & c_2 & c_3 & 0 & 0 \\ c_2 & c_3 & 0 & 0 & 0 \\ c_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_4 & c_5 \\ 0 & 0 & 0 & c_5 & 0 \end{pmatrix} \begin{pmatrix} 1 & \lambda & \lambda^2 & \lambda^3 & \lambda^4 \\ 0 & 1 & 2\lambda & 3\lambda^2 & 4\lambda^3 \\ 0 & 0 & 1 & 3\lambda & 6\lambda^2 \\ 1 & \mu & \mu^2 & \mu^3 & \mu^4 \\ 0 & 1 & 2\mu & 3\mu^2 & 4\mu^3 \end{pmatrix},$$

with a block diagonal matrix C.

Let, as before,

$$Y \equiv XC.$$

Then if the $(k+1)$st leading principal submatrix M_{k+1} of M is written as

$$\begin{aligned} M_{k+1} &= (\, c \quad Jc \quad \cdots \quad J^k c\,)^* \, X^* X \, (\, c \quad Jc \quad \cdots \quad J^k c\,) \\ &= E^*_{:,1:k+1} C^* X^* X C E_{:,1:k+1} = (Y E_{:,1:k+1})^* Y E_{:,1:k+1}, \end{aligned}$$

we immediately obtain, again by using the proof technique of Theorem 1, the formula

$$\|r_k\|^2 = \frac{\sum_{I_{k+1}} |\sum_{J_{k+1}} \det(Y_{I_{k+1},J_{k+1}}) \det(E_{J_{k+1},1:k+1})|^2}{\sum_{I_k} \left|\sum_{J_k} \det(Y_{I_k,J_k}) \det(E_{J_k,2:k+1})\right|^2}. \tag{18}$$

The formula is the same as the one presented in Corollary 1, but of course, Y and E are here generalizations of the Y and E in Corollary 1. E represents all the influence of eigenvalues and Y all the influence of eigenvectors, principal vectors and right-hand side. The remarks in Sections 2 and 3 on the role of $\kappa(X)$ and of $X^{-1}b$ apply to this section, too.

A difference is that the interplay between the distinct eigenvalues will play a role. The determinants of $E_{J_{k+1},1:k+1}$ and $E_{J_k,2:k+1}$ may contain eigenvalue differences. For example, so do most determinants of E involved in forming $\|r_3\|^2$ for the matrix J in (17), see Tables 1 and 2. All determinants in Table 1 have $\mu-\lambda$ as a factor. Hence they may be small if μ is close to λ. This suggests that eigenvalue clusters accelerate convergence whereas outliers cause delay, which is often true (see, e.g., [4]). If $\mu = \lambda$, corresponding to two Jordan blocks with the same eigenvalue, we have early termination, $\|r_3\| = 0$ (in exact arithmetic).

We now investigate whether with non-diagonalizable matrices, GMRES residual norms are slightly less dependent on eigenvalues than with diagonalizable matrices in the sense that not all summands in (18) depend upon eigenvalues. We have seen with Theorem 3 that this holds for matrices with a single Jordan block.

For simplicity, we first we address the case $k = 1$. Let us consider the determinants in the numerator of (18), i.e. the determinants of $E_{J_2,\{1,2\}}$ for the set of indices J_2. There are $n!/(2(n-2)!)$ of them. But the rows that are involved are only of three different types whatever the dimension n is. The first type that we can denote as $t_1(\lambda_i)$ is $t_1(\lambda_i) = (\, 1 \quad \lambda_i\,)$, for an eigenvalue λ_i. The two other types are $t_2 = (\, 0 \quad 1\,)$ and $t_3 = (\, 0 \quad 0\,)$. The two last types may or may not exist depending on the values of n_i, $i = 1, \ldots, m$. We have only three kinds of non-zero determinants

$$\begin{vmatrix} 1 & \lambda_i \\ 1 & \lambda_j \end{vmatrix} = \lambda_j - \lambda_i, \quad \begin{vmatrix} 1 & \lambda_i \\ 0 & 1 \end{vmatrix} = 1, \quad \begin{vmatrix} 0 & 1 \\ 1 & \lambda_i \end{vmatrix} = -1. \tag{19}$$

Then in the terms

$$\left| \sum_{J_2} \det(Y_{I_2,J_2}) \det(E_{J_2,1:2}) \right|^2 ,$$

of the numerator of (18), the sum runs over the set of indices such that $\det(E_{J_2,1:2}) \neq 0$ that is, such that we have one of the three kinds of determinant listed above. With the second and third kind there is no dependence on eigenvalues. For the denominator of (18) we can proceed similarly. Thus, depending on the sizes of the individual Jordan blocks, a number of summands is independent from the spectrum.

Table 1 Determinants of $E_{J_4,1:4}$ for the numerator in (18) with $k = 3$, for the matrix J in (17).

Indices in J_4	value
{1,2,3,4}	$(\mu - \lambda)^3$
{1,2,3,5}	$3(\mu - \lambda)^2$
{1,2,4,5}	$(\mu - \lambda)^4$
{1,3,4,5}	$-2(\mu - \lambda)^3$
{2,3,4,5}	$3(\mu - \lambda)^2$

Table 2 Determinants of $E_{J_3,2:4}$ for the denominator in (18) with $k = 3$, for the matrix J in (17).

Indices in J_3	value	Indices in J_3	value
{1,2,3}	λ^3	{1,4,5}	$\lambda\mu^2(\mu - \lambda)^2$
{1,2,4}	$\lambda^2\mu(\mu - \lambda)^2$	{2,3,4}	$\mu[(\mu - \lambda)^2 + \lambda(2\lambda - \mu)]$
{1,2,5}	$\lambda^2(\mu - \lambda)(3\mu - \lambda)$	{2,3,5}	$3(\mu - \lambda)^2$
{1,3,4}	$\lambda\mu(\mu - \lambda)(\mu - 2\lambda)$	{2,4,5}	$\mu^2(\mu - \lambda)(\mu - 3\lambda)$
{1,3,5}	$\lambda[2(\mu - \lambda)^2 + \mu(\mu - 2\lambda)]$	{3,4,5}	$\mu^2(3\lambda - 2\mu)$

For $k > 1$ we have the following straightforward result.

Proposition 3 *If $k < \max_i(n_i)$, then in formula (18) there are determinants of both $E_{J_{k+1},1:k+1}$ and $E_{J_k,2:k+1}$ that are equal to 1.*

Proof The result is obvious since some of the submatrices are upper triangular with ones on the diagonal.

It is not difficult to see that when an eigenvalue approaches zero, this gives determinants tending to be independent on that eigenvalue. Similarly to the previous section, the influence of the corresponding Jordan block on GMRES is then fully determined by the right-hand side and eigenvectors and/or principal vectors and consequently, eigenvalues close to the origin do not seem to necessarily hamper convergence.

5 Conclusion

We presented the solution of the minimization problem (1) for GMRES residual norms generated with general diagonalizable and with non-diagonalizable matrices. It is explicitly formulated in a closed form, unlike the norms of the GMRES residuals in GMRES computations. The solution is not simple and has no immediate practical application but it completely describes the mechanism of forming the residual norm from eigenvalues, eigenvectors or principal vectors and the right-hand side. It shows in what (complicated) way eigenvalues influence GMRES convergence. Other objects than eigenvalues may lead to more elegant formulaes, but if we wish to know the exact influence of eigenvalues, the presented closed-form expressions give the answer. In the diagonalizable case, it is eigenvalue products and products of eigenvalue differences that influence the residual norm. In the non-diagonalizable case, more general polynomials in eigenvalues play a role in forming the residual norm and small eigenvalues are less prone to hamper convergence. Eigenvectors (principal vectors) influence residual norms in two ways. Determinants of the eigenvector (principal vector) matrix play the most important role. The inverse of this matrix contributes only in the form of its product with the right-hand side. As for the right-hand side, it contributes only through its components in the eigenvector (principal vector) basis. The degree to which GMRES is governed by eigenvalues is not so much determined by the departure from diagonalizability of the system matrix, but in general more by the departure from orthogonality of the eigenvector (principal vector) matrix X. With a small value of $\kappa(X)$, GMRES is governed by the spectrum even if the system matrix is defective; with a larger value of $\kappa(X)$ GMRES may or may not be governed by the spectrum, depending on X, $X^{-1}b$ and the interplay between them.

Future work includes extension to other Krylov methods.

Acknowledgements

We thank Zdeněk Strakoš for stimulating work in this direction. The work of the second author was supported by the institutional support RVO:67985807 and by the grant GA13-06684S of the Grant Agency of the Czech Republic.

References

1. M. Arioli, V. Pták, and Z. Strakoš. Krylov sequences of maximal length and convergence of GMRES. *BIT*, 38(4):636–643, 1998.
2. J. Baglama, D. Calvetti, G. H. Golub, and L. Reichel. Adaptively preconditioned GMRES algorithms. *SIAM J. Sci. Comput.*, 20(1):243–269, 1998.
3. M. Bellalij, K. Jbilou, and H. Sadok. New convergence results on the global GMRES method for diagonalizable matrices. *J. Comp. Appl. Math.*, 219:350–358, 2008.
4. S. L. Campbell, I. C. F. Ipsen, C. T. Kelley, and C. D. Meyer. GMRES and the minimal polynomial. *BIT*, 36(4):664–675, 1996.

5. B. Carpentieri, I. S. Duff, and L. Giraud. A class of spectral two-level preconditioners. *SIAM J. Sci. Comput.*, 25(2):749–765 (electronic), 2003.
6. B. Carpentieri, L. Giraud, and S. Gratton. Additive and multiplicative two-level spectral preconditioning for general linear systems. *SIAM J. Sci. Comput.*, 29(4):1593–1612 (electronic), 2007.
7. A. Chapman and Y. Saad. Deflated and augmented Krylov subspace techniques. *Numer. Linear Algebra Appl.*, 4(1):43–66, 1997.
8. J. Duintjer Tebbens and G. Meurant. Any Ritz value behavior is possible for Arnoldi and for GMRES. *SIAM J. Matrix Anal. Appl.*, 33(3):958–978, 2012.
9. J. Duintjer Tebbens and G. Meurant. Prescribing the behavior of early terminating GMRES and Arnoldi iterations. *Num. Alg.*, 65(1):69–90, 2014.
10. J. Duintjer Tebbens, G. Meurant, H. Sadok, and Z. Strakoš. On investigating GMRES convergence using unitary matrices. *Lin. Alg. Appl.*, 450:83–107, 2014.
11. M. Eiermann. Fields of values and iterative methods. *Linear Algebra Appl.*, 180:167–197, 1993.
12. J. Erhel, K. Burrage, and B. Pohl. Restarted GMRES preconditioned by deflation. *J. Comput. Appl. Math.*, 69(2):303–318, 1996.
13. W. Gautschi. On inverses of Vandermonde and confluent Vandermonde matrices. III. *Numer. Math.*, 29:445–450, 1977/78.
14. L. Giraud, S. Gratton, and E. Martin. Incremental spectral preconditioners for sequences of linear systems. *Appl. Numer. Math.*, 57(11-12):1164–1180, 2007.
15. L. Giraud, S. Gratton, X. Pinel, and X. Vasseur. Flexible GMRES with deflated restarting. *SIAM J. Sci. Comput.*, 32(4):1858–1878, 2010.
16. A. Greenbaum. Generalizations of the field of values useful in the study of polynomial functions of a matrix. *Linear Algebra Appl.*, 347:233–249, 2002.
17. A. Greenbaum, V. Pták, and Z. Strakoš. Any nonincreasing convergence curve is possible for GMRES. *SIAM J. Matrix Anal. Appl.*, 17(3):465–469, 1996.
18. A. Greenbaum and Z. Strakoš. Matrices that generate the same Krylov residual spaces. In *Recent advances in iterative methods*, volume 60 of *IMA Vol. Math. Appl.*, pages 95–118. Springer, New York, 1994.
19. M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49:409–436, 1952.
20. M. Huhtanen and O. Nevanlinna. Minimal decompositions and iterative methods. *Numer. Math.*, 86(2):257–281, 2000.
21. I. C. F. Ipsen. Expressions and bounds for the GMRES residual. *BIT*, 40(3):524–535, 2000.
22. S. A. Kharchenko and A. Yu. Yeremin. Eigenvalue translation based preconditioners for the GMRES(k) method. *Numer. Linear Algebra Appl.*, 2(1):51–77, 1995.
23. A. B. J. Kuijlaars. Convergence analysis of Krylov subspace iterations with methods from potential theory. *SIAM Rev.*, 48(1):3–40 (electronic), 2006.
24. C. Le Calvez and B. Molina. Implicitly restarted and deflated GMRES. *Numer. Algorithms*, 21(1-4):261–285, 1999. Numerical methods for partial differential equations (Marrakech, 1998).
25. J. Liesen, M. Rozložník, and Z. Strakoš. Least squares residuals and minimal residual methods. *SIAM J. Sci. Comput.*, 23(5):1503–1525, 2002.
26. J. Liesen and Z. Strakoš. Convergence of GMRES for tridiagonal Toeplitz matrices. *SIAM J. Matrix Anal. Appl.*, 26(1):233–251 (electronic), 2004.
27. J. Liesen and P. Tichý. Convergence analysis of Krylov subspace methods. *GAMM Mitt. Ges. Angew. Math. Mech.*, 27(2):153–173, 2004.
28. J. Liesen and P. Tichý. The worst-case GMRES for normal matrices. *BIT*, 44(1):79–98, 2004.
29. D. Loghin, D. Ruiz, and A. Touhami. Adaptive preconditioners for nonlinear systems of equations. *J. Comput. Appl. Math.*, 189(1-2):362–374, 2006.
30. R. B. Morgan. A restarted GMRES method augmented with eigenvectors. *SIAM J. Matrix Anal. Appl.*, 16(4):1154–1171, 1995.
31. R. B. Morgan. Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations. *SIAM J. Matrix Anal. Appl.*, 21(4):1112–1135, 2000.
32. R. B. Morgan. GMRES with deflated restarting. *SIAM J. Sci. Comput.*, 24(1):20–37 (electronic), 2002.

33. N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen. How fast are nonsymmetric matrix iterations? *SIAM J. Matrix Anal. Appl.*, 13(3):778–795, 1992. Iterative methods in numerical linear algebra (Copper Mountain, CO, 1990).
34. C. C. Paige and M. A. Saunders. LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8(1):43–71, 1982.
35. M. L. Parks, E. de Sturler, G. Mackey, D. D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems. *SIAM J. Sci. Comput.*, 28(5):1651–1674 (electronic), 2006.
36. J. Pestana and A. Wathen. On choice of preconditioner for minimum residual methods for nonsymmetric matrices. In *Householder Symposium XVIII*, Proceedings of the Householder Symposium on Numerical Algebra, Lake Tahoe, USA, June 12-17, pages 180–181. 2011.
37. Y. Saad. *Iterative methods for sparse linear systems.* Society for Industrial and Applied Mathematics, Philadelphia, PA, second edition, 2003.
38. Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7(3):856–869, 1986.
39. H. Sadok. Analysis of the convergence of the minimal and the orthogonal residual methods. *Numer. Algorithms*, 40(2):201–216, 2005.
40. G. W. Stewart. Collinearity and least squares regression. *Statist. Sci.*, 2(1):68–100, 1987.
41. M. Stynes. Steady-state convection-diffusion problems. *Acta Numer.*, 14:445–508, 2005.
42. P. Tichý, J. Liesen, and V. Faber. On worst-case GMRES, ideal GMRES, and the polynomial numerical hull of a Jordan block. *Electron. Trans. Numer. Anal.*, 26:453–473 (electronic), 2007.
43. D. Titley-Peloquin, J. Pestana, and A. Wathen. GMRES convergence bounds that depend on the right-hand side vector. *IMA J. Numer. Anal.*, 34(2):462–479, 2014.
44. L. N. Trefethen and M. Embree. *Spectra and pseudospectra.* Princeton University Press, Princeton, NJ, 2005.
45. J. Zítko. Generalization of convergence conditions for a restarted GMRES. *Numer. Linear Algebra Appl.*, 7:117–131, 2000.