BIT
1998, Vol. 38, No. 4, pp. 636–643

0006-3835/98/3804-0636 $12.00
© Swets & Zeitlinger

KRYLOV SEQUENCES OF MAXIMAL LENGTH AND CONVERGENCE OF GMRES *

M. ARIOLI[1] , V. PTÁK[2] and Z. STRAKOŠ[2] †

[1] *IAN CNR, Via Abbiategrasso 209, 27100 Pavia, Italy*
email: arioli@dragon.ian.pv.cnr.it
[2] *ICS AS CR, Pod Vodárenskou věží 2, 182 07 Praha 8, Czech Republic*
email: ptak@math.cas.cz, strakos@uivt.cas.cz

Abstract.

In most practical cases, the convergence of the GMRES method applied to a linear algebraic system $Ax = b$ is determined by the distribution of eigenvalues of A. In theory, however, the information about the eigenvalues alone is not sufficient for determining the convergence. In this paper the previous work of Greenbaum et al. is extended in the following direction. It is given a complete parametrization of the set of all pairs $\{A, b\}$ for which GMRES(A, b) generates the prescribed convergence curve while the matrix A has the prescribed eigenvalues. Moreover, a characterization of the right hand sides b for which the GMRES(A, b) converges exactly in m steps, where m is the degree of the minimal polynomial of A, is given.

AMS subject classification: 65F10, 65F15.

Key words: Minimal polynomial, Krylov sequences, GMRES method, convergence, Jordan canonical form.

1 Introduction.

In the paper [1] Greenbaum and Strakoš proved, among other results, that any convergence curve that can be generated with the GMRES method [4] can be generated by the method applied to a matrix having any prescribed eigenvalues. The GMRES method minimizes the norm of the residual over the subspaces of increasing dimensions, and its convergence curve is therefore nonincreasing. It remained unclear whether any nonincreasing sequence of residual norms can be given by GMRES. This question was answered positively by Greenbaum, Pták and Strakoš in [2] with the restriction to sequences converging to zero exactly in step n, where n is the dimension of the linear system. There was also described the set of all matrices (with arbitrary spectra) and right hand sides for which GMRES generates the required residual norms.

One may impose an additional restriction and fix the spectrum of the matrix A. Still, any nonincreasing convergence curve can be generated by GMRES

*Received September 1997. Revised March 1998. Communicated by Åke Björck.

†This work was supported by AS CR Grant A2030706. Part of the work was performed while the third author visited Instituto di Analisi Numerica (IAN CNR).

applied to some matrix A and some right hand side b . In analogy to [2], we wish to determine *all such pairs* $\{A, b\}$. A complete parametrization of this problem is given in Section 2.

We also relax the assumption that GMRES converges exactly in step n. Clearly, the number of GMRES steps for a given $\{A, b\}$ is equal to the length (the maximum number of the linearly independent vectors) of the Krylov sequence $b, Ab, A^2b, \ldots$ which is limited by the degree of the minimal polynomial of the matrix A. Given the matrix A, we characterize in Section 3 those right hand sides for which the length of the Krylov sequences is maximal, i.e. equal to the degree of the minimal polynomial.

Throughout the paper, A represents an n by n nonsingular matrix and b an n-dimensional vector (both may be complex). Reformulation of the results for A, b real is straightforward.

2 Parametrization of the problem.

Similarly to [2], we assume without loss of generality that the initial guess x^0 is zero and the right hand side b is the initial residual, $b \equiv r^0$. Let GMRES(A, b) converge to the exact solution x at the step n, where n is the dimension of the problem. The GMRES residuals r^j, $j = 1, \ldots, n$, are determined by the minimization property

$$\|r^j\| = \min_{p \in \mathcal{P}_j} \|p(A)\, r^0\|, \tag{2.1}$$

where $\mathcal{P}_j$ is the class of polynomials $p(\lambda)$ of degree at most j satisfying $p(0) = 1$. Clearly, the assumption $r^{n-1} \neq 0$ implies that the degree of the minimal polynomial of the matrix A must be equal to n. Consequently, no eigenvalue of the matrix A has geometric multiplicity larger than one, and A is a nonderogatory matrix. A classical matrix theory result states that a matrix is nonderogatory if and only if it is similar to its companion (see, e.g., [3, Theorem 3.3.15, p. 147]). One can conclude that if GMRES(A, b) does not converge to the exact solution until the last step n, then A must be similar to its companion matrix. The following theorem and its proof describe this similarity transformation explicitly. Moreover, the statement characterizes the set of all pairs $\{A, b\}$, for which GMRES(A, b) generates the prescribed convergence curve while the matrix A has the prescribed eigenvalues.

THEOREM 2.1. *Suppose 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 zero. Let A be an n by n matrix, b an n-dimensional vector. Then the following assertions are equivalent:

1° The spectrum of A is $\{\lambda_1, \ldots, \lambda_n\}$ and GMRES applied to the pair $\{A, b\}$ yields residuals $r^0, \ldots, r^{n-1}$ such that

$$\|r^j\| = f(j), \qquad j = 0, 1, \ldots, n-1.$$

2° The matrix A is of the form $A = WRCR^{-1}W^*$ and $b = Wh$, where C is the companion matrix corresponding to the polynomial q, W is unitary, and R nonsingular upper triangular such that $Rs = h$.

The polynomial q and the vectors s, h are constructed as follows:

$$\begin{aligned} q(z) &= (z-\lambda_1)\cdots(z-\lambda_n) = z^n - \sum_{j=0}^{n-1} \alpha_j z^j, \\ s &= (\xi_1,\ldots,\xi_n)^T \quad \text{where} \quad 1-(\xi_1 z + \cdots + \xi_n z^n) = \prod_{i=1}^{n}\left(1-\frac{z}{\lambda_i}\right), \\ h &= (\eta_1,\ldots,\eta_n)^T \quad \text{where} \quad \eta_j = (f(j-1)^2 - f(j)^2)^{1/2}, \\ & \qquad\qquad\qquad\qquad\qquad\qquad j=1,\ldots,n, \quad f(n)\equiv 0. \end{aligned}$$

PROOF. Assume that condition 1° is satisfied. Consider, for $j = 1, 2, \ldots, n$, the Krylov spaces $K_j = \operatorname{span}\{b, Ab, \ldots, A^{j-1}b\}$. Since $f(j-1) > 0$, the dimension of each K_j equals j, in particular $\{b, Ab, \ldots, A^{n-1}b\}$ is a basis. Set $B = (Ab, A^2b, \ldots, A^n b)$. Since A is annihilated by q, we obtain the identities

$$AB = BC, \tag{2.2}$$

$$b = \sum_1^n \xi_j A^j b = Bs. \tag{2.3}$$

There exists a unitary $\tilde{W}$ and an upper triangular $\tilde{R}$ such that $B = \tilde{W}\tilde{R}$. Then the condition $\|r^j\| = f(j)$, $j = 0, 1, \ldots, n-1$, implies $Bs = b = \tilde{W}\Gamma h$, where $\Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_n)$, $|\gamma_j| = 1$, $j = 1, \ldots, n$. Set $W = \tilde{W}\Gamma$, $R = \Gamma^*\tilde{R}$. Then by (2.2)

$$AWR = A\tilde{W}\tilde{R} = AB = BC = WRC,$$

whence $A = WRCR^{-1}W^*$. Also, by (2.3)

$$WRs = \tilde{W}\tilde{R}s = Bs = \tilde{W}\Gamma h = Wh,$$

so that $Rs = h$. This proves 2°.

Assume that condition 2° is satisfied. Clearly, A has the eigenvalues $\lambda_1, \ldots, \lambda_n$. For proving 1°, it is sufficient to show that for the given A, b, and any j, $j = 1, \ldots, n$, the first j column vectors $w_1, w_2, \ldots, w_j$ of the matrix W represent the unitary basis of the Krylov residual subspace $AK_j(A, b)$; cf. [2, p. 466]. We will prove this by induction. Consider $b_j = A^j b$, $j = 1, \ldots, n$. From $b = Wh = WRs$,

$$b_1 = Ab = WRCs = WRe_1 = (R)_{11} w_1.$$

Assume that $b_{j-1} = WRe_{j-1}$, $j \le n$. Then

$$b_j = Ab_{j-1} = WRCe_{j-1} = WRe_j = W(R)_{\bullet j},$$

where $(R)_{\bullet j}$ denotes the j-th column of the matrix R. □

Similarly to Theorem 2.1, one can characterize the set of all pairs $\{A, b\}$, for which GMRES(A, b) generates the prescribed convergence curve. In other words, the following theorem fixes the convergence history GMRES(A, b) but does not impose any condition on the spectrum of the matrix A.

THEOREM 2.2. *Using the notation of Theorem 2.1, the following two assertions are equivalent:*

1° Residual vectors at each step of GMRES(A, b) satisfy $\|r^k\| = f(k)$, $k = 0, 1, 2, \ldots, n-1$.

2° The matrix A is of the form $A = W\hat{R}\hat{H}W^*$ and b satisfies $b = Wh$, where W is a unitary matrix, $\hat{R}$ is a nonsingular upper triangular matrix, and

$$\hat{H} = \begin{pmatrix} 0 & \ldots & 0 & 1/\eta_n \\ 1 & & 0 & -\eta_1/\eta_n \\ & \ddots & \vdots & \vdots \\ 0 & \ldots & 1 & -\eta_{n-1}/\eta_n \end{pmatrix}.$$

PROOF. See [2]. $\square$

Denote by $\mathcal{S}_2 = \mathcal{S}_2(f)$ the set of all pairs $\{A, b\}$ described by Theorem 2.2 for a fixed nonincreasing function f defined on the points $\{0, 1, \ldots, n\}$, $f(n-1) > f(n) = 0$, and by $\mathcal{S}_1 = \mathcal{S}_1(f, \{\lambda_1, \ldots, \lambda_n\})$ the set of all pairs $\{A, b\}$ described by Theorem 2.1 for a fixed nonincreasing f, $f(n-1) > f(n) = 0$, and a fixed spectrum $\{\lambda_1, \ldots, \lambda_n\}$, all eigenvalues different from zero. Clearly $\mathcal{S}_1 \subset \mathcal{S}_2$. One may ask how the characterization given by Theorem 2.1 is related to the characterization given by Theorem 2.2.

For any $\{A, b\} \in \mathcal{S}_1$ we may write

$$A = WRCR^{-1}W^*, \qquad b = Wh,$$

for some unitary matrix W and some nonsingular upper triangular matrix R. Set $RC^{-1} \equiv Y$. Considering $Rs = h$,

$$\begin{aligned} Y = RC^{-1} &= R\begin{pmatrix} \boxed{s} & \begin{matrix} 1 & & \\ 0 & \ddots & \\ & & 1 \\ & & 0 \end{matrix} \end{pmatrix} = \begin{pmatrix} \boxed{h} & \begin{matrix} \boxed{R_{1,n-1}} \\ 0 \end{matrix} \end{pmatrix} \\ &= \begin{pmatrix} \boxed{h} & \begin{matrix} 1 & & \\ 0 & \ddots & \\ & & 1 \\ & & 0 \end{matrix} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \boxed{R_{1,n-1}} \end{pmatrix}, \end{aligned}$$

where $R_{1,n-1}$ denotes the $(n-1)$st left principal submatrix of the matrix R.

Then

$$
\begin{aligned}
RCR^{-1} &= R(RC^{-1})^{-1} = R \begin{pmatrix} 1 & 0 \\ 0 & \boxed{R_{1,n-1}} \end{pmatrix}^{-1} \left(\begin{array}{c|cccc} & 1 & & & \\ & 0 & \ddots & & \\ h & & & 1 & \\ & & & & 0 \end{array} \right)^{-1} \\
&= \hat{R}\hat{H}.
\end{aligned}
$$

where $\hat{R}$ is defined as

$$
\hat{R} = R \begin{pmatrix} 1 & 0 \\ 0 & \boxed{R_{1,n-1}^{-1}} \end{pmatrix}. \tag{2.4}
$$

Then we can write $A = W\hat{R}\hat{H}W^*$, $b = Wh$.

On the other hand, given $A = W\hat{R}\hat{H}W^*$, $b = Wh$ where W is a unitary matrix and $\hat{R}$ a nonsingular upper triangular matrix, the decomposition (2.4) always exists and is unique. Repeating the considerations above backwards we receive $A = WR\tilde{C}R^{-1}W^*$, where $\tilde{C}$ is some companion matrix. From $\{A, b\} \in \mathcal{S}_1$ it follows $\tilde{C} = C$ and, consequently, $Rs = h$.

We summarize the relation between the parametrizations of the sets $\mathcal{S}_1$ and $\mathcal{S}_2$. In both cases, one of the parameters is an arbitrary unitary matrix W. Let us stress that this matrix does not play a substantial role. It only represents a change of the basis.

PROPOSITION 2.3. *The set $\mathcal{S}_1$ is parametrized by W and by a nonsingular upper triangular matrix R satisfying the relation*

$$
Rs = h. \tag{2.5}
$$

The set $\mathcal{S}_2$ is parametrized by W and an arbitrary nonsingular upper triangular matrix $\hat{R}$. If, in addition, the spectrum of the matrix A is prescribed, then this additional condition is equivalent to

$$
RCR^{-1} = \hat{R}\hat{H}, \tag{2.6}
$$

or, equivalently, $\hat{R}$ is given by (2.4), *where R is a nonsingular upper triangular matrix satisfying* (2.5).

Clearly, by fixing the spectrum of the matrix A we decrease the number of free parameters describing $\mathcal{S}_1$ in comparison to $\mathcal{S}_2$ by n. In terms of matrices R and $\hat{R}$ this fact can be viewed in the following way. Using (2.5), the last column $(R)_{\bullet n}$ of the matrix R is given by

$$
\xi_n \, (R)_{\bullet n} = h - \begin{pmatrix} R_{1,n-1} \, (\xi_1, \dots, \xi_{n-1})^T \\ 0 \end{pmatrix} \tag{2.7}
$$

where $\xi_n \neq 0$ due to the nonsingularity of the matrix A. Consequently, any nonsingular upper triangular matrix R satisfying (2.5) is given by its $(n-1)$st left principal submatrix representing free parameters, and by the last column determined by (2.7). The corresponding nonsingular upper triangular matrix $\hat{R}$ parametrizing $\mathcal{S}_1$ as the subset of $\mathcal{S}_2$ is determined by (2.4).

Considering the matrix $G = (b, Ab, \ldots, A^{n-1}b)$ and using the identity

$$AG = GC \tag{2.8}$$

we can give another form of the parametrization of the set $\mathcal{S}_1$. Since $B = AG = GC$, it follows that $G = BC^{-1}$. Using $B = WR$ and $Y = RC^{-1}$ we have $G = WRC^{-1} = WY$, and the parametrization of the pair $\{A, b\}$ has the form

$$A = WYCY^{-1}W^*, \qquad b = Wh. \tag{2.9}$$

Recall that the matrix Y is given by

$$Y = \left(\boxed{h} \quad \begin{matrix} \boxed{R_{1,n-1}} \\ 0 \end{matrix} \right), \tag{2.10}$$

where $R_{1,n-1}$ is any $(n-1)$ by $(n-1)$ nonsingular upper triangular matrix. Inversely, given the parametrization (2.9), the matrix R from Theorem 1 has its left principal submatrix equal to $R_{1,n-1}$ and the last column $R_{\bullet n}$ determined by (2.7), which gives $YCY^{-1} = RCR^{-1}$. Summarizing, we have proved the following corollary of Theorem 1.

COROLLARY 2.4. *Using the notation of Theorem 2.1, the assertions 1° and 2° of this theorem are equivalent to*

> *3° Matrix A is of the form $A = WYCY^{-1}W^*$ and $b = Wh$, where W is a unitary matrix, Y is given by* (2.10) *and $R_{1,n-1}$ is any $(n-1)$ by $(n-1)$ nonsingular upper triangular matrix.*

Finally, using the parametrization of Theorem 2.1 we describe the particular matrix $A = (b, W_{n-1})\ C\ (b, W_{n-1})^{-1}$ constructed in [2, pp. 466–467]. After simple manipulations,

$$\begin{aligned} A &= (b, W_{n-1})\, CCC^{-1}\, (b, W_{n-1})^{-1} \\ &= (W_{n-1}, Aw_{n-1})\, C\, (W_{n-1}, Aw_{n-1})^{-1} \\ &= WRCR^{-1}W^*, \end{aligned}$$

where $R = (I_{n-1}, W^*Aw_{n-1})$. Note that here $Aw_{n-1} = (b, W_{n-1})\,(\alpha_0, \ldots, \alpha_{n-1})^T$.

3 Relation to minimal polynomial.

Let GMRES(A, b) converge to the exact solution x at the step $n-l$, $l \geq 1$. This early termination may be caused by some special properties of the eigenvalue–eigenvector structure of the matrix A, and/or by some special properties of the right hand side b.

Using the polynomial formulation of the GMRES minimization property (2.1), it is clear that GMRES(A, b) must converge to the exact solution on or before step m, where m is the degree of the minimal polynomial of A (cf. [4, 5]). Moreover, there exists a right hand side $\tilde{b}$, for which GMRES$(A, \tilde{b})$ converges to x exactly in m steps. In other words, for a given matrix $A \in C^{n,n}$ there always exist a vector $\tilde{b} \in C^n$ such that the minimal polynomial of $\tilde{b}$ is equal to the minimal polynomial of A. This statement dates back to Krylov and the proof (together with the references to the original work) can be found in [6, Chapter VII, §1, §2 (Theorem 2) and §8]. The question is closely related to the structure of invariant subspaces (the Jordan canonical form) of the matrix A.

For the sake of completeness, let us present the characterization of Krylov sequences having the maximal length. Let $\lambda_1, \dots, \lambda_{\tilde{k}}$ be the all distinct eigenvalues of A and n_j be the size of the largest Jordan block corresponding to the eigenvalue λ_j, $j = 1, \dots, \tilde{k}$. Then the minimal polynomial $q_A(\lambda)$ is given by the formula

$$g_A(\lambda) = (\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_{\tilde{k}})^{n_{\tilde{k}}}.$$

Denote the nullspace of the operator $(\lambda_j I - A)^{n_j}$ by $E(\lambda_j)$,

$$E(\lambda_j) = \mathrm{Ker}\,(\lambda_j I - A)^{n_j}, \qquad j = 1, 2, \dots, \tilde{k}.$$

Then any b can be decomposed as

$$b = t_1 + t_2 + \cdots + t_{n_{\tilde{k}}}, \qquad t_j \in E(\lambda_j).$$

Let $J = S^{-1}AS$ be the Jordan canonical form of A. The nullspace $E(\lambda_j)$ is the invariant subspace generated by the vectors of the Jordan canonical basis corresponding to λ_j; therefore the components of t_j in the directions of all the other vectors of the Jordan canonical basis are zero. For any polynomial p,

$$p(A)\, t_j \;=\; S\, p(J)\, S^{-1} t_j.$$

Consequently, $p(A)\, b \;=\; 0$ is equivalent to $p(J)\, S^{-1} t_j \;=\; 0$, $j = 1, \dots, \tilde{k}$, and the polynomial p of the minimal degree satisfying the previous condition must be of the form

$$p(\lambda) = (\lambda - \lambda_1)^{\hat{n}_1} \cdots (\lambda - \lambda_{\tilde{k}})^{\hat{n}_{\tilde{k}}},$$

where $0 \leq \hat{n}_j \leq n_j$, $j = 1, \dots, \tilde{k}$. Finally, the vector b yields the Krylov sequence of length m if and only if

$$(\lambda_j I - A)^{n_j - 1}\, t_j \neq 0\,, \quad \textit{i.e.} \quad (\lambda_j I - J)^{n_j - 1}\, S^{-1} t_j \neq 0 \tag{3.1}$$

for each j, $j = 1, \dots, \tilde{k}$. Equivalently, the vector b has for each j nonzero component in the direction of at least one last Jordan canonical vector conformed to any of the Jordan blocks largest in size corresponding to λ_j.

4 Conclusions.

We have not studied individual points of the sets $\mathcal{S}_1$ and $\mathcal{S}_2$. The question whether the described parametrizations can help in finding interesting subsets and offer some insight in the problem of the rate of convergence remains open and needs further work. The appearance of the minimal polynomial indicates that the structure of the invariant subspaces and the corresponding Jordan canonical form has an intimate relation to convergence of GMRES. Some form of extension of Theorems 2.1 and 2.2 to the general early termination case is desirable and we plan to work in this direction. We point out the role of the last Jordan canonical vectors corresponding to the Jordan blocks largest in size, which has been, to our knowledge, overlooked in favor of the eigenvectors.

Acknowledgement.

We would like to thank the referees for their comments.

REFERENCES

1. 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-Verlag, Berlin, 1994, pp. 95–118.
2. A. Greenbaum, V. Pták and Z. Strakoš, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469.
3. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
4. Y. Saad and M. Schultz, *A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comp., 7 (1986), pp. 856–869.
5. S. L. Campbell, I. C. F. Ipsen, C. T. Keley, and C. D.Meyer, *GMRES and the minimal polynomial*, BIT, 36 (1996), pp. 664–675.
6. F. R. Gantmakher, *Teorija Matric*, Nauka, Moscow, 1988 (in Russian).