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

On the convergence of Q-OR and Q-MR Krylov methods for solving nonsymmetric linear systems

Gérard Meurant · Jurjen Duintjer Tebbens

Received: date / Accepted: date

Abstract This paper addresses the convergence behavior of Krylov methods for nonsymmetric linear systems which can be classified as Q-OR (quasi-orthogonal) or Q-MR (quasi-minimum residual) methods. It explores, more precisely, whether the influence of eigenvalues is the same when using non-orthonormal bases as it is for the FOM and GMRES methods. It presents parametrizations of the classes of matrices with a given spectrum and right-hand sides generating prescribed Q-OR/Q-MR (quasi) residual norms and discusses non-admissible residual norm sequences. It also gives closed-form expressions of the Q-OR/Q-MR (quasi) residual norms as functions of the eigenvalues and eigenvectors of the matrix of the linear system.

Keywords Q-OR method · Q-MR method · BiCG · QMR · CMRH · eigenvalue influence · prescribed convergence

Mathematics Subject Classification (2000) 65F10

1 Introduction

We consider the problem of solving linear systems $Ax = b$ where A is a square nonsingular matrix of order n with real or complex entries and b is a vector of length n. The probably most popular iterative methods for solving such

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.

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

Jurjen 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 and Faculty of Pharmacy in Hradec Králové, Heyrovského 1203, 500 05 Hradec Králové, Czech Republic
E-mail: duintjertebbens@cs.cas.cz

a possibly nonsymmetric linear system are Krylov methods. Assuming the initial guess is zero, they are based on repeated multiplication of b with A to form Krylov subspaces $\mathcal{K}_k(A,b) \equiv \text{span}\{b, Ab, A^2b, \dots, A^{k-1}b\}$, $k = 1, 2, \dots$, of growing dimension. The approximations x_k to the solution of the linear system are extracted from these subspaces. Interesting references for Krylov methods include the books [17], [39] and [25].

A paper describing these methods in an abstract framework is [12]. In [12] it is shown that most Krylov methods can be described as so-called quasi-orthogonal (Q-OR) or quasi-minimum (Q-MR) residual methods. Many numerical methods that can be classified as Q-OR or Q-MR methods have been proposed over the years, starting in the early fifties. Generally they come into Q-OR/Q-MR pairs. Considering only the nonsymmetric case, probably the most famous one is the FOM (Full Orthogonalization Method) and GMRES (Generalized Minimum RESidual) method pair. They are not quasi but true OR and MR methods in the sense that in the kth iteration, the matrix V_k whose columns form a basis of $\mathcal{K}_k(A,b)$ is defined to have orthonormal columns (provided $\mathcal{K}_k(A,b)$ has dimension k). GMRES is an MR method (that is, minimizing the true residual norm, $\|b - Ax_k\|$) proposed by Saad and Schultz [30] and FOM is an OR method introduced by Saad [28], [29]. The matrix V_k and the upper Hessenberg matrix $H_k = V_k^* A V_k$ are computed column by column using the Arnoldi process [2].

Another famous pair is BiCG/QMR. The BiCG (BiConjugate Gradient) algorithm was derived by Fletcher [13] from the nonsymmetric Lanczos algorithm [24]. The BiCG and QMR (Quasi Minimum Residual) methods use a basis that is bi-orthogonal to a basis of the Krylov subspace $\mathcal{K}_k(A^*, b)$, where A^* denotes the conjugate transpose of A (another initial vector than b can be chosen to generate this auxiliary subspace). The practical interest of these methods is that they use only short recurrences contrary to FOM/GMRES. BiCG is a particular implementation of the Q-OR method and QMR introduced by Freund and Nachtigal [14–16] minimizes the quasi-residual norm; see below for details. Note that in this paper we will use the acronym Q-MR in two different ways. The first one (Q-MR) is to denote the general class of quasi-minimum residual methods and the second one (QMR) is related to the particular method introduced in [14].

Another pair of methods is Hessenberg/CMRH. CMRH (Changing Minimal Residual method based on the Hessenberg process), introduced by Sadok [31], is a Q-MR method which uses the Hessenberg basis computed with an LU factorization with partial pivoting of the Krylov matrix without explicitly computing the Krylov matrix. The Hessenberg method [20] is the corresponding Q-OR method. Other examples of Krylov methods that do not use orthonormal bases include CGS (Conjugate Gradient Squared) introduced by Sonneveld [36], BiCGStab (BiCG Stabilized) by van der Vorst [38], truncated methods [41], [21], [22] and restarted methods [34]. Many of them can be classified as a Q-OR- or Q-MR-type method, but BiCGStab, for instance, is a combination of the two types.

Krylov methods using non-orthonormal bases mostly work with short recurrences, although there are exceptions like CMRH; see [32]. The price paid for the relatively low computational costs with short recurrences are less favorable stability properties; Krylov methods using non-orthonormal bases are called non-optimal methods in [35]. From the theoretical point of view, the usage of non-orthonormal bases leads to methods whose convergence behavior is more complicated to analyse than for their orthonormal counterparts. The residual norms generated in the GMRES method can be described, theoretically, by a particularly simple and natural minimization property. This is probably the main cause of the fact that the majority of convergence results about Krylov methods for nonsymmetric matrices concern the GMRES method (and to a lesser extent FOM).

In this paper we make an attempt to generalize some convergence results for the FOM/GMRES pair to other Q-OR/Q-MR pairs which do not employ orthonormal bases. We are interested in a particular type of convergence results about Krylov methods for nonsymmetric matrices, namely those addressing the extent to which eigenvalues influence the convergence behavior. For the GMRES method, answers to this question can be found in the series of papers [1,19,18], [8,9], [11], [27], [26]; restarted GMRES and FOM were addressed in [40], [33] and [10]. They show that in general, any residual norm history (with non-increasing norms) is possible with any spectrum of the system matrix and thus convergence behavior cannot be determined by the eigenvalue distribution alone. It is not clear whether the same holds for Krylov methods using non-orthonormal bases, although they do work with essentially the same Krylov subspaces. Rather than investigating this and related questions method by method we will adopt the general framework of [12] and we will not consider any particular implementation.

The paper is organized as follows. Section 2 describes the general framework for the Q-OR and Q-MR Krylov methods we are considering and introduces a decomposition of Hessenberg matrices which is important for the paper. In section 3 we point out differences between Q-OR/Q-MR methods and FOM/GMRES. Section 4 considers the problem of constructing a matrix, with a freely chosen spectrum, and a right-hand side yielding prescribed Q-OR residual norm or Q-MR quasi-residual norm convergence curves for a given Q-OR or Q-MR method. Depending on the method, this amounts to construct a Hessenberg matrix with a particular non-zero structure. It turns out that for some Q-OR methods not every residual norm history is possible. Section 5 gives expressions for the Q-OR residual norms or Q-MR quasi-residual norms as a function of the eigenvalues and eigenvectors of A, the right-hand side b and the matrix V of the basis vectors. Finally we give some conclusions.

Throughout the paper we assume that the matrix $\begin{pmatrix} b & Ab & A^2b & \cdots & A^{n-1}b \end{pmatrix}$ is of full rank, that is, the grade of b with respect to A, denoted as $d(A, b)$, is equal to n. This implies that the matrix A is non-derogatory. For the sake of simplicity we assume also that $x_0 = 0$ and $\|b\| = 1$. $\|\cdot\|$ denotes the Euclidean norm. The vector e_i will denote the ith column of the identity matrix (of appropriate order). The identity matrix of order n is denoted I_n.

$X_{i:j,k:\ell}$ denotes the submatrix of X with rows from i to j and columns from k to ℓ. In this paper we assume exact arithmetic. However, the only hypothesis which is made on the matrix V whose columns are the basis vectors is that it is nonsingular. Therefore, it can even correspond to the *computed* basis vectors as long as V is of full rank, but this remains a topic for further research.

2 The general framework for Q-OR and Q-MR methods

As in [12], we consider abstract Krylov methods of two types denoted as *quasi-orthogonal residual (Q-OR)* methods and *quasi-minimal residual (Q-MR)* methods. They are all based on the construction of Krylov subspaces $\text{span}\{b, Ab, \ldots, A^{k-1}b\}, k = 1, 2, \ldots$. These subspaces are nested, i.e. $\mathcal{K}_1(A, b) \subset \mathcal{K}_2(A, b) \subset \cdots \subset \mathcal{K}_n(A, b)$. As we assume that all subspaces have full rank, $\mathcal{K}_n(A, b) = \mathbb{C}^n$ and the Krylov matrix defined as

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

is a nonsingular matrix.

Let us assume that we have an ascending basis v_k, $k = 1, \ldots, n$ of unit norm vectors for $\mathcal{K}_n(A, b)$ with $v_1 = b$. This means that $\{v_1, \ldots, v_k\}$ is a basis of $\mathcal{K}_k(A, b)$ for all $k \leq n$. The unit norm vectors are not necessarily orthonormal to each other. Let V be the matrix whose columns are the basis vectors v_k. The matrix V is nonsingular and there exists a nonsingular upper triangular matrix U (which is the matrix representing the change of basis) such that

$$K = VU. \tag{2.1}$$

Let C be the companion matrix associated with the characteristic polynomial of A denoted as

$$C = \begin{pmatrix} 0 & \cdots & 0 & -\alpha_0 \\ & I_{n-1} & & \vdots \\ & & & -\alpha_{n-1} \end{pmatrix}. \tag{2.2}$$

The monic polynomial with coefficients $\alpha_{n-1}, \ldots, \alpha_0$ has the eigenvalues of A as roots. We have the following theorem.

Theorem 2.1 *Let V and U be defined by (2.1) and C be the companion matrix of A. Then*

$$H = UCU^{-1} \tag{2.3}$$

is an unreduced upper Hessenberg matrix and

$$AV = VH. \tag{2.4}$$

Proof It is well known that $AK = KC$. This equality is straightforward for the first $n - 1$ columns. The equality for the last column is a consequence of the Cayley-Hamilton theorem that is, the matrix A satisfies its characteristic polynomial equation. Since U is nonsingular, we have from $AK = KC$ that

$$AVU = VUC, \quad \text{hence} \quad AV = V(UCU^{-1}).$$

Since U is upper triangular, the matrix UCU^{-1} is upper Hessenberg. The fact that $H = UCU^{-1}$ is unreduced will be proved using Theorem 2.2 and the fact that $u_{j,j} \neq 0$, $\forall j$. □

Note that since $v_1 = b$ and b is of unit norm, we have $u_{1,1} = 1$. For a given H and C corresponding to the characteristic polynomial of H, the matrix U in decomposition (2.3) is obtained straightforwardly by successively equating columns $1, \ldots, n-1$ in $HU = UC$. This gives (but is not recommended to compute numerically)

$$U = \begin{pmatrix} e_1 \ He_1 \ H^2e_1 \ \cdots \ H^{n-1}e_1 \end{pmatrix}.$$

Thus U is the Krylov matrix generated from H and e_1.

Let us now define the Krylov methods we are considering. We proceed as in [12]. Since without loss of generality we have chosen a zero starting vector $x_0 = 0$, we define the iterates x_k as

$$x_k = V_k y^{(k)}, \tag{2.5}$$

where V_k is the matrix of the k first columns of V. This means that we look for x_k in $\mathcal{K}_k(A, b)$. Since

$$AV_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^T = V_{k+1} \underline{H}_k, \tag{2.6}$$

where H_k is the principal submatrix of order k of H and $\underline{H}_k$ is the same matrix appended with the k first entries of the $(k+1)$st row of H, the residual vector r_k can be written as

$$r_k = b - Ax_k = V_k e_1 - AV_k y^{(k)} = V_k(e_1 - H_k y^{(k)}) - h_{k+1,k} y_k^{(k)} v_{k+1}. \tag{2.7}$$

The kth iterate $x_k^O = V_k y^{(k)}$ of a Q-OR method is defined (provided that H_k is nonsingular) by computing $y^{(k)}$ as the solution of the linear system

$$H_k y^{(k)} = e_1. \tag{2.8}$$

This annihilates the first term in the rightmost expression of (2.7). Thus the iterates of the Q-OR method are $x_k^O = V_k H_k^{-1} e_1$, the residual vector, which we denote as r_k^O, is proportional to v_{k+1} and $\|r_k^O\| = |h_{k+1,k} y_k^{(k)}|$. In case H_k is singular and x_k^O is not defined, we shall define the residual norm to be *infinite*, $\|r_k^O\| = \infty$. The vector $y^{(k)}$ is usually computed by using Givens rotations to zero the subdiagonal entries of H_k and then solving an upper triangular system.

The rightmost expression in (2.7) can also be written as $V_{k+1}(e_1 - \underline{H}_k y^{(k)})$. In a Q-MR method the vector $y^{(k)}$ is computed as the solution of the minimization problem

$$\min_y \|e_1 - \underline{H}_k y\|. \tag{2.9}$$

The vector $z_k^M = e_1 - \underline{H}_k y^{(k)}$ of length $k+1$ is referred to as the *quasi-residual*, its norm as the *quasi-residual norm*. Thus Q-MR methods minimize the quasi-residual norm. The residual vector r_k^M is $V_{k+1} z_k^M$. The solution of the minimization problem always exists but this method does not minimize the true residual norm unless V_{k+1} is orthonormal. However, this way of computing $y^{(k)}$ can also be seen as a minimization of the residual norm in a different norm; see [12]. Note that since the basis vectors are assumed to be of unit norm we have $\|r_k^M\| \leq \sqrt{k+1}\,\|z_k^M\|$. The solution of the least squares problem (2.9) is usually computed using Givens rotations to zero the subdiagonal entries of $\underline{H}_k$. Let s_j and c_j be the sines and cosines characterizing these rotations. Then, the norms of the quasi-residuals can be expressed using the sines of the rotations.

Proposition 2.1 *The norms of the Q-MR quasi-residuals are*

$$\|z_k^M\| = |s_1 s_2 \cdots s_k|. \tag{2.10}$$

We have the following relation between the residual norms of the Q-OR method and the quasi-residual norms of the corresponding Q-MR method,

$$\frac{1}{\|r_k^O\|^2} = \frac{1}{\|z_k^M\|^2} - \frac{1}{\|z_{k-1}^M\|^2}. \tag{2.11}$$

Proof This result can be obtained in the same way as Proposition 4.1 in [14] where it is proved for a particular Q-MR method in which H is block tridiagonal; see also [12]. □

Equation (2.11) is responsible for the well-known peak-plateau phenomenon; see [7]. When the Q-OR iterates are not defined because H_k is singular, the quasi-residual norms of the corresponding Q-MR method stagnate.

We have just seen that the matrices H_k and $\underline{H}_k$ are at the heart of the Q-OR/Q-MR methods since they define the problems to be solved at each iteration. It turns out that, for $k < n$, the entries of these matrices are completely determined by U in (2.3), not by the companion matrix giving the spectrum of H.

Theorem 2.2 *For $k < n$ the matrix H_k can be written as $H_k = U_k C^{(k)} U_k^{-1}$, U_k being upper triangular and the leading principal submatrix of order k of U and $C^{(k)} = E_k + \begin{pmatrix} 0 & U_k^{-1} U_{1:k,k+1} \end{pmatrix}$ where E_k is a square down-shift matrix of order k,*

$$E_k = \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}.$$

Moreover, the subdiagonal entries of H are $h_{j+1,j} = \frac{u_{j+1,j+1}}{u_{j,j}}$, $j = 1, \ldots, n-1$. The leading submatrix $\underline{H}_k$ of dimension $(k+1) \times k$ is

$$\underline{H}_k = U_{k+1} E_{k+1} \begin{pmatrix} U_k^{-1} \\ 0 \cdots 0 \end{pmatrix}.$$

Proof From $H = UCU^{-1}$ it is straightforward to see that $h_{j+1,j} = \frac{u_{j+1,j+1}}{u_{j,j}}$, $j = 1, \ldots, n-1$. Consider the matrix H_k. Clearly

$$H_k = \begin{pmatrix} I_k & 0 \end{pmatrix} H \begin{pmatrix} I_k \\ 0 \end{pmatrix}.$$

Then, using $H = UCU^{-1}$,

$$H_k = \begin{pmatrix} U_k & U_{1:k,k+1:n} \end{pmatrix} C \begin{pmatrix} U_k^{-1} \\ 0 \end{pmatrix}.$$

Let us partition the companion matrix C as

$$C = \begin{pmatrix} E_k & C_1 \\ F_k & C_2 \end{pmatrix},$$

where F_k is an $(n-k) \times k$ matrix with only one non-zero entry, $(F_k)_{1,k} = 1$. Then

$$C \begin{pmatrix} U_k^{-1} \\ 0 \end{pmatrix} = \begin{pmatrix} E_k U_k^{-1} \\ F_k U_k^{-1} \end{pmatrix}.$$

The matrix $F_k U_k^{-1}$ is zero except for the entry $(F_k U_k^{-1})_{1,k} = (U_k^{-1})_{k,k} = 1/u_{k,k}$. Therefore $H_k = U_k E_k U_k^{-1} + \begin{pmatrix} 0 & \frac{1}{u_{k,k}} U_{1:k,k+1} \end{pmatrix}$. Hence H_k is a rank-one modification of a nilpotent matrix. Let us denote $u_k = U_{1:k,k+1}$. We can write H_k as

$$H_k = U_k \left[E_k + U_k^{-1} \begin{pmatrix} 0 & u_k \end{pmatrix} \right] U_k^{-1} = U_k \left[E_k + \begin{pmatrix} 0 & U_k^{-1} u_k \end{pmatrix} \right] U_k^{-1}.$$

The other assertion is proved by writing

$$\underline{H}_k = \begin{pmatrix} I_{k+1} & 0 \end{pmatrix} H \begin{pmatrix} I_k \\ 0 \end{pmatrix}.$$

□

3 Comparison of Q-OR and Q-MR methods with FOM and GMRES

The only difference between individual Q-OR methods and between individual Q-MR methods is in the type of generated basis for the Krylov subspaces; the subspaces themselves are the same. At first sight, it may therefore seem straightforward to show that eigenvalues, being of course independent on the type of basis used, have exactly the same limited influence on convergence behavior as has been shown for FOM and GMRES (see, e.g., [19] and [18]). But in this section we point out the potentially completely different convergence behavior of Q-OR/Q-MR methods using non-orthonormal bases as compared to the FOM and GMRES methods.

Let us denote with an upper index A matrices related to the FOM/GMRES pair, which generates an orthogonal basis using the Arnoldi orthogonalization process. We have the relation

$$K = VU = V^A U^A.$$

The matrix V^A computed with the Arnoldi process is unitary, therefore

$$W \equiv (V^A)^* V = U^A U^{-1}. \tag{3.1}$$

In the following we will use the matrix W as a measure of departure from the true OR/MR methods, FOM and GMRES. Note that W is upper triangular. The further W is from the identity matrix, the less a corresponding method can be expected to behave like a true OR/MR method.

The Hessenberg matrices H in (2.4) can be transformed with W to the Hessenberg matrix H^A generated in the Arnoldi process. Using $H^A = U^A C (U^A)^{-1}$, we have

$$H = UCU^{-1} = V^{-1} V^A U^A C (U^A)^{-1} (V^A)^* V = W^{-1} H^A W.$$

Of course the two matrices H and H^A are similar since they are both similar to A. For the submatrices $\underline{H}_k$ and H_k of H, which define, with the computed bases for the Krylov subspaces, respectively the Q-MR and Q-OR iterates, we have the following results.

Theorem 3.1 *Let W_k denote the leading principal submatrix of order k of W. Then,*

$$\underline{H}_k^A = W_{k+1} \underline{H}_k W_k^{-1},$$

and

$$H_k^A = W_k H_k W_k^{-1} + [0 \ \frac{u_{k+1,k+1}}{u_{k,k}^A} w_k],$$

where w_k is the vector of the first k components of the $(k+1)$st column of W_{k+1}.

Proof Using (3.1) we obtain $W_k = U_k^A U_k^{-1}$. Considering $AV_k = V_{k+1}\underline{H}_k$ we can write

$$AV_k^A U_k^A U_k^{-1} = V_{k+1}^A U_{k+1}^A U_{k+1}^{-1} \underline{H}_k.$$

Hence $AV_k^A = V_{k+1}^A U_{k+1}^A U_{k+1}^{-1} \underline{H}_k U_k [U_k^A]^{-1}$, and

$$\underline{H}_k^A = U_{k+1}^A U_{k+1}^{-1} \underline{H}_k U_k [U_k^A]^{-1} = W_{k+1} \underline{H}_k W_k^{-1}.$$

Using a partitioning of W_{k+1} we have

$$\underline{H}_k^A = \begin{pmatrix} W_k & w_k \\ 0 & \omega_k \end{pmatrix} \begin{pmatrix} H_k \\ h_{k+1,k+1} e_k^T \end{pmatrix} W_k^{-1}.$$

Therefore,

$$H_k^A = W_k H_k W_k^{-1} + w_k h_{k+1,k} e_k^T W_k^{-1}.$$

The second term on the right-hand side is $h_{k+1,k}(W_k^{-1})_{k,k}w_k e_k^T$ because W_k is upper triangular. But with Theorem 2.2 and (3.1),

$$h_{k+1,k} = \frac{u_{k+1,k+1}}{u_{k,k}}, \quad (W_k^{-1})_{k,k} = \frac{u_{k,k}}{u^A_{k,k}},$$

and we get the result. □

Thus $(\underline{H})_k$ is transformed to $\underline{H}_k^A$ with W_k and W_{k+1} and the matrix H_k^A is a rank-one modification of H_k transformed with W_k. Clearly, the transformations and the rank-one modification can have a large impact on the iterates and on convergence behavior. In order to investigate the effect of W_k on (quasi-)residual norms, we now formulate a theorem which is the basis for many results of the paper. It states that the moduli of the entries of the first row of U^{-1} are the inverses of the Q-OR residual norms.

Theorem 3.2 *Let* $\begin{pmatrix} 1 \; g_1 \; \cdots \; g_{n-1} \end{pmatrix}$ *be the first row of* U^{-1}*. The entries* g_k *satisfy*

$$|g_k| = \frac{1}{\|r_k^O\|}, \tag{3.2}$$

where r_k^O *are the Q-OR residual vectors.*

Proof Let

$$G_j = \begin{pmatrix} c_j & -s_j \\ \overline{s_j} & c_j \end{pmatrix}, \qquad c_j^2 + |s_j|^2 = 1,$$

be the unitary complex rotation matrices used to reduce H to upper triangular form $\mathcal{R}$. We have $G_{n-1}\cdots G_1 H = \mathcal{R}$. Therefore $U^{-1}H = U^{-1}G_1^{-1}\cdots G_{n-1}^{-1}\mathcal{R}$ and

$$U^{-1}G_1^{-1}\cdots G_{n-1}^{-1} = CU^{-1}\mathcal{R}^{-1}. \tag{3.3}$$

Since $\mathcal{R}^{-1}$ is upper triangular, the matrix $CU^{-1}\mathcal{R}^{-1}$ has a zero first row except for the last entry. We will need the first row of $U^{-1}G_1^{-1}\cdots G_{n-1}^{-1}$. The product of the rotation matrices in terms of the sines and cosines was given in [3], [4]. For instance, for $n = 6$ the product $G_1^{-1}\cdots G_5^{-1}$ is

$$\begin{pmatrix} c_1 & c_2 s_1 & c_3 s_2 s_1 & c_4 s_3 s_2 s_1 & c_5 s_4 s_3 s_2 s_1 & s_5 s_4 s_3 s_2 s_1 \\ -\overline{s_1} & c_2 c_1 & c_3 s_2 c_1 & c_4 s_3 s_2 c_1 & c_5 s_4 s_3 s_2 c_1 & s_5 s_4 s_3 s_2 c_1 \\ 0 & -\overline{s_2} & c_3 c_2 & c_4 s_3 c_2 & c_5 s_4 s_3 c_2 & s_5 s_4 s_3 c_2 \\ 0 & 0 & -\overline{s_3} & c_4 c_3 & c_5 s_4 c_3 & s_5 s_4 c_3 \\ 0 & 0 & 0 & -\overline{s_4} & c_5 c_4 & s_5 c_4 \\ 0 & 0 & 0 & 0 & -\overline{s_5} & c_5 \end{pmatrix}.$$

The proof of the claim is by induction. Let us prove it for g_1. By comparing the first rows in (3.3), we have $c_1 - \overline{s_1} g_1 = 0$. Since, by our hypothesis that the Arnoldi process does not break down prematurely, $s_1 \neq 0$. This yields $g_1 =$

$c_1/\overline{s_1}$ and with $c_1 = \pm\sqrt{1-|s_1|^2}$ we get $|g_1| = \sqrt{\frac{1}{|s_1|^2} - 1}$. Since $\|z_0^M\| = 1$ and $\|z_1^M\| = |s_1|$, this is nothing other than

$$|g_1| = \left(\frac{1}{\|z_1^M\|^2} - \frac{1}{\|z_0^M\|^2} \right)^{1/2} = \frac{1}{\|r_1^O\|}.$$

Let us assume that we have $g_j = c_j/(\overline{s_1} \cdots \overline{s_j})$ for $j = 1, \ldots, k-1$. Then,

$$\begin{aligned} g_k &= \frac{c_k}{\overline{s_k}} \left[s_{k-1}s_{k-2}\cdots s_1 + s_{k-1}\cdots s_2 \frac{c_1^2}{\overline{s_1}} + \cdots + s_{k-1}\frac{c_{k-2}^2}{\overline{s_1}\cdots\overline{s_{k-2}}} + \frac{c_{k-1}^2}{\overline{s_1}\cdots\overline{s_{k-1}}} \right], \\ &= \frac{c_k}{\overline{s_k}\cdots\overline{s_1}} \left[|s_{k-1}|^2 \cdots |s_1|^2 + |s_{k-1}|^2 \cdots |s_2|^2 c_1^2 + \cdots + |s_{k-1}|^2 c_{k-2}^2 + c_{k-1}^2 \right]. \end{aligned}$$

By using $c_j^2 = 1 - |s_j|^2$, the term within brackets is equal to 1. Therefore $g_k = c_k/(\overline{s_1}\cdots\overline{s_k})$ and using (2.10) and (2.11) we obtain the result. □

Thus the inverses of the norms of the Q-OR residuals of any Q-OR method can be read from the first row of U^{-1}, where U is the upper triangular matrix in (2.1). Moreover, using (2.11), the norms of the Q-MR quasi-residuals can be read from the first row of U^{-1}, too; there holds

$$\|z_k^M\| = \left(1 + \sum_{j=1}^{k} |g_j|^2 \right)^{-1/2}. \tag{3.4}$$

The relation between the first row of U^{-1} and the first row of $(U^A)^{-1}$ is fully determined by the matrix W containing the angles between the basis vectors v_j and v_k^A, $j \le k$: Using $U^{-1} = (U^A)^{-1}W$, see (3.1), we immediately obtain $e_1^T U^{-1} = e_1^T (U^A)^{-1} W$, or, equivalently,

$$\begin{pmatrix} 1 \\ g_1 \\ \vdots \\ g_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & & & \\ w_{1,2} & w_{2,2} & & \\ \vdots & \ddots & \ddots & \\ w_{1,n} & \ldots & w_{n-1,n} & w_{n,n} \end{pmatrix} \begin{pmatrix} 1 \\ g_1^A \\ \vdots \\ g_{n-1}^A \end{pmatrix}. \tag{3.5}$$

Thus, depending on the entries of W, the residual norms of a Q-OR method can in theory be arbitrarily different from those of FOM and similarly for Q-MR methods. The next two theorems express the differences between the convergence of Q-OR/Q-MR methods and FOM/GMRES using W.

Theorem 3.3 *If $g_k^A \neq 0$, then for $k = 1, \ldots, n-1$ the relative difference between g_k^A and g_k is*

$$\frac{g_k - g_k^A}{g_k^A} = w_{k+1,k+1} - 1 + \frac{1}{g_k^A} w_{1,k+1} + \frac{g_1^A}{g_k^A} w_{2,k+1} + \cdots + \frac{g_{k-1}^A}{g_k^A} w_{k,k+1},$$

Proof From the $(k+1)$st row in (3.5) we obtain the claim using

$$g_k - g_k^A = w_{1,k+1} + g_1^A w_{2,k+1} + \cdots + g_{k-1}^A w_{k,k+1} + g_k^A [w_{k+1,k+1} - 1].$$

□

Theorem 3.4 *Let us denote the Q-MR and GMRES residual vectors by r_k^M and r_k^G. Then,*

$$\|r_k^G\| \le \|r_k^M\| \le \kappa((V^A)_{k+1}^* V_{k+1})\, \|r_k^G\|. \tag{3.6}$$

Proof We proceed as in [32] where bounds were obtained for GMRES and CMRH. Obviously we have $\|r_k^G\| \le \|r_k^M\|$. The residual vectors are in $\mathcal{K}_{k+1}(A,b)$. Then we can write

$$\begin{aligned} r_k^M &= V_{k+1} z^M, \\ r_k^G &= V_{k+1} z^G = (V^A)_{k+1} w^G. \end{aligned}$$

Since $(V^A)_{k+1}$ is unitary, the second equation yields $w^G = (V^A)_{k+1}^* V_{k+1} z^G$. Since Q-MR minimizes the norm of the quasi-residual, we have

$$\|z^M\| \le \|z^G\| \le \|[(V^A)_{k+1}^* V_{k+1}]^{-1}\|\, \|w^G\| = \|[(V^A)_{k+1}^* V_{k+1}]^{-1}\|\, \|r_k^G\|.$$

Hence

$$\begin{aligned} \|r_k^M\| \le \|V_{k+1}\|\, \|z^M\| &\le \|V_{k+1}\|\, \|[(V^A)_{k+1}^* V_{k+1}]^{-1}\|\, \|r_k^G\| \\ &= \|(V_A)_{k+1}^* V_{k+1}\|\, \|[(V_A)_{k+1}^* V_{k+1}]^{-1}\|\, \|r_k^G\|. \end{aligned}$$

□

4 The construction of linear systems with a prescribed convergence curve and spectrum

The previous section showed that the convergence behavior of a Q-OR method can be arbitrarily different from that of FOM and similarly for the convergence behavior of a Q-MR method and that of the GMRES method. The question how much eigenvalues influence the behavior of a Q-OR or Q-MR method can therefore most probably not be answered by using the answer to that question for FOM and GMRES.

At the same time a partial answer to the question was provided by Theorem 3.2. We have proved that if the basis of the Krylov subspaces (defined by the matrix V with unit norm columns) is such that $K = VU$ with U upper triangular, then $AV = VH$ with $H = UCU^{-1}$, see Theorem 2.1. Thus we have the factorization $A = VUCU^{-1}V^{-1}$ and $b = Ve_1$. According to Theorem 3.2, the inverses of the residual norms generated in the Q-OR method are the moduli of the entries of the first row of U^{-1} and the quasi-residual norms generated in the Q-MR method depend upon the first row of U^{-1} as described by equation (3.4). Therefore, if in $A = VUCU^{-1}V^{-1}$ we change the last column of C (without making the $(1,n)$ entry zero) and keep everything else the same, we obtain a new matrix $\hat{A}$ (with other nonzero eigenvalues) such that we have the same convergence curve for the Q-OR method when solving $\hat{A}x = b$ with the basis V. Using the Q-MR method we obtain the same norms for the quasi-residual vectors. In this sense, the convergence of any Q-OR and Q-MR methods does not depend on the eigenvalues of the matrix only. If a certain convergence curve is generated with a Q-OR or Q-MR method, the same curve can be obtained with any (prescribed) nonzero eigenvalues. In this section we further investigate what convergence curves are possible and attempt to construct linear systems generating admissible curves with arbitrary nonzero eigenvalues.

For GMRES any non-increasing residual norm history is possible [18], and using (2.11) this gives that for FOM any residual norms can be generated, including infinite residual norms (they correspond to stagnation in the GMRES process). To construct a matrix A with prescribed eigenvalues and a right-hand side b such that we generate a given FOM residual norm convergence curve, we first take any upper triangular matrix U^{-1} such that the moduli of the inverses of the entries of the first row are the given FOM residual norms. Then, C being the companion matrix corresponding to the given eigenvalues, we set $H = UCU^{-1}$. Finally, $A = VHV^*$, $b = Ve_1$, where V is any unitary matrix. The same can be done to create a non-increasing GMRES residual norm history; the only difference is that the first row of U^{-1} must satisfy, instead of (3.2), the conditions (3.4) with $\|z_k^M\| = \|r_k^M\|$. The factorization $A = VUCU^{-1}V^{-1}$ can, together with $b = Ve_1$, also be considered a parametrization of the entire class of matrices with right-hand sides yielding a given residual norm history with a given spectrum for the system matrix. There is freedom in the choice of the unitary matrix V and in the rows 2 till n of U^{-1}. We remark that these rows can be used to prescribe, in addition, the Ritz values generated in the underlying Arnoldi process (see [8], [9] and [10]).

A similar construction can be used for the Hessenberg/CMRH pair. These methods are based on an LU decomposition with partial pivoting of the Krylov matrix

$$P^T K = LU, \tag{4.1}$$

hence the decomposition (2.1) takes the form

$$K = VU, \qquad V = PL,$$

with V a nonsingular row permuted lower triangular matrix. Note that in (2.1) we assumed the columns of V are of unit length. This can always be achieved by appropriate scaling of the two factors in the LU decomposition of $P^T K$. To force prescribed residual norms for the corresponding Q-OR method (i.e. for the Hessenberg method), we can, as before, choose an upper triangular matrix U^{-1} such that the absolute values of the inverses of the entries of the first row are the given Hessenberg residual norms (allowing zero entries to force infinite residual norms). With C being the companion matrix corresponding to the prescribed eigenvalues, we set $H = UCU^{-1}$. A and b are then constructed from $A = VHV^{-1}$ and $b = Ve_1$, where V is any nonsingular row permuted lower triangular matrix with unit norm columns. Using the same construction but with the first row of U^{-1} satisfying the conditions (3.4), we can prescribe quasi-residual norms for the CMRH method, with any nonzero eigenvalues. We therefore have the following parametrization.

Theorem 4.1 *The characteristic polynomial of A is $\chi(\lambda) = \lambda^n + \sum_{k=0}^{n-1} \alpha_k \lambda^k$ and the Hessenberg (CMRH) method generates the residual norms $\|r_1^O\|, \dots, \|r_{n-1}^O\|$ (generates the non-increasing quasi-residual norms $\|z_1^M\|, \dots, \|z_{n-1}^M\|$) if and only if $A = VUCU^{-1}V^{-1}$, where C has the form (2.2), the first row $\begin{pmatrix} 1 & g_1 & \cdots & g_{n-1} \end{pmatrix}$ of U^{-1} satisfies (3.2) (satisfies (3.4)), V is a nonsingular row permuted lower triangular matrix with unit norm columns and $b = Ve_1$.*

With a particular choice of V we can not only prescribe CMRH quasi-residual norms, but even CMRH *residual* norms.

Corollary 4.1 *Any residual norm history (with infinite residual norms being allowed) is possible for the Hessenberg method with any nonzero eigenvalues of the system matrix. Any non-increasing residual norm history is possible for the CMRH method with any nonzero eigenvalues of the system matrix.*

Proof The first claim follows immediately from Theorem 4.1. The second claim follows when we apply the CMRH method to the matrix $H = UCU^{-1}$ with $b = e_1$, where U and C are chosen as in Theorem 4.1. Then both P and L in (4.1) are the identity matrix and so is V. Thus V is in particular unitary and the residual norms equal the prescribed quasi-residual norms. □

The situation is less straightforward if we wish to prescribe convergence curves for the BiCG/QMR pair. A first complication is that the Hessenberg matrix H generated in these methods is in fact tridiagonal, which puts additional conditions in the construction process of $H = UCU^{-1}$. Second, infinite BiCG residual norms cannot be prescribed arbitrarily. In fact, we will show that some convergence curves with infinite BiCG residual norms are inadmissable. Also, the choice of V in the parametrization $A = VHV^{-1}$ is a little more delicate.

The BiCG and QMR methods use the nonsymmetric Lanczos process to generate a pair of bi-orthogonal bases, sometimes also known as the nonhermitian Lanczos or the Bi-Lanczos process. In the following we will use the term Bi-Lanczos process. It generates a pair of bases with three-term recurrences. The first basis is an ascending basis for the Krylov subspaces $\mathcal{K}_k(A, b)$, $k \leq n$, represented by the columns v_k of a nonsingular matrix V, just as for the previous pairs of methods. The second basis is an ascending basis for the Krylov subspaces $\mathcal{K}_k(A^*, s)$, $k \leq n$, represented by the columns $\tilde{v}_k$ of a nonsingular matrix $\tilde{V}$. The shadow vector s must satisfy $s^*b \neq 0$, $d(A^*, s) = n$ and is frequently chosen as $s \equiv b$; in the following we will consider this choice. The bi-orthogonality condition $\tilde{v}_i^* v_j \neq 0$ for $i = j$ and $\tilde{v}_i^* v_j = 0$ for $i \neq j$ can be written as

$$\tilde{V}^* V = \Phi = \mathrm{diag}(\phi_1, \dots, \phi_n), \quad \phi_k \neq 0, \quad k = 1, \dots, n.$$

Now suppose we have any pair of bi-orthogonal bases stored in the nonsingular matrices $V, \tilde{V}$ such that V has unit norm columns. It can be generated by applying the Bi-Lanczos process (a variant that uses the correct column scaling of V) to *any* matrix $\hat{A}$ and starting vector $\hat{b}$ provided the bi-orthogonalization process runs to completion. Also suppose we have a *tridiagonal* matrix of the form $T = UCU^{-1}$ such that the absolute values of the inverses of the entries of the first row of the upper triangular matrix U^{-1} are prescribed Bi-CG residual norms and C is the companion matrix corresponding to prescribed nonzero eigenvalues. We will discuss the existence of such a matrix T below. Then we can define $A = VTV^{-1}$ and $b = Ve_1$. It follows that

$$K = \begin{pmatrix} b & Ab & A^2b & \cdots & A^{n-1}b \end{pmatrix} = V \begin{pmatrix} e_1 & Te_1 & T^2e_1 & \cdots & T^{n-1}e_1 \end{pmatrix} = VU.$$

Moreover, defining $\tilde{T} \equiv \Phi T^* \Phi^*$ and using $V^* = \Phi^* \tilde{V}^{-1}$, we have $A^* = \tilde{V}\tilde{T}\tilde{V}^{-1}$ and therefore

$$\begin{pmatrix} b & A^*b & (A^*)^2b & \cdots & (A^*)^{n-1}b \end{pmatrix} = \tilde{V} \begin{pmatrix} e_1 & \tilde{T}e_1 & \tilde{T}^2e_1 & \cdots & \tilde{T}^{n-1}e_1 \end{pmatrix} = \tilde{V}\tilde{U}.$$

Thus the Bi-Lanczos process applied to A and b generates the bi-orthogonal bases represented by V and $\tilde{V}$, A has the desired spectrum and U has the desired first row forcing the prescribed Bi-CG residual norms. We note that the problem of constructing matrices with a prescribed BiCG convergence curve was also considered in [6]. This was done using the relations between FOM/GMRES and BiCG/QMR. Also, given the results of BiCG on $Ax = b$, in [6] a matrix B with the same eigenvalues as A and a right-hand side c are constructed such that FOM applied to $Bx = c$ yields the same convergence curve. In [37] choices of the shadow vector s are presented which lead, for some selected iteration numbers, to the same Bi-CG residual norms as the residual norms obtained when applying a different Krylov method like, e.g., the GMRES method to the same linear system.

Along the same lines as in the above considerations one can find parametrizations for other Q-OR/Q-MR methods, like truncated and restarted methods.

4.1 An inverse constrained tridiagonal eigenproblem

We have just seen that if we wish to prescribe eigenvalues and Bi-CG residual norms simultaneously, we have to construct an upper triangular matrix U such that the first row of U^{-1} is prescribed and such that the upper Hessenberg matrix $H = UCU^{-1}$ is tridiagonal with prescribed eigenvalues given by the companion matrix C. This can be seen as an inverse nonsymmetric tridiagonal eigenproblem, with an additional constraint for the change of basis matrix U. We will first solve this problem when we prescribe finite Bi-CG residual norms only, i.e. the first row of U^{-1} has no zero entries.

We use the results of Joubert [23] which state that for a non-derogatory complex matrix, almost every starting vector r (i.e. except for a measure zero set of vectors) with shadow vector $s = r$ will generate a Bi-CG process that does not yield any breakdown, especially no so-called hard break-down. Several types of breakdown are possible in the Bi-CG method, most of them originating from a breakdown in the underlying Bi-Lanczos process. The nomenclature is not fully unified in the litterature; a hard break-down according to [23] in Bi-CG is an iteration, say number k, where the generated tridiagonal matrix of size k is singular. Hence with as input matrix the companion matrix C corresponding to the prescribed spectrum, almost every starting vector r with shadow vector $s = r$ will generate, after n iterations, a tridiagonal matrix $\hat{T}$ of size n. Clearly, $\hat{T}$ has the same prescribed spectrum as C and, in addition, no leading principal submatrix of $\hat{T}$ is singular. Then we can generate the upper triangular matrix $\hat{U} = [e_1, \hat{T}e_1, \ldots, \hat{T}^{n-1}e_1]$. It follows from the remarks in Section 2 that

$$\hat{T} = \hat{U}C\hat{U}^{-1}.$$

The matrix $\hat{U}^{-1}$ needs not have the first row equal to $g = \begin{pmatrix}1 \ g_1 \ \cdots \ g_{n-1}\end{pmatrix}$ with the g_k satisfying (3.2) for given residual norms $\|r_k^O\|$. But the first row of $\hat{U}^{-1}$ has no zero entries, otherwise some leading principal submatrix of $\hat{T}$ would be singular. We can always find a nonsingular diagonal matrix D such that the first row of $\hat{U}^{-1}D$ equals g. Then with $U = D^{-1}\hat{U}$,

$$D^{-1}\hat{T}D = D^{-1}\hat{U}C\hat{U}^{-1}D = UCU^{-1},$$

and $T \equiv D^{-1}\hat{T}D$ is the desired tridiagonal matrix. We proved the following.

Theorem 4.2 *Any residual norm history with finite residual norms is possible for the Bi-CG method with any nonzero eigenvalues of the system matrix. Any decreasing residual norm history is possible for the QMR method with any nonzero eigenvalues of the system matrix.*

Proof Both claims follow when we apply the BiCG and QMR method to the matrix $T = UCU^{-1}$ with $b = e_1$ from the preceeding discussion. Then V is the identity matrix and is in particular unitary and the QMR residual norms equal the prescribed quasi-residual norms. □

4.2 Banded Hessenberg matrices

We now show a construction that might enable to solve the constrained inverse eigenproblem for general banded Hessenberg matrices. Banded Hessenberg matrices arise, for example, in look-ahead and truncated variants of Krylov methods [22], [21], [41].

The construction consists of a backward algorithm starting with a random choice of the last column of U^{-1}. For clereness, we explain the algorithm on the example of the inverse tridiagonal eigenproblem, even if we have already solved this problem in the previous section. Then a randomly chosen last column of U^{-1} has zero probability of leading to a breakdown of the algorithm for the following reason. The result of Joubert [23] that every starting vector generates a breakdown-free Bi-CG process except for a measure zero set of vectors implies that, for a given spectrum, the set of tridiagonal matrices of size n with that spectrum and with one or more singular leading principal submatrices, is of measure zero. If we denote this set of measure zero with $\mathcal{T}_0$, then the set of upper triangular matrices of the form $[e_1, Te_1, \ldots, T^{n-1}e_1]^{-1}$, $T \in \mathcal{T}_0$, is of measure zero, too. In particular, the set of last columns of triangular matrices of the above form is of measure zero.

Thus, we can simply arbitrarily prescribe the last column of U^{-1} (except the last entry which must be nonzero) in addition to its first row to compute H. Because we do not consider infinite residual norms, we assume that $g_j \neq 0$, $j = 1, \ldots, n-1$. Let us denote with $\nu_{i,j}$ the entries of U^{-1}, let $\nu_{n,1} = g_{n-1}$ and let H be denoted as

$$H = \begin{pmatrix} \gamma_1 & \beta_2 & 0 & 0 & 0 \\ \rho_2 & \gamma_2 & \beta_3 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \rho_{n-1} & \gamma_{n-1} & \beta_n \\ 0 & 0 & 0 & \rho_n & \gamma_n \end{pmatrix}.$$

The last column of the matrix equation $U^{-1}H = CU^{-1}$ yields

$$\begin{pmatrix} g_{n-2}\beta_n + g_{n-1}\gamma_n \\ \vdots \\ \nu_{n,n}\gamma_n \end{pmatrix} = \begin{pmatrix} -\alpha_0 \nu_{n,n} \\ \vdots \\ \nu_{n,n-1} - \alpha_{n-1}\nu_{n,n} \end{pmatrix}.$$

We use the first and last equations

$$\begin{pmatrix} g_{n-2} & g_{n-1} \\ 0 & \nu_{n,n} \end{pmatrix} \begin{pmatrix} \beta_n \\ \gamma_n \end{pmatrix} = \begin{pmatrix} -\alpha_0 \nu_{n,n} \\ \nu_{n,n-1} - \alpha_{n-1}\nu_{n,n} \end{pmatrix}.$$

The solution of this 2×2 nonsingular linear system yields γ_n, β_n. From the other equations that we discarded we can then compute the unknown entries of column $n-1$ of U^{-1} using

$$\nu_{i,n-1}\beta_n + \nu_{n,i}\gamma_n = \nu_{n,i-1} - \alpha_{i-1}\nu_{n,n}, \; i = 2, \ldots, n-1. \tag{4.2}$$

This yields $\nu_{i,n-1}$, $i = 2, \ldots, n-1$. Thus we know all the entries $\nu_{i,n-1}$ since $\nu_{1,n-1} = g_{n-2}$ and $\nu_{n,n-1} = 0$. Now we consider the one but last column in $U^{-1}H = CU^{-1}$. We have three unknowns $\beta_{n-1}, \gamma_{n-1}$ and ρ_n. We first take the first and the last two equations. This gives us a linear system with an upper triangular system,

$$\begin{pmatrix} g_{n-3} & g_{n-2} & g_{n-1} \\ 0 & \nu_{n-1,n-1} & \nu_{n,n-1} \\ 0 & 0 & \nu_{n,n} \end{pmatrix} \begin{pmatrix} \beta_{n-1} \\ \gamma_{n-1} \\ \rho_n \end{pmatrix} = \begin{pmatrix} 0 \\ \nu_{n-2,n-1} \\ \nu_{n-1,n-1} \end{pmatrix}.$$

Note that the system is singular for a measure zero set of choices of $\nu_{n,i}$ only. After computing $\beta_{n-1}, \gamma_{n-1}$ and ρ_n, the unknown components of column $n-2$ of U^{-1} are computed using the discarded equations with conditions analogue to (4.2). Then, we proceed with column $n-2$ of $U^{-1}H = CU^{-1}$ in the same manner until the third column. The second column gives $\rho_3, \gamma_2, \beta_2$ by solving

$$\begin{pmatrix} 1 & g_1 & g_2 \\ 0 & \nu_{2,2} & \nu_{2,3} \\ 0 & 0 & \nu_{3,3} \end{pmatrix} \begin{pmatrix} \beta_2 \\ \gamma_2 \\ \rho_3 \end{pmatrix} = \begin{pmatrix} 0 \\ g_1 \\ \nu_{2,2} \end{pmatrix}. \tag{4.3}$$

The first column yields ρ_2, γ_1 using

$$\begin{pmatrix} 1 & g_1 \\ 0 & \nu_{2,2} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{4.4}$$

Note that if we have a numerical problem with a too small diagonal entry, we can restart the process with a different last column.

We can extend this algorithm to any larger bandwidth for the upper triangular part of H, albeit without a proof of measure zero probability of a breakdown. We just have to prescribe more columns of U^{-1}. If we would like to have a bandwidth m in the upper triangular part of H, then we have to prescribe the last m columns of U^{-1} to start the algorithm. Let us assume, for instance, that we would like an upper bandwidth of 2. We have another upper diagonal with entries $\delta_3, \dots, \delta_n$ in H. Then, we prescribe arbitrarily the last two columns of U^{-1} (except the first entries which are already prescribed). Equating the last column of $U^{-1}H = CU^{-1}$ we have

$$\begin{pmatrix} \delta_n g_{n-3} + \beta_n g_{n-2} + \gamma_n g_{n-1} \\ \delta_n \nu_{2,n-2} + \beta_n \nu_{2,n-1} + \gamma_n \nu_{2,n} \\ \vdots \\ \delta_n \nu_{n-2,n-2} + \beta_n \nu_{n-2,n-1} + \gamma_n \nu_{n-2,n} \\ \beta_n \nu_{n-1,n-1} + \gamma_n \nu_{n-1,n} \\ \gamma_n \nu_{n,n} \end{pmatrix} = \begin{pmatrix} -\alpha_0 \nu_{n,n} \\ g_{n-1} - \alpha_1 \nu_{n,n} \\ \vdots \\ \nu_{n-3,n} - \alpha_{n-3} \nu_{n,n} \\ \nu_{n-2,n} - \alpha_{n-2} \nu_{n,n} \\ \nu_{n-1,n} - \alpha_{n-1} \nu_{n,n} \end{pmatrix}.$$

Since the entries of the last two columns $\nu_{j,n}$ and $\nu_{j,n-1}$ are known we use the first and the last two equations to compute $\gamma_n, \beta_n, \delta_n$. It yields a linear system

$$\begin{pmatrix} g_{n-3} & g_{n-2} & g_{n-1} \\ 0 & \nu_{n-1,n-1} & \nu_{n-1,n} \\ 0 & 0 & \nu_{n,n} \end{pmatrix} \begin{pmatrix} \delta_n \\ \beta_n \\ \gamma_n \end{pmatrix} = \begin{pmatrix} -\alpha_0 \nu_{n,n} \\ \nu_{n-2,n} - \alpha_{n-2} \nu_{n,n} \\ \nu_{n-1,n} - \alpha_{n-1} \nu_{n,n} \end{pmatrix}.$$

Then we use the other equations to compute the entries $\nu_{2,n-2}, \dots, \nu_{n-2,n-2}$. For the next columns we equate column $n-1$ of $U^{-1}H = CU^{-1}$,

$$\begin{pmatrix} \delta_{n-1} g_{n-4} + \beta_{n-1} g_{n-3} + \gamma_{n-1} g_{n-2} + \rho_n g_{n-1} \\ \delta_{n-1} \nu_{2,n-3} + \beta_{n-1} \nu_{2,n-2} + \gamma_{n-1} \nu_{2,n-1} + \rho_n \nu_{2,n} \\ \vdots \\ \delta_{n-1} \nu_{n-3,n-3} + \beta_{n-1} \nu_{n-3,n-2} + \gamma_{n-1} \nu_{n-3,n-1} + \rho_n \nu_{n-3,n} \\ \beta_{n-1} \nu_{n-2,n-2} + \gamma_{n-1} \nu_{n-2,n-1} + \rho_n \nu_{n-2,n} \\ \gamma_{n-1} \nu_{n-1,n-1} + \rho_n \nu_{n-1,n} \\ \rho_n \nu_{n,n} \end{pmatrix} = \begin{pmatrix} 0 \\ g_{n-2} \\ \vdots \\ \nu_{n-4,n-1} \\ \nu_{n-3,n-1} \\ \nu_{n-2,n-1} \\ \nu_{n-1,n-1} \end{pmatrix}.$$

We use the first and the last three equations to compute $\rho_n, \gamma_{n-1}, \beta_{n-1}, \delta_{n-1}$, leading to an upper triangular linear system of size four. If there is no division by zero we can go on solving upper triangular systems of order 4 until column 3 which yields $\rho_4, \gamma_3, \beta_3, \delta_3$ and $\nu_{2,2}$ and compute the last two columns of H as in (4.3) and (4.4).

4.3 Infinite residual norms

The previous sections showed that we can generate any Bi-CG convergence curve with any spectrum, provided we do not prescribe infinite residual norms. We next show that infinite residual norms cannot be forced at arbitrary predetermined iteration numbers. This result is a nonsymmetric analogue of a well-known fact for the symmetric Lanczos process applied to indefinite systems, see, e.g., [5, Section 2.1]. We formulate it for the situation where the underlying Bi-Lanczos process runs till completion, but it holds also when the Bi-Lanczos process terminates prematurely, having found an A-invariant or A^*-invariant Krylov subspace.

Theorem 4.3 *Let the Bi-CG method be applied to a linear system $Ax = b$ with a nonsingular matrix A such that $d(A, b) = n$ and $d(A^*, s) = n$. The Bi-CG residual norms of two consecutive iteration numbers cannot be both infinite.*

Proof Let C be the companion matrix corresponding to the characteristic polynomial of A, let T be the tridiagonal matrix produced by the Bi-Lanczos process and let $U = [e_1, Te_1, \dots, T^{n-1}e_1]$, then $U^{-1}T = CU^{-1}$. The first row of U^{-1} denoted by $g = [1, \dots, g_{n-1}]$ contains the Bi-CG residual norms through $|g_k|^{-1} = \|r_k^{Bi-CG}\|$. Equating the kth column and the first row in $U^{-1}T = CU^{-1}$ for some k, $1 < k < n$, we obtain

$$g_{k-2}\beta_k + g_{k-1}\gamma_k + g_k\rho_{k+1} = 0. \tag{4.5}$$

Per assumption, all entries on the lower and upper subdiagonals of T are nonzero, otherwise the Bi-Lanczos procedure would have terminated early. Therefore equation (4.5) cannot be satisfied if $g_{k-2} = g_{k-1} = 0$ in combination with $g_k \neq 0$ or if $g_{k-2} \neq 0$ in combination with $g_{k-1} = g_k = 0$. By consecutively using the same rule for neighboring columns we conclude that either (1) all g_k, $1 \leq k \leq n-1$, are nonzero, (2) all g_k, $1 \leq k \leq n-1$, are zero, or (3) no two consecutive g_{k-1}, g_k are zero. The claim follows if we can exclude the case (2) where $g_k = 0$ for all $1 \leq k \leq n-1$. This would lead, with the equation $\gamma_1 + g_1\rho_2 = 0$ for the $(1,1)$-entry of $U^{-1}T = CU^{-1}$, to $\gamma_1 = 0$. Then

$$\det \begin{pmatrix} 0 & \beta_2 \\ \rho_2 & \gamma_2 \end{pmatrix} \neq 0,$$

contradicting $g_1 = 0$. □

Thus if we allow infinite residual norms, then not every convergence curve is admissable for Bi-CG. This is in contrast with the situation for FOM and the Hessenberg method. Similarly, in contrast with GMRES and CMRH, two consecutive steps of stagnation of quasi-residual norms are impossible for QMR.

The question whether we can prescribe with a given spectrum of the system matrix, infinite Bi-CG residual norms at given, non-consecutive iteration numbers amounts to solving an inverse tridiagonal eigenproblem where principal leading submatrices of predetermined sizes must be singular. We did not find any text addressing this problem in the literature. Though we were able to derive partial solutions not reported here, we consider prescribing particular g_k's to be zero as an open problem.

5 Influence of eigenvectors, right-hand side and Krylov basis

The results in the previous section show that there are sets of matrices with different, even arbitrary eigenvalue distributions and right-hand sides giving the same residual norms for Q-OR methods or the same quasi-residual norms for Q-MR methods. Of course, this does not mean that the behaviour of these methods is not influenced by the eigenvalues at all. It means that if we modify eigenvalues, then to have the same (quasi-)residual norms, other objects related to the linear system (eigenvectors, the right-hand side) must be modified and indeed can be modified in an appropriate way. Formulaes which reveal the complicated but precise interplay between eigenvalues, eigenvectors and right-hand side when forming GMRES residual norms were given in [27] (and in [11] for normal matrices). They are closed form expressions (unlike the norms of the GMRES residual vectors in actual GMRES computations) and as such have no immediate practical application but they give insight into when residual norms can be expected to be governed by eigenvalues and when not.

In this section we present similar closed form expressions for general Q-OR and Q-MR methods. For simplicity, we will not consider the case of defective matrices (expressions for GMRES residual norms in this case can be found in [27, Sections 3 and 4]). The main ingredient to write the (quasi)-residual norms as a function of the eigenvalues and eigenvectors of A, is the following generalization of a well-known result for GMRES; see [11] and the references therein.

Theorem 5.1 *Let us define the moment matrix $M = U^*U$ and let M_{k+1} be the principal submatrix of M given by $M_{k+1} = U_{k+1}^*U_{k+1}$. Then the norms of the Q-OR residual and of the Q-MR quasi-residual vectors are given by*

$$\|r_k^O\|^2 = \frac{1}{(M_{k+1}^{-1})_{1,1} - (M_k^{-1})_{1,1}}, \quad \|z_k^M\|^2 = \frac{1}{(M_{k+1}^{-1})_{1,1}}. \tag{5.1}$$

Proof Since $M_{k+1}^{-1} = U_{k+1}^{-1}U_{k+1}^{-*}$, the $(1,1)$ entry of this matrix is

$$(M_{k+1}^{-1})_{1,1} = 1 + |g_1|^2 + \cdots + |g_k|^2.$$

Using (3.2) and (3.4) this gives the claim. □

We now use that U in the definition of the moment matrix $M = U^*U$ satisfies $U = V^{-1}K$ and plug in the spectral decomposition of A for the columns $A^k b$ of the Krylov matrix K. As before, it is assumed that K is nonsingular, but the following result is essentially the same if A does not have distinct eigenvalues or when components of the right-hand side in the eigenvector basis are zero.

Theorem 5.2 *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 $Z = V^{-1}X$ with V defined in (2.1). Then for $k < n$,*

$$(M_{k+1}^{-1})_{1,1} = \mu_k^N / \mu_{k+1}^D,$$

where $\mu_1^N = \sum_{i=1}^n \left| \sum_{j=1}^n Z_{i,j}\, c_j\, \lambda_j \right|^2$, for $k \geq 2$

$$\mu_k^N = \sum_{I_k} \left| \sum_{J_k} \det(Z_{I_k,J_k})\, c_{j_1} \cdots c_{j_k}\, \lambda_{j_1} \cdots \lambda_{j_k} \prod_{j_\ell < j_p \in J_k} (\lambda_{j_p} - \lambda_{j_\ell}) \right|^2, \quad \text{and}$$

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

The summations are over all sets of indices $I_{k+1}, J_{k+1}, I_k, J_k$ defined as I_ℓ (or J_ℓ) to be a set of ℓ indices $\big(i_1, i_2, \ldots, i_\ell\big)$ such that $1 \leq i_1 < \cdots < i_\ell \leq n$ and Z_{I_ℓ,J_ℓ} is the submatrix of Z whose row and column indices of entries are defined respectively by I_ℓ and J_ℓ.

Proof We would like to use (5.1). We have $M = U^*U = K^*V^{-*}V^{-1}K$. The matrix A has been assumed to be diagonalizable with $A = X\Lambda X^{-1}$. Then $K = X\left(c\ \Lambda c \cdots \Lambda^{n-1}c\right)$ with $c = X^{-1}b$ and

$$M = \left(c\ \Lambda c \cdots\ \Lambda^{n-1}c\right)^* X^*V^{-*}V^{-1}X\left(c\ \Lambda c \cdots\ \Lambda^{n-1}c\right).$$

Let D_c be the diagonal matrix with diagonal entries c_i and let $\mathcal{V}_{k+1}$ denote the $k+1$ first columns of the square Vandermonde matrix for $\lambda_1, \ldots, \lambda_n$ with entries $(\mathcal{V}_{k+1})_{i,j} = \lambda_i^{j-1}$. Then the matrix M_{k+1} can be written as

$$M_{k+1} = \mathcal{V}_{k+1}^* D_c^* X^* V^{-*} V^{-1} X D_c \mathcal{V}_{k+1}.$$

With $F = V^{-1}XD_c\mathcal{V}_{k+1}$, it is the product F^*F of two $n \times (k+1)$ matrices. We can use Cramer's rule and twice the Cauchy-Binet formula to compute the determinants exactly as in [11], [27]. Therefore we omit the details. □

The theorem, in combination with Theorem 5.1, shows that the (quasi)-residual norms in Q-OR/Q-MR methods depend upon the following three types of objects: Eigenvalues, components of the right-hand side in the eigenvector basis and determinants of Z, i.e. of the eigenvector basis multiplied with the inverse of the generated basis for the Krylov subspaces. As for true OR/MR methods, the contribution of the right-hand side is restricted to its components in the eigenvector basis. A difference with true OR/MR methods is that the involved Krylov subspace basis, which is not *a priori* given, plays a role. Even if A is normal, i.e. $X^*X = I$, this needs not imply that the determinants in μ_k^N, μ_{k+1}^D can be simplified, unless V is unitary. Bounds on the (quasi)-residual norms involving $\kappa(Z)$ or $\kappa(Z\text{diag}(c_1, \ldots, c_n))$ might be derived along the same lines as was done in [27]. For low condition numbers, they would imply a dominating influence of eigenvalues on convergence behavior. Nevertheless, Theorem 5.2 suggests that, in general, eigenvectors, the right-hand side and the generated Krylov subspace basis can play a similarly important role.

6 Conclusion

It is a well-known result that any non-increasing residual norm history is possible for GMRES with any nonzero spectrum. We showed that it can be generalized to any Krylov method classifiable as a Q-OR or Q-MR method with a few slight modifications: For Q-OR methods the residual norm history needs not be non-increasing and when using long recurrences any residual norms including infinite norms (i.e. non-definable iterates) are possible. With three-term recurrences any finite residual norm history is possible but two consecutive infinite norms are not allowed. An open problem is whether any residual norm history without consecutive infinite norms is possible. The proofs are constructive and give parametrizations of the classes of matrices with a given nonzero spectrum and right-hand sides such that prescribed residual norms (or quasi-residual norms for Q-MR methods) are generated. We also derived expressions of the Q-OR residual and Q-MR quasi-residual norms as functions of the eigenvalues and eigenvectors. They are the same as those for FOM and GMRES except for that the matrix of the Krylov subspace basis vectors comes in.

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. W. E. Arnoldi. The principle of minimized iteration in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.*, 9:17–29, 1951.
3. P. N. Brown. A theoretical comparison of the Arnoldi and GMRES algorithms. *SIAM J. Sci. Statist. Comput.*, 12(1):58–78, 1991.
4. P. N. Brown and A. C. Hindmarsh. Reduced storage matrix methods in stiff ODE systems. *Appl. Math. Comput.*, 31:40–91, 1989.

5. S-C. T. Choi, C. C. Paige, and M. A. Saunders. MINRES-QLP: a Krylov subspace method for indefinite or singular symmetric systems. *SIAM J. Sci. Comput.*, 33(4):1810–1836, 2011.
6. J. Cullum. Iterative methods for solving $Ax = b$, GMRES/FOM versus QMR/BiCG. *Advances in Computational Mathematics*, 6:1–24, 1996.
7. J. Cullum and A. Greenbaum. Relations between Galerkin and norm-minimizing iterative methods for solving linear systems. *SIAM J. Matrix Anal. Appl.*, 17(2):223–247, 1996.
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 and G. Meurant. On the admissible convergence curves for restarted GMRES. submitted, 2015.
11. J. Duintjer Tebbens, G. Meurant, H. Sadok, and Z. Strakoš. On investigating GMRES convergence using unitary matrices. *Lin. Alg. Appl.*, 450:83–107, 2014.
12. M. Eiermann and O. G. Ernst. Geometric aspects of the theory of Krylov subspace methods. *Acta Numer.*, 10:251–312, 2001.
13. R. Fletcher. Conjugate gradient methods for indefinite systems. In *Numerical analysis (Proc 6th Biennial Dundee Conf., Univ. Dundee, Dundee, 1975)*, pages 73–89. Lecture Notes in Math., Vol. 506. Springer, Berlin, 1976.
14. R. W. Freund and N. M. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Numer. Math.*, 60(3):315–339, 1991.
15. R. W. Freund and N. M. Nachtigal. An implementation of the QMR method based on coupled two-term recurrences. *SIAM J. Sci. Comput.*, 15(2):313–337, 1994.
16. R. W. Freund and N. M. Nachtigal. Software for simplified Lanczos and QMR algorithms. *Appl. Numer. Math.*, 19(3):319–341, 1995.
17. A. Greenbaum. *Iterative methods for solving linear systems*, volume 17 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
18. 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.
19. 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.
20. M. Heyouni and H. Sadok. On a variable smoothing procedure for Krylov subspace methods. *Linear Algebra Appl.*, 268:131–149, 1998.
21. K. C. Jea and D. M. Young. On the simplification of generalized conjugate-gradient methods for nonsymmetrizable linear systems. *Linear Algebra Appl.*, 52/53:399–417, 1983.
22. Z. Jia. On IGMRES: an incomplete generalized minimal residual method for large unsymmetric linear systems. *Sci. China Ser. A*, 41(12):1278–1288, 1998.
23. W. Joubert. Lanczos methods for the solution of nonsymmetric systems of linear equations. *SIAM J. Matrix Anal. Appl.*, 13(3):926–943, 1992.
24. C. Lanczos. Solution of systems of linear equations by minimized iterations. *J. Research Nat. Bur. Standards*, 49:33–53, 1952.
25. J. Liesen and Z. Strakoš. *Krylov Subspace Methods, Principles and Analysis.* Oxford University Press, 2012.
26. G. Meurant. GMRES and the Arioli, Pták, and Strakoš parametrization. *BIT*, 52(3):687–702, 2012.
27. G. Meurant and J. Duintjer Tebbens. The role eigenvalues play in forming GMRES residual norms with non-normal matrices. *Numerical Algorithms*, 68:143–165, 2015.
28. Y. Saad. Krylov subspace methods for solving large unsymmetric linear systems. *Math. Comp.*, 37(155):105–126, 1981.
29. Y. Saad. The Lanczos biorthogonalization algorithm and other oblique projection methods for solving large unsymmetric systems. *SIAM J. Numer. Anal.*, 19(3):485–506, 1982.
30. 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.
31. H. Sadok. CMRH: a new method for solving nonsymmetric linear systems based on the Hessenberg reduction algorithm. *Numer. Algorithms*, 20(4):303–321, 1999.

32. H. Sadok and D. B. Szyld. A new look at CMRH and its relation to GMRES. *BIT*, 52(2):485–501, 2012.
33. M. Schweitzer. Any cycle-convergence curve is possible for restarted FOM. *preprint BUW-IMACM 14/19*, Bergische Universität Wuppertal, 2014.
34. V. Simoncini. On the convergence of restarted Krylov subspace methods. *SIAM J. Matrix Anal. Appl.*, 22(2):430–452, 2000.
35. V. Simoncini and D. B. Szyld. The effect of non-optimal bases on the convergence of Krylov subspace methods. *Numer. Math.*, 100(4):711–733, 2005.
36. P. Sonneveld. CGS, a fast Lanczos-type solver for nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 10(1):36–52, 1989.
37. P. Tichý. *O některých otevřených problémech v krylovovských metodách*. PhD thesis (in Czech). Charles University, Prague, 2002.
38. H. A. van der Vorst. Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 13(2):631–644, 1992.
39. H. A. van der Vorst. *Iterative Krylov methods for large linear systems*, volume 13 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2003.
40. E. Vecharynski and J. Langou. Any admissible cycle-convergence behavior is possible for restarted GMRES at its initial cycles. *Num. Lin. Algebr. Appl.*, 18:499–511, 2011.
41. P.K.W. Vinsome. Orthomin, an iterative method for solving sparse sets of simultaneous linear equations. In *Proceedings of the fourth symposium on reservoir simulation*, pages 149–159. SPE, 1976.