BIT Numer Math (2012) 52:687–702
DOI 10.1007/s10543-012-0374-4

BIT

GMRES and the Arioli, Pták, and Strakoš parametrization

Gérard Meurant

Received: 11 June 2011 / Accepted: 14 January 2012 / Published online: 28 January 2012

Abstract In this paper we study the parametrization proposed by Arioli, Pták and Strakoš (BIT Numerical Mathematics, v 38, 1998) for the class of matrices having the same GMRES residual norm convergence curve. We give expressions for the Hessenberg matrix and the orthonormal basis vectors constructed by GMRES as well as for iterates and error vectors. The iterates do not depend on the eigenvalues in the sense that changing the coefficients of the characteristic polynomial in the parametrization does not change the GMRES iterates as well as the residuals. However, the error vectors do depend on these coefficients.

Keywords GMRES · Approximation · Residual norm · Convergence curve · Error

Mathematics Subject Classification (2000) 15A06 · 65F10

1 Introduction

We consider solving a linear system

$$Ax = b \tag{1.1}$$

where A is a real nonsingular matrix of order n with the Generalized Minimum RESidual method (GMRES) which is an iterative Krylov method based on the Arnoldi orthogonalization process; see Saad [8, 9] and Saad and Schultz [10]. The initial residual is denoted as $r_0 = b - Ax_0$ where x_0 is the starting vector. The Krylov subspace of order k based on A and r_0, denoted as $\mathcal{K}_k(A, r_0)$, is span$\{r_0, Ar_0, \dots, A^{k-1}r_0\}$. The approximate solution x_k at iteration k is sought as

Communicated by Axel Ruhe.

G. Meurant (✉)
30, rue du Sergent Bauchat, 75012 Paris, France
e-mail: gerard.meurant@gmail.com

$x_k \in x_0 + \mathscr{K}_k(A, r_0)$ such that the residual vector $r_k = b - Ax_k$ has a minimal Euclidian norm; this amounts to satisfy an orthogonality condition with the Krylov subspace.

It was and it is still believed by many people that the convergence of the GMRES residual norm is governed by the eigenvalue distribution. There is some direct influence of the eigenvalues on convergence when the matrix A is normal. But, in [5] Greenbaum and Strakoš proved that any convergence curve for the residual norm that can be generated with GMRES can be obtained with a non-derogatory matrix having prescribed eigenvalues. Therefore, there are many cases for which residual norm convergence is not directly related to the eigenvalue distribution and their relationship is unclear. Greenbaum, Pták and Strakoš [6] showed later that any nonincreasing sequence of residual norms can be given by GMRES. Finally, Arioli, Pták and Strakoš [1] gave a complete parametrization of all pairs $\{A, b\}$ generating a prescribed residual norm convergence curve. These results are interpreted by some researchers working on Krylov methods by saying that "GMRES convergence does not depend on the eigenvalues of A". Many papers were published looking for some other quantities than eigenvalues explaining GMRES convergence: field of values, pseudo-eigenvalues, polynomial numerical hull, etc. In fact, the authors of the papers [1, 5, 6] never wrote that convergence does not depend on the eigenvalues. Their results show that there are sets of matrices with different eigenvalue distributions and right-hand sides giving the same GMRES residual norms.

The aim of this paper is to give more details on matrices or quantities that are involved in GMRES. Building on the parametrization in [1] we will establish some relations between the matrices that are involved in the parametrization and we will characterize the Hessenberg matrix H that is generated by GMRES as well as the matrix V whose columns are the vectors of the orthonormal basis of the Krylov subspace. We will also provide formulas for the GMRES iterates x_k and the error vector $\varepsilon_k = x - x_k$. We will see that the matrix V, the first $n-1$ columns of the matrix H and the iterates do not depend on the eigenvalues, in the sense that in the parametrization we can change the companion matrix C corresponding to the eigenvalues of A (and thus the matrix A) without changing them. Hence, we have the same GMRES iterates for non-derogatory matrices with different eigenvalue distributions and the same residual norm convergence curve. However, our results will show that the error vectors do depend on the matrix C and therefore on the eigenvalues of A through the exact solution of the linear system (1.1).

In this sense, one would be tempted to say that "GMRES convergence does depend on the eigenvalue distribution". However, the negative and the positive statements are not completely true (or false). In addition to the results of [1] everything depends on the measures we are using for understanding "convergence": the residual norm or the error norm.

Throughout the paper we will assume that the matrix A is non-derogatory and that the right-hand side is such that GMRES terminates at iteration n. This implies that all the Krylov subspaces are of maximal rank. For a square matrix B we denote as B_k the principal submatrix of order k (except for the matrices V and W for which it denotes the matrices build with the first k columns) and e_i denotes the ith column of the identity matrix of appropriate dimension. The norm $\|\cdot\|$ is the Euclidean norm.

Since this paper can be seen as a complement of [1] we will use the same notation as in the paper by Arioli, Pták and Strakoš.

The outline of the paper is as follows. Section 2 recalls the result from Arioli, Pták and Strakoš [1]. Section 3 describes some properties of matrices involved in GMRES, for instance the Hessenberg matrix produced by the Arnoldi process and the matrix V of the orthonormal basis vectors. Even though some of these relations are not needed to compute the iterates or the errors, they are of interest per se. Section 4 computes the GMRES iterates x_k as functions of the matrices involved in the parametrization. Section 5 provides expressions for the error at a given iteration of the GMRES algorithm. In Sect. 6 we revisit the GMRES residual norms. Finally we give some conclusions and perspectives.

2 The Arioli, Pták and Strakoš parametrization

We recall the following results that were proved in [1] (Theorem 2.1 and Corollary 2.4) using the same notation.

Theorem 2.1 *Assume we are given $n+1$ positive numbers*

$$f_0 \geq f_1 \geq \cdots \geq f_{n-1} > 0, \qquad f_n = 0$$

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

1. *The spectrum of A is $\{\lambda_1, \ldots, \lambda_n\}$ and GMRES applied to A and b yields residuals r_j, $j = 0, \ldots, n-1$ such that*

$$\|r_j\| = f_j, \quad j = 0, \ldots, n-1.$$

2. *The matrix A is of the form $A = WYCY^{-1}W^*$ and $b = Wh$, where W is a unitary matrix, Y is given by*

$$Y = \begin{pmatrix} h & \begin{matrix} R \\ 0 \end{matrix} \end{pmatrix}, \tag{2.1}$$

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

$$h = (\eta_1, \ldots, \eta_n)^T, \quad \eta_j = (f_{j-1}^2 - f_j^2)^{1/2}$$

and C is the companion matrix corresponding to the polynomial q,

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

We will call the parametrization, $(A, b) = (WYCY^{-1}W^*, Wh)$, the APS parametrization. In the parametrization of Theorem 2.1, the prescribed residual norm convergence curve is implicitly contained in the vector h which is prescribed by the given residual norms defined by f_j, $j = 0, \ldots, n$. The degrees of freedom defining the class of matrices are the upper triangular matrix R, the unitary matrix W and the companion matrix C. Thus we can change R to obtain another matrix in the class for which we have the same residual norm convergence curve as long as the right-hand side is $b = Wh$. Changing W also changes the right-hand side b but not the residual convergence curve. Keeping everything else (that is, W, R, h) the same, we can change C and hence the eigenvalues without changing the norms of the residuals. So we have a class of matrices all having different eigenvalue distributions and the same residual norms.

In this paper we are interested in real matrices and right-hand sides. In this case all the quantities defined in Theorem 2.1 are real except, of course, the eigenvalues. Without loss of generality, we will choose $x_0 = 0$ and $\|b\| = 1$; this yields $r_0 = b$ and $\|h\| = 1$.

3 Properties of the APS parametrization

The Arnoldi process used in GMRES gradually computes an upper Hessenberg matrix H with positive entries on the first subdiagonal using the modified Gram–Schmidt algorithm for computing the orthonormal matrix V of the orthonormal basis vectors for the Krylov subspace $\mathcal{K}_n(A,b)$; see [10]. At iteration n we have $AV = VH$ and, of course, the factorization $A = VHV^T$. In this section we prove some general properties for the matrices that are involved in the GMRES algorithm using the APS parametrization. In particular, we characterize the Hessenberg matrix H and the orthonormal matrix V.

Let K be the Krylov matrix generated from A and b,

$$K = (b \quad Ab \quad A^2b \quad \cdots \quad A^{n-1}b),$$

whose columns are the natural basis vectors of the Krylov space $\mathcal{K}_n \equiv \mathcal{K}_n(A,b)$. The first interesting property is that the Krylov matrix can be factorized as $K = WY$ where W and Y are defined in Theorem 2.1; this result was proven in [1].

There are several QR factorizations that are directly linked to GMRES and the APS parametrization. We first consider a factorization of K.

Proposition 3.1 *The Krylov matrix K can be factorized as*

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

where V is the orthonormal matrix of the orthonormal basis vectors for the Krylov subspace $\mathcal{K}_n(A,b)$ and the matrix U is upper triangular with positive diagonal elements. Moreover,

$$U = (\, e_1 \quad He_1 \quad \cdots \quad H^{n-1}e_1\,). \tag{3.2}$$

Proof Let us prove by induction that $A^jV = VH^j$, $j = 1, \ldots, n-1$. This is true for $j = 1$ since $AV = VH$. If we assume $A^{j-1}V = VH^{j-1}$, we have

$$A^jV = A(A^{j-1}V) = AVH^{j-1} = VH^j.$$

Therefore, since $b = Ve_1$,

$$K = (b \quad Ab \quad \cdots A^{n-1}b\,) = V\,(\,e_1 \quad He_1 \quad \cdots \quad H^{n-1}e_1\,).$$

The matrix H being upper Hessenberg, one can prove easily that the matrix $U = (e_1\ He_1\ \cdots\ H^{n-1}e_1)$ is upper triangular since multiplying H^j by H from the left gives one more subdiagonal than in H^j. Moreover, since H has a positive first subdiagonal, the diagonal entries of U are positive. □

Then, we introduce a QR factorization of the upper Hessenberg matrix H as

$$H = Q\mathscr{R}, \tag{3.3}$$

where Q is upper Hessenberg orthogonal and $\mathscr{R}$ is upper triangular, the signs being chosen such that the entries of the first row of Q are positive. We will see in Theorem 3.1 that the first row of Q is directly related to GMRES convergence.

The orthonormal matrix W in the APS parametrization defines a basis of the space $A\mathscr{K}_n$ and we have

$$AK = W\tilde{\mathscr{R}}, \tag{3.4}$$

where $\tilde{\mathscr{R}}$ is upper triangular. Starting the Arnoldi process with Ab instead of b yields the orthonormal matrix W and an upper Hessenberg matrix $\mathscr{H}$ such that

$$AW = W\mathscr{H}. \tag{3.5}$$

Using these definitions and notation, we have the following results that characterize some of the matrices previously defined.

Theorem 3.1 *The lower triangular matrix U^T defined by (3.1) and (3.2) is the Cholesky factor of the matrix Y^TY which is equal to K^TK, where Y is the matrix involved in the APS factorization and K is the Krylov matrix constructed from A and b. The Hessenberg matrix H of the Arnoldi process for $\{A, b\}$ is given by*

$$H = UCU^{-1}. \tag{3.6}$$

The orthonormal matrix Q and the matrix $\mathscr{R}$ in the QR factorization of H defined in (3.3) are

$$Q = V^TW = UY^{-1} = U^{-T}Y^T, \qquad \mathscr{R} = YCU^{-1}. \tag{3.7}$$

This implies the QR factorization $Y = Q^TU$. The matrix Q is upper Hessenberg and its first row is h^T where h is defined in Theorem 2.1. The matrices Q and $\mathscr{R}$ are also related to the APS parametrization by the upper Hessenberg matrix $\mathscr{H} = \mathscr{R}Q = YCY^{-1}$ defined in (3.5).

The first $n-1$ columns of the upper triangular matrix $\tilde{\mathscr{R}}$ in (3.4) are

$$\begin{pmatrix} R \\ 0 \end{pmatrix}$$

where R of order $n-1$ is defined in Theorem 2.1 and the matrices $\mathscr{R}$ and $\tilde{\mathscr{R}}$ are related by

$$\tilde{\mathscr{R}} = \mathscr{R}U. \tag{3.8}$$

We also have the following factorizations

$$V^TA^TAV = \mathscr{R}^T\mathscr{R}, \tag{3.9}$$

$$K^T A^T A K = \tilde{\mathscr{R}}^T \tilde{\mathscr{R}}, \tag{3.10}$$

$$A = W \mathscr{R} V^T. \tag{3.11}$$

Proof Let us first consider the upper triangular matrix U in (3.1). From $K = WY = VU$, we have

$$K^T K = Y^T W^T W Y = Y^T Y = U^T V^T V U = U^T U.$$

From Proposition 3.1 the matrix U is upper triangular with positive diagonal entries. Therefore, the matrix U^T is the common Cholesky factor of $Y^T Y$ and $K^T K$. The relation (3.6) for H that was given in [2] can, in fact, be proved independently of the APS factorization. The proof is so simple that we give it for the convenience of the reader. It is well-known that we have $AK = KC$. Using (3.1) we obtain $AVU = VUC$ and multiplying on the left by V^T,

$$HU = UC.$$

As we said before, the columns of the matrix W in the APS parametrization give a basis of the Krylov space $A\mathscr{K}_n$ and we defined $AK = W\tilde{\mathscr{R}}$ where the matrix $\tilde{\mathscr{R}}$ is upper triangular. We now characterize this matrix. From $K = VU$ it follows that

$$AK = AVU = VHU = W\tilde{\mathscr{R}}.$$

But, since $K = WY$, we have $AK = AWY = (WYCY^{-1}W^T)WY = WYC$. Therefore, $\tilde{\mathscr{R}} = YC$. From the structures of Y and C, this yields that the first $n-1$ columns of $\tilde{\mathscr{R}}$ are

$$\begin{pmatrix} R \\ 0 \end{pmatrix}.$$

The equality $VHU = W\tilde{\mathscr{R}}$ gives

$$H = V^T W \tilde{\mathscr{R}} U^{-1} = Q\mathscr{R},$$

where $Q = V^T W$ is orthogonal and $\mathscr{R} = \tilde{\mathscr{R}} U^{-1} = YCU^{-1}$ is upper triangular. This is a QR factorization of H. It gives the relation between both basis since $W = VQ$. Moreover, it proves (3.8). To obtain other characterizations of Q we use the other relation linking V and W (that is, $WY = VU$) and we obtain

$$Q = UY^{-1} = U^{-T} Y^T.$$

This relation implies that $Y = Q^T U$, which is a QR factorization of Y. The orthogonal factor Q^T is just the transpose of that of H. Moreover, Q is upper Hessenberg. Now we have to prove that the entries of the first row of Q are positive. Let $\hat{h}$ be the vector of the first $n-1$ components of h. Note that with the hypothesis of Theorem 2.1 we have $\eta_n > 0$. The inverse of the matrix Y is

$$Y^{-1} = \begin{pmatrix} 0 & \cdots & 0 & 1/\eta_n \\ & R^{-1} & & -R^{-1}\hat{h}/\eta_n \end{pmatrix}. \tag{3.12}$$

Let $\hat{L}$ be the Cholesky factor of

$$\hat{L}\hat{L}^T = R^T R - R^T \hat{h}\hat{h}^T R. \tag{3.13}$$

It exists since the symmetric matrix on the right-hand side is $R^T(I-\hat{h}\hat{h}^T)R$ and $I-\hat{h}\hat{h}^T$ is positive definite because $\|\hat{h}\|<1$. Then

$$U=\begin{pmatrix} 1 & \hat{h}^T R \\ 0 & \\ \vdots & \hat{L}^T \\ 0 & \end{pmatrix}. \tag{3.14}$$

We have this result for U since

$$Y^TY=\begin{pmatrix} 1 & \hat{h}^T R \\ R^T\hat{h} & R^T R \end{pmatrix}.$$

The results for Y^{-1} and U in (3.12) and (3.14) imply that

$$Q^Te_1=Y^{-T}U^Te_1=Y^{-T}\begin{pmatrix} 1 \\ R^T\hat{h} \end{pmatrix},$$

and therefore

$$Q^Te_1=\begin{pmatrix} \hat{h} \\ \frac{1}{\eta_n}-\frac{\|\hat{h}\|^2}{\eta_n} \end{pmatrix}.$$

We remark that since $\|h\|^2=\|\hat{h}\|^2+\eta_n^2=1$,

$$\frac{1}{\eta_n}-\frac{\|\hat{h}\|^2}{\eta_n}=\eta_n.$$

Therefore, the first row of Q is h^T and its elements are positive. Hence, the GMRES residual norm convergence curve is fully described by the first row of the orthogonal matrix in the QR factorization of H defined in (3.3).

Since $H=Q\mathscr{R}=V^TAV$, we have a factorization of the matrix A as

$$A=VQ\mathscr{R}V^T=VQ(\mathscr{R}Q)Q^TV^T=W\mathscr{R}QW^T=W\mathscr{H}W^T,$$

where the matrix $\mathscr{H}=\mathscr{R}Q=YCY^{-1}$ is upper Hessenberg. Since

$$H=UCU^{-1}=QYCY^{-1}Q^T,$$

we obtain the relation $H=Q\mathscr{H}Q^T$, between the Hessenberg matrices H and $\mathscr{H}$.

Finally, let us prove relations (3.9) to (3.11). Since

$$W=VQ=VH\mathscr{R}^{-1}=AV\mathscr{R}^{-1} \tag{3.15}$$

and from the orthonormality of W we have

$$\mathscr{R}^T\mathscr{R}=V^TA^TAV.$$

Relation (3.15) also gives

$$A=W\mathscr{R}V^T.$$

This means that using the orthonormal bases of $\mathscr{K}_n$ and $A\mathscr{K}_n$ we can reduce A to upper triangular form. For the matrix $\tilde{\mathscr{R}}$, we have

$$\tilde{\mathscr{R}}^T\tilde{\mathscr{R}}=K^TA^TAK.$$ □

One can also prove other relevant relations. Some of the matrices we have seen before are Krylov matrices. We have already seen that $U=(e_1\ He_1\ \cdots\ H^{n-1}e_1)$. The matrices $\tilde{\mathscr{R}}$ and Y are also Krylov matrices as stated in the following proposition.

Proposition 3.2 *The matrix $\tilde{\mathscr{R}}$ in the QR factorization of AK given by (3.4) is*

$$\tilde{\mathscr{R}} = (\, \mathscr{H} h \quad \cdots \quad \mathscr{H}^n h \,).$$

The matrix Y in the APS factorization is

$$Y = (\, h \quad \mathscr{H} h \quad \cdots \quad \mathscr{H}^{n-1} h \,).$$

Proof We have

$$\tilde{\mathscr{R}} = W^T \,(\, Ab \quad \cdots \quad A^n b \,).$$

Since $A = W\mathscr{H}W^T$, we obtain $A^j = W\mathscr{H}^j W^T$. The relation $h = W^T b$ yields the result for $\tilde{\mathscr{R}}$. The relation for $\tilde{\mathscr{R}}$ immediately gives the result for Y since $\tilde{\mathscr{R}} = YC = \mathscr{H}Y$.

Note that the expressions for $\tilde{\mathscr{R}}$ and U give the relation

$$(\, \mathscr{H} h \quad \cdots \quad \mathscr{H}^n h \,) = \mathscr{R}\,(\, e_1 \quad He_1 \quad \cdots \quad H^{n-1} e_1 \,).$$ □

Theorem 3.2 *The GMRES residual norm convergence curve described by h is characterized by the following relation,*

$$[b^T Ab,\, b^T A^2 b,\, \ldots,\, b^T A^{n-1} b] = \hat{h}^T R, \qquad \eta_n = (1 - \hat{h}^T \hat{h})^{1/2}, \tag{3.16}$$

where $\hat{h}$ is the vector of the first $n-1$ components of h defined in Theorem 2.1 and the upper triangular matrix R in the APS parametrization is such that

$$R^T R = \begin{pmatrix} b^T A^T \\ b^T (A^2)^T \\ \vdots \\ b^T (A^{n-1})^T \end{pmatrix} (\, Ab \quad A^2 b \quad \ldots \quad A^{n-1} b \,). \tag{3.17}$$

Proof Relations (3.16) and (3.17) are obtained by comparing the matrices $Y^T Y$ and $K^T K$. □

Note that the matrix R^T is not necessarily the Cholesky factor of the Gram matrix given by (3.17) unless the diagonal entries of R are positive. Theoretically (that is, in exact arithmetic) GMRES residual convergence is contained in (3.16) and (3.17). They give the relation between A, b and h. Of course, numerically, the vector h must not be computed using (3.16).

It is also worth considering the matrix H in greater detail.

Theorem 3.3 *Let $\hat{L}$ be the Cholesky factor defined by (3.13), l be the vector $l = R^T \hat{h}$ and $\tilde{l} = -\hat{L}^{-1} l$. Let β_0 and β be given by $\beta_0 = -\alpha_0 (\hat{L}^{-T})_{n-1,n-1}$, $\beta = (\beta_1 \ \cdots \ \beta_{n-1})^T$ with $\beta_1 = \tilde{l}_{n-1}^T - \alpha_1 (\hat{L}^{-T})_{n-1,n-1}$, and $\beta_{i-1} = (\hat{L}^{-T})_{i-2,n-1} - \alpha_{i-1} (\hat{L}^{-T})_{n-1,n-1}$, $i = 3, \ldots, n$. Then there exists a nonsingular lower triangular matrix $\tilde{L}$, such that*

$$CU^{-1} = \begin{pmatrix} 0 & \beta_0 \\ \tilde{L}^{-T} & \beta \end{pmatrix}. \tag{3.18}$$

The Hessenberg matrix H given by the Arnoldi process for A and b is

$$H = \begin{pmatrix} l^T \tilde{L}^{-T} & \beta_0 + l^T \beta \\ \hat{L}^T \tilde{L}^{-T} & \hat{L}^T \beta \end{pmatrix}. \tag{3.19}$$

Proof The inverse of the matrix U is

$$U^{-1} = \begin{pmatrix} 1 & -l^T \hat{L}^{-T} \\ 0 & \hat{L}^{-T} \end{pmatrix}.$$

The matrix CU^{-1} is upper Hessenberg. Its entries are obtained from l and $\hat{L}$. Since the first row of C is zero except for the last element (which is $-\alpha_0$), the first row of CU^{-1} is (using Matlab-like notations)

$$(CU^{-1})_{1,:} = (0 \quad \cdots \quad 0 \quad \beta_0).$$

The second row of CU^{-1} is the same as the first row of U^{-1} except for the last element,

$$(CU^{-1})_{2,:} = (1 \quad \tilde{l}_1^T \quad \cdots \quad \tilde{l}_{n-2}^T \quad [\tilde{l}_{n-1}^T - \alpha_1 (\hat{L}^{-T})_{n-1,n-1}]).$$

The submatrix of CU^{-1} for the rows 3 to n and the columns 1 to $n-1$ is given by

$$(CU^{-1})_{3:n,1:n-1} = (0 \quad (\hat{L}^{-T})_{1:n-2,1:n-2}),$$

since CU^{-1} is upper Hessenberg. The elements of the last column from 3 to n are

$$\beta_{i-1} = (CU^{-1})_{i,n} = (\hat{L}^{-T})_{i-2,n-1} - \alpha_{i-1} (\hat{L}^{-T})_{n-1,n-1}, \quad i = 3, \ldots, n.$$

Then we can write CU^{-1} in blockwise form as in (3.18) where $\tilde{L}$ is lower triangular. Note that only the last column of CU^{-1} depends on the coefficients α_j in the companion matrix through β_0 and β.

Finally, we have to multiply from the left by U to obtain (3.19). □

It is interesting to remark that only the last column of H depends on the coefficients α_j that define the companion matrix C. Hence, keeping everything else the same, if we change the coefficients α_j (and therefore the eigenvalues of A), only the last column of H will be changed. Moreover, the principal submatrices H_k, $k = 1, \ldots, n-1$ of H do not depend on α_j, $j = 0, \ldots, n-1$. Note that the eigenvalues of the matrices H_k are used as approximations of the eigenvalues of A in the Arnoldi algorithm. Our results do not mean that the Arnoldi algorithm does never deliver good approximations of the eigenvalues. This is a subtle point. Assume that we have a matrix A for which the Ritz values converge towards the eigenvalues. The APS parametrization and our results show that, in the class of matrices having the same residual norms as A, we can construct matrices with other sets of eigenvalues, choosing them far from the Ritz values which are the same for all the matrices in the class of A, constructed with the same matrices W and Y. In fact, it is shown in [2] that one can construct matrices having prescribed residual norms as well as prescribed Ritz values and eigenvalues.

We end this section by considering the matrix V whose columns are the orthonormal basis vectors of $\mathcal{K}_n$. From (3.7) we know that $V = WQ^T = WYU^{-1}$. We already know that the first row of Q is h^T but the orthonormal matrix Q is fully characterized in the following theorem.

Theorem 3.4 (Theorem 4 of [7]) *Let $\check{L}$ be the lower triangular Cholesky factor of the positive definite matrix $I - \hat{h}\hat{h}^T$ and $\mathscr{S}$ be a diagonal matrix whose diagonal entries are ± 1 such that the diagonal entries of $\mathscr{S}R$ are positive. Then*

$$Q = \begin{pmatrix} \hat{h}^T & \eta_n \\ \mathscr{S}\check{L}^T & -\frac{\mathscr{S}\check{L}^T\hat{h}}{\eta_n} \end{pmatrix}. \tag{3.20}$$

Moreover, the entries of $\check{L}^T$ for $j \geq i$ are

$$(\check{L}^T)_{i,j} = -\frac{\eta_i\eta_j}{\sqrt{\eta_{i+1}^2 + \cdots + \eta_n^2}\sqrt{\eta_i^2 + \cdots + \eta_n^2}}, \qquad (\check{L}^T)_{i,i} = \frac{\sqrt{\eta_{i+1}^2 + \cdots + \eta_n^2}}{\sqrt{\eta_i^2 + \cdots + \eta_n^2}}. \tag{3.21}$$

Proof This is proved by computing YU^{-1} and using the results of [4]; see [7]. □

Using this result for the orthonormal matrix Q we obtain a characterization of V.

Theorem 3.5 *With the notation of Theorem 3.4, the matrix V of the orthonormal basis vectors for $\mathscr{K}_n(A, b)$ is*

$$V = WQ^T = W\begin{pmatrix} \hat{h} & \check{L}\mathscr{S} \\ \eta_n & -\frac{\hat{h}^T\check{L}\mathscr{S}}{\eta_n} \end{pmatrix}. \tag{3.22}$$

We see that if we only change the companion matrix C in the APS parametrization, keeping everything else the same, the matrix V is not changed.

4 The GMRES iterates

In this section we express the iterates x_k given by the GMRES algorithm using the matrices involved in the parametrization of [1]. This can be done in at least two different ways. We have assumed that the starting vector is $x_0 = 0$. Then the iterates are computed as

$$x_k = V_k z_k,$$

where V_k is the matrix of the first k columns of V, that is, the first k orthonormal basis vectors of the Arnoldi process. The vector z_k is computed by minimizing the norm of the residual r_k. We have characterized V in Theorem 3.5. Writing the solution of the minimization problem using the normal equation, one can show that

$$z_k = \tilde{L}_k^T R_k^{-1} h_k, \tag{4.1}$$

with the lower triangular matrix $\tilde{L}$ defined in Theorem 3.3, h_k the vector of the first k components of h and R_k the principal matrix of order k of R. It gives that

$$x_k = WYU^{-1}\begin{pmatrix} \tilde{L}_k^T R_k^{-1} h_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{4.2}$$

However, Z. Strakoš suggested an easier way for obtaining the iterates as a function of the matrices in the parametrization.

Theorem 4.1 *Let h_k be the vector of the first k components of h defined in Theorem 2.1 and R_k be the principal matrix of order k of the matrix R defined by (2.1). The GMRES iterates are given by*

$$x_k = WYC^{-1}\begin{pmatrix} 0 \\ R_k^{-1}h_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} = WY\begin{pmatrix} R_k^{-1}h_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad k < n, \tag{4.3}$$

with only one zero element at the top of the vector on the first right-hand side.

Proof The residual vector $r_k = b - Ax_k$ can be written as

$$r_k = b - W_k h_k,$$

where W_k is the matrix of the first k columns of W. This is obtained because the residual is such that

$$\|r_k\| = \min_{u \in A\mathcal{K}_k} \|b - u\|,$$

see [3, 6]. Then we have

$$x_k = A^{-1}W_k h_k = A^{-1}W\begin{pmatrix} \eta_1 \\ \vdots \\ \eta_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

But $A^{-1}W = WYC^{-1}Y^{-1}$. Using the expression of Y^{-1} in (3.12) we have

$$Y^{-1}\begin{pmatrix} \eta_1 \\ \vdots \\ \eta_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ R_k^{-1}h_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

This gives the first result. The inverse of the companion matrix C is

$$C^{-1} = \begin{pmatrix} -\frac{\alpha_1}{\alpha_0} & 1 & 0 & \cdots & 0 \\ -\frac{\alpha_2}{\alpha_0} & 0 & 1 & 0 & \cdots \\ \vdots & 0 & \ddots & \ddots & \ddots \\ -\frac{\alpha_{n-1}}{\alpha_0} & & \ddots & \ddots & 1 \\ -\frac{1}{\alpha_0} & 0 & \cdots & 0 & 0 \end{pmatrix}. \tag{4.4}$$

Therefore,

$$C^{-1}\begin{pmatrix} 0 \\ R_k^{-1}h_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} R_k^{-1}h_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and this yields the second relation. □

Theorem 4.1 has the important consequence that the GMRES iterates x_k for $k < n$ have no explicit relationship with the coefficients α_j of the characteristic polynomial. It is now worth considering the matrix YC^{-1} since it is involved in (4.3) and it will also be needed in the next section.

Theorem 4.2 *We have*

$$YC^{-1} = \left(z(\alpha) \quad h \quad \begin{pmatrix} R_{n-2} \\ 0 \\ 0 \end{pmatrix} \right),$$

where $z(\alpha)$ is a vector (depending on the α_j, $j = 0, \dots, n-1$) given by

$$z(\alpha) = \begin{pmatrix} -\frac{\alpha_1}{\alpha_0}\hat{h} + Rt \\ -\frac{\alpha_1}{\alpha_0}\eta_n \end{pmatrix} = -\frac{\alpha_1}{\alpha_0}h + \begin{pmatrix} Rt \\ 0 \end{pmatrix}$$

with $t \in \mathbb{R}^{n-1}$, $t = (-\frac{\alpha_2}{\alpha_0}, \cdots -\frac{\alpha_{n-1}}{\alpha_0}, -\frac{1}{\alpha_0})^T$.

Proof From (4.4) the inverse of the companion matrix can be written in blockwise form as

$$C^{-1} = \begin{pmatrix} -\frac{\alpha_1}{\alpha_0} & e_1^T \\ t & S \end{pmatrix},$$

where S is the shift matrix with ones on the first superdiagonal. Multiplying from the left with Y we obtain

$$YC^{-1} = \begin{pmatrix} -\frac{\alpha_1}{\alpha_0}\hat{h} + Rt & \hat{h}e_1^T + RS \\ -\frac{\alpha_1}{\alpha_0}\eta_n & \eta_n e_1^T \end{pmatrix}.$$

Since RS is obtained from R by shifting by one column to the right, its first column is zero and, obviously, the second column of YC^{-1} is just h. □

This result shows that only the first column of YC^{-1} depends on the coefficients α_j. Note that, in the expression (4.3), this first column is multiplied by zero.

5 The GMRES error vectors

Since from the previous section we have an expression for the iterates x_k we may want to compute the error vector $\varepsilon_k = x - x_k$. For this we have to obtain the exact solution $x = A^{-1}b$. Using the APS parametrization of A we have

$$x = (WYCY^{-1}W^T)^{-1}b = WYC^{-1}Y^{-1}W^Tb = WYC^{-1}e_1,$$

since $W^Tb = h$ and $Y^{-1}h = e_1$.

Theorem 5.1 *Let h_k the vector of the first k components of h defined in Theorem 2.1. The error vector in GMRES is given by*

$$\varepsilon_k = WYC^{-1}\begin{pmatrix} 1 \\ -R_k^{-1}h_k \\ 0 \end{pmatrix} = W\left(z(\alpha) - \left(h \quad \begin{pmatrix} R_{k-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}\right) R_k^{-1}h_k\right), \tag{5.1}$$

where $z(\alpha)$ is defined in Theorem 4.2.

Proof From the previous results, we have

$$\varepsilon_k = x - x_k = WYC^{-1}\left(e_1 - \begin{pmatrix} 0 \\ R_k^{-1}h_k \\ 0 \end{pmatrix}\right).$$

This directly gives the result. The second expression is obtained with Theorem 4.2. Note that $Wz(\alpha)$ is the exact solution of the linear system. □

Theorem 5.1 shows that the error vector ε_k depends on the eigenvalues through the vector $z(\alpha)$.

6 The GMRES residual vectors

As we claimed before, the iterates $x_k, k < n$ do not depend on the coefficients α_j, $j = 0, \ldots, n-1$ in the sense that if we keep the given W and Y and change C we obtain the same iterates. Since we have an expression for x_k, we can compute the residual $r_k = b - Ax_k$ using the results of Theorem 4.1.

Theorem 6.1 *Using the notation of Theorem 3.4 the GMRES residual vector r_k, $0 < k < n$ is given by*

$$r_k = W\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \eta_{k+1} \\ \vdots \\ \eta_n \end{pmatrix} = K\begin{pmatrix} 1 \\ -R_k^{-1}h_k \\ 0 \end{pmatrix} = V\begin{pmatrix} \eta_{k+1}^2 + \cdots + \eta_n^2 \\ -\mathscr{S}_k \check{L}_k^T h_k \\ 0 \end{pmatrix}.$$

Proof We have already seen in section 4 that

$$r_k = W\left(h - \begin{pmatrix} h_k \\ 0 \end{pmatrix}\right) = W \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \eta_{k+1} \\ \vdots \\ \eta_n \end{pmatrix} \tag{6.1}$$

and $\|r_k\|^2 = \eta_{k+1}^2 + \cdots + \eta_n^2$. From (5.1), using $WY = K$ and $KC^{-1} = A^{-1}K$, we have

$$\varepsilon_k = A^{-1}K \begin{pmatrix} 1 \\ -R_k^{-1} h_k \\ 0 \end{pmatrix}.$$

Since $A\varepsilon_k = r_k$, this directly gives

$$r_k = K \begin{pmatrix} 1 \\ -R_k^{-1} h_k \\ 0 \end{pmatrix}. \tag{6.2}$$

This is the decomposition of the residual vector on the basis given by the Krylov vectors (the columns of K). Since $K = VU$ and

$$U \begin{pmatrix} 1 \\ -R_k^{-1} h_k \\ 0 \end{pmatrix} = \begin{pmatrix} \eta_{k+1}^2 + \cdots + \eta_n^2 \\ -\hat{L}_k^T R_k^{-1} h_k \\ 0 \end{pmatrix},$$

remarking that $-\hat{L}_k^T R_k^{-1} h_k = -\mathscr{S}_k \check{L}_k^T h_k$, we obtain

$$r_k = V \begin{pmatrix} \eta_{k+1}^2 + \cdots + \eta_n^2 \\ -\mathscr{S}_k \check{L}_k^T h_k \\ 0 \end{pmatrix}, \quad k > 0, \qquad r_0 = V e_1. \tag{6.3}$$

This gives the decomposition of r_k on the basis of the Arnoldi vectors which are the columns of V. It is interesting to note that r_k can be described in terms of the first $k+1$ Arnoldi vectors as well as the last $n-k$ basis vectors of $A\mathscr{K}_n$ given by W. $\square$

We can compute what are the elements of $\check{L}_k^T h_k$ from the results in (3.21). It is easily seen that

$$(\check{L}_k^T h_k)_j = \frac{\eta_j (\eta_{k+1}^2 + \cdots + \eta_n^2)}{\sqrt{\eta_j^2 + \cdots + \eta_n^2}\sqrt{\eta_{j+1}^2 + \cdots + \eta_n^2}}.$$

The last element is

$$(\check{L}_k^T h_k)_k = \eta_k \frac{\sqrt{\eta_{k+1}^2 + \cdots + \eta_n^2}}{\sqrt{\eta_k^2 + \cdots + \eta_n^2}}.$$

Summarizing, we have the following result.

$$r_k = (\eta_{k+1}^2 + \cdots + \eta_n^2)\, V \begin{pmatrix} 1 \\ \mathscr{S}_{1,1} \frac{\eta_1}{\sqrt{\eta_1^2+\cdots+\eta_n^2}\sqrt{\eta_2^2+\cdots+\eta_n^2}} \\ \vdots \\ \mathscr{S}_{k,k} \frac{\eta_k}{\sqrt{\eta_k^2+\cdots+\eta_n^2}\sqrt{\eta_{k+1}^2+\cdots+\eta_n^2}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{6.4}$$

with $\mathscr{S}_{j,j} = \pm 1$.

In (6.4) the residual vector r_k depends on its own norm $\|r_k\|$. However, remember that the η_j's were prescribed and thus the previous result just shows what r_k is when all the residual norms are prescribed. Note that

$$\frac{\eta_j}{\sqrt{\eta_j^2 + \cdots + \eta_n^2}\sqrt{\eta_{j+1}^2 + \cdots + \eta_n^2}} = \sqrt{\frac{1}{\|r_j\|^2} - \frac{1}{\|r_{j-1}\|^2}}.$$

7 Conclusion

In this paper we have further discussed the parametrization introduced in [1] related to prescribing the residual norms in GMRES. In particular we provided expressions for the GMRES iterates and error vectors. This showed that the iterates do not depend on the coefficients of the characteristic polynomial (and therefore on the eigenvalues). However, the error vectors do depend on the eigenvalues through the exact solution of the linear system. We have also shown that the principal submatrices of the Hessenberg matrix H computed by the Arnoldi process do not depend on the eigenvalues of A. All these results have to be understood in the sense that in the parametrization described in [1], keeping W and Y, we can change the companion matrix C without changing the iterates, the Arnoldi basis vectors and the principal submatrices of H.

In a forthcoming paper we will explain how, in the class of real matrices defined by the APS parametrization, we can choose the spectrum (or more exactly the coefficients of the characteristic polynomial) to prescribe the exact solution of the linear system (1.1). Additionally one can compute the coefficients of the characteristic polynomial (and therefore the spectrum) to have a prescribed error vector at a given iteration. We will also consider prescribing the norm of the error at every iteration. It turns out that there is a condition that has to be satisfied for the construction of real matrices having prescribed error norms as well as prescribed residual norms.

Acknowledgements This paper was started in 2010 during a visit to the Nečas Center of Charles University in Prague supported by the grant Jindrich Nečas Center for mathematical modeling project LC06052 financed by MSMT. The author thanks particularly Zdeněk Strakoš and Miroslav Rozložník for their kind hospitality.

The author thanks the referees for detailed remarks that help to improve the exposition and for suggesting Proposition 3.1.

References

1. Arioli, M., Pták, V., Strakoš, Z.: Krylov sequences of maximal length and convergence of GMRES. BIT Numer. Math. **38**(4), 636–643 (1998)
2. Duintjer Tebbens, J., Meurant, G.: Any Ritz value behavior is possible for Arnoldi and for GMRES with any convergence curve (2011, submitted)
3. Eiermann, M., Ernst, O.: Geometric aspects of the theory of Krylov subspace methods. Acta Numer. **10**, 251–312 (2001)
4. Gill, P.E., Golub, G.H., Murray, W., Saunders, M.A.: Methods for modifying matrix factorizations. Math. Comput. **28**(126), 505–535 (1974)
5. Greenbaum, A., Strakoš, Z.: Matrices that generate the same Krylov residual spaces. In: Golub, G.H., Greenbaum, A., Luskin, M. (eds.) Recent Advances in Iterative Methods, pp. 95–118. Springer, Berlin (1994)
6. Greenbaum, A., Pták, V., Strakoš, Z.: Any nonincreasing convergence curve is possible for GMRES. SIAM J. Matrix Anal. Appl. **17**, 465–469 (1996)
7. Meurant, G.: Necessary and sufficient conditions for GMRES complete and partial stagnation (2011, submitted)
8. Saad, Y.: Krylov subspace methods for solving large unsymmetric linear systems. Math. Comput. **37**, 105–126 (1981)
9. Saad, Y.: Practical use of some Krylov subspace methods for solving indefinite and unsymmetric linear systems. SIAM J. Sci. Stat. Comput. **5**, 203–228 (1984)
10. Saad, Y., Schultz, M.H.: GMRES: a generalized minimum residual algorithm for solving nonsymmetric linear systems. SIAM J. Sci. Stat. Comput. **7**(3), 856–869 (1986)