NECESSARY AND SUFFICIENT CONDITIONS FOR GMRES COMPLETE AND PARTIAL STAGNATION

GÉRARD MEURANT*

Abstract. In this paper we give necessary and sufficient conditions for the complete or partial stagnation of the GMRES iterative method for solving real linear systems. Our results rely on a paper by Arioli, Pták and Strakoš [2], characterizing the matrices having a prescribed convergence curve for the residual norms. We show that we have complete stagnation if and only if the matrix A is orthonormally similar to an upper or lower Hessenberg matrix having a particular first row or column or a particular last row or column. Partial stagnation is characterized by a particular pattern of the matrix Q in the QR factorization of the upper Hessenberg matrix generated by the Arnoldi process.

1. Introduction. The Generalized Minimum RESidual iterative method (GMRES) was introduced by Saad and Schultz [9] to solve linear systems $Ax = b$, where A is a nonsingular real matrix of order n. It is a Krylov method based on the Arnoldi orthogonalization process using an orthogonal basis of a Krylov space. 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 $\mathcal{K}_k(r^0, A)$ is $span\{r^0, Ar^0, \dots, A^{k-1}r^0\}$. The approximate solution x^k at iteration k is sought as $x^k \in x^0 + \mathcal{K}_k(r^0, A)$, such that the norm of the residual vector $r^k = b - Ax^k$ is minimized.

In this paper we study the problem of stagnation of the GMRES algorithm for real matrices and right-hand sides. Partial stagnation is defined for a given matrix A and right-hand side b as having $\|r^k\| < \|r^{k-1}\|$, $k = 1, \dots, m$, $\|r^k\| = \|r^{k-1}\|, k = m+1, \dots, m+p-1$, and $\|r^k\| < \|r^{k-1}\|$, $k = m+p, \dots, n$. Hence the norms of the residual stay the same for p iterations starting from $k = m$. Complete stagnation corresponds to $m = 0$ and $p = n$. Thus $\|r^k\| = \|r^0\|$, $k = 1, \dots, n-1$, and $\|r^n\| = 0$. Since $\|r^{n-1}\| \neq 0$ implies that the degree of the minimal polynomial of A is equal to n, we assume that the matrix A is nonderogatory. Without loss of generality we will use $x^0 = 0$; therefore $r^0 = b$ and we will assume that $\|b\| = 1$.

Complete stagnation of GMRES has been studied by Zavorin, O'Leary and Elman in [13] and Zavorin in [12]; see also Simoncini and Szyld [11] and Simoncini [10] who studied conditions for non–stagnation, Liesen and Tichý [8] for the study of worst-case GMRES for normal matrices and Arioli [1]. Up to our knowledge a theoretical study of partial stagnation does not exist in the literature. Note that the definition of partial stagnation in [13] is different from ours. Their definition corresponds to $m = 0$; that is, stagnation at the beginning of GMRES iterations. We will study the stagnation phenomenon using results established for the convergence curve of the residual norms. In [6] Greenbaum and Strakoš proved that any convergence curve for the residual norm that can be generated with GMRES can be obtained with a matrix having prescribed eigenvalues. Greenbaum, Pták and Strakoš [7] showed later that any nonincreasing sequence of residual norms can be given by GMRES. To complement these results, Arioli, Pták and Strakoš [2] gave a complete parametrization of all pairs $\{A, b\}$ generating a prescribed residual norm convergence curve.

Of course, having complete or partial stagnation is a very special case of prescribed residual norm convergence curve. We will build upon Arioli, Pták and Strakoš parametrization to obtain necessary and sufficient conditions for complete or partial stagnation. In the course of our study we will see that the orthogonal matrix Q in the

*(gerard.meurant@gmail.com) started in Prague, June 2010. Version Feb 28, 2011

QR factorization of the upper Hessenberg matrix H generated by the Arnoldi process has a very special structure in case of stagnation.

The contents of the paper are as follows. Section 2 recall the results from Arioli, Pták and Strakoš that we need for our purposes. Section 3 proves general results that are useful for studying stagnation. In section 4 we prove an essential result for the orthogonal factor of a QR factorization of the Hessenberg matrix generated by the Arnoldi process. This result is useful to study complete and partial stagnation. In section 5 we characterize the matrices and right-hand sides leading to complete stagnation. In section 6 we study partial stagnation. Finally we give some conclusions.

Throughout the paper e^j will denote the jth column of the identity matrix of appropriate order.

2. The Arioli, Pták and Strakoš parametrization. We recall the following results that were proved in [2] (Theorem 2.1 and Corollary 2.4).

THEOREM 2.1. *Assume we are given n positive numbers*

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

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

1- *The spectrum of A is $\{\lambda_1, \ldots, \lambda_n\}$ and GMRES applied to A and b 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*

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

R being any nonsingular upper triangular matrix of order $n-1$, h a vector 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, $A = WYCY^{-1}W^*$, the APS parametrization. Here we are only interested in real matrices A and right-hand sides b, so we can replace the $*$ above, that means conjugate transpose, by the transpose. In the parametrization, the prescribed residual norm convergence curve is implicitly contained in the vector h which is the first column of Y. The squares of its elements are the differences of the squares of the residual norms at successive iterations. Whatever the values of W orthogonal, R upper triangular and C a companion matrix are, we will obtain with all these matrices A the same convergence curve for GMRES as long as the right-hand side is $b = Wh$.

When we have partial stagnation some of the residual norms are the same. Hence we have $\eta_j = 0, j = m+1, \ldots, m+p-1$. The vector h has therefore $p-1$ consecutive zero components. In case of complete stagnation all the residual norms are the same except for the last one. In this case we have $\eta_j = 0, j = 1, \ldots, n-1$. The vector h has only the last component which is nonzero. Moreover, since $h = W^* b$ and W is orthogonal, we have that $\|h\| = \|b\| = 1$ from our assumption on the norm of b and therefore $h = e^n$.

3. Some properties of the GMRES algorithm. In this section we prove some general properties for the matrices that are involved in the GMRES algorithm using the APS parametrization. The following results do not assume stagnation. Let K be the Krylov matrix, $K = \begin{pmatrix} b & Ab & A^2b & \cdots & A^{n-1}b \end{pmatrix}$, whose columns are the natural basis vectors of the Krylov space $\mathcal{K}_n$. We use the notation of Theorem 2.1. The first interesting property is that $K = WY$; this was proven in [2]. Let V be the orthogonal matrix whose columns are the mutually orthogonal Arnoldi basis vectors. Then let $K = VU$ be the QR factorization of K where the matrix U is upper triangular. Eventually modifying the signs of the Arnoldi vectors one can obtain a matrix U with positive diagonal elements. The Arnoldi process gradually computes an upper Hessenberg matrix H. At the end of the iterations we have $AV = VH$.

THEOREM 3.1. *The matrix U^T is the Cholesky factor of the matrix Y^TY. The Hessenberg matrix H of the Arnoldi process is given by*

$$H = UCU^{-1}.$$

A QR factorization of H is $H = Q\mathcal{R}$ where $Q = V^TW$ is upper Hessenberg orthogonal and $\mathcal{R}$ is upper triangular. Moreover, we have

$$Q = UY^{-1} = U^{-T}Y^T, \quad \mathcal{R} = YCU^{-1}.$$

The matrices Q and $\mathcal{R}$ are also related to the APS parametrization by $\mathcal{R}Q = YCY^{-1}$.

Proof. We have

$$K^TK = Y^TW^TWY = Y^TY = U^TV^TVU = U^TU.$$

The matrix U being upper triangular with positive diagonal elements, U^T is the Cholesky factor of Y^TY. Since $K = WY = VU$ we have $Y = W^TVU$ and

$$H = V^TAV = V^TWYCY^{-1}W^TV = (W^TV)^TW^TVUCU^{-1}(W^TV)^TW^TV = UCU^{-1}.$$

The columns of the matrix W in the APS parametrization give a basis of the Krylov space $A\mathcal{K}_n(A,b)$ and we have $AK = W\tilde{R}$ where the matrix $\tilde{R}$ is upper triangular. We have also that $K = VU$. Hence

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

But, since $K = WY$, we have $AK = AWY = (WYCY^{-1}W^T)WY = WYC$. Therefore, $\tilde{R} = YC$. The previous equality gives

$$H = V^TW\tilde{R}U^{-1} = Q\mathcal{R},$$

where $Q = V^TW$ is orthogonal and $\mathcal{R} = YCU^{-1}$ is upper triangular. This is a QR factorization of H. It gives the relation between both basis since $W = VQ$. Using the other relation linking V and W (that is, $WY = VU$) 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.

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

$$A = VQ\mathcal{R}V^T = VQ(\mathcal{R}Q)Q^T V^T = W\mathcal{H}W^T.$$

The matrix

$$\mathcal{H} = \mathcal{R}Q = YCY^{-1}$$

is also upper Hessenberg. This is an RQ factorization of $\mathcal{H}$. Since

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

we obtain the relation between H and $\mathcal{H}$, $H = Q\mathcal{H}Q^T$. The matrix Q transforms the upper Hessenberg matrix $\mathcal{H}$ to the upper Hessenberg matrix H. □

4. What is the matrix Q?. In this section we characterize the orthogonal matrix Q that was introduced in Theorem 3.1 as a factor of the decomposition $H = Q\mathcal{R}$. The matrix Q is given by $Q = UY^{-1}$. Therefore we first compute U and the inverse of Y.

LEMMA 4.1. *Using the notation of Theorem 2.1, let $\hat{h}$ be the vector of the first $n-1$ components of h. For the inverse of the matrix Y, we have*

$$Y^{-1} = \begin{pmatrix} 0 & \cdots & 0 & 1/\eta_n \\ & R^{-1} & & -R^{-1}\hat{h}/\eta_n \end{pmatrix}. \tag{4.1}$$

Let $\hat{L}$ be the Cholesky factor of $R^T R - R^T \hat{h}\hat{h}^T R$. Then

$$U = \begin{pmatrix} 1 & \hat{h}^T R \\ 0 & \\ \vdots & \hat{L}^T \\ 0 & \end{pmatrix}. \tag{4.2}$$

Proof. One can easily check that the matrix in (4.1) is the inverse of the matrix Y. The matrix U^T is the Cholesky factor of the symmetric positive definite matrix $Y^T Y$. From the definition of Y and $\|h\| = 1$, we have

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

It is easy to see that the Cholesky factor L of the matrix $Y^T Y$ is

$$L = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ R^T \hat{h} & & \hat{L} & \end{pmatrix}$$

where $\hat{L}\hat{L}^T = R^T R - R^T \hat{h}\hat{h}^T R$ is the Cholesky factorization of the positive definite matrix in the right-hand side. Since $U = L^T$ this proves the result for the matrix U. □

Since we know the matrices U and Y^{-1} from Lemma 4.1, we can immediately write down Q as

$$Q = \begin{pmatrix} \hat{h}^T & \frac{1}{\eta_n} - \frac{\|\hat{h}\|^2}{\eta_n} \\ \hat{L}^T R^{-1} & -\frac{\hat{L}^T R^{-1}\hat{h}}{\eta_n} \end{pmatrix}. \tag{4.3}$$

However this expression can be considerably simplified as we see in the following theorem.

THEOREM 4.2. *Let $\check{L}$ be the lower triangular Cholesky factor of the matrix $I-\hat{h}\hat{h}^T$ and $\mathcal{S}$ be a diagonal matrix whose diagonal entries are ± 1 according to the signs of the diagonal entries of R. Then*

$$Q = \begin{pmatrix} \hat{h}^T & \eta_n \\ \mathcal{S}\check{L}^T & -\frac{\mathcal{S}\check{L}^T\hat{h}}{\eta_n} \end{pmatrix}. \tag{4.4}$$

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

$$(\check{L}^T)_{i,j} = -\frac{\eta_i\eta_j}{\sqrt{\eta_{i+1}^2+\cdots+\eta_n^2}\sqrt{\eta_i^2+\cdots+\eta_n^2}}, \quad (\check{L}^T)_{i,i} = \frac{\sqrt{\eta_{i+1}^2+\cdots+\eta_n^2}}{\sqrt{\eta_i^2+\cdots+\eta_n^2}}. \tag{4.5}$$

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

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

From Lemma 4.1 we have $\hat{L}\hat{L}^T = R^TR - R^T\hat{h}\hat{h}^TR$. This implies

$$\hat{L}^TR^{-1} = \hat{L}^{-1}R^T(I-\hat{h}\hat{h}^T).$$

Hence, we have

$$(R^{-T}\hat{L})(\hat{L}^TR^{-1}) = I - \hat{h}\hat{h}^T = \check{L}\check{L}^T.$$

However we cannot identify $\hat{L}^TR^{-1}$, which appears in Q, with $\check{L}^T$ since this last matrix is the Cholesky factor of $I-\hat{h}\hat{h}^T$ and has positive diagonal entries whence the entries of the upper triangular matrix $\hat{L}^TR^{-1}$ may be negative. The matrices $\hat{L}^T$ and R^{-1} being both upper triangular, we have

$$(\hat{L}^TR^{-1})_{i,i} = \frac{(\hat{L}^T)_{i,i}}{(R)_{i,i}}.$$

Since $(\hat{L}^T)_{i,i} > 0$, the diagonal entries $(\hat{L}^TR^{-1})_{i,i}$ have the sign of $(R)_{i,i}$. Let $\mathcal{S}$ be a diagonal matrix whose diagonal entry $(\mathcal{S})_{i,i}$ is plus or minus one according to the sign of $(R)_{i,i}$. Then we have $\hat{L}^TR^{-1} = \mathcal{S}\check{L}^T$. This together with (4.3) proves the result.

Moreover, since $\check{L}$ is the lower triangular Cholesky factor of a rank-one modification of the identity matrix, its entries can be computed. This can be done using the results of [5], where the authors considered a factorization $\mathcal{L}\mathcal{D}\mathcal{L}^T$ with $\mathcal{L}$ lower triangular and $\mathcal{D}$ diagonal. The matrix $\mathcal{L}$ has a very special structure with ones on the diagonal and

$$(\mathcal{L})_{i,j} = \eta_i\gamma_j,\ i = 2,\ldots,n,\ j = 1,\ldots,n-1.$$

The values γ_j and the diagonal elements d_j of the diagonal matrix $\mathcal{D}$ are computed by recurrences

$$\delta_1 = -1 = -\frac{1}{\|r^0\|^2},$$

and for $j = 1, \ldots, n$,

$$d_j = 1 + \delta_j(\eta_j)^2, \quad \gamma_j = \delta_j \frac{\eta_j}{d_j}, \quad \delta_{j+1} = \frac{\delta_j}{d_j}.$$

It is easy to see that

$$d_j = \frac{\eta_{j+1}^2 + \cdots + \eta_n^2}{\eta_j^2 + \cdots + \eta_n^2} = \frac{\|r^j\|^2}{\|r^{j-1}\|^2}, \quad \delta_j = -\frac{1}{\eta_j^2 + \cdots + \eta_n^2} = -\frac{1}{\|r^{j-1}\|^2},$$

and

$$\gamma_j = -\frac{\eta_j}{\eta_{j+1}^2 + \cdots + \eta_n^2} = -\frac{\eta_j}{\|r^j\|^2}.$$

We have to take the square roots of the elements d_j to obtain the Cholesky factor $\check{L} = \mathcal{L}\sqrt{\mathcal{D}}$. The multiplying factor for the column j below the diagonal element is

$$\gamma_j \sqrt{d_j} = -\frac{\eta_j}{\sqrt{\eta_{j+1}^2 + \cdots + \eta_n^2}\sqrt{\eta_j^2 + \cdots + \eta_n^2}} = -\frac{\eta_j}{\|r^j\|\,\|r^{j-1}\|},$$

and therefore $(\check{L})_{i,j} = \eta_i \gamma_j \sqrt{d_j}$. The diagonal element of the column j is

$$(\check{L})_{j,j} = \sqrt{d_j} = \frac{\sqrt{\eta_{j+1}^2 + \cdots + \eta_n^2}}{\sqrt{\eta_j^2 + \cdots + \eta_n^2}} = \frac{\|r^j\|}{\|r^{j-1}\|}.$$

□

From Theorem 4.2 we see that the first row of the upper Hessenberg orthogonal matrix Q is h^T. Therefore the GMRES residual norm convergence (described by h) can be read from the first row of Q, the orthogonal factor of a QR factorization of H. Note that, in this factorization, the signs are such that the entries of the first row of Q are positive.

The matrix $\check{L}$ depends only on the components η_j of h. So, up to the signs (which depend on the signs of the diagonal entries of R), it is the same for Q. We remark that, for $i < j < n$, the entries $Q_{i,j}$ are proportional to $\eta_{i+1}\eta_j$. The results of Theorem 4.2 give a complete description of the matrix Q in terms of the components η_j.

In practical implementations of GMRES, the upper Hessenberg matrix H is reduced to upper triangular form by Givens rotations. An expression of the matrix Q using the sines and cosines of the rotations was given by P.N. Brown, see [3] and [4].

Note that one can also consider the matrix $\mathcal{R} = YCU^{-1}$. Unfortunately, the expression of $\mathcal{R}$ is not as nice as the one for Q and since it is not relevant for the study of stagnation, we do not consider it furthermore.

5. Complete stagnation. Let us see what we obtain from the relations of the previous section when we have complete stagnation that is, h the first column of Y is e^n. This implies that the vector $\hat{h}$ of the first $n-1$ components is identically zero and $\check{L} = I$.

THEOREM 5.1. *In case of complete stagnation, the orthogonal matrix Q in $H = Q\mathcal{R}$ is*

$$Q = \begin{pmatrix} 0 & 1 \\ \mathcal{S} & 0 \end{pmatrix}, \tag{5.1}$$

where $\mathcal{S}$ is a diagonal matrix with ± 1 as diagonal entries according to the signs of the diagonal entries of R.

Note that Theorem 5.1 implies that $H = Q\mathcal{R}$ is an upper Hessenberg matrix with a first row proportional to $(e^n)^T$; that is, all the elements of the first row are zero, except the last one which is equal to $-\alpha_0(\hat{L}^{-T}e^{n-1})_{n-1}$.

Note also that since $h = e^n$, we have

$$b^T K = b^T WY = (e^n)^T Y = (e^1)^T.$$

Therefore, as it is well-known (see [13]), we have $b^T A^j b = 0$, $j = 2, \ldots, n-1$. This is a necessary and sufficient condition to have complete stagnation. However, with Theorem 5.1, we are now able to characterize in a different way the matrices and right-hand sides for which we have complete stagnation.

THEOREM 5.2. *We have complete stagnation in GMRES if and only if the nonderogatory real matrix A can be written as $A = W\mathcal{U}PW^T$, where W is orthogonal, $\mathcal{U}$ is upper triangular and nonsingular and*

$$P = \begin{pmatrix} 0 & 1 \\ I & 0 \end{pmatrix}, \tag{5.2}$$

is a permutation matrix. The right-hand side giving stagnation is $b = We^n$.

Proof. Assume that we have complete stagnation. From Theorems 2.1 and 3.1

$$\begin{aligned} A &= WYCY^{-1}W^T \\ &= WYU^{-1}UCU^{-1}UY^{-1}W^T \\ &= WYU^{-1}HUY^{-1}W^T \\ &= WQ^T HQW^T \\ &= W\mathcal{R}QW^T. \end{aligned}$$

The matrix Q has the structure given in (5.1). Such a matrix can be written as $Q = \check{\mathcal{S}}P$ with P defined in (5.2) and $\check{\mathcal{S}}$ is a diagonal matrix with $(\check{\mathcal{S}})_{1,1} = 1$ and $(\check{\mathcal{S}})_{j,j} = (\mathcal{S})_{j-1,j-1}$, $j = 2, \ldots, n$. The sign matrix $\check{\mathcal{S}}$ can be absorbed in the upper triangular matrix by defining $\mathcal{U} = \mathcal{R}\check{\mathcal{S}}$. Moreover $b = Wh = We^n$ in the APS parametrization.

Conversely, let us assume that $A = W\mathcal{U}PW^T$ and $b = We^n$. From [7], section 2 page 466, we just need to show that the columns of W are a basis of $A\mathcal{K}_n$. We proceed by induction. Since $W^T b = e^n$ and $Pe^n = e^1$ we have

$$Ab = W\mathcal{U}PW^T b = W\mathcal{U}e^1.$$

But $\mathcal{U}e^1$ is a vector proportional to e^1. Hence Ab is proportional to w^1, the first column of W. More generally, assume that $A^{j-1}b = Wq$ where q is a vector with only the first $j-1$ components non zero. Then,

$$A^j b = A(A^{j-1}b) = W\mathcal{U}Pq.$$

The j first components of the vector Pq are $0, q_1 \ldots, q_{j-1}$. Multiplying the vector Pq by the nonsingular upper triangular matrix $\mathcal{U}$, we obtain a vector with the first j components non zero showing that $A^j b$ is in the span of the j first columns of W. ☐

The result of Theorem 5.2 could have been obtained from Theorem 2.2 in [2] page 639. In case of complete stagnation the matrix denoted $\hat{H}$ in [2] is the same as P. We note that $\mathcal{U}P$ is an upper Hessenberg matrix whose last column is proportional to e^1.

If we have complete stagnation for A and b, then we have also complete stagnation for A^T and b. This was shown in [13] and follows from

$$b^T A^j b = b^T (A^j)^T b = b^T (A^T)^j b = 0,\ j = 1, \dots, n-1.$$

The transpose of A is written as $WP^T\mathcal{U}^TW^T$. The matrix P^T is such that $P^Te^1 = e^n$ and $\mathcal{U}^T$ is lower triangular. The matrix $P^T\mathcal{U}^T$ is lower Hessenberg with a last row proportional to $(e^1)^T$.

Another characterization of the matrices leading to complete stagnation is obtained by using the matrices in the Arnoldi process. We have $AV = VH$ and thus

$$A = VHV^T.$$

We have seen that the upper Hessenberg matrix H is $H = Q\mathcal{R}$ with Q defined in (5.1). Hence the matrix A can also be written as

$$A = VQ\mathcal{R}V^T$$

and $b = v^1 = Ve^1$. Notice that W and V are linked by $W = VQ$. Thus $b = We^n = Ve^1$. The matrix Q can be written as $Q = P\breve{\mathcal{S}}$ where P is defined in (5.2) and $\breve{\mathcal{S}}$ is a diagonal matrix with $(\breve{\mathcal{S}})_{j,j} = (\mathcal{S})_{j,j}$, $j = 1, \dots, n-1$ and $(\breve{\mathcal{S}})_{n,n} = 1$. The sign matrix $\breve{\mathcal{S}}$ can be absorbed in the upper triangular matrix by defining $\mathcal{U} = \breve{\mathcal{S}}\mathcal{R}$. It gives $A = VP\mathcal{U}V^T$.

As before we have also stagnation for matrices

$$A^T = V\mathcal{U}^TP^TV^T.$$

The previous discussion is summarized in the following theorem.

THEOREM 5.3. *We have complete stagnation in GMRES if and only the nonderogatory real matrix A and the real right-hand side b can be written as in one of the four cases:*

$$\begin{aligned} A &= W\mathcal{U}PW^T, \quad b = We^n, \\ A &= VP\mathcal{U}V^T, \quad b = Ve^1, \\ A &= WP^T\mathcal{L}W^T, \quad b = We^n, \\ A &= V\mathcal{L}P^TV^T, \quad b = Ve^1. \end{aligned}$$

where W and V are orthogonal matrices, $\mathcal{U}$ is an upper triangular matrix, $\mathcal{L}$ is a lower triangular matrix and P is the permutation matrix defined in (5.2).

This theorem essentially says that we have complete stagnation if and only if the matrix A is orthogonally similar to an upper Hessenberg matrix with either the first row proportional to $(e^n)^T$ or the last column proportional to e^1 or to a lower Hessenberg matrix with the last row proportional to $(e^1)^T$ or the first column proportional to e^n and the right-hand sides are chosen properly.

6. Partial stagnation. Assume that we have partial stagnation as defined in section 1. Then, we can characterize the orthogonal matrix Q in the QR factorization of H.

THEOREM 6.1. *In case of partial stagnation of GMRES the columns $j = m + 1, \dots, m + p - 1$, of Q are zero except for the subdiagonal entries $(j + 1, j)$ which are ± 1. The rows $i = m + 2, \dots, m + p$, are zero for columns i to n.*

Proof. We apply Theorem 4.2. We have seen that $Q_{i,j}$ is proportional to $\eta_{i+1}\eta_j$ for $1 < i \leq j < n$. Therefore, the columns $m+1$ to $m+p-1$ are zero except for the subdiagonal entries which are ± 1 depending on the sign of the corresponding diagonal entry of R. The rows i from $m+2$ to $m+p$ are zero for columns i to $n-1$. The entries in the last column for these rows are zero because of the structure of $\check{L}^T$ and $\hat{h}$. □

Then we can characterize the matrices and right-hand sides leading to partial stagnation.

THEOREM 6.2. *We have partial stagnation in GMRES for iteration $m \geq 0$ to iteration $m+p-1$ if and only the non-derogatory real matrix A and the real right-hand side b can be written as*

$$A = W\mathcal{R}QW^T, \quad b = WQ^T e^1,$$

where W is orthogonal, $\mathcal{R}$ is upper triangular, Q is orthogonal with the sparsity structure defined in Theorem 6.1. We have also the same partial stagnation if and only if $A = VQ\mathcal{R}V^T$ and $b = Ve^1$ with V orthogonal and Q and $\mathcal{R}$ as before.

Proof. Assume that we have partial stagnation as defined above. From theorems 2.1 and 3.1,

$$A = WYCY^{-1}W^T = W\mathcal{R}QW^T.$$

We have seen from Theorems 4.2 and 6.1 that Q has the required structure and values. Moreover $b = Wh = WQ^T e^1$ in the APS parametrization. The proof of the converse is essentially the same as in Theorem 5.2. The proof of the other assertion follows from the relation between W and V in the APS parametrization. □

Of course, one can easily extend this theorem to the case where we have several sequences of stagnation separated with phases of strict decrease of the residual norm. We also remark that partial stagnation does not lead to orthogonality of b with respect to columns of K (that is, $b^T A^j b = 0$) like in the case of complete stagnation.

An interesting question is to know what happens if we have near stagnation; that is, successive norms of the residual are not exactly the same but almost the same. Since we still have $Q^T e^1 = h$, the corresponding entries in the first row of Q are small being the components of h. For the other rows in these columns of Q, the answer to the question is given in the discussion in Theorem 4.2. Let us look at column j and assume that η_j is small compared to 1. The entry in row $i+1$ of column j is

$$\pm \frac{\eta_j \eta_i}{\sqrt{\eta_{i+1}^2 + \cdots + \eta_n^2}\sqrt{\eta_i^2 + \cdots + \eta_n^2}},$$

for $i < j$. The absolute value of this entry depends on the convergence history of the residual norms before iteration j. The subdiagonal entry is

$$\pm \frac{\sqrt{\eta_{j+1}^2 + \cdots + \eta_n^2}}{\sqrt{\eta_j^2 + \cdots + \eta_n^2}}.$$

Its absolute value is almost one if η_{j+1} is small.

7. Conclusions. In this paper we have given necessary and sufficient conditions for the complete and partial stagnation of GMRES. Unfortunately, given a matrix

A, these conditions are not easy to check in practice. However, it is hoped that our results could lead to a better understanding of GMRES behavior, particularly near-stagnation. Theorems 5.2 and 6.2 can also be used to generate examples with complete or partial stagnation.

Acknowledgments.
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.

REFERENCES

[1] M. Arioli, *An analysis of GMRES worst case convergence*, manuscript, (2010).
[2] M. Arioli, V. Pták and Z. Strakoš, *Krylov sequences of maximal length and convergence of GMRES*, BIT Numerical Mathematics, v 38 n 4 (1998), pp. 636–643.
[3] P.N. Brown, *A theoretical comparison of the Arnoldi and GMRES algorithms*, SIAM J. Sci. Stat. Comput., v 12 n 1 (1991), pp. 58–78.
[4] P.N. Brown and A.C. Hindmarsh, *Reduced storage methods in stiff ODE systems*, J. Appl. Math. Comput., v 31 (1989), pp. 40–91.
[5] P.E. Gill, G.H. Golub, W. Murray and M.A. Saunders, *Methods for modifying matrix factorizations*, Math. Comp., v 28 n 126 (1974), pp. 55–535.
[6] A. Greenbaum and Z. Strakoš, *Matrices that generate the same Krylov residual spaces*, in Recent advances in iterative methods, G.H. Golub, A. Greenbaum and M. Luskin, eds., Springer, (1994), pp. 95–118.
[7] A. Greenbaum, V. Pták and Z. Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., v 17 (1996), pp. 465–469.
[8] J. Liesen and P. Tichý, *The worst-case GMRES for normal matrices*, BIT Numerical Mathematics, v 44 (2004), pp. 79–98.
[9] Y. Saad and M.H. Schultz, *GMRES: a generalized minimum residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., v 7 n 3 (1986), pp. 856–869.
[10] V. Simoncini, *On a non–stagnation condition for GMRES and application to saddle point matrices*, Elec. Trans. Numer. Anal., v 37 (2010), pp. 202–213.
[11] V. Simoncini and D.B. Szyld, *New conditions for non–stagnation of minimal residual methods*, Numer. Math., v 109 n 3 (2008), pp. 477–487.
[12] I. Zavorin, *Analysis of GMRES convergence by spectral factorization of the Krylov matrix*, Ph.D. thesis, University of Maryland, (2001).
[13] I. Zavorin, D.P. O'Leary and H. Elman, *Complete stagnation of GMRES*, Linear Alg. Appl., v 367 (2003), pp. 165–183.