SIAM J. NUMER. ANAL.
Vol. 21, No. 2, April 1984

011

NECESSARY AND SUFFICIENT CONDITIONS FOR THE EXISTENCE OF A CONJUGATE GRADIENT METHOD*

VANCE FABER† AND THOMAS MANTEUFFEL†

Abstract. We characterize the class $CG(s)$ of matrices A for which the linear system $A\mathbf{x}=\mathbf{b}$ can be solved by an s-term conjugate gradient method. We show that, except for a few anomalies, the class $CG(s)$ consists of matrices A for which conjugate gradient methods are already known. These matrices are the Hermitian matrices, $A^*=A$, and the matrices of the form $A=e^{i\theta}(dI+B)$, with $B^*=-B$.

A. Introduction. The conjugate gradient method for symmetric positive definite matrices has proved very effective, especially when coupled with preconditioning techniques. A generalization of this method to nonsymmetric matrices has long been sought. One generalization, which we call the conjugate gradient method, involves optimizing at each step over a Krylov space in some inner product norm. In this paper we characterize the class, $CG(s)$, of matrices for which this iteration can be carried out using an s-term recursion. (This answers a question of Golub reported in Signum Newsletter [7, p. 7].) We show that, except for a few anomalies, the class $CG(s)$ consists of matrices A for which conjugate gradient methods are already known. Another generalization, which we call the orthogonal residual method, involves forcing the residual to be orthogonal to a Krylov space at each step. We shall present a characterization of the class, $OR(s)$, of matrices for which this iteration can be carried out using an s-term recursion in a subsequent paper.

In §§ B–D, we define what we mean by a conjugate gradient method. Section B discusses *gradient methods*; § C shows that optimality leads to *conjugate* gradient methods, § D discusses *finite term* conjugate gradient methods.

The main theorem is:

THEOREM. *An s-term conjugate gradient method exists for the solution of* $A\mathbf{x}=\mathbf{b}$ *if and only if either*

(i) *the minimal polynomial of* A *has degree* $\leqq s$, *or*

(ii) A^* *is a polynomial of degree* $\leqq s-2$ *in* A.

Here A^* is the adjoint of A with respect to a certain inner product. The condition that A^* is a polynomial in A of some degree is equivalent to the condition that A is normal with respect to this inner product (cf. Gantmacher [5, p. 272]).

B. Gradient methods. Given the linear system $A\mathbf{x}=\mathbf{b}$ and an initial guess $\mathbf{x}_0$, let $\mathbf{r}_0=\mathbf{b}-A\mathbf{x}_0$ be the initial residual. A gradient method is defined to be a method in which

$$\mathbf{x}_{i+1}=\mathbf{x}_i+\sum_{j=0}^{i}\eta_{ij}\mathbf{r}_j,\ \mathbf{r}_j=\mathbf{b}-A\mathbf{x}_j; \tag{1}$$

that is, at each step the partial solution is incremented by some linear combination of the previous residual vectors (Rutishauser [4]). The term gradient comes from the fact that if A is symmetric, $\mathbf{r}=\mathbf{b}-A\mathbf{x}$ is the gradient of the bilinear form $Q(\mathbf{x})=\frac{1}{2}((A^{-1}\mathbf{b}-\mathbf{x}), A(A^{-1}\mathbf{b}-\mathbf{x}))$. If A is symmetric positive definite, then $Q(\mathbf{x})$ is minimized at $\mathbf{x}=A^{-1}\mathbf{b}$.

* Received by the editors July 19, 1982, and in revised form March 4, 1983. This research was performed under the auspices of the U.S. Department of Energy.

† Los Alamos National Laboratory, Los Alamos, New Mexico 87545.

If A is not symmetric positive definite, we may still consider iterations of the form in (1). Each $\mathbf{r}_j$ is easily computable, and if A is nonsingular, the solution $\mathbf{x}$ can be obtained by such an iteration. To see this, we shall show by induction that

$$\mathbf{r}_j = p_j(A)\mathbf{r}_0,$$

where $p_j(z)$ is a polynomial of degree at most j. We have

$$\mathbf{x}_1 = \mathbf{x}_0 + \eta_{00}\mathbf{r}_0.$$

Thus,

$$\mathbf{r}_1 = (I - \eta_{00}A)\mathbf{r}_0 = p_1(A)\mathbf{r}_0.$$

In general,

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \sum_{j=0}^{i} \eta_{ij}\mathbf{r}_j.$$

Thus,

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \sum_{j=0}^{i} \eta_{ij}A\mathbf{r}_j = \left[(I - \eta_{ii}A)p_i(A) - A\sum_{j=0}^{i-1} \eta_{ij}p_j(A)\right]\mathbf{r}_0 = p_{i+1}(A)\mathbf{r}_0.$$

Suppose $\eta_{ii} \neq 0$ for every i. Then $p_i(z)$ is of exact degree i and the residual vectors form a basis for the Krylov space of dimension $i+1$ generated by A and $\mathbf{r}_0$; that is,

$$V_{i+1} = \{\mathbf{r}_0, A\mathbf{r}_0, \cdots, A^i\mathbf{r}_0\} = \{\mathbf{r}_0, \mathbf{r}_1, \cdots, \mathbf{r}_i\}.$$

From (1) we see that with the proper choice of the η_{ij}'s we may make $\mathbf{x}_{i+1} - \mathbf{x}_i$ any element of V_{i+1}. Since A is of finite dimension, say N, there is some $l \leqq N$ such that $V_l = V_{l+1}$. Let k be the least such l. The vectors $\{\mathbf{r}_0, \cdots, A^k\mathbf{r}_0\}$ are linearly dependent and we may choose constants $\beta_0, \cdots, \beta_k$ such that

$$\sum_{i=0}^{k} \beta_i A^i \mathbf{r}_0 = \mathbf{0}.$$

If A is nonsingular, either $\beta_0 \neq 0$ or k may be chosen smaller. If $\beta_0 \neq 0$, then the solution to the system $A\mathbf{x} = \mathbf{b}$ is given by

$$\mathbf{x} = \mathbf{x}_0 - \frac{1}{\beta_0}\sum_{i=1}^{k} \beta_i A^{i-1}\mathbf{r}_0.$$

Since the residual vectors span the Krylov space, the proper choice of the n_{ij}'s in (1) will yield the solution. If A is singular with simple zero eigenvalues, and the system is consistent, one can also obtain a solution.

C. Optimality. The problem, of course, is choosing the η_{ij}'s properly. One way is to enforce an optimality condition. Let $\mathbf{e}_i = A^{-1}\mathbf{b} - \mathbf{x}_i = \mathbf{x} - \mathbf{x}_i$ be the error vector. Suppose we have a norm associated with an inner product:

$$\|\mathbf{x}\|^2 = [\mathbf{x}, \mathbf{x}].$$

We would like to choose $\mathbf{x}_i$ so that $\|\mathbf{e}_i\|$ is minimized over all possible iterations of the form in (1). In other words, at each step i let

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

where $\mathbf{p}_i \in V_{i+1}$. We want to choose $\alpha_i \mathbf{p}_i$ such that $\|\mathbf{e}_{i+1}\|$ is as small as possible. We have

$$\mathbf{x}_{i+1} = \mathbf{x}_0 + \sum_{j=0}^{i} \alpha_j \mathbf{p}_j;$$

thus

$$\mathbf{e}_{i+1} = \mathbf{e}_0 - \sum_{j=0}^{i} \alpha_j \mathbf{p}_j.$$

Taking inner products, we have

$$\|\mathbf{e}_{i+1}\|^2 = [\mathbf{e}_0, \mathbf{e}_0] - \left[\mathbf{e}_0, \sum_{j=0}^{i} \alpha_j \mathbf{p}_j\right] - \left[\sum_{j=0}^{i} \alpha_j \mathbf{p}_j, \mathbf{e}_0\right] + \left[\sum_{j=0}^{i} \alpha_j \mathbf{p}_j, \sum_{j=0}^{i} \alpha_j \mathbf{p}_j\right].$$

If we let $\alpha_j = x_j + iy_j$ and take the derivative with respect to each x_j, y_j and set it equal to zero, we have the linear system

$$\begin{bmatrix} [\mathbf{p}_0, \mathbf{p}_0] & [\mathbf{p}_1, \mathbf{p}_0] & \cdots & [\mathbf{p}_i, \mathbf{p}_0] \\ [\mathbf{p}_0, \mathbf{p}_1] & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots \\ [\mathbf{p}_0, \mathbf{p}_{i-1}] & \cdot & \cdot & [\mathbf{p}_i, \mathbf{p}_{i-1}] \\ [\mathbf{p}_0, \mathbf{p}_i] & \cdot & \cdot & [\mathbf{p}_i, \mathbf{p}_i] \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \cdot \\ \vdots \\ \alpha_{i-1} \\ \alpha_i \end{bmatrix} = \begin{bmatrix} [\mathbf{e}_0, \mathbf{p}_0] \\ \cdot \\ \vdots \\ [\mathbf{e}_0, \mathbf{p}_{i-1}] \\ [\mathbf{e}_0, \mathbf{p}_i] \end{bmatrix}.$$

Since this is true for every i, in particular for $i-1$, we have

$$\alpha_i \begin{bmatrix} [\mathbf{p}_0, \mathbf{p}_i] \\ \vdots \\ [\mathbf{p}_{i-1}, \mathbf{p}_i] \end{bmatrix} = 0, \qquad \alpha_i = \frac{[\mathbf{e}_0, \mathbf{p}_i]}{[\mathbf{p}_i, \mathbf{p}_i]}.$$

If $\alpha_i \neq 0$, then $[\mathbf{p}_j, \mathbf{p}_i] = 0$, $j < i$. If $\alpha_i = 0$, then $\|\mathbf{e}_{i+1}\| = \|\mathbf{e}_i\|$. If, on the one hand, $V_i = V_{i+1}$, we have seen above that V_i yields the solution, which implies that $\mathbf{e}_{i+1} = 0$. If, on the other hand, $V_i \neq V_{i+1}$, there is a unique (up-to-scale) vector $\mathbf{p}_i \in V_{i+1}$ such that $[\mathbf{p}_i, \mathbf{z}] = 0$ for every $\mathbf{z} \in V_i$.

To summarize, if each $\mathbf{x}_{i+1} = \mathbf{x}_0 + \mathbf{y}_i$ is chosen such that $\|\mathbf{e}_{i+1}\|$ is optimal over all possible $\mathbf{y}_i \in V_{i+1}$, then $\mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$, where $\mathbf{p}_i$ is the unique (up-to-scale) vector $\mathbf{p}_i \in V_{i+1}$ such that $[\mathbf{p}_i, \mathbf{z}] = 0$ for every $\mathbf{z} \in V_i$, and $\alpha_i = [\mathbf{e}_0, \mathbf{p}_i]/[\mathbf{p}_i, \mathbf{p}_i]$. Notice that

$$(2) \qquad \alpha_i = \frac{[\mathbf{e}_i + \sum_{j=0}^{i-1} \alpha_j \mathbf{p}_j, \mathbf{p}_i]}{[\mathbf{p}_i, \mathbf{p}_i]} = \frac{[\mathbf{e}_i, \mathbf{p}_i]}{[\mathbf{p}_i, \mathbf{p}_i]}.$$

Since the $\mathbf{p}_i$'s are "conjugate" with respect to the inner product, we say this is a conjugate gradient method.

Notice that (2) involves the usually unknown quantity $\mathbf{e}_i$. In order for the algorithm to be viable, the inner product must be chosen so that α_i is computable.

D. Recursive calculation of $\mathbf{p}_i$'s. From above, we know that since $\mathbf{p}_i \in V_{i+1}$ and $\mathbf{p}_i$ is orthogonal to V_i, that $\mathbf{p}_i = p_i(A)\mathbf{p}_0$, where $p_i(z)$ is a polynomial of exact degree i. Thus, $A\mathbf{p}_i$ is of exact degree $i+1$ and so $V_{i+2} = \{A\mathbf{p}_i, \mathbf{p}_0, \cdots, \mathbf{p}_i\}$. The vector $\mathbf{p}_{i+1}$ can be computed as

$$\mathbf{p}_{i+1} = A\mathbf{p}_i - \sum_{j=0}^{i} \beta_{ij} \mathbf{p}_j.$$

The β_{ij}'s are uniquely determined by the conditions

$$[\mathbf{p}_{i+1}, \mathbf{p}_j] = 0, \qquad j = 0, \cdots, i,$$

which yields

$$\beta_{ij} = \frac{[A\mathbf{p}_i, \mathbf{p}_j]}{[\mathbf{p}_j, \mathbf{p}_j]}, \qquad j = 0, \cdots, i.$$

Now if A is self-adjoint with respect to this inner product, we have

$$\beta_{ij} = \frac{[\mathbf{p}_i, A^*\mathbf{p}_j]}{[\mathbf{p}_j, \mathbf{p}_j]} = \frac{[\mathbf{p}_i, A\mathbf{p}_j]}{[\mathbf{p}_j, \mathbf{p}_j]}.$$

Since $\mathbf{p}_j \in V_{j+1}$, then $A\mathbf{p}_j \in V_{j+2}$. Since $\mathbf{p}_i$ is orthogonal to V_i, we have $\beta_{ij} = 0$ for $j < i-1$. Thus

$$\mathbf{p}_{i+1} = A\mathbf{p}_i - \beta_{ii}\mathbf{p}_i - \beta_{ii-1}\mathbf{p}_{i-1}. \tag{3}$$

This is referred to as a three-term recursion for computing $\mathbf{p}_{i+1}$. Similarly, we may consider an s-term recursion

$$\mathbf{p}_{i+1} = A\mathbf{p}_i - \sum_{j=i-s+2}^{i} \beta_{ij}\mathbf{p}_j, \tag{4}$$

and ask for what class of matrices an s-term recursion exists for every $\mathbf{p}_0$.

We must clarify a fine point here. We use the following notation. Let $d(\mathbf{p})$ be the degree of the minimal polynomial of $\mathbf{p}$ with respect to A. This is the dimension of the Krylov space generated by $\mathbf{p}$ and A. Let $d(A)$ be the degree of the minimal polynomial of A. Since V_{i+1} is of dimension $i+1$, if $d(\mathbf{p}_0) = i+1$ then it is not necessary to compute $\mathbf{p}_{i+1}$ because V_{i+1} contains the solution. Thus, β_{ij} need not be computed. With this in mind, we make the following definition:

An s-term conjugate gradient iteration exists for the matrix A if for every $\mathbf{p}_0$, $[A\mathbf{p}_i, \mathbf{p}_j] = 0$ *for every* i, j *such that* $j + s - 1 \leqq i \leqq d(\mathbf{p}_0) - 2$. *We will denote this as* $A \in CG(s)$.

The most commonly used form of the three-term conjugate gradient method applied to symmetric positive definite A has general steps

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i\mathbf{p}_i, \quad \mathbf{r}_{i+1} = \mathbf{r}_i - \alpha_i A\mathbf{p}_i, \quad \mathbf{p}_{i+1} = \mathbf{r}_{i+1} + \beta_i\mathbf{p}_i.$$

Notice that $\mathbf{r}_i = \mathbf{p}_i - \beta_{i-1}\mathbf{p}_{i-1}$ so that

$$\mathbf{r}_{i+1} = \mathbf{p}_i - \beta_{i-1}\mathbf{p}_{i-1} - \alpha_i A\mathbf{p}_i.$$

Substituting for $\mathbf{r}_{i+1}$ above yields

$$\mathbf{p}_{i+1} = -\alpha_i A\mathbf{p}_i + (1+\beta_i)\mathbf{p}_i - \beta_{i-1}\mathbf{p}_{i-1},$$

which is the same as (3) up to a scale factor.

The generalized conjugate gradient methods of Concus, Golub, and O'Leary [2] make use of a symmetric positive definite preconditioning matrix M. Their method is equivalent to the conjugate gradient method as we have defined it applied to a system involving $\hat{A} = M^{-1/2}AM^{-1/2}$. In the next section, we characterize $CG(s)$.

E. Characterization. The inner product $[\cdot,\cdot]$ determines the class of orthogonal bases on the space. The characterization that we present is with respect to this inner product. Consider A^*. A nonsingular matrix C exists such that

$$[\mathbf{x}, \mathbf{y}] = \langle C\mathbf{x}, C\mathbf{y}\rangle,$$

where $\langle\cdot,\cdot\rangle$ is the standard inner product,

$$\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{j=1}^{N} x_j \bar{y}_j.$$

Now

$$\begin{aligned}[A\mathbf{x}, \mathbf{y}] &= \langle CAC^{-1}C\mathbf{x}, C\mathbf{y}\rangle = \langle C\mathbf{x}, \bar{C}^{-T}\bar{A}^T\bar{C}^TC\mathbf{y}\rangle \\ &= \langle C\mathbf{x}, C(\bar{C}^TC)^{-1}\bar{A}^T\bar{C}^TC\mathbf{y}\rangle = [\mathbf{x}, A^*\mathbf{y}].\end{aligned}$$

Thus, $A^* = (\bar{C}^TC)^{-1}\bar{A}^T(\bar{C}^TC)$. Since $\bar{C}^T C$ is Hermitian positive definite, there exists a Hermitian positive definite matrix B such that $B^2 = \bar{C}^TC$.

Let us change the basis so that

$$\hat{A} = BAB^{-1}, \quad \hat{\mathbf{x}} = B\mathbf{x}, \quad \text{and} \quad \hat{\mathbf{y}} = B\mathbf{y}.$$

Then

$$[A\mathbf{x}, \mathbf{y}] = \langle \hat{A}\hat{\mathbf{x}}, \hat{\mathbf{y}}\rangle.$$

The adjoint of $\hat{A}$ with respect to $\langle\cdot,\cdot\rangle$ is $\hat{A}^* = \bar{A}^T$. In this new basis, the original problem becomes $\hat{A}\hat{\mathbf{x}} = \hat{\mathbf{b}}$. A conjugate gradient iteration on this new system that optimizes with respect to the standard inner product will produce iterates corresponding to those produced using $[\cdot,\cdot]$. However, it may be computationally feasible to optimize with respect to $[\cdot,\cdot]$ (for example when C is sparse), but impractical to compute the change of basis matrix B, much less B^{-1}.

For convenience we will use only $\langle\cdot,\cdot\rangle$ and the definition $A^* = \bar{A}^T$ for the remainder of this paper with the understanding that the change of basis has occurred.

It is easy to see that if $d(A) \leqq s$, then $A \in CG(s)$. The condition in the definition is vacuously true because for every $\mathbf{p}_0$, $s-2 \geqq d(\mathbf{p}_0)-2$. The iteration converges in s or less steps.

Another sufficient condition is expressed by Lemma 1.

LEMMA 1. *If A is such that for every* $\mathbf{p}$,

$$A^*\mathbf{p} \in \{\mathbf{p}, A\mathbf{p}, \cdots, A^{s-2}\mathbf{p}\},$$

then $A \in CG(s)$.

Proof. Consider

$$\langle A\mathbf{p}_i, \mathbf{p}_j\rangle = \langle \mathbf{p}_i, A^*\mathbf{p}_j\rangle.$$

Now $A^*\mathbf{p}_j = q(A)\mathbf{p}_j$ for some polynomial $q(z)$ of degree $s-2$ or less. (Note that the polynomial may depend upon $\mathbf{p}_0$.) If $j+s-2<i$, then $q(A)\mathbf{p}_j \in V_{j+s-1} \subseteq V_i$. Since $\mathbf{p}_i$ is orthogonal to V_i, then $\langle A\mathbf{p}_i, \mathbf{p}_j\rangle = 0$.

The next lemma characterizes those matrices that satisfy the hypothesis of Lemma 1. First, we need some results about normal matrices.

The matrix A is normal if and only if $A^*A = AA^*$ and if and only if A has a complete set of orthonormal eigenvectors. Let us write $A = U\Lambda U^*$, where U is unitary and $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_N)$. Let $q(z)$ be a polynomial such that

$$q(\lambda_i) = \bar{\lambda}_i \qquad i = 1, \cdots, N.$$

Then,

$$q(A) = Uq(\Lambda)U^* = U\bar{\Lambda}U^* = A^*.$$

Thus, A is normal if and only if $A^* = q(A)$ for some polynomial $q(z)$. We will denote the degree of the polynomial of smallest degree that satisfies $A^* = q(A)$ by $n(A)$.

LEMMA 2. *A is such that for every* $\mathbf{p}$

$$A^*\mathbf{p} \in \{\mathbf{p}, A\mathbf{p}, \cdots, A^{s-2}\mathbf{p}\}$$

if and only if A is normal and $n(A) \leqq s-2$.

Proof. If A is normal with $n(A) \leqq s-2$, then

$$A^*\mathbf{p} = q(A)\mathbf{p} \in \{\mathbf{p}, A\mathbf{p}, \cdots, A^{s-2}\mathbf{p}\}.$$

Now assume $A^*\mathbf{p} \in \{\mathbf{p}, A\mathbf{p}, \cdots, A^{s-2}\mathbf{p}\}$ for every $\mathbf{p}$. Let $\mathbf{v}_i$ be an eigenvector of A such that $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$. Then $A^*\mathbf{v}_i \in \{\mathbf{v}_i, A\mathbf{v}_i, \cdots, A^{s-2}\mathbf{v}_i\} = \{\mathbf{v}_i\}$, so $A^*\mathbf{v}_i = \mu\mathbf{v}_i$ for some μ. Now

$$\mu\langle\mathbf{v}_i, \mathbf{v}_i\rangle = \langle\mu\mathbf{v}_i, \mathbf{v}_i\rangle = \langle A^*\mathbf{v}_i, \mathbf{v}_i\rangle = \langle\mathbf{v}_i, A\mathbf{v}_i\rangle = \bar{\lambda}_i\langle\mathbf{v}_i, \mathbf{v}_i\rangle.$$

Thus, $\mu = \bar{\lambda}_i$. Now suppose A has a nonlinear elementary divisor associated with λ_i. Then, there is a vector $\mathbf{v}_j$ such that $(A - \lambda_i I)\mathbf{v}_j = \mathbf{v}_i$. Now

$$\langle A\mathbf{v}_j, \mathbf{v}_i\rangle = \lambda_i\langle\mathbf{v}_j, \mathbf{v}_i\rangle + \langle\mathbf{v}_i, \mathbf{v}_i\rangle,$$

$$\langle A\mathbf{v}_j, \mathbf{v}_i\rangle = \langle\mathbf{v}_j, A^*\mathbf{v}_i\rangle = \lambda_i\langle\mathbf{v}_j, \mathbf{v}_i\rangle.$$

Thus $\langle\mathbf{v}_i, \mathbf{v}_i\rangle = 0$, which is a contradiction. We may conclude that A has a complete set of eigenvectors. Now suppose $\lambda_i \neq \lambda_j$. Then,

$$\lambda_i\langle\mathbf{v}_i, \mathbf{v}_j\rangle = \langle\lambda_i\mathbf{v}_i, \mathbf{v}_j\rangle = \langle A\mathbf{v}_i, \mathbf{v}_j\rangle = \langle\mathbf{v}_i, A^*\mathbf{v}_j\rangle = \lambda_j\langle\mathbf{v}_i, \mathbf{v}_j\rangle,$$

and we have $\langle\mathbf{v}_i, \mathbf{v}_j\rangle = 0$. Since A has a complete set of orthonormal eigenvectors, A is normal.

Since A has exactly $d(A)$ distinct eigenvalues, there is an interpolating polynomial, $q(z)$, of degree at most $d(A)-1$ such that $q(\lambda_i) = \bar{\lambda}_i$, $i = 1, \cdots, d(A)$. Thus $n(A) \leqq d(A)-1$. If $d(A) = s-1$, then $n(A) \leqq s-2$.

Suppose $d(A) \geqq s$. Choose $\mathbf{p}$ with $d = d(\mathbf{p}) = d(A)$. The vectors $\{\mathbf{p}, A\mathbf{p}, \cdots, A^{d-1}\mathbf{p}\}$ are linearly independent. By hypothesis

$$A^*\mathbf{p} = q(A)\mathbf{p} \in \{\mathbf{p}, A\mathbf{p}, \cdots, A^{s-2}\mathbf{p}\}.$$

Thus, $q(x)$ must have degree $\leqq s-2$.

Lemmas 3 and 4 characterize those matrices for which $n(A)$ is small.

LEMMA 3. *If A is normal and $n(A) \leqq 1$, then $d(A) = 1$, or $A^* = A$, or*

$$A = e^{i\theta}\left(\frac{r}{2}I + B\right),$$

where r is real and $B = -B^*$.

Proof. If A has all real eigenvalues, then $A^* = A$. Suppose A has a complex eigenvalue, say λ. Then, the linear polynomial $q(z) = az - b$ must satisfy

$$a\lambda - b = \bar{\lambda} \quad \text{and} \quad \bar{a}\bar{\lambda} - \bar{b} = \lambda,$$

which yields $\bar{a}(a\lambda - b) - \bar{b} = \lambda$, or $(\bar{a}a - 1)\lambda - (\bar{a}b + \bar{b}) = 0$.

In general, there is only one root, which means that $d(A) = 1$. There is more than one root only if $a = -b/\bar{b}$. Let $b = re^{-i\theta}$; then

$$q(z) = -e^{-2i\theta}z - re^{-i\theta} = -e^{i\theta}(ze^{-i\theta} - r).$$

If $q(z)=\bar{z}$, then

$$-(ze^{-i\theta}-r)=\bar{z}e^{i\theta}=(\overline{ze^{-i\theta}}),$$

which yields

$$r=ze^{-i\theta}+(\overline{ze^{-i\theta}}).$$

Thus, if λ is an eigenvalue of A, the real part of $\lambda\, e^{-i\theta}$ is $r/2$. This implies that

$$B=\left(e^{-i\theta}A-\frac{r}{2}I\right)$$

has only pure complex eigenvalues. Since A is normal, B is skew symmetric.

LEMMA 4. *If A is normal, then $n(A)\leqq d(A)-1$. If, in addition, $n(A)>1$, then $d(A)\leqq n(A)^2$.*

Proof. Since A is normal, it has exactly $d(A)$ distinct eigenvalues. Using Lagrange interpolating polynomials, we can construct a polynomial of degree $d(A)-1$ or less such that

$$q(\lambda_i)=\bar{\lambda}_i \qquad i=1,\cdots,d(A).$$

Thus $n(A)\leqq d(A)-1$.

Now suppose $n(A)>1$. How many distinct complex numbers satisfy $q(\lambda)=\bar{\lambda}$? Notice that

$$\bar{q}(\bar{\lambda})=\lambda \quad \text{or} \quad \bar{q}(q(\lambda))=\lambda.$$

Since q is not linear, then $\bar{q}(q(\lambda))-\lambda=0$ has at most $n(A)^2$ roots. Thus $d(A)\leqq n(A)^2$.

We have shown that sufficient conditions for $A\in CG(s)$ are $d(A)\leqq s$ or A is normal and $n(A)\leqq s-2$. We shall now show that these are necessary conditions as well. First we establish some results that will facilitate the proof.

Define the set:

$$D=\{\mathbf{p}: d(\mathbf{p})<d(A)\}.$$

This set is a union of subspaces of dimension at most $N-1$. Each subspace is generated by a proper divisor of the minimal polynomial of A. Thus there are a finite number of these subspaces. The set D is therefore a set of measure zero under the topology imposed by the inner product.

Consider the calculation of the orthogonal basis of the Krylov space described in § D. Given $\mathbf{p}$, let

$$\begin{aligned}
&\mathbf{p}_0=\mathbf{p},\\
&\mathbf{p}_1=A\mathbf{p}_0-\beta_{00}\mathbf{p}_0,\\
&\quad\vdots\\
&\mathbf{p}_{i+1}=A\mathbf{p}_i-\sum_{j=0}^{i}\beta_{ij}\mathbf{p}_j, \quad \text{where } \beta_{ij}=\frac{\langle A\mathbf{p}_i,\mathbf{p}_j\rangle}{\langle \mathbf{p}_j,\mathbf{p}_j\rangle}.
\end{aligned}$$

If $d(\mathbf{p})>i$, then $\mathbf{p}_i$ can be considered as a function of $\mathbf{p}$. Let us extend this definition to all $\mathbf{p}$ by setting $\mathbf{p}_i=\mathbf{0}$ for $\mathbf{p}$ such that $d(\mathbf{p})\leqq i$. We now show that for all $i\leqq d(A)$, $\mathbf{p}_i$ is a continuous function of $\mathbf{p}$.

LEMMA 5. *If $\mathbf{p}_i$ is defined as above, then for all $i\leqq d(A)$, $\mathbf{p}_i$ is a continuous function of $\mathbf{p}$ and $\|\mathbf{p}_i\|\leqq\|A\|\cdot\|\mathbf{p}_{i-1}\|$.*

Proof. The proof proceeds by induction. Clearly, $\mathbf{p}_0 = \mathbf{p}$ is continuous. Now consider $\mathbf{p}_1 = A\mathbf{p}_0 - \beta_{00}\mathbf{p}_0$. For $\mathbf{p}_0 \neq \mathbf{0}$, we have

$$|\beta_{00}| = \left|\frac{\langle A\mathbf{p}_0, \mathbf{p}_0\rangle}{\langle \mathbf{p}_0, \mathbf{p}_0\rangle}\right| \leqq \|A\|.$$

Thus, for $\mathbf{p}_0 \neq \mathbf{0}$, β_{00} is continuous and bounded. Since the space upon which $\mathbf{p}_0 = \mathbf{0}$ is closed and of smaller dimension than the entire space, we can continuously extend the product $\beta_{00}\mathbf{p}_0$ to the entire space. Finally, it is clear that

$$\|\mathbf{p}_1\|^2 = \|A\mathbf{p}_0\|^2 - \langle A\mathbf{p}_0, \mathbf{p}_0\rangle^2 \leqq \|A\|^2\|\mathbf{p}_0\|^2.$$

Now assume that for $j \leqq i < d(A)$, $\mathbf{p}_j$ is a continuous function of $\mathbf{p}$ and $\|\mathbf{p}_j\| \leqq \|A\|\,\|\mathbf{p}_{j-1}\|$. We have

$$\mathbf{p}_{i+1} = A\mathbf{p}_i - \sum_{j=0}^{i} \beta_{ij}\mathbf{p}_j.$$

Consider

$$\beta_{ij} = \frac{\langle A\mathbf{p}_i, \mathbf{p}_j\rangle}{\langle \mathbf{p}_j, \mathbf{p}_j\rangle}.$$

For $\mathbf{p}_j \neq \mathbf{0}$,

$$|\beta_{ij}| \leqq \frac{\|A\|\,\|\mathbf{p}_i\|}{\|\mathbf{p}_j\|} \leqq \|A\|^{i-j}.$$

For $\mathbf{p}_j \neq \mathbf{0}$, β_{ij} is a continuous and bounded function of $\mathbf{p}_j$ and thus a continuous and bounded function of $\mathbf{p}$. Since $j < d(A)$, the set upon which $\mathbf{p}_j = \mathbf{0}$ is closed and of smaller dimension than the entire space. We can continuously extend the product $\beta_{ij}\mathbf{p}_j$ to the entire space. Since $\mathbf{p}_{i+1}$ is a sum of continuous functions, it is continuous. Now

$$\|\mathbf{p}_{i+1}\|^2 = \|A\mathbf{p}_i\|^2 - \sum_{j=0}^{i} \langle A\mathbf{p}_i, \mathbf{p}_j\rangle^2 \leqq \|A\|^2\|\mathbf{p}_i\|^2.$$

This completes the proof.

We now prove the main result.

THEOREM. *$A \in CG(s)$ if and only if $d(A) \leqq s$ or A is normal and $n(A) \leqq s-2$.*

Proof. Sufficiency has been established above. We now assume that $A \in CG(s)$ and $d(A) > s$. Let $\mathbf{p}$ be any vector such that $d(\mathbf{p}) > s$. From the definition we know that

$$\langle A\mathbf{p}_i, \mathbf{p}\rangle = \langle \mathbf{p}_i, A^*\mathbf{p}_0\rangle = 0, \qquad s-1 \leqq i \leqq d(\mathbf{p}) - 2.$$

In particular, $\langle \mathbf{p}_{s-1}, A^*\mathbf{p}\rangle = 0$. Let $F(\mathbf{p}) = \langle \mathbf{p}_{s-1}, A^*\mathbf{p}\rangle$ be considered as a function of $\mathbf{p}$. Since $\mathbf{p}_{s-1}$ is a continuous function of $\mathbf{p}$, then $F(\mathbf{p})$ is a continuous function of $\mathbf{p}$. We have $F(\mathbf{p}) = 0$ for every $\mathbf{p}$ such that $d(\mathbf{p}) > s$. Since $d(A) > s$, the set upon which $d(\mathbf{p}) \leqq s$ is closed and of smaller dimension. Therefore, $F(\mathbf{p}) = 0$ for all $\mathbf{p}$. This is trivially true for $d(\mathbf{p}) < s$ since $\mathbf{p}_{s-1} = \mathbf{0}$ for these $\mathbf{p}$. However, if $d(\mathbf{p}) = s$, then $\mathbf{p}_{s-1} \neq \mathbf{0}$.

Let $\mathbf{p}$ be such that $d(\mathbf{p}) = s$ and let

$$V_s = \{\mathbf{p}, A\mathbf{p}, \cdots, A^{s-1}\mathbf{p}\}$$

be the invariant subspace generated by $\mathbf{p}$. Let $\tilde{A}$ be the restriction of A to V_s. If Q is an orthogonal projection onto V_s, then

$$\tilde{A} = AQ \quad \text{and} \quad \tilde{A}^* = QA^*.$$

We now show that $\tilde{A}$ is normal on V_s. We have $d(\tilde{A})=s$. Let $\mathbf{p}$ be any generator of V_s; that is, let $\mathbf{p}\in V_s$ be such that $d(\mathbf{p})=s$. We know from above that

$$F(\mathbf{p})=\langle \mathbf{p}_{s-1}, A^*\mathbf{p}\rangle=\langle \mathbf{p}_{s-1}, QA^*\mathbf{p}\rangle=\langle \mathbf{p}_{s-1}, \tilde{A}^*\mathbf{p}\rangle=0.$$

Since $\tilde{A}^*\mathbf{p}\in V_s$, we must have

$$\tilde{A}^*\mathbf{p}\in\{\mathbf{p}_0,\cdots,\mathbf{p}_{s-2}\}=\{\mathbf{p}, A\mathbf{p},\cdots, A^{s-2}\mathbf{p}\},$$

for every $\mathbf{p}\in V_s$ such that $d(\mathbf{p})=s$. Now consider

$$W(\mathbf{p})=\mathbf{p}\wedge A\mathbf{p}\wedge\cdots\wedge A^{s-2}\mathbf{p}\wedge\tilde{A}^*\mathbf{p},$$

where $\wedge$ is the wedge product (cf. Mostow, Sampson, and Meyer [6, pp. 553–560]). Since the wedge product is a multilinear function from one vector space into another, $W(\mathbf{p})$ is a continuous function of $\mathbf{p}$.

For every $\mathbf{p}\in V_s$ such that $d(\mathbf{p})=s$, $W(\mathbf{p})=\mathbf{0}$ because the vectors are linearly dependent. The set $\{\mathbf{p}\in V_s: d(\mathbf{p})<s\}$ is a closed subset of smaller dimension, therefore $W(\mathbf{p})=\mathbf{0}$ for every $\mathbf{p}\in V_s$. This is trivially true for $\mathbf{p}$ with $d(\mathbf{p})<s-1$ since the first $s-1$ vectors are linearly dependent. However, let $\mathbf{p}\in V_s$ have $d(\mathbf{p})=s-1$. Then, the first $s-1$ vectors are independent. This implies that if $d(\mathbf{p})=s-1$, then

$$\tilde{A}^*\mathbf{p}\in\{\mathbf{p},\cdots,A^{s-2}\mathbf{p}\}.$$

Now let $V_{s-1}^{(1)}\subseteq V_s$ be an invariant subspace of dimension $s-1$. Since $d(\tilde{A})=s$, $V_{s-1}^{(1)}$ must be generated by a $\mathbf{p}$ with $d(\mathbf{p})=s-1$. This implies that $V_{s-1}^{(1)}$ has a basis, say $\mathbf{b}_1,\cdots,\mathbf{b}_{s-1}$, such that $d(\mathbf{b}_i)=s-1$, $i=1,\cdots,s-1$. From above we know that

$$\tilde{A}^*\mathbf{b}_i\in V_{s-1}^{(1)},\qquad i=1,\cdots,s-1.$$

Thus, $V_{s-1}^{(1)}$ is an invariant subspace of $\tilde{A}^*$ as well.

Let $\mathbf{q}_1\in V_s$ be the unique vector $\mathbf{q}_1$ orthogonal to $V_{s-1}^{(1)}$. We have

$$\langle A\mathbf{q}_1,\mathbf{y}\rangle=\langle \mathbf{q}_1,\tilde{A}^*\mathbf{y}\rangle=0$$

for every $\mathbf{y}\in V_{s-1}^{(1)}$. Therefore $A\mathbf{q}_1=\lambda_1\mathbf{q}_1$ for some λ_1. Now let $V_{s-1}^{(2)}\subseteq V_s$ be any invariant subspace of dimension $s-1$ such that $\mathbf{q}_1\in V_{s-1}^{(2)}$. Let $\mathbf{q}_2\in V_s$ be the unique vector $\mathbf{q}_2$ orthogonal to $V_{s-1}^{(2)}$. Using the same argument as above we have $A\mathbf{q}_2=\lambda_2\mathbf{q}_2$ for some λ_2 and $\langle\mathbf{q}_1,\mathbf{q}_2\rangle=0$.

If we continue in this fashion, we see that V_s has a complete set of orthonormal eigenvectors of A. We now show that A has a complete set of orthonormal eigenvectors. Suppose there exists $\mathbf{v}$ such that for some λ_i

$$(A-\lambda_i I)\mathbf{v}\neq\mathbf{0},\qquad (A-\lambda_i I)^2\mathbf{v}=\mathbf{0}.$$

Let $\mathbf{v}$ and $(A-\lambda_i I)\mathbf{v}$ be included in an invariant subspace, V, with $s-2$ eigenvectors associated with distinct eigenvectors. We know that there are at least s such eigenvectors from the argument above. V is an invariant subspace of dimension s. Let $\tilde{A}$ be the restriction of A to V. We have $d(\tilde{A})=s$. Repeating the argument above, we see that V contains s orthogonal eigenvectors of A, which is a contradiction.

Similarly if we assume that $\mathbf{q}_i$ and $\mathbf{q}_j$ are eigenvectors of A associated with distinct eigenvalues, we can include them in an invariant subspace of dimension s upon which $d(\tilde{A})=s$. The above argument will yield $\mathbf{q}_i$ orthogonal to $\mathbf{q}_j$.

We have established that A is normal. This implies that there is a polynomial $q(z)$ such that $A^*=q(A)$. From Lemma 4, $n(A)\leqq d(A)-1$. We now show that in fact $n(A)\leqq s-2$. From the definition of $CG(s)$ above we know that if $d(\mathbf{p})=d(A)$ then

$$F_i(\mathbf{p})=\langle \mathbf{p}_i, A^*\mathbf{p}\rangle=0,\ s-1\leqq i\leqq d(A)-2. \tag{5}$$

Now $F_i(\mathbf{p})$ is a continuous function of $\mathbf{p}$, and so $F_i(\mathbf{p})=0$, $s-2\leqq i\leqq d(A)-2$ for all $\mathbf{p}$. If $d(\mathbf{p})<d(A)$, then from (5)

$$A^*\mathbf{p}\in\{\mathbf{p}_0,\mathbf{p}_1,\cdots,\mathbf{p}_{s-2}\}=\{\mathbf{p},\cdots,A^{s-2}\mathbf{p}\}.$$

Now let $d(\mathbf{p})=d(A)-1$. The polynomial that annihilates $\mathbf{p}$ is the product of all but one factor. Let $d=d(A)$,

$$p_k(z)=\prod_{i\neq k}(z-\lambda_i)=z^{d-1}-\left(\sum_{j\neq k}\lambda_i\right)z^{d-1}+\cdots+\cdots,$$

and let

$$q(z)=\gamma_{d-1}z^{d-1}+\gamma_{d-2}z^{d-2}+\cdots+\cdots.$$

We know that if $d(\mathbf{p})=d(A)-1$, then the vectors $\{\mathbf{p},\cdots,A^{d-2}\mathbf{p}\}$ are linearly independent. Now $A^*\mathbf{p}=q(A)\mathbf{p}=\gamma_{d-1}A^{d-1}\mathbf{p}+\gamma_{d-2}A^{d-2}\mathbf{p}+\cdots$. Because $p_k(A)\mathbf{p}=\mathbf{0}$, we know

$$A^{d-1}\mathbf{p}=\left(\sum_{j\neq k}\lambda_i\right)A^{d-2}\mathbf{p}+\cdots+\cdots.$$

Thus,

$$A^*\mathbf{p}=\left(\gamma_{d-1}\left(\sum_{i\neq k}\lambda_i\right)+\gamma_{d-2}\right)A^{d-2}\mathbf{p}+\cdots+\cdots,$$

but $A^*\mathbf{p}\in\{\mathbf{p},\cdots,A^{s-2}\mathbf{p}\}$ implies

$$\left(\gamma_{d-1}\left(\sum_{i\neq k}\lambda_i\right)+\gamma_{d-2}\right)=0.$$

Since this is true for every k, we must have $\gamma_{d-1}=\gamma_{d-2}=0$. If $\gamma_{d-1}=0$, then we must have $\gamma_i=0$, $i>s-2$ since $\{\mathbf{p},\cdots,A^{d-2}\mathbf{p}\}$ are linearly independent and $A^*\mathbf{p}\in\{\mathbf{p},\cdots,A^{s-2}\mathbf{p}\}$. Thus, $q(z)$ is of degree at most $s-2$.

F. Remarks and conclusions. These results depend upon the choice of an inner product, which in turn implies a change of basis. The main theorem implies that there exists an inner product for which $A\in CG(3)$ if and only if $d(A)\leqq 3$ or if A has a complete set of eigenvectors and eigenvalues lying on some straight line in the complex plane. However, finding such an inner product may be very difficult.

One might consider gradient iterations that are optimal at each step in a norm that is not associated with an inner product. As an example, consider $\mathbf{r}_0$ such that $d(\mathbf{r}_0)=d$. Let $V_d=\{\mathbf{r}_0,A\mathbf{r}_0,\cdots,A^{d-1}\mathbf{r}_0\}$ and let P be the set of all polynomials of degree less than or equal to $d-1$. For every $\mathbf{p}\in V_d$ we have $\mathbf{p}=p(A)\mathbf{r}_0$ for some $p(z)\in P$. Let

$$\|\mathbf{p}\|=\max_{\lambda_i}|p(\lambda_i)|,$$

where the maximum is taken over all eigenvalues of A. In general, this norm is not associated with an inner product.

Elman [3] has demonstrated a finite term iteration for matrices with positive definite symmetric part that is convergent but suboptimal at each step. It may be possible to find such an iteration for general matrices as well.

Our results depend upon a particular form of finite term recursion. A more general form might yield a larger class of matrices for which a conjugate gradient iteration exists. We have considered the following generalization. Let

$$Q_i = (\mathbf{q}_i^{(1)}, \cdots, \mathbf{q}_i^{(t)}),$$

where Q_i is an $n \times t$ matrix and the orthogonal vectors we seek are $\mathbf{p}_i = \mathbf{q}_i^{(1)}$. Let $\mathbf{q}_i^{(j)} \in V_{i+1}$ and let Q_{i+1} be given by

$$Q_{i+1} = Q_i R_i + A Q_i S_i,$$

where R_i and S_i are $t \times t$ matrices. Clearly, this generalization includes the recursion used in (4). It also includes many other finite term recursions. We conjecture, however, that any such recursion is of the type in (4) for some $s \geqq t$.

REFERENCES

[1] P. Concus and G. H. Golub, *A generalized conjugate gradient method for nonsymmetric systems of linear equations*, in Lecture Notes in Economics and Mathematical Systems, 13A, R. Glowinski and J. L. Lions, eds., Springer-Verlag, Berlin, 1976, pp. 56–65.

[2] P. Concus, G. H. Golub and D. P. O'Leary, *A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations*, in Sparse Matrix Computations, James R. Bunch and Donald J. Rose, eds., Academic Press, New York, 1976, pp. 309–332.

[3] H. C. Elman, *Interactive methods for large, sparse, nonsymmetric systems of linear equations*, Yale Univ. Research Report 229, New Haven, CT, April 1982.

[4] M. Engeli, T. Ginsburg, H. Rutishauser and E. Stiefel, *Refined Interactive Methods for Computation of the Solution and the Eigenvalues of Self-Adjoint Boundary Value Problems*, Birkhauser, Verlag, Basel/Stuttgart, 1959.

[5] F. R. Gantmacher, *The Theory of Matrices*, Vol. I, Chelsea, New York, 1977.

[6] G. D. Mostow, J. H. Sampson and Jean-Pierre Meyer, *Fundamental Structures of Algebra*, McGraw-Hill, New York, 1963.

[7] Signum Newsletter, 16, 4 (1981), p. 7.