Numerische Mathematik 12, 47–56 (1968)

Some Aspects of Consistent Ordering

C. G. BROYDEN

Received June 16, 1967

1. Introduction

The concept of "consistent ordering" is central to the theory of solving sets of linear equations by the method of SOR, and several definitions of this property have been given. The earliest was due to YOUNG [6], in 1954, who used an ordering vector related to the disposition of zero and non-zero elements in the matrix. FORSYTHE and WASOW [2] showed in 1960 that YOUNG'S definition was equivalent to saying that the matrix could be transformed in a certain way to a block tri-diagonal form.

In 1962, VARGA [5] introduced a more general definition of consistent ordering that included YOUNG'S as a special case. If $\boldsymbol{A}$ is the matrix of the linear system concerned, normalised so that the elements along the principal diagonal have the value unity, then

$$\boldsymbol{A}=\boldsymbol{I}-\boldsymbol{B} \tag{1.1}$$

where

$$\boldsymbol{B}=\boldsymbol{L}+\boldsymbol{U} \tag{1.2}$$

and $\boldsymbol{L}$, $\boldsymbol{U}$ are strictly lower and strictly upper triangular respectively. VARGA ([5], p. 101) defined $\boldsymbol{A}$ (or $\boldsymbol{B}$) to be consistently ordered if

(V1) the eigenvalues of $\alpha\boldsymbol{L}+\alpha^{-(p-1)}\boldsymbol{U}$, where p is an integer $\geqq 2$, are independent of α for all $\alpha \neq 0$ and

(V2) the matrix $\boldsymbol{B}$ is weakly cyclic of index p.

This definition is extremely general and powerful from a theoretical point of view. Condition (V1) however does not impart any immediate information about the structure of the matrix and it is often difficult to determine exactly whether or not it is satisfied unless $\boldsymbol{B}$ is of known form, e.g. block-bidiagonal.

KJELLBERG [4] in 1961 gave what was in effect an extremely elegant definition of a generalisation of the concept of consistent ordering. He showed that the results obtained by YOUNG and VARGA followed from making separate specific assumptions about the matrices $\boldsymbol{L}$ and $\boldsymbol{U}$, and his definition was as follows: "Let R be a vector space, which is the direct sum of subspaces $v_1, v_2, \ldots, v_m$, and let B be a bounded linear transformation from R to R that we may write $B=L+U$. Assume that L and U satisfy

$$\begin{aligned} L v_i &\in v_{i+h}, \\ U v_i &\in v_{i-k}, \end{aligned} \qquad i=1, 2, \ldots, m$$

where h and k are positive integers, relatively prime to p and with $h+k=p$. (We use the convention that $v_i=0$ whenever $i<1$ or $>m$.)"

The idea of treating $\boldsymbol{L}$ and $\boldsymbol{U}$ independently was also used by BROYDEN [1] in 1964. His definition of a generalisation of consistent ordering involved a complete set of projection matrices $\boldsymbol{S}_i$, $i=1, 2, \ldots, m$ (see section 2 below), and required that

$$\boldsymbol{S}_i \boldsymbol{U} = \boldsymbol{U} \boldsymbol{S}_{i+q}, \tag{1.3}$$

$$\boldsymbol{S}_i \boldsymbol{L} = \boldsymbol{L} \boldsymbol{S}_{i-r} \tag{1.4}$$

where r and q are positive integers and the convention that $\boldsymbol{S}_j$ is null for $j<1$, $j>m$ is assumed throughout.

The definitions of KJELLBERG and BROYDEN as given, which are shown to be equivalent in section 2 (below), are so general that the essential idea of ordering is lost. This may be retrieved by imposing the condition that the projection matrices $\boldsymbol{S}_i$ of BROYDEN's definition are diagonal, and with this restriction Eqs. (1.3) and (1.4) provide an alternative definition of consistent ordering. That this revised definition is equivalent to that of YOUNG if $q=r=1$ is comparatively easy to establish, and it is shown subsequently that if $\boldsymbol{B}$ is irreducible and non-negative, then with the appropriate choice of q and r the revised definition is equivalent to that of VARGA. It is further shown that if $\boldsymbol{B}$ is irreducible and non-negative VARGA's second condition (V2) is implicit in (V1). If, however, $\boldsymbol{B}$ is not irreducible and non-negative the definitions are not equivalent, and we show by example that VARGA's definition with $p=2$ is not equivalent to that of YOUNG.

The revised definition lends itself to the construction of simple numerical algorithms to determine whether or not a matrix is consistently ordered. These algorithms, which are based on determining the diagonal projection matrices concerned, are described in the final section of this note.

2. Implications of the Kjellberg-Broyden Definition

In order to define a generalisation of consistent ordering BROYDEN made use of matrices referred to in [1] as "defining matrices". These were a set of non-null $n \times n$ matrices $\boldsymbol{S}_i$, $i=1, 2, \ldots, m$, $m \leqq n$, that satisfied the following equations:

$$\boldsymbol{S}_i^2 = \boldsymbol{S}_i, \tag{2.1}$$

$$\boldsymbol{S}_i \boldsymbol{S}_j = \boldsymbol{0}, \quad i \neq j, \tag{2.2}$$

$$\sum_{i=1}^{m} \boldsymbol{S}_i = \boldsymbol{I}. \tag{2.3}$$

Since each matrix $\boldsymbol{S}_i$ may be identified as a generalised projection matrix ([3], p. 10) a set of matrices that satisfy all three Eqs. (2.1)—(2.3) is subsequently referred to as a complete set of projection matrices. Further properties of such sets may be found in [1].

The identification of the "defining matrices" of [1] with generalised projection matrices enables the essential equivalence of KJELLBERG's and BROYDEN's definitions to be seen. Eqs. (1.3) and (1.4) imply that a sequence of vector sub-spaces exists such that a vector lying in the $(i+q)$-th sub-space is projected into the i-th by $\boldsymbol{U}$, and that one lying in the $(i-r)$-th sub-space is projected into the i-th by $\boldsymbol{L}$. It is thus apparent that BROYDEN's definition of a generalisation of consistent ordering is that of KJELLBERG expressed in matrix terms.

We now show that if $\boldsymbol{B}$ satisfies the revised definition it also for appropriate values of q and r satisfies VARGA's. We first though define the concepts of reducibility and weak cyclicity in terms of a complete set of diagonal projection matrices.

It follows from (2.1) and (2.2) that diagonal projection matrices can only have zero or unity diagonal elements, and it is this property that makes their use possible in this context.

Definition 1. The $n \times n$ matrix $\boldsymbol{B}$ is reducible if $\exists$ an $n \times n$ diagonal projection matrix $\boldsymbol{S}$, which is neither the null nor the unit matrix, such that $\boldsymbol{SB}(\boldsymbol{I}-\boldsymbol{S})=\boldsymbol{0}$. Conversely, if no such matrix $\boldsymbol{S}$ exists, $\boldsymbol{B}$ is irreducible.

Definition 2. The $n \times n$ matrix $\boldsymbol{B}$ is weakly cyclic of index p if $\exists$ a complete set of p diagonal projection matrices $\boldsymbol{S}_i$, $1 \leq i \leq p$, such that

$$\begin{aligned} \boldsymbol{S}_i \boldsymbol{B} &= \boldsymbol{B}\boldsymbol{S}_{i+1}, \quad 1 \leq i \leq p-1, \\ \boldsymbol{S}_p \boldsymbol{B} &= \boldsymbol{B}\boldsymbol{S}_1. \end{aligned}$$

That these definitions are equivalent to the conventional ones (see e.g. [5]) is elementary. The form given here has been chosen since it is more suited to the subsequent development of the theory.

To show that the revised definition implies VARGA's we note that it was established in [1] that Eqs. (1.3) and (1.4) imply that the eigenvalues of $\lambda^{-\left(\frac{q}{q+r}\right)}\boldsymbol{U}+\lambda^{\frac{r}{q+r}}\boldsymbol{L}$ are independent of λ for all $\lambda \neq 0$. If now $q/r=p-1$, and we put $\lambda^{1/p}=\alpha$ then (V1), the first of VARGA's conditions, follows. To show that (V2) is also satisfied we prove the following theorem.

Theorem 1. Let $\boldsymbol{T}_i$, $i=1, 2, \ldots, m$ be a complete set of projection matrices and let

$$\boldsymbol{T}_i \boldsymbol{L} = \boldsymbol{L}\boldsymbol{T}_{i-1}, \tag{2.4}$$

$$\boldsymbol{T}_i \boldsymbol{U} = \boldsymbol{U}\boldsymbol{T}_{i+p-1}, \tag{2.5}$$

where $p \leq m$ and $\boldsymbol{T}_j$ is assumed null for $j<1$, $j>m$. Then, if $\boldsymbol{B}=\boldsymbol{L}+\boldsymbol{U}$, $\exists$ a complete set of p projection matrices $\boldsymbol{S}_i$ such that $\boldsymbol{S}_i\boldsymbol{B}=\boldsymbol{B}\boldsymbol{S}_{i+1}$, $1 \leq i \leq p$, where $p+1$ is interpreted modulo p.

Proof. Let

$$\boldsymbol{S}_i = \sum_{j=1}^{\infty} \boldsymbol{T}_{pj+1-i}, \quad i=1, 2, \ldots, p, \tag{2.6}$$

where the number of non-null terms in the sum is finite since $\boldsymbol{T}_j$ is assumed null for $j>m$. Then, from (2.4) and (2.6)

$$\begin{aligned} \boldsymbol{S}_i \boldsymbol{L} &= \boldsymbol{L} \sum_{j=1}^{\infty} \boldsymbol{T}_{pj-i}, \quad i=1, 2, \ldots, p, \\ &= \boldsymbol{L}\boldsymbol{S}_{i+1}, \quad i=1, 2, \ldots, p, \end{aligned}$$

where $p+1$ is interpreted modulo p. Similarly

$$\boldsymbol{S}_i \boldsymbol{U} = \boldsymbol{U}\boldsymbol{S}_{i+1}$$

and since $\boldsymbol{B}$ is the sum of $\boldsymbol{L}$ and $\boldsymbol{U}$ the theorem follows.

If the matrices $\boldsymbol{S}_i$, $1 \leq i \leq p$, are diagonal Definition 2 asserts that $\boldsymbol{B}$ is weakly cyclic of index p. Since the revised definition requires that the matrices $\boldsymbol{T}_i$ are diagonal, the matrices $\boldsymbol{S}_i$ are also diagonal and both VARGA'S conditions are satisfied.

It is possible, however, to satisfy these conditions when the projection matrices $\boldsymbol{T}_i$ are *not* diagonal since this last requirement is sufficient but not necessary for $\boldsymbol{B}$ to be weakly cyclic. We demonstrate this with the following example.

Example. Let the Jacobi operator $\boldsymbol{B}$ be given by

$$\boldsymbol{B}=\begin{bmatrix} \boldsymbol{0} & \boldsymbol{F} & \boldsymbol{0} & \boldsymbol{G} \\ \boldsymbol{H} & \boldsymbol{0} & \boldsymbol{F} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H} & \boldsymbol{0} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{H} & \boldsymbol{0} \end{bmatrix}$$

where $\boldsymbol{0}$ is the null matrix of order 2 and

$$\boldsymbol{F}=\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

$$\boldsymbol{G}=\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$\boldsymbol{H}=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Clearly $\boldsymbol{B}$ is weakly cyclic of index 2, and it can be shown that the eigenvalues of $\boldsymbol{B}(\alpha)\equiv\alpha\boldsymbol{L}+\alpha^{-1}\boldsymbol{U}$ are independent of α. $\boldsymbol{B}$ therefore satisfies VARGA'S definition of being consistently ordered despite the fact that it contravenes that of YOUNG, and the two definitions are thus not equivalent when $p=2$.

3. Implications of Varga's Definition

We now show that if $\boldsymbol{B}$ is non-negative, irreducible and satisfies VARGA'S definition of consistent ordering then it also satisfies the revised definition. In so doing we prove a number of theorems about non-negative matrices.

Definition 3. A matrix $\boldsymbol{B}=[b_{ij}]$ is an α-matrix if $b_{ij}=c_{ij}\alpha^{r_{ij}}$, where the c_{ij} are independent of α, $\alpha\neq 0$ and the r_{ij} are positive or negative integers, or zero. If $c_{ij}\geqq 0$ and $\alpha>0$ then $\boldsymbol{B}$ will be called a non-negative α-matrix.

Theorem 2. The eigenvalues of an irreducible non-negative $n\times n$ α-matrix $\boldsymbol{B}$ are independent of α only if the matrix $\sum_{i=1}^{m} a_i\boldsymbol{B}^i$ is an α-matrix for any $a_i\geqq 0$ and positive integer m.

Proof. If $\sum_{i=1}^{m} a_i\boldsymbol{B}^i$ is not an α-matrix at least one of its elements will have the form $\alpha^r p(\alpha)$, where $p(\alpha)$ is a polynomial in α of degree $\geqq 1$ whose coefficients are all non-negative with at least two being non-zero. Since $\boldsymbol{B}$ is non-negative and $a_i\geqq 0$ the matrix $\boldsymbol{I}+\sum_{i=1}^{m}(k_i a_i+b_i)\boldsymbol{B}^i$, where $k_i>0$ and $b_i\geqq 0$, is itself a non-negative matrix with at least one element having the form of a polynomial in α, and thus $(\boldsymbol{I}+\boldsymbol{B})^m$ has at least one such element. Now a well-known theorem (see e.g. [5], p. 26) states that if $\boldsymbol{B}$ is an irreducible non-negative $n\times n$ matrix then $(\boldsymbol{I}+\boldsymbol{B})^k > \boldsymbol{0}$, $k \geqq n-1$, and combining this with the previous result implies the existence of a least positive integer M such that $(\boldsymbol{I}+\boldsymbol{B})^M$ is strictly positive and has at least one element of polynomial form. It follows immediately that $(\boldsymbol{I}+\boldsymbol{B})^{2M}$ has at least one diagonal element of polynomial form, and since it is non-negative its trace is thus a polynomial in α of degree $\geqq 1$. The eigenvalues of $(\boldsymbol{I}+\boldsymbol{B})^{2M}$, and hence those of $\boldsymbol{B}$, thus depend upon α and the theorem is proved.

Theorem 2 shows that if the eigenvalues of an irreducible non-negative α-matrix are independent of α the elements of the matrix having a given power of α occur in precise zones, and these zones are invariant for any polynomial function of the original matrix. We shall see subsequently that these zones may be specified concisely by a complete set of diagonal projection matrices.

Definition 4. Two $n \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ have non-zero elements in common, or, more briefly, elements in common if for at least one pair of integers i and j, $a_{ij} \neq 0$ and $b_{ij} \neq 0$.

If for every integer pair i, j for which $a_{ij} \neq 0$, b_{ij} is also non-zero we say that $\boldsymbol{A}$ is common with $\boldsymbol{B}$. Note that if $\boldsymbol{A}$ is common with $\boldsymbol{B}$ the converse is not necessarily true, since $\boldsymbol{B}$ may have more non-zero elements than $\boldsymbol{A}$.

Theorem 3. If $\boldsymbol{C}$ is an irreducible non-negative matrix, $\sum_{i=1}^{p} \boldsymbol{C}^i > \boldsymbol{0}$, and of the first $(p+1)$ positive powers of $\boldsymbol{C}$ only $\boldsymbol{C}$ and $\boldsymbol{C}^{p+1}$ have elements in common then $\boldsymbol{C}$ is weakly cyclic of index p.

Proof. Let $\boldsymbol{S}_i$ be the diagonal projection matrix whose unity diagonal elements occur in those columns occupied by the non-zero entries of the first row of $\boldsymbol{C}^i$, $i \geqq 1$. Since $\boldsymbol{C}$ is irreducible no $\boldsymbol{S}_i$ can be null and it follows from the conditions of the theorem that $\boldsymbol{S}_i$, $i = 1, 2, \ldots, p$, form a complete set of diagonal projection matrices.

Consider now the first row of the matrix $\boldsymbol{C}^{j+1}$, $j = 1, 2, \ldots, p-1$. It must have zeroes in those columns where the first row of $\boldsymbol{C}_i$, $1 \leqq i \leqq p$, $i \neq j+1$, has non-zero elements and since $\boldsymbol{C}^{j+1} \equiv \boldsymbol{C}^j \boldsymbol{C}$ the non-negativity of $\boldsymbol{C}$ implies that

$$\boldsymbol{S}_j \boldsymbol{C}(\boldsymbol{S}_1 + \boldsymbol{S}_2 + \cdots + \boldsymbol{S}_j + \boldsymbol{S}_{j+2} + \cdots + \boldsymbol{S}_p) = \boldsymbol{0}, \qquad j = 1, 2, \ldots, p-1.$$

Since, however, the first p matrices $\boldsymbol{S}_i$ form a complete set of projection matrices the above equation reduces to

$$\boldsymbol{S}_j \boldsymbol{C}(\boldsymbol{I} - \boldsymbol{S}_{j+1}) = \boldsymbol{0}, \qquad j = 1, 2, \ldots, p-1,$$

or

$$\boldsymbol{S}_j \boldsymbol{C} = \boldsymbol{S}_j \boldsymbol{C} \boldsymbol{S}_{j+1}, \qquad j = 1, 2, \ldots, p-1, \tag{3.1}$$

and similar considerations applied to the matrix $\boldsymbol{C}^{p+1} \equiv \boldsymbol{C}^p \boldsymbol{C}$ give

$$\boldsymbol{S}_p \boldsymbol{C} = \boldsymbol{S}_p \boldsymbol{C} \boldsymbol{S}_1.$$

Now from (2.1)—(2.3) and (3.1)

$$\boldsymbol{C}\boldsymbol{S}_{j+1} = \left(\sum_{i=1}^{p} \boldsymbol{S}_i\right) \boldsymbol{C}\boldsymbol{S}_{j+1} = \left(\sum_{i=1}^{p} \boldsymbol{S}_i \boldsymbol{C}\boldsymbol{S}_{i+1}\right) \boldsymbol{S}_{j+1} = \boldsymbol{S}_j \boldsymbol{C} \boldsymbol{S}_{j+1}, \qquad j = 1, 2, \ldots, p-1,$$

so that, from (3.1),

$$\boldsymbol{S}_j \boldsymbol{C} = \boldsymbol{C}\boldsymbol{S}_{j+1}, \qquad j = 1, 2, \ldots, p-1. \tag{3.2}$$

Similarly,

$$S_p C = C S_1$$

so that C is weakly cyclic of index p, proving the theorem.

The converse of Theorem 3 is elementary. That the first p powers of a weakly cyclic of index p matrix occupy different zones follows immediately from Definition 2, but they do not necessarily occupy all the space at their disposal. This may be seen by considering the n-th order matrix $B = [b_{ij}]$, where $b_{ij} = 1$ if $i = j+1$ or $j-1$ and is zero otherwise, for $n > 3$.

Lemma 1. If B is an irreducible non-negative $n \times n$ matrix then

$$\sum_{i=1}^{n} B^i > 0.$$

Proof. Since

$$\sum_{i=0}^{n-1} \binom{n-1}{i} B^i \equiv (I+B)^{n-1} > 0 \qquad ([5], \text{ p. } 26)$$

it follows that $\sum_{i=0}^{n-1} B^i > 0$, and multiplication of this last result by B yields the lemma.

Theorem 4. If $B = L + U$ is an irreducible non-negative $n \times n$ matrix and the eigenvalues of $B(\alpha) \equiv \alpha L + \alpha^{-(p-1)} U$ are independent of α then B is weakly cyclic of index p.

Proof. Assume that $\alpha > 0$ so that, from Lemma 1, $\sum_{i=1}^{n} [B(\alpha)]^i > 0$. Define C by

$$C = \sum_{i=0} [B(\alpha)]^{(pi+1)},$$

the sum being extended over all i for which $pi + 1 \leqq n$. It then follows that $\sum_{i=1}^{p} C^i > 0$ since $\sum_{i=1}^{p} C^i$ is equal to $\sum_{i=1}^{n} [B(\alpha)]^i$ plus non-negative terms. Moreover since $B(\alpha)$ is common with C the latter is irreducible, and it is clearly non-negative.

Consider now the powers of α occurring in the matrix C^i. Only those from the sequence $\{i - rp\}$, $r = 0, 1, 2, \ldots$, appear so that the powers of α present in C^i are distinct from those of C^j, $1 \leqq i < j \leqq p$. Similarly the powers of α found in C^{p+1} are distinct from those of C^i, $2 \leqq i \leqq p$. Now if the eigenvalues of $B(\alpha)$ are independent of α it follows from Theorem 2 that of the first $p+1$ powers of C and C, only C^{p+1} may have elements in common so that, from Theorem 3, C is weakly cyclic of index p. But since $B(\alpha)$ is common with C it also is weakly cyclic of index p, proving the theorem.

Theorem 5. If $B = L + U$ is an irreducible non-negative $n \times n$ matrix and the eigenvalues of $B(\alpha) \equiv \alpha L + \alpha^{-(p-1)} U$ are independent of α then $\exists$ a complete set of non-null diagonal projection matrices S_i, $i = 1, 2, \ldots, m$ such that $S_i L = L S_{i-1}$ and $S_i U = U S_{i+p-1}$.

Proof. Assume that $\alpha > 0$ and let $C = \sum_{i=1}^{n} [B(\alpha)]^i$. Then from Lemma 1 $C > 0$ and from Theorem 2 it follows that C is an α-matrix. Let $r_1 < r_2 < \cdots < r_m$, $m \leqq n$, be the powers of α that occur in the first row of C. Denote by R the set of integers

r_k, $1 \leqq k \leqq m$, some of which may be negative and which are not assumed to be consecutive. Let $\boldsymbol{T}_{r_k}$ be the diagonal projection matrix whose unity diagonal elements occur in the same columns as the elements in the first row of $\boldsymbol{C}$ whose powers of α are equal to r_k. Since $\boldsymbol{C} > \boldsymbol{0}$ the projection matrices $\boldsymbol{T}_r$, $r \in R$, form a complete set. Now the eigenvalues of $\boldsymbol{B}(\alpha)$ are independent of α so that, by Theorem 2, both $\boldsymbol{C}^2$ and $\boldsymbol{C} + \boldsymbol{C}^2$ must be α-matrices and this implies that the disposition of powers of α in the first row of $\boldsymbol{C}^2$ is identical with that of $\boldsymbol{C}$. Hence, since $\boldsymbol{C} > \boldsymbol{0}$, all the non-zero elements of the matrix $\boldsymbol{T}_i \boldsymbol{C} \boldsymbol{T}_j$, $i \in R$, $j \in R$, must have powers of α equal to $j - i$.

Define $\boldsymbol{C}_q$ to be the matrix $\boldsymbol{C}$ with elements involving powers of α other than the q-th replaced by zeroes. Then

$$\boldsymbol{C}_q = \sum_{j \in R} \boldsymbol{T}_j \boldsymbol{C} \boldsymbol{T}_{j+q}, \tag{3.3}$$

with the convention that if $k \notin R$, $\boldsymbol{T}_k$ is null. Thus, with this convention, premultiplication of (3.3) by $\boldsymbol{T}_i$ and postmultiplication plication by $\boldsymbol{T}_{i+q}$ gives

$$\boldsymbol{T}_i \boldsymbol{C}_q = \boldsymbol{T}_i \boldsymbol{C} \boldsymbol{T}_{i+q},$$

$$\boldsymbol{C}_q \boldsymbol{T}_{i+q} = \boldsymbol{T}_i \boldsymbol{C} \boldsymbol{T}_{i+q}$$

for any integer i so that

$$\boldsymbol{T}_i \boldsymbol{C}_q = \boldsymbol{C}_q \boldsymbol{T}_{i+q}. \tag{3.4}$$

Now $\boldsymbol{C} = \boldsymbol{B}(\alpha) +$ non-negative terms, and $\boldsymbol{B}(\alpha) \equiv \alpha \boldsymbol{L} + \alpha^{-(p-1)} \boldsymbol{U}$. Since $\boldsymbol{C}$ is an α-matrix this implies that $\boldsymbol{L}$ is common with $\boldsymbol{C}_1$ and $\boldsymbol{U}$ is common with $\boldsymbol{C}_{-(p-1)}$. Thus, from Eq. (3.4),

$$\boldsymbol{T}_i \boldsymbol{L} = \boldsymbol{L} \boldsymbol{T}_{i+1}, \tag{3.5}$$

$$\boldsymbol{T}_i \boldsymbol{U} = \boldsymbol{U} \boldsymbol{T}_{i-(p-1)}. \tag{3.6}$$

Suppose now that the integers $r_1, r_2, \ldots, r_m$ are not consecutive. Let $r_1, r_2, \ldots, r_k$ be consecutive but assume that $r_{k+1} \neq r_k + 1$. Denote by R_1 the set of integers r_i, $1 \leqq i \leqq k$ and R_2 the set of integers r_j, $k+1 \leqq j \leqq m$.

Since $r_k + 1 \notin R$ it follows that if $i \in R_1$ and $j \in R_2$ then $j > i + 1$. Thus, from (2.2), (3.5) and (3.6),

$$\Big(\sum_{i \in R_1} \boldsymbol{T}_i\Big) \boldsymbol{L} \Big(\sum_{j \in R_2} \boldsymbol{T}_j\Big) = \boldsymbol{0}$$

and

$$\Big(\sum_{i \in R_1} \boldsymbol{T}_i\Big) \boldsymbol{U} \Big(\sum_{j \in R_2} \boldsymbol{T}_j\Big) = \boldsymbol{0}.$$

Thus

$$\Big(\sum_{i \in R_1} \boldsymbol{T}_i\Big) \boldsymbol{B} \Big(\sum_{j \in R_2} \boldsymbol{T}_j\Big) = \boldsymbol{0}$$

but since

$$\sum_{i \in R_1 \cup R_2} \boldsymbol{T}_i = \boldsymbol{I}$$

this contradicts the hypothesis that $\boldsymbol{B}$ is irreducible (Definition 1). The integers $r_1, r_2, \ldots, r_m$ must therefore be consecutive so that if $\boldsymbol{S}_i$ is defined by

$$\boldsymbol{S}_i = \boldsymbol{T}_{r_{m+1-i}}, \quad 1 \leqq i \leqq m$$

the matrices $\boldsymbol{S}_i$, $1 \leqq i \leqq m$, form a complete set of non-null diagonal projection matrices which, from (3.5) and (3.6), satisfy

$$\boldsymbol{S}_i \boldsymbol{L} = \boldsymbol{L} \boldsymbol{S}_{i-1},$$
$$\boldsymbol{S}_i \boldsymbol{U} = \boldsymbol{U} \boldsymbol{S}_{i+p-1},$$

proving the theorem.

These final theorems establish the assertions that if $\boldsymbol{B}$ is irreducible and non-negative then Varga's first condition (V1) implies his second, and is equivalent to the revised definition. Since the revised definition with $q=r=1$ is equivalent to that of Young the latter is also equivalent to (V1) with $p=2$. The example in the previous section shows, however, that this equivalence only holds when $\boldsymbol{B}$ is irreducible and non-negative.

4. Determination of Consistent Ordering

The algorithms discussed here for determining whether or not a particular ordering is consistent are based on finding the diagonal projection matrices $\boldsymbol{S}_i$ of the revised definition.

Case 1. Symmetric Matrices

Let

$$\boldsymbol{B} = \boldsymbol{U}^T + \boldsymbol{U}$$

where $\boldsymbol{U} = [u_{ij}]$ is a strictly upper triangular $n \times n$ matrix. Denote by σ_i the set of integers such that $j \in \sigma_i$ if and only if $u_{ij} \neq 0$ and τ_j the set of integers such that $i \in \tau_j$ if and only if $u_{ij} \neq 0$. Thus $\tau_j = \bigcup_{j \in \sigma_i} i$ with a similar expression for σ_i. Let N denote the set of integers $1, 2, \ldots, n$.

To determine the projection matrices we construct a sequence of sets $\{S_i\}$ recursively using the following operations.

(i) Forward Sweep

$$S_{i+1} := \bigcup_{j \in S_i} \sigma_j, \qquad i = 1, 2, \ldots,$$

the computation being terminated either when the sets cease to be disjoint or when S_{i+1} becomes empty.

(ii) Backward Sweep

$$S_{i-1} := \bigcup_{j \in S_i} \tau_j, \qquad i = m, m-1, \ldots, 1,$$

where m is the largest value of i for which S_i is non-empty.

The computation is initialised by setting $S_1 = 1$, and then forward and backward sweeps are carried out alternately until either (a) the sets S_i cease to be disjoint or (b) the sequences $\{S_i\}$ obtained by two successive sweeps are identical. Since the sweeps are equivalent to determining sequences of projection matrices $\{\boldsymbol{T}_i\}$ such that

$$\boldsymbol{T}_i \boldsymbol{U} = \boldsymbol{T}_i \boldsymbol{U} \boldsymbol{T}_{i+1}, \qquad \text{(Forward)}$$
$$\boldsymbol{U} \boldsymbol{T}_{i+1} = \boldsymbol{T}_i \boldsymbol{U} \boldsymbol{T}_{i+1}, \qquad \text{(Backward)}$$

the occurrence of (a) indicates inconsistent ordering whereas that of (b) indicates consistent ordering. If $\boldsymbol{B}$ is irreducible and consistently ordered and $\{S_i\}$ is the final computed sequence then $\bigcup_i S_i = N$.

Case 2. Non-symmetric Matrices

In this case

$$B = L + U$$

and we proceed exactly as in the symmetric case to obtain from U a sequence of sets of integers $\{S_i^{(1)}\}$. Again if these are not disjoint B is inconsistently ordered but it is possible if B is consistently ordered that $\bigcup_i S_i^{(1)} \neq N$. If this occurs let k be the least positive integer for which $k \notin \bigcup_i S_i^{(1)}$ and $\sigma_k \neq 0$. If we define $S_1^{(2)} = k$ we can construct a second sequence of sets $\{S_i^{(2)}\}$ in the same way as $\{S_i^{(1)}\}$. Again if these are mutually disjoint B may be consistently ordered, although in any case

$$\left(\bigcup_i S_i^{(2)}\right) \cap \left(\bigcup_i S_i^{(1)}\right) = 0.$$

In this way we construct r sequences of disjoint sets such that if $1 \leqq k \leqq n$ then either

(a) $$k \in \bigcup_{j=1}^{r} \left(\bigcup_i S_i^{(j)}\right)$$

or

(b) $$\sigma_k = 0.$$

The set $S_i^{(1)}$ will correspond to the diagonal projection matrices $T_{1+(i-1)q}$ and the sets $S_i^{(j)}$ to the matrices $T_{g_j+(i-1)q}$, where q and the g_j are as yet undetermined.

Similarly, for the upper triangular matrix L^T, we construct a sequence of disjoint sets $\{P_i^{(j)}\}$, $1 \leqq j \leqq s$, where the set $P_i^{(1)}$ now corresponds to the diagonal projection matrix $Q_{1+(i-1)r}$. If the sets $S_{k+1}^{(1)}$ and $P_{m+1}^{(1)}$, $k \geqq 1$, $m \geqq 1$, are not disjoint then for the matrix to be consistently ordered according to the revised definition these sets must refer to the same projection matrix, so that $kq = mr$, and q and r are then the relatively prime integers that satisfy this equation. Once q and r have been established it is a simple matter to obtain the complete set of projection matrices and verify that they satisfy Eqs. (1.3), (1.4) and (2.1) to (2.3).

Example.

$$B = [b_{ij}] = \begin{bmatrix} 0 & 0 & X & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X & 0 & 0 \\ 0 & 0 & 0 & X & 0 & 0 & 0 \\ 0 & X & 0 & 0 & 0 & 0 & X \\ X & 0 & 0 & 0 & 0 & X & 0 \\ 0 & 0 & X & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X & 0 & 0 \end{bmatrix}$$

$$\{S_i^{(1)}\} = 1, 3, 4, 7;$$
$$\{S_i^{(2)}\} = 2, 5, 6;$$
$$\{P_i^{(1)}\} = 1, 5, 7;$$
$$\{P_i^{(2)}\} = 2, 4;$$
$$\{P_i^{(3)}\} = 3, 6.$$

Since $7 = S_4^{(1)}$ and $7 = P_3^{(1)}$, $k+1=4$ and $m+1=3$ so that $3q=2r$. Hence

$$q=2, \quad r=3.$$

Thus the sets $\{S_i^{(1)}\}$ and $\{P_i^{(1)}\}$ define the 7 projection matrices corresponding to the 7 sets

$$1;\ 0;\ 3;\ 5;\ 4;\ 0;\ 7;$$

and filling in this sequence from $\{S_i^{(2)}\}$ gives

$$1;\ 2;\ 3;\ 5;\ 4;\ 6;\ 7.$$

Since this does not contravene the requirements of $\{P_i^{(2)}\}$ and $\{P_i^{(3)}\}$ the matrix is consistently ordered according to the revised definition.

Acknowledgement. The author is glad to acknowledge his indebtedness to Mr. B. L. MEEK of Queen Elizabeth College, London, for reading the first draft of this paper and suggesting many improvements.

References

1. BROYDEN, C. G.: Some generalisations of the theory of successive over-relaxation. Num. Math. **6**, 269 (1964).
2. FORSYTHE, G. E., and W. R. WASOW: Finite-difference methods for partial differential equations. New York: Wiley 1960.
3. HOUSEHOLDER, A. S.: The theory of matrices in numerical analysis. London: Blaisdell 1964.
4. KJELLBERG, G.: On the successive over-relaxation method for cyclic operators. Num. Math. **3**, 87 (1961).
5. VARGA, R. S.: Matrix iterative analysis. London: Prentice-Hall 1962.
6. YOUNG, D.: Iterative methods for solving partial differential equations of elliptic type. Trans. Amer. Math. Soc. **76**, 92 (1954).

Dr. C. G. BROYDEN
Computing Centre
University of Essex
Wivenhoe Park
Colchester, Essex, Great Britain