Numerische Mathematik 6, 269—284 (1964)

Some Generalisations of the Theory of Successive Over-Relaxation

By

C. G. BROYDEN

A new definition of "consistent ordering" is used to generalise the theory of successive over-relaxation to certain matrices that cannot be permuted into block tri-diagonal form. The matrices described by YOUNG, VARGA and KJELLBERG appear as special cases of the more general theory.

1. Introduction

The theory of the method of successive over-relaxation, which will subsequently be abbreviated to the method of SOR, has shown that for a certain class of matrices there exists a relationship between the eigenvalues of the SOR error operator and those of the Jacobi error operator (see MARTIN and TEE, 1961 for a definition of terms and a summary of results, and FORSYTHE and WASOW, 1960 and VARGA, 1962 for a more complete exposition of the theory). This class of matrices can be defined in terms of the possibility of performing certain permutations of their rows and columns in order to transform the matrix into a standard form, notably the block tri-diagonal form into which YOUNG's original form (YOUNG, 1954) can be permuted, or the p-cyclic form described by VARGA. The theory also indicates that the Jacobi operators of this class of matrices are, in some sense, cyclic. One of the properties of cyclic matrices is that their non-zero eigenvalues occur in groups, where the eigenvalues belonging to a particular group are closely related. These results depend on a fundamental property shared by these matrices, the property of being consistently ordered.

This concept was introduced by YOUNG in 1954, who defined a consistently ordered matrix in terms of an ordering vector which was immediately related to the disposition of zero and non-zero elements in the matrix. FORSYTHE and WASOW, in 1960, showed that this was equivalent to the possibility of permuting the rows and columns of the matrix so that the transformed matrix was of block tri-diagonal form. One important feature of these permutations was that they were not permitted to shift non-zero elements across the principal diagonal. The block tri-diagonal representation of a matrix of this type is not unique, and provided a permitted permutation into any one of these representations is possible, the matrix is consistently ordered. Matrices that can be transformed by permitted permutations into a particular block tri-diagonal representation are said to be consistently ordered with respect one to another.

VARGA, (VARGA, 1959) gave an alternative definition of consistent ordering which depended upon the eigenvalues of a certain matrix remaining unaltered when a scalar parameter was varied. This definition was more general than those that preceded it, although not so immediately related to the block structure of the matrix concerned.

The present paper introduces yet another definition of consistent ordering of a matrix, one that relies upon the existence of a set of matrices having special properties. If certain relationships between these defining matrices and the original matrix exist, then the latter is said to be generally consistently ordered. These relationships involve two parameters, which govern both the form of the final eigenvalue relationships and the cyclicity of the Jacobi operator. It is shown that the set of defining matrices can themselves be generated from a further matrix, and if this latter happens to be the unit matrix, then the consistently ordered matrix assumes the characteristic block tri-diagonal form. Assigning particular values to the two parameters then gives the results of YOUNG, VARGA or KJELLBERG.

2. Some Preliminary Results

In order to define a "generally consistently ordered" (GCO) $n \times n$ matrix, a set of p non-null $n \times n$ matrices S_i, $i=1, 2, \dots, p$, is required, where the members of the set satisfy the following conditions

$$S_i^2 = S_i, \qquad \text{c}(1)$$

$$S_i S_j = 0, \qquad j \neq i, \qquad \text{c}(2)$$

$$\sum_{i=1}^{p} S_i = I. \qquad \text{c}(3)$$

A set of matrices satisfying these three conditions will be referred to subsequently as a complete set of defining matrices. A set satisfying only conditions c(1) and c(2) will be called an incomplete set of defining matrices.

Theorem 1. A complete set of defining matrices may be derived from any non-singular matrix; conversely, any complete set of defining matrices may be derived from some non-singular matrix.

Proof. To demonstrate the first part of the theorem, let G be any non-singular matrix having an inverse H, and partition G into p sub-matrices, where the i-th partition G_i has $m(i)$ columns, and all have n rows. Similarly, partition H so that the i-th partition H_i^T has $m(i)$ rows and n columns. Then, since

$$GH = I, \tag{2.1}$$

$$HG = I, \tag{2.2}$$

it follows that

$$\sum_{i=1}^{p} G_i H_i^T = I_n, \tag{2.3}$$

$$H_i^T G_i = I_{m(i)}, \tag{2.4}$$

$$H_i^T G_j = 0, \qquad j \neq i. \tag{2.5}$$

(The subscript to the unit matrix is used to indicate its order.)

Let

$$S_i = G_i H_i^T, \qquad i = 1, 2, \dots, p. \tag{2.6}$$

Then, from (2.3) to (2.6), the set of matrices S_i satisfies the conditions c(1) to c(3).

To show that the converse is true, let S_i have rank $m(i)$. Then it may be expressed as the product of two rectangular matrices, the first having n rows and $m(i)$ linearly independent columns, and the second having n columns and $m(i)$ linearly independent rows. Let these matrices be G_i and H_i^T respectively. Then

$$S_i = G_i H_i^T. \tag{2.7}$$

Since S_i satisfies c(1),

$$G_i H_i^T = G_i H_i^T G_i H_i^T \tag{2.8}$$

and this, coupled with the linear independence of the columns of G_i and the rows of H_i^T implies that

$$H_i^T G_i = I_{m(i)}. \tag{2.9}$$

Similarly, from c(2) and the linear independence it follows that

$$H_i^T G_j = 0, \qquad j \neq i. \tag{2.10}$$

Let

$$G = [G_1, G_2, \ldots, G_p] \tag{2.11}$$

and

$$H = \begin{bmatrix} H_1^T \\ H_2^T \\ \vdots \\ H_p^T \end{bmatrix}. \tag{2.12}$$

G will be an $n \times m$ matrix, and H an $m \times n$ matrix, where

$$m = \sum_{i=1}^{p} m(i). \tag{2.13}$$

Equations (2.9) and (2.10) may now be expressed as the single equation

$$H G = I_m. \tag{2.14}$$

However, the matrices S_i must also satisfy c(3) and this may be written

$$G H = I_n. \tag{2.15}$$

Equations (2.14) and (2.15) can only be satisfied concurrently if G and H are both square, and hence reciprocal, and this completes the proof.

In addition to Theorem 1, the following Lemmas will be useful in the development of the theory.

Lemma 1. The rank of the sum of any two defining matrices is equal to the sum of their ranks.

Proof. The sum of S_i and S_j may, from Theorem 1, always be written

$$S_i + S_j = [G_i\, G_j] \begin{bmatrix} H_i^T \\ H_j^T \end{bmatrix}.$$

Since all the columns of G and rows of H are linearly independent, the rank of $S_i + S_j$ is equal to $m(i) + m(j)$, the total number of columns in $[G_i G_j]$. But the rank of S_i is $m(i)$ and that of S_j is $m(j)$, and the Lemma follows.

Lemma 2. If the set of p matrices S_i is divided into q subsets, and R_j denotes the sum of the S_i's contained in the j-th subset, then the q matrices R_j are themselves defining matrices.

Lemma 3. The sum of the ranks of all the matrices S_i is equal to n.

3. General Consistent Ordering

The set of matrices S_i discussed in the previous section will now be used to define a GCO matrix. Since, however, this concept is intimately related to both SOR and SBOR (successive block over-relaxation), it cannot usefully be discussed independently of the error operators appropriate to these methods, or of the corresponding Jacobi error operators. Consider the solution of the equation

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

where

$$A = D - E - F \tag{3.2}$$

and

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,m} \\ A_{2,1} & A_{2,2} & \dots & A_{2,m} \\ \vdots & & & \\ A_{m,1} & A_{m,2} & \dots & A_{m,m} \end{bmatrix} \tag{3.3}$$

$$D = \begin{bmatrix} A_{1,1} & 0 & \dots & 0 \\ 0 & A_{2,2} & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & A_{m,m} \end{bmatrix} \tag{3.4}$$

$$E = -\begin{bmatrix} 0 & 0 & \dots & 0 \\ A_{2,1} & 0 & \dots & 0 \\ \vdots & & & \\ A_{m,1} & A_{m,2} & \dots & 0 \end{bmatrix} \tag{3.5}$$

$$F = -\begin{bmatrix} 0 & A_{1,2} & \dots & A_{1,m} \\ 0 & 0 & \dots & A_{2,m} \\ \vdots & & & \\ 0 & 0 & \dots & 0 \end{bmatrix} \tag{3.6}$$

The elements $A_{i,j}$ are sub-matrices, or blocks, and are not necessarily square except when they lie on the principal block diagonal, when they must be both square and non-singular. The equation giving the $(m+1)$-st estimate of the solution in terms of the m-th for the method of SBOR is given by (VARGA, 1962, p. 80)

$$(D - \omega E) X^{(m+1)} = [\omega F - (\omega - 1) D] X^{(m)} + \omega K, \qquad m \geqq 0. \tag{3.7}$$

The corresponding equation for the block Jacobi method is (ibid, p. 105)

$$D X^{(m+1)} = (E + F) X^{(m)} + K, \qquad m \geqq 0. \tag{3.8}$$

From these the two error operators may be derived, these being

$$K = D^{-1}(E + F) \qquad \text{Block Jacobi,} \tag{3.9}$$

$$H = (D - \omega E)^{-1}[\omega F - (\omega - 1) D] \qquad \text{SBOR.} \tag{3.10}$$

Suppose now that equation (3.1) is pre-multiplied by D^{-1}. If L and U are defined by

$$L = D^{-1} E, \tag{3.11}$$

$$U = D^{-1} F, \tag{3.12}$$

equation (3.1) becomes

$$(I - L - U) X = D^{-1} K \tag{3.13}$$

and (3.9) and (3.10) may be written

$$K = L + U, \tag{3.14}$$

$$H = (I - \omega L)^{-1}[\omega U - (\omega - 1) I]. \tag{3.15}$$

Equations (3.13) to (3.15) show that K and H are identical to the point Jacobi and SOR operators associated with the matrix $D^{-1}A$, so the methods of Block Jacobi and SBOR are formally equivalent to the corresponding point method applied to the modified matrix equation (3.13). The block structure of the iteration is specified by the choice of D, and the modified matrix has a unit principal diagonal. The concept of General Consistent Ordering will be defined in terms of this modified matrix, and so will be valid both for SOR and SBOR.

Definition 1. Let S_i, $i = 1, 2, \ldots, p$, be a complete set of defining matrices and let S_i be defined as the $n \times n$ null matrix for $i < 1$ and $i > p$. Then a matrix A is a GCO(q, r) matrix for a scheme of SBOR specified by the non-singular block-diagonal matrix D if two positive integers q and r, $1 \leqq q, r < p$, exist such that

$$\begin{gathered} D^{-1} A = I - L - U, \\ S_i U = U S_{i+q}, \\ S_i L = L S_{i-r} \end{gathered} \tag{3.16}$$

for all integral i, where U, L are strictly upper and lower triangular respectively.

Further relationships of U and L may be deduced from the above definition. Condition c(3) and (3.16) give

$$U = \sum_{i=1}^{p} S_i U = \sum_{i=1}^{p-q} U S_{i+q}. \tag{3.17}$$

Similarly,

$$L = \sum_{i=1}^{p} S_i L = \sum_{i=r+1}^{p} L S_{i-r}. \tag{3.18}$$

4. The Jacobi Method and the Method of SOR

In general little can be said about the properties of the Jacobi error operator K and the SOR error operator H defined by equations (3.14) and (3.15), but if the matrix A is a GCO matrix it is possible to derive a functional relationship between their eigenvalues. Before this is done, however, two Lemmas valid for all matrices will be proved.

Lemma 4. The SOR operator has a zero eigenvalue if and only if ω, the over-relaxation factor, is equal to unity.

Proof. The SOR error operator consists, from (3.15), of two factors of which the first is non-singular for all finite ω. The second factor is singular if and only if ω equals unity, and since the product of two non-singular matrices is itself non-singular, and any matrix product involving a singular matrix is itself singular, the Lemma is proved.

Lemma 5. The SOR operator has a non-zero eigenvalue λ, and a corresponding eigenvector Z if and only if the matrix F, where

$$F=(\lambda^{-\gamma}U+\lambda^{1-\gamma}L) \tag{4.1}$$

and γ is some arbitrary scalar, itself has an eigenvector Z and an eigenvalue η/λ^{γ}, where

$$\eta=\frac{\lambda+\omega-1}{\omega}. \tag{4.2}$$

Proof. The definition of the SOR operator gives

$$(I-\omega L)^{-1}[\omega U-(\omega-1)I]Z=\lambda Z. \tag{4.3}$$

Pre-multiplying this by $(I-\omega L)$ and rearranging terms gives

$$(U+\lambda L)Z=\left(\frac{\lambda+\omega-1}{\omega}\right)Z \tag{4.4}$$

and dividing (4.4) by λ^{γ} gives

$$FZ=\frac{\eta Z}{\lambda^{\gamma}} \tag{4.5}$$

proving one part of the Lemma. The converse is proved by carrying out the above reasoning in the reverse order.

It will now be shown that if A is a GCO matrix then, for any non-zero value of λ and a particular value of γ, there exists a similarity transformation of the Jacobi operator K into F.

Let A be a GCO(q,r) matrix for some scheme of SBOR specified by the block diagonal matrix D. Then A, and its two associated operators K and H, satisfy equations (3.14) to (3.16). Consider now the matrix P, where

$$P=\sum_{i=1}^{p}\alpha_i S_i, \qquad \alpha_i\neq 0 \tag{4.6}$$

and the set S_i consists of the defining matrices of A. Then, from conditions c(1) to c(3),

$$P\sum_{i=1}^{p}\alpha_i^{-1}S_i=\sum_{i=1}^{p}S_i=I. \tag{4.7}$$

Hence

$$P^{-1}=\sum_{i=1}^{p}\alpha_i^{-1}S_i. \tag{4.8}$$

Consider now the transformations $U_1 = P U P^{-1}$, $L_1 = P L P^{-1}$.

$$\begin{aligned} U_1 &= \left(\sum_{i=1}^{p} \alpha_i S_i\right) U \left(\sum_{i=1}^{p} \alpha_i^{-1} S_i\right) \\ &= \left(\sum_{i=1}^{p-q} \alpha_i U S_{i+q}\right) \left(\sum_{i=1}^{p} \alpha_i^{-1} S_i\right) \end{aligned} \tag{4.9}$$

$$U_1 = \sum_{i=1}^{p-q} \frac{\alpha_i}{\alpha_{i+q}} U S_{i+q}. \tag{4.10}$$

Since the values of α_i may be chosen to be any arbitrary non-zero number, let

$$\alpha_i = \varrho^{i-1}. \tag{4.11}$$

Equation (4.10) then becomes, by virtue of equation (3.17),

$$U_1 = \varrho^{-q} U. \tag{4.12}$$

Similarly,

$$L_1 = \varrho^{r} L. \tag{4.13}$$

Now the Jacobi operator will have been transformed into F if

$$\begin{aligned} \varrho^{-q} &= \lambda^{-\gamma} \\ \varrho^{r} &= \lambda^{1-\gamma} \end{aligned} \tag{4.14}$$

and these two equations are satisfied if

$$\varrho = \lambda^{\left(\frac{1}{q+r}\right)}, \tag{4.15}$$

$$\gamma = \frac{q}{q+r}. \tag{4.16}$$

Thus, if A is a GCO(q, r) matrix and γ satisfies equation (4.16), K may be transformed into F, where λ may assume any non-zero value. Since a similarity transformation leaves the eigenvalues of the transformed matrix unchanged, all the matrices F will have eigenvalues identical with those of K. Now if λ is chosen to satisfy the equation

$$\mu \lambda^{\gamma} = \eta, \tag{4.17}$$

where μ is an eigenvalue of K, then η/λ^{γ} will be an eigenvalue of F, and hence, from Lemma 5, λ will be an eigenvalue of H. Conversely, if λ is an eigenvalue of H then, from Lemma 5, η/λ^{γ} is an eigenvalue of F for any γ and non-zero λ. If γ is now chosen to satisfy (4.16) then F may be transformed into K, and hence η/λ^{γ} is an eigenvalue of K. These results do not follow if λ equals zero, since in this case F is not defined. However, Lemma 4 states that λ can only be zero if ω is equal to unity, and in this case (4.17) reduces to

$$\mu \lambda^{\gamma} = \lambda \tag{4.18}$$

and since, from Lemma 4, zero is an eigenvalue of H when ω is equal to one, then even in this case if μ is an eigenvalue of K, and λ satisfies (4.18), then λ is an eigenvalue of H. The converse, however, does not hold for if λ is put equal to zero in (4.18), *any* value of μ will satisfy the equation, a result which is manifestly absurd. These results may be summed up in

Theorem 2. If A is a $\mathrm{GCO}(q, r)$ matrix, and if λ, μ satisfy the equation

$$\omega\mu\lambda^{\frac{q}{q+r}} = \lambda + \omega - 1 \tag{4.19}$$

then if μ is an eigenvalue of the (Block) Jacobi operator, λ is an eigenvalue of the S(B)OR operator. Furthermore, if λ is a non-zero eigenvalue of the S(B)OR operator, any μ satisfying (4.19) is an eigenvalue of the (Block) Jacobi operator.

The similarity transformation of K into F may be used to deduce a relationship between the eigenvectors of K and those of H. Lemma 5 states that H and F have a common set of eigenvectors, so it is only necessary to derive the relationship between the eigenvectors of F and those of K. Let X be the eigenvector of K corresponding to μ, and Z be the eigenvector of F corresponding to λ, the eigenvalue of H related to μ by equation (4.19). Then

$$KX = \mu X, \tag{4.20}$$

$$FZ = \eta/\lambda^{\gamma} Z. \tag{4.21}$$

Pre-multiplying (4.20) by P gives

$$PKP^{-1}PX = \mu PX. \tag{4.22}$$

This, from the similarity transformation and equation (4.19), may be written

$$FPX = \eta/\lambda^{\gamma} PX. \tag{4.23}$$

Hence,

$$Z = PX \tag{4.24}$$

or, expanding P by equation (4.6), (4.11) and (4.15),

$$Z = \sum_{i=1}^{p} \left(\lambda^{\frac{1}{q+r}}\right)^{i-1} S_i X \tag{4.25}$$

and this transformation is valid for all non-zero λ.

5. The Jacobi Operators of GCO Matrices

It was shown in the previous section that a relationship exists between the eigenvalues of the Jacobi and SOR operators derived from a GCO matrix. In this section the eigenvalue structure of the Jacobi operator itself will be examined, and it will be shown that the eigenvalues form a definite pattern. This pattern is the same as the pattern formed by the weakly cyclic matrices discussed by VARGA (VARGA, 1962), and this similarity suggests that the term "cyclic" is not inappropriate if used to describe these more general matrices. The precise form of the pattern will now be specified.

Definition 2. A matrix will be said to be a Cyclic (p), or $\mathrm{C}(p)$, matrix if for every non-zero eigenvalue μ there exist $p-1$ other eigenvalues $\vartheta^j\mu$, where $j = 1, 2, \ldots, p-1$, and ϑ is a primitive p-th root of unity.

It will now be shown that the Jacobi operators derived from GCO matrices are cyclic in accordance with the above definition.

Theorem 3. Let K be the Jacobi operator of a $\mathrm{GCO}(q, r)$ matrix, and let t be the greatest common divisor of q and r. If m is defined by

$$mt = q + r \tag{5.1}$$

then K is a $\mathrm{C}(m)$ matrix.

Proof. Let ϑ be a primitive m-th root of unity, and define k by

$$kt = q. \tag{5.2}$$

Now if μ is an eigenvalue of K, then $\vartheta^{jk}\mu$ is an eigenvalue of $\vartheta^{jk}K$, where j is an integer in the range $1 \leqq j \leqq m-1$. If $\vartheta^{jk}K$ is denoted by F, then from equation (3.14)

$$F = \vartheta^{jk}(U + L). \tag{5.3}$$

Now since ϑ is an m-th root of unity, $\vartheta^{jm} = 1$, and

$$\vartheta^{jk} = \vartheta^{-j(m-k)}. \tag{5.4}$$

From this equation, and equations (5.1) and (5.2), (5.3) may be written

$$F = \left(\vartheta^{-\frac{j}{t}}\right)^{-q} U + \left(\vartheta^{-\frac{j}{t}}\right)^{r} L. \tag{5.5}$$

Now K is the Jacobi operator of a $\mathrm{GCO}(q, r)$ matrix, and it was shown in Section 4 that it could be transformed into F by a similarity transformation provided that $\vartheta^{-\frac{j}{t}} \neq 0$. Since ϑ is a root of unity, it can never be zero, and since a similarity transformation leaves the eigenvalues of a matrix unchanged, then $\vartheta^{jk}\mu$, $j = 1, 2, \ldots, m-1$ must also be eigenvalues of K. But t is the greatest common divisor of q and r, so that k and m are relatively prime. Hence the $m-1$ derived eigenvalues of K are a particular sequence of $\vartheta^{j}\mu$, $j = 1, 2, \ldots, m-1$, and Theorem 3 is proved.

Corollary. The Jacobi operator of a $\mathrm{GCO}(q, r)$ matrix, where q and r are relatively prime, is a $\mathrm{C}(q+r)$ matrix.

Theorem 4. If K is the Jacobi operator of a $\mathrm{GCO}(q, r)$ matrix, and t is the greatest common divisor of q and r, then K may be expressed as the sum of t matrices K_j, $j = 1, 2, \ldots, t$, where

$$K_j K_i = 0, \qquad i \neq j \tag{5.6}$$

and each K_j is a $\mathrm{C}(m)$ matrix, where m is defined by equation (5.1).

Proof. Let S_i, $i = 1, 2, \ldots, p$ be the complete set of defining matrices of K, and let $S_i = 0$ for $i < 1, > p$. Define R_j, $j = 1, 2, \ldots, t$ by

$$R_j = \sum S_{j+(h-1)t} \tag{5.7}$$

where the sum is taken over all integral values of h for which $S_{j+(h-1)t}$ is not null. Since K is derived from a $\mathrm{GCO}(q, r)$ matrix,

$$S_{j+(h-1)t} U = U S_{j+(h-1)t+q}. \tag{5.8}$$

If k is defined by equation (5.2), this becomes

$$S_{j+(h-1)t} U = U S_{j+(h+k-1)t}. \tag{5.9}$$

Now the defining matrix on the r.h.s. of this equation is itself a term in R_j, so that post-multiplication of r.h.s. by R_j will leave it unchanged. Hence

$$S_{j+(h-1)t} U = S_{j+(h-1)t} U R_j. \tag{5.10}$$

Summing this equation for all terms of R_j gives

$$R_j U = R_j U R_j. \tag{5.11}$$

The identical relationship similarly holds for L, and since K is the sum of L and U,

$$R_j K = R_j K R_j. \tag{5.12}$$

Now it follows from Lemma 2 and the definition of R_j's that they themselves form a complete set of defining matrices, and hence sum to the unit matrix. If

$$K_j = R_j K \tag{5.13}$$

then clearly K is equal to the sum of the t K_j's, and, from (5.12) and condition c(2), (5.6) must hold.

To show that each K_j is a C(m) matrix, all that is necessary is to note that R_j and the matrix effecting the similarity transformation of K into F commute. Hence $R_j(L+U)$ can be transformed into $\left(\left(\vartheta^{-\frac{j}{t}}\right)^{-q} R_j U + \left(\vartheta^{-\frac{j}{t}}\right)^{r} R_j L\right)$ by the same similarity transformation that takes K into the F of equation (5.5), and a similar argument to that used in establishing the previous theorem completes the proof of Theorem 4.

6. Special Cases

In general, a GCO matrix will not have a well defined block structure, nor indeed will it be possible to achieve this structure by permutation of the rows and columns. Only if the defining matrices are generated from particularly simple matrices will permutation into this block structure be possible, and these cases will now be considered. The simplest possible non-singular matrix is the unit matrix, and if, in Section 2 above, the matrix G is taken to be the unit matrix then the form of any GCO matrix finally obtained is a generalisation of the block tridiagonal matrix discussed by FORSYTHE and WASOW.

To show this partition the matrix G_i of Section 2 into p blocks, where the j-th block consists of $m(j)$ rows and all have $m(i)$ columns. Then

$$G_i = \begin{bmatrix} G_{i1} \\ G_{i2} \\ \vdots \\ G_{ip} \end{bmatrix}. \tag{6.1}$$

If G is the unit matrix, then

$$\begin{aligned} G_{ii} &= I_{m(i)}, \\ G_{ij} &= 0, \quad j \neq i \end{aligned} \tag{6.2}$$

and H_i^T is merely the transpose of G_i. Hence, if S_i is partitioned into p block-rows and p block-columns as follows

$$S_i = \begin{bmatrix} S_{11} & S_{12} \dots S_{1p} \\ S_{21} & S_{22} \dots S_{2p} \\ \vdots & \vdots \quad\quad \vdots \\ S_{p1} & S_{p2} \dots S_{pp} \end{bmatrix}, \tag{6.3}$$

where S_{jk} is an $m(j) \times m(k)$ submatrix, then

$$\begin{aligned} &S_{ii} = I_{m(i)}, \\ &S_{jk} = 0, \qquad j \neq i \quad \text{or} \quad k \neq i. \end{aligned} \tag{6.4}$$

Consider now the definition of General Consistent Ordering. The second equation of the set (3.16) states that

$$S_i U = U S_{i+q}. \tag{6.5}$$

Now the effect of pre-multiplying a matrix by the particular form of S_i defined by (6.4) is to annihilate a large number of its rows. In fact, if U is partitioned according to the scheme of equation (6.3) into p block-rows and block-columns, then pre-multiplying it by S_i annihilates all block rows with the exception of the i-th, which remains unchanged. Similarly, post-multiplication of U by S_i annihilates all block-columns with the exception that the i-th is unaltered. Consider now equation (6.5). This indicates that for a GCO matrix defined by this particular set of defining matrices, annihilation of all block-rows but the i-th is equivalent to the annihilation of all block-columns but the $(i+q)$-th. This is true if and only if the only non-zero elements in the i-th block-row and $(i+q)$-th block-column are common to both, i.e. occur at the intersection of this row and column. Hence if the defining matrices are generated from the unit matrix, equation (6.5) may be written

$$U_{ij} = 0, \qquad j \neq i + q, \qquad 1 \leqq i \leqq p - q, \tag{6.6}$$

where U_{ij} is the $m(i) \times m(j)$ sub-matrix common to the i-th block-row and j-th block-column. This shows that the non-zero elements of U occur along some upper block-diagonal, where the distance of this diagonal from the principal, in terms of number of blocks, is given by the parameter q. Similarly L consists of some lower block-diagonal whose position is determined by the parameter r. Now the Jacobi operator K is the sum of U and L, and thus consists of a general block bi-diagonal matrix. This is shown schematically for $p=7$ and various combinations of q and r in Figs. 1—4.

The theory developed by YOUNG, VARGA and KJELLBERG (KJELLBERG, 1961), however treats not only block-diagonal matrices but matrices that can be transformed into them by suitable row and column permutations. It is now shown that if these permutations are possible, then an appropriate set of defining matrices exists.

Let U be an upper block-diagonal matrix whose defining matrices S_i are generated from the unit matrix. Let U_1 be a strictly upper triangular matrix and P be some permutation matrix such that

$$U = P^T U_1 P. \tag{6.7}$$

Then, from equation (6.5)

$$S_i P^T U_1 P = P^T U_1 P S_{i+q}. \tag{6.8}$$

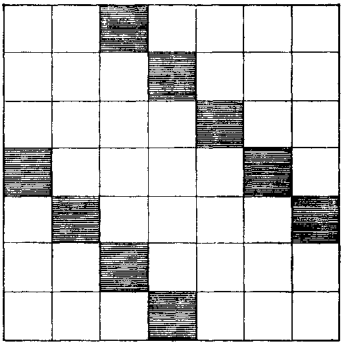

Fig. 1. Typical Case. $q=2$, $r=3$

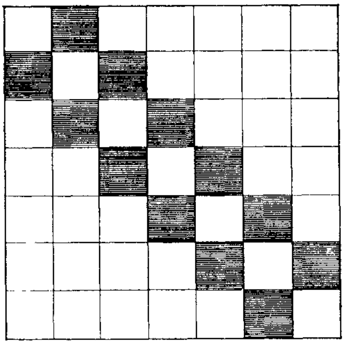

Fig. 2. YOUNG's Form. $q=r=1$

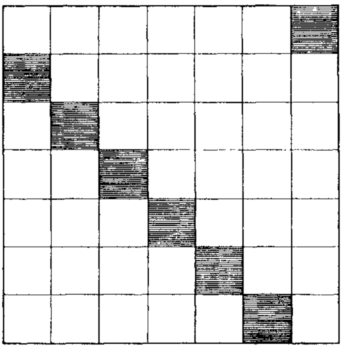

Fig. 3. VARGA's Form. $r=1$, $q=p-1=6$

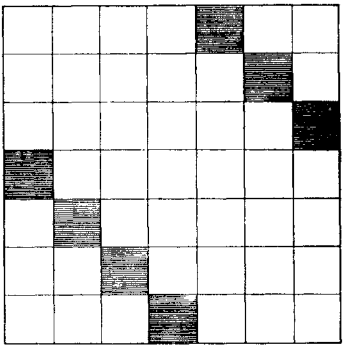

Fig. 4. KJELLBERG's Form. $q=4$, $r=p-q=3$

Pre-multiplying (6.8) by P, postmultiplying by P^T and remembering that the transpose of a permutation matrix is also the inverse gives

$$T_i U_1 = U_1 T_{i+q}, \tag{6.9}$$

where

$$T_i = P S_i P^T. \tag{6.10}$$

It is simple to verify that since the matrices S_i form a complete set of defining matrices, then so do the matrices T_i.

Now if the strictly lower triangular matrix L_1 and the lower block-diagonal matrix L are related in the same way, the matrix $(I - L_1 - U_1)$ will be a GCO(q, r) matrix with defining matrices T_i. Hence any matrix that can be permuted into a block tri-diagonal GCO(q, r) matrix without permitting any element to cross the principal diagonal is itself a GCO(q, r) matrix. Furthermore, the eigenvalues of its SOR operator will be the same as the eigenvalues of the SOR operator of the block tri-diagonal matrix. For let K_1 be the appropriate Jacobi operator. Then

$$K = P^T K_1 P. \tag{6.11}$$

Since a similarity transformation preserves the eigenvalues, the eigenvalues of K and K_1 will be identical, and since the eigenvalues of both H and H_1 may be derived by the same version of equation (4.19) from the same set of Jacobi eigenvalues, they too must be identical.

The above analysis, from equation (6.7) onward, is also valid if P and P^T are replaced by any non-singular matrix and its inverse. Provided that the lower and upper triangular matrices are preserved under the transformation, the SOR operators of the two matrices will have an identical set of eigenvalues.

The forms and equations associated with Young, Varga and Kjellberg will now be derived. It will be assumed that the permutations into block-diagonal form have been carried out, since these do not affect the essential eigenvalue structure of the operators.

Figs. 1—4 depict schematically the block structure of the Jacobi operators associated with four GCO matrices whose defining matrices are generated from the unit matrix. In each case p is taken to be seven.

Fig. 1 shows the case where $q=2$ and $r=3$, and the equation relating the Jacobi and SOR eigenvalues is

$$\omega \mu \lambda^{\frac{2}{5}} = \lambda + \omega - 1. \tag{6.12}$$

Fig. 2 shows the well-known block tri-diagonal form of Young, as described by Forsythe and Wasow. For this case, $q=r=1$, and (4.19) becomes

$$\omega \mu \lambda^{\frac{1}{2}} = \lambda + \omega - 1. \tag{6.13}$$

Fig. 3 depicts the p-cyclic form of Varga. Putting $r=1$ and $q=p-1$ gives the equation

$$\omega \mu \lambda^{\frac{p-1}{p}} = \lambda + \omega - 1. \tag{6.14}$$

Kjellberg's form, illustrated in Fig. 4, permits q to take any integer value in the range $1 \leqq q \leqq p-1$, but puts $r=p-q$. It is thus a more general form of Varga's case. The constraint upon the choice of r ensures that there is no "overlap" of the blocks. In Fig. 4 each block-row and block-column has one and one only non-zero block, whereas in Fig. 1 overlap occurs in two rows and columns, and in Fig. 2 there is the maximum amount of overlap. Kjellberg's equation is

$$\omega \mu \lambda^{q/p} = \lambda + \omega - 1. \tag{6.15}$$

Although it is possible to permute Young's form into Varga's "2-cyclic" form (i.e., σ_1 ordering) this can only be done by carrying blocks across the principal diagonal. Transformations involving block migration of this nature also leave the eigenvalues of the SOR operator unaltered provided that both forms are generally consistently ordered, the same value of $q/(q+r)$ applies in both cases, and the eigenvalues of the Jacobi operator are unaltered. The reasoning is similar to that when no block migration occurs.

Matrices for which it is not possible to find a permutation into some block tri-diagonal form have not so far, to the Author's knowledge, been included in the theory of successive over-relaxation. An example of a matrix for which this permutation is not possible but for which the theory still holds will now be given by way of illustration.

Consider the matrix A, where

$$A=\begin{bmatrix} 1 & -a & a & -a \\ -a & 1 & 0 & -a \\ a & 0 & 1 & -a \\ -a & -a & -a & 1 \end{bmatrix} \tag{6.16}$$

and the defining matrices S_1 and S_2, where

$$\begin{aligned} S_1&=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ S_2&=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{6.17}$$

It may easily be verified that

$$\begin{aligned} S_1 U &= U S_2, \\ S_2 L &= L S_1. \end{aligned} \tag{6.18}$$

Hence, A is a GCO(1, 1) matrix whose Jacobi operator will have eigenvalues occurring in pairs $\pm\mu$, and whose Jacobi and SOR operators will have eigenvalues related by equation (6.13). These facts may be verified by direct computation.

7. Further Properties of some Special Cases

A matrix A is said to be reducible if it is possible to obtain some elements of the solution X of the equation

$$AX=B \tag{7.1}$$

by considering only a certain partition of A. If this is possible, the most economical way of solving (7.1) is to find these elements initially, and then to solve for the remainder of X using the other partitions of A. More formally,

Definition 3. An $n\times n$ matrix is said to be reducible if it is possible to select from it m rows, where $m<n$, to form a sub-matrix having $(n-m)$ zero columns.

It will now be shown that certain types of GCO matrix are reducible.

Theorem 5. A GCO(q, r) matrix is reducible if its defining matrices can be generated from a permutation matrix, and q and r are not relatively prime.

Proof. From Theorem 4, the Jacobi operator of a matrix of this type can be expressed as the sum of matrices K_j, where

$$K_j K_i = 0, \qquad i \neq j \tag{7.2}$$

and

$$K_j = R_j K = K R_j. \tag{7.3}$$

Consider now the matrix A of which K is the Jacobi operator. Then

$$A = I - K. \tag{7.4}$$

Pre-multiplying (7.4) by R_j and post-multiplying by $(I - R_j)$ gives, from (7.3),

$$R_j A (I - R_j) = 0. \tag{7.5}$$

Now if the defining matrices of A can be generated from a permutation matrix, then R_j itself may be expressed as

$$R_j = G_j G_j^T, \tag{7.6}$$

where G_j consists of the first m columns of some $n \times n$ permutation matrix P. Similarly, $(I - R_j)$ satisfies the equation

$$I - R_j = G_c G_c^T, \tag{7.7}$$

where

$$[G_j G_c] = P. \tag{7.8}$$

Hence, equation (7.5) may be written

$$G_j G_j^T A\, G_c G_c^T = 0 \tag{7.9}$$

and pre-multiplying this by G_j^T and post-multiplying by G_c gives

$$G_j^T A\, G_c = 0. \tag{7.10}$$

Now G_j is an $n \times m$ partition of a permutation matrix, and G_c is, from (7.8), an $n \times (n - m)$ partition of the same matrix. Now multiplying a matrix by a partition of a permutation matrix is equivalent to selecting certain rows or columns, so that equation (7.10) indicates that if the appropriate m rows of A are selected, and of these the appropriate $n - m$ columns are taken, the result is null matrix. This is precisely the condition that A is reducible, and so Theorem 5 is proved.

Since this is true for all j, this theorem indicates that the solution of equation (7.1) could be accomplished by solving the t sets of $m(j)$ equations

$$(G_j^T A\, G_j)\, G_j^T X = G_j^T B, \tag{7.11}$$

where $m(j)$ is the rank of R_j, and t is the greatest common divisor of q and r. If a matrix of this type arose from a finite-difference mesh, it would indicate that the n mesh-points could be divided into t sets, with the finite-difference operator linking only those points within a particular set. An example would be an operator that connected each row of the mesh with the next row but one, an improbable but not impossible situation.

8. Conclusions

This attempt at the unification of the theory of successive over-relaxation has emphasised the importance of consistent ordering in the derivation of the Jacobi-SOR eigenvalue relationship. In particularly simple cases this ordering appears as a characteristic block structure of the matrix, which then consists of a principal, an upper and lower block diagonal, and the parameters in the Jacobi-SOR eigenvalue relationship are governed solely by the position of the outer block diagonals. The concept of consistent ordering has been generalised to matrices that have no apparent block structure, and the Jacobi and SOR operators of these matrices have been shown to be related in a similar manner to those operators obtained from matrices in which the block structure is evident.

Another feature exhibited by the Jacobi operators of these ordered matrices is the cyclic property, the eigenvalues of the Jacobi operator occurring in particularly simple patterns. This cyclic property obtains for all GCO matrices, although it is quite possible to construct cyclic matrices that are not generally consistently ordered.

Whether or not the extension of the theory to matrices having a more sophisticated form of ordering than those described by YOUNG and VARGA will have any practical applications is a matter of some doubt. Although it may be possible to construct finite-difference operators deliberately to have these properties, and an example of this already exists (TEE, 1963), there is no guarantee that the use of these operators and the theoretically optimum value of ω will provide a more rapidly converging process than more conventional forms of operator, which give rise to matrices that are not ordered in any way and for which the existence of an optimum value of ω has not yet been theoretically established.

The Author is grateful to his colleagues, especially G. J. TEE, for their helpful criticism of the various drafts of this paper, to Dr. D. M. PARKYN for sanctioning the study and to the directors of the English Electric Co. Ltd. for permission to publish the results.

References

FORSYTHE, G. E., and W. R. WASOW: Finite-Difference Methods for Partial Differential Equations. New York: Wiley 1960.

KJELLBERG, G.: On the successive over-relaxation method for cyclic operators. Num. Math. **3**, 87 (1961).

MARTIN, D. W., and G. J. TEE: Iterative Methods for Linear Equations with Symmetric Positive Definite Matrix. The Computer Journal **4**, 242 (1961).

TEE, G. J.: A Novel Finite-Difference Approximation to the Biharmonic Operator. The Computer Journal **6**, 177 (1963).

VARGA, R. S.: p-cyclic matrices: a generalization of the Young-Frankel successive over-relaxation scheme. Pacific J. Math. **9**, 617 (1959).

— Matrix Iterative Analysis. London: Prentice-Hall 1962.

YOUNG, D.: Iterative Methods for solving Partial Differential Equations of Elliptic Type. Trans. Amer. Math. Soc. **76**, 92 (1954).

Applied Analysis Section
Atomic Power Division
English Electric Co. Ltd.
Cambridge Road
Whetstone, Nr. Leicester (Great Britain)

(Received March 10, 1964)